

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

SOBRE DINÂMICA CAÓTICA E CONVERGÊNCIA EM ALGORITMOS DE EQUALIZAÇÃO AUTODIDATA

Autor

Romis Ribeiro de Faissol Attux

Orientador

Prof. Dr. João Marcos Travassos Romano

Banca Examinadora:

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

Prof. Dr. Max Gerken (EP/USP)

Prof. Dr. Fernando José Von Zuben (FEEC/UNICAMP)

Prof. Dr. João Batista Destro Filho (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, Abril de 2001

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

At87s Attux, Romis Ribeiro de Faissol
 Sobre dinâmica caótica e convergência em algoritmos
de equalização autodata / Romis Ribeiro de Faissol
Attux. --Campinas, SP: [s.n.], 2001.

 Orientador: João Marcos Travassos Romano
 Dissertação (mestrado) - Universidade Estadual de
Campinas, Faculdade de Engenharia Elétrica e de
Computação

 1. Filtros adaptativos. 2. Processamento de sinais –
Técnicas digitais. 3. Comportamento caótico nos
sistemas. I. Romano, João Marcos Travassos. II.
Universidade Estadual de Campinas. Faculdade de
Engenharia Elétrica e de Computação. III. Título.

Resumo

O objetivo deste trabalho é realizar uma análise dinâmica da expressão iterativa do algoritmo do módulo constante (CMA), uma das mais populares técnicas de equalização cega. Esta análise, realizada para as versões determinística e estocástica do algoritmo, mostra que o CMA pode apresentar, além dos comportamentos usualmente enfocados de convergência para um ponto fixo e divergência, um cenário dinâmico bastante rico, incluindo duplicação de período e caos. Tais características são verificadas com detalhe para diversas configurações de canal e equalizador lineares, o que propicia uma satisfatória abrangência das constatações com relação ao modelo do sistema. É feita uma análise teórica, fundamentada nos conceitos da teoria de sistemas dinâmicos não-lineares e em algumas idéias mais recentes, como a de saltos markovianos, de forma a se estabelecer um suporte sólido para a discussão dos resultados obtidos. Por fim é proposto um método, baseado no estudo realizado, que pode aumentar a convergência global de algoritmos de Bussgang em algumas situações. A aplicabilidade do método é investigada para o CMA e também para o algoritmo de decisão direta (DD), através de simulações envolvendo diferentes modelos de canal.

Abstract

The objective of this work is to perform a dynamical analysis of the constant modulus algorithm (CMA) iterative expression. The analysis shows that the CMA, in both deterministic and stochastic versions, can present, besides the usual “convergence to fixed point” and “divergence” types of behavior, a very rich dynamical scenario, which includes period-doubling and chaos. These features are verified in detail for many linear channel/equalizer configurations, what provides a satisfactory scope for the conclusions in terms of the system model.

A theoretical analysis, founded on the classical nonlinear dynamical systems theory and on more recent ideas, like markovian jumps, is carried out in order to establish a solid support to the discussion concerning the obtained results.

Finally, we propose a method, based on some conclusions taken from the study, that may improve Busgang algorithms global convergence in some situations. The applicability of the method is investigated for the CMA and also for the decision-directed (DD) algorithm, through simulations concerning different channel models.

O Reino de Deus não vem de modo visível,
nem se dirá: 'Aqui está ele', ou 'Lá está ele';
porque o Reino de Deus está em vocês.

Lucas 17, 20-21

Não rir, não lamentar, nem amaldiçoar,
mas compreender.

Baruch de Spinoza

Baste a quem baste o que lhe basta

O bastante de lhe bastar!

A vida é breve, a alma é vasta;

Ter é tardar.

Fernando Pessoa

A cabeça dos outros é um péssimo lugar
para ser a sede da verdadeira felicidade de um homem.

Arthur Schopenhauer

Thus let me live, unseen, unknown;

Thus unlamented let me die,

Steal from the world, and not a stone

Tell where I lie.

Alexander Pope

A Nilmar Lins Pimenta, *in memoriam*

Agradecimentos

Agradeço

A Deus.

Aos inúmeros cidadãos que financiaram esta pesquisa.

À CAPES, pelo suporte financeiro.

Ao Prof. João Marcos Romano, pelo incentivo, pela paciência, pela compreensão e pelas produtivas discussões.

A meus pais, Dina e Romis e à minha irmã, Cecília, pelo enorme apoio, em todos os sentidos.

A Dilmara Reischak, por seu carinho e incentivo.

A meus avós, Isabel e Afrânio, pelo suporte em todos os momentos, e às amigas Dedê e Verônica.

A Alice, Arthur, Leonardo e Rodrigo, pelo convívio fraterno.

A Aline Neves, pela paciente revisão, pelas valiosas contribuições e pelas enriquecedoras discussões sobre equalização cega.

A Cristiano Cruz, pelas sugestões e pelo incentivo.

A Cristiano Panazio, pelas inspiradoras conversas sobre convergência e pelo apoio técnico.

A Alexandre Osorio, Alim Pedro Gonçalves, Charles Cavalcante, Cynthia Junqueira, Danilo Zanatta Filho, Helder Bertan, Lais Macêdo, Maurício Sol e Moisés Ribeiro, pelo apoio técnico e pelas valiosas sugestões.

Ao Prof. Fernando Von Zuben, pela paciência e pela inestimável ajuda na fase de correção desta tese.

Ao Prof. João Batista Destro Filho, pelo companheirismo científico em todos os momentos deste trabalho.

Ao Prof. Max Gerken, por suas valiosas sugestões.

A Lúcia, Noêmia, Eloisa, Celi, Lena, Licinha, Gislaine, Giane, Mazé, Washington, Warley e a todos os funcionários e funcionárias da FEEC, pelo carinho e apoio imprescindíveis.

Ao Prof. Reginaldo Palazzo Jr., por sua amizade, por sua paciência e pela sua disposição em discutir e ensinar.

A Andrezza Gusmão, Flávia Gigli e Juliana Pedrosa, pelo incentivo constante.

A Alysson, Jane e Patricia, pela presença amiga na defesa desta tese.

A todos amigos do DSPCom, pela convivência sempre frutífera.

A minha família.

A todos amigos do Departamento de Comunicações (DECOM).

A todos amigos do Amor Exigente (AE) e da Associação Promocional Oração e Trabalho (APOT).

A todos amigos de Goiânia.

A todos amigos do Movimento Abrindo Portas (MAP) e dos demais projetos da Moradia.

A todos amigos da Paróquia Divino Salvador.

A todos amigos da Pastoral Universitária da Unicamp.

A todos amigos de Porto Alegre.

A todos amigos da Turma EE95.

A todos amigos.

A todos.

Aos professores Arvid Näess, Phillip Regalia e Marconi Madrid, pela atenção e abertura.

Lista de Abreviaturas

- AM: Modulação em Amplitude (*Amplitude Modulation*)
- AR: Auto-Regressivo
- ARMA: Auto-Regressivo / Média Móvel (*Moving Average*)
- AWGN: Ruído Aditivo, Branco e Gaussiano (*Additive White Gaussian Noise*)
- BER: Taxa de Erro de Bits (*Bit Error Rate*)
- CDMA: Múltiplo Acesso por Divisão em Código (*Code Division Multiple Access*)
- CM: Módulo Constante (*Constant Modulus*)
- CMA: Algoritmo do Módulo Constante (*Constant Modulus Algorithm*)
- CRU: Círculo de Raio Unitário
- DD: Decisão Direta
- DFE: Equalizador com Decisão Realimentada (*Decision Feedback Equalizer*)
- DSCI: Dependência Sensitiva das Condições Iniciais
- EOS: Estatísticas de Ordem Superior
- FIR: Resposta ao Impulso de Duração Finita (*Finite Impulse Response*)
- FM: Modulado em Frequência (*Frequency Modulated*)
- GSA: *Generalized Sato Algorithm*
- IIR: Resposta ao Impulso de Duração Infinita (*Infinite Impulse Response*)
- IIS: Interferência Intersimbólica
- LIT: Linear e Invariante no Tempo
- LMS: *Least Mean Square*
- MA: Média Móvel (*Moving-Average*)
- MAP: Máximo *a Posteriori*
- ML: Máxima Verossimilhança (*Maximum Likelihood*)
- MSE: Erro Quadrático Médio (*Mean Square Error*)
- OFDM: Multiplexação por Divisão Ortogonal em Frequência (*Orthogonal Frequency Division Multiplexing*)
- PAM: Modulação por Amplitude de Pulso (*Pulse Amplitude Modulation*)
- PSK: *Phase Shift Keying*

QAM: Modulação de Amplitude em Quadratura (*Quadrature Amplitude Modulation*)

RLS: *Recursive Least Squares*

SCA: *Standard Cumulant Algorithm*

SIR: Relação Sinal-Interferência (*Signal-to-Interference Ratio*)

SNR: Relação Sinal-Ruído (*Signal-to-Noise Ratio*)

SW: Shalvi-Weinstein

SWA: Algoritmo de Shalvi-Weinstein (*Shalvi-Weinstein Algorithm*)

ZF: *Zero-Forcing*

Índice

1. INTRODUÇÃO	1
1.1. ORGANIZAÇÃO DA TESE	4
2. UMA INTRODUÇÃO À TEORIA DOS SISTEMAS DINÂMICOS	5
2.1. SISTEMAS DINÂMICOS: DEFINIÇÃO	5
2.2. O ESPAÇO DE ESTADO	8
2.3. SISTEMAS CONTÍNUOS: CONDIÇÃO DE LIPSCHITZ	11
2.4. SISTEMAS CONTÍNUOS: SEÇÃO E MAPAS DE POINCARÉ	11
2.5. PONTOS DE EQUILÍBRIO E CICLOS-LIMITE EM SISTEMAS DINÂMICOS	12
2.5.1. DEFINIÇÕES DE ESTABILIDADE	14
2.5.2. ANÁLISE DA ESTABILIDADE DE UM PONTO FIXO DE UM SISTEMA CONTÍNUO	14
2.5.3. ANÁLISE DA ESTABILIDADE DE UM PONTO FIXO DE UM SISTEMA DISCRETO	17
2.5.4. CICLOS-LIMITE EM SISTEMAS DINÂMICOS	19
2.6. BIFURCAÇÕES	21
2.6.1. BIFURCAÇÃO SELA-NÓ	22
2.6.2. BIFURCAÇÃO DE HOPF	22
2.6.3. BIFURCAÇÃO FLIP	23
2.7. ALGUNS COMENTÁRIOS SOBRE SISTEMAS DINÂMICOS A SALTOS MARKOVIANOS	24
2.8. DISCUSSÃO	24
3. DINÂMICA CAÓTICA	27
3.1. O SISTEMA DE LORENZ E A DEPENDÊNCIA SENSITIVA DAS CONDIÇÕES INICIAIS	28
3.2. EXPOENTES DE LYAPUNOV	30
3.2.1. CÁLCULO DO EXPOENTE DE LYAPUNOV PARA UM MAPA UNIDIMENSIONAL	32
3.2.2. CÁLCULO DOS EXPOENTES DE LYAPUNOV PARA UM MAPA BIDIMENSIONAL	34
3.3. A ENTROPIA DE KOLMOGOROV-SINAI	34
3.4. ATRADORES ESTRANHOS	38
3.4.1. O MAPA DE HÉNON	42
3.4.2. DIMENSÃO FRACTAL, DIMENSÕES GENERALIZADAS E A CONJECTURA DE KAPLAN-YORKE	43
3.5. CARACTERIZAÇÃO DO CAOS DETERMINÍSTICO	47
3.6. O MAPA LOGÍSTICO	48
3.6.1. OBTENÇÃO DOS PONTOS DE EQUILÍBRIO E ANÁLISE DE ESTABILIDADE	50
3.6.2. OS CICLOS-LIMITE	51
3.6.3. CAOS, JANELAS DE PERIODICIDADE E CRISES	52
3.6.4. ANÁLISE POR DIAGRAMAS DE TEIA	53
3.6.5. UNIVERSALIDADE	58
3.7. DISCUSSÃO	61
4. FILTRAGEM ADAPTATIVA E EQUALIZAÇÃO SUPERVISIONADA	63

4.1.	INTRODUÇÃO	63
4.2.	MODELOS DE CANAIS	64
4.2.1.	CANAIS AUTO-REGRESSIVOS	67
4.2.2.	CANAIS MOVING-AVERAGE	68
4.2.3.	CANAIS ARMA	69
4.3.	EQUALIZADORES LINEARES	69
4.4.	FILTRAGEM DE WIENER	72
4.5.	O ALGORITMO LMS	79
4.5.1.	O MÉTODO STEEPEST DESCENT	79
4.5.2.	DERIVAÇÃO E PROPRIEDADES DO ALGORITMO LMS	84
4.5.3.	A ESCOLHA DO PASSO DE ADAPTAÇÃO	94
4.6.	OUTROS MÉTODOS BASEADOS NA TEORIA DE WIENER	98
4.7.	INTRODUÇÃO AOS MÉTODOS AUTODIDATAS	99
4.7.1.	TÉCNICAS MAIS UTILIZADAS	100

5. O ALGORITMO DO MÓDULO CONSTANTE **103**

5.1.	A PROPOSTA ORIGINAL DE GODARD	103
5.2.	TREICHLER E AGEE ENTRAM EM CENA	106
5.3.	ANÁLISE DE CONVERGÊNCIA: EQUALIZADORES DUPLAMENTE INFINITOS	107
5.4.	PRIMEIRAS EXTENSÕES PARA COMPRIMENTO FINITO	109
5.5.	CONSTATAÇÃO DA CONVERGÊNCIA PARA MÍNIMOS SUB-ÓTIMOS	109
5.5.1.	OBTENÇÃO DOS MÍNIMOS E ANÁLISE DE ESTABILIDADE	110
5.5.2.	EXTENSÕES DOS RESULTADOS	112
5.6.	SOBRE A POSIÇÃO DOS MÍNIMOS DE $J^{(2)}$	112
5.7.	EQUALIZADORES DE COMPRIMENTO FINITO REVISITADOS	115
5.8.	ESTABILIDADE DOS MÍNIMOS LOCAIS E COMPRIMENTO DO EQUALIZADOR	117
5.9.	EXTRAINDO INFORMAÇÃO DA SUPERFÍCIE DE ERRO	118
5.9.1.	A CLASSE DOS MÁXIMOS	121
5.9.2.	A CLASSE DE MAZO	121
5.9.3.	A CLASSE DE DING	122
5.9.4.	DISCUSSÃO	123
5.10.	EXTRAINDO INFORMAÇÃO DA SUPERFÍCIE DE ERRO II: SOBRE A LOCALIZAÇÃO DO MÍNIMO GLOBAL	124
5.11.	EXTRAINDO INFORMAÇÃO DA SUPERFÍCIE DE ERRO III: OS LIMITANTES DAS REGIÕES DE CONCAVIDADE E CONVEXIDADE	127
5.12.	INFLUÊNCIA DOS PONTOS DE SELA SOBRE A CONVERGÊNCIA	129
5.13.	BREVE DISCUSSÃO: AS PRIMEIRAS ABORDAGENS	130
5.14.	EFEITO DA KURTOSIS INICIAL NA CONVERGÊNCIA DO ALGORITMO E CONSIDERAÇÕES SOBRE A POSIÇÃO DOS MÍNIMOS	130
5.15.	MÍNIMOS DEPENDENTES DO COMPRIMENTO E DO CUSTO E CONSIDERAÇÕES SOBRE A PRESENÇA DE RUÍDO	135
5.16.	EQUALIZADORES SUBPARAMETRIZADOS	137
5.17.	A EQUIVALÊNCIA ENTRE OS CRITÉRIOS DE GODARD E DE SHALVI-WEINSTEIN	139
5.17.1.	A ABORDAGEM DE LI E DING	139
5.17.2.	A ABORDAGEM DE REGALIA	140
5.18.	RELAÇÃO ENTRE RECEPTORES BASEADOS NOS CRITÉRIOS CM E WIENER	144
5.19.	BREVE DISCUSSÃO: O PARADIGMA GEOMÉTRICO	148
5.20.	UMA ANÁLISE DINÂMICA DE EQUALIZADORES CEGOS	149
5.21.	BREVE DISCUSSÃO: ANÁLISE DINÂMICA	152

6. DINÂMICA CAÓTICA EM ALGORITMOS DO MÓDULO CONSTANTE **153**

6.1.	INTRODUÇÃO À ANÁLISE DINÂMICA	154
6.2.	OCORRÊNCIA DE CAOS NO ALGORITMO DETERMINÍSTICO	155
6.2.1.	ESTUDO DE UM CASO SIMPLES: UM CANAL DE ORDEM ZERO	155
6.2.2.	CANAL MA(1) DE FASE MÍNIMA	160
6.2.3.	CANAL MA(1) DE FASE MÁXIMA	170
6.2.4.	CANAL MA(2) DE FASE MISTA	176
6.2.5.	CANAL AR(1)	180
6.2.6.	CONSIDERAÇÕES SOBRE O CASO DETERMINÍSTICO	187
6.3.	ALGORITMO ESTOCÁSTICO: PRINCÍPIOS DE ANÁLISE	188
6.4.	OCORRÊNCIA DE CAOS NO ALGORITMO ESTOCÁSTICO	192
6.5.	CONSIDERAÇÕES SOBRE O CASO ESTOCÁSTICO	198

7. UMA PROPOSTA PARA AUMENTAR A TAXA DE CONVERGÊNCIA GLOBAL DO CMA **201**

7.1.	BASE TEÓRICA	201
7.1.1.	APLICAÇÃO AO ALGORITMO ESTOCÁSTICO	202
7.2.	O MÉTODO	207
7.3.	DIVISÃO EM FASES	209
7.3.1.	PRIMEIRA FASE	209
7.3.2.	SEGUNDA FASE	209
7.3.3.	TERCEIRA FASE	210
7.4.	RESULTADOS	210
7.4.1.	CONFIGURAÇÕES GERAIS	210
7.4.2.	CANAL MA(1) DE FASE MÍNIMA – ALGORITMO ESTOCÁSTICO	212
7.4.3.	CANAL MA(1) DE FASE MÁXIMA – ALGORITMO ESTOCÁSTICO	213
7.4.4.	CANAL MA(1) DE FASE MÍNIMA – ALGORITMO DETERMINÍSTICO	213
7.4.5.	CANAL MA(1) DE FASE MÁXIMA – ALGORITMO DETERMINÍSTICO	214
7.4.6.	CANAL MA(1) DE FASE MÍNIMA – ALGORITMO DD	215
7.4.7.	CANAL MA(1) DE FASE MÁXIMA – ALGORITMO DD	216

8. CONCLUSÃO **219**

9. BIBLIOGRAFIA **221**

10. ÍNDICE DE AUTORES **227**

APÊNDICE A: TRABALHOS REALIZADOS DURANTE O MESTRADO **231**

1. Introdução

A comunicação sempre foi um aspecto primordial da vida em sociedade. De fato, desde que começou a se agrupar em cidades e reinos, o homem tem buscado formas de comunicação que sejam adequadas a diversas atividades.

Com o advento da eletricidade e a consolidação da teoria eletromagnética por Maxwell e Hertz, estavam lançadas as bases de alguns inventos e técnicas que alargaram imensamente a possibilidade de comunicação. A invenção do telégrafo, do telefone e a primeira transmissão de informação por ondas de rádio influenciaram profundamente o mundo no século XX.

É certo que o avanço técnico das comunicações não significa que o homem de hoje se comunique mais com o seu semelhante, ou que a tecnologia conseguiu romper as barreiras que nos separam. Pensar a ciência e a tecnologia num contexto social e ético é fundamental para que o avanço do conhecimento e sua aplicação não se tornem instrumentos de opressão e desigualdade.

Embora tenham sido abertas muitas perspectivas, as mesmas trouxeram consigo um novo conjunto de problemas relacionado à real aplicabilidade das técnicas concebidas. Trabalhos como os de Wiener, Shannon e Nyquist foram fundamentais para que as concepções acerca da comunicação se concretizassem em sistemas aplicáveis às condições de “mundo real”.

O canal de comunicação caracteriza o meio pelo qual se propaga o sinal portador de informação, no trajeto entre o transmissor e o receptor. Este canal atua sobre o sinal transmitido, alterando suas características, possivelmente até um ponto em que haja uma corrupção significativa da mensagem enviada. Em determinados sistemas, é simplesmente inviável repassar diretamente ao destinatário a mensagem extraída do sinal recebido. É preciso, portanto, encontrar formas de reduzir os efeitos nocivos do canal até um patamar aceitável para a aplicação visada.

Uma das técnicas mais conhecidas e utilizadas para compensar a atuação do canal é a *equalização*. Equalizar significa operar sobre o sinal recebido de forma a deixá-lo o mais próximo possível do sinal original, ou seja, atuar sobre a mensagem de forma a cancelar os

efeitos do canal. O dispositivo que realiza esta tarefa é denominado *equalizador*, normalmente um filtro que tem como entrada o sinal recebido.

Pela definição acima, nota-se que há uma íntima ligação entre o equalizador e o canal. O filtro que faz o papel de equalizador deve ter seus parâmetros ajustados adequadamente de modo a cumprir sua tarefa. Se o canal for conhecido, há técnicas de projeto que podem ser utilizadas na confecção do dispositivo.

Em diversos sistemas, porém, não é possível conhecer o canal precisamente, por exemplo, devido a uma rápida variação de suas características. Neste caso, torna-se muito importante encontrar um método de ajuste de parâmetros *adaptativo*, ou seja, capaz de se adaptar a modificações sofridas pelo canal. Tal método precisa se basear em algum critério capaz de produzir uma função objetivo, cuja otimização leve a um ajuste adequado dos parâmetros do equalizador. O uso de um procedimento iterativo de otimização é o que proporciona a capacidade de rastreamento do equalizador.

Caso se disponha de amostras do sinal transmitido, pode-se procurar aproximar a saída do equalizador o mais possível do que se conhece acerca deste sinal. Esta idéia é a raiz da filtragem de Wiener, a pedra angular da teoria de filtragem adaptativa. A partir da solução de Wiener surgiram os algoritmos LMS e RLS, sendo o primeiro, principalmente, bastante utilizado na prática.

Porém, só tem sentido tal procedimento de equalização durante um período limitado, dito de treinamento. Também é bom ressaltar que introduzir sequências de treinamento na transmissão significa um dispêndio de espectro, o que nem sempre é conveniente do ponto de vista prático. Estas limitações levaram ao desenvolvimento de um outro paradigma: o da filtragem não-supervisionada, autodidata ou cega, em contraposição ao paradigma de Wiener, que dispunha da “supervisão” de um sinal “piloto”, desejado. Dentre os primeiros frutos deste paradigma temos os algoritmos da decisão-direta (DD), normalmente utilizado após uma fase de treinamento, e de Sato. No começo dos anos 80, porém, surgiu um algoritmo que é uma referência clássica dentro das técnicas autodidatas [Johnson et al., 1998], por ser bastante eficaz, de formulação simples e elegante e bastante aplicado na prática: o CMA (Algoritmo do Módulo Constante). Tal algoritmo, como os outros de sua classe, a das técnicas de Bussgang, utiliza uma estimativa não-linear para suprir a ausência de um sinal desejado, expediente este que acaba por levar a uma

expressão iterativa também não-linear. Esta característica, como se sabe da teoria de sistemas dinâmicos, pode levar a uma riqueza de cenários completamente estranha aos sistemas lineares (como os algoritmos derivados do filtro de Wiener).

Qual é, porém, o interesse que isto pode despertar? Ora, é sabido que um algoritmo deve ter a sua fundamentação teórica estabelecida nas bases mais sólidas possíveis. Caso contrário, será muito difícil para o projetista de um sistema de comunicações avaliar a real utilidade da técnica e a adequação dos parâmetros de projeto do equalizador adaptativo. Será difícil, ainda, propor melhorias que possam tornar o método ainda mais adequado a determinadas aplicações práticas sem um bom conhecimento acerca de suas limitações, equivalências em relação a outros métodos e peculiaridades.

A análise teórica das técnicas supervisionadas é algo já bastante bem estabelecido. Já para as técnicas autodidatas, ainda há grande abertura para estudos no âmbito teórico. O motivo para isto, como já afirmamos, é a presença de não-linearidades, que tornam a análise teórica bem mais difícil. Especificamente no caso do CMA, os anos 90 viram uma grande quantidade de trabalhos teóricos enfocando vários pontos relativos tanto à caracterização dos pontos de equilíbrio da expressão iterativa como à evolução dinâmica do estado do filtro. Estes trabalhos permitiram que se aprofundasse bastante o conhecimento acerca da técnica, o que também teve implicações práticas positivas (por exemplo, o desenvolvimento de procedimentos de inicialização como o “*center-spike*”).

Em todos estes trabalhos, porém, excetuando uma pequena menção em [Frater et al., 1995], a tônica de análise se restringiu ao binômio “convergência para um ponto fixo ou divergência”. Este binômio é suficiente no âmbito de um algoritmo linear, mas não é necessariamente verdadeiro quando se lida com um algoritmo não-linear.

O objetivo central deste trabalho é buscar caracterizar os possíveis comportamentos do algoritmo do módulo constante que transcendam tal binômio, procurando assim demonstrar que há uma complexidade inerente ao algoritmo que se manifesta quando da sua aplicação a vários modelos de canal. Iremos focar o CMA tanto em sua versão determinística quanto na estocástica (convencional), usando técnicas de análise distintas em ambos os casos. Desta forma, esperamos ter alcançado uma certa abrangência em relação ao domínio de nosso estudo, demonstrando que o CMA pode ter seu estado adaptado

segundo órbitas caóticas em diversas situações de interesse, justamente na região intermediária em relação aos dois cenários do binômio citado acima.

Desta análise teórica e do estudo de alguns artigos sobre otimização surgiram várias idéias que procuramos compilar na forma de um método para se aumentar a taxa de convergência global do CMA. Foram realizados alguns testes do mesmo em vários modelos de canal, sendo a aplicação também estendida ao algoritmo DD.

Para que as propostas deste trabalho pudessem estar razoavelmente situadas no panorama geral das análises de convergência, e ainda com o intuito de que se tivesse um texto que discutisse alguns dos principais trabalhos sobre a convergência do algoritmo, procuramos comentar os resultados obtidos pelos pesquisadores da área, na forma de um tutorial.

1.1. Organização da Tese

No **capítulo 2** é apresentada uma revisão de diversos conceitos sobre sistemas dinâmicos, para melhor situar o leitor que não teve contato com as ferramentas de análise usuais desta área.

No **capítulo 3** discutimos diversos tópicos versando sobre dinâmica caótica, para que possamos entender o que significa um sistema operar em caos.

No **capítulo 4** são expostos alguns conceitos-chave sobre equalização adaptativa, incluindo a clássica análise de convergência para o LMS.

No **capítulo 5** comentamos vários trabalhos versando sobre diversos aspectos estáticos e dinâmicos do CMA, buscando assim tecer um panorama acerca da evolução do conhecimento teórico sobre o método.

No **capítulo 6** expomos e comentamos os resultados principais do trabalho, a saber, a verificação e confirmação teórica da ocorrência de comportamento complexo no CMA, em suas versões determinística e estocástica.

No **capítulo 7** apresentamos os elementos de um método que pode aumentar, em alguns casos, a taxa de convergência global do algoritmo, sendo discutidos alguns resultados de simulação.

No **capítulo 8** estão colocadas as conclusões e considerações finais.

2. Uma Introdução à Teoria dos Sistemas Dinâmicos

Observa-se na natureza, de modo geral, uma incessante variação. Da agitação das moléculas à peregrinação dos astros, podemos afirmar que o Universo se afigura em constante mudança. Poderíamos, como alguns filósofos o fizeram, questionar a realidade de nossas observações. Mas, pelo menos no campo da Física, tornou-se imperativo introduzir em modelos aplicados a diversos fenômenos um formalismo adequado para expressar a variação de determinadas grandezas.

Podemos citar, como um grande exemplo desta necessidade, o desenvolvimento do Cálculo Diferencial por Leibniz e Newton, independentemente (e antes deles, em diversos aspectos, por Fermat [Boyer e Merzbach, 1991]). No caso de Newton, o Cálculo foi um dos pilares de sua proposta de Física, sendo que a própria força, em seu modelo, era definida como sendo a derivada do momento linear em relação ao tempo. A derivada, neste caso, nada mais é do que uma expressão da taxa de *variação* de uma dada grandeza física.

A matemática, portanto, começou a fornecer ferramentas adequadas às novas concepções do Universo, na esteira de uma verdadeira revolução científica. Era preciso lidar, cada vez mais, com modelos *dinâmicos* dos diversos sistemas estudados, fossem eles galáxias, circuitos, populações... Isto levou à formulação de uma teoria matemática bastante sólida, cujos resultados encontram aplicação nas mais diversas áreas e constituem um dos pilares da ciência moderna.

Neste capítulo, serão apresentados alguns conceitos e resultados desta teoria, os quais serão muito úteis em nossa análise posterior.

2.1. *Sistemas Dinâmicos: Definição*

Dado o modelo de um sistema dinâmico, podemos procurar dentre suas variáveis um conjunto especial, no sentido que o seu valor num certo instante contém toda a informação sobre o passado e o presente do sistema que pode afetar seu futuro. Em suma, conhecer o valor deste conjunto de variáveis é suficiente para se conhecer toda a história subsequente do sistema dinâmico [Von Zuben, 2001]. A este conjunto é dado o nome de

variáveis de estado [Haykin, 1994]. O conjunto destas variáveis é denominado *estado* do sistema.

Supondo uma variável t independente e contínua (pertencente ao conjunto dos reais), podemos escrever as variáveis de estado como $x_1(t), x_2(t), \dots, x_K(t)$. O valor de K é a ordem do sistema. Caso a variável independente seja discreta (pertencente ao conjunto dos inteiros), será denominada n , e as variáveis de estado serão $x_1(n), x_2(n), \dots, x_K(n)$. Podemos definir um vetor com todas as variáveis de estado, que será chamado o vetor de estado $\mathbf{x}(t)$ ou $\mathbf{x}(n)$, para os casos *contínuo* ou *discreto*, respectivamente.

Um sistema dinâmico é aquele no qual o estado muda com a evolução da variável independente. Quando esta variável é uma medida de tempo, é bastante natural utilizar o termo “dinâmico”.

Uma ampla classe destes sistemas pode ser modelada por um conjunto de equações de primeira ordem [Ruelle, 1980, Haykin, 1994]. As equações serão *diferenciais*, se o sistema for contínuo,

$$\begin{bmatrix} dx_1 / dt \\ dx_2 / dt \\ \dots \\ dx_k / dt \end{bmatrix} = \begin{bmatrix} F_1(x_1, \dots, x_k) \\ F_2(x_1, \dots, x_k) \\ \dots \\ F_k(x_1, \dots, x_k) \end{bmatrix} \quad (2.1)$$

ou *a diferenças*, se o sistema for discreto.

$$\begin{bmatrix} x_1(n+1) \\ x_2(n+1) \\ \dots \\ x_k(n+1) \end{bmatrix} = \begin{bmatrix} F_1[x_1(n), \dots, x_k(n)] \\ F_2[x_1(n), \dots, x_k(n)] \\ \dots \\ F_k[x_1(n), \dots, x_k(n)] \end{bmatrix} \quad (2.2)$$

As equações descritas em (2.1) e (2.2) são chamadas de *equações de estado* do sistema. O segundo membro destas equações define um campo vetorial $\mathbf{F}(\mathbf{x})$, que caracteriza o sistema dinâmico enfocado. Se algum dos componentes do campo vetorial tiver uma dependência não-linear com respeito a pelo menos uma das variáveis de estado, então o sistema dinâmico é dito *não-linear* [Haykin, 1994]. Caso contrário, é considerado

linear. Como os campos vetoriais $\mathbf{F}(\mathbf{x})$ de (2.1) e (2.2) não dependem de t ou n , o sistema é dito *autônomo*. Caso dependessem, seria *não-autônomo* [Fiedler-Ferrara e do Prado, 1994].

A análise de sistemas não-lineares é bem mais trabalhosa que a dos lineares. Em muitos casos, é preciso realizar uma linearização em torno de um certo ponto do estado do sistema, para se ter mais informação a seu respeito. Porém, é justamente a não-linearidade que dá origem a uma série de comportamentos dinâmicos complexos presentes nos mais variados sistemas. Portanto, buscar uma análise mais detalhada destes sistemas é fundamental para uma melhor compreensão de diversos fenômenos.

Um sistema linear autônomo, segundo a concepção expressa em (2.1) e (2.2), pode ser escrito como:

$$d\mathbf{x}(t)/dt = \mathbf{A} \cdot \mathbf{x}(t) \quad (2.3)$$

ou

$$\mathbf{x}(n+1) = \mathbf{B} \cdot \mathbf{x}(n) \quad (2.4)$$

onde $\mathbf{x}(t)$ é um vetor coluna com K elementos e \mathbf{A} e \mathbf{B} são matrizes $K \times K$. Usualmente, todos os entes são definidos em espaços reais.

Esta forma matricial é amplamente utilizada, pois permite uma notação compacta, além de muito proveitosa matematicamente, como iremos verificar mais adiante.

Podemos perceber que, uma vez obtida uma expressão para o campo vetorial $\mathbf{F}(\mathbf{x})$, o estado do sistema em qualquer instante de tempo, passado ou futuro, dada uma condição inicial $\mathbf{x}(0)$, fica perfeitamente determinado, caso sejam obedecidas as condições de existência e unicidade da solução, que discutiremos na seção 2.3.

É interessante notar que, caso o comportamento dos mais diversos sistemas do nosso Universo pudesse ser expresso através de equações como (2.1) e (2.2) e se tivesse um exato conhecimento da condição inicial, tudo estaria perfeitamente determinado. Este tipo de pensamento, porém, perdeu muita força com o advento da Física Quântica (que limita a precisão acerca do conhecimento de determinadas grandezas físicas) e da teoria de sistemas dinâmicos caóticos, que mostra que a inexatidão arbitrariamente pequena acerca do

conhecimento das condições pode levar a uma imprevisibilidade do estado do sistema a “longo prazo”. Além disso, há comportamentos dinâmicos que não estão sujeitos a uma descrição matemática precisa, com base no aparato de formalização hoje disponível.

2.2. O Espaço de Estado

Uma vez estabelecida a noção de estado de um sistema, podemos buscar maneiras de “vislumbrar” a sua evolução dinâmica, tanto para o caso linear como para o não-linear.

O conceito de *espaço de estado* é muito útil para esta finalidade. Para entendermos esta idéia, voltemos à definição de estado do sistema, um conjunto de variáveis que, em um certo instante, constituem a informação necessária e suficiente para se determinar os valores passados e futuros destas mesmas variáveis.

Sendo o sistema de ordem K , o vetor de estado terá K componentes. Podemos então imaginar este vetor como um ponto em um espaço K -dimensional. Este espaço é o chamado espaço de estado.

Tentemos perceber a relevância desta representação. Ora, dado o estado do sistema em um certo instante, vimos que ele corresponde a um ponto no espaço de estado. As características de um sistema dinâmico estão diretamente ligadas à maneira pela qual a posição deste ponto varia. No espaço de estado, esta variação provoca o surgimento de trajetórias “percorridas” pelo estado do sistema, indicando como se dá a sua variação temporal. Tomemos como exemplo um circuito RLC série autônomo. As equações de estado são:

$$di/dt = (-1/L).(Ri + v) \quad (2.5)$$

$$dv/dt = i/C \quad (2.6)$$

O vetor $[i(t) \ v(t)]^T$, para este caso, é o estado do sistema no instante t . Fazendo $R = L = C = 1$ e $v(0) = i(0) = 1$ (as unidades foram omitidas por simplicidade de notação), a trajetória percorrida no espaço de estado está apresentada na Fig. 2.1.

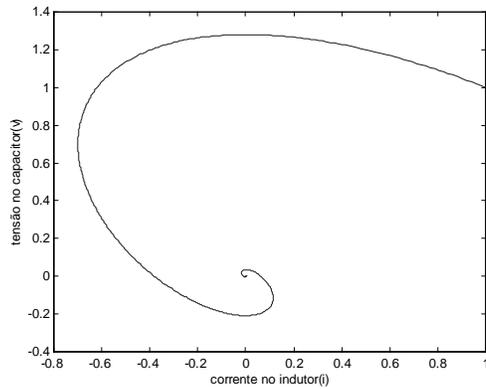


Figura 2.1: Trajetória no Espaço de Estado

Esta trajetória é uma espiral em direção ao ponto $v(\infty) = i(\infty) = 0$. Isto significa que o capacitor e o indutor tendem a “perder” toda a sua energia inicial, dissipada no resistor. A forma pela qual a trajetória converge também é uma informação sobre o tipo de ponto de equilíbrio do sistema, como discutiremos mais adiante.

O conjunto das trajetórias do estado para *todas* as possíveis condições iniciais é chamado *retrato de fase* do sistema [Haykin, 1994]. Embora a definição inclua a palavra *todas*, iremos usá-la mesmo quando tratar-se apenas de algumas condições iniciais representativas. Vejamos, para o nosso exemplo, o retrato de fase:

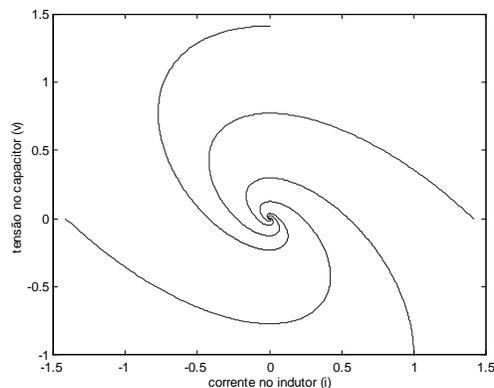


Figura 2.2: Retrato de Fase

Tal retrato mostra uma tendência do sistema em levar o estado para a origem, o que está ligado ao sentido físico citado, de dissipação de energia. O ponto $[0 \ 0]^T$ lembra um

“ralo”, atraindo as trajetórias, e sempre através de espirais. É bom notar que as trajetórias são sempre orientadas. No caso, esta orientação não foi explicitada nas figuras, pois o texto não dá margem a dúvidas.

Vejamos agora o caso de um sistema dinâmico discreto. No caso, foi escolhido o seguinte sistema linear:

$$x(n+1) = a.x(n) + b.y(n) \quad (2.7)$$

$$y(n+1) = c.x(n) + d.y(n) \quad (2.8)$$

onde $a = d = 0.9$, $b = 0.5$ e $c = -0.1$.

A Fig. 2.3 mostra o retrato de fase do sistema acima. Como era esperado, as trajetórias no espaço de estado de um sistema discreto não são traçados contínuos, mas sequências de pontos. Apesar desta diferença de natureza, é notável a semelhança de comportamento em relação ao exemplo anterior. Novamente, a origem funciona como um atrator de trajetórias.

Em síntese, trabalhar com o espaço de estado fornece um elo muito forte e útil entre a teoria de sistemas dinâmicos e a análise topológica e geométrica. Esta “interdisciplinaridade” foi e é fundamental para o avanço deste campo.

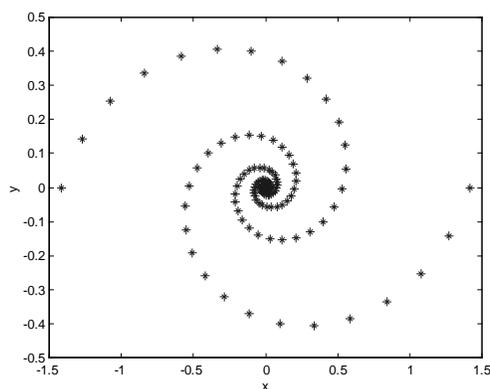


Figura 2.3: Retrato de Fase

2.3. *Sistemas Contínuos: Condição de Lipschitz*

Seja um sistema contínuo da forma de (2.1). Podemos imediatamente perguntar se existe uma solução para o mesmo, e caso a mesma exista, se é única. A resposta a estas perguntas está no campo vetorial $\mathbf{F}(\mathbf{x})$. Para que uma solução exista, é suficiente que $\mathbf{F}(\mathbf{x})$ seja contínuo em todos os seus argumentos. Porém, isto não garante a unicidade da solução. Para tanto, é preciso haver uma condição adicional, a condição de Lipschitz [Haykin, 1994], que expomos a seguir.

Sejam \mathbf{x} e \mathbf{u} dois vetores em um conjunto aberto W de um espaço de estado. De acordo com a condição de Lipschitz, existe uma constante G tal que:

$$\|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{u})\| \leq G \cdot \|\mathbf{x} - \mathbf{u}\| \quad (2.9)$$

para todo \mathbf{x} e \mathbf{u} em W .

Esta condição implica que $\mathbf{F}(\mathbf{x})$ é contínuo e que a solução para o sistema de equações é única.

2.4. *Sistemas Contínuos: Seção e Mapas de Poincaré*

Um sistema dinâmico discreto também pode ser denominado *mapa*. Em certas circunstâncias, pode ser muito interessante estudar algumas propriedades de um sistema dinâmico contínuo, não diretamente através de suas trajetórias no espaço de estado, mas através de sequências de pontos que, de alguma forma, forneçam informação acerca das mesmas. É esta idéia, de buscar um paradigma ao mesmo tempo simples e representativo de análise, que está por trás da seção e dos mapas de Poincaré [Parker e Chua, 1987, Fiedler-Ferrara e do Prado, 1994].

Imaginemos um sistema dinâmico de terceira ordem, por exemplo. As trajetórias correspondentes à evolução do estado estão em um espaço tridimensional. Pensemos agora num plano específico contido neste espaço. Cada vez que o estado cruzar o plano, temos um ponto no mesmo. Com o passar do tempo, devido a sucessivos cruzamentos, teremos marcado diversos pontos, correspondentes a várias intersecções. Podemos pensar na

sequência de pontos como um mapa bidimensional, que chamamos de *mapa de Poincaré*. Em nosso exemplo, o plano fez papel da *seção de Poincaré*.

Podemos resumir esta discussão em uma frase [Fiedler-Ferrara e do Prado, 1994]: a seção de Poincaré é uma ferramenta que permite que analisemos um sistema contínuo de ordem K por meio de um sistema discreto de ordem $K-1$. E o que é fundamental: este sistema discreto permitirá que se tenha uma idéia do tipo de comportamento apresentado pelo sistema contínuo K -dimensional.

2.5. *Pontos de Equilíbrio e Ciclos-Limite em Sistemas Dinâmicos*

Um dos aspectos mais relevantes quando se analisa os sistemas dinâmicos é a obtenção de seus *pontos de equilíbrio* ou *pontos fixos*.

Para termos uma idéia sobre tais pontos, imaginemos um pêndulo amortecido. Se o deslocarmos e soltarmos de diversas posições, e com diversas velocidades, o pêndulo vai perdendo energia cinética rumo a uma configuração vertical e estática. Uma vez atingida tal configuração, a menos de uma intervenção externa, não haverá mais modificação nem na posição nem na velocidade do pêndulo. Para o caso do circuito RLC série, visto na seção 2.2, ocorre um comportamento análogo: uma vez descarregados os bipolos armazenadores, não se observará mais nenhuma atividade no sistema, ou seja, os elementos permanecerão descarregados. Em ambos os casos, a conservação da energia torna naturais estas considerações: de onde poderia surgir energia num circuito sem fontes? O que poderia mover um pêndulo estático? Podemos afirmar que o sistema atingiu um equilíbrio, uma configuração especial para a qual não mais ocorre variação em seu estado.

Podemos definir, em termos intuitivos, um ponto de equilíbrio como um ponto de onde o estado jamais escapa, uma vez nele. Em termos matemáticos mais rigorosos, podemos escrever, para sistemas contínuos, que \mathbf{x}_e é um ponto de equilíbrio se, e somente se:

$$\mathbf{F}(\mathbf{x}_e) = \mathbf{0} \tag{2.10}$$

e para sistemas discretos:

$$\mathbf{F}(\mathbf{x}_e) = \mathbf{x}_e \quad (2.11)$$

Ambas as fórmulas são bastante razoáveis do ponto de vista da intuição. Olhando para (2.1), vemos que a condição (2.10) implica todas as derivadas nulas, o que nos dá uma idéia de invariância. Observando (2.2), percebemos que (2.11) nos mostra que a próxima iteração é sempre igual à anterior, o que também está em harmonia com a definição dada acima.

As condições (2.10) e (2.11) são sistemas de equações algébricas que, uma vez resolvidos, fornecem os pontos de equilíbrio do sistema dinâmico. Nem sempre é simples obter uma solução analítica para estes sistemas, sendo necessário muitas vezes recorrer a métodos numéricos.

Vejamos, a título de exemplo, como obter os pontos de equilíbrio dos sistemas da seção 2.2. A partir das equações (2.5) e (2.6), temos as condições:

$$(-1/L).(Ri_e + v_e) = 0 \quad (2.12)$$

$$i_e/C = 0 \quad (2.13)$$

que nos dão unicamente a solução trivial $i_e = v_e = 0$.

A partir de (2.7) e (2.8) temos:

$$x_e = a.x_e + b.y_e = (a-1).x_e + b.y_e = 0 \quad (2.14)$$

$$y_e = c.x_e + d.y_e = c.x_e + (d-1).y_e = 0 \quad (2.15)$$

Novamente, a única solução para o sistema acima é a trivial, $x_e = y_e = 0$. Em ambos os casos, já eram esperadas tais soluções, a partir do que constatamos anteriormente.

Porém, não basta obter os pontos de equilíbrio de um sistema, é preciso analisá-los quanto à sua estabilidade. Por exemplo, sabemos que um pêndulo poderia, idealmente,

permanecer parado numa posição invertida (“de pé”), mas nunca se teve notícia de algum caso em que esta configuração tenha durado mais que alguns instantes. De fato, trata-se de um ponto de equilíbrio (e o modelo do sistema nos garante isto), mas de um ponto de equilíbrio *instável*. Intuitivamente, sabemos o que é instabilidade...imaginemos que o pêndulo esteja na configuração descrita. Se uma pequena mosca pousar em um dos lados do peso, o mesmo irá pender e sair do equilíbrio. Com o passar do tempo, ele acabaria chegando à configuração de equilíbrio estável.

Com base nesta intuição, tentaremos definir *estabilidade* de um ponto fixo.

2.5.1. Definições de Estabilidade

Um ponto de equilíbrio, ou fixo, é dito *assintoticamente estável* [Fiedler-Ferrara e do Prado, 1994, Haykin, 1994] se a resposta do sistema a uma pequena perturbação (“em torno” do ponto em questão) tende ao ponto fixo quando o tempo tende a infinito. O ponto é então denominado *atrator* ou *sorvedouro*, por atrair trajetórias em sua vizinhança. Pensemos novamente no pêndulo: se o mesmo estiver em repouso e introduzirmos uma pequena perturbação, o mesmo tenderá a voltar ao repouso com o passar do tempo.

Um ponto fixo é dito *estável* (no sentido neutro ou de Lyapunov) se a resposta a uma pequena perturbação permanece pequena com o tempo tendendo a infinito. Caso a perturbação seja amplificada com o passar do tempo, o ponto é dito *instável*.

Quando o contexto não permitir interpretações dúbias, podemos utilizar apenas a denominação *estável* para qualificar um ponto de equilíbrio que rigorosamente deveria ser denominado *assintoticamente estável*.

2.5.2. Análise da Estabilidade de um Ponto Fixo de um Sistema Contínuo

Um sistema contínuo linear pode ser escrito como em (2.3). Em tal caso, a solução do sistema é:

$$\mathbf{x}(t) = \mathbf{x}_0 \cdot \exp(\mathbf{A} \cdot t) \quad (2.16)$$

onde \mathbf{x}_0 é um vetor com K elementos que contém a informação sobre a condição inicial do sistema.

É possível mostrar, a partir de (2.16), que há uma estreita dependência entre $\mathbf{x}(t)$ e os autovalores da matriz \mathbf{A} , que se evidencia, por exemplo, no fato que são tais entes matemáticos que determinam a estabilidade do ponto de equilíbrio (no caso linear, só temos um possível ponto). Para que o ponto seja assintoticamente estável, é preciso que todos os autovalores tenham suas partes reais negativas. Se pelo menos um dos autovalores tiver sua parte real maior que zero, o sistema não será estável sob nenhuma definição [Parker e Chua, 1987].

Tomando como exemplo um sistema de ordem dois, podemos produzir a seguinte tabela, encontrada em [Fiedler-Ferrara e do Prado, 1994]:

Autovalores	Sinal da Parte Real	Classificação	Estabilidade
$\lambda_1 \neq \lambda_2, \lambda_1, \lambda_2 \neq 0$ λ_1, λ_2 real	$\lambda_1, \lambda_2 > 0$ $\lambda_1, \lambda_2 < 0$	Nó (hiperbólico)	Assint. Estável
$\lambda_1 \neq \lambda_2, \lambda_1, \lambda_2 \neq 0$ λ_1, λ_2 real	$\lambda_1, \lambda_2 > 0$ $\lambda_1, \lambda_2 > 0$	Nó (hiperbólico)	Instável
$\lambda_1 \neq \lambda_2, \lambda_1, \lambda_2 \neq 0$ λ_1, λ_2 real	$\lambda_1, \lambda_2 < 0$	Sela (hiperbólica)	Instável
$\lambda_1 = \lambda_2^*$ λ_1, λ_2 complexos	Positivo	Foco (hiperbólico)	Instável
$\lambda_1 = \lambda_2^*$ λ_1, λ_2 complexos	Negativo	Foco (hiperbólico)	Assint. Estável
$\lambda_1 = \lambda_2^*$ λ_1, λ_2 complexos	Nulo	Centro (elíptico)	Estável (mas não assintoticamente)
$\lambda_1, \lambda_2 = 0$	Nulo	Caso Degenerado	
$\lambda_1 = \lambda_2$	Positivo	<i>Inflected Node</i> (hiperbólico)	Instável
$\lambda_1 = \lambda_2$	Negativo	<i>Inflected Node</i> (hiperbólico)	Assint. Estável

Tabela 2.1: Classificação dos Pontos de Equilíbrio para Sistemas Contínuos 2-D

Voltemos ao exemplo do circuito RLC série, focado anteriormente na seção 2.2. A matriz \mathbf{A} do sistema é:

$$\mathbf{A} = \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix} \quad (2.17)$$

Os autovalores desta matriz são $\frac{-1 + \sqrt{3}.j}{2}$ e $\frac{-1 - \sqrt{3}.j}{2}$. Portanto, segundo a Tab. 2.1, a origem é um foco estável, como já intuíamos a partir dos resultados anteriores.

A análise da estabilidade de um ponto de equilíbrio pode ser estendida a sistemas não-lineares, em caráter local, através da linearização em torno de um ponto de equilíbrio. Vejamos como isto pode ser feito. Escrevamos:

$$\mathbf{x}(t) = \mathbf{x}_e + \Delta\mathbf{x}(t) \quad (2.18)$$

Usando uma aproximação de primeira ordem, através da série de Taylor, obtemos:

$$\mathbf{F}(\mathbf{x}) \approx \mathbf{F}(\mathbf{x}_e) + \mathbf{A} \cdot \Delta\mathbf{x}(t) \quad (2.19)$$

Onde \mathbf{A} , neste caso, é a *matriz jacobiana* do sistema calculada em \mathbf{x}_e . A definição da matriz jacobiana, que genericamente chamaremos de \mathbf{J} é:

$$\mathbf{J} = \begin{bmatrix} \partial F_1 / \partial x_1 & \dots & \dots & \partial F_1 / \partial x_k \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \partial F_k / \partial x_1 & \dots & \dots & \partial F_k / \partial x_k \end{bmatrix} \quad (2.20)$$

Sabendo que $\mathbf{F}(\mathbf{x}_e) = \mathbf{0}$, e usando (2.18) e (2.19), temos:

$$d\Delta\mathbf{x}(t)/dt \approx \mathbf{A} \cdot \Delta\mathbf{x}(t) \quad (2.21)$$

O sistema pode então ser analisado com o mesmo procedimento do caso linear, observando-se os autovalores da matriz **A**. Porém, quando o método indica que o ponto é um *centro*, não quer dizer que de fato o seja. A isto chamamos *problema dos falsos centros* [Fiedler-Ferrara e do Prado, 1994]. Para os outros casos, não há ambiguidade alguma.

É importante atentar para o fato de que, como fizemos uso de uma aproximação em torno do ponto de equilíbrio analisado, iremos ter informações *locais* sobre sua estabilidade, para o caso de sistemas não-lineares.

2.5.3. Análise da Estabilidade de um Ponto Fixo de um Sistema Discreto

Um sistema discreto linear pode ser escrito como em (2.4). A solução de tal sistema tem por componentes funções do tipo:

$$\mathbf{x}(n) = \mathbf{x}_0 \cdot \lambda^n \quad (2.22)$$

onde λ representa um autovalor genérico da matriz **B**. Novamente, os autovalores determinam a estabilidade do ponto de equilíbrio do sistema linear. Caso todos os autovalores estejam dentro do círculo de raio unitário (CRU), o ponto será assintoticamente estável. Caso um dos autovalores esteja fora do CRU, o ponto não será estável em nenhum sentido.

Novamente, tomando um sistema de segunda ordem como modelo, vejamos na Tabela 2.2 a classificação dos pontos de equilíbrio, conforme exposta em [Fiedler-Ferrara e do Prado, 1994].

Autovalores	Módulo	Classificação	Estabilidade
λ_1, λ_2 reais	$ \lambda_1 $ e $ \lambda_2 < 1$	Atrator	Assint. Estável
λ_1, λ_2 reais	$ \lambda_1 $ e $ \lambda_2 > 1$	Repulsor	Instável
λ_1, λ_2 reais	$ \lambda_1 > 1$ e $ \lambda_2 < 1$ ou $ \lambda_2 > 1$ e $ \lambda_1 < 1$	Ponto de Sela	Instável
$\lambda_1 = \lambda_2^*$ λ_1, λ_2 complexos	$ \lambda_1 , \lambda_2 > 1$	Foco	Instável
$\lambda_1 = \lambda_2^*$ λ_1, λ_2 complexos	$ \lambda_1 , \lambda_2 < 1$	Foco	Assint. Estável
$\lambda_1 = \lambda_2^*$ λ_1, λ_2 complexos	$ \lambda_1 , \lambda_2 = 1$	Centro	Estável (mas não assintoticamente)

Tabela 2.2: Classificação dos Pontos de Equilíbrio para Sistemas Discretos 2-D

Analisemos o sistema discreto que nos serviu de exemplo na seção 2.2. A matriz **B** (2.4) pode ser escrita como:

$$\mathbf{B} = \begin{bmatrix} 0.9 & 0.5 \\ -0.1 & 0.9 \end{bmatrix} \quad (2.23)$$

Os autovalores são $\frac{18 + \sqrt{20}.j}{20}$ e $\frac{18 - \sqrt{20}.j}{20}$. O módulo de ambos é $\frac{\sqrt{86}}{10} \approx 0.9274$.

A partir da Tab. 2.2, podemos perceber que a origem é um foco estável.

Para estendermos esta análise a sistemas não-lineares, novamente podemos utilizar uma linearização em torno do ponto que se deseja analisar. Para tanto, sejam:

$$\mathbf{x}(n) = \mathbf{x}_e + \Delta \mathbf{x}(n) \quad (2.24)$$

e

$$\mathbf{x}(n+1) = \mathbf{x}_e + \Delta \mathbf{x}(n+1) \quad (2.25)$$

Temos que:

$$\mathbf{x}(n+1) = \mathbf{F}[\mathbf{x}_e + \Delta \mathbf{x}(n)] \approx \mathbf{F}(\mathbf{x}_e) + \mathbf{M} \cdot \Delta \mathbf{x}(n) \quad (2.26)$$

Lembrando que $F(\mathbf{x}_e) = \mathbf{x}_e$ e usando (2.26), obtemos:

$$\Delta \mathbf{x}(n+1) \approx \mathbf{M} \cdot \Delta \mathbf{x}(n) \quad (2.27)$$

Nesta expressão, \mathbf{M} é a matriz jacobiana calculada no ponto de equilíbrio. Tal matriz também é chamada de *matriz de Floquet*. Seus autovalores são chamados de *multiplicadores de Floquet*.

2.5.4. Ciclos-Limite em Sistemas Dinâmicos

Na seção anterior discutimos a convergência para pontos de equilíbrio em sistemas dinâmicos. Ressaltamos, porém, que este tipo de comportamento é apenas um dos possíveis em tais sistemas.

Uma outra possibilidade são soluções periódicas, denominadas *ciclos-limite*. Estes tipos de oscilação estão presentes em uma ampla gama de fenômenos físicos, de sistemas neurais a circuitos elétricos.

É interessante notar que, até agora, só havíamos falado de atratores pontuais. A trajetória descrita em um ciclo-limite é também um atrator, embora de dimensão não-nula (no caso contínuo). Trata-se de um atrator no qual o estado não permanece constante, mas varia de forma periódica. Tal comportamento só é encontrado em sistemas não-lineares.

Vejamos um exemplo que pode nos ajudar a entender melhor o conceito discutido acima. Seja o sistema contínuo e não-linear de segunda ordem:

$$dx/dt = y \quad (2.28)$$

$$dy/dt = -\lambda \cdot (1 - x^2) \cdot y - x \quad (2.29)$$

Foi feito $\lambda = 0.5$, e traçadas algumas trajetórias, para diferentes condições iniciais. O retrato de fase obtido está na Fig. 2.4 (novamente, foi omitida a orientação das trajetórias, sendo que todas convergem para o ciclo).

Podemos notar, pela figura, que a estabilidade do ciclo limite se processa através de um movimento “para fora” de configurações de estado com pequena amplitude e de um movimento oposto para configurações com grande amplitude. Este comportamento produz uma região de fronteira estável que caracteriza o movimento periódico. Como consequência, sempre que houver um único ciclo limite estável envolvendo a origem, a mesma será necessariamente instável.

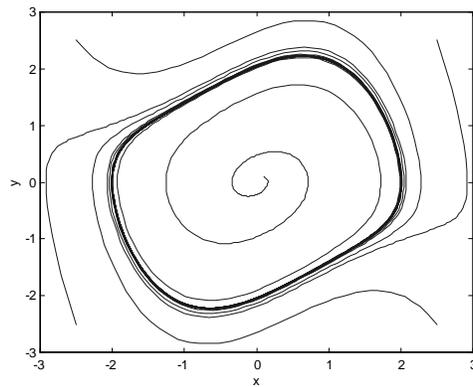


Figura 2.4: Retrato de Fase para um Ciclo-Limite Estável

Ciclos-limite podem ainda ocorrer em sistemas discretos, o que iremos verificar através de um exemplo. Seja o seguinte sistema discreto unidimensional e não-linear:

$$x(n+1) = \mu \cdot x(n) \cdot [1 - x(n)] \quad (2.30)$$

Tal sistema é denominado *mapa logístico* (ver seção 3.6). Ele apresenta um comportamento dinâmico muito rico, com várias possibilidades dinâmicas, de acordo com o valor de μ . Usaremos, no exemplo, o valor $\mu = 3.1$. Para este valor, temos as seguintes 50 primeiras iterações:

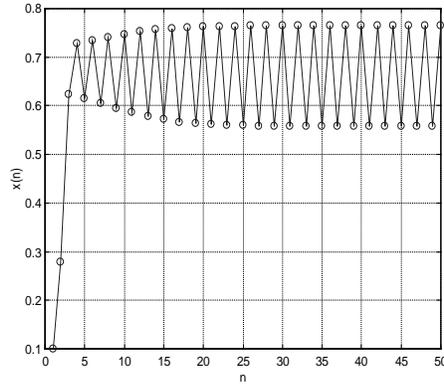


Figura 2.5: Ciclo-Limite do Mapa Logístico para $\mu = 3.1$

Em regime permanente, o estado oscila entre dois valores: 0.7646 e 0.5583. Estes dois valores constituem um atrator, um ciclo-limite para o qual converge o estado do sistema após cessarem os transitórios.

2.6. Bifurcações

Certamente, um dos conceitos centrais da teoria de sistemas dinâmicos é o de *bifurcação*. Intuitivamente, bifurcação é uma mudança qualitativa nas características de um sistema causada por uma pequena variação quantitativa em seus parâmetros [Guckenheimer e Holmes, 1983, Fiedler-Ferrara e do Prado, 1994]. Por exemplo, há uma bifurcação quando um sistema que só apresentava pontos de equilíbrio estáveis passa, graças a uma variação em seus parâmetros, a ter comportamento periódico devido ao surgimento de um ciclo-limite estável. Este novo ente espelha uma mudança muito destacada no retrato de fase do sistema.

Em resumo, e procurando adotar uma terminologia mais formal, ocorre uma bifurcação quando a mudança de um ou mais parâmetros de um sistema produz uma alteração significativa nas características topológicas de seu fluxo.

Matematicamente, seja $\mathbf{F}(\mathbf{x})$ o campo vetorial associado a um certo sistema dinâmico. Se uma pequena modificação neste campo levar a uma mudança qualitativa na topologia do fluxo no espaço de estados, falamos de uma *bifurcação*. A mudança no campo

$\mathbf{F}(\mathbf{x})$ usualmente se dá através da modificação do valor dos parâmetros livres, ou *parâmetros de bifurcação*.

Há diversos tipos de bifurcação, classificados de acordo com o tipo de modificação ocorrida na dinâmica do sistema. Analisemos alguns deles, que serão mais importantes para o desenvolvimento posterior deste trabalho.

2.6.1. Bifurcação Sela-Nó

Uma bifurcação sela-nó [Guckenheimer e Holmes, 1983, Fiedler-Ferrara e do Prado, 1994], como o próprio nome indica, ocorre quando a variação das características do sistema leva ao surgimento de um par de pontos de equilíbrio: um ponto de sela e um nó estável. Para que possamos perceber como isto se dá, analisemos um exemplo simples. Consideremos o seguinte sistema:

$$dx/dt = \mu - x^2 \quad (2.31)$$

Aplicando (2.10), obtemos os pontos de equilíbrio $x_{e1} = \sqrt{\mu}$ e $x_{e2} = -\sqrt{\mu}$. A matriz jacobiana, no caso um escalar, vale respectivamente $-2\sqrt{\mu}$ e $2\sqrt{\mu}$.

Ora, se $\mu < 0$, x_{e1} e x_{e2} assumirão valores complexos. Caso o estado seja inicializado em um valor real e μ também pertença a este domínio, tais pontos não terão influência alguma sobre a evolução dinâmica do sistema. Portanto, $\mu = 0$ é o valor para o qual os pontos de equilíbrio passam a, de fato, existir no espaço de estado real do sistema.

Para $\mu > 0$, a matriz jacobiana nos mostra que o primeiro ponto (nó) é estável e que o segundo (sela) é instável. Considerando um aumento progressivo de μ a partir de um valor inicial negativo, podemos afirmar que em $\mu = 0$ ocorre uma bifurcação sela-nó.

2.6.2. Bifurcação de Hopf

No primeiro tipo de bifurcação estudado, só havia pontos de equilíbrio envolvidos. Na *bifurcação de Hopf*, um ponto de equilíbrio estável dá lugar a um ciclo-limite estável, com a variação do campo vetorial $\mathbf{F}(\mathbf{x})$.

Um exemplo desta bifurcação pode ser visto através do seguinte sistema:

$$\frac{dx}{dt} = -y + x \cdot (\mu - x^2 - y^2) \quad (2.32)$$

$$\frac{dy}{dt} = x + y \cdot (\mu - x^2 - y^2)$$

Demonstra-se que em $\mu = 0$ há uma bifurcação de Hopf. Observa-se que a matriz jacobiana calculada no ponto de bifurcação apresenta dois autovalores puramente imaginários, o que é característico desta bifurcação [Fiedler-Ferrara e do Prado, 1994].

Vejamos algumas trajetórias (novamente a orientação é omitida, estando subentendida a atração) para $\mu = -1$ e $\mu = 1$:

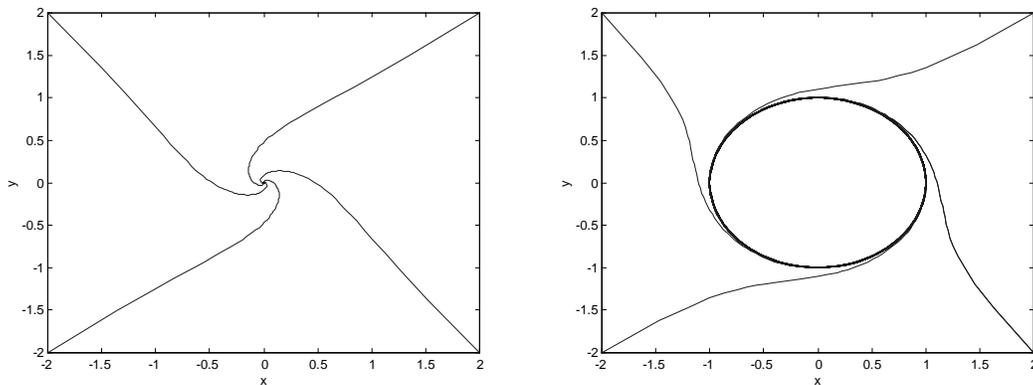


Figura 2.6: Bifurcação de Hopf

O que observamos na Fig. 2.6 clarifica a idéia básica deste tipo de bifurcação: uma transição entre atratores estáveis distintos (do pontual para o cíclico).

2.6.3. Bifurcação Flip

Na *bifurcação flip*, um ciclo limite estável perde sua estabilidade e surge uma nova órbita com o dobro do período. Este é o motivo pelo qual esta bifurcação também é denominada *bifurcação de duplicação de período*. Caso ocorram sucessivamente n destas

bifurcações, a partir de um ciclo-limite original de período T surgirá um ciclo limite com período $2^n.T$.

Para que a bifurcação flip ocorra em sistemas contínuos, são necessárias 3 dimensões, no mínimo. Para sistemas discretos não há restrições quanto à ordem do sistema para a ocorrência de tais fenômenos.

2.7. *Alguns Comentários Sobre Sistemas Dinâmicos a Saltos Markovianos*

Até agora, vínhamos estudando sistemas dinâmicos determinísticos, cuja evolução era descrita de forma inequívoca por um conjunto fixo de equações diferenciais ou a diferenças. No entanto, este tipo de modelo pode não ser razoável em circunstâncias onde ocorrem eventos cuja descrição matemática se encaixa bem no contexto da teoria de probabilidade e processos estocásticos.

Em tal classe se encaixam os modelos de sistemas a saltos markovianos, que têm despertado crescente atenção da comunidade científica [Gonçalves et al., 2000]. A idéia destes sistemas pode ser resumida da seguinte forma: o sistema que se deseja estudar é modelado como um conjunto de sistemas determinísticos associados a estados numa cadeia de Markov [Palazzo, 1999]. A transição entre estados liga-se a uma mudança de comportamento dinâmico, sendo que o sistema determinístico associado ao estado presente sempre “rege” a evolução do estado.

Esta idéia será por nós retomada, no contexto da análise do CMA, na seção 6.3.

2.8. *Discussão*

Neste capítulo, procuramos expor e discutir alguns conceitos fundamentais da teoria de sistemas dinâmicos. O objetivo desta discussão foi começar a preparar o cenário para a análise posterior do CMA, a qual empregará muitos dos conceitos apresentados. Basicamente, procuramos fundamentar as idéias de sistema dinâmico e de estado, de importância capital para todo o desenvolvimento subsequente. A seguir, foi exposto o conceito de espaço de estado, que fornece um elo entre a teoria de sistemas dinâmicos e a geometria. Passamos, após mais algumas idéias, à definição e caracterização dos pontos de

equilíbrio de um sistema. A convergência para tais pontos constitui uma classe muito importante das possibilidades dinâmicas do sistema, sendo que, em análise de algoritmos adaptativos, é certamente a classe mais relevante do ponto de vista prático.

A seguir, foi introduzido um novo atrator: o ciclo-limite, que ocorre apenas em sistemas não-lineares. Trata-se, como iremos mostrar, de um atrator presente na dinâmica do CMA em todos os casos analisados. Analisamos então a idéia de bifurcação, estudando, dentre todos os possíveis tipos, aqueles que aplicaremos diretamente neste trabalho e, por fim, tecemos alguns comentários sobre sistemas dinâmicos a saltos markovianos.

3. Dinâmica Caótica

A palavra *caos* tem o sentido usual de desordem e irregularidade. A palavra grega original é $\chi\alpha\omicron\zeta$, que se refere ao vazio que havia antes da criação de todas as coisas [Schuster, 1988].

No contexto de sistemas dinâmicos, a idéia de caos já aparece nos trabalhos de Henri Poincaré, um eminente físico-matemático francês e James Clerk Maxwell [Hunt e Yorke, 1993], um dos pais da moderna teoria eletromagnética. Basicamente, os dois perceberam que determinados sistemas dinâmicos, em algumas condições, podiam evoluir de forma sensivelmente distinta mesmo a partir de condições iniciais muito próximas. Este tipo de comportamento não despertou grande interesse nos pesquisadores da época, e ficou vários anos sem ser adequadamente estudado.

Apenas em 1963, um meteorologista chamado Edward Lorenz descobriu que um sistema não-linear de três equações diferenciais acopladas apresentava, para alguns valores de um parâmetro de controle, um comportamento pouco usual. Seu trabalho não teve um grande impacto imediatamente, mas logo ajudou a desencadear uma verdadeira revolução científica, na qual se percebeu que esta nova faceta dos sistemas dinâmicos não era uma mera curiosidade, mas um tipo de fenômeno comum aos mais diversos contextos.

O fato curioso sobre os sistemas de Poincaré, Maxwell e Lorenz é que embora sendo determinísticos e possuindo um número finito de graus de liberdade, apresentavam *comportamento caótico, com dependência sensitiva das condições iniciais (DSCI)*.

A origem da “descoberta” de Lorenz é bastante emblemática, bem como o seu famoso “efeito borboleta”. Iremos contar esta história de forma resumida, como é relatada em [Gleick, 1987]. O intuito de Lorenz era criar um modelo atmosférico simplificado. Simulando as equações de seu modelo em um computador digital, ele percebeu, por acaso, que pequenas mudanças nas condições iniciais logo evoluíam para estados subsequentes completamente distintos. A conclusão que se pode tirar disto é que a previsão do tempo através destes modelos não nos dá grandes garantias de segurança, sendo a longo prazo inviável. O resumo desta história está no enunciado do efeito borboleta (nome derivado do título de uma palestra proferida por Lorenz): “Uma borboleta batendo as asas no Brasil pode causar um tornado no Texas?”.

Para buscarmos compreender o conceito de DSCI, analisemos o próprio sistema de Lorenz.

3.1. *O Sistema de Lorenz e a Dependência Sensitiva das Condições Iniciais*

O cenário do modelo de Lorenz é o seguinte [Fiedler-Ferrara e do Prado, 1994]: um fluido entre duas placas horizontais, entre as quais há uma diferença de temperatura ΔT (a placa superior está a uma maior temperatura). O valor de ΔT tem grande relevância para a determinação do comportamento físico do sistema. Para ΔT pequeno, há condução de calor, e com o aumento da diferença, começam a aparecer rolos de convecção. Trabalhando com as equações de Navier-Stokes, da condução de calor e da continuidade, e empregando condições de contorno e simplificações adequadas, Lorenz obteve um sistema não-linear de três equações diferenciais acopladas:

$$dx/dt = -\sigma \cdot (x - y) \quad (3.1)$$

$$dy/dt = r \cdot x - y - x \cdot z \quad (3.2)$$

$$dz/dt = x \cdot y - b \cdot z \quad (3.3)$$

Neste sistema, $x(t)$ é proporcional à velocidade circulatória do fluido, $y(t)$ é proporcional à diferença de velocidade entre seus elementos ascendentes e descendentes e $z(t)$ é proporcional aos desvios do perfil de temperatura vertical em relação a um perfil linear.

Supondo $\sigma = 10$ e $b = 8/3$, deixemos r (o número de Rayleigh relativo) ser nosso parâmetro livre (de bifurcação). Para valores pequenos de r , o sistema apresenta convergência para um ponto de equilíbrio $(x,y,z) = (0,0,0)$. Acima de um certo valor, este ponto perde a estabilidade, e dois outros pontos passam a atrair as trajetórias. Fisicamente, trata-se do regime de convecção. A partir de um certo valor, passam a coexistir as duas soluções estacionárias e um atrator não-periódico. Aumentando ainda mais o valor de r ,

apenas este último atrator permanece estável. Neste contexto, os rolos de convecção não mais descrevem o que ocorre, sendo que a movimentação do fluido é mais complexa e irregular.

Usando $r = 28$, podemos visualizar na Fig. 3.1 o atrator comentado acima:

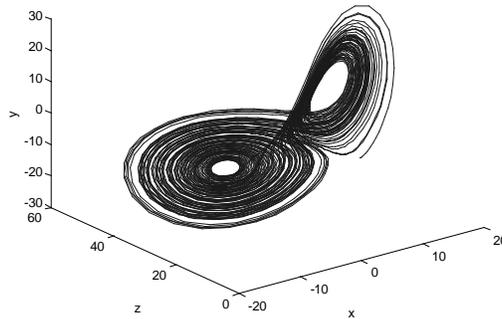


Figura 3.1: Atrator de Lorenz

Trata-se de um atrator bastante diferente dos que observamos até agora, lembrando um novelo de trajetórias em torno de dois “centros”, que correspondem ao par de pontos de equilíbrio correspondentes ao estado de convecção. Como o sistema dinâmico é determinístico, não há cruzamento de trajetórias, pois isto significaria que, a partir de um mesmo estado, o sistema poderia evoluir para estados diferentes. Iremos discutir com mais detalhe a estrutura deste tipo de atrator mais adiante.

Tentemos “reproduzir” a observação de Lorenz. Para tanto, utilizaremos a configuração dada acima e duas condições iniciais muito próximas.

Sejam $\mathbf{x}_{01} = (0.25, 2.1, 1.100000)$ e $\mathbf{x}_{02} = (0.25, 2.1, 1.100001)$. Na Fig. 3.2, foi traçada a diferença entre os valores da variável $x(t)$ nos dois casos. Podemos perceber claramente que os valores, inicialmente próximos, em $t = 20$ s já divergem claramente, não sendo aparente que as duas trajetórias se originaram de um estado inicial tão próximo. Trata-se de um exemplo clássico de DSCI, que está na essência do comportamento caótico.

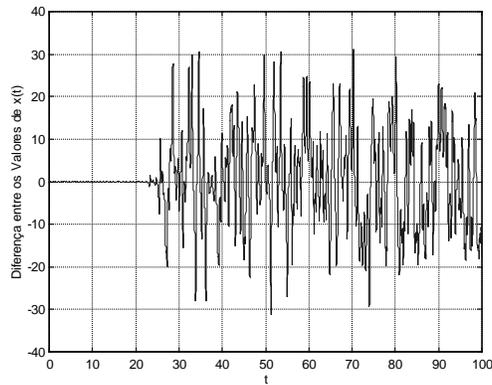


Figura 3.2: Diferença entre os Valores de $x(t)$ para Duas Condições Iniciais Distintas

Este efeito tem implicações físicas bastante concretas. Trata-se de uma grande limitação relativa à informação que podemos obter acerca de um sistema caótico: um pequeno “engano” sobre o estado do sistema logo leva à quase completa ignorância sobre o mesmo. Considerações baseadas na teoria de informação ajudam a elucidar este relevante aspecto da DSCI, como veremos adiante. Uma vez feita a introdução ao caos, aproximadamente seguindo as próprias constatações pioneiras de Poincaré, Maxwell e Lorenz, podemos tentar aprofundar nossos conceitos. Um conceito matemático muito importante para a análise de sistemas dinâmicos em geral, e de sistemas caóticos em particular, é o de expoentes de Lyapunov.

3.2. *Expoentes de Lyapunov*

Quando analisamos a DSCI, percebemos que ela pode ser entendida como uma espécie de situação em que há divergência entre trajetórias inicialmente próximas. Esta intuição encontra justificativa matemática através dos *expoentes de Lyapunov*. Estes entes matemáticos indicam como uma diferença inicial entre estados evolui em termos exponenciais.

Para entendermos a situação, observemos a Fig. 3.3 que mostra a evolução de um conjunto de pontos contido num elemento circular num espaço de duas dimensões [Fiedler-Ferrara e do Prado, 1994].

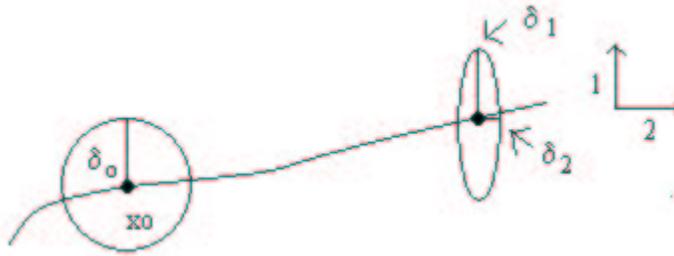


Figura 3.3: Evolução de um Elemento de Volume

onde $\delta_0(x_0)$ é o raio do elemento original e $\delta_1(t)$ e $\delta_2(t)$ são os semi-eixos nas direções 1 e 2 relativos ao elemento elíptico, observado após t segundos. Analisar a evolução de δ_1 e δ_2 proporciona muita informação sobre o fluxo. Procurando ajustar as distorções decorrentes da evolução do elemento circular a um modelo exponencial de variação temporal, chega-se aos expoentes de Lyapunov.

Suponhamos agora um sistema de ordem genérica. Podemos definir os expoentes de Lyapunov como [Fiedler-Ferrara e do Prado, 1994]:

$$\lambda_i = \lim_{t \rightarrow \infty} \lim_{\delta_0(x_0) \rightarrow 0} \frac{1}{t} \cdot \ln \left[\frac{\delta_i(t)}{\delta_0(x_0)} \right], i = 1, \dots, K \quad (3.4)$$

onde K é a ordem do sistema dinâmico.

Esta equação permite a seguinte aproximação (desprezando-se os limites):

$$\delta_i(t) \approx \delta_0(x_0) \cdot \exp(\lambda_i \cdot t) \quad (3.5)$$

Pensando em um sistema contínuo, vemos que expoentes negativos indicam convergência de trajetórias em suas direções correspondentes. Já expoentes nulos associam-se a soluções periódicas enquanto expoentes positivos estão ligados à divergência entre trajetórias (uma característica central em sistemas que apresentam DSCI). A presença de pelo menos um expoente positivo está, portanto, intimamente ligada à ocorrência de caos determinístico.

Definamos um elemento de volume no espaço de dimensão K como:

$$dV(t) = \delta_1(t) \cdot \delta_2(t) \cdot \dots \cdot \delta_k(t) \quad (3.6)$$

Usando (3.5) e (3.6), temos:

$$dV(t) = dV(0) \cdot \exp[(\lambda_1 + \dots + \lambda_k)t] \quad (3.7)$$

A partir de (3.7) percebemos que, caso a soma dos expoentes seja nula, o volume do elemento se conservará. Isto caracteriza um *sistema conservativo*, estando tal resultado ligado ao chamado *teorema de Liouville* [Symon, 1982, Guckenheimer e Holmes, 1983]. Uma soma de expoentes negativa associa-se à contração do elemento de volume, o que caracteriza um *sistema dissipativo*.

A soma dos expoentes deve ser negativa ou nula, para que o elemento não tenda a ter um volume infinito, o que caracterizaria instabilidade.

3.2.1. Cálculo do Expoente de Lyapunov para um Mapa Unidimensional

Suponhamos agora que o sistema de interesse seja um mapa unidimensional. Como fizemos para o caso contínuo, é necessário também definir uma variável que permita quantificar o tipo de evolução dinâmica experimentada por um conjunto de pontos qualquer. Suponhamos que, no instante $n = 0$, tomamos dois pontos, $x(0)$ e $y(0)$. Suponhamos ainda que passemos a acompanhar a evolução temporal dos mesmos. A mudança em suas posições relativas, caso os dois estivessem inicialmente a uma distância suficientemente próxima, fornece informação sobre o comportamento do mapa, de forma análoga ao elemento de volume empregado na seção anterior.

No instante inicial, o elemento é dado por:

$$\varepsilon = y(0) - x(0) \quad (3.8)$$

No instante imediatamente posterior, temos um novo valor para o elemento:

$$\varepsilon' = y(1) - x(1) \quad (3.9)$$

Relacionar estes dois valores é fundamental para que se quantifique matematicamente a distorção imposta pelo sistema ao elemento original. Novamente, utilizamos um ajuste do tipo exponencial:

$$\varepsilon' = \exp(\lambda) \cdot \varepsilon \quad (3.10)$$

onde λ é um expoente. De (3.8), (3.9) e (3.10), deduzimos:

$$|F[x(0) + \varepsilon] - F[x(0)]| = |\varepsilon| \cdot \exp(\lambda) \quad (3.11)$$

Iterando-se o mapa N vezes, obtém-se:

$$|F^N[x(0) + \varepsilon] - F^N[x(0)]| = |\varepsilon| \cdot \exp(N \cdot \lambda) \quad (3.12)$$

onde $F^N(\cdot)$ corresponde a N aplicações da função F . Isolando λ , temos:

$$\lambda = \frac{1}{N} \cdot \ln \left| \frac{F^N[x(0) + \varepsilon] - F^N[x(0)]}{\varepsilon} \right| \quad (3.13)$$

Considerando uma distância inicial ε infinitesimal e um número de aplicações tendendo a infinito, chegamos a:

$$\lambda = \lim_{N \rightarrow \infty} \lim_{\varepsilon \rightarrow 0} \frac{1}{N} \cdot \ln \left| \frac{F^N[x(0) + \varepsilon] - F^N[x(0)]}{\varepsilon} \right| = \lim_{N \rightarrow \infty} \frac{1}{N} \ln \left| \frac{dF^N[x(0)]}{dx(0)} \right| \quad (3.14)$$

Através da regra da cadeia e de uma propriedade dos logaritmos, chegamos enfim a:

$$\lambda[x(0)] = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=0}^{N-1} \ln |F'(x_i)| \quad (3.15)$$

que é a definição do expoente de Lyapunov para um sistema discreto com apenas uma dimensão. Em um mapa unidimensional, o expoente de Lyapunov é negativo quando há convergência para um ponto fixo ou para um atrator periódico, se anula quando ocorre uma bifurcação e é positivo quando há DSCI.

3.2.2. Cálculo dos Expoentes de Lyapunov para um Mapa Bidimensional

Vejamos agora como fica o cálculo dos expoentes de Lyapunov para o caso de um mapa bidimensional. Por analogia com o caso de um mapa unidimensional, temos:

$$\lambda_j = \lim_{N \rightarrow \infty} \frac{1}{N} \ln |\Lambda_j^N|, \quad j = 1, \dots, p \quad (3.16)$$

onde $|\Lambda_j|$ são os módulos dos autovalores da matriz M definida como:

$$\mathbf{M} = \prod_{i=1}^N \mathbf{J}(\mathbf{x}_i) \quad (3.17)$$

sendo $\mathbf{J}(\mathbf{x}_i)$ a matriz jacobiana calculada no estado \mathbf{x}_i .

A analogia entre (3.16) e (3.15) justifica-se pelo fato de a matriz jacobiana ser uma “generalização” da derivada unidimensional para campos vetoriais.

O uso direto de (3.16) costuma, porém, levar a problemas numéricos, mesmo para N não muito grande. Tais problemas podem ser contornados através do uso de um procedimento de ortogonalização, como o de Gram-Schmidt.

3.3. A Entropia de Kolmogorov-Sinai

Seja um atrator no espaço de estado. Recubramos o atrator com hipercubos de lado ϵ . Seja uma sequência de cubos $c_1, c_2, c_3, \dots, c_b$. Suponha que o estado do sistema visite os

cubos listados (na ordem apresentada) em intervalos regulares t . A informação média associada a uma sequência de cubos, segundo a definição de Shannon, é dada por:

$$I_b^{(1)}(\epsilon) = \sum_{c_1, \dots, c_b} p(c_1, \dots, c_b) \cdot \ln[p(c_1, \dots, c_b)] \quad (3.18)$$

onde $p(c_1, \dots, c_b)$ é a probabilidade de que os cubos c_1, \dots, c_b sejam visitados sequencialmente. O somatório refere-se a todas as possibilidades de escolha dos cubos. Intuitivamente, $I_b^{(1)}$ é a quantidade de informação necessária para localizar o estado em uma trajetória especial c_1^*, \dots, c_b^* , dadas apenas as probabilidades $p(c_1, \dots, c_b)$.

A grandeza:

$$I_{b+1}^{(1)}(\epsilon) - I_b^{(1)}(\epsilon) \quad (3.19)$$

determina a informação necessária para se localizar o estado do sistema (determinar o cubo c_{b+1}) sabendo que ele percorreu a sequência de cubos c_1, \dots, c_b . A grandeza expressa por (3.19) quantifica a perda de informação do observador (e o ganho de informação do sistema, conseqüentemente) entre os instantes de tempo bt e $(b+1)t$.

A *entropia de Kolmogorov-Sinai* é definida como a taxa média de criação de informação no sistema (ou de perda de informação do observador) [Fiedler-Ferrara e do Prado, 1994]:

$$K_s = \lim_{t \rightarrow 0} \lim_{\epsilon \rightarrow 0} \lim_{b \rightarrow \infty} \frac{1}{bt} \sum_{c_1, \dots, c_b} p(c_1, \dots, c_b) \cdot \ln[p(c_1, \dots, c_b)] \quad (3.20)$$

sendo que, para mapas, $t=1$.

Qual o motivo de se estudar o fluxo temporal de informação em um sistema dinâmico? Imaginemos um sistema no qual o estado está convergindo para um ponto fixo: estudando o comportamento do sistema por algum tempo, obtemos a informação necessária para antecipar com precisão qual será o seu comportamento num instante de tempo quase que imediatamente posterior. Isto quer dizer que não há produção de informação pelo sistema.

Pensemos agora num sistema caótico. A DSCI já nos faz sentir que com o passar do tempo, a informação que temos sobre o sistema num instante inicial fica cada vez menos significativa. Isto indica que há produção de informação por parte do sistema, o que justifica a definição da entropia de Kolmogorov-Sinai [Parker e Chua, 1987].

A partir do que comentamos, podemos intuir que há uma íntima relação entre os expoentes de Lyapunov e a entropia de Kolmogorov-Sinai. Para percebermos melhor tal fato, iremos tecer uma série de considerações [Fiedler-Ferrara e do Prado, 1994].

Pensemos no espaço de estados. Iremos recobri-lo com uma “malha” de uma certa “espessura”. Quanto mais fina for a malha, mais definição se tem na especificação da posição de um certo estado. Se a malha delimitar uma só região, não há necessidade sequer de especificação. Havendo duas, é necessária uma “coordenada”. Havendo quatro, são necessárias duas especificações. Em suma, podemos escrever:

$$I^{(1)} = \log_2 N \quad (3.21)$$

que é a definição de Boltzmann para a informação. Quanto maior o número de regiões, maior o número de especificações necessárias e maior a informação gerada pela malha.

A informação média ligada à posição do sistema, segundo Shannon, é:

$$I^{(1)} = -\sum_{i=0}^N p_i \cdot \log_2 p_i \quad (3.22)$$

De (3.22), chega-se a (3.21) supondo equiprobabilidade das regiões. Supondo um sistema contínuo, consideremos Δx_i um elemento de comprimento na direção i e ϵ uma unidade de comprimento (muito menor que Δx_i). Então, $\Delta x_i = N \cdot \epsilon$, onde N é o número de subregiões. A informação que se obtém através da especificação é:

$$I_i^{(1)} = \log_2 (\Delta x_i / \epsilon) \quad (3.23)$$

Admitamos que Δx_i corresponda à distância, na direção i , entre dois pontos no espaço de estados, num instante t . A partir de (3.5), temos:

$$\Delta x_i(t) = \Delta x_i(0) \cdot \exp(\lambda_i \cdot t) \quad (3.24)$$

Substituindo (3.24) em (3.23), temos:

$$I_i^{(1)} = \log_2 [\Delta x_i(0) \cdot \exp(\lambda_i \cdot t) / \epsilon] \quad (3.25)$$

Portanto,

$$dI_i^{(1)}/dt \propto \lambda_i \quad (3.26)$$

Quando há expoentes de Lyapunov positivos (DSCI), produz-se informação no sistema (a taxa de variação temporal é positiva). Isto condiz com o que discutimos anteriormente.

Definamos a grandeza:

$$L = \sum_{\lambda_i > 0} \lambda_i \quad (3.27)$$

Pelo que discutimos acima, é natural que a mesma esteja relacionada à taxa de produção de informação no sistema, que é a entropia de Kolmogorov-Sinai. Temos que:

$$K_s \leq L \quad (3.28)$$

sendo que, sob algumas condições, vale a igualdade:

$$K_s = L \quad (3.29)$$

Podemos nos perguntar: qual a relação entre a entropia de Kolmogorov-Sinai e o tipo de sistema dinâmico analisado? Tentemos elucidar, a partir do que já discutimos, este ponto muito relevante.

Num sistema determinístico *regular* (com convergência para pontos de equilíbrio ou ciclos-limite), pontos próximos no espaço de estados evoluem sem uma divergência significativa. Com isto, não se produz informação no sistema e, conseqüentemente, não aumenta a ignorância do observador com o passar do tempo, sendo K_s igual a zero. Num sistema determinístico *caótico*, existe DSCI, havendo portanto divergência. Ocorre então produção de informação no sistema dinâmico, com perda de informação por parte de um observador. Logo, $K_s > 0$, obedecendo a (3.28). Caso o sistema seja *estocástico*, a evolução do estado não seguirá nenhum tipo de continuidade no espaço de estado, havendo uma perda abrupta de informação por parte de um observador. Portanto, $K_s \rightarrow \infty$.

A exposição deste conceito nos traz mais uma informação acerca de sistemas caóticos: os mesmos têm entropia de Kolmogorov-Sinai positiva e finita. Este conceito também destaca a distinção entre um sistema caótico e um sistema estocástico.

3.4. *Atratores Estranhos*

Nesta seção, iremos discutir a idéia de *atrator estranho*, intimamente ligada ao comportamento caótico de sistemas dinâmicos. Dado um conjunto limitado A num espaço de N dimensões, podemos dizer que o mesmo é um atrator associado a um campo vetorial $F(\mathbf{x})$ se há um conjunto U tal que [Ruelle, 1980]:

- 1) U é uma vizinhança N -dimensional de A , ou seja, para cada ponto \mathbf{x} de A há uma pequena esfera centrada em \mathbf{x} e completamente contida em U .
- 2) Para cada ponto inicial \mathbf{x}_0 em U , o ponto \mathbf{x}_t permanece em U para valores positivos de t . O estado se aproxima e permanece tão próximo quanto se queira de A para t grande o suficiente. Isto quer dizer que A é um *atrator*.
- 3) Pode ser escolhido um ponto \mathbf{x}_0 em A , tal que, arbitrariamente próximo de cada outro ponto Y em A , há um ponto \mathbf{x}_t para algum t positivo. Isto nos mostra que A não pode ser decomposto em dois atratores distintos.

Analisemos cada uma das proposições. A primeira tem a finalidade de definir uma vizinhança U que envolva o atrator. Qual o motivo de se trabalhar com uma vizinhança? Ora, é justamente o fluxo *na vizinhança* de A que pode caracterizar a sua capacidade de atrair trajetórias. Por exemplo, um ponto de equilíbrio pode atrair ou repelir, mas este comportamento é observado não através do ponto em si, e sim das trajetórias em sua vizinhança. Imaginemos que o conjunto A em questão seja composto pelos pontos de uma circunferência no plano. A vizinhança U poderia ser uma coroa circular de espessura muito pequena, como nos mostra a Fig. 3.4.

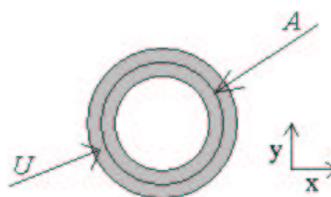


Figura 3.4: Exemplo de definição de U

Na segunda proposição, temos a noção de atração propriamente dita. A idéia é a seguinte: caso o estado do sistema assuma um valor pertencente a U , o mesmo ocorrerá para todos os valores subsequentes. Isto quer dizer que, uma vez na vizinhança de A , não há mais possibilidade de escape por parte do estado. Ainda é dito que, uma vez em U , a tendência é que o estado se aproxime cada vez mais do atrator, o que é bastante intuitivo.

Na terceira proposição, é imposta uma condição sobre a unicidade do atrator. Esta unicidade é garantida pela não-existência de regiões de atração isoladas entre si. A condição é que, dado um ponto \mathbf{x}_0 inicial, a evolução subsequente do estado irá fazer com que ele se aproxime arbitrariamente de *qualquer* ponto de A , caso A seja um único atrator. Portanto, unicidade significa inexistência de zonas de atração distintas e independentes.

Tendo exposto a noção de atrator, passemos a analisar o que quer dizer “estranho”. Isto nos leva a uma quarta proposição:

4) A é um *atrator estranho* se for um atrator e se, quando a condição inicial \mathbf{x}_0 estiver no conjunto U , a evolução temporal subsequente apresentar DSCI.

Portanto, a “estranheza” destes atratores se liga diretamente à ocorrência de dependência sensitiva das condições iniciais. Por que a DSCI torna estes atratores tão diferentes que tenham merecido, da parte de David Ruelle e Floris Takens, tal nome?

Já vimos na seção 3.1 o que significa, do ponto de vista dinâmico, a ocorrência de DSCI. Duas condições iniciais muito próximas passam a evoluir de forma significativamente diferente após alguns instantes. Tal comportamento deve estar de alguma forma espelhado na estrutura topológica dos atratores estranhos. Isto põe no ar uma questão: que tipo de processo *sui generis* ocorre em tais entes geométricos? Antes de discutirmos este ponto, voltemos aos nossos conhecidos atratores “usuais”.

Por exemplo, suponhamos que nosso espaço de estado tenha duas dimensões, e que a origem seja um atrator. Se imaginarmos um elemento Q , composto por alguns pontos e acompanharmos sua evolução, percebemos que ele sofre um processo de contração até que, no limite, tende a um lugar geométrico de dimensão nula (a própria origem) [Fiedler-Ferrara e do Prado, 1994]. Isto pode ser visto na Fig. 3.5.

Em tal figura, a título de ilustração apresentamos ainda Q' , que corresponde a uma observação do conjunto de pontos em um instante qualquer, o que explicita a contração experimentada pelo elemento.

Se o atrator fosse um ciclo-limite, o elemento de volume sofreria rotação, tendendo a um valor unitário de dimensão (que é a dimensão do ciclo).

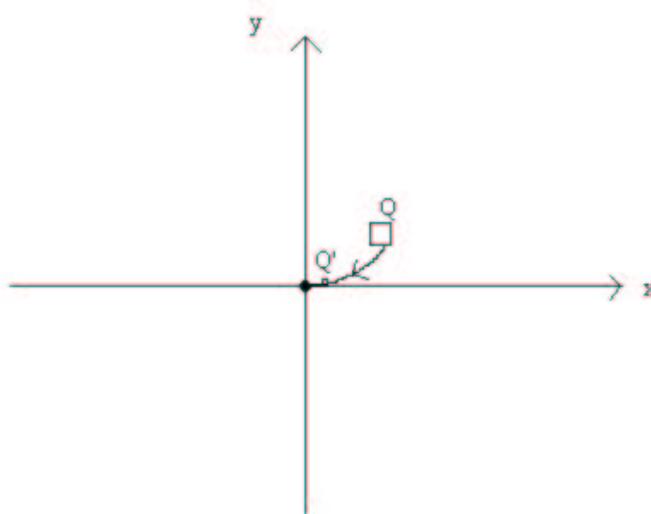


Figura 3.5: Evolução Temporal de um Elemento

Como procuramos ilustrar, usualmente a atração para um ponto fixo se dá por meio de processos de contração e os ciclos-limite estão associados a processos de rotação. Mas serão estes os únicos processos possíveis?

Há algumas questões muito pertinentes que podem surgir e que ainda não foram adequadamente respondidas. Na seção 3.2, mencionamos que sistemas caóticos apresentam pelo menos um expoente de Lyapunov positivo. Ora, vimos que isto significa a ocorrência de uma divergência do tipo exponencial entre trajetórias. Como tal divergência pode se processar sem que haja uma explosão para o infinito? É evidente que há peças faltando neste quebra-cabeça topológico.

A resposta para estas indagações está no chamado *processo de dobra*. Tal processo pode lidar com as duas questões colocadas no parágrafo anterior, ou seja: comportar divergência exponencial sem explosão e “produzir” DSCI.

Um exemplo que pode trazer alguma luz à nossa discussão é o seguinte: imaginemos um pedaço de massa de pão. Marquemos dois pontos nessa massa, e comecemos a trabalhá-la através do procedimento usual dos padeiros, que inclui expansões e dobras sucessivas. Podemos imaginar que, após alguns segundos, os pontos ocuparão posições tão distintas que alguém que chegue durante o processo nem sequer suspeitará que eles tenham estado muito próximos no começo do trabalho. O tipo de deformação imposto pelos processos de expansão e dobra é, portanto, a chave para a DSCI, para a divergência exponencial e também a razão pela qual atratores estranhos são tão diferentes dos demais atratores.

O teorema de Poincaré-Bendixson [Guckenheimer e Holmes, 1983] atesta que só há duas possibilidades de atrator para um sistema dinâmico bidimensional: ponto fixo ou ciclo-limite. Isto exclui categoricamente a possibilidade de ocorrência de um atrator estranho em um sistema contínuo com dimensão menor que três. Vale frisar que esta restrição não se aplica a sistemas discretos, não contemplados pelo teorema. Mesmo em sistemas discretos com apenas uma dimensão pode ocorrer caos, como é o caso do mapa logístico, que veremos na seção 3.6.

3.4.1. O Mapa de Hénon

Hénon, em 1976, propôs um sistema com dinâmica discreta de ordem dois no qual ocorre um atrator estranho. O modelo consiste das seguintes equações [Hénon, 1976]:

$$x(n+1) = 1 - a \cdot x^2(n) + y(n) \quad (3.30)$$

$$y(n+1) = b \cdot x(n) \quad (3.31)$$

A idéia de Hénon foi propor um modelo mais simples, mas que possuísse as propriedades fundamentais do sistema de Lorenz. Em suas palavras, foi uma “abordagem reducionista”.

Inspirado por resultados anteriores sobre o sistema de Lorenz, ele propôs três transformações do plano em si mesmo. A primeira transformação realiza a dobra, a segunda realiza uma contração na direção do eixo x (por um fator b), e a terceira faz uma troca entre os eixos x e y. O encadeamento das três leva a (3.30) e (3.31).

Os valores utilizados em [Hénon, 1976] foram $a = 1.4$ e $b = 0.3$, os mesmos que iremos utilizar nesta seção. A Fig. 3.6 nos dá a evolução do sistema para uma condição inicial arbitrária $x(0) = 1$ e $y(0) = 0$.

Podemos perceber a condição inicial e a segunda iteração, assinaladas por círculos, e depois não mais podemos acompanhar o estado do sistema na figura acima, denominada *atrator de Hénon*. Trata-se de um sistema que apresenta DSCI, como já era esperado a partir da constituição das transformações que originaram o mapa.

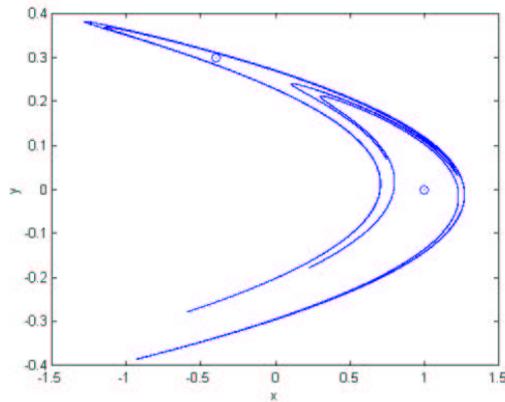


Figura 3.6: Atrator de Hénon

É possível vislumbrar a complexidade da estrutura deste atrator estranho através de ampliações. A Fig. 3.7 nos mostra duas delas.

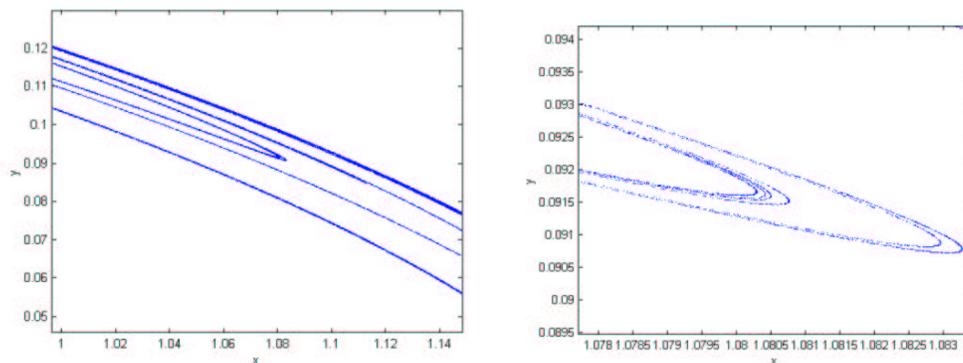


Figura 3.7: Ampliações do Atrator de Hénon

Os dois gráficos nos dão uma idéia da *auto-similaridade* inerente ao atrator. Uma figura auto-similar é caracterizada por ter partes cuja estrutura é semelhante à estrutura do todo. Isto, como veremos a seguir, ocorre tipicamente em *fractais*.

3.4.2. Dimensão Fractal, Dimensões Generalizadas e a Conjectura de Kaplan-Yorke

Estamos acostumados a pensar em dimensão como sendo o número de vetores que compõem uma base ortonormal relativa a um certo espaço [Boldrini et al., 1986]. Esta

definição, muito intuitiva por sinal, parece quase uma verdade indiscutível. Mas o fato é que, a partir dos anos 70 [Gleick, 1987], houve uma grande explosão de pesquisas envolvendo figuras geométricas com uma característica bastante peculiar: sua dimensão não podia ser expressa por um número inteiro. Tais figuras foram denominadas *fractais*.

A dimensão dos fractais não é contemplada pela definição usual de dimensão (como se pode ter uma base com 2.7815 vetores, por exemplo?). Isto exigiu o uso de outras ferramentas matemáticas para caracterizar tais figuras. Uma delas, a única que discutiremos com algum detalhe neste trabalho, é a *dimensão de Hausdorff* ou *dimensão de Hausdorff-Besicovitch* [Parker e Chua, 1987, Fiedler-Ferrara e do Prado, 1994].

Seja um conjunto de pontos A num espaço de dimensão N . Sejam estes pontos recobertos por hiper-cubos de aresta ϵ . A dimensão de Hausdorff é dada por:

$$D_0 = \lim_{\epsilon \rightarrow 0} \frac{\log K(\epsilon)}{\log(1/\epsilon)} \quad (3.32)$$

onde $K(\epsilon)$ é o número mínimo de hiper-cubos de aresta ϵ necessários para cobrir o conjunto de pontos A . Podemos expressar (3.32) de forma inversa, dizendo que $K(\epsilon)$ varia segundo $\exp(-D_0)$ para $\epsilon \rightarrow 0$.

É importante ressaltar que a definição da dimensão de Hausdorff fornece valores adequados não só quando aplicada a figuras com dimensão fracionária, mas também produz resultados coerentes para figuras com dimensão inteira.

Analisemos um exemplo de fractal, para que possamos tomar contato com a riqueza destas figuras à primeira vista tão paradoxais. O exemplo por nós escolhido é o *conjunto de Cantor*, que definiremos a seguir [Parker e Chua, 1987, Fiedler-Ferrara e do Prado, 1994].

Imaginemos um segmento de reta. Dividamos o mesmo em três partes iguais. Agora, retiremos a parte do meio. Retiremos a terça parte central dos dois segmentos restantes, e depois façamos o mesmo com os quatro, oito, enfim, *infinitas* vezes. No infinito, estará formado o conjunto de Cantor. A Fig. 3.8 ilustra a construção deste conjunto.



Figura 3.8: Conjunto de Cantor

Na etapas 1, 2, 3, ou em qualquer outra, a figura será composta por um número de segmentos de reta, e terá dimensão um. Mas isto não garante nada a respeito da dimensão do conjunto de Cantor, como tentaremos mostrar.

Vejamos a aplicação de (3.32): para a situação 1 da Fig. 3.8 é preciso, para recobrir toda a figura, um hiper-cubo (no caso, o “hiper-cubo” é um simples segmento de reta), de aresta L (onde L é o comprimento do segmento original). Na segunda situação, são necessários dois hiper-cubos, cada um com aresta $L/3$. Na terceira, quatro hiper-cubos de aresta $L/9$. Portanto, temos:

$$D_0 = \lim_{n \rightarrow \infty} \frac{\log 2^n}{\log(3^n / L)} \approx \frac{\log 2}{\log 3} \approx 0.6309 \quad (3.33)$$

O cálculo da dimensão de Hausdorff nos mostra que o conjunto de Cantor não é simplesmente “poeira de pontos” com dimensão nula, mas tem um valor fracionário de dimensão.

Neste caso, escolhemos uma disposição ótima para os hiper-cubos, o que nos permitiu um cálculo bastante simples. Caso não possamos conhecer, a priori, uma distribuição ótima, é necessário recobrir os pontos com um número adequado de hiper-cubos de aresta ε bastante pequena para se ter uma boa aproximação de D_0 . Este tipo de procedimento é denominado *contagem de caixas*.

Há um comentário que precisa ser feito. Existem fractais não-uniformes que necessitam, para uma adequada caracterização, de outras dimensões além da dimensão de

Hausdorff. Dentre estas, podemos citar a *dimensão de informação* (D_1) e a *dimensão de correlação* (D_2).

Por fim, vale citar a célebre *conjectura de Kaplan-Yorke* [Fiedler-Ferrara e do Prado, 1994], que liga os expoentes de Lyapunov de um sistema caótico à dimensão fractal de seu atrator estranho. Definamos, em primeiro lugar, a *dimensão de Kaplan-Yorke*, ou *dimensão de Lyapunov* [Parker e Chua, 1987]:

$$D_{KY} = j + \frac{\sum_{i=1}^j \lambda_i}{|\lambda_{j+1}|} \quad (3.34)$$

onde $\lambda_1 > \lambda_2 > \dots > \lambda_N$ são os expoentes de Lyapunov ordenados de forma decrescente, e j é o maior inteiro tal que:

$$\sum_{i=1}^j \lambda_i > 0 \quad (3.35)$$

A conjectura de Kaplan-Yorke, que tem sido confirmada computacionalmente em vários sistemas é que:

$$D_{KY} = D_0 \quad (3.36)$$

Cabe ainda um comentário final. Uma característica marcante dos fractais é a sua auto-similaridade. Podemos perceber no conjunto de Cantor como as partes têm uma estrutura análoga à da figura como um todo. Os atratores caóticos, conforme já havíamos afirmado na seção anterior, também apresentam esta propriedade. A observação desta característica (até um certo ponto) em plantas, flocos de neve, vasos capilares, para ficar em alguns exemplos, levou a uma popularização da idéia de dimensão fracionária. É preciso frisar, no entanto, que um fractal é um ente que tem característica auto-similar para ampliações sucessivas de *qualquer ordem*, o que não ocorre em um objeto físico.

3.5. *Caracterização do Caos Determinístico*

É difícil dar uma definição precisa do que seja comportamento caótico. Não é nosso objetivo dar uma definição completa e rigorosa, mas apontar aspectos relevantes para que possamos entender melhor o que significa, do ponto de vista dinâmico, um sistema determinístico operar em regime de caos.

Vejamos algumas características emblemáticas de sinais e sistemas caóticos [Fiedler-Ferrara e do Prado, 1994]:

- 1) Um sistema caótico apresenta DSCI.
- 2) A evolução temporal das variáveis de estado é rápida e tem uma aparência errática, embora seja possível muitas vezes observar padrões que se repetem aproximadamente, de forma não-periódica.
- 3) O sinal caótico tem uma função de autocorrelação temporal que decresce rapidamente, o que é esperado a partir da idéia de DSCI e de expansão e dobra no espaço de estado. Assim, amostras espaçadas de um certo intervalo de tempo têm uma correlação tanto menor quanto maior for o intervalo, via de regra. Como consequência disto, os processos caóticos têm espectro de potências contínuo e de faixa larga.
- 4) Em um sistema caótico, há pelo menos um expoente de Lyapunov positivo e a entropia de Kolmogorov-Sinai é positiva e finita. Isto quer dizer que há uma produção de informação por parte do sistema.
- 5) Sistemas caóticos dão origem a atratores estranhos, que por sua vez possuem dimensão fracionária.

3.6. O Mapa Logístico

Em 1845, P. F. Verhulst [Schuster, 1988] introduziu um modelo populacional discreto para uma espécie mantida em uma área fechada:

$$x(n+1) = F[x(n)] = \mu \cdot x(n) \cdot [1 - x(n)] \quad (3.37)$$

Tal equação é denominada *mapa logístico* [Schuster, 1988]. Neste mapa, $x(n)$ assume valores entre 0 e 1, sendo que o menor valor representa a ausência de população, e 1 representa a maior população possível. Trata-se de um dos sistemas mais analisados e discutidos dentro da área de caos. Sua rica dinâmica começou a ser melhor compreendida nos anos 70, através dos trabalhos de Robert May e Mitchell Feigenbaum, entre outros [Gleick, 1987]. Nesta seção, iremos discutir alguns aspectos do comportamento de (3.37), que ajudarão a fixar os conceitos discutidos e nos familiarizarão com os procedimentos de análise que usaremos em capítulos posteriores.

Para melhor compreendermos o sentido de (3.37), partamos do famoso modelo malthusiano de população [Schuster, 1988]:

$$x(n+1) = \mu \cdot x(n) \quad (3.38)$$

que conduz a um crescimento exponencial desenfreado de população (para $\mu > 1$) ou a uma contínua redução (para $\mu < 1$), não sendo, portanto, um modelo realista. É preciso levar em conta limitações de recursos e de espaço, o que pode ser feito através da introdução de um termo do tipo $(1-x)$, que leva a uma diminuição de população para valores altos da mesma, o que contemplaria uma possível luta por recursos escassos, doenças, etc.

Trata-se de um sistema com um modelo dinâmico muito simples, facilmente iterável numa calculadora científica, mas que tem, em sua estrutura, bastante complexidade.

Uma das ferramentas mais usadas para se analisar o comportamento de um sistema dinâmico discreto (ou um sistema contínuo discretizado), quando são variados seus parâmetros, é o *diagrama de bifurcação*. Em tal diagrama, é ilustrado o regime permanente do sistema dinâmico em função dos parâmetros de controle ou de bifurcação. Assim, é

possível observar para que valores destes parâmetros ocorrem mudanças substanciais na dinâmica do sistema.

Para o mapa logístico, traçamos um diagrama no qual foram exibidos os 40 últimos valores de $x(n)$, em função do valor de μ . O resultado está na Fig. 3.6.

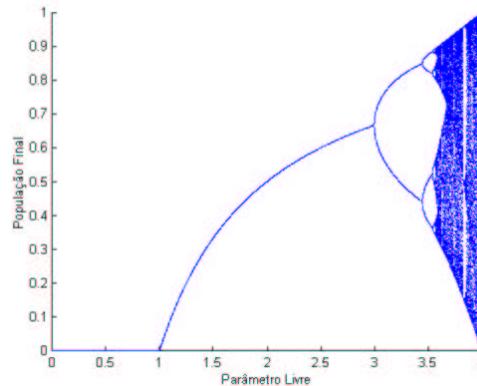


Figura 3.9: Diagrama de Bifurcação

Para $0 < \mu < 1$, a população tende à extinção, como nos indica o gráfico. Para $1 < \mu < 3$, a população tende para um valor de equilíbrio, que varia com o parâmetro de bifurcação. Para $\mu > 3$, observamos que a população oscila periodicamente, primeiro com período 2, depois com período 4, 8, 16, e assim sucessivamente até um ponto de acumulação, a partir do qual começa a ocorrer comportamento caótico, correspondente às regiões mais “cheias” do diagrama. Dentro da faixa de caos, há zonas nas quais o sistema volta a ter comportamento periódico, ilustradas por traços verticais brancos na figura. Para $\mu > 4$ o sistema diverge.

Em suma, temos uma situação de convergência para pontos fixos, ocorrendo então uma bifurcação de Hopf (seção 2.6.2) seguida por uma *cascata de bifurcações flip* (seção 2.6.3) que levam a uma zona caótica, entremeada por *janelas de periodicidade*, nas quais podemos ter inclusive períodos de oscilação ímpares. Após isto, há divergência para o infinito.

Analisemos, passo a passo, este mapa, que apresenta propriedades que encontraremos adiante, quando analisarmos o CMA. Isto aumentará nossa familiaridade, tanto em relação ao método de análise quanto a sistemas discretos caóticos de modo geral.

3.6.1. Obtenção dos Pontos de Equilíbrio e Análise de Estabilidade

A aplicação de (2.11) para se obter os pontos de equilíbrio do mapa logístico nos leva à seguinte equação algébrica:

$$x_e = \mu \cdot x_e \cdot (1 - x_e) \quad (3.39)$$

que por sua vez fornece duas soluções:

$$x_{e1} = 0 \text{ ou } x_{e2} = 1 - \mu^{-1} \quad (3.40)$$

A análise de estabilidade do sistema depende da primeira derivada de $F(x)$, que é a matriz jacobiana do sistema, dada por:

$$F'(x) = \mu \cdot (1 - 2x) \quad (3.41)$$

Para o ponto x_{e1} , temos:

$$F'(0) = \mu \quad (3.42)$$

Para o ponto x_{e2} , temos:

$$F'(1 - \mu^{-1}) = 2 - \mu \quad (3.43)$$

Em ambos os casos, o ponto fixo só será estável se:

$$|F'(x_e)| < 1 \quad (3.44)$$

o que ocorre para o primeiro ponto na faixa:

$$0 < \mu < 1 \quad (3.45)$$

e para o segundo na faixa:

$$1 < \mu < 3 \quad (3.46)$$

Os resultados condizem, como era de se esperar, com as faixas de estabilidade dos pontos de equilíbrio obtidas diretamente a partir do diagrama.

3.6.2. Os Ciclos-Limite

Para $\mu > 3$, o segundo ponto de equilíbrio também perde a estabilidade. Portanto, não há mais pontos fixos atratores no mapa. Isto indica que o comportamento do sistema deve sofrer uma mudança significativa, o que acontece com o surgimento de ciclos-limite estáveis.

O que ocorre, basicamente, é uma bifurcação de Hopf seguida por uma sequência infinita de bifurcações *flip*, sendo que cada uma duplica o período de oscilação do estado. Podemos analisar a estabilidade destes ciclos através de uma aplicação composta de $F(x)$. Por exemplo, se temos um ciclo-limite de período dois, com valores de oscilação iguais a x_1 e x_2 , vale:

$$F(x_1) = x_2 \text{ e } F(x_2) = x_1 \quad (3.47)$$

o que nos leva a:

$$F[F(x_1)] = x_1 \text{ ou } F[F(x_2)] = x_2 \quad (3.48)$$

Portanto, os pontos fixos de $F^{(2)}(x) = F[F(x)]$ incluirão os dois pontos de equilíbrio já conhecidos de $F(x)$ e mais os dois pontos do ciclo-limite de período dois. Para fixar melhor esta idéia, observemos a Fig. 3.10.

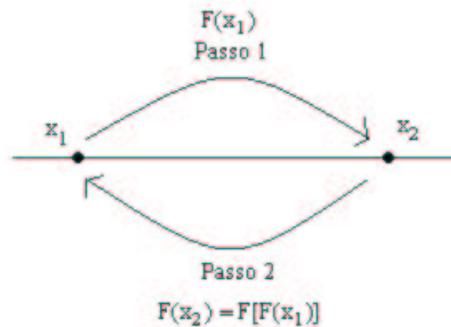


Figura 3.10: Ciclos-Limite como Pontos de Equilíbrio de $F^{(2)}(\cdot)$

Suponhamos que apliquemos a função $F(\cdot)$ ao ponto x_1 (passo 1). Isto levará ao ponto x_2 . Apliquemos agora $F(\cdot)$ a este ponto (passo 2) e teremos como resultado o ponto de partida, x_1 . Portanto, x_1 é ponto de equilíbrio da aplicação conjunta (passo 1 + passo 2) de $F(\cdot)$, ou seja, de $F^{(2)}(\cdot)$. O mesmo raciocínio se aplica se partirmos de x_2 , o outro componente do ciclo-limite.

Analisar a estabilidade destes pontos é analisar a estabilidade do ciclo de período dois. Raciocínio análogo pode ser feito para um ciclo de qualquer periodicidade.

Após sucessivas bifurcações, porém em intervalos cada vez mais curtos, chega-se a um valor do parâmetro de controle para o qual as bifurcações tendem a um período infinito, sendo que a partir daí começa a ocorrer caos.

3.6.3. Caos, Janelas de Periodicidade e Crises

A região caótica é perceptível no diagrama apresentado na Fig. 3.9, pois é uma faixa “cheia”, ao contrário das regiões de convergência para pontos fixos ou ciclos-limite. A explicação para isto é a característica essencialmente aperiódica do sinal caótico, que não dá margem ao surgimento de padrões pontuais no diagrama.

Porém, em algumas regiões, observamos “manchas brancas” que indicam volta a um comportamento periódico. De fato, estas regiões são denominadas *janelas de*

periodicidade. Podemos observar algumas janelas na Fig. 3.11, que é composta por ampliações da Fig. 3.9.

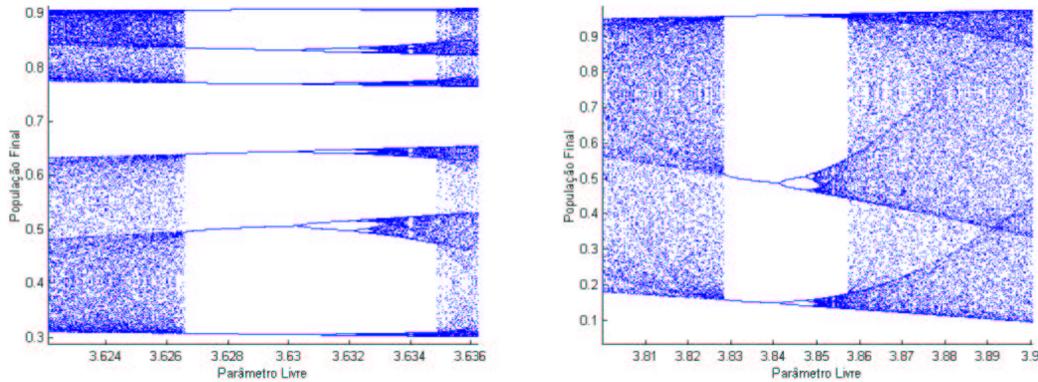


Figura 3.11: Janelas de Periodicidade

Nota-se que há uma mudança abrupta no comportamento do sistema, com o surgimento de pontos entre os quais o estado oscila. O sistema experimenta então uma cascata de duplicações de período, seguida de caos.

A bifurcação que faz surgir os ciclos-limite estáveis é do tipo *sela-nó* (seção 2.6.1), ou seja, para cada valor do ciclo há um segundo valor instável. É interessante perceber que, logo que o caos recomeça, há três zonas caóticas separadas, cada uma correspondente a um dos pontos do ciclo. Para μ próximo de 3.86, há um alargamento súbito das três bandas caóticas, formando uma banda única. A este fenômeno, damos o nome de *crise interior*. Esta crise é causada justamente pela interação entre os pontos de sela instáveis (e portanto, não visíveis no diagrama) e os atratores estranhos ligados à rota para o caos que envolve também os nós [Grebogi et al., 1983].

3.6.4. Análise por Diagramas de Teia

Uma poderosa ferramenta de análise de mapas unidimensionais é o digrama de teia [Aguirre, 1996, Fiedler-Ferrara e Cintra do Prado, 1994]. Neste diagrama é possível acompanhar a evolução da variável de estado do sistema através das sucessivas iterações. A base deste diagrama consiste de duas funções, a função $F(\cdot)$, que nada mais é do que o

campo vetorial discreto da equação (2.2), e a função identidade. A importância da função $F(\cdot)$, como já discutimos anteriormente, está em seu papel de destaque na determinação das características dinâmicas do mapa. Já a função identidade é uma referência para a realização das iterações, introduzindo a realimentação inerente aos mapas discretos.

Para o mapa logístico, o cerne do diagrama está na Fig. 3.12, para $\mu = 3$.

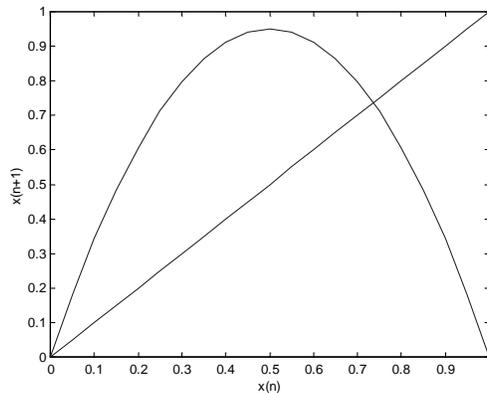


Figura 3.12: Cerne do Diagrama de Teia

Como podemos observar na Fig. 3.12, no eixo das abcissas é representado valor presente do estado, $x(n)$, e no eixo das ordenadas, o valor após uma iteração, $x(n+1)$. Isto nos dá a idéia de como realizar uma vez o processo iterativo. Mas, e depois? Bem, aí entra a função identidade: através dela, pode-se “transformar” um valor de $x(n+1)$ no eixo das ordenadas em um valor $x(n)$ no eixo das abcissas. Utilizando $\mu = 3$ e a condição inicial $x(0) = 0.1$, foi traçada a Fig. 3.13, que mostra as duas primeiras iterações do mapa através do diagrama de teia.

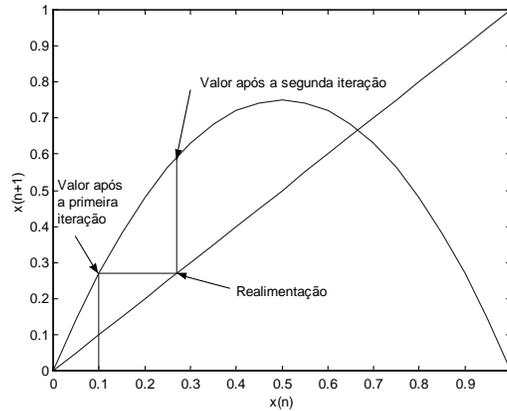


Figura 3.13: Visualização de Duas Iterações do Mapa Logístico

Na primeira iteração, o valor $x(0) = 0.1$, representado no eixo das abscissas, produz $x(1) = 0.27$, numa ascensão vertical até o encontro com a parábola $F(\cdot)$. Isto é o equivalente geométrico da aplicação da função ao um valor $x(0)$. Na próxima iteração, $x(1)$ precisará ser um valor presente, o que se realiza através de um movimento horizontal até que haja um encontro com a função identidade. Temos expressa a idéia de recursão, inerente à classe de mapas que estamos estudando. Para a segunda iteração, o mesmo processo se repete.

Uma vez colocadas estas idéias principais, estamos aptos a aplicar o diagrama à análise do mapa logístico. Manteremos $x(0) = 0.1$ para todos os casos, e variaremos μ para realizar um estudo envolvendo os diversos tipos possíveis de comportamento em estado estacionário apresentados pelo mapa logístico. Em primeiro lugar, tomemos $\mu = 0.5$. O diagrama de teia correspondente está na Fig. 3.14.

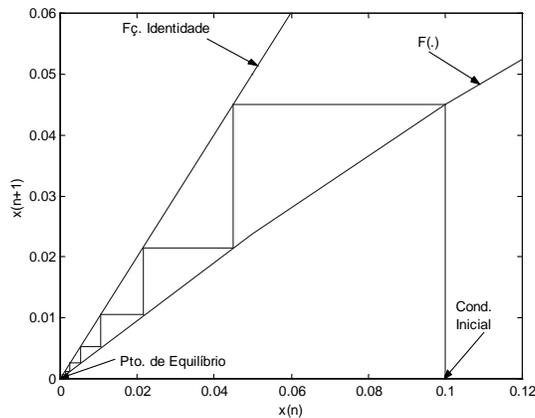


Figura 3.14: Diagrama de Teia para $\mu = 0.5$

Os pontos para os quais há o cruzamento entre $F(\cdot)$ e a função identidade são os pontos de equilíbrio do mapa em questão, o que se pode apreender da análise de (2.11). Para o caso em foco, o cruzamento ocorre apenas na origem, o que demonstra ser $x_e = 0$ o único ponto de equilíbrio para este valor de μ . A estabilidade deste ponto é atestada pela convergência do estado para a origem, conforme indica a Fig. 3.14. Isto era esperado, a partir da análise empreendida por outros meios, como o diagrama de bifurcação da Fig. 3.9.

Analisemos agora uma outra situação, para a qual $\mu = 2.5$, na Fig. 3.15.

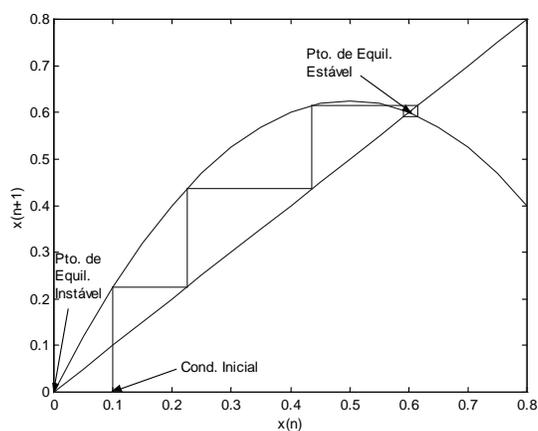


Figura 3.15: Diagrama de Teia para $\mu = 2.5$

Para este caso, temos dois pontos de equilíbrio. O primeiro deles, $x_e = 0$, é instável, como já havíamos inferido de nossa análise anterior. Podemos verificar este fato também no diagrama, verificando que uma reta tangente à parábola $F(\cdot)$, na origem, teria coeficiente angular de magnitude maior que a unitária. Isto quer dizer que $|F'(0)| > 1$ o que nos remete às equações (3.42) e (3.44). Já o outro ponto de equilíbrio, $x_e = 0.6$, será estável, o que já era esperado e de fato é verificado pela convergência ilustrada na Fig. 3.15.

Tomemos agora $\mu = 3.3$. A Fig. 3.16 nos traz o diagrama de teia para este caso.

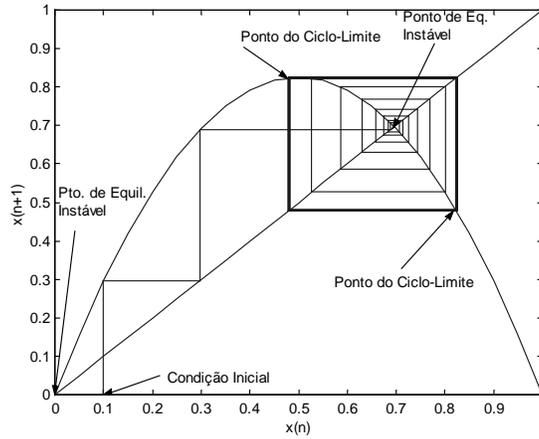


Figura 3.16: Diagrama de Teia para $\mu = 3.3$

De fato, verificamos o comportamento periódico esperado tendo em vista a Fig. 3.9. O estado não mais converge para um ponto fixo, mas tende a oscilar entre dois valores, vértices de um quadrado visível na Fig. 3.16.

Em nossa última análise, utilizamos $\mu = 4$. O diagrama de teia encontra-se na Fig. 3.17.

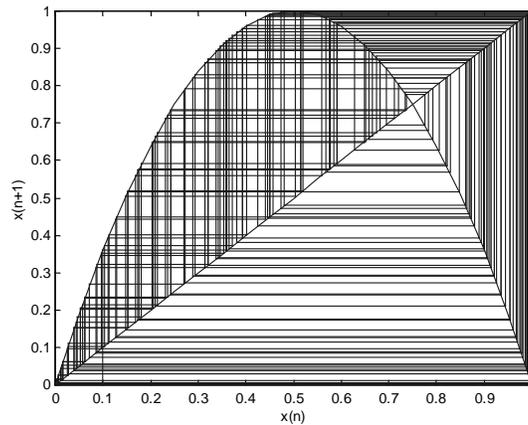


Figura 3.17: Diagrama de Teia para $\mu = 4$

Para tal valor de μ , é esperado comportamento caótico, conforme a análise anteriormente empreendida. Tal comportamento, de natureza aperiódica, leva às

características representadas na Fig. 3.17. Figuras como esta explicam o fato de a nomenclatura destes diagramas estar associada à idéia de uma teia de aranha.

Diagramas de teia, como procuramos mostrar, são ferramentas de grande utilidade para o entendimento da dinâmica de mapas unidimensionais, revelando de forma clara e pictórica o comportamento de tais sistemas.

3.6.5. Universalidade

Mitchell Feigenbaum, ao analisar diferentes mapas unidimensionais, percebeu determinadas propriedades que depois se mostraram válidas para uma ampla classe de sistemas [Feigenbaum, 1978, 1979 e 1983]. Fazendo contas em sua calculadora [Gleick, 1987], o pesquisador começou a notar padrões regulares na rota para o caos de duplicação de período, padrões estes posteriormente qualificados de *universais*, por se repetirem em uma ampla classe de sistemas dinâmicos. A intuição inicial foi depois muito bem formalizada em diversos artigos, revelando aspectos até então desconhecidos mas de grande relevância, tanto matemática quanto física. Vale ressaltar que, embora os aspectos apontados por Feigenbaum fossem inéditos, a idéia de universalidade já havia sido explorada em outro âmbito por Metropolis, Stein e Stein [Gleick, 1987, Schuster, 1988, Fiedler-Ferrara e do Prado, 1994].

Para que possamos entender o que é universalidade, voltemo-nos para o exemplo do mapa logístico. A Fig. 3.9 traz o diagrama de bifurcação deste mapa, onde podemos observar a cascata de bifurcações “flip” que caracteriza o cenário. Na Fig. 3.18 temos uma ampliação da Fig. 3.9, onde podemos observar esta região com mais detalhe.

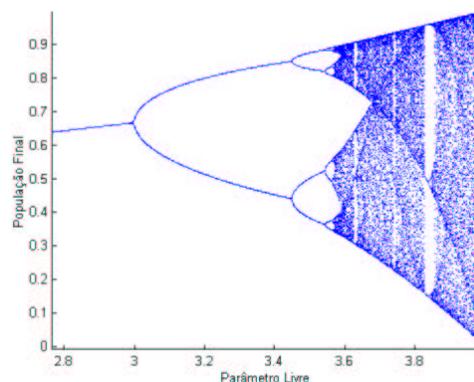


Figura 3.18: Ampliação Mostrando as Bifurcações Flip

Denominemos μ_n o valor do parâmetro livre para o qual ocorre a n-ésima bifurcação flip. A grandeza analisada por Feigenbaum foi a razão entre a extensão de duas zonas de oscilação periódica sucessivas [Feigenbaum, 1978]:

$$\delta_n = \frac{\mu_n - \mu_{n-1}}{\mu_{n+1} - \mu_n} \quad (3.49)$$

O que se percebeu foi que:

$$\lim_{n \rightarrow \infty} \delta_n = \delta \quad (3.50)$$

onde δ é a primeira constante de Feigenbaum, cujo valor é:

$$\delta = 4.669201609... \quad (3.51)$$

passível de obtenção apenas numérica. Este valor é universal dentro de uma ampla classe de mapas.

A segunda constatação de Feigenbaum está ligada ao máximo da função $F(\cdot)$. Basicamente, ele observou que há um fator de escala que liga as sucessivas distâncias entre o ponto de máximo e o ponto integrante do ciclo-limite mais próximo ao máximo, quando este último é um dos pontos componentes do ciclo. Este conceito torna-se mais claro pela observação da Fig. 3.19 onde as sucessivas distâncias citadas acima são representadas por d_1, d_2 .

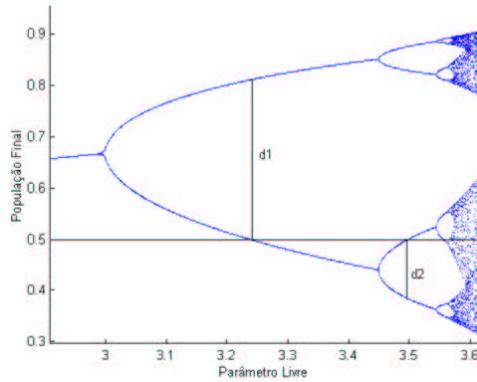


Figura 3.19: Distâncias Sucessivas d_n

Definida a razão entre distâncias sucessivas:

$$\alpha_n = \frac{d_n}{d_{n+1}} \quad (3.52)$$

o que Feigenbaum [Feigenbaum, 1978] demonstrou foi que:

$$\lim_{n \rightarrow \infty} \alpha_n = \alpha \quad (3.53)$$

onde α é uma constante universal para uma classe de mapas, cujo valor é dado por:

$$\alpha = 2.50290787... \quad (3.54)$$

Mais do que apenas uma curiosidade, a universalidade é uma propriedade muito forte, e isto não apenas do ponto de vista teórico. Ela mostra que sistemas distintos podem apresentar determinadas características e até mesmo valor de constantes em comum. Daí a força desta propriedade: ser um termo comum entre sistemas não-lineares distintos que apresentam um cenário de duplicação de período.

3.7. *Discussão*

O objetivo deste capítulo foi expor e analisar alguns conceitos básicos da teoria de sistemas dinâmicos caóticos, formando assim uma base para as discussões posteriores sobre o comportamento do CMA. Tentaremos agora discutir um pouco a idéia geral do que apresentamos.

A característica principal de um sistema caótico é apresentar dependência sensitiva das condições iniciais (DSCI). Esta característica está ligada ao comportamento extremamente complexo de um sistema dinâmico operando em regime caótico, que tem um caráter de aparente aleatoriedade. Na verdade, não há aleatoriedade alguma, apenas um processo de evolução temporal rico que inclui expansão e dobras no espaço de estado.

Uma outra forma de se enxergar a DSCI é através da divergência entre trajetórias próximas, o que nos leva ao conceito de expoente de Lyapunov, central na análise de sinais caóticos. Esta divergência, como comentamos, pode ser vista também do ponto de vista da teoria de informação, o que conduz à aplicação do conceito de entropia de Kolmogorov-Sinai. De posse deste conceito, pode-se tecer um quadro muito enriquecedor a respeito do fluxo de informação associado à evolução temporal do vetor de estado.

A idéia de atrator estranho também está intimamente ligada ao caráter geral de um sistema caótico. Como nos mostra a conjectura de Kaplan-Yorke, pelo menos em muitos casos conhecidos a dimensão fractal do atrator está diretamente relacionada à questão da divergência entre trajetórias, ou seja, ao valor dos expoentes de Lyapunov positivos.

A análise do mapa logístico foi importante, pois nos colocou em contato com as idéias centrais do cenário de Feigenbaum, também chamado de cenário de duplicação de período, que será fundamental na análise posterior do algoritmo do módulo constante.

Por fim, talvez caiba colocar que todas estas idéias são aspectos de uma mesma essência, inerente a muitos sistemas não-lineares. Todos os conceitos apresentados aqui talvez possam ser imaginados como os sons de diferentes instrumentos que se juntam para expressar a misteriosa sonoridade do caos.

4. Filtragem Adaptativa e Equalização Supervisionada

4.1. Introdução

Basicamente, um filtro pode ser definido como um dispositivo que atua em um conjunto de dados, visando a extração de informação relevante [Haykin, 1996]. No problema clássico de filtragem, utiliza-se informação atual e passada de um sinal para a obtenção de características presentes relevantes. Iremos abordar, neste trabalho, uma das aplicações mais importantes dos filtros: a equalização. Para que possamos compreender a utilidade desta aplicação, precisamos nos voltar para os modelos de *sistemas de comunicação*.

Basicamente, o objetivo de um sistema de comunicação é permitir o fluxo de informação confiável entre transmissor e receptor, através de algum *meio de transmissão*. Este meio ou *canal* tem características próprias, independentes das do sinal sobre o qual atua e pode acarretar *distorção* da informação transmitida. Esta distorção pode levar a uma interpretação errônea ou incoerente da mensagem enviada, impossibilitando assim o intercâmbio de dados. Uma possível maneira de tornar a degradação imposta pelo meio limitada a patamares razoáveis está no uso de *equalizadores*. Tais dispositivos se encontram no receptor e atuam no sinal degradado a fim de tentar “compensar” a influência do canal sobre o sinal de informação, tendo uma atuação “inversa” à do meio de transmissão. Desta forma, a informação após o equalizador deve estar mais próxima da que foi transmitida, ocorrendo uma significativa melhoria no desempenho do sistema.

Vejamos, na Fig. 4.1, um modelo simplificado de sistema de comunicação:

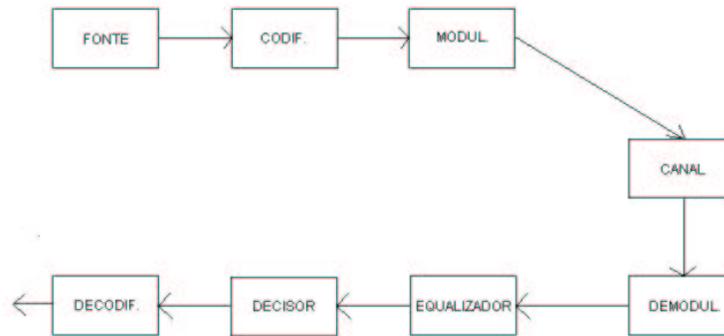


Figura 4.1: Modelo Simplificado de um Sistema de Comunicação

A informação é adequadamente codificada e modulada no transmissor, enviada através do canal de comunicação, e então demodulada, equalizada, “reconstruída” e decodificada no receptor. Idealmente, o que se espera é que a saída do equalizador seja igual a uma versão atrasada (de n_d) do sinal na saída do codificador, a menos de um possível fator de escala G . Matematicamente, temos então como objetivo:

$$y(n) = G.s(n-n_d) \quad (4.1)$$

onde $y(n)$ é o sinal na saída do equalizador e $s(n)$ é o sinal na entrada do canal.

Sendo obedecida esta expressão, é possível uma perfeita recuperação da informação enviada. Neste caso, fala-se em equalização *zero-forcing* (ZF). Nem sempre será possível satisfazer (4.1), sendo nestes casos desejável obter uma aproximação do critério ZF.

4.2. Modelos de Canais

Na Fig. 4.1, o ruído foi considerado parte integrante do bloco do canal, pois, em termos conceituais, o ruído é *parte do canal de comunicação*. Porém, neste trabalho não lidaremos diretamente com o ruído, sendo que a palavra “canal” estará associada a um sistema puramente determinístico.

Há diversos possíveis meios pelos quais se transmite a informação (atmosfera, par de cabos, fibra óptica, etc.), sendo que cada um tem características específicas. No entanto,

em muitos sistemas, há alguns fatores comuns de modelamento, seja por similaridade, seja por simplicidade matemática.

Há duas hipóteses amplamente aceitas na concepção de modelos para inúmeros sistemas de comunicação: *linearidade e invariância no tempo*. Feitas estas hipóteses, obtém-se um modelo bastante tratável matematicamente, e que se ajusta bem a diversos casos. Analisemos o que estamos assumindo. Seja $R(\cdot)$ o operador que fornece a resposta do canal a um sinal de entrada $s(n)$, e seja $x(n)$ a saída do canal.

O sistema é dito *linear* se e somente se [Oppenheim et al., 1997]:

$$R[k_1.s_1(n) + k_2.s_2(n)] = k_1.R[s_1(n)] + k_2.R[s_2(n)] \quad (4.2)$$

Em (4.2), temos a expressão matemática do *princípio da superposição*. Intuitivamente, temos que a resposta a entradas superpostas é a superposição das respostas do sistema a cada uma das entradas, individualmente. Há ainda uma segunda consideração de que, caso a entrada de um sistema linear seja amplificada, a sua saída será amplificada pelo mesmo fator (o termo amplificado refere-se a ganhos genéricos, não apenas a ganhos cujo módulo seja maior que um).

O sistema é dito *invariante no tempo* se [Oppenheim et al., 1997]:

$$x(n + n_0) = R[s(n + n_0)] \quad (4.3)$$

Esta expressão nos mostra que, uma vez que num certo instante n o sistema responde à entrada $s(n)$ com a saída $x(n)$, ele continuará a responder da mesma forma com o passar do tempo. Isto quer dizer que os parâmetros do canal não se alteram temporalmente. Os sistemas que obedecem a estas duas restrições são denominados lineares e invariantes no tempo (LIT). Neles, a saída se relaciona com a entrada através da relação:

$$x(n) = \sum_{k=-\infty}^{\infty} h(k).s(n - k) \quad (4.4)$$

onde $h(n)$ é a resposta do canal ao impulso unitário $\delta(n)$. Em (4.4), temos a expressão da convolução discreta. Podemos escrever, equivalentemente:

$$x(n) = h(n)*s(n) \quad (4.5)$$

A partir de (4.4), podemos escrever:

$$x(n) = h(n_d).s(n - n_d) + \sum_{k=-\infty, k \neq n_d}^{\infty} h(k).s(n - k) \quad (4.6)$$

A primeira parte do segundo membro corresponde ao sinal que se deseja recuperar, e o restante está ligado à influência do passado (e do futuro, pois não fizemos restrições quanto à causalidade). Esta última parcela corresponde a um tipo de interferência, a interferência intersimbólica (IIS). Esta interferência provoca distorções e pode levar a uma severa degradação no desempenho do sistema. Para satisfazer, ainda que aproximadamente, o critério ZF, o equalizador precisa reduzi-la consideravelmente.

Uma visão muito interessante e poderosa pode ser obtida a partir do uso da transformada Z. A transformada Z de um sinal discreto $s(n)$ é descrita por $S(z)$, e definida por:

$$S(z) = \sum_{k=-\infty}^{\infty} s(k).z^{-k} \quad (4.7)$$

A partir desta definição e de (4.4), chega-se a [Oppenheim et al., 1997]:

$$X(z) = H(z).S(z) \quad (4.8)$$

que é a contrapartida da convolução temporal no domínio da transformada. A transformada Z da resposta ao impulso, $H(z)$, se chama *função de transferência*. O uso de um simples produto ao invés de uma convolução faz com que se tenha mais facilidade na análise dos efeitos introduzidos por um canal de comunicação. Caso o *círculo de raio unitário* (CRU)

esteja no domínio da transformada Z de um certo sistema, podemos ainda obter o espectro de Fourier de $x(n)$ fazendo:

$$z = e^{j\omega} \quad (4.9)$$

Desta forma, temos que (4.8) se torna:

$$X(e^{j\omega}) = H(e^{j\omega}).S(e^{j\omega}) \quad (4.10)$$

onde $H(e^{j\omega})$ é a *resposta em frequência* do canal, e $S(e^{j\omega})$ e $X(e^{j\omega})$ são os *espectros de frequência* dos sinais de entrada e saída, respectivamente. Esta forma de análise permite que se tenha uma visão da influência do canal no domínio da frequência. Dado o sinal de informação $s(n)$, o mesmo tem um conteúdo em frequência $S(e^{j\omega})$. Ao ser enviado através de um canal com resposta em frequência $H(e^{j\omega})$, o sinal tem o seu espectro distorcido. Por exemplo, se o canal for *passa-baixas*, o conteúdo em alta frequência de $s(n)$ será muito atenuado. Portanto, o equalizador deve enfatizar esta região do espectro para compensar a distorção imposta pelo meio de transmissão.

Por fim, podemos reescrever a condição ZF (4.1) como:

$$Y(z) = G.S(z).z^{-n_d} \quad (4.11)$$

Vejamos agora alguns modelos de canal que empregaremos neste trabalho.

4.2.1. Canais Auto-Regressivos

Denominaremos esta classe de canais de *auto-regressivos* (AR) por analogia com os modelos de processos estocásticos de estrutura similar. Um canal será AR de ordem p se satisfizer a seguinte equação a diferenças:

$$s(n) = a_0.x(n) + a_1.x(n-1) + \dots + a_p.x(n-p) \quad (4.12)$$

Portanto, o valor da saída $x(n)$ dependerá do valor presente da entrada e de valores passados da própria saída, caracterizando *feedback*. Supondo, sem perda de generalidade, que $a_0 = 1$, podemos aplicar a transformada Z a (4.12):

$$H(z) = \frac{X(z)}{S(z)} = \frac{1}{1 + a_1 \cdot z^{-1} + \dots + a_p \cdot z^{-p}} \quad (4.13)$$

Percebe-se que apenas os p pólos de $H(z)$ trazem a informação sobre os parâmetros do canal. Antes de mais nada, é preciso que todos estejam contidos no círculo de raio unitário para que o canal seja estável e, portanto, realizável na prática. Canais AR são IIR (*infinite impulse response*), pois têm uma resposta ao impulso de duração infinita.

4.2.2. Canais Moving-Average

Esta classe de canais também recebe este nome por analogia com o modelo de processos estocásticos de estrutura similar. Um canal é denominado *moving-average* (MA) de ordem p se obedecer à seguinte equação a diferenças:

$$x(n) = b_0 \cdot s(n) + b_1 \cdot s(n-1) + \dots + b_p \cdot s(n-p) \quad (4.14)$$

O valor da saída do canal depende apenas do presente e do passado das amostras de entrada. Neste caso, não há realimentação, não havendo também problemas de estabilidade. Aplicando a transformada Z e supondo $b_0 = 1$, temos:

$$H(z) = \frac{X(z)}{S(z)} = 1 + b_1 \cdot z^{-1} + \dots + b_p \cdot z^{-p} \quad (4.15)$$

Portanto, a informação sobre o canal está contida no valor dos zeros de $H(z)$. Caso todos os zeros estejam dentro do CRU, o canal é dito de *fase mínima*. Caso todos estejam fora do CRU, o canal é dito de *fase máxima*. Nos demais casos, o canal é dito de *fase mista*.

Os canais MA também podem ser denominados FIR (*finite impulse response*), por terem uma resposta ao impulso de duração finita.

4.2.3. Canais ARMA

Esta classe de canais consiste de uma mistura entre os dois modelos anteriores. Portanto, temos que a ordem do modelo é dada por dois números, um associado à “porção AR” e o outro à “porção MA”. Para um modelo ARMA(p,q) temos a seguinte equação a diferenças:

$$x(n) + a_1.x(n-1) + \dots + a_p.x(n-p) = s(n) + b_1.s(n-1) + \dots + b_q.s(n-q) \quad (4.16)$$

Neste caso, também há realimentação, podendo haver problemas de estabilidade e valendo a mesma restrição do modelo AR quanto aos pólos. A função de transferência pode ser escrita como:

$$H(z) = \frac{1 + b_1.z^{-1} + \dots + b_q.z^{-q}}{1 + a_1.z^{-1} + \dots + a_p.z^{-p}} \quad (4.17)$$

Quanto à classificação relativa à posição dos zeros, valem também as mesmas considerações tecidas na descrição do modelo MA.

4.3. Equalizadores Lineares

Dentre as muitas possíveis estruturas para se realizar equalização, abordaremos neste trabalho apenas um tipo: os *filtros transversais lineares*. A estrutura de tais dispositivos está na Fig. 4.2.

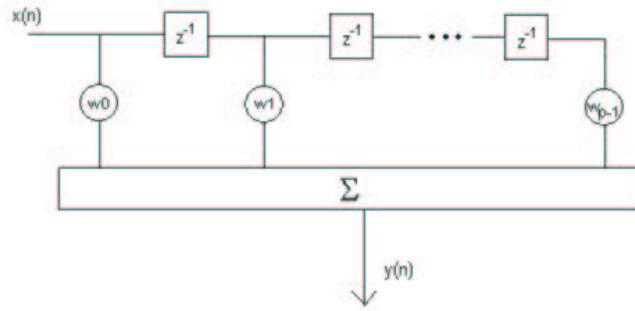


Figura 4.2: Filtro Transversal Linear

Tal equalizador nada mais é do que um filtro FIR (modelo MA), com um certo número de parâmetros livres (w_0 a w_{p-1}), que caracterizam a sua ordem. Esta estrutura tem duas grandes vantagens: não apresenta problemas de estabilidade, por não ser realimentada, e é simples de se tratar matematicamente, quando se busca a otimização de seus parâmetros. A resposta ao impulso destes filtros é dada pela sequência:

$$h_w(n) = \{w_0, w_1, \dots, w_{p-1}\} \quad (4.18)$$

sendo portanto o próprio vetor de parâmetros do dispositivo.

Conforme discutimos na seção anterior, o objetivo do equalizador é ter uma configuração inversa à do canal. Esta estrutura só é capaz de satisfazer o critério ZF se o canal não tiver zeros (sendo portanto AR), o que podemos perceber se observarmos como a saída do equalizador se relaciona com o sinal transmitido:

$$Y(z) = H(z).W(z).S(z) \quad (4.19)$$

Para que haja equalização ZF, é preciso que:

$$H(z).W(z) = G.z^{n_d} \quad (4.20)$$

Portanto, pólos “anulam” zeros e zeros “anulam” pólos. Como a nossa estrutura é caracterizada por zeros, ela precisaria ter uma ordem infinita para anular outros zeros. Porém, apesar de não poder equalizar com perfeição canais MA e ARMA, com o uso de

uma ordem adequada podem ser obtidas boas aproximações do critério ZF empregando-se um filtro transversal linear.

Resta, porém, uma questão muito importante: como ajustar os parâmetros do filtro de modo a se obter a situação desejada de equalização ZF, ou uma boa aproximação sua?

Caso tenhamos informações sobre o canal, e caso estas informações permaneçam válidas com o passar do tempo, pode-se escolher a localização dos zeros que melhor “anula” os pólos do canal. Trata-se de um processo que exige bastante técnica e intuição.

Outra maneira de realizar o projeto é através do uso de um sinal de treinamento, que indica o que se deseja obter após o processo de filtragem, fornecendo assim uma referência para o ajuste dos parâmetros. Este caminho leva à teoria da filtragem de Wiener, e aos algoritmos LMS e RLS.

Uma outra possível forma consiste em se explorar *estatísticas de ordem superior* (EOS) do sinal recebido, tendo algum conhecimento sobre as estatísticas do sinal transmitido, de modo a se extrair informação sobre a atuação do canal sobre o mesmo. Este caminho nos leva à teoria da equalização cega ou autodidata.

Neste trabalho, iremos abordar apenas as duas últimas maneiras, que permitem uma adaptação recursiva dos parâmetros do equalizador, sendo assim mais úteis para canais cujos parâmetros variam no tempo.

Uma ressalva que gostaríamos de fazer é sobre a aparente contradição entre o parágrafo anterior e a exposição do conceito de invariância no tempo. De fato, embora utilizemos a idéia de convolução em um filtro adaptativo, assumimos que os parâmetros do canal e do filtro variam, mas de uma forma que permita que se utilize a propriedade da convolução em cada instante de tempo, com o valor adequado dos parâmetros.

Para que possamos analisar a teoria da equalização autodidata de forma conveniente, passemos a uma atenta análise de alguns aspectos da teoria clássica de filtragem adaptativa supervisionada, que tem seu resultado central associado ao nome de Norbert Wiener.

4.4. Filtragem de Wiener

Quando se projeta um filtro, inicialmente é preciso escolher uma estrutura. Uma vez escolhida, faz-se necessário um ajuste de seus parâmetros, visando obter uma configuração ótima segundo algum critério. No paradigma associado a Wiener, o objetivo é ajustar os parâmetros do filtro de modo a aproximar a sua saída de um sinal desejado conhecido. A Fig. 4.3 contém o esquema deste modelo de filtragem.

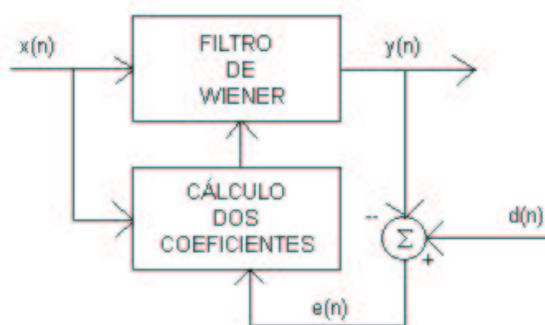


Figura 4.3: Filtro de Wiener

Basicamente, temos um filtro com um vetor de parâmetros \mathbf{w} , de comprimento infinito, em cuja entrada são inseridas amostras de um processo estocástico $x(n)$. A saída deste filtro, $y(n)$, é comparada com um *sinal desejado* $d(n)$, sendo que a diferença entre os dois sinais dá origem a um *erro de estimação* $e(n)$.

O objetivo do ajuste dos parâmetros do filtro é fazer com que $y(n)$ seja a melhor estimativa possível de $d(n)$. Portanto, trata-se de projetar um dispositivo que “molde” o processo estocástico $x(n)$ de modo a se ter uma saída semelhante a um modelo de sinal conhecido *a priori*. O grande objetivo por trás do processo é reduzir o erro de estimação a um mínimo. Esta redução é a linha condutora dos esforços de otimização empreendidos.

Certamente, minimizar o erro implica no tratamento uniforme para erros positivos e negativos, de modo que o que se deseja afinal é aproximar o erro de zero. Também é relevante levar em consideração não apenas situações instantâneas, mas uma minimização no sentido da *média*.

Um critério que contempla estas observações é o do *erro quadrático médio* [Haykin, 1996]. Assim, a função custo pode ser escrita, no domínio dos complexos, como:

$$J = E[e(n).e^*(n)] = E[|e(n)|^2] \quad (4.21)$$

A saída do equalizador será:

$$y(n) = \sum_{k=0}^{\infty} w_k^* . x(n - k) \quad (4.22)$$

Definamos o k-ésimo peso complexo:

$$w_k = a_k + j.b_k \quad (4.23)$$

E o k-ésimo elemento do gradiente complexo:

$$\nabla_k = \partial/\partial a_k + j.(\partial/\partial b_k) \quad (4.24)$$

Para obtermos o ponto de mínimo da função custo, precisamos ter, para todos os valores de k:

$$\nabla_k J = 0 \quad (4.25)$$

Substituindo (4.22) em (4.21), aplicando o operador gradiente definido em (4.24) e levando em conta a regra da cadeia, chegamos a:

$$\nabla_k J = -2.E[x(n - k).e^*(n)] \quad (4.26)$$

Aplicando (4.26) à condição (4.25), obtemos por fim:

$$E[x(n - k).e_o^*(n)] = 0 \quad k=0, 1, 2, \dots \quad (4.27)$$

onde $e_o(n)$ designa o erro na situação de otimalidade, ou seja, de gradiente nulo.

A equação (4.27) permite que digamos [Haykin, 1996]:

A condição necessária e suficiente para que a função J atinja o seu mínimo valor é que o valor do erro de estimação seja ortogonal a cada amostra de entrada empregada na estimativa da resposta desejada no instante n .

Tal resultado é conhecido como *princípio da ortogonalidade*. Um interessante corolário de (4.27) é:

$$E[y_o(n) \cdot e_o^*(n)] = 0 \quad (4.28)$$

que indica que, no filtro de Wiener, a saída $y_o(n)$ é ortogonal ao sinal de erro.

Tentemos agora obter o valor do vetor de pesos para o qual há otimalidade. Para tanto, apliquemos (4.22) a (4.27):

$$E\{x(n-k) \cdot [d^*(n) - \sum_{i=0}^{\infty} w_{oi} \cdot x^*(n-i)]\} = 0 \quad (4.29)$$

onde w_{oi} é o i -ésimo coeficiente do filtro ótimo.

Manipulando (4.29), temos:

$$\sum_{i=0}^{\infty} w_{oi} \cdot E[x(n-k) \cdot x^*(n-i)] = E[x(n-k) \cdot d^*(n)], \text{ para } k = 0, 1, \dots \quad (4.30)$$

O termo $r(i-k) = E[x(n-k) \cdot x^*(n-i)]$ é a função de autocorrelação de $x(n)$ para um atraso de $i-k$. O termo $p(-k) = E[x(n-k) \cdot d^*(n)]$ é a correlação cruzada entre a entrada do filtro $x(n)$ e a resposta $d(n)$ para um atraso de $-k$.

Portanto, temos que (4.30) se torna:

$$\sum_{i=0}^{\infty} w_{oi} \cdot r(i-k) = p(-k), \quad k = 0, 1, \dots \quad (4.31)$$

que são as *equações de Wiener-Hopf*.

Supondo agora um filtro linear com p coeficientes, podemos escrever as equações de Wiener-Hopf na forma matricial:

$$\mathbf{R} \cdot \mathbf{w}_o = \mathbf{p} \quad (4.32)$$

onde $\mathbf{R} = E[\mathbf{x}(n) \cdot \mathbf{x}^H(n)]$ e $\mathbf{p} = E[\mathbf{x}(n) \cdot d^*(n)]$, sendo $\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-p+1)]^T$.

Por fim, obtém-se o valor de \mathbf{w}_o por uma simples inversão:

$$\mathbf{w}_o = \mathbf{R}^{-1} \cdot \mathbf{p} \quad (4.33)$$

Podemos nos perguntar: como é a forma da função custo $J(\mathbf{w})$? Certamente isto é bastante relevante para que possamos caracterizar o esquema de filtragem ótima de Wiener.

Os resultados anteriores nos mostraram que há apenas um ponto, a *solução de Wiener* expressa em (4.33), que anula o gradiente da função. Portanto, a função é *unimodal*.

Uma primeira análise pode ser no sentido de obter o valor mínimo desta função, ou seja, $J(\mathbf{w}_o)$. Temos que:

$$J(\mathbf{w}_o) = E[|e_o(n)|^2] \quad (4.34)$$

Uma vez que:

$$d(n) = y_o(n) + e_o(n) \quad (4.35)$$

Temos ainda, supondo processos de média nula:

$$\sigma_d^2 = \sigma_{y_o}^2 + J(\mathbf{w}_o) \quad (4.36)$$

graças ao corolário do princípio da ortogonalidade. Portanto:

$$J_{\min} = J(\mathbf{w}_o) = \sigma_d^2 - \sigma_{y_o}^2 \quad (4.37)$$

Assim, o valor mínimo da função custo é a diferença entre a variância do sinal desejado e a variância da saída do filtro na condição ótima.

Tomando (4.21) e (4.22), após diversas manipulações, chegamos à expressão geral para a função custo:

$$J(\mathbf{w}) = \sigma_d^2 - \mathbf{w}^H \cdot \mathbf{p} - \mathbf{p}^H \cdot \mathbf{w} + \mathbf{w}^H \cdot \mathbf{R} \cdot \mathbf{w} \quad (4.38)$$

que também pode tomar as formas:

$$J(\mathbf{w}) = \sigma_d^2 - \mathbf{p}^H \cdot \mathbf{R}^{-1} \cdot \mathbf{p} + (\mathbf{w} - \mathbf{R}^{-1} \cdot \mathbf{p})^H \cdot \mathbf{R} \cdot (\mathbf{w} - \mathbf{R}^{-1} \cdot \mathbf{p}) \quad (4.39)$$

ou

$$J(\mathbf{w}) = J_{\min} + (\mathbf{w} - \mathbf{w}_o)^H \cdot \mathbf{R} \cdot (\mathbf{w} - \mathbf{w}_o) \quad (4.40)$$

A função custo $J(\mathbf{w})$ é portanto *quadrática*, e tem a forma de um parabolóide com p graus de liberdade, caracterizado por um único mínimo. Os exemplos a seguir podem fixar melhor os conceitos discutidos até agora.

Exemplo 4.1: Canal MA(1)

Seja um canal MA(1):

$$H_{MA}(z) = 1 + \alpha \cdot z^{-1} \quad (4.41)$$

cujo único zero encontra-se em $z = -\alpha$. Empreguemos um equalizador com dois coeficientes, com vetor de pesos:

$$\mathbf{w}(n) = [w_0(n) \ w_1(n)]^T \quad (4.42)$$

Suponhamos que o sinal $s(n)$ assumira valores $+1$ e -1 com igual probabilidade, e que há independência entre as amostras. Podemos escrever, a partir de (4.41):

$$x(n) = s(n) + \alpha \cdot s(n-1) \quad (4.43)$$

Portanto:

$$r(0) = E[s^2(n) + \alpha^2 \cdot s^2(n-1) + 2 \cdot \alpha \cdot s(n) \cdot s(n-1)] = 1 + \alpha^2 \quad (4.44)$$

E também:

$$r(1) = E\{ [(s(n) + \alpha \cdot s(n-1))] \cdot [s(n-1) + \alpha \cdot s(n-2)] \} = \alpha \quad (4.45)$$

Temos a seguinte propriedade da função de autocorrelação:

$$r(-k) = r(k) \quad (4.46)$$

que nos permite dizer que:

$$r(-1) = r(1) = \alpha \quad (4.47)$$

Portanto:

$$\mathbf{R} = \begin{bmatrix} 1 + \alpha^2 & \alpha \\ \alpha & 1 + \alpha^2 \end{bmatrix} \quad (4.48)$$

Supondo que o sinal desejado é o próprio sinal transmitido, temos:

$$d(n) = s(n) \quad (4.49)$$

Portanto:

$$p(1) = E[d(n).x(n)] = E\{s(n).[s(n) + \alpha.s(n-1)]\} = E[s^2(n)] = 1 \quad (4.50)$$

e

$$p(2) = E[d(n).x(n-1)] = E\{s(n).[s(n-1) + \alpha.s(n-2)]\} = 0 \quad (4.51)$$

O valor do vetor de correlação cruzada é então:

$$\mathbf{p} = [1 \ 0]^T \quad (4.52)$$

A solução de Wiener será, portanto:

$$\mathbf{w}_o = \mathbf{R}^{-1} \cdot \mathbf{p} = \left[\begin{array}{cc} \frac{1 + \alpha^2}{1 + \alpha^2 + \alpha^4} & \frac{-\alpha}{1 + \alpha^2 + \alpha^4} \end{array} \right]^T \quad (4.53)$$

Para podermos obter alguns resultados numéricos, fizemos $\alpha = 0.6$. Para este valor de α , a solução de Wiener é:

$$\mathbf{w}_o = [0.9130 \ -0.4028]^T \quad (4.54)$$

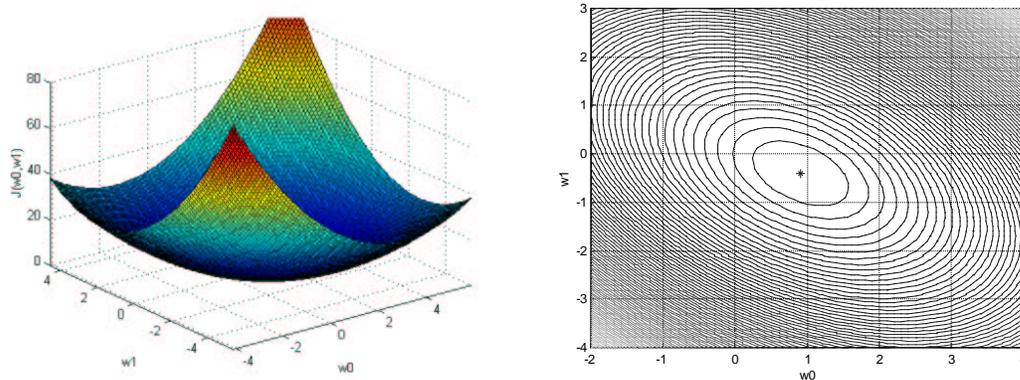


Figura 4.4: Superfície de Erro e Curvas de Nível

Na Fig. 4.4, está traçada a superfície de erro para o modelo utilizado no exemplo. Esta figura é complementada por um conjunto de curvas de nível de $J(\mathbf{w})$, onde a solução de Wiener está indicada por um asterisco, no centro das diversas elipses.

4.5. O Algoritmo LMS

Na filtragem de Wiener, tínhamos um situação eminentemente estática: há o cálculo do vetor ótimo de pesos, o que conclui o projeto do filtro. Porém, em situações práticas, é muito importante tentar atingir a solução ótima por meio de um processo iterativo. Há alguns motivos importantes para isto, como por exemplo o custo computacional. Solucionar as equações de Wiener-Hopf pode ser muito custoso computacionalmente, principalmente para equalizadores com muitos taps e operando em altas taxas [Haykin, 1996].

Uma alternativa pode ser o uso do método *steepest descent*, que descreveremos na seção seguinte.

4.5.1. O Método Steepest Descent

A partir do cálculo diferencial, sabemos que o gradiente de uma função num dado ponto, caso não seja o vetor nulo, tem sempre a direção e o sentido de maior crescimento da mesma [Kreyszig, 1993]. Intuitivamente, percebemos que “caminhar” no sentido contrário ao maior crescimento possível é uma boa estratégia para buscar um ponto de mínimo, pois trata-se da “descida mais íngreme”. Este é o conceito por trás do chamado *método steepest*

descent, introduzido por Cauchy em 1847 [Kreyszig, 1993]. Portanto, dada uma função custo $J(\mathbf{w})$, podemos expressar que a atualização do vetor de parâmetros \mathbf{w} se faz através da seguinte expressão:

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \cdot \nabla J[\mathbf{w}(n)] \quad (4.55)$$

onde o parâmetro μ , denominado *passo de adaptação*, controla o tamanho do passo dado a cada iteração. Repetidas realizações de (4.55) devem levar a um mínimo local, se o valor do passo for adequadamente escolhido.

Para que possamos utilizar este método na busca do valor que otimiza (4.21), precisamos efetuar o cálculo do vetor gradiente de tal função. Aplicando (4.24) às expressões (4.38-4.40), temos:

$$\nabla J = -2 \cdot (\mathbf{p} - \mathbf{R} \cdot \mathbf{w}) \quad (4.56)$$

A partir de (4.55) e (4.56), obtemos a expressão do método *steepest descent* aplicado à busca da solução de Wiener:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + 2 \cdot \mu \cdot [\mathbf{p} - \mathbf{R} \cdot \mathbf{w}(n)] \quad (4.57)$$

Por razões de simplicidade, incorporaremos o fator multiplicativo 2 ao passo de adaptação em (4.57).

Isto produz, finalmente:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \cdot [\mathbf{p} - \mathbf{R} \cdot \mathbf{w}(n)] \quad (4.58)$$

O método *steepest descent* nos levou, portanto, a um sistema dinâmico discreto linear com p variáveis de estado, onde p é o número de parâmetros do filtro. A aplicação da condição de equilíbrio (2.11) nos dá como único ponto justamente a solução de Wiener, como era de se esperar. Isto conclui a *análise estática do algoritmo*. Em termos de análise

dinâmica, compete a nós buscar uma análise de estabilidade do ponto obtido. Para isto, vamos expressar a matriz jacobiana (seção 2.5.2) de (4.58), dada por:

$$\mathbf{J} = \mathbf{I} - \mu \cdot \mathbf{R} \quad (4.59)$$

onde \mathbf{I} é a matriz identidade de ordem p .

Através de (4.59), podemos estabelecer uma relação direta entre os autovalores da matriz jacobiana e os da matriz de autocorrelação:

$$\lambda_J = 1 - \mu \cdot \lambda_R \quad (4.60)$$

onde λ_J é um autovalor genérico da matriz jacobiana e λ_R o autovalor correspondente na matriz de autocorrelação.

Sabemos que a condição de estabilidade para um sistema linear de ordem p é:

$$|\lambda_{Ji}| < 1, i = 0, \dots, p-1 \quad (4.61)$$

Aplicando (4.60), obtemos a condição equivalente:

$$|1 - \mu \cdot \lambda_{Ri}| < 1, i = 0, \dots, p-1 \quad (4.62)$$

A condição (4.62) pode ser desdobrada em:

$$-1 < 1 - \mu \cdot \lambda_{Ri} < 1, i = 0, \dots, p-1 \quad (4.63)$$

que equivale a:

$$0 < \mu < 2/\lambda_{Ri}, i = 0, \dots, p-1 \quad (4.64)$$

Sendo uma propriedade da matriz de autocorrelação ser *definida não-negativa*, ou seja, valendo a restrição:

$$\lambda_{Ri} \geq 0, i = 0, \dots, p-1 \quad (4.65)$$

a condição (4.64) equivale a:

$$0 < \mu < 2/\lambda_{R\text{máx}} \quad (4.66)$$

onde $\lambda_{R\text{máx}}$ é o autovalor da matriz \mathbf{R} com maior magnitude.

A faixa de valores indicada em (4.66) revela a faixa de operação satisfatória para o método *steepest descent*. Fora desta faixa, o algoritmo não mais converge para a solução de Wiener, sendo que o valor de $J[\mathbf{w}(n)]$ tende a valores infinitamente grandes. Graças à linearidade inerente à expressão iterativa (4.58), há apenas estas duas possibilidades.

Uma vez concluídas as análises estática e dinâmica, passemos a um exemplo de aplicação do método.

Exemplo 4.2 – Aplicação do Método *Steepest Descent*

Voltemos ao modelo do exemplo 1: uma canal MA(1) com $\alpha = 0.6$, e um equalizador com dois pesos. Usando as expressões (4.48) e (4.52) em (4.58), obtém-se a expressão do método para esta aplicação específica. A condição inicial escolhida foi $\mathbf{w}(0) = [0 \ 0]^T$.

Foram realizadas 1000 iterações com $\mu = 0.01$. A escolha do passo é razoável pois, para este caso, a faixa de estabilidade obtida a partir de (4.66) foi $0 < \mu < 1.0204$. A trajetória seguida pelo vetor de pesos encontra-se na Fig. 4.5, junto com as curvas de nível. É interessante notar que a trajetória segue aproximadamente uma direção ortogonal à tangente às curvas de nível em cada ponto de cruzamento. Isto decorre diretamente das propriedades do vetor gradiente e da suficiente pequenez do passo de adaptação. Após as sucessivas iterações, o vetor $\mathbf{w}(n)$ encontra-se muito próximo da solução de Wiener (indicada por um asterisco).

Na Fig. 4.6, encontramos a evolução de $J[\mathbf{w}(n)]$, simplesmente expressa por $J(n)$. Percebe-se que há um decrescimento no valor da função custo até um valor mínimo. Tal

comportamento era o esperado por nós. Calculemos o valor de J_{\min} para este caso. Para tanto, podemos utilizar a relação:

$$\sigma_{y_0}^2 = \mathbf{w}_o^H \cdot \mathbf{R} \cdot \mathbf{w}_o = \mathbf{w}_o^H \cdot \mathbf{p} \quad (4.67)$$

Substituindo (4.67) em (4.37), obtemos:

$$J_{\min} = \sigma_d^2 - \mathbf{w}_o^H \cdot \mathbf{p} \quad (4.68)$$

Sabemos que $\sigma_d^2 = 1$. Utilizando os resultados do exemplo 1, obtemos:

$$J_{\min} = 0.0870 \quad (4.69)$$

Podemos perceber que este valor é o valor mínimo da Fig. 4.6, o que reafirma o bom desempenho do método. Com 400 iterações já se encontrava o erro muito próximo do limitante teórico (4.69).

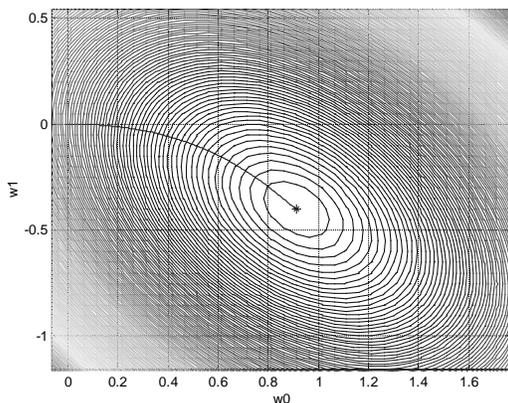


Figura 4.5: Trajetória de Convergência

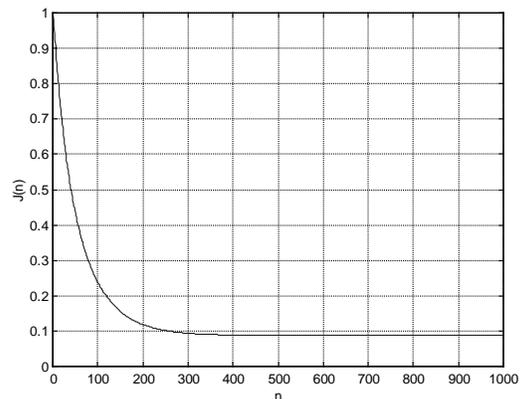


Figura 4.6: Evolução de J(n)

Sendo (4.58) um sistema dinâmico linear, a expressão para $\mathbf{w}(n)$ pode ser obtida analiticamente, e será do tipo exponencial. É preciso, para tanto, resolver a parte homogênea (transitória) e a particular (permanente). A Fig. 4.7 mostra claramente o caráter de combinação de exponenciais discretas das soluções:

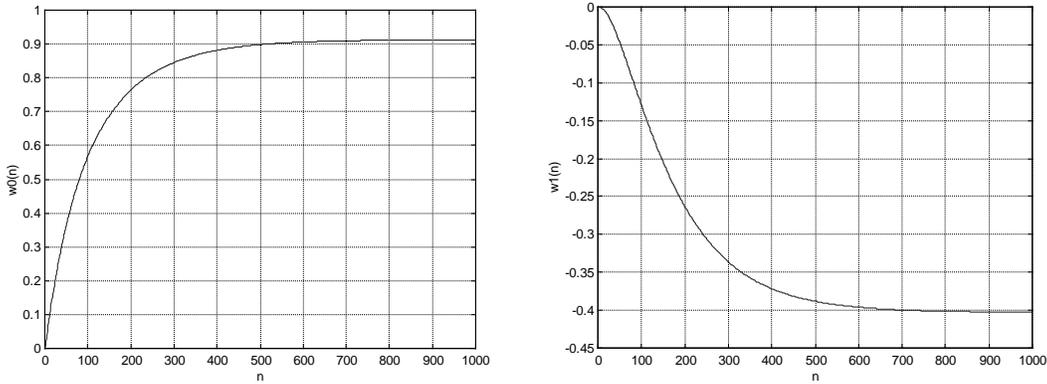


Figura 4.7: Evolução dos Pesos

As bases das exponenciais são os autovalores de \mathbf{J} , expressos em (4.60). Há um termo constante na solução, que corresponde ao regime permanente. Tal termo leva à solução de Wiener, expressa em (4.54), quando o transitório desvanece.

4.5.2. Derivação e Propriedades do Algoritmo LMS

Conforme vemos em (4.58), o emprego do método *steepest descent* exige o cálculo da autocorrelação do sinal de entrada e também da sua correlação cruzada com o sinal desejado. Na maioria das aplicações práticas do procedimento, o canal não é conhecido *a priori*, o que impede o cálculo analítico de \mathbf{R} . Pode-se, no entanto, realizar estimativas temporais da matriz, supondo a ergodicidade do processo de entrada.

Porém, tais processos requerem um custo computacional que pode se tornar indesejavelmente alto em certos casos. Um procedimento mais simples torna-se então bastante interessante.

Com base nestas idéias, surgiu o algoritmo LMS (Least Mean Square), proposto por Widrow e Hoff em 1960 [Haykin, 1996]. Tal algoritmo tem por base uma aproximação: a substituição da esperança por valores instantâneos. Isto dá origem às grandezas aproximadas:

$$\mathbf{R}_{ap} = \mathbf{x}(n) \cdot \mathbf{x}^H(n) \quad (4.70)$$

e

$$\mathbf{p}_{ap} = \mathbf{x}(n).d^*(n) \quad (4.71)$$

Aplicando (4.70) e (4.71) em (4.58), obtemos:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu.[\mathbf{x}(n).d^*(n) - \mathbf{x}^H(n).\mathbf{x}(n).\mathbf{w}(n)] \quad (4.72)$$

que após algumas manipulações nos leva a:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu.\mathbf{x}(n).[d^*(n) - \mathbf{x}^H(n).\mathbf{w}(n)] = \mathbf{w}(n) + \mu.e^*(n).\mathbf{x}(n) \quad (4.73)$$

Em (4.73), temos a expressão final do algoritmo LMS. Trata-se de um algoritmo extremamente simples, sendo portanto bastante atraente do ponto de vista prático. Tendo o equalizador p parâmetros, são necessárias $2p+1$ multiplicações complexas e $2p$ adições complexas por iteração. Portanto, a complexidade é do tipo $O(p)$ [Haykin, 1996].

As aproximações realizadas nos levam a uma versão *estocástica* do gradiente, pois o mesmo irá variar instantaneamente de acordo com $e(n)$ e $\mathbf{x}(n)$, que são processos aleatórios. Portanto, a expressão (4.73) refere-se a um *sistema dinâmico estocástico*, de natureza bastante diferente dos sistemas determinísticos que temos estudado.

Esta aproximação do vetor gradiente faz com que o algoritmo LMS tenha algumas diferenças de desempenho muito significativas em relação ao método *steepest descent*. A introdução de incertezas na adaptação impede que o sistema convirja para um ponto fixo (no caso a solução de Wiener), sendo o comportamento estacionário do vetor de estado caracterizado por uma variação aleatória em torno do ponto de equilíbrio do método *steepest descent*. A aleatoriedade introduzida pela aproximação estocástica pode ser vista como a introdução de um *ruído de gradiente*.

Diremos que o algoritmo LMS será convergente, no sentido da média quadrática, se:

$$J(n) \rightarrow J(\infty), \text{ se } n \rightarrow \infty \quad (4.74)$$

onde $J(\infty)$ é um valor fixo. A diferença entre este e o menor valor possível, J_{\min} , caracteriza o *desajuste (misadjustment)*, que quantifica a diferença de comportamento do método *steepest descent* e do LMS.

A convergência do algoritmo LMS está intimamente ligada ao passo de adaptação μ , como já podíamos antecipar a partir de (4.66). De fato, a escolha do passo está ligada a fatores que iremos discutir mais adiante.

Tentaremos agora mostrar alguns pontos relevantes da análise do LMS, seguindo a abordagem exposta em [Haykin, 1996].

Em primeiro lugar, definamos:

$$\boldsymbol{\varepsilon}(n) = \mathbf{w}(n) - \mathbf{w}_o \quad (4.75)$$

que indica o erro do vetor de pesos em relação ao valor ótimo de Wiener. Portanto, $\boldsymbol{\varepsilon}(n)$ será chamado de *vetor de erro de pesos*. Aplicando (4.75) a (4.73), temos:

$$\boldsymbol{\varepsilon}(n+1) = [\mathbf{I} - \mu \cdot \mathbf{x}(n) \cdot \mathbf{x}^H(n)] \cdot \boldsymbol{\varepsilon}(n) + \mu \cdot \mathbf{x}(n) \cdot e_o^*(n) \quad (4.76)$$

onde

$$e_o(n) = d(n) - \mathbf{w}_o^H \cdot \mathbf{x}(n) \quad (4.77)$$

Supondo um passo de adaptação suficientemente pequeno, podemos afirmar que a solução de (4.76) estará próxima da solução de:

$$\boldsymbol{\varepsilon}(n+1) = [\mathbf{I} - \mu \cdot \mathbf{R}] \cdot \boldsymbol{\varepsilon}(n) + \mu \cdot \mathbf{x}(n) \cdot e_o^*(n) \quad (4.78)$$

A condição sobre a pequenez de μ procede, pois neste caso a variação aleatória de $\boldsymbol{\varepsilon}(n)$ será bastante atenuada.

Discutamos também a chamada *teoria da independência*, baseada em quatro pontos:

1 – Os vetores de entrada $\mathbf{x}(0)$, ..., $\mathbf{x}(n)$ são estatisticamente independentes.

2 – Num instante n , o vetor $\mathbf{x}(n)$ é estatisticamente independente das amostras passadas $d(0), \dots, d(n-1)$.

3 – Num dado n , $d(n)$ depende de $\mathbf{x}(n)$, mas independe de $d(0), \dots, d(n-1)$.

4 – O vetor de entrada $\mathbf{x}(n)$ e o sinal desejado $d(n)$ consistem de variáveis aleatórias Gaussianas e mutuamente distribuídas para todos os valores de n .

Os quatro pontos acima, embora sejam simplificadores na análise do algoritmo, são bastante discutíveis. No primeiro ponto, assume-se independência entre os vetores de entrada. Ora, dois vetores sucessivos $\mathbf{x}(n)$ e $\mathbf{x}(n+1)$ compartilham $p-2$ amostras idênticas, sendo um deles proveniente do deslocamento do outro (com a eliminação de uma amostra antiga e a inclusão de uma nova). Portanto, para uma estrutura de filtragem temporal como a que enfocamos neste trabalho, o primeiro ponto não se sustenta. No caso de um array de antenas, onde as entradas não apresentam uma linha de atrasos, a primeira hipótese é mais razoável.

Os pontos 2 e 3 também podem ser contestados, caso o sinal desejado apresente dependência entre amostras sucessivas, e esteja de alguma forma relacionado ao sinal transmitido. O ponto 4 também é bastante limitador, como podemos imediatamente constatar.

Feitas as hipóteses simplificadoras, partamos para algumas definições que serão úteis.

Dada (4.78), podemos afirmar que haverá convergência, na média, se:

$$E[\boldsymbol{\varepsilon}(n)] \rightarrow \mathbf{0} \quad (4.79)$$

ou seja,

$$E[\mathbf{w}(n)] \rightarrow \mathbf{w}_o \quad (4.80)$$

O fato de haver convergência na média não significa um comportamento satisfatório, pois não elimina a possibilidade de variações substanciais em torno do nível médio. Um critério mais interessante é:

$$D(n) = E[|\varepsilon(n)|^2] \rightarrow \text{constante} \quad (4.81)$$

que leva em conta um desvio quadrático médio $D(n)$. Podemos ainda escrever, no mesmo espírito:

$$J(n) = E[|e(n)|^2] \rightarrow \text{constante} \quad (4.82)$$

Os critérios (4.81) e (4.82) estão intimamente ligados, o que permite que a análise se restrinja a um deles apenas.

Definamos a matriz de autocorrelação do vetor de erro dos pesos:

$$\mathbf{K}(n) = E[\varepsilon(n) \cdot \varepsilon^H(n)] \quad (4.83)$$

Aplicando (4.83) a (4.78), e aplicando a hipótese de independência discutida, temos:

$$\mathbf{K}(n+1) = (\mathbf{I} - \mu \cdot \mathbf{R}) \cdot \mathbf{K}(n) \cdot (\mathbf{I} - \mu \cdot \mathbf{R}) + \mu^2 \cdot \mathbf{J}_{\min} \cdot \mathbf{R} \quad (4.84)$$

A matriz de autocorrelação não irá convergir para $\mathbf{0}$, pelo segundo termo do segundo membro. Isto indica que haverá uma oscilação aleatória cuja potência é determinada por este termo.

Utilizaremos (4.84) para tentar definir o comportamento de $J(n)$. Para tanto, seja:

$$e(n) = d(n) - \mathbf{w}^H(n) \cdot \mathbf{x}(n) = d(n) - \mathbf{w}_o^H \cdot \mathbf{x}(n) - \varepsilon^H(n) \cdot \mathbf{x}(n) = e_o(n) - \varepsilon^H(n) \cdot \mathbf{x}(n) \quad (4.85)$$

Aplicando (4.85) a (4.21), temos:

$$J(n) = J_{\min} + E[\varepsilon^H(n) \cdot \mathbf{x}(n) \cdot \mathbf{x}^H(n) \cdot \varepsilon(n)] \quad (4.86)$$

Sabendo que o traço de um escalar é ele mesmo, e lembrando a hipótese de independência assumida, podemos escrever:

$$E[\boldsymbol{\varepsilon}^H(n) \cdot \mathbf{x}(n) \cdot \mathbf{x}^H(n) \cdot \boldsymbol{\varepsilon}(n)] = E\{\text{tr}[\boldsymbol{\varepsilon}^H(n) \cdot \mathbf{x}(n) \cdot \mathbf{x}^H(n) \cdot \boldsymbol{\varepsilon}(n)]\} = \text{tr}\{E[\mathbf{x}(n) \cdot \mathbf{x}^H(n) \cdot \boldsymbol{\varepsilon}(n) \cdot \boldsymbol{\varepsilon}^H(n)]\} = \text{tr}[\mathbf{R} \cdot \mathbf{K}(n)] \quad (4.87)$$

Portanto, (4.86) equivale, sob as hipóteses feitas, a:

$$J(n) = J_{\min} + \text{tr}[\mathbf{R} \cdot \mathbf{K}(n)] = J_{\min} + J_{\text{ex}}(n) \quad (4.88)$$

Sendo $\mathbf{R} \cdot \mathbf{K}(n)$ sempre definida positiva, podemos afirmar que o algoritmo LMS sempre produz um erro quadrático médio maior que o mínimo teórico J_{\min} .

Sabendo que:

$$\mathbf{Q}^H \cdot \mathbf{R} \cdot \mathbf{Q} = \Lambda \quad (4.89)$$

onde Λ é uma matriz diagonal contendo os autovalores de \mathbf{R} , e \mathbf{Q} é uma matriz cujas colunas são os autovetores associados aos mesmos.

Seja também:

$$\mathbf{Q}^H \cdot \mathbf{K}(n) \cdot \mathbf{Q} = \mathbf{U}(n) \quad (4.90)$$

Em geral, $\mathbf{U}(n)$ não é diagonal. A partir de (4.88), (4.89) e (4.90), temos:

$$\text{tr}[\mathbf{R} \cdot \mathbf{K}(n)] = \text{tr}[\mathbf{Q} \cdot \Lambda \cdot \mathbf{Q}^H \cdot \mathbf{Q} \cdot \mathbf{U}(n) \cdot \mathbf{Q}^H] = \text{tr}[\Lambda \cdot \mathbf{U}(n)] \quad (4.91)$$

onde usamos as propriedades $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$ e $\mathbf{Q} \cdot \mathbf{Q}^H = \mathbf{I}$.

A partir de (4.91), podemos escrever:

$$\mathbf{J}_{\text{ex}}(\mathbf{n}) = \sum_{i=0}^{p-1} \lambda_{Ri} \cdot u_i(\mathbf{n}) \quad (4.92)$$

onde $u_i(\mathbf{n})$, $i = 0, \dots, p-1$, são os elementos da diagonal da matriz $\mathbf{U}(\mathbf{n})$.

A partir de (4.89), (4.90) e (4.84), obtemos:

$$\mathbf{U}(\mathbf{n}+1) = (\mathbf{I} - \mu \cdot \Lambda) \cdot \mathbf{U}(\mathbf{n}) \cdot (\mathbf{I} - \mu \cdot \Lambda) + \mu^2 \cdot \mathbf{J}_{\text{min}} \cdot \Lambda \quad (4.93)$$

Pela dependência de (4.92), nos restringiremos aos termos diagonais de $\mathbf{U}(\mathbf{n})$. A partir de (4.93), obtemos:

$$u_i(\mathbf{n}+1) = (1 - \mu \cdot \lambda_{Ri})^2 \cdot u_i(\mathbf{n}) + \mu^2 \cdot \mathbf{J}_{\text{min}} \cdot \lambda_{Ri} \quad (4.94)$$

Definindo:

$$\mathbf{u}(\mathbf{n}) = [u_0(\mathbf{n}), \dots, u_{p-1}(\mathbf{n})]^T \quad (4.95)$$

e

$$\boldsymbol{\lambda} = [\lambda_{R0}, \dots, \lambda_{Rp-1}]^T \quad (4.96)$$

podemos reescrever (4.94) como:

$$\mathbf{u}(\mathbf{n}+1) = \mathbf{B} \cdot \mathbf{u}(\mathbf{n}) + \mu^2 \cdot \mathbf{J}_{\text{min}} \cdot \boldsymbol{\lambda} \quad (4.97)$$

onde $b_{ij} = (1 - \mu \cdot \lambda_{Ri})^2$, se $i = j$ ou $b_{ij} = \mu^2 \cdot \lambda_i \cdot \lambda_j$, se $i \neq j$.

A matriz \mathbf{B} é real, definida positiva e simétrica. É possível demonstrar [Haykin, 1996] que a solução de (4.97) é dada por:

$$\mathbf{u}(n) = \sum_{i=0}^{p-1} c_i^n \cdot \mathbf{g}_i \cdot \mathbf{g}_i^T \cdot [\mathbf{u}(0) - \mathbf{u}(\infty)] + \mathbf{u}(\infty) \quad (4.98)$$

onde c_i é o i -ésimo autovalor de \mathbf{B} e \mathbf{g}_i o i -ésimo autovetor.

Aplicando (4.98) em (4.92), temos:

$$J_{\text{ex}}(n) = \lambda^T \cdot \mathbf{u}(n) = \sum_{i=0}^{p-1} c_i^n \cdot \lambda^T \cdot \mathbf{g}_i \cdot \mathbf{g}_i^T \cdot [\mathbf{u}(0) - \mathbf{u}(\infty)] + \lambda^T \mathbf{u}(\infty) \quad (4.99)$$

de onde inferimos que:

$$J_{\text{ex}}(\infty) = \lambda^T \cdot \mathbf{u}(\infty) \quad (4.100)$$

Sabendo que:

$$J(n) = J_{\text{min}} + J_{\text{ex}}(n) \quad (4.101)$$

obtemos:

$$J(n) = \sum_{i=0}^{p-1} \gamma_i \cdot c_i^n + J_{\text{min}} + J_{\text{ex}}(\infty) \quad (4.102)$$

onde $\gamma_i = \lambda^T \cdot \mathbf{g}_i \cdot \mathbf{g}_i^T [\mathbf{x}(0) - \mathbf{x}(\infty)]$, $i = 0, \dots, p-1$.

Podemos inferir que o erro quadrático médio, expresso em (4.102), irá evoluir sem nenhum tipo de oscilação de natureza periódica amortecida, consistindo apenas de uma soma de exponenciais e de termos constantes. Vale lembrar que, em *uma única realização*, a curva irá apresentar oscilações aleatórias, que desaparecem quando se toma a esperança e obtém-se $J(n)$.

Os transitórios de $J(n)$ tendem a desaparecer com o aumento de n , ou seja, o LMS converge se e somente se a restrição (4.66) for obedecida. Para que isto aconteça, os autovalores de \mathbf{B} devem ser menores que 1 em magnitude. Pela própria definição de autovalor e autovetor, temos:

$$\mathbf{B} \cdot \mathbf{g} = c\mathbf{g} \quad (4.103)$$

que equivale a:

$$\sum_{j=0}^{p-1} b_{ij} \cdot g_j = c g_i, i = 0, \dots, p-1 \quad (4.104)$$

Lembrando as expressões para b_{ij} , expostas junto a (4.97), e aplicando-as a (4.104), obtemos:

$$(1 - \mu \lambda_{Ri})^2 \cdot g_i + \mu^2 \lambda_{Ri} \sum_{j=0}^{p-1} \lambda_{Rj} \cdot g_j = c g_i, i = 0, \dots, p-1 \quad (4.105)$$

Finalmente, resolvendo para g_i , chegamos a:

$$g_i = \frac{\mu^2 \cdot \lambda_{Ri}}{c - (1 - \mu \cdot \lambda_{Ri})^2} \cdot \sum_{j=0, j \neq i}^{p-1} \lambda_{Ri} \cdot g_j, i = 0, \dots, p-1 \quad (4.106)$$

Utilizemos o teorema de Perron-Frobenius [Haykin, 1996], que afirma que uma matriz definida positiva \mathbf{B} tem um único autovalor máximo (em magnitude). O autovetor correspondente terá apenas componentes positivas. Portanto, fazendo com que este máximo autovalor valha 1, e realizando simplificações, obtemos:

$$g_i = \frac{\mu}{2 - \mu \cdot \lambda_{Ri}} \cdot \sum_{j=0, j \neq i}^{p-1} \lambda_{Ri} \cdot g_j, i = 0, \dots, p-1 \quad (4.107)$$

No caso de convergência do LMS, segundo (4.82), podemos escrever:

$$J_{\text{ex}}(\infty) = J(\infty) - J_{\text{min}} \quad (4.108)$$

A partir de (4.94), obtemos:

$$u_i(\infty) = \frac{\mu \cdot J_{\text{min}}}{2 - \mu \cdot \lambda_{Ri}}, i = 0, \dots, p-1 \quad (4.109)$$

Substituindo em (4.100), temos:

$$J_{\text{ex}}(\infty) = \sum_{i=0}^{p-1} \lambda_{Ri} \cdot u_i(\infty) = J_{\text{min}} \cdot \sum_{i=0}^{p-1} \frac{\mu \cdot \lambda_{Ri}}{2 - \mu \cdot \lambda_{Ri}} \quad (4.110)$$

Se o somatório do segundo membro de (4.110) for menor que um, podemos afirmar que:

$$J_{\text{ex}}(\infty) < J_{\text{min}} \quad (4.111)$$

Uma importante medida de desempenho do algoritmo LMS é o já mencionado desajuste (*misadjustment*), definido como:

$$\Delta = \frac{J_{\text{ex}}(\infty)}{J_{\text{min}}} = \sum_{i=0}^{p-1} \frac{\mu \cdot \lambda_{Ri}}{2 - \mu \cdot \lambda_{Ri}} \quad (4.112)$$

Pela relação estabelecida em (4.111), temos que:

$$\Delta < 1 \quad (4.113)$$

O desajuste fornece o quanto a solução obtida através do LMS está afastada da solução de Wiener. Portanto, quanto maior o desajuste, mais pobre será a estimativa do

algoritmo. Um desajuste de 25 %, por exemplo, indica que o custo da solução do LMS está 25% acima de J_{\min} .

Percebemos que Δ depende do passo de adaptação μ . A escolha do passo, portanto, relaciona-se com a qualidade da estimativa fornecida pelo método. Supondo que o valor do passo é muito menor que o maior (em magnitude) autovalor de \mathbf{R} , podemos escrever:

$$\Delta \approx \frac{\mu}{2} \cdot \sum_{i=0}^{p-1} \lambda_{Ri} \quad (4.114)$$

Sendo a matriz \mathbf{R} Toeplitz, todos os elementos de sua diagonal principal são idênticos e iguais a $r(0)$. Assim,

$$\text{tr}[\mathbf{R}] = p \cdot r(0) = \sum_{k=0}^{p-1} E[|x(n-k)|^2] = \sum_{i=0}^{p-1} \lambda_{Ri} \quad (4.115)$$

A expressão (4.115) indica que a soma dos autovalores corresponde à potência na entrada do filtro.

Para fins práticos, podemos adotar uma restrição para μ mais conservadora que (4.66), pois a obtenção de λ_{\max} não é razoável na prática. Portanto, podemos adotar a condição:

$$0 < \mu < 2/\text{tr}[\mathbf{R}] \quad (4.116),$$

que pode ser mais facilmente determinada num caso prático, através do uso da equação (4.115), que relaciona o traço da matriz à potência na entrada do filtro.

4.5.3. A Escolha do Passo de Adaptação

Conforme pudemos constatar, o passo de adaptação tem um papel preponderante no projeto de um equalizador adaptativo. Dele dependem a velocidade de convergência, o

desajuste, a estabilidade, enfim, trata-se de um parâmetro que deve ser cuidadosamente escolhido.

Basicamente, trata-se de um compromisso clássico entre velocidade de convergência (que aumenta com o aumento de μ) e precisão da solução (que diminui com o aumento de μ , como vimos na seção anterior). Isto levando sempre em consideração a faixa de valores de passo para a qual há convergência. Portanto, cada caso de aplicação deve ser analisado, para que se opte por dar mais ênfase a uma ou outra característica, que podemos dizer, mais se destacam no “dilema” da escolha de parâmetros de um filtro linear adaptativo.

Em algoritmos lineares, como o são os que derivam da filtragem de Wiener, a variação do passo de adaptação leva a dois cenários: convergência para um ponto de equilíbrio (próximo ou igual à solução de Wiener), ou divergência para o infinito. Em algoritmos baseados em outros critérios, há uma riqueza de cenários maior, como veremos no capítulo 6.

Voltemos ao modelo MA(1) para um exemplo envolvendo o uso do algoritmo LMS.

Exemplo 4.3 – Aplicação do Algoritmo LMS

Voltando ao modelo MA(1), que já enfocamos em duas outras oportunidades, apliquemos agora o algoritmo LMS para buscarmos o filtro ótimo. Utilizaremos os parâmetros $\alpha = 0.6$, $\mu = 0.01$ e $\mathbf{w}(0) = [0 \ 0]^T$. Na Fig. 4.8, encontra-se a trajetória percorrida pelo estado do sistema, traçada juntamente com as curvas de nível, para uma melhor referência.

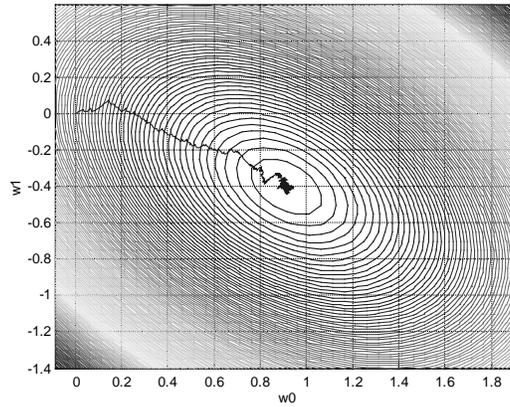


Figura 4.8: Trajetória do Vetor de Pesos no LMS

Em comparação com a Fig. 4.5, percebemos claramente a influência do ruído de gradiente na trajetória, responsável por suas oscilações e pela flutuação em torno da solução de Wiener, como discutimos anteriormente. Na Fig. 4.9, podemos analisar a evolução do erro quadrático médio na realização corrente do algoritmo LMS.

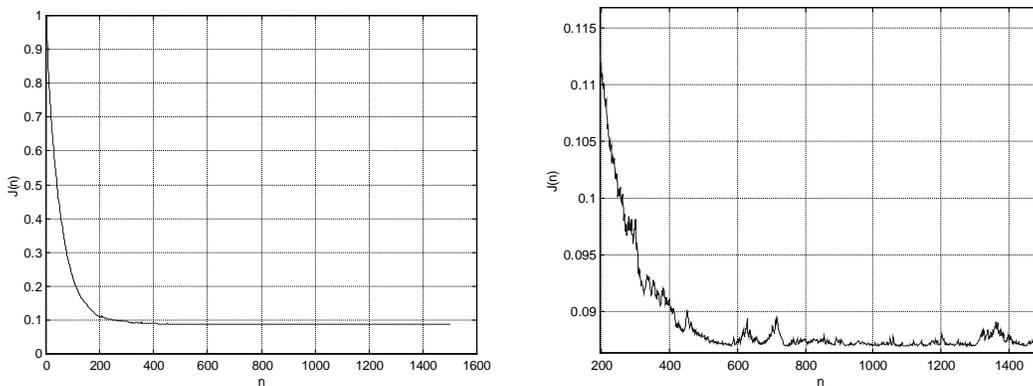


Figura 4.9: Erro Quadrático Médio e Ampliação

A evolução do erro quadrático médio tem um comportamento semelhante ao apresentado na Fig. 4.6. Porém, como era de se esperar, no caso do algoritmo LMS há flutuações na região de regime permanente da curva. Tais flutuações também são consequência da oscilação ruidosa verificada na trajetória do vetor de pesos, comentada no parágrafo anterior.

Utilizando (4.112), obtemos que o desajuste vale:

$$\Delta = 0.0137 \quad (4.117)$$

um valor compatível com uma análise mais detalhada da Fig. 4.9.

Na Fig. 4.10, podemos observar a evolução de $w_0(n)$ e $w_1(n)$. Podemos perceber que tal evolução é semelhante à exibida na Fig. 4.7, a menos das ressalvas feitas quanto à aproximação do gradiente determinístico, que explicam as flutuações aleatórias inexistentes no caso do método *steepest descent*.

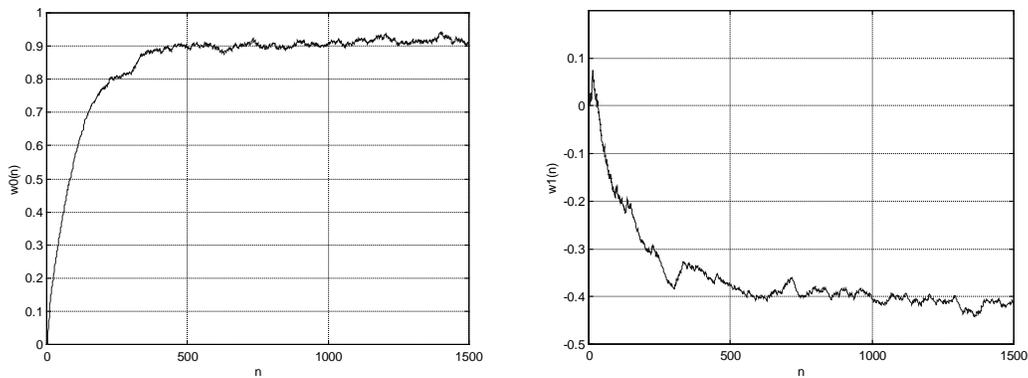


Figura 4.10: Evolução dos Pesos

Por fim, vale notar que uma única realização do algoritmo LMS leva a uma curva de erro quadrático com bastante “ruído”. Uma média entre diferentes realizações vai gradativamente reduzindo tais aleatoriedades, levando a estimativas razoáveis do custo (erro quadrático médio). A Fig. 4.11 confronta uma única realização do erro com uma média feita com 30 realizações.

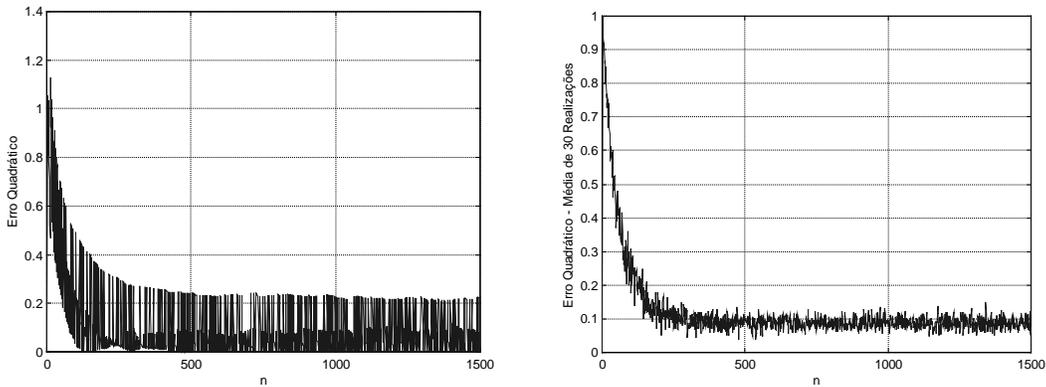


Figura 4.11: Erro Quadrático (em uma e trinta realizações)

4.6. Outros Métodos Baseados na Teoria de Wiener

O algoritmo LMS não é a única forma de se obter, de forma adaptativa, estimativas da solução de Wiener. Foram desenvolvidas algumas outras técnicas, que comentaremos brevemente nesta seção.

O algoritmo RLS (Recursive Least Squares) é talvez o mais conhecido dos algoritmos supervisionados, excetuando-se o LMS. Fundamenta-se na minimização de uma função custo que é a soma ponderada dos valores do erro quadrático. Ou seja, a definição não envolve esperanças matemáticas como no caso da teoria de Wiener. A partir desta definição, chega-se a equações muito semelhantes às de Wiener-Hopf, com a diferença de que não estarão envolvidos o vetor de correlação cruzada e a matriz de autocorrelação, mas sim suas estimativas temporais. Tais estimativas são atualizadas iterativamente, o que é bastante relevante do ponto de vista computacional. Para se evitar a constante inversão de matrizes, que são muitas vezes de grande dimensão, emprega-se um engenhoso resultado algébrico denominado *lema de inversão de matrizes* ou *identidade de Woodbury* [Haykin, 1996]. Através deste resultado, consegue-se também estimar a inversa iterativamente. Isto conclui a obtenção do algoritmo.

Em comparação com o LMS, o RLS apresenta um maior custo computacional, mas tem uma velocidade de convergência tipicamente uma ordem de grandeza maior que a do LMS [Haykin, 1996]. A velocidade de convergência do RLS independe das estatísticas de

segunda ordem do sinal, o que não é verdade para o LMS. Por fim, o valor de $J_{ex}(n)$ converge para zero quando n tende a infinito, o que não ocorria para o LMS.

Baseados nestes dois algoritmos, surgiram métodos modificados, baseados por exemplo em fatoração de matrizes, como a fatoração QR, ou em procedimentos de normalização (como o chamado LMS normalizado).

4.7. *Introdução aos Métodos Autodidatas*

Conforme tivemos a oportunidade de constatar, os métodos baseados na teoria do filtro ótimo de Wiener exigem o conhecimento de uma sequência $d(n)$, correspondente a um sinal desejado. Não é difícil perceber que dispor de amostras do sinal transmitido no receptor é algo complicado, pois o objetivo da comunicação é justamente fazer com que a informação em um transmissor possa ser enviada a um receptor. Portanto, para se gerar a sequência de treinamento do equalizador, faz-se necessária a transmissão de informação já conhecida no destino. O emprego do sistema no envio de mensagens que servem apenas para o ajuste do equalizador pode levar à necessidade de um aumento na taxa de transmissão, o que significa maior ocupação espectral, por exemplo. Há ainda sistemas em que se torna impossível a adição de tais sequências, por impossibilidade de projeto, inviabilizando esses procedimentos de equalização. Um exemplo clássico, e que nos faz entender a motivação inicial de Godard ao propor a classe de algoritmos que hoje leva o seu nome (iremos discutir este algoritmo com detalhes no capítulo 5), é o de uma *rede de computadores multiponto* [Haykin, 1996, Godard, 1980 e da Rocha, 1996]. Supondo uma arquitetura onde haja uma unidade de controle e diversos terminais, torna-se problemático o retreinamento de um destes últimos elementos, que só podem operar quando designados pela unidade de controle. O problema decorre de possíveis variações no canal ou do fato de o terminal não estar ligado durante a sincronização inicial da rede.

Outro caso em que pode não ser prático o uso de sequências de treinamento é num ambiente de comunicações móveis, onde a faixa de transmissão é severamente limitada, e a presença de desvanecimentos severos pode tornar muito difícil a obtenção de um adequado sincronismo.

Enfim, por motivos como estes, torna-se atraente uma técnica que permita o ajuste iterativo do equalizador sem a necessidade de transmissão de uma sequência de treinamento. Tais técnicas são denominadas *autodidatas* ou *cegas*.

Podemos dizer que a equalização autodidata está ligada a um cenário de desconvolução. A razão para isto está em (4.5), onde percebe-se que o sinal transmitido é convoluído com a resposta ao impulso do canal, para dar origem ao sinal recebido. Portanto, para recuperarmos a informação transmitida faz-se necessário realizar uma operação inversa à convolução, a *desconvolução*. Esta idéia não se limita a canais de comunicação, mas a todo e qualquer sistema modelado como LIT.

Dispondo do sinal de saída do canal (sinal recebido), há dois cenários nos quais pode ocorrer a desconvolução: com ou sem conhecimento do canal. Este último caso é o mais interessante para nós, pois nele está a raiz do problema da equalização autodidata.

4.7.1. Técnicas Mais Utilizadas

Basicamente, a desconvolução autodidata tem sido realizada segundo três paradigmas, a saber [Haykin, 1996]:

- 1) *Algoritmos Baseados Explicitamente em Estatísticas de Ordem Superior*, que usam *cumulantes* de ordem maior que dois e suas contrapartidas frequenciais, os *poliespectros*. Tais grandezas, ao contrário da autocorrelação, preservam a informação de fase, informação esta que “faz a diferença” quando se lida com canais de fase não-mínima.
- 2) *Algoritmos Baseados Implicitamente em Estatísticas de Ordem Superior*, grupo no qual se incluem os algoritmos de Busgang, classe da qual o CMA faz parte. Trata-se de algoritmos que utilizam um estimador não-linear para suprir a ausência de um sinal desejado. São algoritmos que apresentam um comportamento dinâmico bem mais rico que o dos algoritmos supervisionados, como procuraremos mostrar neste trabalho.

3) *Algoritmos Baseados em Estatísticas Cicloestacionárias*, que utilizam estatísticas cicloestacionárias de ordem dois do sinal recebido. A razão para o emprego de tais estatísticas está nas propriedades das modulações baseadas em uma portadora senoidal.

O leitor interessado em aprofundar-se no estudo de qualquer um dos três paradigmas pode encontrar em [Haykin, 1996] referências dos trabalhos seminais relativos a cada uma das áreas.

O tema deste trabalho centra-se no segundo paradigma, o mais utilizado e também o de mais simples implementação e operação.

Basicamente, o maior interesse das técnicas de equalização cega reside no projeto de equalizadores para canais de fase não-mínima. A razão para isto é que para se equalizar um canal de fase mínima é preciso simplesmente inverter a resposta de amplitude do canal, o que se faz através de um filtro branqueador, enquanto que para canais de fase não-mínima torna-se imperativo que informações de fase sejam utilizadas para se ter um desempenho satisfatório.

Muito do conhecimento que temos sobre a equalização cega está sintetizado nos teoremas de Benveniste-Goursat-Ruget [Benveniste et al., 1980] e de Shalvi-Weinstein [Shalvi e Weinstein, 1990]. O primeiro teorema estabelece que uma condição necessária e suficiente para a ocorrência de equalização ZF é a igualdade da distribuição de probabilidade entre o sinal transmitido e a saída do equalizador. O segundo “enfraquece” a condição e dá como condição necessária e suficiente para a equalização ZF a igualdade entre estatísticas de segunda ordem e de alguma ordem par superior a dois. Vale enfatizar que estes teoremas são válidos para sinais transmitidos não-gaussianos. Tais trabalhos são dois pilares teóricos das técnicas que empregam estatísticas de ordem superior.

Após esta visão geral de filtros adaptativos, com ênfase num estudo mais profundo das técnicas supervisionadas, convém que agora nos voltemos para o algoritmo que é o tema desta tese: o algoritmo do módulo constante (CMA), uma das técnicas de Bussgang, que “substituem” o sinal desejado por uma estimativa não-linear, produzida a partir da saída do equalizador.

5. O Algoritmo do Módulo Constante

Após um breve panorama geral sobre os diversos paradigmas de equalização autodata, iremos neste capítulo analisar mais especificamente a técnica mais relacionada às principais motivações desta tese: o algoritmo do módulo constante (CMA). Procuraremos expor as principais descobertas sobre os comportamentos estático e dinâmico da aplicação deste algoritmo, desde sua proposta inicial até o presente. Sendo o assunto bastante vasto, é importante apresentá-lo da forma mais didática possível, dividindo-o segundo as diferentes abordagens que pudemos encontrar na literatura e preservando, sempre que for conveniente, a linha histórica na qual evoluíram os resultados. Assim, as treze primeiras seções enfocam a análise do CMA segundo a abordagem mais tradicional, baseada na função custo e em suas derivadas, no espaço dos parâmetros do equalizador.

Da seção 5.14 à 5.19, tratou-se de um novo paradigma que classificamos aqui como “abordagem geométrica”. Tal metodologia concentra seus resultados no chamado espaço conjunto, resultante da convolução entre os elementos da resposta do canal e os parâmetros do equalizador. Na análise dos pontos críticos do critério neste novo espaço, utiliza-se sobretudo determinadas regiões oportuna e elegantemente definidas para este fim. Destacam-se ainda neste setor do capítulo as demonstrações de equivalência entre os critérios do módulo constante e o de Shalvi-Weinstein.

Finalmente, os dois últimos itens referem-se à análise dinâmica do CMA, já antecipando a discussão do capítulo 6, onde se concentram as contribuições mais inovadoras desta tese.

5.1. *A Proposta Original de Godard*

Embora ainda sem receber este nome, o algoritmo do módulo constante apareceu pela primeira vez na literatura técnica em novembro de 1980, num artigo de autoria de Dominique Godard [Godard, 1980]. Godard, à época, trabalhava no Centro de Estudos e Pesquisas da IBM, em Le Gaude, França. Sendo uma das áreas de interesse do autor a transmissão digital, não é de se estranhar que o algoritmo foi proposto dentro de um contexto de comunicação por redes multiponto, onde pode ser impraticável dispor de

sequências de treinamento em certas circunstâncias e é imperativo que se tenha uma convergência rápida do algoritmo de equalização empregado. Outro ponto muito relevante para a aplicação motivadora de Godard é que o método por ele proposto não exigia recuperação da fase da portadora na entrada do equalizador, sendo que tal tarefa poderia ser realizada na saída do mesmo, tendo por base o algoritmo da decisão direta.

Basicamente, Godard expressa os objetivos do método como sendo “reduzir os efeitos das distorções do canal, de tal forma que as decisões do receptor tornem-se seguras o suficiente para se empregar algoritmos de gradiente convencionais de decisão direta”. Para isto, é proposta uma função custo capaz de “caracterizar o nível de IIS na saída do equalizador independentemente da constelação e da fase da portadora”. Sem preservar a notação original do autor, o que certamente daria margem a equívocos, tal função custo se expressa por:

$$J^{(p)} = E[|y(n)|^p - R_p]^2 \quad (5.1)$$

A esta função, Godard dá o nome de *dispersão de ordem p*. A constante R_p é positiva, e seu valor será derivado posteriormente.

Se analisarmos atentamente (5.1), podemos perceber alguns aspectos muito relevantes. O primeiro é que, de fato, a fase do sinal recebido não importa para o cálculo da dispersão, pois apenas o módulo da saída é levado em conta. O segundo é que podemos encarar este critério, para $p = 2$, como um “recuperador de módulo constante”, caso a constelação transmitida tenha esta característica. Daí o nome CMA para esta versão do algoritmo, que como já frisamos não foi usado originalmente por Godard, embora ele tenha percebido esta propriedade. A análise começa com a imposição de duas restrições sobre os sinais transmitidos, a saber, que pertencessem a uma constelação com uma simetria tal que:

$$E[s^2(n)] = 0 \quad (5.2)$$

além da usual condição de estacionariedade e decorrelação entre os símbolos.

A escolha de R_p é feita a partir da análise do gradiente de $J^{(p)}$. Após este cálculo, Godard observa que a variação de R_p provoca uma mudança de amplitude na solução

ótima. Assim sendo, ele observa que esta constante é um controle de amplificação do equalizador. Forçando uma condição ótima, na qual a saída do equalizador é igual ao sinal transmitido a menos de uma distorção de fase, é obtido o valor de R_p que proporciona tal situação, a saber:

$$R_p = \frac{E[|s(n)|^{2p}]}{E[|s(n)|^p]} \quad (5.3)$$

A partir destas considerações e também de uma aproximação estocástica para o gradiente, chega-se então à expressão clássica do algoritmo:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \cdot e^*(n) \cdot \mathbf{x}(n) \quad (5.4)$$

onde

$$e(n) = y(n) \cdot |y(n)|^{p-2} \cdot (R_p - |y(n)|^p) \quad (5.5)$$

e μ é um passo de adaptação.

A partir da proposta, Godard já restringe os seus testes aos valores $p = 1$ e $p = 2$, por problemas com aritmética de precisão finita. De fato, estes dois algoritmos foram, desde o início de seus trabalhos, os mais utilizados.

Godard realiza então a primeira análise da convergência do método, na qual já se obtém algumas conclusões importantes.

Nesta análise, o autor já observa que a função dispersão por ele proposta não é convexa. Isto quer dizer que a inicialização influi profundamente no desempenho do algoritmo, pois há mínimos locais na função custo. É assumido um comprimento infinito para o equalizador, e soluções de equilíbrio no espaço conjunto (canal + equalizador) são procuradas. São procuradas soluções ZF, consideradas como mínimos globais. Godard também expõe, pela primeira vez, uma inicialização baseada em um único “*spike*”, idéia mais explorada por Foschini, como veremos na seção 5.4. Os resultados obtidos mostram

que o algoritmo obteve soluções próximas à de Wiener através de um processo de convergência considerado lento.

5.2. *Treichler e Agee Entram em Cena*

Um artigo de 1983 [Treichler e Agee, 1983], escrito por Treichler e Agee, figura também entre os trabalhos pioneiros sobre o CMA.

Basicamente, os autores propõem o algoritmo, que recebe pela primeira vez o nome de algoritmo do módulo constante, para compensar os efeitos de multipercursos seletivos em frequência e de interferência em sinais com envelope constante. O termo “envelope constante” estende o domínio de aplicabilidade da técnica também a sinais FM, o que se liga ao problema motivador dos autores, a propagação de sinais pelo ar. Vale lembrar que a estrutura do sinal FM está muito ligada à dos sinais do tipo PSK, mais relacionados ao trabalho de Godard.

É explicado com cuidado o problema dos multipercursos seletivos em frequência, capazes de provocar efeitos indesejados de AM e distorções de fase, sendo proposta então uma função custo capaz de mitigar estes danos. Tal função é justamente (5.1), com $p = 2$ e a suposição de módulo unitário, ou seja, $R_2 = 1$. Há ainda a adição de um fator de escala. Foi utilizada a aproximação estocástica, sendo adotada a notação de (5.4), que é semelhante à do LMS. Os autores consideram esta notação muito útil por permitir que se use o mesmo hardware para ambos os algoritmos, o que à época significava uma relevante economia de recursos.

É empreendida então uma análise teórica do algoritmo. Em primeiro lugar é analisada a forma da superfície de erro do CMA, sendo tomadas como variáveis de estudo as partes real e imaginária da saída do equalizador. As figuras obtidas indicam a existência de um vale circular, representando a condição de módulo unitário, e de um ponto de máximo, o que é bastante razoável tendo em vista a estrutura matemática da expressão iterativa. Tal análise, porém, não significa que o vale vá ser sempre atingido. Por exemplo, no caso clássico de equalização de um canal FIR por um filtro FIR, jamais, mesmo no mínimo global, será atingido tal vale. Portanto, analisar o algoritmo através de $y(n)$ é um procedimento bastante limitado. Em seguida, os autores mostram que, para que $y(n)$ tenha

módulo unitário, é preciso que a resposta conjunta canal + equalizador, $c(n)$, tenha apenas um termo não-nulo. A prova é válida para os tipos de sinais usualmente empregados em telecomunicações.

Após mostrar que um deslocamento de fase no vetor de pesos não altera o custo da solução e que qualquer atraso de grupo é equivalente do ponto de vista da recuperação do módulo unitário, ou seja, qualquer configuração ótima (ZF) atingida por $\mathbf{c}(n)$ é equivalente, realiza-se uma comparação da técnica com alguns paradigmas clássicos. Em primeiro lugar, a comparação é feita com os estimadores MAP e ML, sendo apontadas importantes diferenças conceituais entre as técnicas. Em seguida, os autores procuram estabelecer uma relação entre o filtro de Wiener e a solução obtida pela função custo CM, sendo obtidas algumas conclusões que apontam no sentido da equivalência, desde que sejam satisfeitas determinadas hipóteses estatísticas. Análises mais profundas foram realizadas posteriormente, como veremos ao longo do capítulo.

Os autores mostram resultados para um canal com um multipercurso, sendo que o algoritmo comporta-se bem na presença e na ausência de ruído, apresentando uma convergência qualificada como “moderadamente rápida”. No fim do trabalho os autores mencionam o algoritmo de Godard, no contexto do uso da técnica para sinais digitais com módulo constante.

5.3. *Análise de Convergência: Equalizadores Duplamente Infinitos*

Os primeiros trabalhos sobre a convergência do algoritmo do módulo constante fundavam-se principalmente em um equalizador com comprimento duplamente infinito. Isto quer dizer que sua resposta ao impulso seria:

$$h_w = \{ \dots, w_{-1}, w_0, w_1, \dots \} \quad (5.6)$$

Trata-se de um filtro não realizável na prática, não-causal e com infinitos parâmetros. Porém, é uma configuração que permite uma análise bastante simplificada no domínio conjunto canal + equalizador \mathbf{c} . Neste tipo de análise, tem particular interesse

configurações de \mathbf{c} com um único elemento não-nulo, correspondentes às configurações zero-forcing (seção 4.1). Em [Foschini, 1985], Foschini analisou pontos de equilíbrio no domínio \mathbf{c} que tinham M elementos não-nulos, através da matriz hessiana da função $J^{(2)}$ de Godard. Tal matriz é definida por:

$$\mathbf{H}_{\text{ess}} = \frac{\partial^2 J}{\partial c_m \partial c_n} \quad (5.7)$$

Basicamente, foram apontadas três possibilidades para os pontos de equilíbrio:

- 1) $M = 0$, correspondendo a um máximo local (\mathbf{H}_{ess} é semi-definida negativa);
- 2) $M = 1$, correspondendo a mínimos globais (\mathbf{H}_{ess} é semi-definida positiva);
- 3) $M > 1$, correspondendo a pontos de sela (\mathbf{H}_{ess} é indefinida);

Ao apontar estas características, Foschini reconhece que, de fato, não é o interesse primordial de uma análise de convergência se basear no espaço \mathbf{c} , pois os parâmetros de relevância são os pesos do equalizador. Porém, ele estende os resultados que acabamos de discutir ao espaço do parâmetros do equalizador com base em algumas considerações matemáticas envolvendo a características da convolução e a expressão da função custo de Godard.

Pudemos constatar então que a análise do equalizador duplamente infinito nos leva a um cenário em que os pontos de equilíbrio estáveis são mínimos globais segundo o critério CM. Em tal condição, haveria poucos problemas de operação, pois os demais pontos não seriam atratores e todas as condições zero-forcing se equivaleriam, do ponto de vista da teoria de equalização. Mas resta uma pergunta muito pertinente: até que ponto estas conclusões valem para um equalizador finito, que é de fato o que se irá utilizar numa aplicação prática?

5.4. *Primeiras Extensões para Comprimento Finito*

As conclusões de Foschini [Foschini, 1985] sobre a extensão dos resultados mostrados para um equalizador com comprimento finito não são fundamentadas em argumentos matemáticos sólidos, como o próprio autor reconhece. Apesar disto, é importante que enfoquemos dois aspectos de sua análise:

1 – A idéia de que buscar periodicamente uma configuração onde o coeficiente de referência (de maior magnitude) do equalizador esteja no centro de massa do vetor de parâmetros poderia ajudar a “reproduzir” uma situação de comprimento infinito do equalizador. Este é o conceito de “*center-spike*”, sendo importante frisar que configurações com um único parâmetro não-nulo já vinham sendo analisadas desde o trabalho de Godard. Teremos outras ocasiões para discutir esta idéia ao longo do capítulo.

2 – A noção de que o aumento do número de taps do equalizador poderia permitir que se tivesse um comportamento arbitrariamente próximo do comportamento do equalizador duplamente infinito.

As conclusões de Foschini são altamente intuitivas, e por muito tempo foram aceitas sem maiores análises. Alguns anos depois, porém, novos trabalhos lançaram importantes luzes sobre estes e outros aspectos de convergência do CMA, revolucionando o que se entendia até então a respeito do comportamento deste algoritmo.

5.5. *Constatação da Convergência para Mínimos Sub-Ótimos*

Como vimos, as análises mais rigorosas de convergência realizadas por Foschini e Godard limitaram-se a uma situação nada prática: o emprego de um equalizador “duplamente infinito”. Porém, garantir convergência global para este caso não garante que tal característica se manifeste em um modelo finito.

De fato, em 1991, Ding, Kennedy, Anderson e Johnson [Ding et al., 1991] demonstraram que um equalizador finito ajustado pelo critério de Godard poderia convergir para situações em que haveria nenhuma redução de IIS, ou seja, mínimos locais da função custo (5.1). Isto já havia sido demonstrado para o algoritmo de Sato em trabalhos anteriores, como mencionado em [Ding et al., 1991]. Mas análises anteriores do CMA ainda não haviam contemplado tais pontos. Em [Ding et al., 1991] é mostrado que, ao contrário de técnicas supervisionadas e do caso duplamente infinito, o critério CM não apresenta convergência global em equalizadores implementáveis na prática. Isto torna o desempenho do algoritmo altamente dependente da inicialização do estado do filtro, o que passa a ser um importante problema adicional de projeto.

O cenário da demonstração dos resultados é o de um canal AR equalizado por um filtro FIR, sem a presença de ruído aditivo. A primeira constatação é que o mínimo global da função custo necessariamente implica a remoção da IIS. Podemos intuir isto, pois temos uma condição na qual o equalizador pode inverter perfeitamente o canal, ou seja, há a possibilidade de se atingir uma configuração ZF. Tal configuração será, para a hipótese de independência feita sobre os símbolos transmitidos, idêntica à ótima do ponto de vista do critério de Wiener. É razoável imaginar que a mesma coincidirá com o mínimo global de $J^{(2)}$. Além disto, como a equalização é perfeita, temos que $J^{(2)}(\mathbf{w}_{\text{glob}}) = 0$, onde \mathbf{w}_{glob} é o mínimo global de $J^{(2)}$.

A equivalência entre o mínimo global de $J^{(2)}$ e a solução ZF é mostrada através de dois resultados: o primeiro estabelece que a convolução entre a sequência de entrada e a resposta conjunta só terá um módulo constante e igual a um valor ρ para todos os instantes de tempo se a condição ZF tiver sido atingida. Em seguida, mostra-se que isto equivale à minimização de $J^{(2)}$, o que conclui o raciocínio.

5.5.1. Obtenção dos Mínimos e Análise de Estabilidade

Para a obtenção dos pontos de mínimo local do equalizador, é considerado um canal AR(N) com a seguinte função de transferência:

$$H(z) = \frac{1}{1 + \alpha \cdot z^{-N}} \quad (5.8)$$

O equalizador é considerado com sendo MA(N), sendo portanto, a condição de equalização perfeita:

$$\mathbf{w}_{\text{glob}} = \pm [1 \ 0 \ \dots \ \alpha]^T \quad (5.9)$$

Reconhecendo ser muito difícil obter os demais valores de \mathbf{w} que anulam o gradiente de $J^{(2)}$, os autores restringem a busca de outros pontos de equilíbrio a configurações do tipo:

$$\mathbf{w}_{\text{eq}} = \pm \mathbf{K}_{\text{eq}} \cdot [0 \ 0 \ \dots \ 1]^T \quad (5.10)$$

Impondo a condição de gradiente nulo a estas soluções e tecendo algumas considerações de modelo, chega-se ao valor de \mathbf{K}_{eq} para o qual \mathbf{w}_{eq} é ponto crítico de $J^{(2)}$:

$$\mathbf{K}_{\text{eq}} = \left(\frac{E\{|x(n)|^p\}}{E\{|x(n)|^{2p}\}} \cdot \mathbf{R}_p \right)^{1/p} \quad (5.11)$$

Uma vez obtida esta classe de pontos críticos, resta ainda saber se eles são mínimos da função, ou seja, capazes de atrair trajetórias no espaço de estados. Para isto, utiliza-se a matriz hessiana de $J^{(2)}$, definida em (5.7). Calculando a matriz nos pontos de (5.10), percebe-se que eles serão estáveis desde que a constelação de entrada satisfaça a condição de simetria:

$$3 \cdot [E(|s(n)|^2)]^2 > E(|s(n)|^4) \quad (5.12)$$

Os resultados de simulações apresentados confirmaram o proposto pela teoria para o caso de modulações 2 e 8-PAM, com $N = 1$.

5.5.2. Extensões dos Resultados

Uma vez realizado todo o desenvolvimento para o modelo de canal apresentado, fica ainda a pergunta: os mínimos locais irão influir no comportamento de equalizadores FIR quando houver ruído? E quando o canal não for da forma (5.8)?

Através de simulações, mostra-se em [Ding et al., 1991] que a adição de ruído com potência 10 dB menor que a do sinal não altera significativamente a bacia de atração dos mínimos locais, embora haja uma modificação na forma de $J^{(2)}$, assim como na posição e na profundidade dos mínimos. Também não se mostra verdadeira, para o valor de SNR citado, a conjectura de que a presença de ruído contribui inequivocamente para o escape dos mínimos locais, pois o ruído também pode vir a provocar escape dos mínimos globais, que são mais “rasos” em tal caso que no caso sem ruído.

Por fim, os resultados são estendidos a um modelo de canal do tipo MA(2) em que atração para mínimos locais é confirmada por meio de simulações.

Resultados análogos são deduzidos ainda para constelações complexas, confirmando o que se obteve para sinais binários.

5.6. Sobre a Posição dos Mínimos de $J^{(2)}$

Como vimos, a presença de mínimos locais é uma característica inerente ao critério CM, pelo menos em equalizadores realizáveis. Portanto, passa a ser um problema fundamental buscar estratégias que ajudem a evitar soluções indesejadas, melhorando o desempenho do algoritmo.

Em 1992, um artigo de Ding e Kennedy [Ding e Kennedy, 1992] enfocou a análise de posição dos mínimos de um equalizador ajustado por dois critérios: o de Benveniste-Goursat-Ruget (BGR) [Benveniste et al., 1980] e o de Godard [Godard, 1980]. Concentraremos-nos nos resultados obtidos para este último, que é o tema deste trabalho, aplicado no âmbito de um canal do tipo MA com resposta ao impulso $h(n)$.

Escrevamos a função custo de Godard como:

$$J^{(p)} = E\{\psi[y(n)]\} \quad (5.13)$$

onde $\psi[y(n)] = (1/2p).E\{|y(n)|^p - R_p\}^2$.

Na configuração de equilíbrio global, com parâmetros de projeto adequados, temos que:

$$\mathbf{x}^H(n).\mathbf{w}_{glob} = s(n) \quad (5.14)$$

Tomando uma configuração $\mathbf{w}_{dir} = \gamma.\mathbf{w}_{glob}$, podemos escrever $J^{(p)}$ para o caso M-PAM como:

$$J^{(p)}(\mathbf{w}_{dir}) = \frac{2}{M} \sum_{i=1}^{M/2} \psi[(2i-1).\gamma] \quad (5.15)$$

Quando se calcula $\psi[\mathbf{x}^H(n).\mathbf{w}_{dir}]$, percebe-se que a função ψ precisa ser unimodal [Ding e Kennedy, 1992]. Isto porque a multi-modalidade desta função acarretaria a presença de outros mínimos na direção de \mathbf{w}_{dir} , o que não faz sentido para uma função sensível ao módulo. Uma condição adicional, nada restritiva em se tratando de técnicas de Bussgang, é que $\psi(y) \rightarrow \infty$ quando $|y| \rightarrow \infty$.

Analisando vetores de pesos do tipo $\mathbf{w} = \rho.\boldsymbol{\varphi}$, percebe-se que $J^{(p)}$ atinge valores mínimos em:

$$|\rho_1(\boldsymbol{\varphi})|^p = R_p \cdot \frac{E[|\mathbf{x}^H(n).\boldsymbol{\varphi}|^p]}{E[|\mathbf{x}^H(n).\boldsymbol{\varphi}|^{2p}]} \quad (5.16)$$

e em seus simétricos com relação à origem, devido à simetria do critério CM.

Todos os pontos de mínimo e de sela estão na variedade descrita por (5.16). O máximo está em $\rho = 0$. Para analisar as características desta variedade, realiza-se uma normalização em $\boldsymbol{\varphi}$ de forma a se ter:

$$\sum_{k=0}^{\infty} |t(n)|^2 = 1 \quad (5.17)$$

onde $t(n) = h(n) * \varphi(n)$.

É então demonstrado que $\rho_1(\varphi)$ atinge o seu máximo quando a IIS é completamente eliminada, ou seja, quando $t(n)$ tem um único elemento não-nulo. De forma geral, podemos expressar esta conclusão pelo resultado:

$$\rho_1(\varphi) \leq 1 \quad (5.18)$$

com igualdade na configuração ZF. Na obtenção destes resultados foram feitas algumas considerações sobre a simetria das constelações transmitidas.

Portanto, para o CMA, todos os mínimos e selas da função custo estarão sobre esferas com raios $\rho_1(\varphi)$. Dada a normalização (5.17), mostra-se que o mínimo global leva a um valor máximo e unitário para o raio, enquanto os demais pontos encontram-se dentro da esfera unitária. Portanto, foi obtida pelos autores uma condição de diferenciação para os mínimos globais e os demais pontos críticos, baseada no raio $\rho_1(\varphi)$, que pode ser calculado a partir de $t(n)$ e $s(n)$.

Por fim, os autores desencorajam o uso desta condição de raio no projeto de um CMA com restrições, pois o uso deste artifício poderia gerar novos equilíbrios indesejados na esfera, além dos já apontados. Porém, consideram o resultado uma ferramenta possivelmente útil para identificar a qualidade da solução obtida pelo critério do módulo constante, o que poderia direcionar procedimentos destinados a melhorar o desempenho do algoritmo.

O artigo em questão foi um marco para a análise de convergência do CMA, tendo obtido resultados fortes e elucidativos. Outras análises sobre a posição dos pontos de equilíbrio, talvez mais refinadas que a de Ding e Kennedy, foram desenvolvidas posteriormente, e teremos a oportunidade de revisitar este tópico em outras seções.

5.7. Equalizadores de Comprimento Finito Revisitados

Ainda em 1992, um artigo de Ding, Johnson e Kennedy [Ding et al., 1992] estabeleceu um resultado muito importante, que fundamentou mais solidamente o procedimento de análise de algoritmos adaptativos de Bussgang. Podemos afirmar que o grande resultado do artigo é uma prova de que a análise feita no espaço conjunto (canal + equalizador) só equivale à análise feita no domínio dos parâmetros do equalizador se a matriz de convolução do canal tiver um espaço nulo trivial, o que pode ocorrer em dois casos:

- 1 – Quando o equalizador é duplamente infinito.
- 2 – Quando o canal é trivial, ou seja, simplesmente impõe um ganho ao sinal de entrada.

Isto quer dizer que, em situações em que o equalizador não for duplamente infinito e o canal não-trivial, o método de análise no espaço conjunto, empregado em [Godard, 1980], [Foschini, 1985], e também por Shalvi e Weinstein em [Shalvi e Weinstein, 1990], não produz resultados conclusivos sobre convergência global.

O impacto desta afirmação pode ser percebido se observarmos que as primeiras análises dos algoritmos foram feitas em \mathbf{c} , sendo que a partir daí era estudada uma extensão para o caso prático. O que Ding, Johnson e Kennedy mostraram é que tal extensão não é possível exceto em casos particulares. Outro resultado muito relevante do mesmo artigo é que um equalizador com um número arbitrário de parâmetros não é ainda uma garantia de que haja um possível mapeamento bijetor de convergência global entre o espaço combinado e o espaço de parâmetros do equalizador, o que contradiz a intuição exposta por Foschini, vista por nós no segundo item da seção 5.3. Vale lembrar que este último já reconhecia que a sua intuição não era suportada por nenhuma prova matemática rigorosa.

O primeiro questionamento que leva aos resultados expostos é se a obtenção dos pontos críticos da função custo do algoritmo é equivalente nos dois espaços. No espaço do equalizador, os pontos críticos são os que satisfazem a condição de gradiente nulo (em \mathbf{w}):

$$E [\mathbf{x}^*(n).y(n).|y(n)|^{p-2}.(|y(n)|^p - R_p)] = 0 \quad (5.19)$$

e no espaço conjunto, os pontos críticos satisfazem:

$$E [\mathbf{s}^*(n).y(n).|y(n)|^{p-2}.(|y(n)|^p - R_p)] = 0 \quad (5.20)$$

A grande pergunta é:

$$(5.19) \Leftrightarrow (5.20)?$$

pois neste caso, a obtenção dos pontos nos dois espaços seria equivalente.

Uma vez que:

$$\mathbf{x}(n) = \mathbf{H}.s(n) \quad (5.21)$$

onde \mathbf{H} é a matriz de convolução do canal, podemos reescrever (5.19) como sendo:

$$\mathbf{H}^* . E [\mathbf{s}^*(n).y(n).|y(n)|^{p-2}.(|y(n)|^p - R_p)] = 0 \quad (5.22)$$

o que imediatamente nos leva a concluir que:

$$(5.20) \Rightarrow (5.19)$$

Porém, a outra implicação não é necessariamente verdadeira, pois \mathbf{H}^* pode ter um espaço nulo de dimensão maior que zero.

Para entendermos o motivo pelo qual tal condição deve ser satisfeita, analisemos um pouco mais a matriz \mathbf{H} . Vejamos, primeiramente, como ela é definida:

$$\mathbf{H} = \begin{bmatrix} h(0) & h(1) & h(2) & \dots & h(N) & 0 & 0 & 0 \\ 0 & h(0) & h(1) & h(2) & \dots & h(N) & \dots & 0 \\ \vdots & \vdots & \ddots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & h(0) & h(1) & h(2) & \dots & h(N) \end{bmatrix} \quad (5.23)$$

Tal matriz tem $M + 1$ linhas e $M + N + 1$ colunas, onde M é a ordem do equalizador e N é a ordem do canal.

Agora, passamos à pergunta crucial: em que condições pode \mathbf{H} ter um espaço nulo trivial, ou seja, (5.19) \Leftrightarrow (5.20) ?

1 - Quando $N = 0$, ou seja, o canal é apenas um ganho, sem adição de IIS e a matriz \mathbf{H} tem dimensão $M+1 \times M+1$. Neste caso, \mathbf{H} terá apenas um espaço nulo trivial.

2 - E quando $M \rightarrow \infty$, \mathbf{H} terá um espaço nulo trivial? Para sistemas causais isto não ocorrerá, mas se considerarmos o caso de um equalizador duplamente infinito, então teremos tal condição sobre a matriz de convolução.

Portanto, apenas nestes dois casos os pontos obtidos no espaço conjunto irão necessariamente contemplar todas as possibilidades do espaço dos parâmetros do equalizador finito. São dois casos de pouco valor prático, o que indica que, apesar de ser tratável do ponto de vista matemático, a análise por meio do espaço conjunto não corresponde necessariamente à análise bem mais complexa realizada no domínio de \mathbf{w} .

Vale destacar que, mesmo para um equalizador de comprimento infinito, pode não haver a correspondência biunívoca entre os pontos nos dois domínios de análise. Isto vai de encontro à crença de que o aumento do número de taps leva a um comportamento arbitrariamente próximo do caso duplamente infinito, assunto ao qual voltaremos na seção seguinte.

5.8. *Estabilidade dos Mínimos Locais e Comprimento do Equalizador*

Em 1993, um artigo de Ding e Johnson [Ding e Johnson, 1993] procurou analisar uma conjectura feita por alguns pesquisadores da área. Basicamente, o ponto de partida da conjectura era a prova da convergência global de equalizadores duplamente infinitos ajustados pelo critério de Godard. A partir deste resultado, passou a ser uma suposição

popular a que “em equalizadores FIR suficientemente longos, os mínimos locais estáveis poderiam desaparecer (assintoticamente)”, sendo que, neste caso, um comportamento de convergência global seria arbitrariamente aproximado.

O resultado capital do artigo é que, para uma certa classe de canais AR, a análise teórica mostra que os mínimos locais se mantêm estáveis para todo e qualquer comprimento finito escolhido para o equalizador. Este resultado vem questionar a conjectura exposta acima, pois demonstra que não haverá um comprimento, por maior que seja, que permita a eliminação da estabilidade dos mínimos locais. O modelo de canal usado pelos autores é o mesmo da equação (5.8). A caracterização dos pontos de mínimo já foi feita na seção 5.5.

Os autores demonstram que os pontos de mínimo local serão localmente atraentes se for satisfeita uma certa condição que envolve apenas a estatística de entrada e as características do canal, não dependendo, portanto, da ordem do equalizador. Portanto, os mínimos locais não perdem a sua estabilidade para um certo valor de ordem do equalizador suficientemente grande, como afirmavam as conjecturas descritas anteriormente.

Os autores ressaltam, com muita propriedade, que o trabalho não tece considerações sobre a variação da profundidade e da extensão da região de atração dos mínimos com o comprimento do equalizador. O aumento da dimensão do equalizador, embora não possa eliminar os pontos de mínimo local, pode levar, através de uma redução de suas bacias de atração e da altura do ponto de sela mais próximo, a uma diminuição considerável dos efeitos “nocivos ao desempenho” por eles causados.

5.9. *Extraindo Informação da Superfície de Erro*

Johnson e Anderson, em 1995, publicaram um artigo [Johnson e Anderson, 1995] que teve como foco de análise as características da função custo do algoritmo do módulo constante, no domínio dos pesos do equalizador, considerado de comprimento finito. As hipóteses sobre o modelo foram as usuais: canal FIR, fonte binária e branca de média zero (± 1) e equalizador linear, FIR e espaçado de um intervalo de símbolo.

São apresentadas então, após o adequado desenvolvimento matemático, elegantes expressões matriciais para o vetor gradiente e para a matriz hessiana, ferramentas fundamentais para a análise da superfície de erro do CMA.

Em todas as simulações é utilizado o canal $\mathbf{h} = [0.5 \ -0.3 \ 0.25]^T$ e um equalizador de dois taps. Em tal modelo, a superfície de erro apresenta dois mínimos globais, dois locais, dois pares de pontos de sela e um máximo local.

A primeira análise tem a ver com o vetor gradiente, cuja direção, com sinal negativo, é decomposta em um componente radial e um angular. A seguir, são mapeados os pontos que apresentam componente radial para dentro ou para fora e componente angular no sentido horário ou anti-horário. Um aspecto interessante é que os pontos de sela e de mínimo da superfície ficam onde a fronteira entre os dois tipos de componente radial encontra a fronteira entre os dois tipos de componente angular.

A partir da expressão da matriz hessiana, ocorre um novo mapeamento, desta vez relacionando os pontos da superfície ao número de autovalores positivos da matriz (5.7). Vale lembrar que um ponto de mínimo precisa ter dois autovalores positivos, um máximo deve ter ambos negativos, e um ponto de sela terá autovalores com diferentes sinais. Este mapeamento delinea as regiões da função custo, caracterizando-a com respeito aos pontos críticos.

Outro aspecto analisado é a atração ou a repulsão exercidas pela origem. Se um ponto estiver a uma grande distância da origem, o caminho contrário ao gradiente tenderá a aproximá-lo (contração). Se, ao contrário, o ponto estiver muito próximo, o caminho tenderá a afastá-lo (expansão). O mapeamento, realizado através da análise dos parâmetros da expressão iterativa determinística, apresenta as regiões de expansão e contração separadas por uma fronteira que contém os possíveis atratores.

É então introduzida uma classificação muito interessante dos pontos críticos da função custo em três grupos, o que é, em nossa opinião, o ponto alto do trabalho:

- 1) A origem, que é sempre um máximo local (classe dos máximos);
- 2) Mínimos globais e pontos de sela (classe de Mazo);
- 3) Os “falsos mínimos” (classe de Ding);

São consideradas duas novas hipóteses, a saber, a inexistência de nulos espectrais e uma dimensão suficientemente grande do equalizador (bem maior que a de um filtro capaz de produzir uma boa aproximação da inversa do canal).

Conforme afirmam os autores, o que diferencia as três classes é a “forma como elas anulam o gradiente da função”. Tal definição não traz grandes luzes, de forma que tentaremos agora entender um pouco melhor o que se quis dizer com isto.

Já havíamos dito, nesta seção, que é utilizada em [Johnson e Anderson, 1995] uma notação muito elegante e concisa. É com base nesta notação que a análise se desenrola. Partindo da expressão do gradiente (com respeito aos parâmetros do equalizador) da função custo do CMA e realizando inúmeras manipulações, os autores chegam por fim à expressão de tal vetor:

$$\nabla_w J = \mathbf{H} \cdot \Lambda \cdot \mathbf{c} \quad (5.24)$$

onde \mathbf{H} foi definida em (5.23), \mathbf{c} é o vetor de parâmetros no espaço conjunto e Λ é uma matriz definida por:

$$\Lambda = \text{diag} (\Lambda_0, \Lambda_1, \dots, \Lambda_g) \quad (5.25)$$

onde g é a soma do número de parâmetros do canal e do equalizador e:

$$\Lambda_i = 4 \cdot \left(3 \cdot \sum_{q=0}^g c_q^2 - 1 - 2 \cdot c_i^2 \right) \quad (5.26)$$

Discutamos, separadamente, as características de cada classe de pontos.

5.9.1. A Classe dos Máximos

A primeira afirmação categórica a respeito desta classe é que a origem do espaço dos parâmetros do equalizador, ou seja, o vetor $\mathbf{w} = \mathbf{0}$, é o único máximo local da função custo de Godard. Isto é demonstrado pelos autores em seu trabalho.

Como o vetor de parâmetros do equalizador é nulo, o vetor \mathbf{c} também o será, como consequência da operação de convolução de \mathbf{w} com a resposta ao impulso do canal de comunicação. Esta é a forma pela qual o máximo anula o gradiente (5.24), sendo característica de sua classe.

5.9.2. A Classe de Mazo

J. E. Mazo tem seu nome associado ao estudo da convergência do algoritmo de decisão direta (DD). Ele empresta seu nome à classe que envolve dois tipos de pontos críticos: mínimos globais e pontos de sela. A associação destes dois tipos de pontos com o ponto de máximo nos remete à visão de Foschini [Foschini, 1985], que já analisamos na seção 5.3. Naquele cenário, do equalizador duplamente infinito, os únicos pontos críticos além do máximo eram os da classe de Mazo: mínimos globais e pontos de sela. Estes pontos são características comuns à função custo de Godard no caso duplamente infinito e nos demais casos, os quais incluem todos os equalizadores realizáveis na prática. A partir do que discutimos na seção 5.7, podemos intuir que os pontos da classe de Mazo, justamente por estarem presentes também no caso do equalizador duplamente infinito, devem anular o gradiente tanto no espaço dos parâmetros do equalizador como também no espaço conjunto (lembrando que ambos levam a resultados idênticos no caso duplamente infinito). Os autores expressam o gradiente de J com relação aos parâmetros \mathbf{c} de forma matricial como:

$$\nabla_{\mathbf{c}} J = \Lambda \cdot \mathbf{c} \quad (5.27)$$

Como \mathbf{c} é diferente do vetor nulo, para que o gradiente se anule no espaço conjunto é necessário que $\Lambda.\mathbf{c}$ se anule. Como já era esperado a partir do que vimos na seção 5.7, isto implica:

$$\nabla_{\mathbf{w}}J = \mathbf{H}.\Lambda.\mathbf{c} = \mathbf{H}.\mathbf{0} = \mathbf{0} \quad (5.28)$$

Portanto, os pontos de Mazo anulam o gradiente de J tanto se o mesmo for tomado com respeito a \mathbf{w} quanto se for tomado com respeito a \mathbf{c} . Esta é a principal característica destes pontos.

5.9.3. A Classe de Ding

Os pontos da classe de Ding são aqueles que não anulam o gradiente no espaço conjunto, mas o fazem no espaço dos parâmetros do equalizador. Isto está, novamente, em perfeita harmonia com o que discutimos na seção 5.7 e com o fato dos mesmos terem sido ignorados na análise de Foschini. Tais considerações nos levam, imediatamente, à condição:

$$\nabla_{\mathbf{c}}J = \Lambda.\mathbf{c} \neq \mathbf{0} \quad (5.29)$$

Porém, como tais pontos *são mínimos de J* no espaço dos parâmetros do equalizador, temos, necessariamente:

$$\nabla_{\mathbf{w}}J = \mathbf{H}.\Lambda.\mathbf{c} = \mathbf{0} \quad (5.30)$$

Como podem ser obedecidas, conjuntamente, as condições (5.29) e (5.30)? A resposta, no fundo, já nos foi dada pela seção 5.7: se $\Lambda.\mathbf{c}$ existir no espaço nulo de \mathbf{H} . Portanto, os mínimos de Ding se caracterizam por não serem mínimos no espaço conjunto canal + equalizador, mas apenas no espaço dos parâmetros do equalizador.

Após algumas considerações de natureza matemática, os autores expõem um resultado muito interessante:

“Os mínimos de Ding caracterizam-se por apresentar valores (em c) significativos no início e/ou final do vetor de parâmetros, sendo que, para que eles sejam estáveis, é necessário que não haja mais que um valor significativo na região central do vetor.”

Este resultado ajuda-nos a ter algum *insight* sobre as características dos mínimos de Ding, lembrando que para esta análise supõe-se que o equalizador tenha uma ordem suficientemente alta. Ele nos mostra que tais mínimos apresentam uma distribuição de coeficientes significativos, ou seja, de maior magnitude, nos extremos do vetor c . Isto encoraja fortemente a inicialização “*center-spike*” pois, a partir desta inicialização, seria necessário muito tempo para que os extremos do vetor de parâmetros atingissem um valor significativo, o que seria característico de uma convergência para um mínimo local. Assim, é de se esperar que a inicialização “*center-spike*” favoreça a convergência para o mínimo global.

5.9.4. Discussão

Não seria exagero dizer que o trabalho [Johnson e Anderson, 1995] representa um marco: a síntese de alguns dos esforços pioneiros empreendidos na busca pela caracterização teórica do algoritmo do módulo constante. É visível um amadurecimento acerca das possibilidades do método, desde a proposta inicial de Godard, passando pelo trabalho vigoroso de Foschini até os trabalhos de Ding e outros, que ajudaram a construir bases mais sólidas para o entendimento do método. As três próximas seções trazem novos resultados baseados neste mesmo paradigma de análise, baseado nas derivadas da função custo do critério CM.

5.10. *Extraindo Informação da Superfície de Erro II: Sobre a Localização do Mínimo Global*

Em 1996, Li e Ray Liu [Li e Ray Liu, 1996] obtiveram novos resultados sobre a localização dos pontos de mínimo de um critério de equalização autodidata. Num contexto bastante generalista, é enfocada uma função custo que engloba quase todos os algoritmos autodidatas conhecidos. Podemos escrever tal função como sendo:

$$J_{\text{GEN}} = E\{\Phi[y(n)]\} \quad (5.31)$$

Definamos ainda

$$\phi[y(n)] = \Phi'[y(n)] \quad (5.32)$$

que é tão somente a primeira derivada da função custo com respeito à saída do equalizador, considerado um filtro FIR não-causal com $2N+1$ coeficientes.

A análise de algumas propriedades de $\phi(\cdot)$ fornece a base teórica para os resultados capitais deste trabalho, dentre os quais iremos destacar um, bastante intuitivo:

Quando os parâmetros do equalizador estão numa configuração de mínimo da função custo, a saída $y_{\min}(n)$ satisfaz:

$$E\{\phi[y_{\min}(n)].\mathbf{x}(n)\} = \mathbf{0} \quad (5.33)$$

Discutamos brevemente este resultado. Em primeiro lugar, a partir de (5.32), percebemos que $\phi[y(n)]$ funciona como uma estimativa do sinal de erro a partir da saída do equalizador, num modo de operação que está na raiz de todas as técnicas de Bussgang (seção 4.7.1). O que temos em (5.33) é uma espécie de princípio da ortogonalidade (seção 4.4) para algoritmos cegos. Ou seja, em um mínimo da função custo, a função erro de predição será ortogonal ao vetor de entrada do equalizador $\mathbf{x}(n)$. Isto decorre diretamente da

condição de gradiente nulo, necessária para que um ponto seja um mínimo. Tais considerações são tecidas supondo que os parâmetros do equalizador tenham uma configuração próxima à de inversão perfeita do canal. Isto quer dizer simplesmente que a saída $y_{\min}(n)$ estará muito próxima do sinal transmitido $s(n)$. Uma vez assumida esta hipótese, torna-se razoável a seguinte aproximação, básica para toda a análise feita no artigo:

$$\phi[y_{\min}(n)] = \phi[s(n)] + \phi'[s(n)] \cdot [y_{\min}(n) - s(n)] \quad (5.34)$$

que é simplesmente uma aproximação de primeira ordem de $\phi[y_{\min}(n)]$ em torno de $s(n)$.

Definindo ainda matriz de convolução \mathbf{K} , para este caso, de forma ligeiramente diferente da matriz \mathbf{H} definida em (5.23)

$$\mathbf{K} = \begin{bmatrix} h(-M) & 0 & \dots & 0 \\ \vdots & h(-M) & \ddots & \vdots \\ h(0) & \vdots & \ddots & 0 \\ \vdots & h(0) & \ddots & h(-M) \\ h(M) & \vdots & \ddots & \vdots \\ 0 & h(M) & \ddots & h(0) \\ \vdots & 0 & \ddots & \vdots \\ 0 & \dots & \dots & h(M) \end{bmatrix} \quad (5.35)$$

onde $2M+1$ é o número de coeficientes não-nulos da resposta ao impulso do canal, chega-se a:

$$\mathbf{w}_{\text{glob}} = f(0) \cdot \mathbf{R}_f^{-1} \cdot \mathbf{h} \quad (5.36)$$

onde

$$\mathbf{h} = [0, \dots, 0, h(M), \dots, h(0), \dots, h(-M), 0, \dots, 0]^T \quad (5.37)$$

e, por definição:

$$\mathbf{R}_f = \mathbf{x}^T \cdot \mathbf{F} \cdot \mathbf{x} \quad (5.38)$$

com

$$\mathbf{F} = \text{diag}[f(1), \dots, f(1), f(0), f(1), \dots, f(1)] \quad (5.39)$$

onde aparece $M+N$ vezes o valor $f(1)$, uma vez $f(0)$ e outras $M+N$ vezes o valor $f(1)$. Vale lembrar que o equalizador tem $2N+1$ parâmetros neste trabalho.

Os valores $f(0)$ e $f(1)$ são definidos como:

$$f(0) = (1/\sigma_s^2) \cdot E\{ \phi'[s(n)] \cdot s^2(n) \} \text{ e } f(1) = E\{ \phi'[s(n)] \} \quad (5.40)$$

sendo σ_s^2 a variância do sinal transmitido.

A expressão (5.36) é o resultado capital deste artigo. Ela nos mostra que é possível expressar o valor do mínimo global do critério de Godard por meio de uma equação matricial muito parecida com (4.33), que expressa a solução de Wiener. É importante frisar, no entanto, que a validade de (5.36) depende da proximidade do mínimo global em relação à inversa do canal, como já foi ressaltado.

E o que podemos apreender de todas estas fórmulas, no final das contas? O que consideramos mais importante é a expressão do mínimo global do critério CM de uma maneira bastante elegante e similar às equações de Wiener-Hopf. Trata-se de uma iniciativa louvável pois, tendo a função custo do critério CM ordem quatro, percebemos o quão difícil seria obter analiticamente todos os seus pontos críticos e, mais ainda, distinguir um mínimo dos demais tipos de pontos. O “espírito” de (5.36), por assim dizer, está na aproximação (5.34) e nas propriedades de $\phi(\cdot)$, expostas com detalhes na referência estudada, a qual sugerimos ao leitor que deseje se aprofundar neste tipo de análise.

5.11. *Extraindo Informação da Superfície de Erro III: Os Limitantes das Regiões de Concavidade e Convexidade*

Em 1997, Choi, Song e Park publicaram um artigo [Choi et al., 1997] que lidava com uma análise de convergência centrada nas propriedades da função custo do algoritmo “quártico” de Godard, nome aplicável ao CMA, por sua função custo ter ordem quatro.

A análise é realizada para sinais binários reais, i.i.d., e considera-se um canal sem ruído aditivo. O estudo se funda nos conceitos de estabilidade local vistos no capítulo 2, tendo porém por base a matriz hessiana. A chave de todo o trabalho é uma expansão de Taylor multidimensional, de primeira ordem, do gradiente da função custo em torno do ponto de mínimo global \mathbf{w}_{glob} :

$$\frac{\partial J^{(2)}}{\partial \mathbf{w}} = \frac{\partial J^{(2)}}{\partial \mathbf{w}}(\mathbf{w}_{\text{glob}}) + \mathbf{H}_{\text{ess}}(\mathbf{w}_{\text{glob}}) \cdot (\mathbf{w} - \mathbf{w}_{\text{glob}}) \quad (5.41)$$

onde \mathbf{H}_{ess} é a matriz hessiana, definida em (5.7).

Os primeiros resultados expostos pelos autores enfocam basicamente uma comparação entre algumas características de convergência do algoritmo do módulo constante e do algoritmo da decisão direta (DD), cuja função custo é:

$$J_{\text{DD}} = E\{\text{dec}[y(n)] - y(n)\}^2 \quad (5.42)$$

onde $\text{dec}(\cdot)$ representa a atuação do decisor.

A comparação é interessante pois liga duas das mais populares técnicas empregadas no âmbito da equalização autodidata. Basicamente, os autores mostram que, dadas algumas restrições de modelo e uma normalização adequada, é possível provar que o máximo valor do passo de adaptação μ para o qual há convergência é menor, no caso do CMA, na vizinhança dos mínimos globais das respectivas funções custo. Outro resultado obtido é que, sob as hipóteses aventadas, o algoritmo do módulo constante converge mais rapidamente que o algoritmo DD.

Não entraremos em pormenores sobre as considerações dos autores, pois isto exigiria uma exposição demasiadamente longa. O que queremos deixar claro é apenas a idéia geral da análise empreendida e de seus objetivos.

Uma outra contribuição relevante é o estudo das regiões de concavidade e convexidade da superfície de erro do algoritmo do módulo constante. A região próxima ao máximo local é uma região côncava, como podemos intuir. Afastando o vetor de pesos ainda mais da origem, chega-se à região dos mínimos. Prosseguindo com o afastamento, logo se estará, por certo, em uma região convexa, caracterizada por um rápido aumento no custo para um aumento no valor da norma do vetor de parâmetros do equalizador. Para que o leitor possa visualizar melhor estas regiões, sugerimos a observação da Fig. 6.4, que traz as curvas de nível da função custo para um exemplo que discutiremos no próximo capítulo.

O trabalho dos autores visou o estabelecimento de limitantes teóricos para as regiões de concavidade e convexidade. Estes limitantes são lugares geométricos que contém a região sob análise, apresentando no entanto, como é bastante razoável, uma geometria mais simples que a de sua respectiva região.

Em síntese, são obtidos limitantes esféricos do tipo:

$$\|\mathbf{w}\| \leq \text{constante} \text{ ou } \|\mathbf{w}\| \geq \text{constante} \quad (5.43)$$

para as regiões de concavidade e convexidade, respectivamente.

É obtido ainda um limitante do mesmo tipo para a região de atração do mínimo global, ou seja:

$$\|\mathbf{w} - \mathbf{w}_{\text{glob}}\| \leq \text{constante} \quad (5.44)$$

Estes limitantes permitem que se faça um esboço das características da função e da região de atração de mínimo global, fornecendo importantes subsídios para a formação de uma visão mais intuitiva sobre o critério CM.

5.12. Influência dos Pontos de Sela sobre a Convergência

Em 1997, um artigo de Lambotharan, Chambers e Johnson [Lambotharan et al., 1997] analisou a influência dos pontos de sela, presentes na função custo, sobre a convergência do algoritmo do módulo constante.

A ocorrência deste fenômeno se faz notar através de porções planas da curva de erro, que indicam uma variação muito lenta na adaptação do vetor de pesos. A estas porções planas associa-se um período de atração por um ponto de sela que, como vimos no capítulo 2, tem direções estáveis e instáveis.

No artigo, é mostrado um exemplo no qual os autores utilizam um canal espaçado de metade do período de símbolo, um equalizador fracionário com inicialização *center-spike*, onde o estado do equalizador é atraído por duas selas antes de convergir para um ponto de mínimo. Portanto, pode haver atração para múltiplas selas numa inicialização padrão, o que torna o problema ainda mais delicado. Mostra-se ainda que o tempo de escape das selas depende das características do canal e também do tipo de ponto de sela em questão.

Os autores propõem então uma modificação do CMA que ajuda a acelerar a sua convergência, tanto mais quanto a atração por selas for um problema relevante. Basicamente, eles introduzem o termo $E[y^2(n)]$ como uma “medida” da qualidade do valor presente do estado do algoritmo. Assim, o algoritmo assume a forma, lembrando que o sinal transmitido é binário, de módulo unitário e tem amostras i.i.d.:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu}{E[y^2(n)]} \cdot y(n) \cdot [1 - y^2(n)] \cdot \mathbf{x}(n) \quad (5.45)$$

Assim, quando se está próximo de uma solução ótima, o valor da esperança estará próximo de um, e a velocidade do algoritmo terá poucas modificações. Quando, porém, o estado estiver próximo de uma sela, a esperança será menor que um, o que irá aumentar o

passo de adaptação do algoritmo, acelerando o escape. O valor da esperança pode ser estimado através de uma média temporal tomada com um fator de esquecimento adequado.

5.13. *Breve Discussão: As Primeiras Abordagens*

As doze primeiras seções deste capítulo podem ser agrupadas em torno de uma análise do CMA segundo o paradigma mais tradicional, baseado principalmente na função custo e em suas derivadas. Os resultados apresentados constituem uma base muito sólida, podemos até dizer que o alicerce fundamental que inspirou outros autores a buscarem novos paradigmas de análise.

Podemos dizer que as *abordagens geométricas* constituem um outro paradigma de análise muito frutífero, sendo que analisaremos a seguir alguns trabalhos que envolvem diversos aspectos do critério CM, dentre os quais destacamos a equivalência entre os critérios CM e de Shalvi-Weinstein e a comparação entre o critério CM e os critérios de Wiener e ZF. Gostaríamos de ressaltar que, quando utilizamos o termo *abordagem geométrica* não queremos afirmar que o estudo da matriz hessiana não envolva geometria ou que o paradigma geométrico não possa utilizar também as derivadas da função custo.

5.14. *Efeito da Kurtosis Inicial na Convergência do Algoritmo e Considerações sobre a Posição dos Mínimos*

Um artigo de Li e Ding [Li e Ding, 1995] deu origem a uma abordagem conceitual interessante, baseada na idéia de *cones de mínimo global único*. Procuraremos expor, nesta seção, os conceitos básicos deste paradigma de estudo, para que o leitor tenha alguma idéia sobre o processo de análise empregado pelos autores. Uma vez exposta esta base, não iremos nos ater aos detalhes matemáticos das demonstrações, sendo que apenas comentaremos os principais resultados obtidos.

O cerne de todo o paradigma de análise de Li e Ding centra-se em três conceitos:

1) *Attainable set* T : é o conjunto de todas as configurações \mathbf{c} no espaço conjunto que podem ser obtidas a partir de um equalizador finito. O que isto quer dizer? Trata-se do conjunto de todos os valores de \mathbf{c} produzidos a partir da convolução de todos os possíveis valores de \mathbf{w} com a resposta ao impulso do canal. Dado que o canal e o equalizador são filtros FIR, jamais se poderá atingir a condição ZF, por exemplo. Isto quer dizer que não haverá nenhum vetor \mathbf{c} pertencente a T com um único elemento não-nulo, o que corresponderia à condição de equalização perfeita. Isto mostra bem que não é possível obter qualquer valor de \mathbf{c} para um dado sistema, mas apenas aqueles ligados ao vetor \mathbf{w} de parâmetros do equalizador por meio da convolução deste com a resposta ao impulso do canal.

2) *Cones de mínimo global único* são definidos como:

$$S_n^+: \{ \mathbf{c} \in l^1(\mathbb{R}) / c(n) > 0 \text{ e } c(n) > |c(k)|, k \neq n \} \quad (5.46)$$

$$S_n^-: \{ \mathbf{c} \in l^1(\mathbb{R}) / c(n) < 0 \text{ e } -c(n) > |c(k)|, k \neq n \} \quad (5.47)$$

onde $l^1(\mathbb{R})$ é um espaço vetorial real.

Basicamente, um cone de mínimo global único é definido pela posição do elemento de maior magnitude do vetor \mathbf{c} e por seu sinal. Uma característica marcante destes cones é que há apenas um mínimo global pertencente a cada um deles, sendo o mesmo uma solução ZF. Por exemplo, se tomarmos o cone S_2^+ , notamos que $\mathbf{c} = [\dots 0 0 1 \text{ (posição 2)} 0 \dots]^T$ é um de seus pontos, sendo também uma configuração ZF. Já para o cone S_2^- , a única solução ótima seria $\mathbf{c} = [\dots 0 0 -1 \text{ (posição 2)} 0 \dots]^T$. Em ambas soluções mostradas, tínhamos que o único elemento não-nulo era $c(2)$, o que já é esperado a partir do índice numérico que caracteriza o cone. Esta propriedade, de haver apenas um mínimo global em cada um dos cones é essencial, como a denominação destes entes já indica.

3) *Super-Esfera* $\Phi(r)$, com $r > 0$, é definida por:

$$\Phi(r) = \{ \mathbf{c} \in l^1(\mathbf{R}) / \|\mathbf{c}\| = r \} \quad (5.48)$$

onde $\Phi(r)$ é simplesmente uma esfera de raio r no espaço conjunto canal + equalizador. O termo “super” simplesmente torna a definição aplicável a espaços com dimensão maior que três.

Tendo definido tais entes, iremos expor os principais resultados obtidos pelos autores. Para tanto, definamos a *kurtosis normalizada* como sendo:

$$K_{nm}(\mathbf{b}) = K(\mathbf{b})/\sigma_b^4 \quad (5.49)$$

onde a *kurtosis* $K(\mathbf{b})$ é definida como:

$$K(\mathbf{b}) = E[|\mathbf{b}|^4] - 2.E^2(|\mathbf{b}|^2) - |E(\mathbf{b}^2)|^2 \quad (5.50)$$

e σ_b^4 é o quadrado da variância de \mathbf{b} .

O que os autores mostram é que o valor desta grandeza no instante inicial de adaptação tem enorme influência na evolução subsequente do estado do equalizador. Podemos perguntar: que tipo de influência? Seja então a seguinte condição:

$$\frac{K_{nm}[y(n)]}{K_{nm}[s(n)]} > 0.5 \quad (5.51)$$

Suponhamos que a saída do equalizador obedeça à condição (5.51) no instante inicial, ou seja, que o valor $\mathbf{w}_{inicial}$, associado a um certo $\mathbf{c}_{inicial}$ no espaço conjunto, leve a saída do equalizador a satisfazer (5.51). Se o valor de $\mathbf{c}_{inicial}$ estiver contido em um certo cone de mínimo local único S_n^\pm , e o passo de adaptação for suficientemente pequeno, então é demonstrado que o vetor \mathbf{c} irá convergir para um ponto de mínimo contido neste mesmo cone, ou seja, em $T \cap S_n^\pm$. Esta intersecção indica apenas que o valor de \mathbf{c} para o qual o equalizador irá convergir estará contido também no *attainable set*, o que não nos surpreende, pois o próprio fato de o equalizador ser capaz de atingir tal configuração obrigatoriamente faz dela um elemento de T .

A linha geral da demonstração acima é mostrar que caso a condição (5.51) seja obedecida e $\mathbf{c}_{\text{inicial}}$ esteja em um cone S_n^\pm , então o custo inicial será menor que o menor custo dentre todos os valores da fronteira do cone. Ora, isto mostra que, sendo usado um método do tipo *steepest descent* com um passo adequado, não há forma pela qual \mathbf{c} possa cruzar a fronteira, o que demandaria uma migração para soluções com um custo maior, *num método que busca, idealmente, soluções com custo menor*.

Fica no ar uma outra questão, que tem grande relevância, mas não é discutida neste paper: o que é, exatamente, um passo de adaptação “pequeno” ou “suficientemente pequeno”? Entendemos o que os autores buscam afirmar, quando usam tais termos, que se trata de uma situação em que, primeiramente, não haja instabilidade e, ainda, em que as flutuações aleatórias mantenham-se em patamares nos quais sua influência não leve a nenhum efeito “fora do script”. Trata-se certamente de uma hipótese restritiva, mas isto não diminui a importância do estudo teórico proposto, pois seus resultados, embora talvez não nos levem a uma compreensão exata do comportamento do algoritmo em todas as possíveis situações, nos permitem entender melhor determinados aspectos estruturais da técnica, o que é deveras relevante por si só.

Há ainda alguns outros resultados que iremos comentar. Por exemplo, qual a localização dos mínimos? É mostrado que o paradigma de análise utilizado pode lançar algumas luzes sobre este problema.

Denominaremos uma solução ZF do tipo $[0 \ 0 \ \dots \ 1 \ (\text{posição } n) \ \dots \ 0 \ 0]^T ZF_n^+$ e $[0 \ 0 \ \dots \ -1 \ (\text{posição } n) \ \dots \ 0 \ 0]^T ZF_n^-$. Se $ZF_n^\pm \in T \cap S_n^\pm$, então esta solução será o único mínimo em $T \cap S_n^\pm$ e não haverá pontos de mínimo em $T \cap B_n^\pm$, onde B_n^\pm é a fronteira de S_n^\pm . Discutamos um pouco estas afirmações. Suponha que o equalizador seja capaz de inverter perfeitamente o canal. Esqueçamos o caso do equalizador duplamente infinito, para o qual já sabemos ser as soluções ZF os únicos mínimos possíveis. No caso de um equalizador finito, pode haver perfeita inversão de canal se este for AR e o equalizador for MA, por exemplo, num problema de “zeros cancelando pólos”. Neste caso, soluções ZF farão parte do *attainable set* e, portanto, pode ser satisfeita uma condição do tipo $ZF_n^\pm \in T \cap S_n^\pm$ para algum valor de n . Há apenas uma pequena novidade quanto à notação: a introdução de B_n^\pm . Este ente expressa simplesmente o lugar geométrico correspondente à fronteira do respectivo cone S_n^\pm . Por exemplo, B_7^+ é a fronteira do cone

S_7^+ . O que é demonstrado é que, caso haja perfeita inversão de canal, ou seja, $ZF_n^\pm \in T \cap S_n^\pm$ para algum valor de n , então não haverá pontos de mínimo em $T \cap B_n^\pm$ ou seja, nos pontos da fronteira contidos no *attainable set*. Isto não contradiz o trabalho de Ding [Ding et al., 1991], mas apenas indica que os mínimos locais existentes não estarão associados ao mesmo cone que contém o mínimo global ou à sua fronteira.

E se não for possível uma perfeita inversão, como é o caso de equalizador FIR e modelo de canal MA? Neste caso, também podem ser tecidas algumas considerações. A primeira delas é que se ZF_n^\pm estiver próxima a, mas não contida em, $T \cap S_n^\pm$, então deve haver apenas um mínimo em $T \cap S_n^\pm$ perto de ZF_n^\pm , enquanto outros possíveis mínimos em $T \cap S_n^\pm$ deverão estar próximos à fronteira do cone S_n^\pm .

Toda a análise teórica que expusemos, vale frisar, foi desenvolvida para o algoritmo de Shalvi-Weinstein. Porém, como discutiremos mais adiante (seção 5.17), os resultados são extensíveis ao algoritmo de Godard, graças ao forte elo que une as duas técnicas.

Vamos agora comentar algumas idéias expostas pelos autores acerca da inicialização do algoritmo. Basicamente, eles propõem um processo de inicialização baseado em três etapas e nos conceitos que discutimos. Tentemos entender o que foi feito.

Suponhamos que a inicialização satisfaça a condição (5.51) e pertença a um certo cone. Como vimos, isto quer dizer que o estado do sistema convergirá para um mínimo contido no mesmo cone ao qual pertence $\mathbf{c}_{inicial}$. Sabemos que este mínimo deve estar próximo a uma condição ZF e, caso o equalizador tenha um tamanho suficiente, podemos afirmar que a solução ZF deverá estar muito próxima da obtida, que pertence ao *attainable set*.

Em suma, temos três condições:

- 1) A resposta conjunta inicial do sistema, $\mathbf{c}_{inicial}$, deve estar contida em algum cone de mínimo local único S_n (os sinais + e - foram omitidos por clareza).
- 2) A condição inicial deve levar a saída a satisfazer (5.51).
- 3) A solução ZF_n deve estar contida em (caso possa haver equalização perfeita), ou muito próxima a, $T \cap S_n$.

Os autores afirmam que a inicialização *center-spike* (todos os elementos de \mathbf{w} nulos, exceto o central, que vale um) irá garantir a obediência a 1) caso o valor máximo da resposta ao impulso do canal ocorra para um único valor de n , o que é comum. Não se pode garantir a obediência a 2) *a priori*, pois o canal não é conhecido, e 3) pode ser satisfeita com a escolha de um comprimento suficientemente grande para o equalizador.

Os autores sugerem então que a inicialização *center-spike* aliada a procedimentos de correção periódica do “centro de massa” dos pesos do equalizador e a eventuais alterações no comprimento do filtro pode levar a uma situação bastante próxima das condições ideais descritas mais acima, para uma ampla gama de canais.

O artigo termina com a extensão dos resultados, obtidos no âmbito de um sinal transmitido PAM, ao caso mais genérico de sinais QAM.

5.15. *Mínimos Dependentes do Comprimento e do Custo e Considerações sobre a Presença de Ruído*

Num artigo publicado em 1996, Li, Ray Liu e Ding [Li et al., 1996] utilizaram o mesmo paradigma de análise empregado na seção anterior para lidar com o estudo de dois fatores distintos que levam à ocorrência de mínimos locais num equalizador cego: finitude de seu comprimento e a própria função custo escolhida.

Para elucidar os resultados mais importantes deste trabalho, vamos partir das seguintes perguntas: a existência de mínimos locais é universal dentro da família de algoritmos cegos? Pode haver algum critério convexo de equalização autodidata? Dentro de que condições?

Para que entendamos a abordagem proposta pelos autores, partamos de duas condições genéricas envolvendo um critério de equalização autodidata qualquer:

C1) As soluções do tipo ZF são os únicos mínimos globais da função custo do critério no espaço conjunto canal + equalizador. Note que isto não significa que estas soluções pertençam sempre ao *attainable set*.

C2) A função custo do critério é contínua no espaço conjunto real canal + equalizador.

Estas condições não são, de fato, muito restritivas. Aliás, são condições que devem ser obedecidas por qualquer critério razoável. Isto constitui um ponto muito positivo da análise dos autores, pela generalidade das hipóteses de trabalho.

Um primeiro resultado forte dos autores é a prova, usando os conceitos geométricos comentados na seção anterior, de que todo equalizador finito adaptado por um critério que satisfaça C1 e C2 pode ter seu estado atraído por um mínimo local. Esta afirmação é, certamente, muito forte, e praticamente “sela o destino” da filtragem adaptativa cega: não se deve esperar que um critério autodidata, mesmo que projetado de maneira razoável, seja convexo no espaço de parâmetros do equalizador, caso este tenha um número finito de graus de liberdade. Vale frisar que funções custo convexas e não-supervisionadas baseadas em restrições [Kennedy e Ding, 1992] não se encaixam no cenário descrito acima, pois normalmente não satisfazem C1.

Estes mínimos que surgem quase que universalmente nos critérios autodidatas são chamados de *mínimos locais dependentes do comprimento*, pois estão associados à finitude da resposta ao impulso do equalizador.

Os autores mostram que há uma outra classe de mínimos, associada a técnicas que empregam não-linearidades de natureza descontínua. Um exemplo desta classe de técnicas é o algoritmo DD, que tem em seu cerne a não-linearidade do decisor, tipicamente caracterizada por apresentar “saltos”. Outro exemplo é o chamado algoritmo *stop-and-go* (SGA), sendo que podemos ainda mencionar, em termos estruturais, o próprio *decision feedback equalizer* (DFE). Nestes casos, os pontos de mínimo local ocorrem mesmo para o caso duplamente infinito, quer dizer, *não são dependentes do comprimento do equalizador* e sim da *estrutura da função custo*, sendo chamados de *mínimos locais dependentes do custo*.

Ora, podemos inferir que há um tipo de mínimo local compartilhado por todos os critérios autodidatas considerados razoáveis, e há um tipo de mínimo presente apenas em algumas técnicas. Isto produz uma clara distinção entre os métodos, a saber, entre aqueles que só possuem mínimos dependentes do comprimentos e os que possuem ambos os tipos de mínimos.

No trabalho há ainda uma análise da influência do ruído no comportamento de um algoritmo cego. São analisados o CMA e o *standard cumulant algorithm* (SCA), de quem o

algoritmo de Shalvi-Weinstein é um caso particular. É demonstrado que a adição de ruído branco e gaussiano faz com que as condições ZF não mais sejam os mínimos globais das funções custo no espaço conjunto, o que é razoável pois a condição ZF tem a ver apenas com a IIS, não se ligando ao ruído. Mostra-se, no entanto, que para altas SNRs o desvio da solução ótima em relação à solução ZF é muito pequeno.

Um outro resultado interessante é uma análise comparativa de desempenho para o caso ruidoso. Os autores afirmam que, em linhas gerais, o desempenho dos algoritmos cegos é pior que o da solução de Wiener, mas melhor que uma simples inversão do canal, que não leve em conta o ruído.

É interessante notar que, como veremos a seguir, Regalia e Mboup [Regalia e Mboup, 1999] questionam algumas constatações expostas em [Li et al., 1996], no decorrer prova de que equalizadores autodidatas finitos obedecendo C1 e C2 devem apresentar mínimos locais. Teremos a chance de discutir um pouco mais esta “controvérsia” na próxima seção.

5.16. Equalizadores Subparametrizados

O artigo de Regalia e Mboup desenvolve uma análise estática envolvendo o conceito de *equalização subparametrizada*. O significado deste termo é que não é possível uma perfeita inversão do canal, ou seja, não é possível atingir alguma configuração ZF. A análise por eles empreendida leva a muitos resultados interessantes. Tentaremos expor, de maneira simples, alguns resultados essenciais.

O critério estudado pelos autores, proposto por Donoho [Regalia e Mboup, 1999] baseia-se na maximização de uma relação entre cumulantes do sinal de saída do tipo:

$$\frac{\text{cum}_{2p}[y(n)]}{\{\text{cum}_2[y(n)]\}^p} \quad (5.52)$$

onde $\text{cum}_{2p}(\cdot)$ é o cumulante normalizado de ordem $2p$. Maximizar (5.52), como mostram os autores, equivale a maximizar a seguinte grandeza:

$$D_{2p}(\mathbf{s}) = \frac{\|\mathbf{c}\|_{2p}}{\|\mathbf{c}\|_2} \quad (5.53)$$

$$\text{onde } \|\mathbf{c}\|_{2p} = \begin{cases} (\sum_n |c(n)|^{2p})^{1/2p}, & 2p < \infty \\ \sup_n |c(n)|, & p \rightarrow \infty \end{cases}$$

O interesse de se estudar (5.53) provém do fato de tal função ter como caso particular o critério de Shalvi-Weinstein, quando $2p = 4$, o que nos trará também informação sobre o critério CM, como veremos na próxima seção.

Talvez o principal resultado do artigo seja a constatação de que os máximos de (5.53) estarão tanto mais próximos das correspondentes soluções de Wiener quanto maior for o valor de p empregado. Isto indica que o uso de estatísticas de ordem cada vez mais elevada leva a pontos críticos mais próximos dos receptores de Wiener. Trata-se de um resultado muito forte e interessante, que traz novas luzes à concepção da relação entre critérios supervisionados e autodidatas.

Pode ser então feita a pergunta: qual a razão de não serem usados então critérios com estatísticas de ordem muito elevada? Um motivo é que a otimização de tais critérios envolve estimadores com variância muito grande, o que pode ser muito problemático do ponto de vista prático.

Podemos ainda citar, como um resultado relevante, uma condição necessária obtida para que um certo ponto, pertencente ao *attainable set*, seja um máximo de D_{2p} , para p finito.

Por fim, os autores contestam um argumento presente no trabalho de Li, Ray Liu e Ding [Li et al., 1996], mostrando que é possível que o estado do sistema penetre um dado cone mesmo que *não haja* um mínimo a ser atingido em seu interior, ao contrário ao que dão a entender os autores em seu trabalho. Tal argumento não foi explicitado em nossa análise.

5.17. A Equivalência entre os Critérios de Godard e de Shalvi-Weinstein

5.17.1. A Abordagem de Li e Ding

No mesmo artigo analisado em 5.14 [Li e Ding, 1995], os autores demonstram, pela primeira vez, a equivalência entre os critérios de Shalvi-Weinstein [Shalvi e Weinstein, 1990] e Godard. O que isto quer dizer?

Não é correto imaginar que tal equivalência signifique uma igualdade entre os critérios ou menos ainda entre os algoritmos CMA e de Shalvi-Weinstein (SWA). Na verdade, o grande resultado obtido pelos autores foi a demonstração de que há uma correspondência biunívoca entre os mínimos dos dois critérios, para todas as possíveis ordens de equalizador. Tentemos entender como tal resultado foi obtido.

A base teórica desta demonstração é a mesma dos resultados expostos na seção 5.14, e é feita a suposição típica de que a fonte gera símbolos i.i.d. . Em primeiro lugar, é enfocado o caso do equalizador duplamente infinito. Ora, como vimos na seção 5.3 e tivemos a chance de discutir em diversas seções posteriores, para este caso, a função custo CM só tem mínimos globais, sendo que cada um deles corresponde a uma solução ZF. Este mesmo resultado foi obtido para o SWA em [Shalvi e Weinstein, 1990]. Desta forma, para este caso particular, a equivalência já estava estabelecida, pelo menos para os pontos de mínimo. Faltava estabelecê-la para equalizadores com comprimento finito, e foi justamente isto que Li e Ding fizeram.

Antes de prosseguirmos, convém definir o critério de Shalvi e Weinstein (SW):

$$\max_w |K[y(n)]| \quad (5.54)$$

onde o operador $K(\cdot)$ é a *kurtosis*, definida em (5.50). Há ainda, superposta ao critério, uma restrição de igualdade de potência entre o sinal na saída do equalizador e na entrada do canal:

$$E[|y(n)|^2] = E[|s(n)|^2] \quad (5.55)$$

Para realizar a tarefa a que se propuseram, eles partiram dos conceitos expostos em 5.14 e de uma propriedade que não explicitamos, a saber, que se um \mathbf{c} pertence a um cone S_n^\pm , então $\alpha\mathbf{c}$ também pertence a este mesmo cone, caso α seja positivo. O que Li e Ding fazem, então, é mostrar que, se \mathbf{c}_0 é um ponto de mínimo do critério de Shalvi-Weinstein, então ele estará direta e unicamente associado a um ponto de mínimo do critério CM. Esta análise envolve tomar derivadas de primeira e segunda ordem, porém, de natureza escalar, de forma que se possa considerar como parâmetro de otimização o fator de escala α (daí o uso da propriedade que acabamos de expor). A seguir, é percorrido o caminho inverso, ou seja, é mostrado que um mínimo do critério CM também se associa de forma única a um mínimo do critério de Shalvi-Weinstein. Não achamos conveniente entrar nos pormenores da demonstração, preferindo dar uma visão geral do caminho adotado pelos autores, o que se adequa mais ao escopo deste trabalho.

5.17.2. A Abordagem de Regalia

O problema da equivalência entre os critérios CM e SW voltou a ser abordado em 1999, através de um trabalho de Phillip Regalia [Regalia, 1999]. O trabalho de Regalia aborda o problema de uma forma bastante distinta da que observamos em [Li e Ding, 1995], não sendo, de forma alguma, redundante. Qual foi então a abordagem proposta?

De maneira muito engenhosa, o autor escreve as funções custo CM e SW em termos de uma coordenada radial e $p - 1$ ângulos de rotação, que formam o vetor Θ , onde p é o número de parâmetros do equalizador. Para o caso complexo, o vetor Θ deve incluir também as $p-1$ fases dos coeficientes do filtro.

Assim, as funções custo são expressas como:

$$J_{CM}(\mathbf{r}, \Theta) \text{ e } J_{SW}(\mathbf{r}, \Theta) \quad (5.56)$$

Vale aqui ressaltar a forma como é definida a função custo de SW, diferente da exposta em (5.54) e (5.55):

$$J_{sw}(r, \Theta) = \frac{c_4[y(n)]}{\{E[y^2(n)]\}^2} \quad (5.57)$$

onde $c_4(b) = E(|b|^4) - 2.E^2[|b|^2] - E(b^2).E[(b^*)^2]$.

Em seguida, o autor escreve o vetor de pesos do equalizador como:

$$\mathbf{w} = r.\mathbf{w}_{norm} \quad (5.58)$$

onde \mathbf{w}_{norm} é o vetor de pesos normalizado, de tal forma que $\|\mathbf{w}_{norm}\| = 1$. Isto permite escrever:

$$y(n) = r.y_{norm}(n) \quad (5.59)$$

É mostrado então que a função custo de SW, expressa em (5.57), fornece o mesmo valor quando aplicada a $y(n)$ ou a $y_{norm}(n)$. Isto quer dizer que tal função é insensível ao componente radial do vetor de pesos. Portanto, podemos escrever:

$$J_{sw}(r, \Theta) = J_{sw}(\Theta) \quad (5.60)$$

o que certamente é um resultado muito importante.

Destacamos que, no entanto, o mesmo raciocínio não se aplica à função custo de Godard, que *depende* do componente radial de \mathbf{w} . Como é possível estudar conjuntamente estas funções, apesar desta diferença tão marcante? Talvez esta seja a idéia fundamental de todo o artigo. Regalia opta por analisar não a função custo em si, mas sim a *superfície de erro reduzida* do critério CM. Do que consiste esta superfície?

Como vimos, o componente radial é o aparente “pomo da discórdia” entre os dois critérios. Haverá alguma forma de “eliminá-lo” do cenário, de forma a se buscar então relações entre os critérios? A idéia proposta é, *primeiramente*, otimizar a função custo do critério CM *com respeito* ao componente radial. Isto significa impor:

$$\frac{\partial J_{\text{CM}}(r, \Theta)}{\partial r} = 0 \quad (5.61)$$

obtendo assim os valores ótimos de r . O autor obtém três soluções para (5.61). A primeira, $r = 0$, corresponde ao máximo local, que não nos interessa aqui. As duas outras soluções compõem o par:

$$r_{\text{ot}} = \pm \sqrt{R_2 \cdot \frac{E(|y_{\text{norm}}(n)|^2)}{E(|y_{\text{norm}}(n)|^4)}} \quad (5.62)$$

Como na função custo CM só nos importará o quadrado de r_{ot} , as duas soluções são equivalentes. A função:

$$J_{\text{CMred}}(\Theta) = J_{\text{CM}}(\pm r_{\text{ot}}, \Theta) \quad (5.63)$$

é a superfície de erro reduzida, ou seja, a função custo CM já otimizada para o componente radial. Obtemos, então, um função dependentes apenas dos componentes angulares, que era o que queríamos, para ter uma base adequada de comparação com o critério SW.

Chega-se então ao principal resultado formal do artigo, expresso por:

$$J_{\text{CMred}}(\Theta) = R_2^2 \left(1 - \frac{1}{3 + J_{\text{SW}}(\Theta)} \right) \quad (5.64)$$

para $y(n)$ pertencente ao conjunto dos números reais.

O que (5.64) nos diz? Ora, vemos explicitada a íntima relação entre as funções-custo CM e SW. Nas palavras de Regalia, a superfície de erro reduzida associada ao critério CM é uma deformação da superfície de erro associada ao critério SW. O operador responsável pela deformação tem duas propriedades interessantes: é de natureza monotônica, o que se infere da relação extraída de (5.64):

$$\frac{dJ_{\text{CMred}}(\Theta)}{dJ_{\text{SW}}(\Theta)} = \frac{R_2^2}{[3 + J_{\text{SW}}(\Theta)]^2} > 0 \quad (5.65)$$

e preserva ordem, ou seja,

$$J_{\text{CMred}}(\Theta_1) > J_{\text{CMred}}(\Theta_2) \Leftrightarrow J_{\text{SW}}(\Theta_1) > J_{\text{SW}}(\Theta_2) \quad (5.66)$$

De (5.65), também pode ser inferido que todos os pontos críticos da função custo CM são também pontos críticos da função SW. Isto, matematicamente, é expresso pela relação:

$$\frac{dJ_{\text{SW}}(\Theta)}{d\Theta} = 0 \Leftrightarrow \frac{dJ_{\text{CMred}}(\Theta)}{d\Theta} = 0 \quad (5.67)$$

Adicionando a condição (5.66), mostra-se que, além de terem em comum os pontos críticos, a classificação (mínimo global, sela, etc.) de cada ponto coincide para os dois critérios.

Isto conclui a análise para o caso real. Para o caso complexo, são impostas *condições de circularidade* sobre a saída $y(n)$, que levam a:

$$E[y_{\text{norm}}^2(n)] = E\{[y_{\text{norm}}^*(n)]^2\} = 0 \quad (5.68)$$

Feita esta restrição, é obtida uma expressão ligeiramente diferente para a deformação:

$$J_{\text{CMred}}(\Theta) = R_2^2 \left(1 - \frac{1}{2 + J_{\text{SW}}(\Theta)} \right) \quad (5.69)$$

que, no entanto, também obedece às propriedades de monotonicidade e preservação de ordem. Portanto, as mesmas conclusões obtidas para o caso real se aplicam ao caso complexo obedecendo a (5.68).

Caso a saída não obedeça às restrições de (5.68), como pode ocorrer em certos processamentos não-lineares, pode não haver equivalência entre os pontos críticos.

Cabe agora uma recapitulação sobre os resultados deste artigo profundamente rico. Em primeiro lugar, observemos as hipóteses feitas pelo autor. Para que tudo o que discutimos valha, é preciso pouco mais do que a condição de estacionariedade até quarta ordem para o sinal de entrada do equalizador. Não importa se os símbolos transmitidos são i.i.d. ou correlacionados, se o canal é linear ou não-linear, se o ruído é aditivo ou multiplicativo. Isto mostra que o escopo de validade das constatações apresentadas é muito amplo.

Quanto aos resultados, cabe aqui aproveitar uma discussão feita pelo próprio autor, que esclarece qual tipo de equivalência foi obtida entre os métodos CM e SW. Basicamente, eles têm em comum: susceptibilidade à existência de mínimos locais biunivocamente associados, mesma capacidade de reduzir IIS, mesma degradação média do desempenho na presença de ruído ou de correlação da fonte. Pode, entretanto, haver diferença no desajuste (*misadjustment*) ou na velocidade de convergência, em decorrência do tipo de deformação que relaciona as funções.

Em nossa opinião, o trabalho mencionado aborda de forma profunda e muito elegante a relação entre os dois critérios, com a grande vantagem de ser relativamente pouco restritivo com respeito às hipóteses de trabalho.

5.18. *Relação entre Receptores Baseados nos Critérios CM e Wiener*

Há uma tríade de trabalhos [Zeng et al., 1998, Zeng et al., 1999 e Gu e Tong, 1999] que destacaremos pela abordagem de um problema muito interessante: a relação entre diferentes critérios de equalização. Basicamente, o que se deseja entender é: qual a relação entre um ponto de mínimo de um critério autodidata (como o CM, por exemplo) e um mínimo de outro critério essencialmente distinto, como o de Wiener ou mesmo o critério ZF? Trata-se, por certo, de uma questão muito relevante, e que permeou, ao menos de forma mais “pontual”, outros trabalhos por nós analisados. Qual a relevância desta questão?

Vimos que, para um equalizador implementável, há mínimos globais e locais na função custo do critério CM. Podemos nos perguntar: qual a “qualidade” destas soluções, ou seja, a que tipo de desempenho elas conduzem? Para respondermos esta pergunta, precisamos de uma base comparativa com um outro paradigma conhecido. A tríade citada utiliza como referência dois receptores clássicos: o de Wiener e o ZF, que já discutimos várias vezes ao longo deste trabalho. Portanto, podem ser usados como termo de comparação o erro quadrático médio (MSE) de diferentes mínimos do critério autodidata estudado ou a sua relação com uma certa solução ZF.

Iremos analisar, separadamente, cada trabalho da tríade. Começemos por [Zeng et al., 1998], que foi o primeiro submetido à publicação. Nas palavras dos próprios autores, o objetivo do seu trabalho foi “revelar conexões entre receptores CM e de Wiener”. As conexões por eles procuradas podem ser sumarizadas em três questões:

Q1) Qual a localização dos mínimos locais do critério CM?

Q2) Qual o desempenho, em termos de erro quadrático médio (MSE), do receptor CM?

Q3) Qual a relação entre o receptor de Wiener e o receptor CM?

Não é o objetivo dos autores responder perfeitamente às questões acima, mas buscar, de alguma forma, aprender mais sobre o critério CM a partir das respostas parciais obtidas.

Em seu trabalho, os autores não contemplaram o modelo que vínhamos analisando com mais frequência: o de um equalizador cujos pesos estão espaçados de um período de símbolo (*baud rate*). O escopo de sua análise engloba estruturas como o equalizador fracionário e um *array* de antenas. Os sinais são supostos binários, i.i.d., e o ruído é suposto AWGN.

A base da análise deste artigo é, nas palavras dos próprios autores, “geométrica”. A idéia básica é baseada em um teorema, dito de Weierstrass, e pode ser sumarizada como: encontrar uma região compacta B com uma fronteira ∂B e uma referência interna \mathbf{w}_{ref} tal que o custo (em termos do critério CM) associado aos pontos da fronteira seja maior que o associado à referência \mathbf{w}_{ref} . Caso isto ocorra, se terá mostrado que há pelo menos um

mínimo em B . Trata-se de uma idéia bastante intuitiva, que pode ser muito útil caso B e \mathbf{w}_{ref} sejam escolhidas de forma a se proporcionar elos entre os critérios de Wiener e CM.

Uma vez exposta, em linhas gerais, a base teórica empregada pelos autores, passemos a uma discussão dos principais resultados obtidos. Consideramos que a melhor forma de apresentá-los é por meio das respostas dos próprios autores às três questões colocadas.

R1) O critério CM apresenta mínimos locais na presença de ruído. Aqui vale um comentário: lembramos ao leitor que *não foram estudados neste artigo equalizadores baud rate*, que são as estruturas que vínhamos analisando nesta dissertação, para as quais o comentário acima seria de pouca relevância, pois há mínimos locais também na ausência de ruído.

É interessante notar que, se o mínimo erro quadrático médio for menor que um certo valor, a existência de um receptor CM na vizinhança do receptor de Wiener pode ser estabelecida. Vemos aqui uma típica abordagem comparativa, pois a idéia de erro quadrático médio está associada exclusivamente ao receptor de Wiener e, no entanto, é utilizada também para se obter resultados acerca de um equalizador cego. Mínimos do critério CM em outras regiões não são enfocados no artigo, sendo obtidos, portanto, mínimos “bons”, com baixo erro quadrático médio.

R2) Foi obtido um limitante para o erro quadrático médio destes “bons” receptores CM, o qual depende das características do receptor de Wiener associado. Os resultados do artigo mostram que o limitante é bastante razoável, em todos os exemplos estudados.

R3) Mostrou-se que o equalizador CM e o de Wiener são aproximadamente colineares para as hipóteses de trabalho feitas no artigo. É obtida uma expressão para o erro quadrático médio em excesso do receptor CM com relação ao de Wiener.

Gostaríamos de mencionar um aspecto interessante: há, para um mesmo canal, vários possíveis receptores de Wiener. A diferença entre eles provém da escolha do sinal desejado. Por exemplo, se tivermos sinais provenientes de múltiplos usuários incidindo em

um *array* de antenas, podemos escolher o sinal transmitido por cada um deles como sinal desejado e cada escolha nos conduz a uma solução de Wiener distinta. Os autores mostram que há uma associação entre estes receptores de Wiener e os diferentes mínimos da função custo CM.

Um outro trabalho dos mesmos autores, publicado em 1999 [Zeng et al., 1999], pode ser visto como uma extensão do trabalho que acabamos de comentar. Nele, os autores também consideram, em sua análise, os equalizadores *baud rate*. Isto nos traz de volta ao universo da estrutura que temos analisado com mais detalhe ao longo desta tese. Isto é, de fato, muito enfatizado pelos autores, que consideram a principal contribuição de seu trabalho “o desenvolvimento de um procedimento sistemático para a análise de receptores CM”. A maior generalidade deste trabalho é também atestada pela consideração de uma fonte real, e não apenas binária como no trabalho anterior [Zeng et al., 1998]. A idéia geral de análise foi mantida, sendo também baseada no teorema de Weierstrass e na escolha de uma referência e do conjunto B de forma a se estabelecer um elo entre os dois receptores.

O caso genérico conduziu a resultados similares aos do trabalho anterior, a saber: o estabelecimento de um limitante superior, em termos de erro quadrático médio, para o desempenho do receptor CM; considerações sobre o erro quadrático médio “em excesso” do receptor CM, e a constatação de que os dois receptores são aproximadamente colineares no espaço de estados.

Estas considerações levam à conclusão de que um receptor cego projetado segundo o critério CM alcança quase que o mesmo desempenho, em termos de taxa de erro de bits (BER), do receptor linear ótimo, para altas SNRs.

Um artigo de Gu e Tong [Gu e Tong, 1999], o mais recente em termos da data de submissão, vem trazer algumas novas luzes a este problema. A análise do artigo se limita ao caso de um canal ortogonal e ao caso bidimensional, ou seja, aquele no qual a resposta conjunta canal + equalizador é dada por um vetor bidimensional. O canal ortogonal é aquele no qual as colunas da matriz de convolução são ortogonais. Ambos os casos não se encaixam no modelo *baud rate*. Destacamos também as hipóteses usuais, a saber, sinais transmitidos i.i.d. e ruído AWGN.

A base matemática do trabalho é a associação entre critérios de equalização e normas vetoriais distintas. Assim, otimização dos critérios de equalização passa a ser um

problema de minimizar/maximizar uma certa norma no espaço conjunto canal + equalizador. São abordados os receptores de Wiener, ZF e SW, o qual, pela equivalência discutida em 5.17, estende o escopo da análise também ao critério CM.

Os autores mostram que, para o canal ortogonal, um modelo adequado para sistemas OFDM e CDMA síncrono, os receptores CM, Wiener, SW e ZF são colineares e correspondentes um-a-um, caso não haja ruído.

Para o caso bidimensional, mostram-se ainda alguns resultados interessantes:

- 1) Há pelo menos um, e não mais que dois receptores CM/SW. Isto indica que não há *soluções espúreas*, ou seja, ótimos de CM/SW sem associação com algum receptor de Wiener. É interessante notar que a definição deixa aberta a possibilidade de que só haja um receptor CM/SW, ou seja, de que haja uma solução de Wiener sem um receptor CM/SW correspondente. Isto quer dizer que, sem correspondência com uma solução de Wiener, não há existência de um receptor CM/SW, o que mostra que tais critérios tem uma íntima ligação com o receptor de Wiener e um desempenho razoavelmente próximo ao deste.
- 2) Os receptores CM/SW estão sempre localizados em setores determinados pelos receptores de Wiener. Mostra-se que a relação sinal-interferência (SIR) associada a um receptor CM/SW é sempre menor que a associada ao receptor de Wiener correspondente. Esta relação permite a obtenção de um limitante para a posição dos receptores CM/SW.
- 3) Um resultado surpreendente é também obtido: a diminuição da SNR aproxima a SIR do receptor CM/SW da apresentada pelo receptor de Wiener. Isto quer dizer que a diminuição da SNR torna menos díspare o desempenho dos dois receptores.

5.19. Breve Discussão: o Paradigma Geométrico

As cinco seções anteriores formam o cerne dos trabalhos envolvendo o paradigma que denominamos *geométrico*. Os trabalhos de Li e outros, que introduziram a idéia de cones de mínimo global único lançaram, por assim dizer, as bases do enfoque geométrico. Destacamos também a demonstração de Regalia para a equivalência entre os critérios CM e SW, de uma elegância matemática marcante. Os trabalhos de Zeng et al. e de Gu e Tong trouxeram novas e importantes luzes acerca das relações entre receptores cegos e

supervisionados. Suas novas abordagens aumentaram ainda mais as possibilidades da análise estática dos critérios autodidatas. Por fim, o trabalho de Regalia e Mboup voltou aos conceitos expostos por Li et al. trazendo novos resultados de grande relevância.

5.20. *Uma Análise Dinâmica de Equalizadores Cegos*

Enquanto a análise estática de um algoritmo adaptativo procura obter a localização de seus pontos críticos, em especial de seus mínimos, a análise dinâmica diz respeito à convergência do estado para uma determinada solução.

O artigo de 1996, de autoria de Li e Ray Liu [Li e Ray Liu, 1996], que já analisamos na seção 5.10, traz também resultados sobre a dinâmica do algoritmo e algumas considerações sobre o comprimento do equalizador. Iremos analisar estes dois aspectos nesta seção.

Uma hipótese forte feita pelos autores, mas que é imprescindível para a análise, é supor que $\varepsilon(n)$ e $s(n)$ são estatisticamente independentes, sendo $\varepsilon(n)$ dado por:

$$\varepsilon(n) = \mathbf{w}(n) - \mathbf{w}_{\min} \quad (5.70)$$

Trata-se, certamente, de uma aproximação nos moldes das aproximações de independência tomadas para a análise dinâmica do LMS em 4.5.2. Mas, como os autores afirmam, tais aproximações são usuais em estudos teóricos envolvendo algoritmos adaptativos, pela grande simplicidade por elas proporcionada.

Um resultado muito interessante obtido pelos autores diz respeito à clássica restrição sobre o passo de adaptação μ para que o equalizador convirja adequadamente. Vimos este tipo de restrição em nosso estudo sobre o LMS, mais especificamente, em (4.66). Para o caso do CMA, procura-se obter a condição sobre o passo para que haja convergência na vizinhança do mínimo global. Os dois pilares da análise empreendida são a hipótese de independência que acabamos de discutir e a aproximação de primeira ordem (5.34). A condição obtida sobre o passo para haver convergência nas proximidades de \mathbf{w}_{\min} é:

$$0 < \mu < 2/\lambda_{F_{\max}} \quad (5.71)$$

onde $\lambda_{F_{\max}}$ é o maior autovalor de $\mathbf{R}_{f_{\min}}$, definida por:

$$\mathbf{R}_{f_{\min}} = \mathbf{K}^T \cdot \mathbf{F}_{\min} \cdot \mathbf{K} \quad (5.72)$$

onde \mathbf{K} é a matriz de convolução definida em (5.35) e \mathbf{F}^{ot} é definida como:

$$\mathbf{F}_{\min} = \text{diag}[f_{\min}(1), \dots, f_{\min}(1), f_{\min}(0), f_{\min}(1), \dots, f_{\min}(1)] \quad (5.73)$$

onde aparece $M+N$ vezes o valor $f_{\min}(1)$, uma vez $f_{\min}(0)$ e outras $M+N$ vezes o valor $f(1)$. Vale lembrar que o canal tem $2M+1$ parâmetros e o equalizador tem $2N+1$.

Os valores $f_{\min}(0)$ e $f_{\min}(1)$ são definidos como:

$$f_{\min}(0) = (1/\sigma_s^2) \cdot E\{ \phi'[y_{\min}(n)] \cdot s^2(n) \} \text{ e } f_{\min}(1) = E\{ \phi'[y_{\min}(n)] \} \quad (5.74)$$

onde $y_{\min}(n)$ é saída do equalizador para a configuração \mathbf{w}_{\min} e σ_s^2 é a variância do sinal. Novamente, consideramos todas estas expressões matemáticas como acessórias, sendo que o resultado (5.71) é o que consideramos mais relevante expor. Neste trabalho, convergir para o mínimo global significa ter:

$$E[\mathbf{w}(n)] \rightarrow \mathbf{w}_{\min} \quad (5.75)$$

Os autores ainda obtém uma equação que define unicamente a matriz de autocorrelação de $\varepsilon(n)$, envolvendo também $\mathbf{R}_{f_{\min}}$ e uma outra matriz. É apresentado também o domínio de validade desta equação em relação a μ . Estudar $\varepsilon(n)$ é muito importante, pois permite que sejam obtidas informações valiosas sobre o aprendizado do algoritmo. É obtida uma expressão envolvendo este erro que julgamos válido reproduzir:

$$E[\varepsilon(n)] = (\mathbf{I} - \mu \cdot \sigma_s^2 \cdot \mathbf{R}_{f_{\min}})^n \cdot E[\varepsilon(0)] \quad (5.76)$$

onde o passo μ precisa obedecer à restrição (5.71) para que a expressão (5.76) acabe por conduzir à situação desejada, dada por (5.75).

Por fim, os autores analisam a influência do comprimento do equalizador no desempenho do algoritmo. Há dois fatores a analisar, um envolvendo um aspecto estático e o outro envolvendo um aspecto dinâmico do algoritmo. Definamos a *distorção devida ao comprimento finito do equalizador* como sendo:

$$D_f = \|\mathbf{c}_{\min} - \mathbf{c}_{ZF}\| \quad (5.77)$$

onde \mathbf{c}_{\min} é o ponto no espaço conjunto canal + equalizador associado a \mathbf{w}_{\min} . D_f nos dá uma medida de quão próxima está a configuração do equalizador daquela que sabe-se ser ótima. É intuitivo que, quanto maior for o comprimento do equalizador, menor será o valor desta medida de distorção. Isto poderia levar-nos a concluir que quanto maior for o comprimento do equalizador, melhor será o desempenho do algoritmo. Isto não é de todo correto, como tentam mostrar os autores, pois há um outro tipo de distorção que também precisa ser levada em conta.

Esta segunda distorção é causada pelo chamado ruído de gradiente, já discutido no contexto do LMS em 4.5.2. Este ruído decorre da estimativa instantânea utilizada no cálculo do gradiente, a qual dá origem às expressões (5.4) e (5.5). Tal distorção pode ser definida como:

$$D_g = E [\|\mathbf{c} - \mathbf{c}_{\min}\|^2] \quad (5.78)$$

Esta expressão indica quão próxima está a estimativa obtida pelo algoritmo do melhor valor que ela poderia obter. Quanto maior o ruído de gradiente, mais distante estará o estado do sistema, em média, do valor ótimo da função custo.

Através da análise dinâmica empreendida, os autores obtêm uma expressão aproximada para D_g :

$$D_g \approx \mu \cdot (2N+1) \cdot \delta \quad (5.79)$$

onde δ é uma constante que depende das características do sinal transmitido.

O que (5.79) nos indica? Ora, supondo um valor fixo de μ , percebemos que o aumento da ordem do equalizador leva a um aumento do ruído de gradiente, o que piora, em média, a qualidade da solução obtida pelo algoritmo. Assim, aumentar o comprimento do equalizador indiscriminadamente não é necessariamente uma boa política. Certamente D_f diminuirá, mas D_g tenderá a aumentar, o que acaba por comprometer o panorama geral do desempenho. Pode-se argumentar que μ pode ser diminuído de modo a “acomodar” o aumento do equalizador, mas isto significaria uma diminuição na velocidade de convergência do algoritmo. É preciso, pois, analisar *conjuntamente* os efeitos dos parâmetros de projeto sobre o valor das duas distorções.

5.21. Breve Discussão: Análise Dinâmica

Um resultado muito interessante obtido em [Li e Ray Liu, 1996] foi a obtenção de um limitante para o passo de adaptação, para que o estado convirja nas vizinhanças do mínimo global. De forma geral, foi um trabalho que muito contribuiu para o entendimento da dinâmica estocástica do algoritmo, estando, de certa forma, ligado a alguns resultados obtidos anteriormente em [Cusani e Laurenti, 1995].

Porém, em todos os trabalhos que tivemos a oportunidade de ler, a análise dinâmica de um algoritmo adaptativo sempre se restringiu ao binômio “convergência para um ponto de mínimo/divergência”. Este binômio aplica-se muito bem a algoritmos supervisionados como o LMS, mas não necessariamente esgota as possibilidades de um sistema não-linear.

No próximo capítulo, tentaremos explorar alguns aspectos da convergência do algoritmo do módulo constante que não são enfocados usualmente pelas análises de convergência por nós conhecidas. Mostraremos que o CMA, sob algumas condições, pode apresentar convergência para ciclos-limite e atratores estranhos, o que deixa claras as limitações da visão acerca de convergência “herdada” da análise clássica de algoritmos supervisionados.

6. Dinâmica Caótica em Algoritmos do Módulo Constante

No capítulo anterior, apresentamos um vasto panorama de estudo sobre o critério e algoritmo do módulo constante, desde sua concepção até a evolução dos conceitos relativos à sua análise de convergência. Nele vimos diversas idéias que foram formando grande parte do conhecimento atual acerca desta técnica, na forma de um breve tutorial.

Como já pudemos perceber, a partir de uma comparação entre os capítulos 4 e 5, a não-linearidade inerente aos algoritmos de Bussgang torna-os muito mais complexos que os algoritmos baseados no paradigma de Wiener. Iremos mostrar que esta maior complexidade dá origem a comportamentos estranhos à intuição que muitas vezes temos, baseada quase sempre nos resultados do capítulo 4, relativos ao algoritmo LMS. Para tanto, fundaremos nossa investigação no padrão de análise usual de sistemas dinâmicos, exposto no capítulo 2. Nesta análise, iremos focar a expressão iterativa do CMA como um sistema dinâmico N -dimensional, onde N é o número de parâmetros do equalizador. Caso estejamos lidando com o algoritmo na forma direta ou determinística, sua expressão será:

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \cdot \nabla J^{(2)}(n) \quad (6.1)$$

onde $J^{(2)}(n)$, como vimos antes, é a função custo do critério CM. Aproximando a esperança matemática pelo valor instantâneo, obtemos a versão convencional ou estocástica:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \cdot e^*(n) \cdot \mathbf{x}(n) \quad (6.2)$$

onde

$$e(n) = y(n) \cdot [R_2 - |y(n)|^2] \quad (6.3)$$

É importante observar que, embora utilizemos o mesmo símbolo μ em (6.1) e (6.2), há uma diferença de fator de escala entre eles, da qual iremos nos abstrair por questão de simplicidade.

Usualmente, como vimos, as análises de convergência têm seu foco em dois aspectos: a convergência para um ponto fixo (um dos mínimos da função custo) ou a divergência (que corresponde a uma “explosão para o infinito”). Mostraremos que o algoritmo CMA, em ambas as versões, pode apresentar um comportamento dinâmico muito mais rico e situado, em termos do passo de adaptação, entre estes dois casos. A única afirmação levantada por nós neste sentido, que supunha a possibilidade de ocorrer caos no algoritmo (6.1), está em [Frater et al., 1995].

Nosso objetivo foi o de buscar caracterizar os possíveis comportamentos do algoritmo nesta região intermediária pouco explorada, mas para nós relevante do ponto de vista do estudo da convergência da técnica. Consideramos ainda digna de nota a complexidade que pode emergir de um critério conceitualmente simples, mas que devido à utilização de EOS dá origem a um cenário dinâmico de significativa riqueza. É muito importante destacar que toda a dinâmica do processo *está relacionada ao processo iterativo de otimização*, sendo a função custo em cada caso fixa e independente do tipo de busca empregado ou de seus parâmetros.

6.1. Introdução à Análise Dinâmica

A análise dinâmica que realizaremos objetiva a caracterização dos possíveis comportamentos e cenários apresentados pelo CMA, visto como um sistema não-linear. Nossa análise se dividirá em duas partes distintas, uma dedicada ao algoritmo determinístico e uma ao algoritmo estocástico. Em ambos os casos, tomaremos como parâmetro de bifurcação o passo de adaptação μ , que é, de fato, uma variável de projeto.

Alguns aspectos são comuns a todos os modelos por nós adotados neste trabalho, e serão destacados já:

- 1) As amostras do sinal transmitido $s(n)$ são i.i.d. (independentes e identicamente distribuídas), assumindo os valores $+A$ e $-A$ com igual probabilidade.

2) O equalizador é um filtro transversal linear, com ordem igual à do canal, adaptado pelo algoritmo do módulo constante em uma das duas versões já citadas.

Passemos aos resultados por nós obtidos. Começaremos por um caso simples, para que possamos nos familiarizar com o processo de análise e com a aplicação dos métodos descritos nos capítulos 1 e 2. Passaremos então à análise do algoritmo determinístico em diversos modelos de canal. Em seguida estudaremos o algoritmo estocástico, também em vários contextos, onde algumas idéias necessárias para a análise serão propostas. Um breve sumário conclui as discussões relativas tanto ao caso determinístico quanto ao estocástico.

6.2. *Ocorrência de Caos no Algoritmo Determinístico*

6.2.1. *Estudo de um Caso Simples: um Canal de Ordem Zero*

Em nosso primeiro modelo, a função de transferência do canal é dada por:

$$H_1(z) = K \quad (6.4)$$

onde K é um ganho real. O estado de nosso equalizador, neste caso, é dado por um único valor, $w(n)$. Este equalizador é capaz de compensar perfeitamente a atuação do canal (que no caso, se restringe a uma alteração de ganho).

Para o modelo proposto de canal, temos:

$$x(n) = K.s(n) \quad (6.5)$$

e

$$y(n) = w(n).x(n) = K.s(n).w(n) \quad (6.6)$$

Aplicando as equações (6.6) e (6.5) a (6.2), obtemos:

$$w(n+1) = w(n) + \mu.K^2.A^4.w(n).[1 - K^2.w^2(n)], \quad (6.7)$$

que pode ser escrita como:

$$w(n+1) = F[w(n)] \quad (6.8)$$

Percebe-se que, neste caso particular extremamente simples, mesmo a aplicação de (6.2) leva a uma expressão determinística, o que acaba unificando as duas versões da técnica.

Impondo a condição de equilíbrio (2.11), obtemos três pontos:

$$w_{e1} = 0 \text{ e } w_{e2,3} = \pm 1/K \quad (6.9)$$

O primeiro ponto é um máximo local e os outros dois são pontos de mínimo global, correspondentes à condição ZF. Tendo obtido os pontos de equilíbrio, passemos ao estudo da estabilidade dos mesmos. Para tanto, recorreremos aos conceitos expostos no capítulo 2. No caso do sistema acima, a matriz jacobiana é na verdade um escalar, dado por:

$$F'[w(n)] = 1 + \mu.K^2.A^4 - 3.\mu.K^4.A^4.w^2(n) \quad (6.10)$$

Para o ponto w_{e1} , temos:

$$F'(w_{e1}) = 1 + \mu.K^2.A^4 \quad (6.11)$$

Percebemos que tal ponto não será estável para nenhum valor aceitável de μ , pois

$$\mu > 0 \rightarrow F'(w_{e1}) > 1 \quad (6.12)$$

Para os outros dois pontos, temos:

$$F'(w_{e2,3}) = 1 - 2.\mu.K^2.A^4 \quad (6.13)$$

A condição de estabilidade local (seção 2.5.3) é $|F'(w_{e2,3})| < 1$, o que nos leva à seguinte faixa de valores para o passo de adaptação:

$$0 < \mu < \frac{1}{K^2 \cdot A^4} \quad (6.14)$$

Portanto, em (6.14) obtivemos uma zona de operação satisfatória para o equalizador, na qual os pontos correspondentes às soluções ZF são localmente atraentes. Para valores de μ acima do limitante superior, certamente não se atingirá a solução ideal. Mas fica a pergunta: o que acontece?

Podemos pensar, se nos basearmos na intuição construída pela análise mais usual de algoritmos supervisionados, que terá lugar a divergência para o infinito. Porém, se formos mais criteriosos em nosso estudo, veremos que o algoritmo do módulo constante reserva em sua estrutura uma riqueza de comportamento típica dos sistemas não-lineares. Conforme vimos no capítulo 3, podemos fazer uso de um diagrama de bifurcação para melhor compreender o comportamento do sistema. Tal diagrama mostra a evolução da dinâmica do sistema quando é variado um certo parâmetro livre (em nosso caso, o passo de adaptação). Para $K = 2$, $A = 1$ e $w(0) = 0.1$, realizando 1000 iterações para cada valor de μ e retendo as 50 últimas para a apresentação, obtivemos a Fig. 6.1.

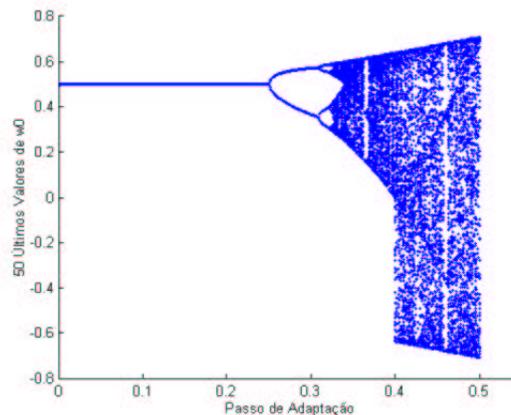


Figura 6.1: Diagrama de Bifurcação de $w(n)$

Na faixa de valores de μ que vai de 0 a 0.25, há uma operação satisfatória. Isto confirma a validade de (6.14) para o caso estudado. A convergência para 0.5, ao invés de -0.5, dá-se pela inicialização feita em um valor positivo.

Quando μ sai da faixa de estabilidade do ponto fixo, no entanto, não ocorre a divergência, e sim uma bifurcação de Hopf (seção 2.6.2). Aumentando ainda mais o passo, surge uma cascata de bifurcações de duplicação de período (seção 2.6.3), que só cessam, após infinitas ocorrências, para um certo valor de μ próximo a 0.32. Após este valor, inicia-se uma região de comportamento caótico, que podemos distinguir no diagrama, observando suas partes “cheias”, onde ocorre dependência sensível das condições iniciais. Dentro desta região, há janelas de periodicidade, indicando o aparecimento de ciclos-limite estáveis. Para valores de μ levemente maiores que 0.5 dá-se, enfim, divergência.

O cenário que acabamos de descrever não é estranho ao leitor, tendo sido descrito na seção 3.6, no contexto de um outro sistema dinâmico, o mapa logístico. Recomendamos que o leitor atente para a análise apresentada em tal seção, pois a essência da evolução dinâmica do mapa unidimensional é a mesma do CMA em sua versão determinística, para todas as ordens por nós testadas.

Um outro aspecto interessante merece ser ressaltado. Para μ próximo de 0.4, há uma mudança substancial no atrator estranho associado ao mínimo global. Podemos ver isto através do alargamento da banda de oscilação caótica, observável no diagrama. Isto ocorre justamente quando o estado começa a assumir valores negativos. Na verdade, se olharmos com atenção, veremos que o que acontece é uma “fusão” entre atratores estranhos: o associado ao ponto de mínimo 0.5 e o associado ao ponto -0.5, que não víamos em virtude da inicialização, antes da ocorrência desta espécie de crise [Grebogi et al., 1983].

Buscando um maior embasamento matemático para todas estas conclusões, calculamos o expoente de Lyapunov associado ao mapa unidimensional expresso em (6.7). Para tanto, foram utilizadas 5000 iterações da fórmula direta (3.15), para cada valor de μ , a partir da mesma condição inicial do diagrama. O resultado está na Fig. 6.2.

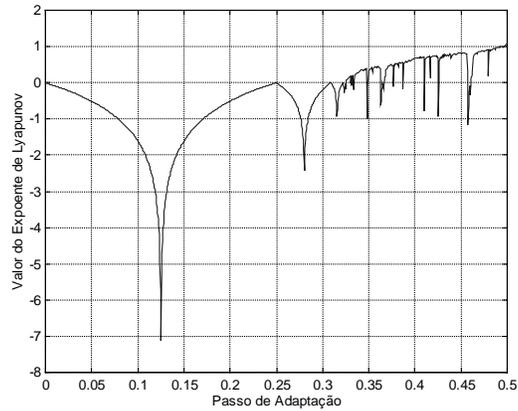


Figura 6.2: Exponente de Lyapunov

Conforme era esperado, o expoente é negativo para as zonas de convergência para pontos fixos ou ciclos-limite, nulo nas bifurcações e positivo na região de caos, o que confirma a nossa análise anterior, fundada apenas no diagrama. Algumas janelas de periodicidade podem ser percebidas como picos negativos na região caótica.

Este primeiro caso confirmou algumas de nossas suposições: de fato, há um comportamento dinâmico bastante rico entre a região de convergência para um dos pontos de equilíbrio e a região de divergência. Nesta região intermediária, há infinitos ciclos-limite estáveis, e uma zona na qual há DSCI (que também contém infinitos ciclos-limite, como discutimos anteriormente), o que confirma a suspeita de que a execução do algoritmo do módulo constante pode apresentar caos.

Para que possamos melhor entender como será o comportamento do algoritmo na região de caos, vejamos a evolução de $w(n)$ para um valor de μ pertencente à zona caótica. Utilizamos 500 iterações do algoritmo para produzir os dados, apresentadas na Fig. 6.3.

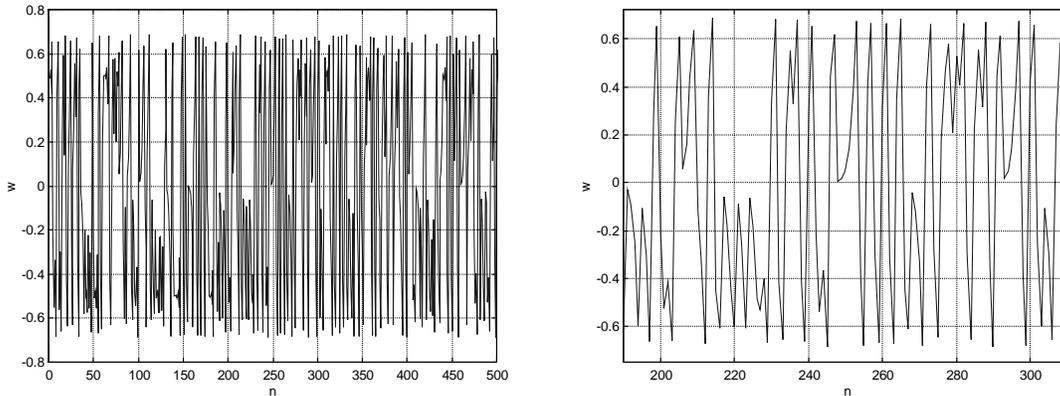


Figura 6.3: Evolução Temporal de $w(n)$

Observamos nesta figura as características usuais de uma série temporal caótica: aperiodicidade, uma certa auto-similaridade, variações rápidas, etc. Certamente, a operação de um equalizador para tal valor de μ seria um completo desastre do ponto de vista da equalização.

6.2.2. Canal MA(1) de Fase Mínima

Uma vez estabelecidas nossas linhas básicas de análise, passemos agora a modelos mais complexos e próximos dos adotados para sistemas reais. Voltemos, primeiramente, nossa atenção para um canal MA(1) com a seguinte função de transferência:

$$H_2(z) = 1 + \alpha.z^{-1} \quad (6.15)$$

cujo único zero, suposto real, está localizado em $-\alpha$. Nesta seção, adotaremos um modelo de fase mínima, ou seja, com $|\alpha| < 1$. Especificamente, escolhemos $\alpha = 0.6$. Portanto, $x(n)$ será expresso pela seguinte equação a diferenças:

$$x(n) = s(n) + \alpha.s(n-1) \quad (6.16)$$

O equalizador, que terá dois taps, tem seu estado no instante n representado pelo vetor:

$$\mathbf{w}(n) = [w_0(n), w_1(n)]^T \quad (6.17)$$

O uso de um equalizador de tal ordem faz com que a função custo $J^{(2)}$ assuma a seguinte forma:

$$J^{(2)}(w_0, w_1) = 1 + \theta_1 \cdot w_0^4 + 4 \cdot \theta_2 \cdot w_0^3 \cdot w_1 + 6 \cdot \theta_4 \cdot w_0^2 \cdot w_1^2 + 4 \cdot \theta_3 \cdot w_0 \cdot w_1^3 + \theta_1 \cdot w_1^4 - 2 \cdot \theta_5 \cdot w_0^2 - 4 \cdot \theta_6 \cdot w_0 \cdot w_1 - 2 \cdot \theta_5 \cdot w_1^2 \quad (6.18)$$

Os diversos θ_i são momentos e momentos cruzados dos componentes de $\mathbf{x}(n)$. Os momentos $\theta_1, \theta_2, \theta_3$ e θ_4 têm grau 4, sendo o primeiro o quarto momento de $x(n)$, e os demais, momentos cruzados de $x(n)$ e $x(n-1)$. Os momentos θ_5 e θ_6 são de grau 2, sendo o primeiro o segundo momento de $x(n)$, e o último expresso por $E[x(n) \cdot x(n-1)]$.

Calculando o gradiente a partir da expressão (6.18), pode-se aplicar (6.1), o que caracteriza o uso do algoritmo determinístico na busca de um dos mínimos da função custo. Embora não seja usado na prática, estudar este algoritmo é relevante para a compreensão do algoritmo estocástico, particularmente no contexto do presente capítulo.

O primeiro passo é procurar os pontos de equilíbrio de (6.18), que são os pontos críticos de $J^{(2)}$. Trata-se de um problema complexo, sendo que em [Gu e Tong, 1999] os autores até questionam a viabilidade da obtenção de uma solução analítica. Iremos nos limitar a uma análise da superfície de custo e à obtenção dos pontos para este modelo de canal específico. Na Fig. 6.4.a encontram-se as curvas de nível de (6.18), enquanto que na Fig. 6.4.b temos um gráfico que traz em si a informação acerca dos pontos críticos que se extrai do vetor gradiente. De fato, há duas curvas na Fig. 6.4.b, uma com o rótulo grad0 e a outra com o rótulo grad1. A primeira curva corresponde ao lugar geométrico no plano para o qual é anulado o primeiro elemento do vetor gradiente de (6.18). A segunda curva corresponde ao lugar geométrico para o qual é nulo o segundo elemento. A interseção entre as duas marca portanto os pontos críticos.

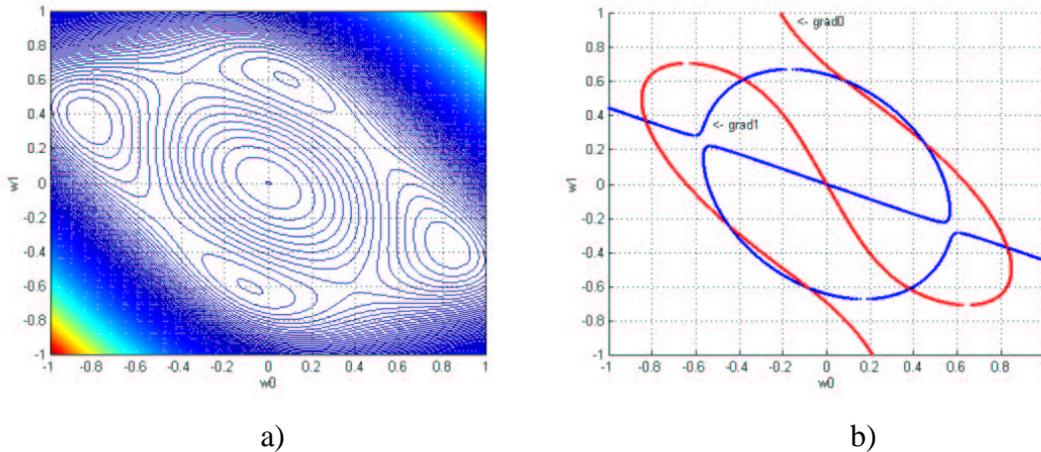


Figura 6.4: Curvas de Nível da Função Custo (a) e Lugares Geométricos de Gradiente Nulo (b)

É válido frisar que quando se lida com o problema da equalização cega, *não é possível* dispor da função custo, pois isto implicaria conhecimento do canal de comunicação. No entanto, para fins analíticos, é muito válido analisar gráficos como os mostrados na Fig. 6.4.

Observando a Fig. 6.4.b, percebemos que há 9 pontos críticos: dois pares de mínimos, dois pares de pontos de sela e um máximo. Tanto as selas quanto o máximo não são estáveis (seção 2.5.3), não sendo portanto soluções capazes de atrair o estado do equalizador. Apenas os mínimos o são, obviamente para valores de passo maiores que zero, que são os únicos aceitáveis num processo *steepest descent*. Portanto, limitaremos a análise a um elemento de cada par de mínimos. Vale recordar que os elementos de um par de mínimos são *sempre* simétricos, o que torna redundante a análise de ambos. O conjunto de análise se restringirá então a um mínimo global e a um local, que estão, aproximadamente, em:

$$\mathbf{P} = \{[0.831, -0.372]^T, [0.084, 0.607]^T\} \quad (6.19)$$

Em primeiro lugar, analisemos o mínimo global, primeiro elemento de \mathbf{P} . Para verificar a sua estabilidade em função de μ , utilizamos o cálculo do autovalor de maior magnitude da matriz de Floquet do sistema, calculada no ponto que se deseja analisar.

Como vimos na seção 2.5.3, caso este autovalor (ou multiplicador de Floquet) esteja dentro do círculo de raio unitário (CRU), o ponto será localmente estável. A relação da estabilidade local com o valor de μ decorre da dependência deste parâmetro por parte da matriz de Floquet. Na Fig. 6.5 encontramos um gráfico relacionando o autovalor de maior magnitude ao passo de adaptação.

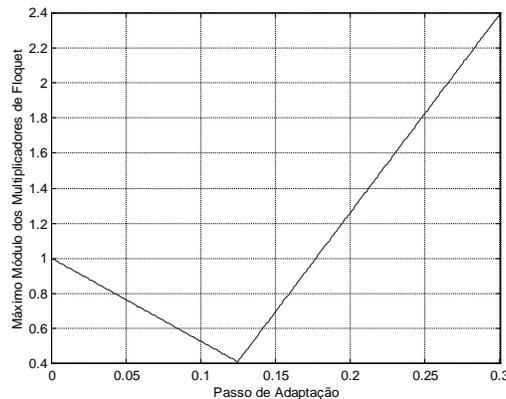


Figura 6.5: Multiplicador de Floquet de Maior Magnitude

Podemos perceber que a condição de estabilidade é obedecida para $0 < \mu < 0.1765$. Portanto, esta é a faixa de operação satisfatória para o mínimo global. Comparando a obtenção de uma faixa de estabilidade neste caso com o caso do método *steepest descent* aplicado num contexto de filtragem supervisionada (seção 4.5.1), verificamos que o limitante superior para μ dependerá do ponto de equilíbrio analisado, o que não ocorria de forma explícita para a técnica linear. É possível, portanto, explicitar uma faixa em termos algébricos, mas sempre em relação ao máximo autovalor da matriz de Floquet, que por sua vez dependerá do ponto crítico estudado.

Novamente, fica a questão do que ocorre quando o ponto de equilíbrio deixa de atrair trajetórias próximas. Utilizando uma condição inicial na vizinhança deste mínimo, obtivemos o diagrama de bifurcação para os dois pesos do filtro. Novamente, foram usadas 1000 iterações por valor de μ e exibidos os 50 últimos valores das variáveis w_0 e w_1 . Frisamos que esta configuração será mantida para todos os demais diagramas exceto onde for explicitado o contrário. O resultado da simulação está na Fig. 6.6.

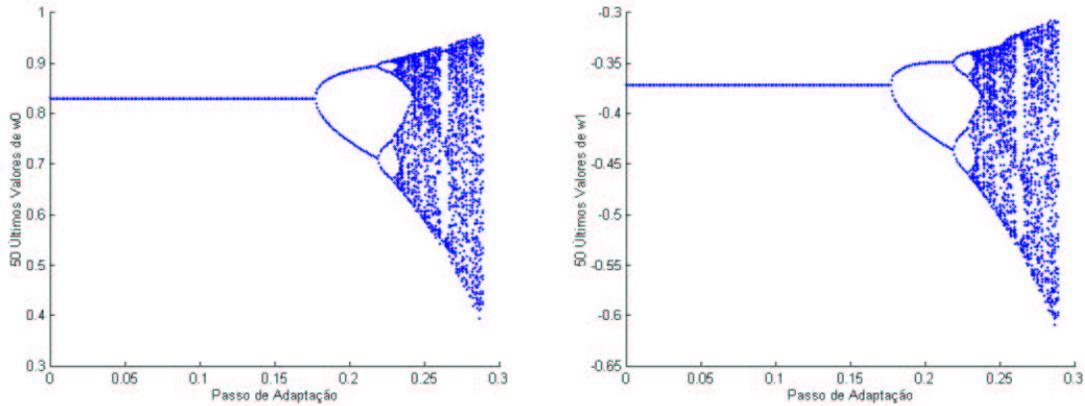


Figura 6.6: Diagramas de Bifurcação para o Mínimo Global ($\mathbf{w}(0) = [0.831, -0.372]^T$)

Após a perda de estabilidade por parte do mínimo, observamos mais uma vez a ocorrência do cenário de Feigenbaum, ou seja, de sucessivas bifurcações flip seguidas por uma zona caótica entremeada por regiões de periodicidade. Após a região de caos, há divergência. Vale ressaltar que trata-se do mesmo tipo de comportamento verificado no caso anterior.

Novamente, utilizamos os expoentes de Lyapunov. Desta vez, por ser um mapa bidimensional, haverá dois expoentes. Para a nossa análise, bastará o maior deles. O gráfico mostrando o valor deste expoente em função do passo de adaptação está na Fig. 6.7.

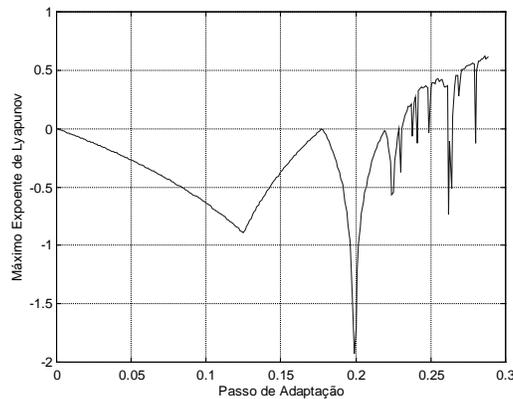


Figura 6.7: Máximo Expoente de Lyapunov

Novamente o expoente calculado confirma as nossas observações, apresentando valores condizentes com os diagramas da Fig. 6.6.

Para entendermos melhor como irá se comportar o sistema durante a operação em caos, vejamos na Fig. 6.8 como será o atrator estranho associado ao sistema para dois valores distintos de μ .

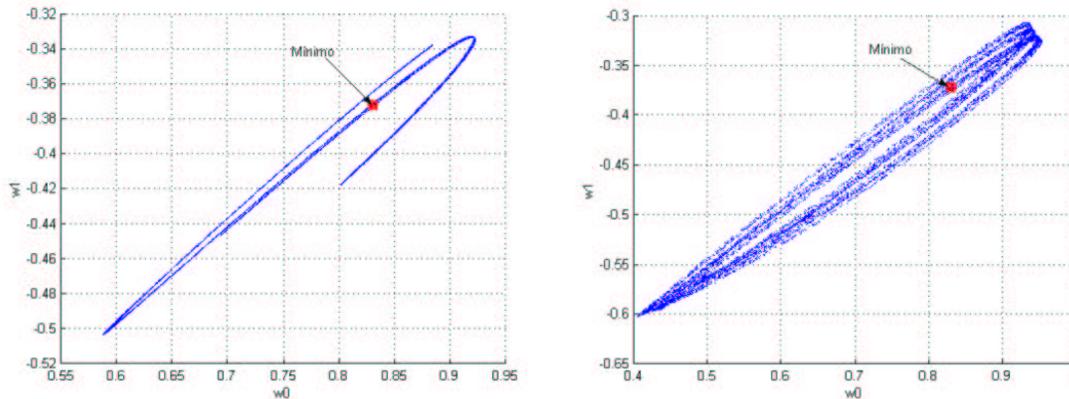


Figura 6.8: Atratores Estranhos Associados ao Mínimo Global para $\mu = 0.25$ e $\mu = 0.285$.

Podemos notar que o atrator estranho encontra-se nas proximidades do mínimo global, assinalado por um quadrado, convenção que adotaremos em todos os gráficos deste tipo. A estrutura de ambos apresenta algumas similaridades com a do atrator de Hénon (seção 3.4.1), emblemático dentro do subconjunto dos sistemas discretos. Vejamos agora, no domínio do tempo, como será a evolução do peso $w_0(n)$ para o valor $\mu = 0.25$. Na Fig. 6.9 temos 500 iterações desta variável juntamente com a ampliação de uma região específica do gráfico.

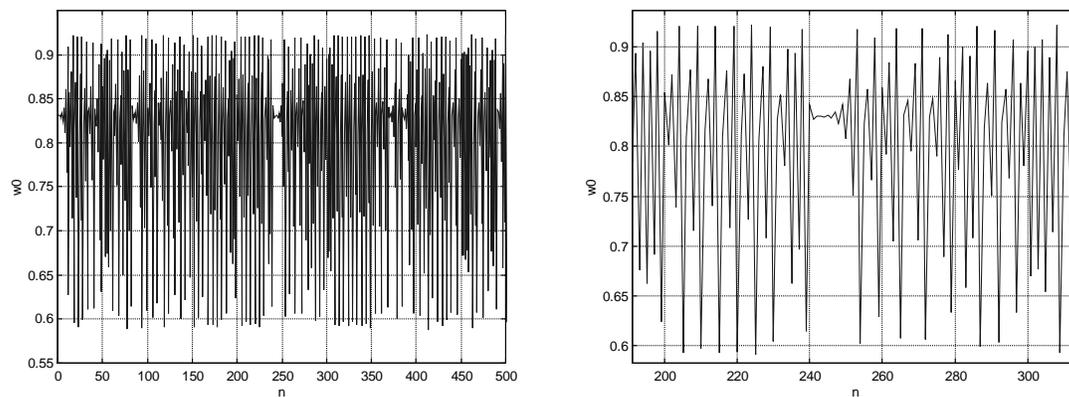


Figura 6.9: Evolução de $w_0(n)$ para $\mu = 0.25$

Neste diagrama, além das características gerais, vale ressaltar um aspecto interessante. Para n pouco maior que 240, observamos que o estado fica próximo do mínimo global, o que é razoável pela posição do atrator estranho. Porém, como o ponto já não é estável, o sistema vai escapando da vizinhança deste ponto, oscilando com amplitude cada vez maior, até ir finalmente para outra parte do atrator estranho. Trata-se de um comportamento similar ao de cenários do tipo Pomeau-Maneville [Schuster, 1988]. É portanto bastante aparente que um equalizador operando na região de caos, como será regra em todos os casos estudados, não atinge os objetivos da equalização.

Nossa análise ainda não está terminada. É preciso ainda estudar o mínimo local que escolhemos. Em primeiro lugar, vejamos sua faixa de estabilidade, através do multiplicador de Floquet de maior magnitude. A Fig. 6.10 traz as informações necessárias.

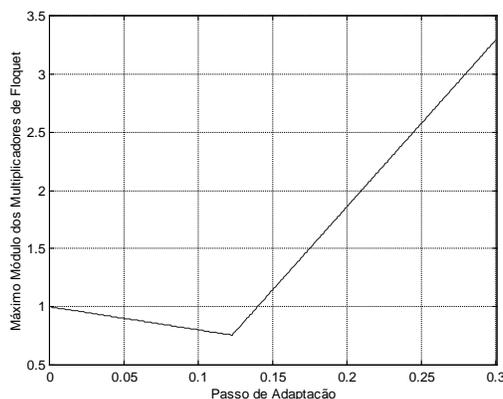


Figura 6.10: Multiplicador de Floquet de Maior Magnitude

A faixa de estabilidade para este mínimo foi de $0 < \mu < 0.1400$, tendo portanto um limitante menor que no caso anterior. Inicializando o algoritmo em um valor próximo ao mínimo local, foram obtidos os diagramas de bifurcação para as duas variáveis de estado, mostrados na Fig. 6.11. Novamente, a faixa de estabilidade do mínimo foi confirmada.

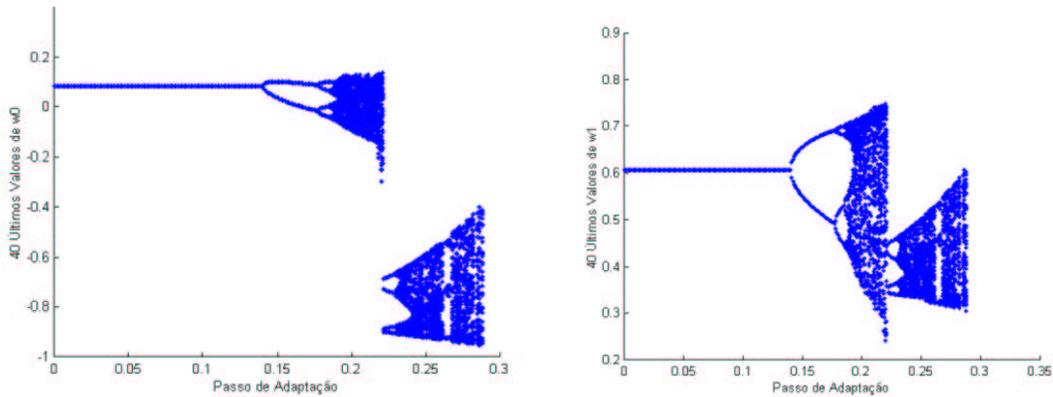


Figura 6.11: Diagramas de Bifurcação para o Mínimo Local ($\mathbf{w}(0) = [0.084 \ 0.607]^T$)

Tais diagramas expressam o mesmo tipo de evolução dinâmica dos casos anteriores, mas com um fato marcante. Para μ próximo de 0.22, o atrator estranho associado ao mínimo local perde a sua estabilidade. Porém, ao contrário do que ocorria nos casos anteriores, isto não provoca divergência para o infinito. Ocorre no entanto que o sistema passa a oscilar num ciclo-limite de período quatro, sofrendo novamente toda uma evolução subsequente até o caos e a explosão para o infinito. Porém, esta “segunda evolução” está intimamente ligada ao mínimo global, e não ao local, embora $\mathbf{w}(0)$ esteja muito próximo a este último mínimo. Portanto, o sistema passa a evoluir de forma análoga a quando a inicialização era feita próxima ao mínimo global. Trata-se de um fato muito interessante, que terá algumas implicações práticas descritas no capítulo 7.

Novamente, foi calculado o máximo expoente de Lyapunov do sistema para diversos valores de μ . O resultado está na Fig. 6.12.

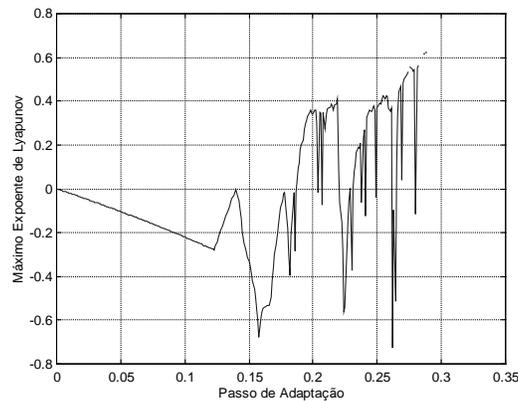


Figura 6.12: Máximo Expoente de Lyapunov

O expoente mostra bem a descontinuidade de comportamento por nós apontada, havendo um retorno para valores negativos quando o atrator estranho associado ao mínimo global perde a estabilidade. Os atratores estranhos associados ao mínimo local, para $\mu = 0.2$ e $\mu = 0.215$ estão na Fig. 6.13.

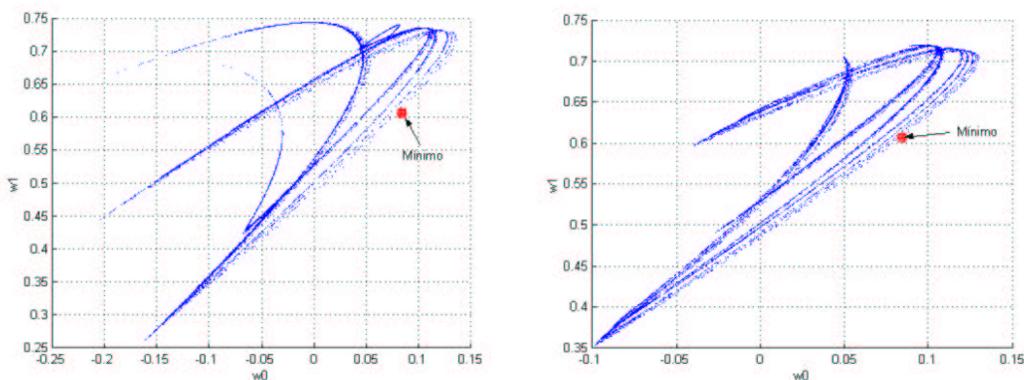


Figura 6.13: Atratores Estranhos Associados ao Mínimo Local para $\mu = 0.2$ e $\mu = 0.215$

Na Fig. 6.14, traçamos os dois atratores estranhos associados aos mínimos globais (usando $\mu = 0.25$) e os dois associados aos mínimos locais (usando $\mu = 0.2$), tendo ao fundo as curvas de nível da função custo, de forma a se estabelecer uma melhor noção da distribuição espacial de tais entes. A proximidade entre um dado atrator e seu respectivo mínimo é bastante evidente, como já havíamos comentado.

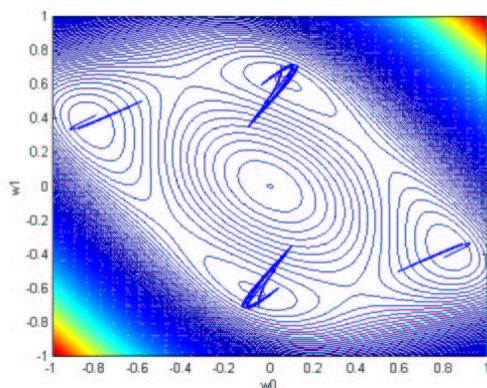


Figura 6.14: Posição dos Atratores Estranhos

Na Fig. 6.15, novamente exibimos a evolução temporal do peso w_0 num valor de passo para o qual o sistema apresenta comportamento caótico. Isto nos mostra as consequências deste regime para a convergência dos pesos do filtro.

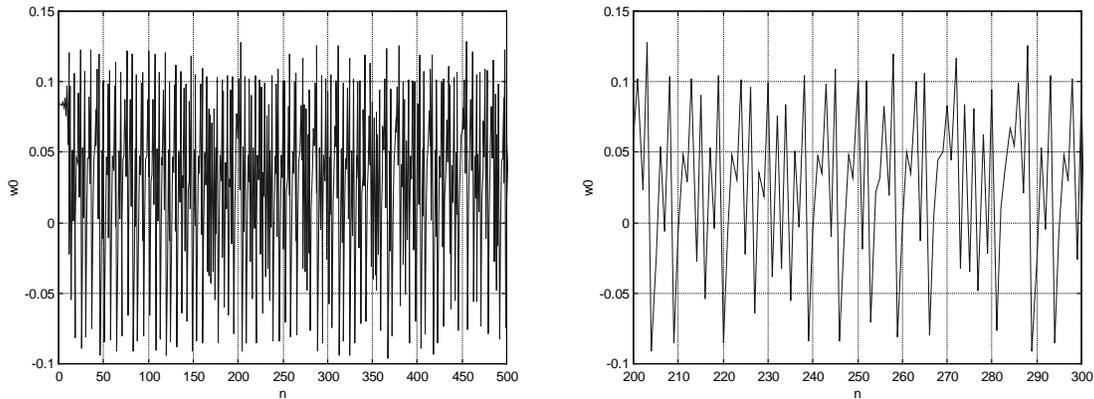


Figura 6.15: Evolução de $w_0(n)$ para $\mu = 0.2$

Antes de encerrarmos a seção, cabe um comentário. Muito da análise que fizemos é de caráter essencialmente local. Por exemplo, quando obtivemos as faixas de estabilidade, utilizamos a matriz de Floquet. Tal matriz, como vimos no primeiro capítulo, decorre de uma linearização em torno de um ponto de equilíbrio. Portanto, os resultados deste procedimento são precisos na medida em que se apliquem à vizinhança do ponto estudado.

O mesmo é válido para os diagramas de bifurcação, que expressam cenários dependentes da inicialização. E isto não apenas em termos da trajetória caótica, que como vimos é altamente dependente das condições iniciais, mas também em relação aos diversos tipos de comportamento e ao valor de μ para o qual começa a ocorrer divergência. A título de exemplo, traçamos os diagramas, na Fig. 6.16, para $\mathbf{w}(0) = [0.5 \ 0.5]^T$.

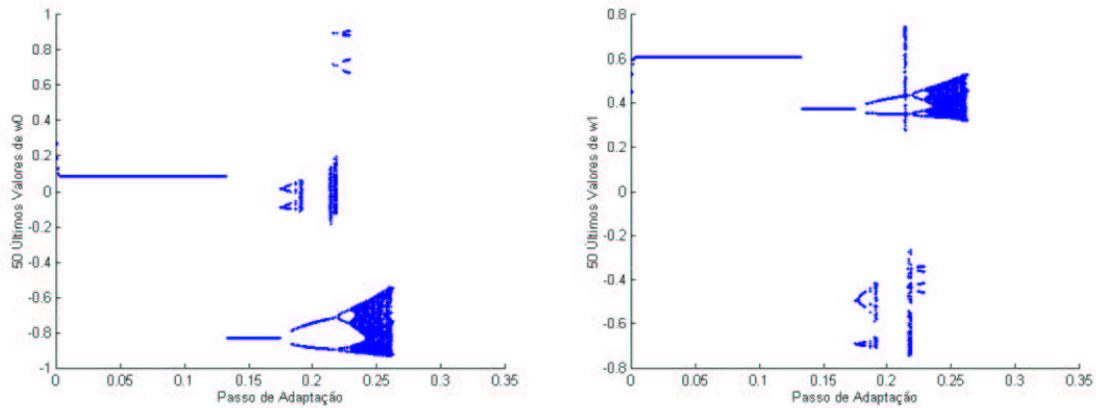


Figura 6.16: Diagramas de Bifurcação para $\mathbf{w}(0) = [0.5 \ 0.5]^T$

Como podemos imediatamente notar, a divergência para o infinito começa a ocorrer para um valor visivelmente menor que nos dois casos anteriores. Mas o que mais chama a atenção é a descontinuidade do diagrama, que contém trechos correspondentes à evolução da dinâmica do sistema a partir das vizinhanças de cada um dos quatro mínimos que, por simplicidade, passaremos a denominar “evolução dos mínimos”. Podemos ver segmentos similares aos diagramas traçados para o mínimo global e para o mínimo local, bem como trechos que estão ligados a diagramas que não mostramos (relativos aos dois mínimos que não analisamos). Utilizamos em nossa análise esta terminologia de “evolução”, mas vale deixar bem claro que esta evolução se dará de forma distinta para condições iniciais distintas.

6.2.3. Canal MA(1) de Fase Máxima

Neste caso, a função de transferência do canal é também dada por (6.15), mas com $|\alpha| > 1$. O valor por nós escolhido foi $\alpha = 1.5$. Vejamos as características da função custo através de suas curvas de nível, mostradas na Fig. 6.17.

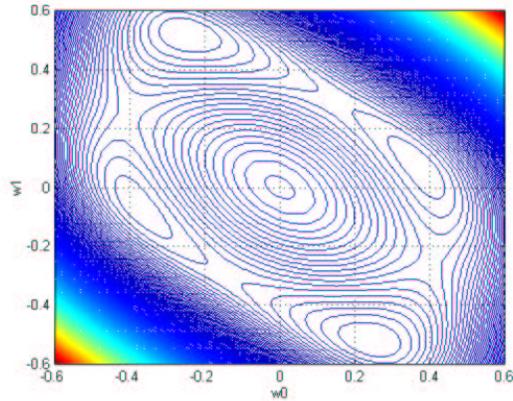


Figura 6.17: Curvas de Nível da Função Custo

As características gerais da superfície foram preservadas, havendo novamente 9 pontos representativos: dois pares de mínimos, dois pares de pontos de sela e um máximo. Os mínimos por nós escolhidos para análise foram:

$$\mathbf{P} = \{ [0.246 \ -0.522]^T, [0.378 \ 0.065]^T \} \quad (6.20)$$

O primeiro é global e o segundo, local. Passemos à análise do mínimo global. A Fig. 6.18 traz a relação entre o multiplicador de Floquet de maior magnitude e o passo de adaptação.

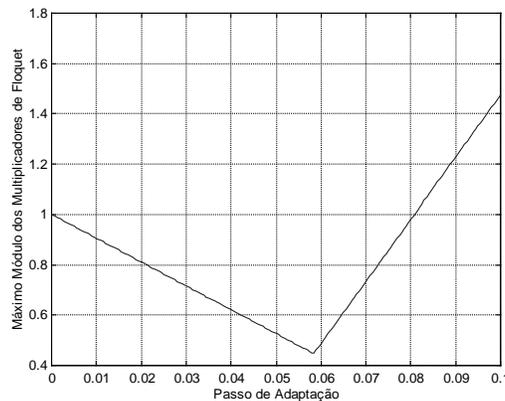


Figura 6.18: Multiplicador de Floquet de Maior Magnitude

A partir desta figura, obtemos a faixa de estabilidade $0 < \mu < 0.0808$. Utilizando um valor inicial próximo ao mínimo global, obtivemos os diagramas de bifurcação para as duas variáveis de estado, que estão na Fig. 6.19.

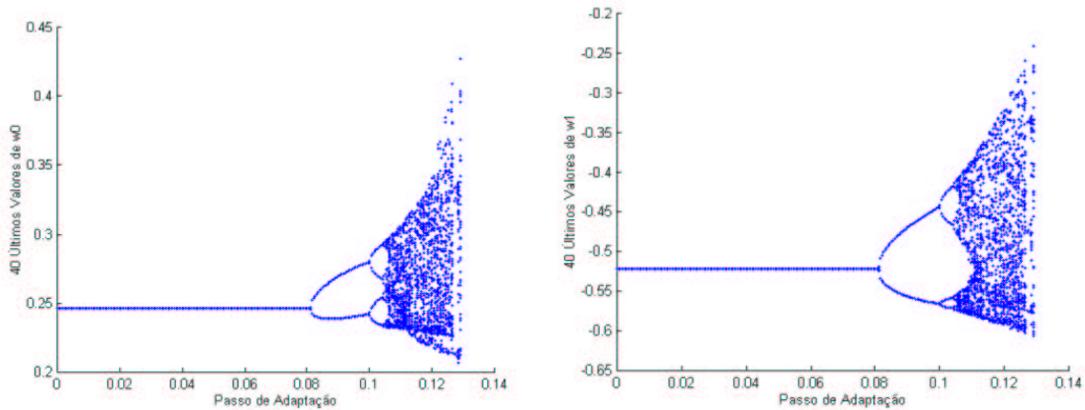


Figura 6.19: Diagramas de Bifurcação ($\mathbf{w}(0) = [0.246 \ -0.522]^T$)

Novamente, verifica-se uma região inicial de convergência para o ponto fixo, cuja extensão concorda com o valor obtido através da matriz de Floquet, tendo lugar em seguida o cenário de Feigenbaum. O máximo expoente de Lyapunov, para este caso, está na Fig. 6.20. Os resultados expressos neste gráfico confirmam as observações por nós feitas a partir da Fig. 6.19.

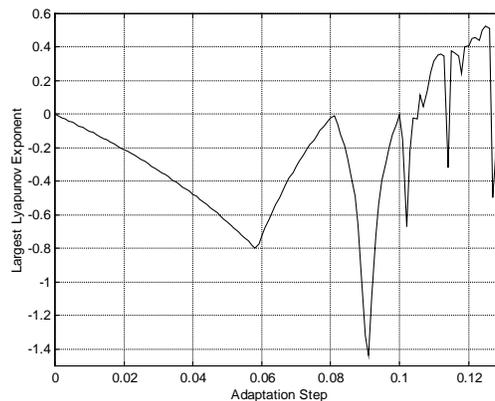


Figura 6.20: Máximo Expoente de Lyapunov

Na Fig. 6.21 estão os atratores estranhos associados ao mínimo global para dois valores distintos de μ (0.12 e 0.123).

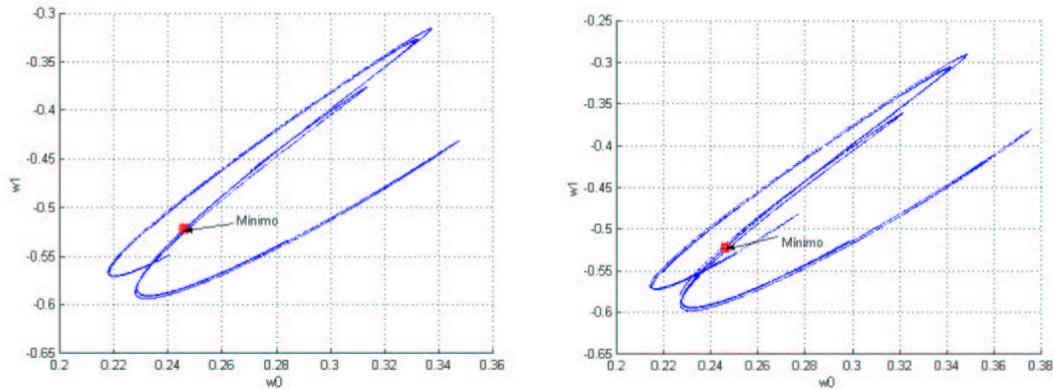


Figura 6.21: Atratores Estranhos Associados ao Mínimo Global para $\mu = 0.12$ e $\mu = 0.123$

Podemos notar uma interessante similaridade entre a estrutura destes atratores e a dos atratores da Fig. 6.8, também associados a um mínimo global. Por fim, vejamos como é a evolução temporal dos pesos quando variam segundo o primeiro dos atratores da Fig. 6.20 ($\mu = 0.12$). O gráfico correspondente e sua ampliação numa determinada faixa estão na Fig. 6.22.

Esta figura revela novamente o comportamento aperiódico típico, com alguns instantes de aproximação do mínimo global e um subsequente escape do mesmo, de forma análoga à da Fig. 6.9.

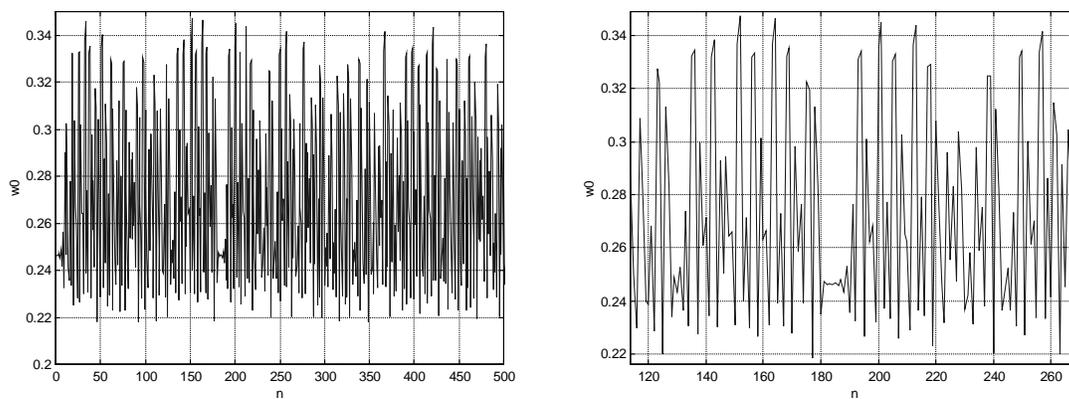


Figura 6.22: Evolução Temporal de $w_0(n)$ para $\mu = 0.12$

Passemos à análise do mínimo local escolhido. Em primeiro lugar traçamos o gráfico que relaciona o multiplicador de Floquet de máxima magnitude ao passo de adaptação. O resultado está na Fig. 6.23.

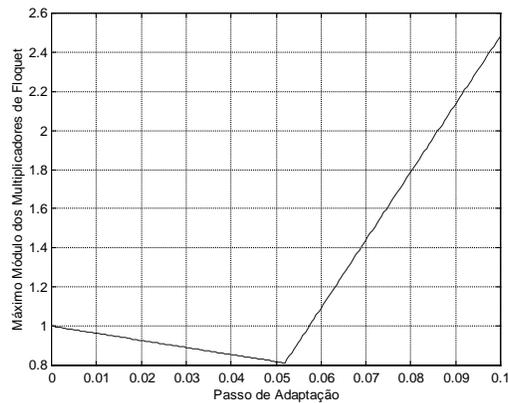


Figura 6.23: Multiplicador de Floquet de Maior Magnitude

A partir da Fig. 6.23, extraímos a faixa de estabilidade obtida para o mínimo local, que foi $0 < \mu < 0.0574$. Utilizando uma condição inicial próxima a este mínimo, obtivemos os diagramas de bifurcação relativos às duas variáveis de estado, ilustrados na Fig. 6.24.

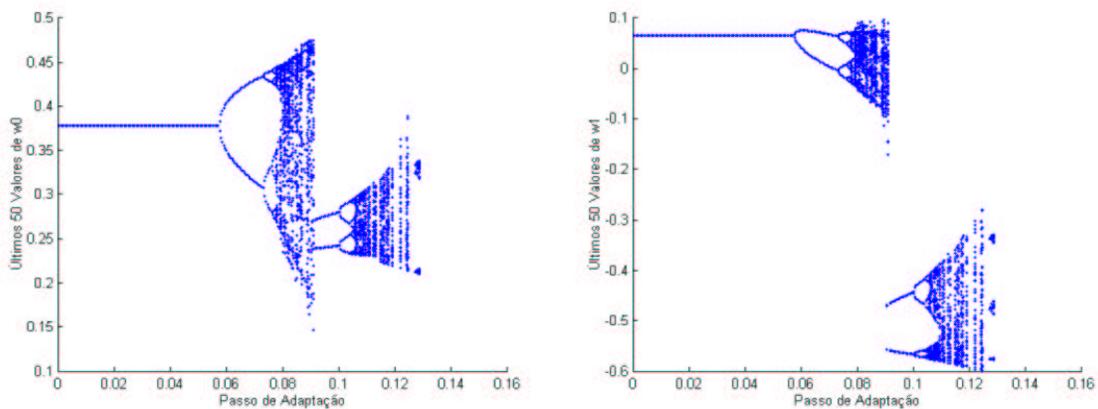


Figura 6.24: Diagramas de Bifurcação ($\mathbf{w}(0) = [0.378 \ 0.065]^T$)

Novamente, observamos uma descontinuidade na “evolução” dinâmica do sistema, sendo que o atrator estranho associado ao mínimo local perde sua estabilidade quando ainda é estável o ciclo-limite de período dois associado ao mínimo global. Isto faz com que

o estado convirja para um região do espaço de estados próxima a este ponto, levando a um comportamento similar ao verificado na Fig. 6.19. É interessante notar que este tipo de descontinuidade foi comum aos dois canais MA(1) que estudamos. É interessante notar, nos últimos valores de μ antes da divergência, o surgimento de uma pequena região cujas ramificações se alargam formando pequenas bandas.

Para este sistema, também calculamos o máximo expoente de Lyapunov associado para diversos valores de μ . O resultado está indicado na Fig. 6.25.

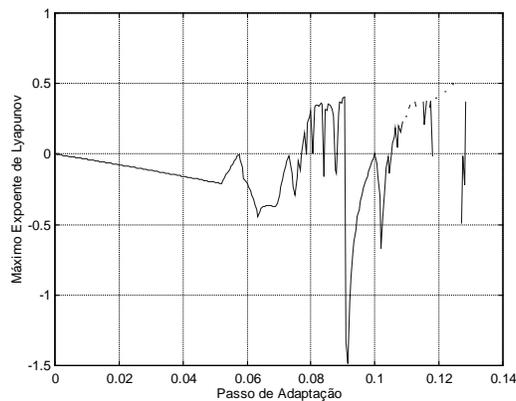


Figura 6.25: Máximo Expoente de Lyapunov

Novamente, os valores do expoente dão suporte à nossa análise e refletem adequadamente a perda “precoce” de estabilidade por parte do atrator estranho associado ao mínimo local. Na região entre $\mu = 0.11$ e $\mu = 0.13$, podemos notar descontinuidades no expoente, devidas à divergência para o infinito, já explicitadas pelo diagrama da Fig. 6.24.

Na Fig. 6.26 representamos o atrator estranho correspondente ao mínimo local para $\mu = 0.082$ e $\mu = 0.085$. O mínimo local está assinalado por um quadrado.

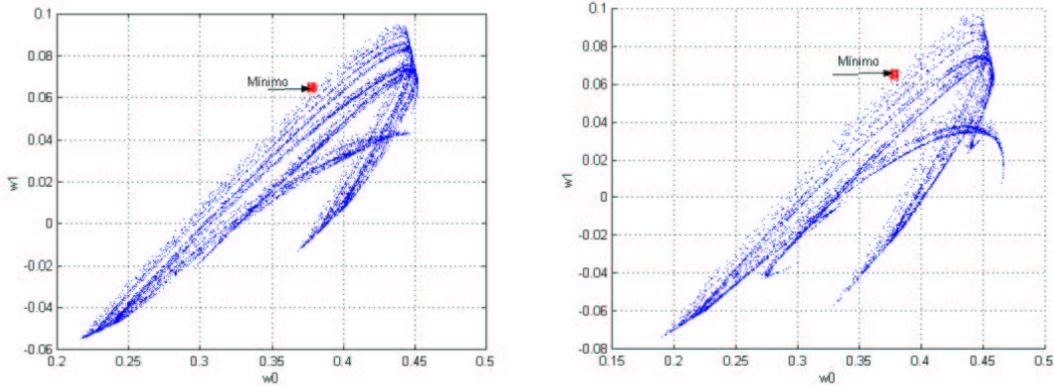


Figura 6.26: Atrator Estranho Associado ao Mínimo Local para $\mu = 0.082$ e $\mu = 0.085$

Finalmente, a evolução temporal de $w_0(n)$ para $\mu = 0.085$ encontra-se na Fig. 6.27, tendo características semelhantes às obtidas anteriormente para este tipo de curva.

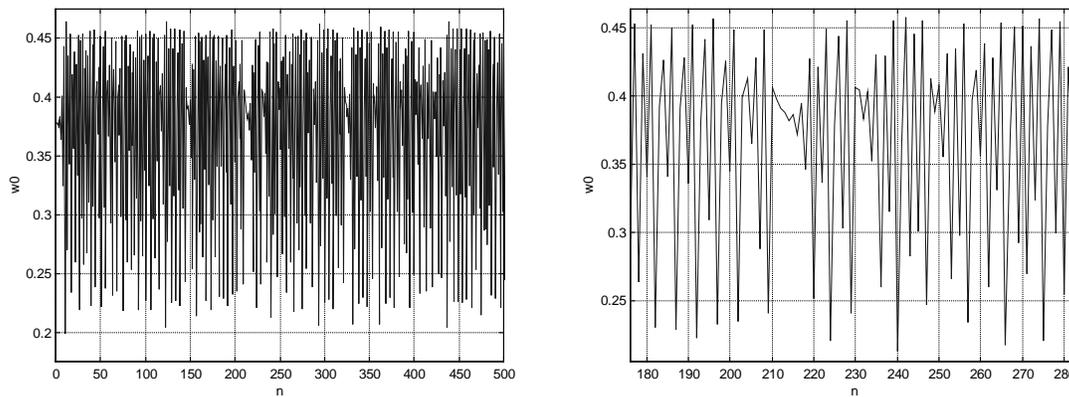


Figura 6.27: Evolução Temporal de $w_0(n)$ para $\mu = 0.085$

6.2.4. Canal MA(2) de Fase Mista

Nas duas seções anteriores, analisamos canais MA(1) com fase mínima e com fase máxima. Podemos afirmar que os dois exemplos formam um conjunto bastante abrangente dentro do modelo analisado, permitindo que se tenha uma idéia geral do tipo de cenário dinâmico imposto por este tipo de canal. Porém, da teoria clássica de filtragem sabemos que modelos MA(1) de canal (excetuando, certamente, canais com um nulo espectral) são passíveis de equalização por meio de um filtro branqueador, cuja determinação necessita

apenas de estatísticas de ordem dois. Portanto, os mesmos não são, por assim dizer, os canais que mais tornam necessário o uso de estatísticas de ordem superior, mesmo que implicitamente, como é o caso do CMA.

Escolhemos então um canal que impõe esta necessidade quando da equalização por meio de técnicas não-supervisionadas: um canal MA(2) de fase mista. A função de transferência de um canal MA(2) genérico é dada por:

$$H_3(z) = 1 + \alpha.z^{-1} + \beta.z^{-2} \quad (6.21)$$

Escolhemos $\alpha = 1.6$ e $\beta = 0.55$, fazendo com que os zeros da função fossem -1.1 e -0.5 . Estes valores fazem com que um dos zeros de $H_3(z)$ esteja fora do CRU e o outro dentro dele, caracterizando fase mista.

Como o equalizador irá ser de ordem dois, não é possível visualizar a superfície gerada a partir da função custo do CMA. Iremos, neste caso, analisar três mínimos, um global e dois locais com profundidades distintas, tendo o segundo menor custo que o terceiro. Analisemos, em primeiro lugar, o mínimo global, localizado aproximadamente em $\mathbf{w}_{e1} = [0.241 \ -0.589 \ 0.124]^T$.

Os diagramas de bifurcação associados a este ponto encontram-se na Fig. 6.28. Observamos que a evolução também se dá segundo o cenário de Feigenbaum, sendo que a faixa de estabilidade levantada foi aproximadamente $0 < \mu < 0.045$. A região de caos apresenta algumas descontinuidades nas proximidades da região de divergência, o que indica convergência para atratores associados a outros mínimos. Portanto, comprova-se que o mínimo com menor custo não tem necessariamente o atrator estranho com maior faixa de estabilidade, como se poderia pensar a partir de um raciocínio indutivo baseado nos modelos MA(1). Por motivos de simplicidade, optamos por não apresentar o cálculo do expoente de Lyapunov para este caso, e consideramos que a visualização do atrator estranho discreto em três dimensões não traria tantas contribuições significativas.

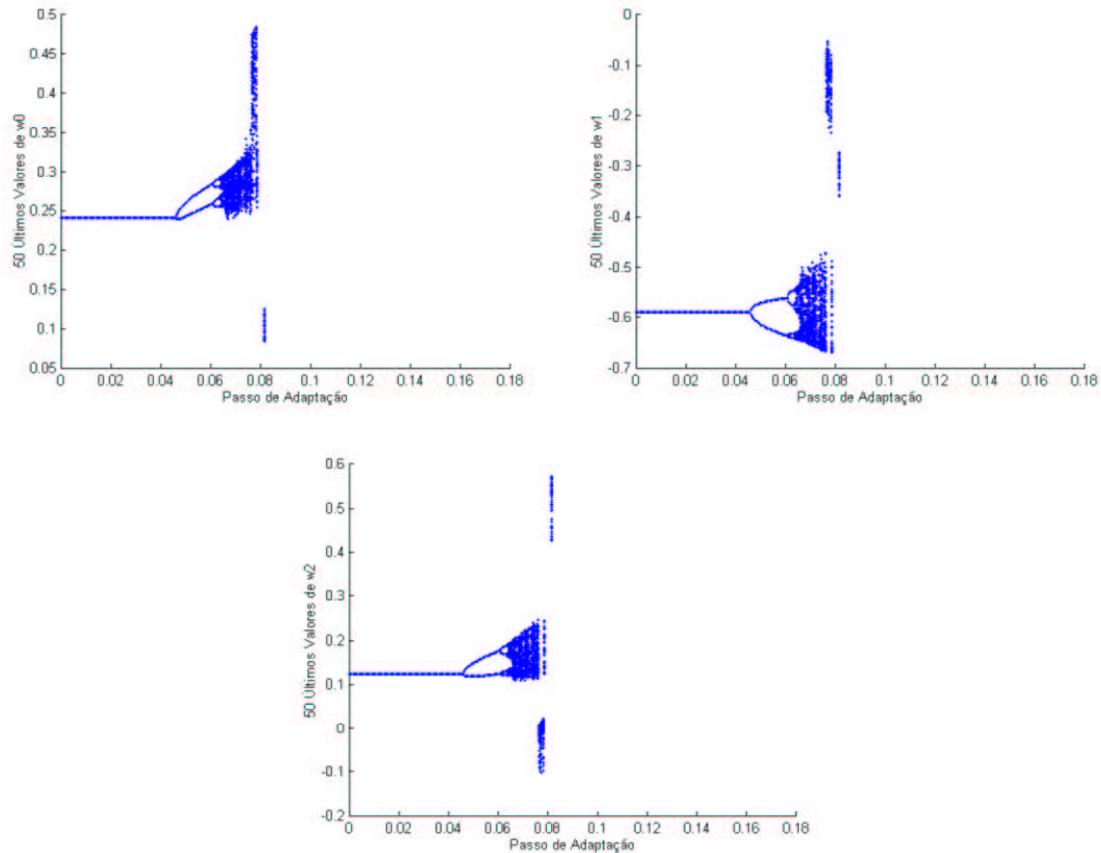
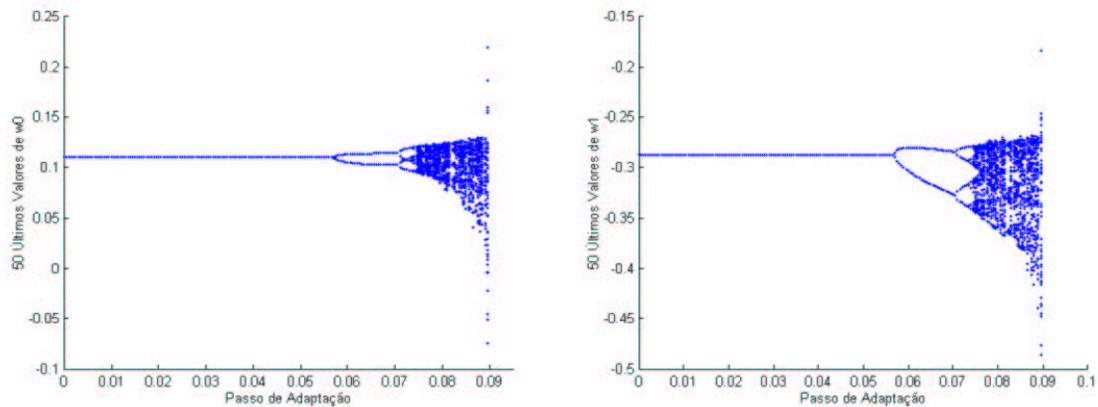


Figura 6.28: Diagramas de Bifurcação ($\mathbf{w}(0) = [0.241 \ -0.589 \ 0.124]^T$)

Podemos entender um pouco mais o comportamento do sistema analisando o segundo “melhor” mínimo, localizado em $\mathbf{w}_{e2} = [0.110 \ -0.287 \ 0.524]^T$. A faixa de estabilidade associada a este ponto foi $0 < \mu < 0.056$.

Na Fig. 6.29, estão os diagramas de bifurcação associados a este mínimo.



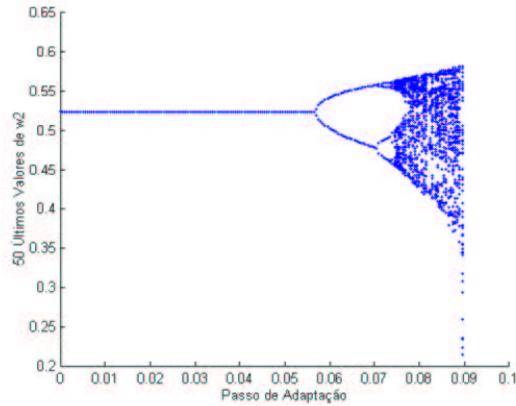
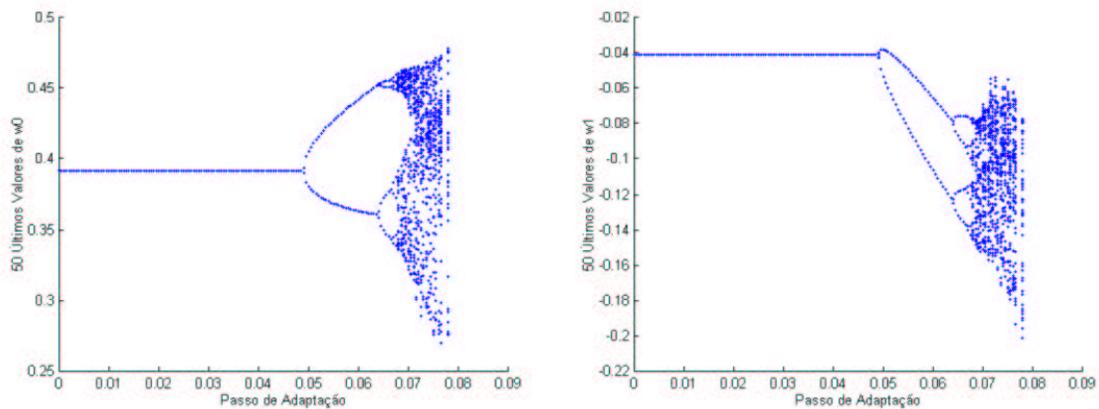


Figura 6.29: Diagramas de Bifurcação ($\mathbf{w}(0) = [0.110 \ -0.287 \ 0.524]^T$)

Podemos perceber que este mínimo tem um comportamento mais “regular” na região próxima à zona de divergência. Há apenas um espalhamento de banda que ocorre em um intervalo bastante pequeno, verificável na Fig. 6.29. Vale ressaltar que este tipo de comportamento, manifestado para uma faixa muito estreita de μ , poderia não ser observável no diagrama, pelo fato deste ser produzido através de uma discretização em μ .

Por fim, vejamos um mínimo com maior custo, localizado em $\mathbf{w}_{e3} = [0.392 \ -0.041 \ -0.028]^T$. Os diagramas de bifurcação estão na Fig. 6.30. A faixa de estabilidade associada foi $0 < \mu < 0.049$.



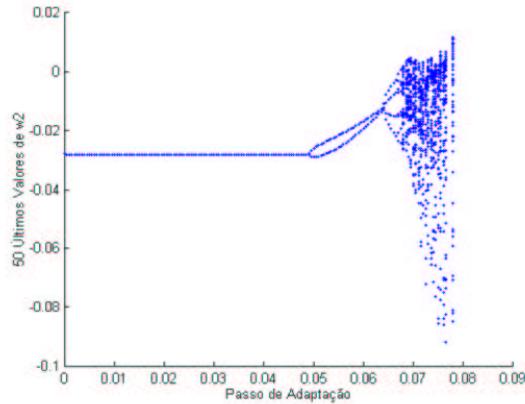


Figura 6.30: Diagramas de Bifurcação ($\mathbf{w}(0) = [0.392 -0.041 -0.028]^T$)

A evolução deste mínimo também se deu segundo o cenário de Feigenbaum, sendo que na região próxima à zona de divergência não foi observada convergência para outros atratores estranhos. O diagrama de w_2 tem um aspecto curioso, um “cruzamento” na segunda duplicação de período.

É interessante ver que o melhor mínimo teve o diagrama associado mais descontínuo na zona de caos e a menor faixa de estabilidade, o que contradiz qualquer intuição de que a “sua dinâmica” deva apresentar sempre regiões estáveis mais amplas.

6.2.5. Canal AR(1)

Até agora, vínhamos analisando modelos de canal com resposta ao impulso finita. De fato, estes modelos são os mais aplicados em sistemas reais de comunicação. Porém, no estudo da convergência do CMA um outro modelo de canal teve uma grande importância histórica: o modelo AR(1). Como vimos na seção 5.5, este foi o modelo utilizado por Ding et al. no trabalho [Ding et al., 1991] onde foi demonstrado que um equalizador linear finito ajustado pelo CMA pode ter seu estado atraído para mínimos locais. O grande atrativo teórico deste modelo de canal é a possibilidade de ser perfeitamente equalizado pela estrutura que enfocamos, o filtro linear e transversal.

A função de transferência deste canal é expressa por:

$$H_4(z) = \frac{1}{1 + \alpha \cdot z^{-1}} \quad (6.22)$$

onde α deve estar necessariamente dentro do CRU, para que o canal seja estável. Utilizamos o mesmo valor de [Ding et al., 1991], ou seja, $\alpha = 0.6$. Como de costume, analisemos as curvas de nível da superfície de custo, que se encontram na Fig. 6.31.

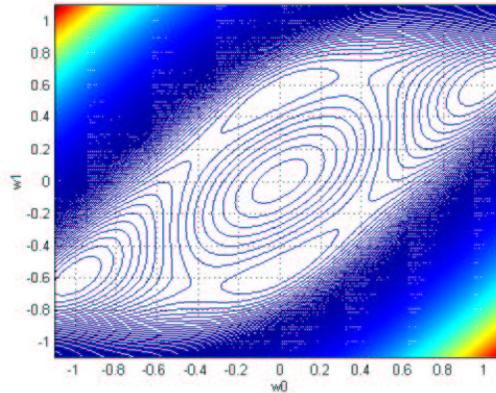


Figura 6.31: Curvas de Nível da Função Custo

Novamente, temos o mesmo número de pontos críticos. Os dois mínimos por nós escolhidos são:

$$\mathbf{P} = \{[1 \ 0.6]^T, [0 \ 0.5575]^T\} \quad (6.23)$$

sendo o primeiro correspondente à situação de perfeita inversão, e o segundo um mínimo local. Como vimos na seção 5.5, são mínimos de profundidades muito distintas, pois o primeiro corresponde à perfeita recuperação do módulo constante e o segundo introduz apenas um atraso e uma atenuação, não retirando IIS.

Utilizando novamente a análise da matriz de Floquet, expressamos na Fig. 6.32 a relação entre o multiplicador de maior magnitude e o passo de adaptação.

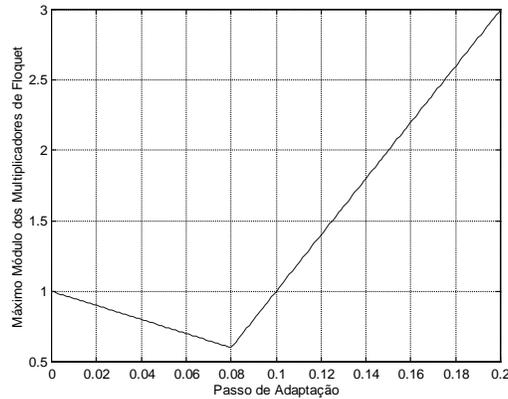


Figura 6.32: Multiplicador de Floquet de Maior Magnitude

Por meio desta análise, a faixa de estabilidade obtida para o mínimo global foi $0 < \mu < 0.100$. Fazendo uma inicialização próxima ao mínimo global, obtivemos os diagramas de bifurcação para as duas variáveis de estado, que estão na Fig. 6.33.

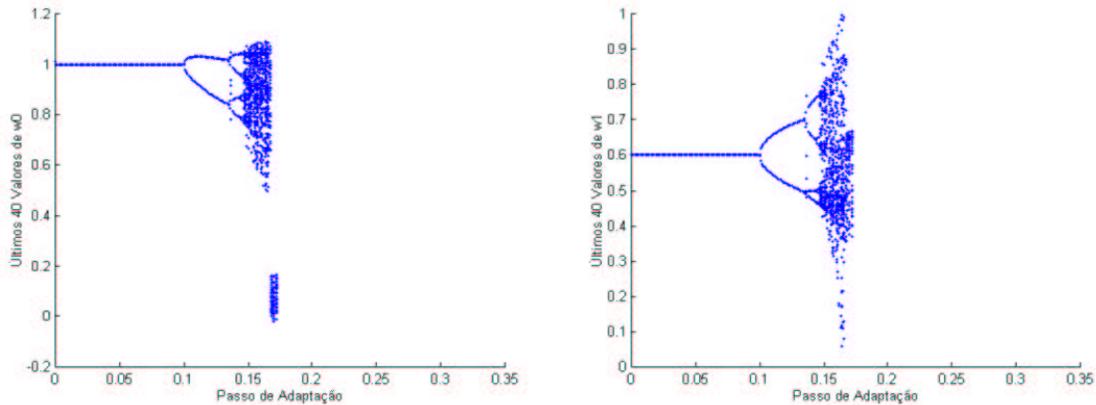


Figura 6.33: Diagramas de Bifurcação ($\mathbf{w}(0) = [1.01 \ 0.6]^T$)

De imediato, verificamos a adequação da faixa de estabilidade do mínimo global. Quando este perde a sua estabilidade, tem lugar o usual cenário de duplicação de período seguida de caos. Porém, para μ próximo de 0.13, há uma estreita banda de oscilação que surge na faixa correspondente ao ciclo-limite de período quatro. Uma outra peculiaridade merece destaque: a última região antes da divergência para o infinito não corresponde ao atrator estranho que vinha continuamente evoluindo, mas é uma banda caótica que emerge abruptamente, possivelmente associada ao mínimo local.

Calculamos o máximo expoente de Lyapunov para os diversos valores de μ , sendo que o resultado está na Fig. 6.34.

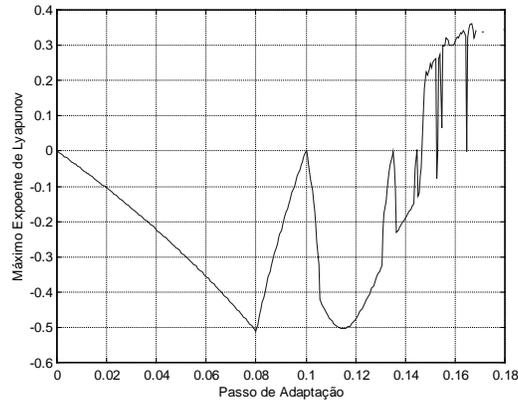


Figura 6.34: Máximo Expoente de Lyapunov

O cálculo do expoente de Lyapunov confirmou o panorama apontado pelo diagrama de bifurcação. Vejamos agora, na Fig. 6.35, como é o atrator estranho associado ao mínimo global, para dois valores distintos de μ .

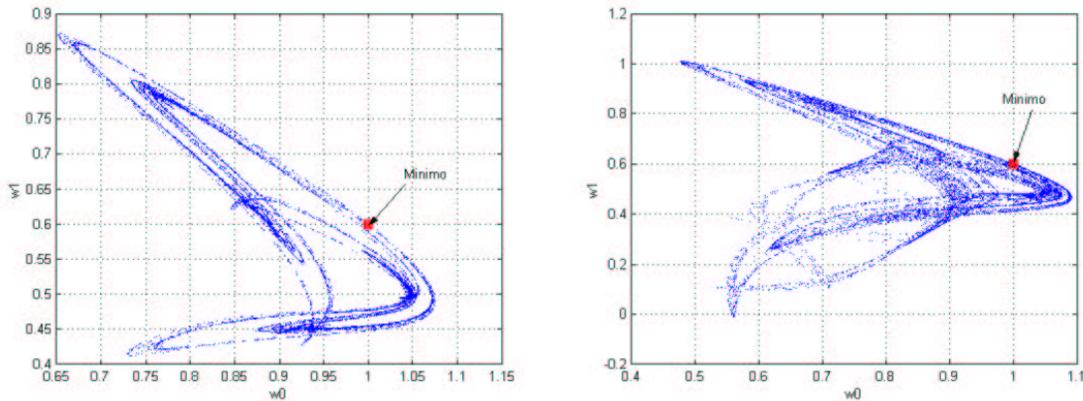


Figura 6.35: Atrator Estranho Associado ao Mínimo Global para $\mu = 0.15$ e $\mu = 0.165$

O atrator estranho associado ao sistema apresenta um aspecto bastante distinto dos vistos anteriormente para modelos de canal MA, o que era de se esperar, tendo em vista que a mudança substancial no modelo do canal leva a mudanças nas estatísticas do sinal recebido, com influências na distribuição dos mínimos da função custo.

A evolução temporal de $w_0(n)$ encontra-se na Fig. 6.36.

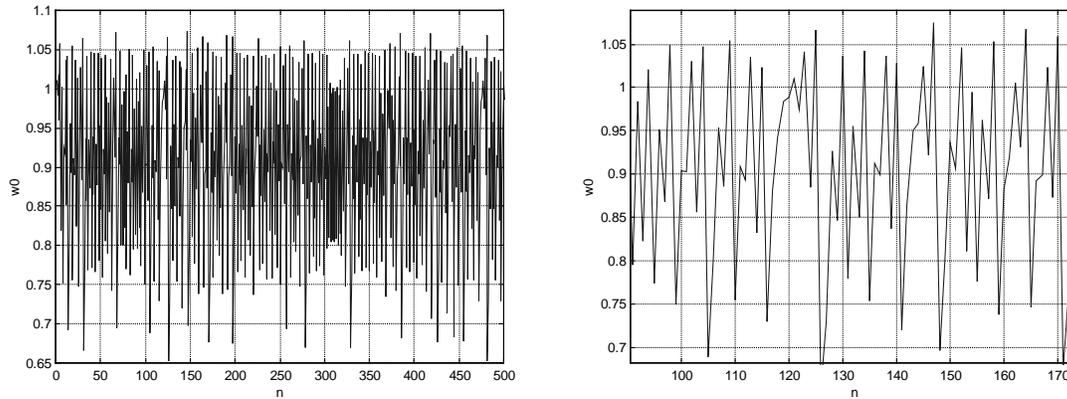


Figura 6.36: Evolução Temporal de $w_0(n)$ para $\mu = 0.15$

Passemos à análise do mínimo local. Em primeiro lugar, vejamos na Fig. 6.37 a relação entre o multiplicador de Floquet de maior magnitude e o passo de adaptação.

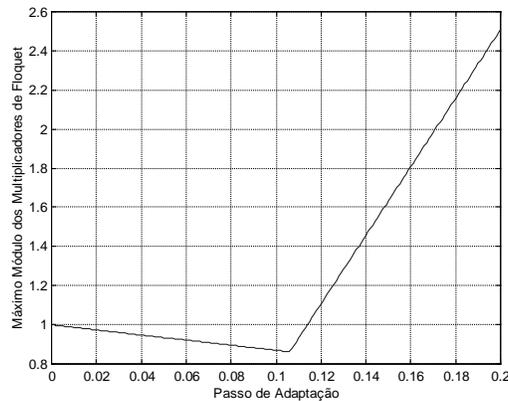


Figura 6.37: Multiplicador de Floquet de Maior Magnitude

A faixa de estabilidade neste caso foi $0 < \mu < 0.114$. Vale observar que a faixa é ligeiramente maior que a obtida para o mínimo global, ao contrário do que ocorreu para os canais MA(1). Os diagramas obtidos a partir de uma condição inicial próxima ao mínimo local estão na Fig. 6.38.

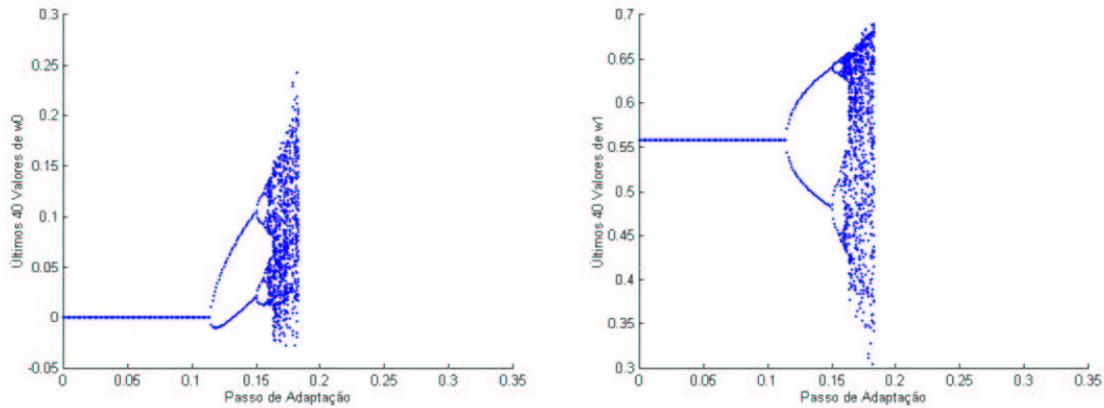


Figura 6.38: Diagramas de Bifurcação ($\mathbf{w}(0) = [0 \ 0.5575]^T$)

Os diagramas apresentam o mesmo tipo de evolução dinâmica dos casos anteriores, mas sem nenhum tipo de irregularidade ou descontinuidade. Deve ser ressaltado que, neste caso, houve uma inversão em relação ao modelo MA(1): a faixa de estabilidade foi maior para o mínimo local e o atrator estranho correspondente teve uma maior região de estabilidade.

A evolução do máximo expoente de Lyapunov em função do passo de adaptação está apresentada na Fig. 6.39.

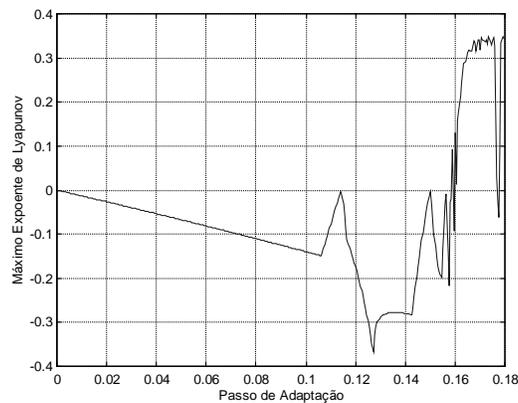


Figura 6.39: Máximo Expoente de Lyapunov

Os valores do expoente novamente mostram-se em perfeita concordância com a análise realizada através do diagrama de bifurcação. Obtivemos ainda o atrator estranho associado ao mínimo local para $\mu = 0.17$ e $\mu = 0.175$, sendo ambos mostrados na Fig. 6.40.

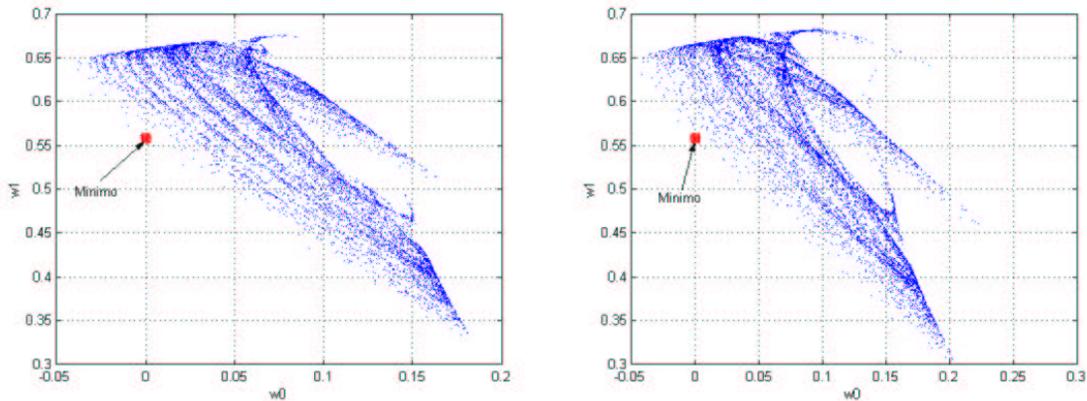


Figura 6.40: Atrator Estranho Associado ao Mínimo Local para $\mu = 0.17$ e $\mu = 0.175$

Como ocorreu em todos os casos analisados, atratores estranhos associados a mínimos distintos apresentam diferenças em sua estrutura. Como nos demais casos, o mínimo encontra-se na região do atrator.

Finalmente, a evolução temporal de $w_0(n)$ para $\mu = 0.17$ está na Fig. 6.41, mantendo o aspecto geral dos casos anteriores.

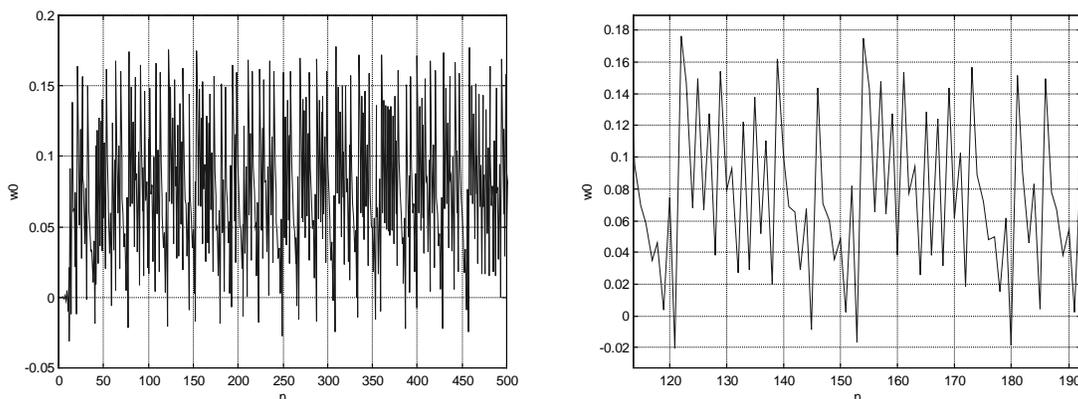


Figura 6.41: Evolução Temporal de $w_0(n)$

6.2.6. Considerações Sobre o Caso Determinístico

Antes de passarmos à análise do algoritmo estocástico, convém esboçar algumas conclusões. Primeiramente, verificou-se em todos os cenários testados a ocorrência de comportamento complexo, a saber: oscilação em ciclos-limite de infinitos períodos e caos, sendo portanto confirmadas as suposições que motivaram nosso trabalho. Além da inspeção dos diagramas de bifurcação tivemos a confirmação matemática fornecida pelos expoentes de Lyapunov nos modelos de primeira ordem, havendo sempre perfeita concordância entre as técnicas. Esta concordância, aliada à questão da simplicidade, nos levou a omitir o cálculo deste expoente para o modelo MA(2).

Constatamos que a evolução dinâmica dos mínimos apresentou algumas diferenças significativas ligadas ao modelo de canal adotado. Nos canais MA(1), os mínimos globais apresentaram uma faixa de estabilidade maior que a dos mínimos locais, enquanto que nos canais AR(1) e MA(2) ocorreu o contrário. Um outro aspecto relevante é que nos canais MA(1) verificamos que, para inicialização próxima aos pontos de mínimos, a última zona não-divergente estava associada ao mínimo global. Tal fato não ocorreu para os outros dois tipos de canal, o que invalida uma suposição de que tal comportamento seja uma regra geral ou mesmo um cenário inerente ao modelo moving-average genérico.

Os atratores estranhos associados aos diferentes mínimos também apresentaram diferenças, não só em posição espacial, mas em suas estruturas. Todas as evoluções temporais obtidas na zona de caos nos permitiram ter uma idéia do comportamento catastrófico, neste regime, no que se refere ao desempenho do equalizador. Certamente é uma evolução temporal prejudicial ao desempenho do sistema. Mesmo quando o estado encontra-se num ciclo-limite estável pode haver uma grande perda de desempenho, em termos de abertura do diagrama de olho.

O objetivo de toda a nossa análise foi abordar um comportamento do sistema que não se enquadra nas situações clássicas de convergência para um ponto fixo ou de divergência para o infinito, e que contém em si uma estrutura extremamente rica e com implicações, que pouco, ou nada, foram exploradas. Uma vez estudada a versão determinística do algoritmo, que está na base do CMA estocástico, convém passarmos à

análise direta deste último, para que possamos verificar também nele a possibilidade de ocorrência de comportamento complexo.

6.3. *Algoritmo Estocástico: Princípios de Análise*

Quando lidamos com o algoritmo do módulo constante em sua versão prática, precisamos estudar um sistema dinâmico estocástico. Analisar tais sistemas é certamente tarefa muito mais difícil do que estudar um sistema determinístico, como vínhamos fazendo até agora.

Nossa análise, para este caso bem mais complexo, terá por único objetivo demonstrar a possibilidade concreta de ocorrência de comportamento caótico no CMA estocástico. Mais especificamente, analisaremos tal algoritmo quando aplicado a um modelo de canal FIR e procuraremos mostrar que o estado do sistema pode variar caoticamente em determinados casos.

O motivo desta escolha está no cerne do método que iremos usar. Em linhas gerais, este método consiste em desmembrar um sistema dinâmico estocástico em um conjunto de subsistemas determinísticos que se revezam ao longo das iterações que modificam o estado. Avaliando separadamente estes sistemas determinísticos, podemos tirar algumas conclusões sobre o sistema estocástico global. Vejamos então como surge esta divisão.

Suponhamos que o sinal transmitido seja binário, como temos feito até agora. Tomemos o modelo de canal/equalizador MA(1) expresso em (6.15). Sabemos que o CMA possui, neste caso, a seguinte expressão:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \cdot y(n) \cdot [R_2 - y^2(n)] \quad (6.24)$$

onde

$$y(n) = w_0(n) \cdot x(n) + w_1(n) \cdot x(n-1) \quad (6.25)$$

Lembrando que

$$x(n) = s(n) + \alpha \cdot s(n-1) \quad (6.26)$$

e

$$x(n-1) = s(n-1) + \alpha \cdot s(n-2) \quad (6.27)$$

percebemos que, caso α não seja igual a um, o vetor $\mathbf{x}(n)$ assumirá tantos valores quanto forem as combinações de $s(n)$, $s(n-1)$ e $s(n-2)$. De modo geral, podemos afirmar que o número destas combinações é a quantidade máxima de estados assumidos pelo vetor $\mathbf{x}(n)$.

Para que possamos perceber o que ocorre, construímos a Tab. 6.1, na qual supomos, sem perda de generalidade, que $s(n)$ assume os valores $+1$ e -1 .

Configuração	$s(n)$	$s(n-1)$	$s(n-2)$	$x(n)$	$x(n-1)$
1	+1	+1	+1	$1+\alpha$	$1+\alpha$
2	+1	+1	-1	$1+\alpha$	$1-\alpha$
3	+1	-1	+1	$1-\alpha$	$-1+\alpha$
4	+1	-1	-1	$1-\alpha$	$-1-\alpha$
5	-1	+1	+1	$-1+\alpha$	$1+\alpha$
6	-1	+1	-1	$-1+\alpha$	$1-\alpha$
7	-1	-1	+1	$-1-\alpha$	$-1+\alpha$
8	-1	-1	-1	$-1-\alpha$	$-1-\alpha$

Tabela 6.1: Possíveis Configurações de $\mathbf{x}(n)$

A Tab. 6.1 nos dá oito possibilidades para o vetor de entrada $\mathbf{x}(n)$. Porém, como podemos observar, para uma dada configuração $\mathbf{x}(n)$ há também a simétrica $-\mathbf{x}(n)$. A expressão (6.24), porém, será a mesma para os dois casos. Isto mostra que as oito configurações, quando aplicadas a (6.24), dão origem não a oito, mas a quatro subsistemas dinâmicos determinísticos distintos. Podemos associar estes quatro sistemas, por exemplo, às quatro primeiras configurações da Tab. 6.1.

A aleatoriedade inerente a (6.24), para este modelo de canal, é essencialmente discreta. Há um certo número de possibilidades de subsistemas dinâmicos, sendo que cada um deles assume temporariamente o papel de atualizar o vetor de pesos. Os sinais transmitidos $s(n)$ acabam por determinar qual subsistema é o escolhido a cada iteração.

Um ponto interessante é que isto mostra que a evolução de $\mathbf{w}(n)$ se dá sempre através de um conjunto discreto de direções no espaço de fase. Trata-se de algo muito interessante, pois nos faz pensar, de maneira um tanto quanto poética, que a convergência do estado para um ponto de equilíbrio se faz através da compensação da finitude do repertório de direções de adaptação por uma engenhosa “escolha” do módulo do gradiente.

Outro ponto muito importante é mostrar que se estes subsistemas determinísticos apresentarem comportamento complexo, o sistema estocástico global também pode manifestar tais tipos de comportamento.

O nosso método de análise irá consistir de uma análise determinística das componentes de um sistema dinâmico estocástico, o que permitirá um tipo de estudo não muito distinto do que temos realizado. Certamente esta análise tem um alcance limitado, mas se presta bem ao objetivo central deste trabalho.

Em termos gerais, podemos afirmar que, caso $s(n)$ pertença a um alfabeto M -ário e o número de pesos do equalizador e do canal seja N , haverá um número máximo de possibilidades para $x(n)$ igual a $N_{x\max}$, dado por:

$$N_{x\max} = M^{2N-1} \quad (6.28)$$

No caso de $M = 2$ (com valores $+A$ e $-A$), temos que para cada valor do vetor de entrada do equalizador estará associado o seu simétrico. Isto quer dizer que, neste caso, teremos um número máximo de sistemas igual a:

$$N_{\text{sis max}} = \frac{2^{2N-1}}{2} = 2^{2N-2} = 4^{N-1} \quad (6.29)$$

Embora tenhamos mostrado, através da Tab. 6.1, que a iteração no algoritmo do módulo constante só se dá por meio de quatro sistemas determinísticos que se alternam

nesta função, é preciso discutir com mais profundidade esta idéia. Em primeiro lugar, vejamos como são as transições entre estados.

Suponhamos que o equalizador se encontre na configuração 1 da Tab. 6.1. Se no próximo instante n o sinal transmitido for -1 , o equalizador passará à configuração 5. Se, porém, o sinal transmitido for $+1$, o equalizador permanecerá na configuração 1. Para que possamos esgotar as possibilidades desta análise, construamos a Tab. 6.2.

Configuração atual	Se for transmitido $+1$, passa à configuração	Se for transmitido -1 , passa à configuração
1	1	5
2	1	5
3	2	6
4	2	6
5	3	7
6	3	7
7	4	8
8	4	8

Tabela 6.2: Possíveis Transições entre Configurações

A Tab. 6.2 é a expressão típica de uma cadeia de Markov. Tais cadeias são compostas por diversos estados e suas probabilidades de transição. Uma cadeia de Markov caracteriza-se pelo fato da probabilidade de transição entre estados quaisquer i e j depender exclusivamente destes estados [Palazzo, 1999]. Ora, este é justamente o caso da Tab. 6.2, como podemos averiguar.

Portanto, as configurações do vetor de entrada correspondem a estados de uma cadeia de Markov, cuja transição é determinada pelo sinal $s(n)$. A dependência com relação a $s(n)$ está embutida nas probabilidades de transição entre estados, como também é perceptível pela Tab. 6.2. Em nosso caso, a distribuição da fonte que produz o sinal $s(n)$ é quem determina tais probabilidades.

A cada uma das configurações está associado um sistema dinâmico determinístico, sendo que a cada par simétrico estará associado o mesmo sistema, responsável pela iteração no instante n . Portanto, os saltos na cadeia são responsáveis pela escolha do “sistema da vez”. Sistemas deste tipo têm sido estudados no âmbito de controle, com resultados interessantes [Gonçalves et al., 2000], como vimos na seção 2.7. No nosso caso, iremos utilizar esta abordagem de *saltos markovianos* simplesmente para desmembrar o sistema global em seus elementos determinísticos. A aleatoriedade é retirada do sistema em si e passa a estar contida nas transições entre os estados. Isto permite que possamos aplicar ferramentas usuais de análise dinâmica a cada um dos blocos isolados e tirar assim conclusões sobre *possibilidades* de comportamento, condicionadas a determinadas trajetórias dentro da cadeia.

Portanto, se pudermos demonstrar que um determinado subsistema apresenta um atrator estranho para determinada configuração, estará demonstrado que o sistema global pode apresentar comportamento caótico, desde que haja um número de iterações suficiente para que o estado atinja a vizinhança do atrator. Portanto, a análise que realizaremos se limitará a mostrar “possibilidades de comportamento”, condicionadas à ocorrência de determinadas sequências de $s(n)$.

6.4. Ocorrência de Caos no Algoritmo Estocástico

Utilizaremos, para ilustrar este procedimento e obter algumas conclusões sobre o comportamento do CMA, o modelo de canal (6.15). Optamos pelo valor $\alpha = 0.6$, por já o termos focado no caso determinístico. Supomos também que o módulo do sinal bipolar transmitido vale 1.

Aplicando (6.29), percebemos que as configurações dão origem a quatro subsistemas. Utilizando o índice k para rotular estes sistemas, podemos escrever:

$$\mathbf{w}_k(n+1) = \mathbf{w}_k(n) + \mu \cdot [\mathbf{w}_k^T(n) \cdot \mathbf{x}_k] \cdot \{1 - [\mathbf{w}_k^T(n) \cdot \mathbf{x}_k]^2\} \cdot \mathbf{x}_k \quad (6.30)$$

O vetor $\mathbf{x}_k(n)$ corresponde à k -ésima configuração da Tab. 6.1. Analisaremos cada um destes sistemas em separado, passando então a uma conclusão conjunta.

Primeiro Sistema:

Fazendo $k = 1$, temos:

$$\mathbf{x}_1 = [1.6 \ 1.6]^T \quad (6.31)$$

Há três condições de equilíbrio para este sistema:

$$\mathbf{w}_{e1} = [0 \ 0]^T \quad (6.32)$$

$$\mathbf{w}_{e2}^T \mathbf{x}_1 = 1 \quad (6.33)$$

$$\mathbf{w}_{e3}^T \mathbf{x}_1 = -1 \quad (6.34)$$

Inicializando o algoritmo na vizinhança do mínimo global, temos os diagramas de bifurcação da Fig. 6.42.

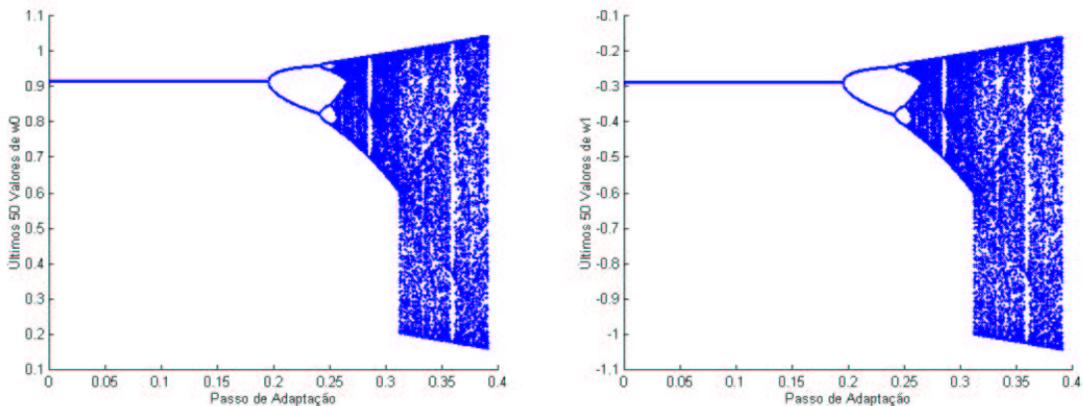


Figura 6.42: Diagramas de Bifurcação ($\mathbf{w}(0) = [0.831, -0.372]^T$)

Para este sistema ocorre o cenário usual em nossas análises: o de Feigenbaum. O ponto de equilíbrio para o qual o sistema converge foi $\mathbf{w}_e = [0.914 \ -0.289]^T$. Tal ponto satisfaz a condição (6.33). A faixa de estabilidade para este ponto foi $0 < \mu < 0.195$.

Podemos observar ainda uma mudança na banda de oscilação do atrator estranho, para μ próximo a 0.313, como na Fig. 6.1. Há uma identidade entre os dois diagramas da Fig. 6.42, o que não é de causar espanto, pois de (6.30) percebemos que a variação $\Delta \mathbf{w}(n) = \mathbf{w}(n+1) - \mathbf{w}(n)$ é a mesma para w_0 e w_1 .

Segundo Sistema:

Fazendo $k = 2$, temos:

$$\mathbf{x}_2 = [1.6 \ 0.4]^T \quad (6.35)$$

Há três condições de equilíbrio para o sistema:

$$\mathbf{w}_{e1} = [0 \ 0]^T \quad (6.36)$$

$$\mathbf{w}_{e2}^T \mathbf{x}_2 = 1 \quad (6.37)$$

$$\mathbf{w}_{e3}^T \mathbf{x}_2 = -1 \quad (6.38)$$

Utilizando a mesma inicialização do caso anterior, obtivemos os diagramas de bifurcação da Fig. 6.43.

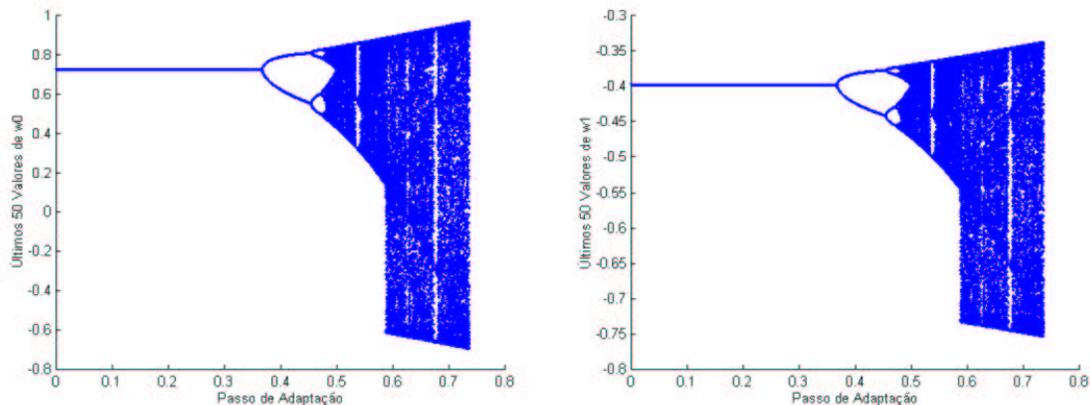


Figura 6.43: Diagramas de Bifurcação ($\mathbf{w}(0) = [0.831, -0.372]^T$)

Verificamos o mesmo tipo de cenário dinâmico, sendo que o ponto de equilíbrio para o qual o sistema converge, para baixos valores de μ , é $[0.725 \ -0.399]^T$. A faixa de estabilidade para este ponto foi $0 < \mu < 0.367$. Tal ponto corresponde à condição (6.37), por questões de inicialização. Novamente, observamos o alargamento da banda caótica.

Terceiro Sistema:

Fazendo $k = 3$, temos:

$$\mathbf{x}_3 = [0.4 \ -0.4]^T \quad (6.39)$$

Há três condições de equilíbrio para o sistema:

$$\mathbf{w}_{e1} = [0 \ 0]^T \quad (6.40)$$

$$\mathbf{w}_{e2}^T \mathbf{x}_3 = 1 \quad (6.41)$$

$$\mathbf{w}_{e3}^T \mathbf{x}_3 = -1 \quad (6.42)$$

Utilizando a mesma inicialização do caso anterior, obtivemos os diagramas de bifurcação apresentados na Fig. 6.44. Tais diagramas têm também a mesma forma geral dos obtidos nos casos anteriores. O ponto de equilíbrio para o qual o sistema convergiu, para $0 < \mu < 3.11$ foi $[1.479 \ -1.021]^T$, correspondente à condição (6.41).

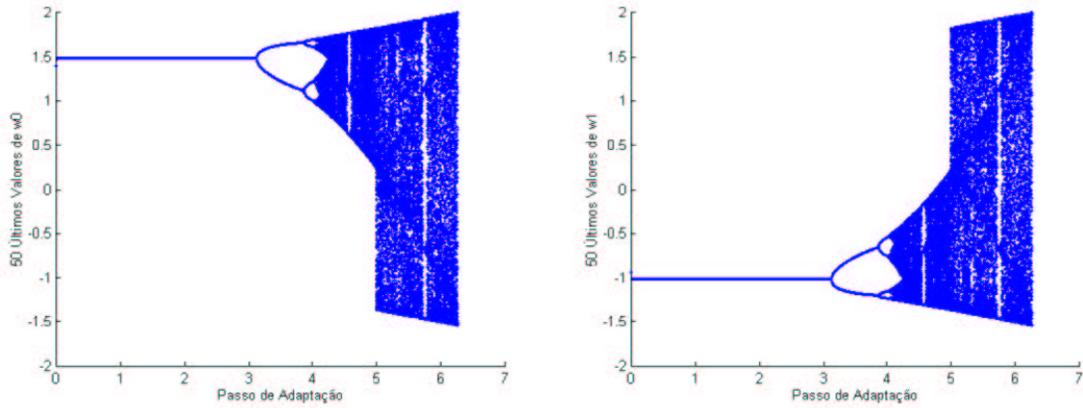


Figura 6.44: Diagramas de Bifurcação ($\mathbf{w}(0) = [0.831, -0.372]^T$)

Quarto Sistema:

Fazendo $k = 4$, temos:

$$\mathbf{x}_4 = [0.4 \ -1.6]^T \quad (6.43)$$

Há três condições de equilíbrio para o sistema:

$$\mathbf{w}_{e1} = [0 \ 0]^T \quad (6.44)$$

$$\mathbf{w}_{e2}^T \mathbf{x}_4 = 1 \quad (6.45)$$

$$\mathbf{w}_{e3}^T \mathbf{x}_4 = -1 \quad (6.46)$$

Utilizando a mesma inicialização do caso anterior, chegamos aos diagramas de bifurcação da Fig. 6.45.

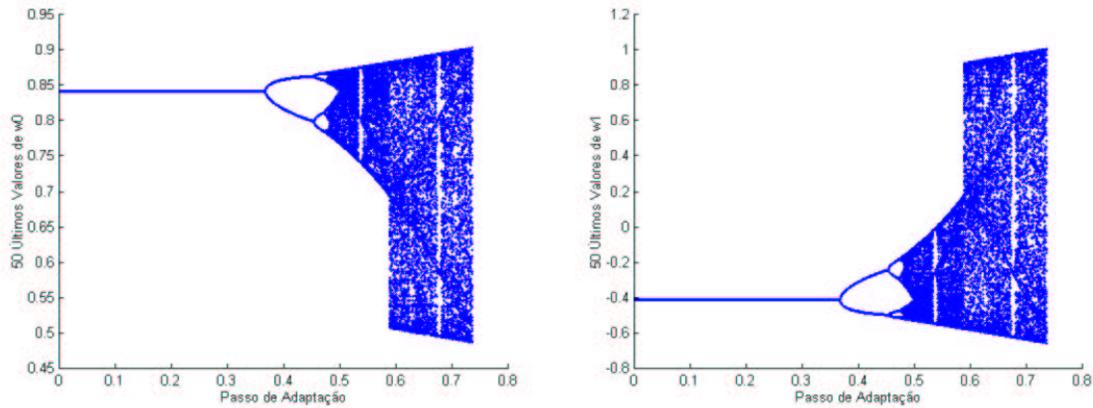


Figura 6.45: Diagramas de Bifurcação ($\mathbf{w}(0) = [0.831, -0.372]^T$)

Novamente, os diagramas apontam um tipo de cenário idêntico ao de todos os casos anteriores. O ponto de equilíbrio, para o qual convergiu o estado para a faixa $0 < \mu < 0.366$ foi $[0.842 \ -0.415]^T$, o que corresponde à condição (6.45).

Uma vez analisados os sistemas em separado, passemos a uma análise que aproxima uma certa situação real. Suponhamos, a título de exemplo, que entre os valores do sinal transmitido haja uma sequência de símbolos +1. Isto fará com que o algoritmo estocástico se reduza, por um dado período, ao primeiro subsistema que analisamos, o que permite uma análise determinística do sistema global. Novamente, utilizamos o canal MA(1) de fase mínima (6.15) com $\alpha = 0.6$. O estado do sistema foi inicializado bastante próximo ao mínimo global, conforme o procedimento padrão deste capítulo, e o valor do passo foi $\mu = 0.35$, cuja adequação é comprovada pela Fig. 6.42. Partindo do primeiro símbolo transmitido, a sequência de +1 teve 25 elementos, uma quantidade propositalmente exagerada para ressaltar as características que buscamos analisar. Do símbolo 26 até o 70, os símbolos +1 e -1 foram produzidos com igual probabilidade.

Na Fig. 6.46, temos a evolução temporal de $w_0(n)$. Lembramos que a adaptação dos pesos só começa em $n = 3$, quando os acumuladores do canal e do equalizador estão totalmente preenchidos, e que a partir de $n = 27$ o estado não mais evolui segundo o primeiro subsistema.

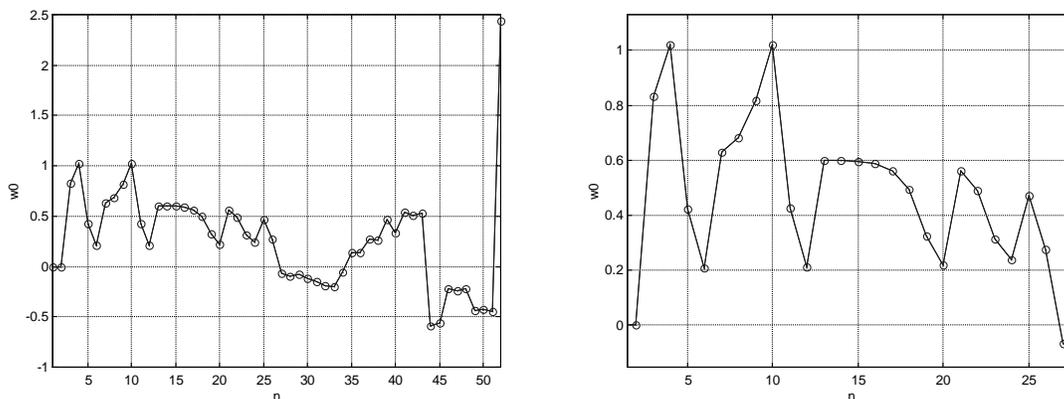


Figura 6.46: Evolução Temporal de $w_0(n)$ e Ampliação da Região de Caos (de $n = 3$ a 26)

Podemos perceber que até $n = 26$, evolução temporal apresenta um padrão comum aos sistemas caóticos que temos analisado. Através da Fig. 6.42 temos a confirmação do tipo de comportamento do sistema, bem como podemos verificar a largura da banda de oscilação caótica para o correspondente valor de passo. Esta quantidade de amostras permite o desaparecimento dos transitórios, clarificando a análise. Para $n > 26$, no entanto, a aparência da série muda, pois não mais corresponde a um dado sistema dinâmico mas à alternância de vários. Para n pouco maior que 50 começa a explosão para o infinito.

Mesmo para comprimentos menores de sequência pode haver comportamento caótico, pois o atrator estranho será estável, ficando a evolução temporal do estado intimamente ligada ao número de iterações necessário para a sua convergência.

6.5. Considerações Sobre o Caso Estocástico

Conforme explicitamos no início desta seção, a análise estocástica por nós realizada tem um escopo restrito e bem definido: mostrar a possibilidade de ocorrência de caos no processo iterativo responsável pela adaptação cega de um filtro FIR. A utilização da idéia dos saltos markovianos mostrou-se adequada, permitindo uma linha de análise simples e totalmente embasada pelo ferramental usual da teoria de sistemas dinâmicos determinísticos.

Observamos caos em todos os subsistemas, que apresentam em sua evolução dinâmica um cenário de Feigenbaum típico, com os possíveis comportamentos distribuídos ao longo de diferentes faixas de valores de passo.

Uma vez comprovada a possibilidade de caos no algoritmo estocástico, passamos à implementação de um exemplo que pôde mostrar com mais detalhe como o sistema dinâmico global pode efetivamente apresentar comportamento caótico. Este exemplo envolveu uma sequência de bits transmitidos que levou ao “monopólio” do primeiro subsistema nas primeiras iterações, sendo que nas iterações posteriores os dados transmitidos voltaram a ser gerados aleatoriamente. Poderiam ser ainda procurados cenários mais complexos, envolvendo a alternância de diferentes subsistemas, o que é bastante razoável se tomarmos por base os diagramas de bifurcação das Figs. 6.42 a 6.45. Embora tenhamos restringido nossa análise ao modelo MA(1), acreditamos não haver nenhuma restrição quanto à ordem do modelo para que ocorra caos, tanto no sistema determinístico quanto no estocástico, pelas características do algoritmo, que naturalmente conduzem ao cenário de Feigenbaum.

7. Uma Proposta para Aumentar a Taxa de Convergência Global do CMA

Os resultados obtidos no capítulo anterior, referentes à análise dinâmica do algoritmo do módulo constante, levaram-nos a identificar uma proposta visando o aumento de sua convergência global. Não é a nossa idéia realizar uma análise extensiva do método proposto, por não ser este o tema central da tese, mas iremos expor alguns resultados que podem ilustrar o seu potencial em alguns casos.

Veremos, em primeiro lugar, as idéias que deram origem ao método, para em seguida apresentar sua implementação e os resultados obtidos em diferentes modelos de canais.

7.1. *Base Teórica*

No que se refere aos resultados do capítulo 6, um fato nos chama de imediato a atenção: a disparidade das evoluções dos mínimos. Quando lidamos com os canais MA(1), por exemplo, notamos que o atrator estranho associado ao mínimo local colapsava para um valor de μ menor que o do atrator associado ao mínimo global. O que isto quer dizer?

Suponhamos que a inicialização seja feita próxima a um certo mínimo w_{e1} . Como vimos, para um certo valor de passo este mínimo perderá sua estabilidade, tendo lugar uma cascata de bifurcações seguida de caos. Tanto nas oscilações periódicas como nas caóticas, constatamos nos casos por nós estudados que os atratores permanecem mais próximos do mínimo a que estão associados. Ousamos estender esta idéia e afirmar a partir do que vimos que, tanto em regime periódico quanto caótico, em um certo instante n a maior probabilidade é que o estado esteja na bacia de atração do mínimo que deu origem ao regime em que o sistema se encontra. Portanto, se o sistema oscila num ciclo-limite ou num atrator estranho associado ao mínimo global, é mais provável que num certo momento o estado esteja na bacia de atração deste mínimo, e não do local. Não demonstramos isto, e também não analisamos este fato com grande rigor, mas os resultados que obtivemos

apontam nesta direção. No capítulo 6, por exemplo, sempre marcamos junto ao atrator estranho o mínimo correspondente, o que permite constatar o fato que acabamos de expor.

Uma vez posto isto, voltemos ao cenário descrito no início desta seção. Enfoquemos o caso do canal MA(1), descrito nas seções 6.2.2 e 6.2.3. Quando a inicialização era feita próxima ao mínimo local w_L , ele perdia sua estabilidade, dando início às bifurcações, seguidas pela zona de caos. Quando o atrator estranho associado a w_L perdia sua estabilidade, o sistema voltava a oscilar num ciclo-limite, mas desta vez associado ao mínimo global w_G . Com o sistema oscilando em tal ciclo, como vimos, a maior probabilidade é que o estado esteja na bacia de atração do mínimo global correspondente. O mesmo vale para o atrator estranho que surge para um valor ainda maior de μ , sendo seguido da zona de divergência. Isto quer dizer o seguinte: mesmo com uma inicialização muito próxima ao mínimo local, um passo de adaptação escolhido adequadamente pode levar o estado a oscilar nas proximidades do mínimo global.

Como fazer, no entanto, o estado convergir para uma região ainda mais próxima do mínimo global? A resposta para isto é derivada do conceito de *annealing*, sendo necessário um “resfriamento” do passo de adaptação, ou seja, a sua redução a um valor no qual o mínimo global volte a ser estável e atraia o estado em sua vizinhança. Esta idéia nos foi inspirada pelos trabalhos [Zhou e Chen, 1997] e [Tokuda et al., 1998].

É muito importante, porém, ressaltar que a validade deste método se restringe a cenários para os quais os atratores estranhos associados aos mínimos locais perdem sua estabilidade para um valor menor que o associado ao mínimo global, dentro do contexto das distintas condições iniciais. Isto ocorre para os canais MA(1) estudados, mas não para os canais AR(1) e MA(2), o que mostra que o método não tem aplicabilidade universal.

7.1.1. Aplicação ao Algoritmo Estocástico

O método mostrou-se eficaz também para o caso estocástico. Sua fundamentação, entretanto, não é a mesma do caso determinístico, que vinha sendo nossa base de argumentação. O bom funcionamento do método na versão estocástica pode ser atribuído à diferença de profundidade dos mínimos que, aliada ao ruído de gradiente, pode fazer com que o aumento do passo de adaptação leve a uma divergência prematura de um dado

mínimo local em relação ao global. Podemos visualizar este processo como um escape de natureza aleatória, valendo inclusive uma certa analogia com os resultados obtidos em [Frater et al., 1995]. Resolvemos analisar os diagramas de bifurcação de uma realização estocástica do CMA, para o modelo de canal MA(1) com $\alpha = 0.6$. Os diagramas associados ao mínimo global estão na Fig. 7.1, enquanto os associados ao mínimo local estão na Fig. 7.2. As inicializações são feitas nos moldes do capítulo anterior, sempre próximas ao mínimo que se deseja analisar. Vale lembrar que, para uma configuração idêntica de canal e de equalizador, os mínimos do caso determinístico coincidem com os do caso estocástico, pois a função custo a ser otimizada é a mesma.

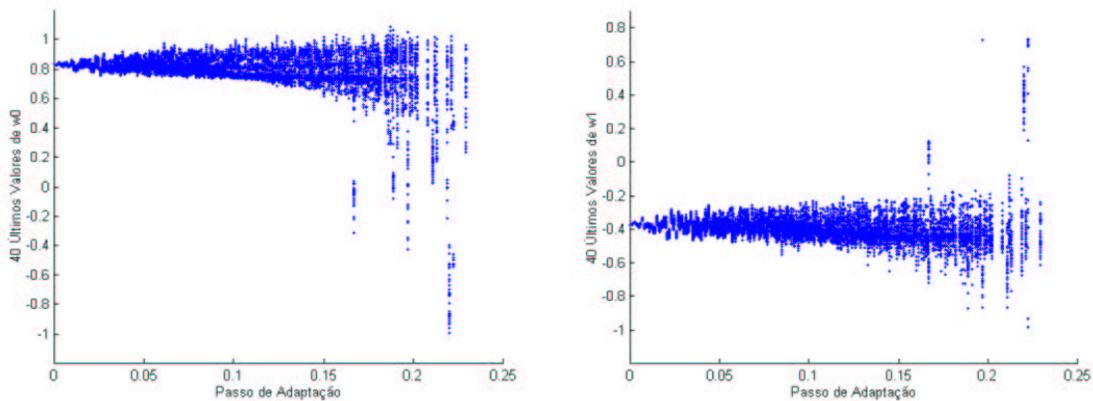


Figura 7.1: Diagramas de Bifurcação ($\mathbf{w}(0) = [0.831, -0.372]^T$)

A Fig. 7.1 mostra que os diagramas de bifurcação associados ao mínimo global mantêm-se aproximadamente contínuos até a divergência, salvo em alguns valores onde há descontinuidades em uma faixa estreita. Ao longo do diagrama o estado permanece sempre próximo ao mínimo ou ao seu simétrico, salvo nas pequenas faixas comentadas.

Quando analisamos os diagramas associados ao mínimo local, mostrados na Fig. 7.2, vemos que há um comportamento diferente: o sistema passa, a partir de um valor de μ pouco menor que 0.1, a apresentar convergência para o mínimo global, de forma análoga ao verificado na seção 6.2.2. Portanto, um cenário deste tipo também é propício à aplicação do método descrito na seção 7.1.

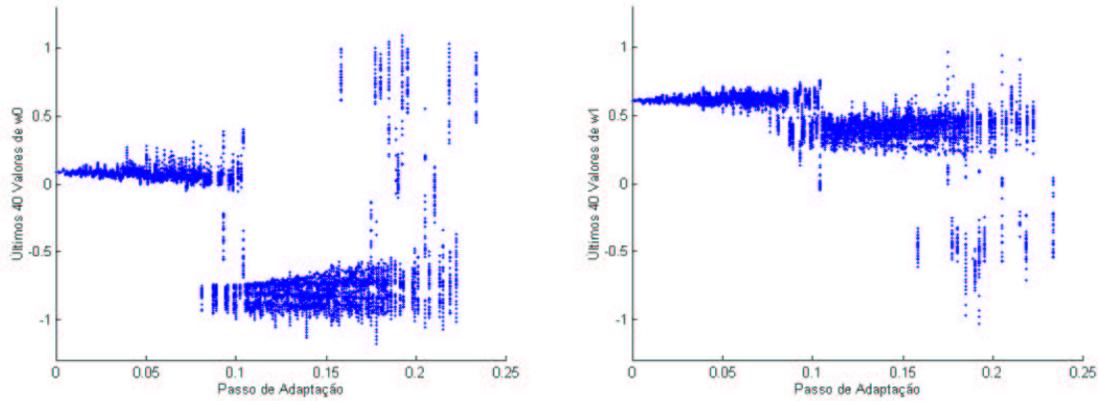


Figura 7.2: Diagramas de Bifurcação ($\mathbf{w}(0) = [0.084 \ 0.607]^T$)

Passemos, pois, ao caso do canal AR(1), já exposto em 6.2.5. Utilizaremos as mesmas configurações do caso determinístico. Para uma inicialização próxima ao mínimo global, obtivemos os diagramas de bifurcação apresentados na Fig. 7.3. Podemos observar que não há escape dos mínimos globais, havendo apenas regiões de convergência e divergência.

Para uma inicialização próxima ao mínimo local, foram obtidos os diagramas da Fig. 7.4. Tais diagramas indicam uma oscilação em torno do mínimo local para valores baixos de μ , o que indica sua pouca profundidade, confirmada pelo escape para o mínimo global com um ligeiro aumento de μ .

Trata-se de um cenário perfeitamente adequado à aplicação do método, que deve levar a resultados muito bons.

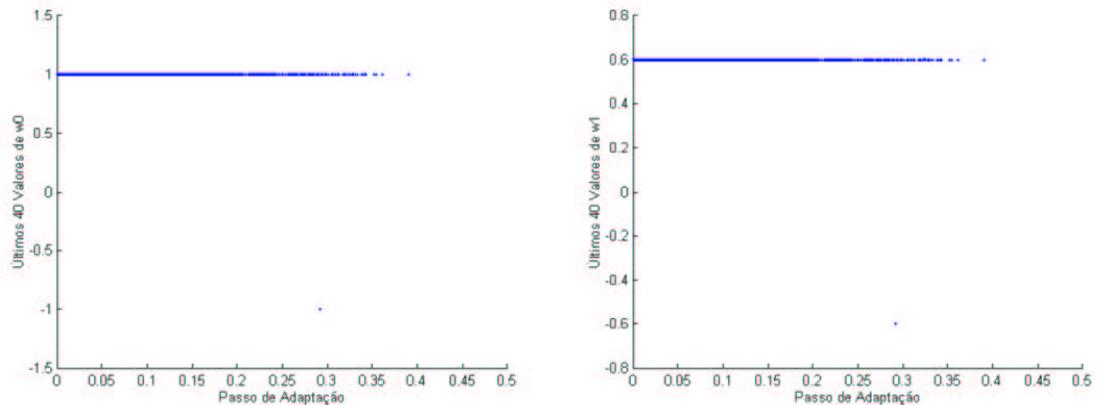


Figura 7.3: Diagramas de Bifurcação ($\mathbf{w}(0) = [1.01 \ 0.6]^T$)

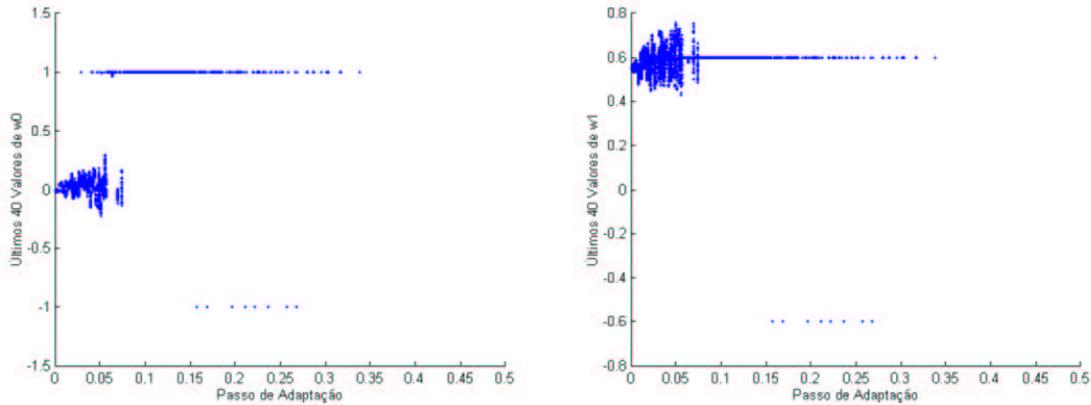


Figura 7.4: Diagramas de Bifurcação ($\mathbf{w}(0) = [0 \ 0.5575]^T$)

Analisemos agora uma realização do sistema estocástico quando é adotado o modelo de canal MA(2) descrito em 6.2.4. Para uma inicialização próxima ao mínimo global, foi obtida a Fig. 7.5.

Podemos notar que para alguns valores de passo (por exemplo, para μ variando entre 0.025 e 0.03 aproximadamente), há uma flutuação aleatória na vizinhança de um outro mínimo. Isto ocorre, para faixas de tamanho distinto, em outras regiões do diagrama. Portanto, não há, para a realização acima, uma constante convergência para o mínimo global. Quando o valor de μ está próximo a 0.025, as flutuações do algoritmo já apresentam uma possibilidade razoável de escapar do mínimo global, como observamos a partir de outras realizações. Vejamos agora, através dos diagramas da Fig. 7.6, como se comporta o sistema para uma inicialização próxima ao mínimo local mais profundo.

O melhor mínimo local que, como afirmamos em 6.2.4, tem um custo muito próximo ao do mínimo global, apresenta um comportamento semelhante. Podemos observar uma banda na qual o estado oscila em torno dele, havendo portanto, escape nesta realização.

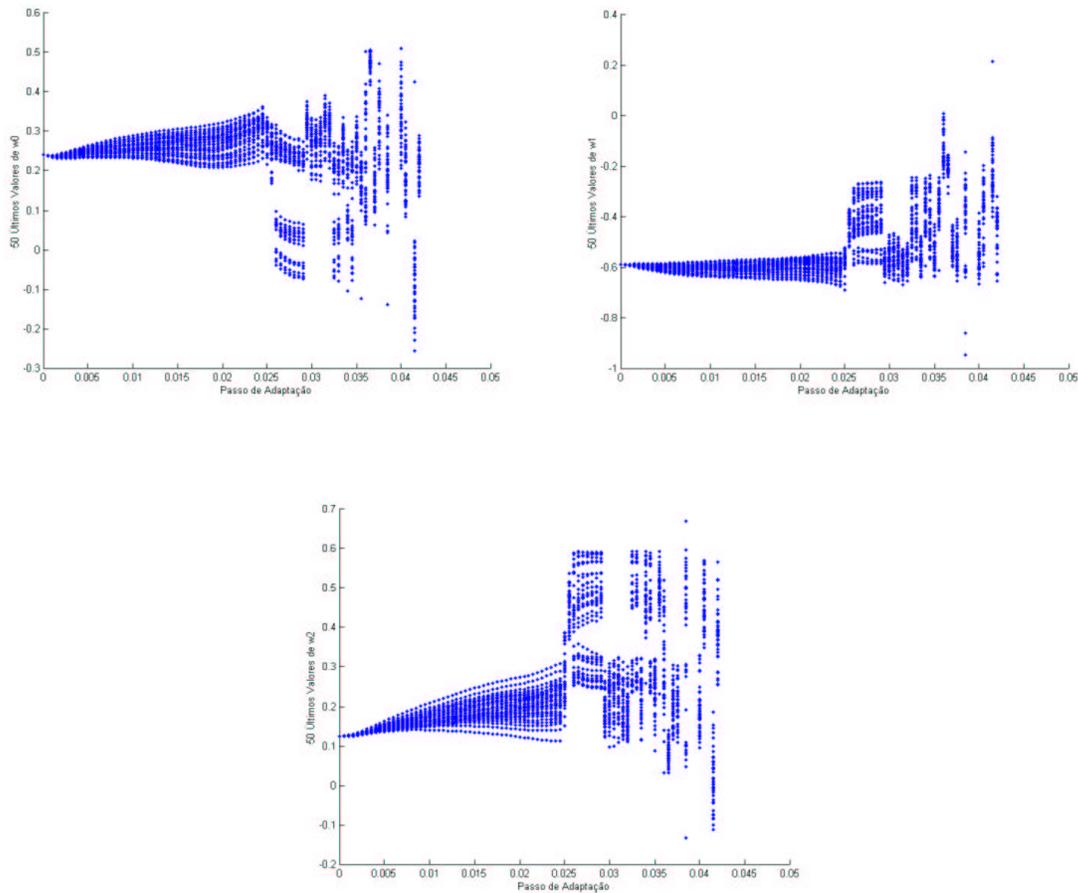


Figura 7.5: Diagramas de Bifurcação ($\mathbf{w}(0) = [0.241 \ -0.589 \ 0.124]^T$)

O cenário exposto mostra não ser este um caso onde o método deve conseguir substanciais elevações na taxa de convergência global, por não haver uma diferença pronunciada entre o padrão de estabilidade dos dois mínimos. Porém, operar em qualquer um dos dois é praticamente equivalente, neste caso, no que se refere ao desempenho em termos de abertura de olho. Vale ressaltar que estes dois mínimos não são os únicos para este caso, sendo que os analisamos por serem os mais significativos do ponto de vista do custo.

Assim, por motivos distintos, o método pode ser útil tanto no caso determinístico quanto no estocástico, sendo que neste último caso o potencial de aplicação é bem mais significativo do ponto de vista da prática.

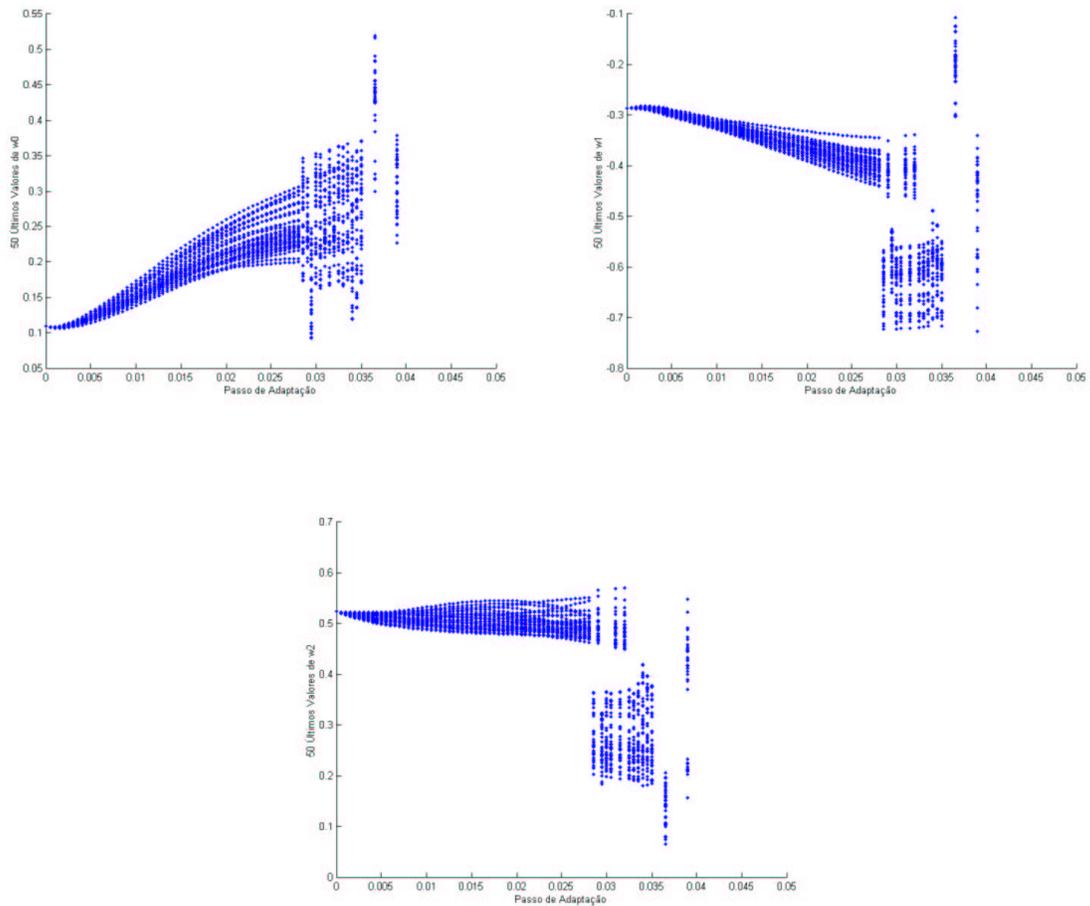


Figura 7.6: Diagramas de Bifurcação ($\mathbf{w}(0) = [0.110 \ -0.287 \ 0.524]^T$)

7.2. O Método

Uma vez expostos os princípios básicos do método, resta discutir alguns aspectos da implementação algorítmica. O primeiro deles é: como obter um passo adequado para que o sistema opere na condição satisfatória descrita na seção anterior? Uma escolha viável e satisfatória nos casos adequados à aplicação do método é utilizar o maior valor de passo para o qual não há divergência para o infinito. Este valor, caso o cenário seja favorável, deverá corresponder a uma região associada ao mínimo global. Como, porém, obter este valor? Uma idéia é iniciar a operação com um passo para o qual há divergência e ir gradualmente diminuindo-o, repetindo o processo até que o sistema não mais divirja, para a condição inicial escolhida. Pode-se então deixar o sistema oscilar por algumas iterações

para que ele convirja para as vizinhanças do atrator devido. Esta redução de passo pode ser feita de forma geométrica, fazendo uma iteração do tipo:

$$\mu(k+1) = \xi \cdot \mu(k) \quad (7.1)$$

onde $\xi < 1$. Quando não houver divergência, armazena-se o valor μ_{op} para o qual tal fato foi verificado. É bom lembrar que o “não haver divergência” é uma expressão de natureza indutiva, correspondendo à asserção “não haver divergência para um dado número, considerado suficientemente grande, de iterações”. Neste processo pode ainda ser introduzido ainda uma outra grandeza, um fator de correção κ que indica um refinamento do valor de operação μ_{op} . Este refinamento se dá da seguinte forma: quando se obtém o primeiro valor de $\mu(k)$ para o qual não há divergência, não se faz ainda $\mu_{op} = \mu(k)$, mas ainda se repete o método (iterando o passo segundo (7.1)) até o valor $\mu(k+\kappa)$, que então é escolhido como μ_{op} . Isto traz maior robustez à estimativa do passo quando o algoritmo em questão tem natureza estocástica.

Uma vez obtido o valor μ_{op} , roda-se o algoritmo por algumas iterações. Isto garante que o estado terá convergido para um atrator associado ao mínimo global. Para que o valor de $\mathbf{w}(n)$ possa então convergir para uma região ainda mais próxima do mínimo desejado, deve-se reduzir ainda mais o valor do passo. Esta nova redução se faz de forma abrupta, para um valor suficientemente pequeno, de modo que o ponto volte a ser estável e atraia as trajetórias em sua vizinhança. Após algumas iterações, o estado estará suficientemente próximo deste ponto, e pode-se usar um valor final das iterações como condição inicial \mathbf{w}_{inic} . Este valor é o objetivo final do método e pode então ser usado em definitivo. Espera-se que este método possa aumentar a taxa de convergência global do algoritmo, fornecendo condições iniciais melhores que aquelas que alimentam o primeiro passo do mesmo, pelo menos em uma boa parte dos casos que levariam espontaneamente à convergência local.

7.3. *Divisão em Fases*

A partir das linhas gerais dadas na seção anterior, podemos melhor sistematizar o método através de uma divisão em três fases distintas, descritas a seguir.

7.3.1. *Primeira Fase*

O objetivo da primeira fase é obter o valor de μ_{op} . Para tanto, começa-se com uma condição inicial qualquer e um passo inicial $\mu(0)$ que será iterado segundo (7.1). Vale aqui destacar que usamos a variável k , e não n , para a iteração do passo, pois esta variável conta as realizações do algoritmo e não as iterações no tempo de uma dada realização. Toda vez que se efetua uma nova realização, ou seja, quando k muda, apenas o passo de adaptação se modifica.

Basicamente, são determinadas a condição inicial e os fatores ξ e κ . Após esta escolha, o algoritmo é iterado até que uma das duas condições ocorra: ou o valor absoluto de um dos componentes do vetor de pesos ultrapassa um limiar, afigurando divergência, ou é atingido um número limite de iterações, configurando assim convergência para um certo atrator. No primeiro caso, k é atualizado e o processo repetido. No segundo, há a repetição por κ vezes, quando então, se não houver divergência em nenhum caso, é obtido μ_{op} como o valor final de passo.

7.3.2. *Segunda Fase*

Uma vez obtido o valor μ_{op} , utiliza-se o mesmo no algoritmo, com a mesma condição inicial da fase anterior. Após algumas iterações, que permitem o desvanecimento dos transitórios, reduz-se abruptamente o passo até um valor pequeno, mas que permita uma convergência razoavelmente rápida para as vizinhanças do ponto fixo. Armazena-se então o valor do estado, que é o objetivo a ser atingido nesta fase. Ele será chamado de \mathbf{w}_{inic} e, como veremos, será a condição inicial do algoritmo na próxima fase.

7.3.3. Terceira Fase

A terceira fase consiste tão somente da operação normal do CMA, com o passo escolhido pelo projetista, utilizando a condição inicial \mathbf{w}_{inic} .

7.4. Resultados

7.4.1. Configurações Gerais

Primeiro Passo:

A condição inicial variou segundo uma “grade” de 441 pontos uniformemente distribuídos, centrada na origem e com lado unitário. Tal grade está na Fig. 7.7.

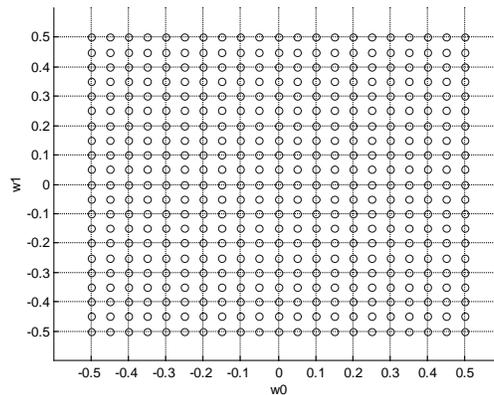


Figura 7.7: “Grade” de Pontos

Todos os pontos foram utilizados em cada uma das simulações, para que pudéssemos estimar a taxa de convergência global do sistema. O critério de divergência foi:

$$|w_0(n)| > 2 \quad (7.2)$$

e o número de iterações limitadas para que a convergência fosse aceita foi:

$$N_{\text{conv}} = 100 \quad (7.3)$$

O fator ξ foi escolhido como sendo $\frac{1}{1.01}$, sendo $\mu(0) = 0.5$.

Segundo Passo:

No segundo passo, o valor de μ foi mantido em μ_{op} por 20 iterações, sendo depois mantido em 0.005 por 480 iterações, sendo o último valor do estado retido como \mathbf{w}_{inic} .

Terceiro Passo:

Utilizou-se $\mathbf{w}(0) = \mathbf{w}_{\text{inic}}$ e foram rodadas 700 iterações do algoritmo. A partir do último valor, utilizava-se uma medida de distância entre o módulo dos pesos e o valor ótimo para determinar se a convergência foi global ou local. Também se fez uma medida para averiguar uma possível divergência.

A partir destas configurações foram calculadas duas taxas de convergência: uma dada por:

$$T_1 = \frac{N_{\text{glob}}}{N_{\text{pontos}}} \quad (7.4)$$

onde N_{glob} é o número de condições iniciais que conduziram à convergência global e N_{pontos} o total de condições iniciais. A segunda taxa é calculada através da fórmula:

$$T_2 = \frac{N_{\text{glob}}}{N_{\text{conv}}} \quad (7.5)$$

onde N_{conv} é o número de condições iniciais que levaram a uma solução final limitada, ou seja, a um comportamento de convergência.

Também foi calculada a taxa de convergência global sem a aplicação do método, para que se pudesse ter um parâmetro de avaliação de sua eficiência. No caso do algoritmo estocástico, para se ter um padrão de comparação mais fiel, foi utilizado o mesmo sinal transmitido tanto no levantamento da taxa sem o método quanto nos passos dois e três.

Avaliamos a aplicação do método às duas versões do algoritmo do módulo constante: a determinística e a estocástica. Os resultados serão expostos e discutidos a seguir.

7.4.2. Canal MA(1) de Fase Mínima – Algoritmo Estocástico

Em primeiro lugar, foi utilizado o modelo MA(1), descrito em 6.2.2, com $\alpha = 0.6$. Esta escolha de α configura, como já vimos, um canal de fase mínima. As taxas T_1 e T_2 foram obtidas para κ variando de 0 a 110, em intervalos de comprimento 5, totalizando 23 pontos. Os resultados estão na Fig. 7.8.

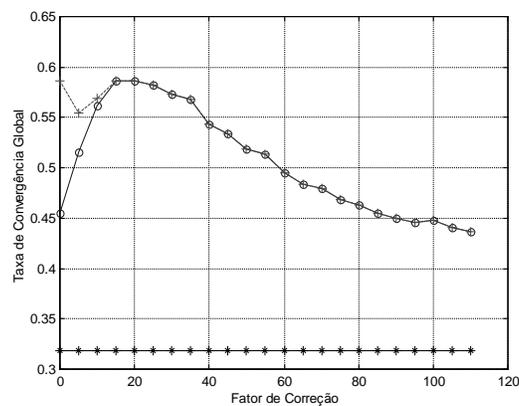


Figura 7.8: Taxas de Convergência Global (o - T_1 , + - T_2 e v- sem método)

A Fig. 7.8 mostra que o método elevou a taxa de convergência global em todos os casos, sendo que no melhor deles a elevação foi de aproximadamente 27%. Para valores baixos de κ , houve alguns casos divergentes, o que é verificável pela discrepância entre as curvas de T_1 e T_2 . De modo geral, o método mostrou-se bastante adequado a este canal, para as configurações escolhidas.

7.4.3. Canal MA(1) de Fase Máxima – Algoritmo Estocástico

Novamente adotamos o modelo MA(1), porém com $\alpha = 1.5$. A Fig. 7.9 traz os resultados.

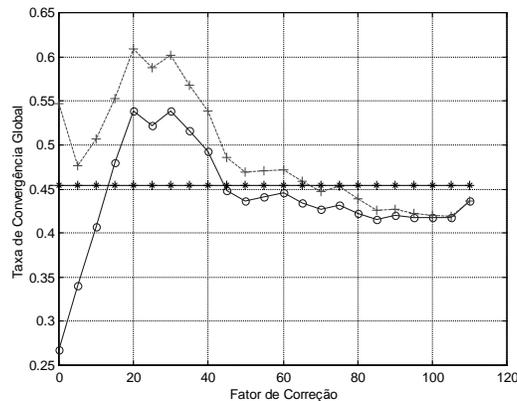


Figura 7.9: Taxas de Convergência Global (o - T₁, + - T₂ e ∇- sem método)

Neste caso, há uma região (de $\kappa = 15$ a 40) para o qual o método representa um ganho. Se desprezarmos os casos divergentes, a faixa é mais larga, dada por $\kappa = 0$ a 65. Ainda assim, é vantajoso o uso do método, pois mesmo na maior parte da região para a qual ele não consegue elevar a taxa de convergência, T₁ e T₂ permanecem próximas aos valores sem o método.

7.4.4. Canal MA(1) de Fase Mínima – Algoritmo Determinístico

Embora o algoritmo determinístico não seja empregado na prática, tendo em vista que ele deu origem à idéia do método, é bastante interessante que o mesmo seja aplicado a tal procedimento, para que possamos verificar se a intuição original de fato era fundamentada. Utilizamos, como de costume, $\alpha = 0.6$. O resultado está na Fig. 7.10.

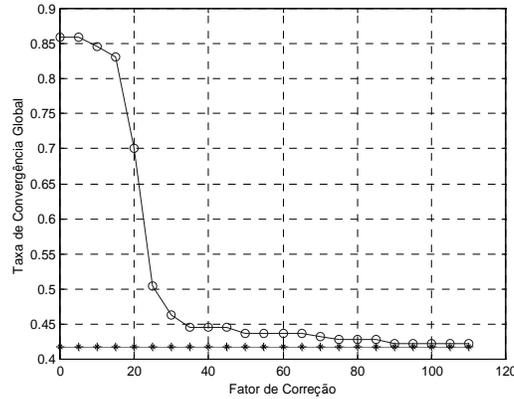


Figura 7.10: Taxas de Convergência Global (o - T_1/T_2 e v- sem método)

Notamos novamente um bom aumento (da ordem de pouco mais do que 40%) da taxa de convergência global. T_1 e T_2 foram coincidentes para todos os valores testados. Vale destacar dois padrões de regularidade associados ao fato do algoritmo ser de natureza determinística:

- 1) Quando o fator de correção é nulo, o método apresenta um valor muito bom para a taxa de convergência global. Isto pode ser explicado pelo fato de não haver mais as incertezas de natureza aleatória em relação ao máximo valor para o qual não há divergência.
- 2) Quando o valor do fator de correção tende a valores altos, T_1 se iguala ao valor da taxa de convergência sem a aplicação do método, indicando operação em valores para os quais há uma chance cada vez menor de w_{inic} estar na bacia de atração do mínimo global. O limite seria a utilização de um fator de correção tão alto que levaria a um μ_{op} na região de estabilidade dos mínimos, caso no qual o método não surte mais efeito.

7.4.5. Canal MA(1) de Fase Máxima – Algoritmo Determinístico

Ainda no âmbito do algoritmo determinístico, utilizamos o modelo de canal MA(1) de fase máxima, com $\alpha = 1.5$. Os resultados obtidos estão sumarizados na Fig. 7.11.

Podemos observar que o método produz resultados muito bons também para o canal de fase máxima. O aumento máximo da taxa, para $\kappa = 10$, foi de 30%, sendo que para

nenhum valor de κ houve uma degradação de desempenho imposta pelo método. Novamente, observamos que T_1 e T_2 foram coincidentes.

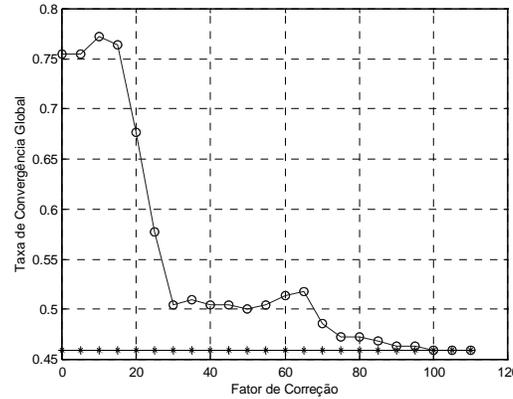


Figura 7.11: Taxas de Convergência Global (o - T_1/T_2 e ∇- sem método)

7.4.6. Canal MA(1) de Fase Mínima – Algoritmo DD

A aplicabilidade do método quando empregado o algoritmo do módulo constante levou-nos a tentar utilizá-lo no contexto de um outro algoritmo, o da decisão direta. Como neste trabalho não fizemos nenhum tipo de análise teórica envolvendo tal técnica, iremos apenas mostrar os resultados de sua aplicação. Iremos ainda limitar a aplicação do método aos canais MA(1) de fase mínima e máxima.

Em primeiro lugar, fizemos $\alpha = 0.6$, caracterizando o canal como de fase mínima. Para o parâmetro κ foi utilizada uma variação de 0 a 110, com espaçamento de 5 entre pontos sucessivos. As taxas T_1 e T_2 , além da taxa de convergência global sem a aplicação do método estão na Fig. 7.12. Como T_1 e T_2 apresentaram os mesmos valores para todos os parâmetros κ , as duas estão representadas por uma única curva.

Neste caso o sistema teve um desempenho ainda melhor que quando foi utilizado o CMA, atingindo quase 100% de convergência global para $\kappa = 25$. Isto representou um aumento de pouco mais de 65% em relação à não aplicação do procedimento, o que é deveras significativo.

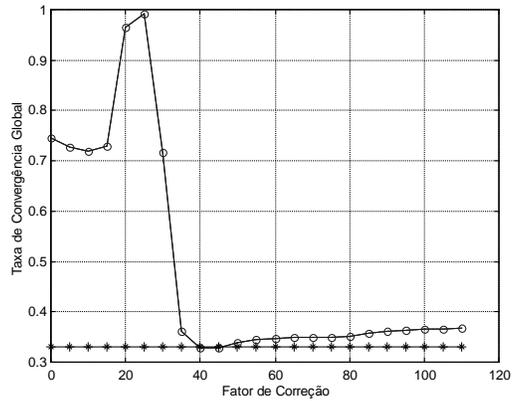


Figura 7.12: Taxas de Convergência Global (o - T_1/T_2 e ∇- sem método)

7.4.7. Canal MA(1) de Fase Máxima – Algoritmo DD

Escolhemos para este caso, como de costume, $\alpha = 1.5$. Após alguns ensaios, achamos justificável operar com κ variando de 0 a 20, com intervalo unitário entre os valores sucessivos. O resultado está apresentado na Fig. 7.13.

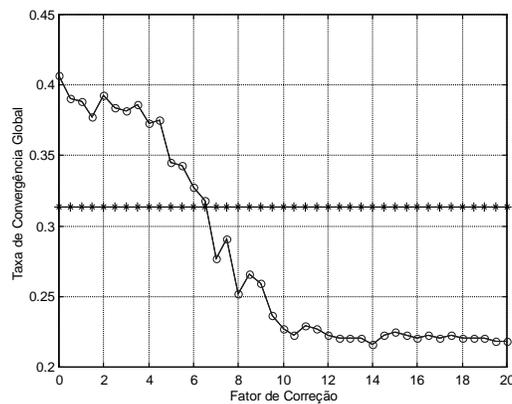


Figura 7.13: Taxas de Convergência Global (o - T_1/T_2 e ∇- sem método)

Novamente o canal não propiciou um desempenho tão bom quanto o verificado para o canal de fase mínima. O algoritmo trouxe algum benefício para κ entre 0 e 8 (com uma elevação máxima próxima de 10%), sendo que nos demais valores houve quedas de desempenho que chegaram a patamares próximos a 10%. Isto indica que neste caso a escolha de κ tem um papel muito relevante para que o método seja de algum proveito.

Embora de aplicação ainda não delimitada, o método produz elevações significativas na taxa de convergência global não só do algoritmo do módulo constante como na do algoritmo DD, levando-nos a crer que seja aplicável a técnicas de Bussgang em geral. Aliar a inicialização *center-spike* a um método deste tipo deve produzir resultados ainda melhores.

8. Conclusão

A análise de convergência de um algoritmo adaptativo é um aspecto teórico fundamental para a avaliação de seu desempenho e a delimitação de seu domínio de aplicabilidade. O objetivo deste trabalho foi analisar o comportamento dinâmico de um algoritmo cego, segundo uma abordagem que foge ao domínio usual de análise, baseado no binômio “convergência para ponto fixo” / “divergência”. Tais comportamentos decorrem da não-linearidade, inerente a algoritmos de Bussgang, que dá origem a um cenário típico de Feigenbaum, caracterizado pela ocorrência de ciclos-limite e caos, conforme se pôde observar ao longo do capítulo 6 deste trabalho.

Estes dois tipos de comportamento foram verificados para a versão determinística do algoritmo CMA em modelos de canal tanto FIR quanto IIR, de fase mínima e não-mínima, o que propicia uma certa generalidade às constatações do trabalho. A base desta análise foram as ferramentas usuais de teoria de sistemas dinâmicos não-lineares, expostas nos capítulos 2 e 3, além dos conceitos de filtragem adaptativa, apresentados no capítulo 4 de forma a se introduzir as técnicas não-supervisionadas.

Constatamos que a evolução dinâmica do algoritmo, para um aumento do valor do passo de adaptação, é altamente dependente da condição inicial, sendo que verificamos, ao inicializar o algoritmo com uma condição próxima a mínimos distintos, que há diferentes atratores estranhos e ciclos-limite que podem ser visitados pelo estado do sistema. Isto mostra que há uma diferenciação essencial entre o tipo de comportamento associado a cada um dos mínimos envolvidos, o que constitui um aspecto bastante relevante.

Para a versão estocástica do algoritmo, não pudemos realizar uma análise tão aprofundada quanto no caso determinístico, pela enorme complexidade envolvida no estudo de um sistema dinâmico não-linear com parâmetros aleatórios. Limitamo-nos a provar que, para uma dada configuração, o algoritmo do módulo constante pode ter seu estado adaptado segundo uma órbita caótica. Apesar do domínio restrito da análise, as características da expressão iterativa do CMA leva-nos a afirmar que vai ocorrer comportamento caótico, mesmo para modelos de canal/equalizador de ordens superiores às aqui estudadas.

A partir dos resultados obtidos, propusemos no capítulo 7 um método que, para alguns casos, pode aumentar a taxa de convergência global de algoritmos de Bussgang. Em

todos os modelos analisados, o método permitiu que a taxa de convergência global fosse aumentada, sendo que os melhores resultados, mesmo que apenas indicativos, foram obtidos para o CMA determinístico e para o DD. A validade do procedimento proposto precisa ser investigada com mais detalhe, para que seja verificada a gama de aplicabilidade do método e, caso esta seja razoavelmente ampla, para que sejam resolvidos problemas de ordem prática como a escolha do fator de correção.

Além das contribuições mais específicas, citadas acima, procuramos agregar a esta dissertação um valor tutorial, para que o leitor pudesse ter uma melhor noção histórica do estudo teórico sobre o CMA, desde a sua formulação original em 1980. Apresentamos, assim, no capítulo 5 uma discussão acerca de alguns resultados teóricos expostos na literatura ao longo de 20 anos. É interessante perceber como evoluiu a concepção acerca do método, dos primeiros estudos envolvendo o equalizador duplamente infinito, passando pela constatação de convergência para mínimos locais, até as atuais abordagens baseadas em idéias geométricas.

A complexidade inerente aos sistemas não-lineares nos encoraja a formular uma última observação: ainda há muito para se aprender sobre a convergência de algoritmos adaptativos não-lineares como os de Bussgang, e é necessário que não nos atenhamos apenas às concepções herdadas da clássica análise dos algoritmos supervisionados lineares.

9. Bibliografia

L. A. Aguirre “A Tutorial Introduction to Nonlinear Dynamics and Chaos, Part I: Tools and Benchmarks”, *Controle & Automação*, Vol. 7, no. 1, pp. 29-49, 1996.

L. A. Aguirre “A Tutorial Introduction to Nonlinear Dynamics and Chaos, Part II: Modeling and Control”, *Controle & Automação*, Vol. 7, no. 1, pp. 50-66, 1996.

A. Benveniste, M. Goursat, and G. Ruget “Robust Identification of a Nonminimum Phase System: Blind Adjustment of a Linear Equalizer in Data Communications”, *IEEE Trans. on Automat. Control*, AC-25, pp. 385-399, 1980.

J. L. Boldrini, S. I. R. Costa, V. L. Figueiredo, H. G. Wetzler *Álgebra Linear*, 3^a Edição, Editora Harbra, 1986.

C. Boyer, U. Merzbach *A History of Mathematics*, Segunda Edição, Wiley, 1991.

J. Choi, I. Song, R. Park “Some Convergence Properties of Godard’s Quartic Algorithm”, *Signal Processing*, Vol. 56, pp. 313-320, 1997.

R. Cusani, A. Laurenti “Convergence Analysis of the CMA Blind Equalizer”, *IEEE Transactions on Communications*, Vol. 43, No. 2/3/4, pp. 1304-1307, 1995.

Z. Ding, R. A. Kennedy, B. D. O. Anderson, and C. R. Johnson “Ill-Convergence of Godard Blind Equalizers in Data Communication Systems”, *IEEE Transactions on Communications*, vol. COM-39, pp. 1313-1327, 1991.

Z. Ding, C. R. Johnson, 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, 1992.

Z. Ding, R. A. Kennedy “On the Whereabouts of Local Minima for Blind Adaptive Equalizers”, IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing, Vol. 39, No. 2, pp. 119-123, 1992.

Z. Ding, C. R. Johnson “On the Nonvanishing Stability of Undesirable Equilibria for FIR Godard Blind Equalizers”, IEEE Transactions on Signal Processing, Vol. 41, No. 5, pp. 1940 – 1944, 1993.

M. J. Feigenbaum “Quantitative Universality for a Class of Nonlinear Transformations”, Journal of Statistical Physics, Vol. 19, No. 1, 1978.

M. J. Feigenbaum “The Universal Metric Properties of Nonlinear Transformations”, Journal of Statistical Physics, Vol. 21, No. 6, 1979.

M. J. Feigenbaum “Universal Behavior in Nonlinear Systems”, Physica 7D, pp. 16-39, 1983.

N. Fiedler-Ferrara, C. C. do Prado *Caos: uma Introdução*, Edgard Blücher, 1994.

G. J. Foschini “Equalizing Without Altering or Detecting Data”, AT&T Technical Journal, Vol. 64, No. 8, pp. 1885-1911, 1985.

M. R. Frater, R. R. Bitmead, C. R. Johnson “Local Minima Escape Transients by Stochastic Gradient Descent Algorithms in Blind Adaptive Equalizers”, Automatica, Vol. 31, No. 4, pp. 637-641, 1995.

J. Gleick *Caos: A Criação de uma Nova Ciência*, 4ª Edição, Editora Campus, 1990.

A. P. C. Gonçalves, J. C. Geromel e J. B. R. do Val “Sistemas Lineares Discretos Sujeitos a Saltos Markovianos”, Relatório Final de Iniciação Científica, FEEC/Unicamp, 2000.

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, 1980.

C. Grebogi, E. Ott, J. A. Yorke “Crises, Sudden Changes in Chaotic Attractors, and Transient Chaos”, Physica 7D, pp. 181-200, 1983.

M. Gu, L. Tong “Geometrical Characterization of Constant Modulus Receivers”, IEEE Transactions on Signal Processing, Vol. 47, No. 10, pp. 2745 – 2756, 1999.

J. Guckenheimer, P. Holmes *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Springer Verlag, 1983.

M. Hasler “Chaos”, Anais do International Symposium on Signals, Systems and Electronics (ISSSE), Paris, França, 1992.

S. Haykin *Adaptive Filter Theory*, Prentice Hall, 1996.

S. Haykin *Neural Networks: a Comprehensive Foundation*, MacMillan, 1994.

M. Hénon “A Two-dimensional Mapping with a Strange Attractor” Communications in Mathematical Physics, Vol. 50, pp. 69-77, 1976.

B. R. Hunt, J. A. Yorke, “Maxwell on Chaos”, Nonlinear Science Today, Vol. 3, No. 1, pp. 1 – 4, 1993.

C. R. Johnson, B. D. O. Anderson “Godard Blind Equalizer Error Surface Characteristics: White, Zero-mean, Binary Source Case”, *International Journal of Adaptive Control and Signal Processing*, no. 9, pp. 301-324, 1995.

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, 1998.

R. A. Kennedy, Z. Ding “Blind Adaptive Equalizers for Quadrature Amplitude Modulated Communication Systems Based on Convex Cost Functions”, *Optical Engineering*, Vol. 31, No. 6, pp. 1189-1199, 1992.

E. Kreyszig *Advanced Engineering Mathematics*, Wiley, 1993.

S. Lambotharan, J. Chambers, C. R. Johnson “Attraction of Saddles and Slow Convergence in CMA Adaptation”, *Signal Processing*, Vol. 59, pp. 335-340, 1997.

Y. Li, Z. Ding “Convergence Analysis of Finite Length Blind Adaptive Equalizers”, *IEEE Transactions on Signal Processing*, Vol. 43, No. 9, pp. 2120-2129, 1995.

Y. Li, K. J. Ray Liu, Z. Ding “Length- and Cost-dependent Local Minima of Unconstrained Blind Equalizers”, *IEEE Transactions on Signal Processing*, Vol. 44, No. 11, pp. 2726-2735, 1996.

Y. Li, K. J. Ray Liu “Static and Dynamic Convergence Behavior of Adaptive Blind Equalizers”, *IEEE Transactions on Signal Processing*, Vol. 44, No. 11, pp. 2736-2745, 1996.

A. V. Oppenheim, A. S. Willsky, S. H. Nawab *Signals and Systems*, Prentice Hall, 1997.

R. Palazzo Jr. *Transmissão Digital: Fundamentos e Aplicações*, Material Didático do Curso de EE088, Unicamp, 1999.

T. S. Parker, L. O. Chua “Chaos: a Tutorial for Engineers”, Proceedings of the IEEE, Vol. 75, No. 8, pp. 982 – 1008, 1987.

P. Regalia “On the Equivalence Between the Godard and Shalvi-Weinstein Schemes of Blind Equalization”, Signal Processing, No. 73, pp. 185-190, 1999.

P. 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, 1999.

C. A. F. da Rocha *Técnicas Preditivas para Equalização Autodidata*, Tese de Doutorado, FEEC/UNICAMP, 1996.

D. Ruelle “Strange Attractors”, The Mathematical Intelligencer 2, Vol. 3, pp. 126-137, 1980.

H. G. Schuster *Deterministic Chaos: an Introduction*, VCH Publishers, 1988.

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, 1990.

K. R. Symon *Mecânica*, 6^a Edição, Editora Campus, 1982.

J. R. Treichler, B. G. Agee “A New Approach to Multipath Correction of Constant Modulus Signals”, IEEE Transactions on Acoustic, Speech and Signal Processing, Vol. ASSP – 31, no. 2, pp. 459 – 472, 1983.

I. Tokuda, K. Onodera, R. Tokunaga, K. Aihara, T. Nagashima “Global Bifurcation Scenario for Chaotic Dynamic Systems that Solve Optimization Problems and Analysis of Their Optimization Capability” – Electronics and Communications in Japan - Part 3, Vol. 81, No. 2, 1998.

F. J. Von Zuben *Notas de Aula do Curso EA932 – Sistemas de Controle II*, FEEC/Unicamp, 2001.

H. Zeng, L. Tong, C. R. Johnson “An Analysis of Constant Modulus Receivers” – IEEE Transactions on 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 Transactions on Information Theory, Vol. 44, No. 4, pp. 1523 – 1538, 1998.

C. Zhou, T. Chen “Chaotic Annealing for Optimization”, Physical Review E, Vol. 55, pp. 2580-2587, 1997.

10. Índice de Autores

L. A. Aguirre 53
B. G. Agee 106
K. Aihara 202
B. D. O. Anderson 110, 112, 118, 120, 123, 133, 180, 181
J. D. Behm 2
A. Benveniste 101, 112
R. R. Bitmead 3, 154, 203
J. L. Boldrini 43
C. Boyer 5
D. R. Brown 2
R. A. Casas 2
J. Chambers 129
T. Chen 202
J. Choi 127
L. O. Chua 11, 15, 36, 44, 46
S. I. R. Costa 43
R. Cusani 152
Z. Ding 110, 112, 113, 115, 117, 130, 133, 135, 136, 137, 138, 139, 140, 180, 181
T. J. Endres 2
M. J. Feigenbaum 58, 59, 60
N. Fiedler-Ferrara 7, 11, 12, 14, 15, 17, 21, 22, 23, 28, 30, 31, 35, 36, 40, 44, 47, 53, 58
V. L. Figueiredo 43
G. J. Foschini 108, 109, 115, 121
M. R. Frater 3, 154, 203
J. C. Geromel 24, 192
J. Gleick 27, 44, 48, 58
D. N. Godard 99, 103, 112, 115
A. P. C. Gonçalves 24, 192

C. Grebogi 53
M. Goursat 101, 112
M. Gu 144, 147, 161
J. Guckenheimer 21, 22, 32, 41
S. Haykin 6, 9, 11, 14, 63, 73, 74, 79, 84, 85, 86, 90, 92, 98, 99, 100, 101
M. Hénon 42
P. Holmes 21, 22, 32, 41
B. R. Hunt 27
C. R. Johnson 2, 110, 112, 115, 117, 118, 120, 123, 133, 144, 145, 147, 180, 181
R. A. Kennedy 110, 112, 113, 115, 133, 136, 180, 181
E. Kreyszig 79, 80
S. Lambbotharan 129
A. Laurenti 152
Y. Li 124, 130, 135, 137, 138, 139, 140, 149, 152
M. Mboup 137
U. Merzbach 5
T. Nagashima 202
S. H. Nawab 65
K. Onodera 202
A. V. Oppenheim 65
E. Ott 53
R. Palazzo Jr. 24, 191
R. Park 127
T. S. Parker 11, 15, 36, 44, 46
C. C. do Prado 7, 11, 12, 14, 15, 17, 21, 22, 23, 28, 30, 31, 35, 36, 40, 44, 47, 53, 58
K. J. Ray Liu 124, 135, 137, 138, 149, 152
P. Regalia 137, 140
C. A. F. da Rocha 99
D. Ruelle 6, 38
G. Ruget 101, 112
P. Schniter 2

H. G. Schuster 27, 48, 58, 166
O. Shalvi 101, 115, 139
I. Song 127
K. R. Symon 32
R. Tokunaga 202
J. R. Treichler 106
I. Tokuda 202
L. Tong 144, 145, 147, 161
F. J. Von Zuben 5
E. Weinstein 101, 115, 139
H. G. Wetzler 43
A. S. Willsky 65
J. Yorke 27
H. Zeng 144, 145, 147
C. Zhou 202

Apêndice A: Trabalhos Realizados Durante o Mestrado

R. R. F. Attux, J. M. T. Romano “Um Estudo sobre Dinâmica Caótica em Equalizadores Autodidatas”, Anais do XVIII Simpósio Brasileiro de Telecomunicações (SBrT), Gramado, Brasil, 2000.

C.M. Panazio, H.C. Bertan, e R.R.F. Attux “Emprego de Mapas Auto-Organizáveis de Kohonen no Projeto de Decisores em Sistemas de Comunicação Digital”, Anais do XVIII Simpósio Brasileiro de Telecomunicações (SBrT), Gramado, Brasil, 2000.

A seguir anexamos uma cópia do primeiro destes trabalhos, ligado diretamente ao tema desta tese.

UM ESTUDO SOBRE DINÂMICA CAÓTICA EM EQUALIZADORES AUTODIDATAS

Romis Ribeiro de Faissol Attux e João Marcos Travassos Romano

DSPCom - Lab. de Proc. de Sinais para Comunicações Móveis - Depto. de Comunicações
Fac. de Eng. Elétrica e de Computação – Universidade Estadual de Campinas (Unicamp)
Caixa Postal 6101, CEP: 13083-970, Campinas - SP - Brasil
Tel.: (19) 7883703, Fax: (19) 2891395
{romisri, romano}@decom.fee.unicamp.br

SUMÁRIO

Neste trabalho, buscou-se estudar a ocorrência de comportamento caótico em um equalizador ajustado pelo critério do módulo constante. Para este fim, foram feitas duas análises do comportamento do processo iterativo de ajuste dos parâmetros do equalizador: uma estática, buscando a caracterização dos pontos de equilíbrio do processo, e uma dinâmica, avaliando a estabilidade destes pontos em função do passo de adaptação do método. Através da variação do passo de adaptação, verifica-se que, além de convergência para um dos mínimos e divergência para o infinito, o algoritmo do módulo constante apresenta zonas de comportamento periódico e também regiões onde há caos. O emprego de multiplicadores de Floquet e expoentes de Lyapunov serviu para dar maior embasamento matemático às conclusões do trabalho. Por fim, propõe-se um método que pode levar a um aumento da taxa de convergência global de algoritmos cegos.

1. INTRODUÇÃO

As técnicas de equalização conhecidas como *cegas* ou *autodidatas* têm despertado grande interesse desde que surgiram. O motivo é que as mesmas dispensam o uso de sequências de treinamento, que nem sempre são possíveis de se obter em uma aplicação prática.

Dentre os algoritmos adaptativos cegos, podemos destacar, como um dos mais populares, o CMA (Constant Modulus Algorithm), proposto por Godard em 1980 [7]. Este algoritmo tem sido extensivamente estudados em vários de seus aspectos, por diversos autores [1-4, 10, 11], na busca de um maior entendimento acerca de diversas características de convergência do mesmo.

O CMA pertence à família dos algoritmos de Bussgang [9]. Tais algoritmos têm, em sua fórmula iterativa, termos não-lineares em relação à saída do equalizador. A presença de não-linearidades na expressão de atualização dos coeficientes do filtro torna-a uma possível candidata a mostrar comportamento complexo, inclusive caos. A motivação deste trabalho foi a de investigar a ocorrência destes comportamentos em algoritmos baseados no critério do módulo constante.

Para a realização da análise do algoritmo, foram empregados dois modelos de canal/equalizador (ambos filtros FIR de mesma ordem): um ganho real e um modelo MA(1). Em ambos, foram feitas duas análises: uma estática, centrada na obtenção dos pontos de equilíbrio do método iterativo, e uma dinâmica, na qual o comportamento do sistema é estudado para diferentes valores do passo de adaptação. Na análise dinâmica, enfoca-se a questão da estabilidade dos pontos

de equilíbrio, bem como os comportamentos que emergem quando os pontos de mínimo perdem sua estabilidade.

Além dos comportamentos usualmente discutidos no contexto da equalização adaptativa, a saber, convergência para um ponto de mínimo e divergência para o infinito, foram observadas oscilações periódicas nos pesos do filtro e caos. Os resultados da análise nos levaram a propor as linhas gerais de um método de obtenção de convergência global, baseado nas características dinâmicas do algoritmo.

O artigo está estruturado como descreveremos a seguir. Na seção 2, são apresentados o problema geral da desconvolução autodidata e o algoritmo baseado no critério do módulo constante (CMA). Na seção 3, o critério é aplicado à equalização de dois canais, um ganho real e um filtro FIR com dois taps. São mostrados e comentados os resultados da análise estática e da análise dinâmica, para os dois modelos adotados. Na seção 4, são discutidos alguns elementos resultantes da análise precedente que podem ser explorados para melhorar a taxa de convergência global do algoritmo empregado. Finalmente, a seção 5 traz algumas conclusões e perspectivas de trabalhos futuros.

2. O PROBLEMA DA DESCONVOLUÇÃO AUTODIDATA

No âmbito da teoria de comunicações, pode-se dizer que um dos grandes problemas consiste em se obter uma boa estimativa de um sinal transmitido a partir de uma versão do mesmo distorcida por um canal. A versão distorcida, caso o canal seja considerado linear e invariante no tempo, será a *convolução* do sinal original com a resposta ao impulso do canal de transmissão. Portanto, para se obter uma estimativa do sinal original no receptor, é preciso realizar uma operação de *desconvolução*.

Quando este processo ocorre sem que se conheça a resposta ao impulso do canal de transmissão, fala-se de *desconvolução autodidata ou cega* [9].

Podemos visualizar isto em um modelo genérico do sistema de comunicações descrito acima, encontrado na Fig. 1. Neste modelo, $s(n)$ é o sinal transmitido, $x(n)$ é o sinal distorcido pelo canal e presente à entrada do equalizador, enquanto $y(n)$ é o sinal equalizado. Os parâmetros do equalizador transversal FIR são expressos pelo vetor $w(n)$. Podemos dizer que o objetivo da desconvolução é obter um valor de $y(n)$ tão próximo quanto possível de $s(n)$ ou de uma versão atrasada deste.

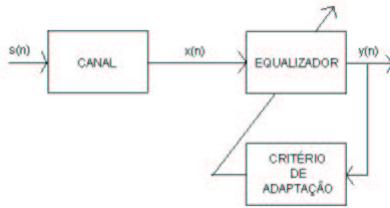


Figura 1. Modelo Canal-Equalizador

Um dos mais populares algoritmos para se realizar a desconvolução cega é o CMA (Constant Modulus Algorithm), pertencente à classe dos algoritmos de Bussgang [9]. Por ser autodidata, esta técnica não dispõe de um sinal desejado para comparar com a saída, como no caso da filtragem de Wiener. Porém, supondo que o sinal transmitido pertença a uma constelação bidimensional, pode-se definir uma dispersão em torno de um certo valor médio que, no caso de constelações com módulo constante, é o próprio quadrado do módulo. Assim, minimizar esta dispersão será fazer com que o sinal de saída tenha características de módulo semelhantes às da constelação do sinal transmitido. A expressão para a dispersão, no CMA, é a seguinte:

$$J(n) = E[|y(n)|^2 - R_2]^2 \quad (1)$$

onde

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

Sendo o equalizador um filtro FIR, podemos escrever:

$$y(n) = \mathbf{w}^H(n) \cdot \mathbf{x}(n) \quad (3)$$

Utilizando o método *steepest descent*, que fundamenta-se numa busca seguindo a direção contrária à do gradiente, temos a seguinte expressão de busca do ponto de operação do equalizador:

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \cdot \nabla J[\mathbf{w}(n)] \quad (4)$$

Utilizando a aproximação estocástica para o gradiente, temos:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \cdot \mathbf{e}^*(n) \cdot \mathbf{x}(n) \quad (5)$$

onde $\mathbf{e}(n)$ é um “erro” definido como:

$$\mathbf{e}(n) = y(n) \cdot (R_2 - |y(n)|^2) \quad (6)$$

3. ALGORITMO DO MÓDULO CONSTANTE E DINÂMICA CAÓTICA

Com o fim de analisar a expressão iterativa do algoritmo como um sistema dinâmico, iremos aplicar o critério do módulo constante em dois casos. Primeiramente consideraremos um “canal” que consiste apenas de um fator ganho real e, em seguida, assumiremos um modelo MA(1). Embora demasiado simples em relação aos casos práticos de equalização de canais discretos, tais modelos permitem que se tenha uma boa noção dos tipos possíveis de comportamento presentes na adaptação de equalizadores autodidatas em sistemas

como o da Fig.1. É bom ressaltar que a não-linearidade inerente ao método abre um grande “leque” de possibilidades para o sistema em questão.

Em ambos os casos, o equalizador tem a mesma ordem do canal e as amostras do sinal são consideradas i.i.d., valendo $+A$ e $-A$ com igual probabilidade.

3.1 Caso simples: canal MA(0)

Neste modelo, o efeito do canal resume-se a uma alteração na potência do sinal transmitido. Considera-se, portanto, o canal como um ganho real de valor K . Para que o equalizador seja o inverso deste canal, o valor de $w(n)$ deve ser igual a $\pm 1/K$. O equalizador irá de fato atuar como um CAG (Controle Automático de Ganho) cego. Trata-se de um caso bastante simples, sem grandes implicações práticas, cuja grande vantagem é a viabilidade de se obter uma expressão recursiva totalmente determinística para o parâmetro $w(n)$, permitindo assim uma determinação analítica dos pontos de equilíbrio e de um limitante de operação satisfatória, expresso em termos do passo de adaptação.

3.1.1 Análise da expressão iterativa

A partir da Fig. 1, podemos escrever:

$$y(n) = w(n) \cdot x(n) = K \cdot w(n) \cdot s(n) \quad (7)$$

A aplicação de (5) nos leva à seguinte expressão para $w(n)$:

$$w(n+1) = w(n) + \mu \cdot K^2 \cdot A^4 \cdot w(n) \cdot [1 - K^2 \cdot w^2(n)] \quad (8)$$

A equação (8) encaixa-se na forma geral $w(n+1) = F[w(n)]$, que caracteriza sistemas dinâmicos unidimensionais. O sistema é não-linear, pois a função F mostra esta dependência em relação à variável $w(n)$. Sistemas dinâmicos não-lineares da forma de (8) são candidatos naturais à ocorrência de caos. Como primeiro passo de nossa análise, iremos determinar os pontos de equilíbrio de nosso sistema e a sua estabilidade em função do passo de adaptação μ . Isto irá fornecer informação sobre as possibilidades de comportamento de nosso equalizador para diferentes valores do parâmetro livre. Tais pontos satisfazem a equação:

$$w_e = F(w_e) \quad (9)$$

De (8), conclui-se facilmente que os pontos de equilíbrio de nosso sistema são $w_{e1} = 0$ e $w_{e2,3} = \pm 1/|K|$. Para verificarmos, como comentamos acima, a influência do parâmetro livre μ no comportamento do sistema, observemos a primeira derivada de $F[w(n)]$ [5]:

$$F'[w(n)] = 1 + \mu \cdot K^2 \cdot A^4 - 3\mu K^4 \cdot A^4 \cdot w^2(n) \quad (10)$$

Aplicando (10) aos pontos de equilíbrio, temos:

$$F'(w_{e1}) = 1 + \mu \cdot K^2 \cdot A^4 \quad (11)$$

$$F'(w_{e2,3}) = 1 - 2 \cdot \mu \cdot K^2 \cdot A^4 \quad (12)$$

É sabido que a capacidade de atração da vizinhança de um ponto de equilíbrio só existe se $|F'(w_e)| < 1$. Concluímos então que o primeiro

ponto não será estável para nenhum valor de $\mu > 0$. Já os outros dois o serão apenas para:

$$0 < \mu < \frac{1}{K^2 \cdot A^4} \quad (13)$$

Podemos entender (13) como a condição sobre o passo de adaptação para que haja performance satisfatória do equalizador, ou seja, convergência para w_{e2} ou w_{e3} . Quando μ ultrapassa o valor limite, ocorre uma mudança substancial na dinâmica do sistema, pois cessa a convergência para os pontos de equilíbrio. A tal mudança dá-se o nome de *bifurcação*.

Uma forma de se tentar observar o que ocorre é através de um diagrama de bifurcação, que indica o comportamento do sistema após o fim dos transitórios, para diversos valores de μ . Empregamos, na análise subsequente, os valores $K=2$ e $A=1$. Nas simulações, foi usado $w(0) = 0.1$. Na Fig. 2, foram rodadas 1000 iterações de (8), e representados os últimos 50 valores para cada μ (tomado aqui de 0.0001 a 0.5496, a intervalos uniformes de 0.0005).

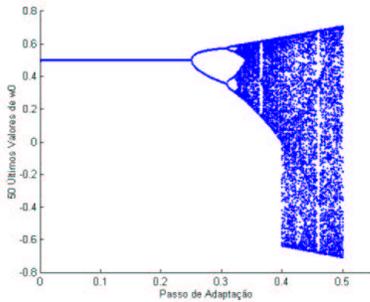


Figura 2. Diagrama de Bifurcação

Podemos então verificar os diferentes comportamentos apresentados pelo nosso CAG, para diferentes valores de μ . A primeira região é a de convergência para o ponto de equilíbrio $1/K$ (converge-se para este ponto e não para seu oposto por causa da inicialização), como já mostramos em (13). Quando o valor crítico de μ é ultrapassado, o sistema não mais converge para um ponto de equilíbrio, mas passa a oscilar entre dois valores, em um ciclo-limite estável. Aumentando ainda mais μ , ocorrem ciclos de período 4, 8, 16, até que chega-se a um ponto de acumulação, a partir do qual ocorre caos (para μ próximo de 0.32). Porém, mesmo nesta região, há “janelas” de periodicidade, onde ciclos-limite voltam a ser estáveis. Após a região de caos, há divergência para o infinito.

Outra característica notável é a mudança significativa na estrutura do atrator estranho, para um valor de μ próximo de 0.4. Associamos isto à ocorrência de uma *crise* [8] no atrator, ocorrida quando os valores de $w(n)$ começam a ser negativos. Podemos atribuir esta crise à “colisão” entre o ponto de equilíbrio instável $w_{e1} = 0$ e o atrator estranho. A região caótica inferior, que aparece no diagrama apenas após a crise, é “proveniente” da “evolução” do outro ponto de equilíbrio $(-1/K)$.

Por fim, foi calculado o expoente de Lyapunov deste sistema, para os valores de μ do diagrama. No cálculo do expoente, foram utilizadas 5000 iterações para cada passo de adaptação. Este expoente, no nosso sistema, assumirá valores negativos nas regiões de convergência para pontos de equilíbrio ou ciclos-limite, será nulo

nas bifurcações e positivo na região de caos. A Fig. 3 mostra os resultados obtidos e permite verificar matematicamente o que se comentou no parágrafo anterior.

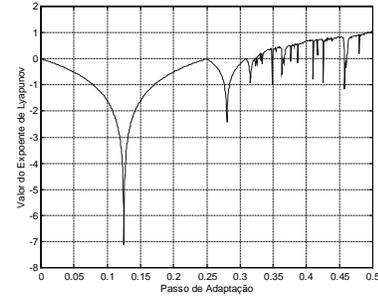


Figura 3. Expoente de Lyapunov

3.2 Modelo de canal MA(1)

Tendo caracterizada a existência de fenômenos caóticos através do exemplo anterior, podemos estender nosso estudo a um caso de maior interesse prático que é um canal FIR de primeira ordem, ou modelo MA(1). Normalizando-o em relação ao primeiro coeficiente, podemos escrever para o sistema da Fig. 1:

$$x(n) = s(n) + b.s(n-1) \quad (14)$$

$$y(n) = w_0(n).x(n) + w_1(n).x(n-1) \quad (15)$$

onde, novamente, consideraremos $A = 1$.

3.2.1 Análise estática

De posse do modelo, obter a expressão de $J(n)$ torna-se um trabalho matemático. Isto porém só é prático, via de regra, num processo de análise, pois exige conhecimento acerca do canal, por definição não disponível.

A partir de (1), obtemos:

$$J(w_0, w_1) = 1 + \theta_1 \cdot w_0^4 + 4 \cdot \theta_2 \cdot w_0^3 \cdot w_1 + 6 \cdot \theta_4 \cdot w_0^2 \cdot w_1^2 + 4 \cdot \theta_3 \cdot w_0 \cdot w_1^3 + \theta_1 \cdot w_1^4 - 2 \cdot \theta_5 \cdot w_0^2 - 4 \cdot \theta_6 \cdot w_0 \cdot w_1 - 2 \cdot \theta_5 \cdot w_1^2 \quad (16)$$

Onde os valores θ_i expressam diferentes momentos ou momentos cruzados de $x(n)$ e $x(n-1)$, sendo funções do zero do canal.

O gradiente de tal superfície é composto por:

$$\frac{\partial J}{\partial w_0} = 4 \cdot \theta_1 \cdot w_0^3 + 12 \cdot \theta_2 \cdot w_0^2 \cdot w_1 + 12 \cdot \theta_4 \cdot w_0 \cdot w_1^2 + 4 \cdot \theta_3 \cdot w_1^3 - 4 \cdot \theta_5 \cdot w_0 - 4 \cdot \theta_6 \cdot w_1 \quad (17)$$

$$\frac{\partial J}{\partial w_1} = 4 \cdot \theta_2 \cdot w_0^3 + 12 \cdot \theta_4 \cdot w_0^2 \cdot w_1 + 12 \cdot \theta_3 \cdot w_0 \cdot w_1^2 + 4 \cdot \theta_1 \cdot w_1^3 - 4 \cdot \theta_6 \cdot w_0 - 4 \cdot \theta_5 \cdot w_1 \quad (18)$$

Uma vez determinado o gradiente, podemos buscar o ponto de operação de nosso equalizador através de (4). Os pontos de equilíbrio deste sistema satisfazem:

$$\nabla J[\mathbf{w}(n)] = 0 \quad (19)$$

Em nossa análise subsequente, adotaremos $b = 0.6$. Para este valor, foram obtidas as curvas de nível de $J(n)$, que estão na Fig. 4. Como podemos perceber, $J(n)$ não é convexa, apresentando como pontos que satisfazem (19): dois mínimos globais, dois mínimos locais, quatro pontos de sela e um máximo, totalizando 9 pontos “significativos”. Para o nosso exemplo específico, os pontos citados acima constituem, respectivamente, o conjunto :

$$P = \{ [0.831 \ -0.372], [-0.831 \ 0.372] [0.084 \ 0.607], [-0.084 \ -0.607], [0.450 \ 0.241], [-0.382 \ 0.613], [-0.450 \ -0.241], [0.382, -0.613], [0 \ 0] \} \quad (20)$$

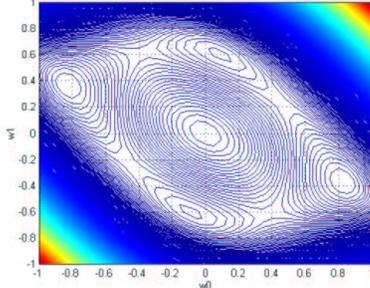


Figura 4. Curvas de Nível de $J(n)$

Como se pode perceber, sempre que um ponto satisfaz (19), o seu simétrico em relação à origem também o faz. Para se obter os pontos acima, foi preciso solucionar as equações $\partial J/\partial w_0 = 0$ e $\partial J/\partial w_1 = 0$. Observando (17) e (18), podemos perceber que a condição de gradiente nulo corresponde à intersecção de duas curvas no plano $w_0 \times w_1$, uma vez que cada elemento da equação vetorial (19) é uma relação implícita entre os dois taps do equalizador. Por métodos numéricos, conseguimos traçar as duas curvas no plano, e obtivemos a Fig. 5. As intersecções entre as curvas correspondem exatamente aos pontos de (20), o que confirma nossos comentários.

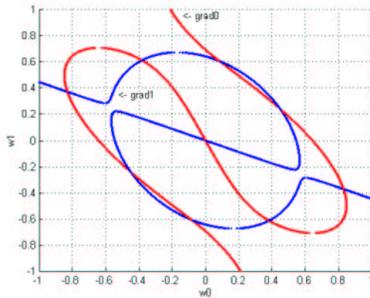


Figura 5. Curvas no Plano $w_0 \times w_1$ (grad0 corresponde à condição $\partial J/\partial w_0 = 0$ e grad1 à condição $\partial J/\partial w_1 = 0$)

3.2.1 Análise dinâmica dos mínimos

Passemos agora à análise dos pontos de equilíbrio do algoritmo. Analisaremos apenas um ponto de mínimo global e um ponto de mínimo local (os outros dois pontos de mínimo são simétricos aos primeiros, o que torna desnecessária a sua análise). O máximo não é estável para $\mu > 0$, e os pontos de sela não o são para nenhum valor de μ .

Na análise dinâmica realizada nesta seção, foi considerado o algoritmo sem aproximação estocástica para o gradiente, conforme a equação (4). O motivo desta escolha é que esta expressão não

contém parâmetros aleatórios, o que permite o uso de ferramentas padrão de análise. Consideramos que, mesmo sem estarmos lidando diretamente com (5), a análise da expressão determinística permite inferir algumas características médias de convergência do CMA.

Tomemos então o mínimo global $w_G = [0.831 \ -0.372]$. Por se tratar de um sistema de dimensão maior que um, a análise de estabilidade está relacionada à matriz de Floquet do sistema. Mais especificamente, é preciso observar a posição dos autovalores, também conhecidos como multiplicadores de Floquet, desta matriz, calculada no ponto analisado. Caso ambos estejam dentro do círculo de raio unitário (CRU), o ponto em questão irá atrair as trajetórias de sua vizinhança.

A figura a seguir mostra o valor do máximo módulo dos autovalores da matriz de Floquet, $\max(|\lambda_1|, |\lambda_2|)$, em função do passo de adaptação μ .

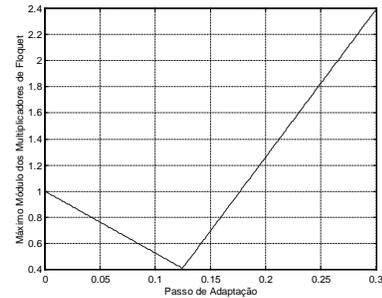


Figura 6. Valor Máximo dentre os Módulos dos Multiplicadores de Floquet (condição de mínimo global)

A partir da Fig. 6, podemos perceber que ambos autovalores estarão dentro do CRU para $0 < \mu < 0.1765$. Isto significa que o ponto analisado será estável para esta faixa de valores do passo de adaptação. Novamente, para percebermos o que ocorre para outros valores de μ , é interessante traçar um diagrama de bifurcação. Para tanto, escolhemos o peso w_0 . Foi escolhida uma condição inicial que assegurasse convergência global. O valor de μ variou de 0 a 0.3, com um passo de 0.001. Foram rodadas 1000 iterações. A Fig. 7 mostra o diagrama. Podemos observar o mesmo cenário do modelo anterior, com uma cascata de infinitas bifurcações de duplicação de período e uma faixa de caos (com janelas de periodicidade). Após esta faixa, há divergência para o infinito.

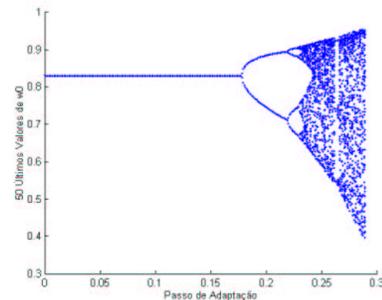


Figura 7. Diagrama de Bifurcação para w_0

Por fim, foi calculado o maior expoente de Lyapunov do sistema. Para isto, foi utilizada a condição $w(0) = [0.831 \ -0.372]^T$ e foram rodadas 5000 iterações para cada valor de μ (tomado de 0 a 0.3, com

passo de 0.001). A Fig. 8 traz a relação entre o máximo expoente de Lyapunov e o passo de adaptação.

Podemos observar uma região onde o maior expoente é nulo ou negativo, correspondendo a um comportamento estacionário ou periódico, e uma faixa para o qual tal expoente é positivo, caracterizando comportamento caótico. Novamente, na região de caos há infinitas “janelas de periodicidade”, nas quais o sistema volta a oscilar de forma ordenada. Quando o sistema apresenta divergência para o infinito, o expoente também diverge.

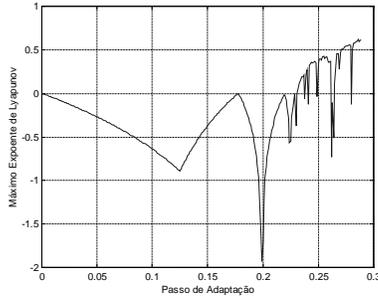


Figura 8. Maior Expoente de Lyapunov (condição de mínimo global)

Refaçamos agora a análise para o ponto de mínimo local dado por $w_L = [0.084, 0.607]^T$. Na Fig. 9, foi traçado o máximo módulo dos autovalores da matriz de Floquet. Nota-se que tal ponto será estável para $0 < \mu < 0.1400$ pois, fora desta faixa, pelo menos um dos autovalores estará fora do CRU. É interessante observar que neste caso o mínimo local mantém-se estável para uma faixa de valores menor que no caso do mínimo global.

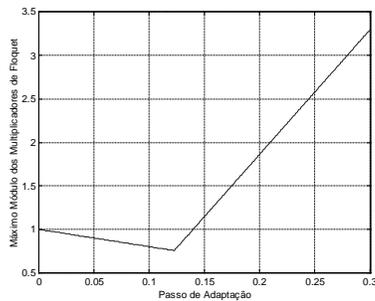


Figura 9. Valor Máximo dentre os Módulos dos Multiplicadores de Floquet (condição de mínimo local)

A Fig. 10 mostra o diagrama de bifurcação (de w_0) associado ao mínimo local, obtido de forma análoga ao relativo ao mínimo global. Para a obtenção deste diagrama, foi utilizada uma condição inicial que levou o sistema a convergir para um mínimo local.

Para μ próximo de 0.22, há uma “descontinuidade” no diagrama, quando o atrator estranho associado ao mínimo local perde sua estabilidade, sendo que o sistema passa então a apresentar uma evolução idêntica a de um mínimo global (pela prevalência da estabilidade desta região).

Observando o máximo expoente de Lyapunov associado ao sistema, mostrado na Fig. 11, verificamos que o mesmo começa a ser maior

do que zero em um valor de μ próximo de 0.18, mas depois sofre uma mudança abrupta. Tal mudança corresponde à perda de estabilidade do atrator estranho, comentada anteriormente.

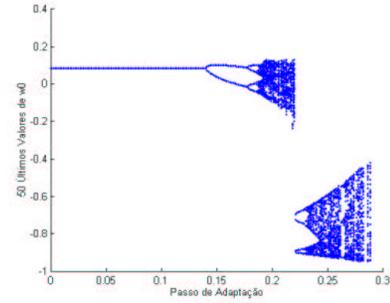


Figura 10. Diagrama de Bifurcação para w_0

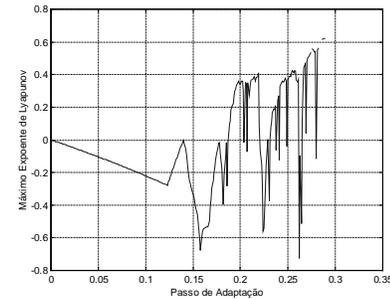


Figura 11. Máximo Expoente de Lyapunov (condição de mínimo local)

4. UMA PROPOSTA PARA A OBTENÇÃO DE CONVERGÊNCIA GLOBAL

A partir dos resultados apresentados neste trabalho, surgiu uma possibilidade de aplicação prática das características dinâmicas inerentes à técnica de equalização pelo critério do módulo constante. Esta possibilidade foi inspirada por trabalhos de otimização baseada em caos [13,14]. Podemos dividir a técnica em duas partes: uma fase de varredura e uma fase de diminuição no valor do passo de adaptação.

O princípio por trás da primeira fase é intuitivo: testar diversos valores de w e verificar, segundo um certo critério, qual é o melhor deles. Caso estejamos no caso determinístico, o critério pode ser a própria função custo $J(n)$. Caso estejamos operando com o CMA propriamente dito, é preciso obter um outro critério que aproxime o papel desempenhado pela função custo. Chamemos este critério escolhido de G .

Quando o sistema opera na zona caótica, ele pode percorrer uma ampla região do espaço de estados, como podemos perceber pelas Figs. 2, 7 e 10. Munidos de nosso critério G , podemos usar esta “varredura caótica” para buscar um valor de w tão próximo quanto possível do mínimo global, simplesmente comparando, através de nosso critério, os diversos valores varridos.

Este método é limitado pela faixa percorrida, pois caso a mesma não englobe valores considerados bons, os resultados não devem ser satisfatórios.

A segunda fase do método consiste em escolher um valor de μ para o qual o sistema opere em caos. Em seguida, reduz-se gradativamente este valor, até que o sistema convirja para um ponto de equilíbrio que, pela diminuição do passo, volta a ser estável. Tal ponto é então avaliado segundo G e comparado com o ótimo da primeira fase.

A Fig. 10 ajuda a compreender a lógica desta fase. Podemos ver que a região caótica do mínimo local perde a estabilidade para um valor de μ menor do que no caso do mínimo global. Se o valor do passo for tal que esteja na região caótica associada ao mínimo global, quando o mesmo for diminuído, o sistema *não irá convergir para o mínimo local*, como seria se o diagrama fosse seguido, mas para o global. Portanto, para canais como o apresentado, a diminuição do valor de μ a partir de um valor adequado leva a um valor próximo do ótimo. Porém, como foi comentado, a segunda fase será eficaz em cenários como o apresentado aqui, que não são regra para todos os canais.

Para inicialização conveniente na região caótica, pode ser usado o maior valor de passo para o qual não há divergência para o infinito. Tal escolha parece adequada para ambas as fases.

Uma vez obtida uma estimativa de w por este método, tal valor pode ser usado como condição inicial do algoritmo adaptativo, visando refinar esta estimativa. Assim, deve haver uma melhoria na taxa de convergência global em relação a uma escolha aleatória.

5. CONCLUSÕES

Neste trabalho, procurou-se analisar a expressão iterativa de atualização dos parâmetros de um equalizador autodidata como um sistema dinâmico. A não-linearidade inerente ao critério empregado dá origem a ciclos-limite e a comportamentos caóticos, pouco investigados na análise de algoritmos adaptativos, que usualmente se restringe a verificar a convergência ou não para um ponto de mínimo da função custo.

Nosso estudo consistiu em duas partes básicas: uma análise estática, ligada à localização dos pontos de equilíbrio do sistema, e uma análise dinâmica, que relacionou diferentes comportamentos ao valor do parâmetro livre (no caso o passo de adaptação μ). As observações foram suportadas matematicamente por técnicas usuais de análise de sistemas determinísticos e confirmaram a suposição inicial do trabalho de que algoritmos de Bussgang podem operar em regime caótico.

Os resultados obtidos levaram ainda à proposta preliminar de um método que pode aumentar as taxas de convergência global do algoritmo CMA.

Os próximos passos deste trabalho são a busca de uma análise mais refinada para o caso de canais e equalizadores de qualquer ordem, na busca de soluções para os pontos significativos da função custo e de uma melhor caracterização dos fenômenos observados no CMA. A partir de uma análise estática mais detalhada, poderemos ter mais informações sobre a dinâmica do sistema. Um outro passo importante é um aprofundamento em métodos de análise dinâmica estocástica, que se apresenta como a ferramenta mais adequada para o caso de filtros de ordens mais elevadas. Por fim, é preciso estabelecer uma análise mais detalhada do desempenho e da aplicabilidade do método de convergência global, cujos princípios foram aqui propostos.

6. BIBLIOGRAFIA

- [1] Ding Z., Kennedy R. A., Anderson B. D. O., Johnson C.R., "Ill-convergence of Godard blind equalizers in data communications Systems", IEEE Trans. on Comm., Vol. 39, pp.1313-27, 1991.
- [2] Ding Z., Kennedy R. A., Johnson C. R., "On the (non)existence of undesirable equilibria of Godard blind equalizer", IEEE Trans. on Sig. Processing, vol. 40, pp. 2425-32, 1992.
- [3] Ding Z., Johnson C.R., "On the nonvanishing stability of undesired equilibria for FIR Godard blind equalizers", IEEE Trans. on Sig. Proc., Vol. 41, pp.1940-44, 1993.
- [4] Ding Z., Kennedy R. A., "On the whereabouts of local minima for blind adaptive equalizers", IEEE Trans. on Circ. Syst. II, Vol. 39, pp. 119-23, 1992.
- [5] Fiedler-Ferrara N., Cintra do Prado C.P., Caos: uma Introdução, Primeira reimpressão, Edgard Blücher, 1999.
- [6] Feigenbaum M., "Universal behavior in nonlinear systems", Physica 7D, pp. 16-39, 1983.
- [7] Godard D.N., "Self-recovering equalization and carrier tracking in two-dimensional data communication systems", IEEE Trans. on Comm., vol. COM-28, pp. 1867-75, 1980.
- [8] Grebogi C., Ott E., "Crises, sudden changes in strange attractors, and transient chaos", Physica 7D, pp. 181-200, 1983.
- [9] Haykin S., Adaptive Filter Theory, Third Edition, Prentice Hall, 1996.
- [10] Li Y., Ding Z., "Convergence analysis of finite length blind equalizers", IEEE Trans. on Sig. Proc., Vol. 41, pp. 2120-28, 1995.
- [11] Li Y., Ray Liu K.J., "Static and dynamic convergence behavior of adaptive blind equalizers", IEEE Trans. on Sig. Proc., Vol. 44, pp. 2736-45, 1996.
- [12] Schuster H. G., Deterministic Chaos: An Introduction, First reprint of second revised edition, VCH, 1989.
- [13] Tokuda I., Onodera K., Tokunaga R., Aihara K., Nagashima T., "Global bifurcation scenario for chaotic dynamical systems that solve optimization problems and analysis of their optimization capability", Electronics and Communications in Japan, Part 3, Vol. 81, No. 2, 1998.
- [14] Zhou C., Chen T., "Chaotic annealing for optimization", Physical Review E, Vol. 55, pp. 2580-87, 1997.

Agradecimentos : Gostaríamos de agradecer à CAPES que financia parcialmente esta pesquisa, a Márcio Eisenkraft, da Escola Politécnica da USP, pelas valiosas discussões e informações sobre dinâmica caótica, e a Cristiano Panazio, Helder Bertan e Alexandre Osorio, pelo apoio técnico.