

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

**NOVOS PARADIGMAS PARA EQUALIZAÇÃO E  
IDENTIFICAÇÃO DE CANAIS BASEADOS EM  
ESTRUTURAS NÃO-LINEARES E ALGORITMOS  
EVOLUTIVOS**

Autor

**Romis Ribeiro de Faissol Attux**

Orientador

**João Marcos Travassos Romano**

Banca Examinadora:

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

Prof. Dr. João César Moura Mota (DTI/UFC)

Prof. Dr. Rui Seara (DEE/UFSC)

Prof. Dr. Amauri Lopes (FEEC/UNICAMP)

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

Prof. Dr. Wagner Caradori do Amaral (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 Doutor em Engenharia Elétrica.

Campinas, 26 de abril de 2005

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

At87n            Attux, Romis Ribeiro de Faissol  
                    Novos paradigmas para equalização e identificação de  
                    canais baseados em estruturas não-lineares e algoritmos  
                    evolutivos / Romis Ribeiro de Faissol Attux. --Campinas,  
                    SP: [s.n.], 2005.

                    Orientador: João Marcos Travassos Romano.  
                    Tese (doutorado) - Universidade Estadual de Campinas,  
                    Faculdade de Engenharia Elétrica e de Computação.

                    1. Filtros digitais (Matemática). 2. Identificação de  
                    sistemas. 3. Redes neurais (Computação). 4. Sistemas  
                    difusos. 5. Algoritmos genéticos. 6. Volterra, Séries de. I.  
                    Romano, João Marcos Travassos. II. Universidade Estadual  
                    de Campinas. Faculdade de Engenharia Elétrica e de  
                    Computação. III. Título.

Titulo em Inglês: New paradigms for channel equalization and  
identification based on nonlinear structures and  
evolutionary algorithms

Palavras-chave em Inglês: Digital filters (mathematics), System  
identification, Neural networks (computer  
science), Fuzzy systems, Genetic algorithms e  
Volterra Series

Área de concentração: Telecomunicações e Telemática

Titulação: Doutor em Engenharia Elétrica

Banca examinadora: João César Mota, Rui Seara, Amauri Lopes, Fernando  
José Von Zuben e Wagner Caradori do Amaral

Data da defesa: 26/04/2005

## Resumo

O objetivo deste trabalho é investigar a aplicação de estruturas não-lineares e de técnicas de otimização baseadas em computação evolutiva a problemas de equalização e identificação de canal. O relato se divide em duas partes: a primeira voltada à análise dos fundamentos do problema de filtragem, e a segunda, à apresentação de novas abordagens para sua solução. A primeira parte, inaugurada pelas noções primordiais de comunicação, abrange os diferentes aspectos do projeto de um filtro. Permeia toda a exposição uma idéia fundamental: o estabelecimento de um paradigma genérico de filtragem adaptativa. Na segunda parte, apresentamos contribuições originais que se encaixam de diversas formas no espírito desse paradigma. Os problemas abordados são: equalização linear cega, equalização e pré-distorção baseadas em redes neurais, identificação cega, identificação de plantas recursivas, busca cega do receptor de máxima verossimilhança e equalização não-linear cega baseada em predição. Tais propostas, além de possuírem um valor intrínseco, podem ser entendidas como um *corpus* de evidências da validade das idéias unificadoras pertencentes ao arcabouço teórico erigido.

## Abstract

The objective of this work is to investigate the use of nonlinear structures and optimization techniques based on evolutionary computation in channel equalization and identification problems. The text is structured according to a twofold division: the first part is dedicated to the analysis of the foundations of the filtering problem, and the second, to the exposition of new approaches to its solution. The first part, whose starting point is the primordial idea of communication, encompasses the various aspects of the design of a signal-processing device. The entire explanation is permeated by a fundamental idea: the constitution of a generic paradigm for adaptive filtering. In the second part, a number of original contributions are presented that are related in various ways to the essence of this paradigm. These contributions involve the problems of blind linear equalization, neural network-based equalization and pre-distortion, blind and supervised identification of feedforward and feedback plants, maximum-likelihood sequence estimation and nonlinear blind equalization based on prediction. Such proposals, apart from possessing an intrinsic value, can also be understood as a *corpus* of evidences of the validity of the unifying ideas belonging to the established theoretical framework.

*“Duas coisas há que sobretudo contribuem para nos dar força de ânimo: a fé na verdade, a confiança em nós mesmos.”*

Lúcio Aneu Sêneca, Cartas a Lucílio, 94, 46<sup>1</sup>

---

<sup>1</sup> Da versão publicada pela Fundação Calouste Gulbenkian.

*A Dilmara, Lester, Dina, Romis e Cecília.*

*Aos que partiram, Alice e Naby, e a quem vier.*

## **Agradecimentos**

Gostaria de manifestar minha gratidão

Ao meu orientador, Prof. João Marcos Romano, pelo apoio e pela compreensão em todos os momentos.

À FAPESP, pelo apoio financeiro.

À minha esposa, Dilmara, ao meu filho canino, Lester, e à minha sogra, Jecy.

Aos meus pais, Romis e Dina, à minha irmã, Cecília, e a todos os meus familiares.

Aos meus amigos do Laboratório de Processamento de Sinais para Comunicações (DSPCom).

Aos membros da banca, pelas valiosas sugestões.

Aos Profs. Fernando Von Zuben, Leandro de Castro e Maurice Bellanger, pela proveitosa cooperação científica.

Ao amigo Cristiano Cruz, especialmente pelo apoio no curso que juntos ministramos.

A todos os meus amigos.

Aos amigos da turma EE95.

Aos amigos da querida “turma da pós-graduação”.

Aos amigos da FEEC em geral e da representação discente.

Aos alunos do curso de EE500.

Às amigas Celi, Eloísa, Giane, Gislaine, Jaqueline, Jerusa, Juraci, Licinha, Lúcia, Mazé e Noêmia.

A todos que me apoiaram desde que cheguei a Campinas.

A todos que puderam comparecer à defesa de tese.

A todos os funcionários da FEEC.

## Lista de Abreviações e Acrônimos

AG: Algoritmo Genético  
AM: Modulação em Amplitude  
ASD: Algoritmo *Steepest Descent*  
ASK: *Amplitude-Shift Keying*  
AWGN: Ruído Aditivo, Branco e Gaussiano (*Additive White Gaussian Noise*)  
BER: Taxa de Erro de Bit (*Bit Error Rate*)  
BGR: Benveniste-Goursat-Ruget  
BP: *Back-Propagation*  
CM: Módulo Constante (*Constant Modulus*)  
CMV: Critério de Máxima Verossimilhança  
CRU: Círculo de Raio Unitário  
DD: Decisão Direta (a rigor, *Decision-Directed*)  
DEP: Densidade Espectral de Potência  
DFE: *Decision-feedback Equalizer*  
DPD: Dispositivo de Pré-distorção  
EEF: *Equation-error Formulation*  
EQM: Erro Quadrático Médio  
FEP: Filtro de Erro de Predição  
FIR: *Finite Impulse Response*  
FM: Modulação em Frequência  
FSK: *Frequency-Shift Keying*  
IIS: Interferência Intersimbólica  
IIR: *Infinite Impulse Response*  
LMS: *Least Mean Square*  
MLP: *Multilayer Perceptron*  
MV: Máxima Verossimilhança  
OEF: *Output-error Formulation*  
PAM: Modulação por Amplitude de Pulso  
PD: Pré-distorção  
PSK: *Phase-Shift Keying*  
QAM: *Quadrature Amplitude Modulation*  
RBF: Função de Base Radial (*Radial-basis Function*)  
RLS: *Recursive Least Squares*  
RN: Rede Neural  
SER: Taxa de Erro de Símbolo (*Symbol Error Rate*)  
SNR: Relação Sinal-ruído (*Signal-to-noise Ratio*)  
SISO: *Single-Input / Single-Output*  
SW: Shalvi-Weinstein  
ZF: *Zero-forcing*

<b>1. INTRODUÇÃO</b>	<b>1</b>
<hr/>	
<b>2. SISTEMAS DE COMUNICAÇÃO DIGITAL</b>	<b>7</b>
<hr/>	
2.1. INTRODUÇÃO	7
2.2. UMA VISÃO GERAL DOS ELEMENTOS DE UM SISTEMA DE COMUNICAÇÃO	8
2.3. MODERNOS SISTEMAS DE COMUNICAÇÃO	9
2.4. SISTEMAS ANALÓGICOS	9
2.5. SISTEMAS DIGITAIS: O TEOREMA DA AMOSTRAGEM	11
2.6. SISTEMAS DIGITAIS: QUANTIZAÇÃO	12
2.7. SISTEMAS DIGITAIS: CODIFICAÇÃO	13
2.8. SISTEMAS DIGITAIS: MODULAÇÃO	14
2.8.1. MODULAÇÃO POR AMPLITUDE DE PULSO (PAM):	14
2.8.2. MODULAÇÕES EM BANDA PASSANTE	16
2.9. SISTEMAS DIGITAIS: CONSIDERAÇÕES SOBRE O CANAL DE COMUNICAÇÃO	18
2.9.1. INTERFERÊNCIA INTERSIMBÓLICA (IIS)	18
2.9.2. RUÍDO	18
2.10. SISTEMAS DIGITAIS: DETECÇÃO ÓTIMA DE SINAIS EM RUÍDO BRANCO	25
2.11. MODELO DISCRETO DE CANAL	30
2.11.1. SINAL TRANSMITIDO	31
2.11.2. CANAL	32
2.12. CONCLUSÕES E MODELO BÁSICO DE CANAL	37
<b>3. EQUALIZAÇÃO, IDENTIFICAÇÃO E PREDIÇÃO</b>	<b>39</b>
<hr/>	
3.1. INTRODUÇÃO	39
3.2. INTRODUÇÃO ÀS ESTRUTURAS DE FILTRAGEM	40
3.2.1. ESTRUTURAS LINEARES	41
3.2.2. ESTRUTURAS NÃO-LINEARES	44
3.3. ESCOLHA DOS PARÂMETROS	45
3.3.1. ETAPA 1: ESCOLHA DO CRITÉRIO	46
3.3.2. ETAPA 2: OBTENÇÃO DOS PARÂMETROS	46
3.3.3. ETAPA 3: ADAPTAÇÃO DOS PARÂMETROS	47
3.4. EQUALIZAÇÃO	48
3.4.1. EQUALIZAÇÃO VISTA COMO INVERSÃO DO CANAL – O CRITÉRIO <i>ZERO-FORCING</i> (ZF)	48
3.4.2. EQUALIZAÇÃO A PARTIR DE UMA MEDIDA DE ERRO: O CRITÉRIO DE WIENER	51
3.4.3. EQUALIZAÇÃO A PARTIR DE ESTATÍSTICAS DE ORDEM SUPERIOR: CRITÉRIOS NÃO-SUPERVISIONADOS	60
3.4.4. EQUALIZAÇÃO VISTA COMO UM PROBLEMA DE CLASSIFICAÇÃO DE PADRÕES	69
3.4.5. FINALE: ALGUNS COMENTÁRIOS SOBRE O EQUALIZADOR DE MÁXIMA VEROSSIMILHANÇA	80
3.5. IDENTIFICAÇÃO	80
3.5.1. ESCOLHA DA ESTRUTURA	81
3.5.2. IDENTIFICAÇÃO SUPERVISIONADA	81

3.5.3.	IDENTIFICAÇÃO NÃO-SUPERVISIONADA E <i>CUMULANT MATCHING</i>	84
3.5.4.	DISCUSSÃO	86
<b>3.6.</b>	<b>PREDIÇÃO</b>	<b>87</b>
3.6.1.	PREDIÇÃO E EQUALIZAÇÃO	88
<b>4.</b>	<b><u>ESTRUTURAS NÃO-LINEARES DE FILTRAGEM</u></b>	<b>91</b>
<b>4.1.</b>	<b>REDES NEURAIS</b>	<b>94</b>
<b>4.2.</b>	<b>REDES MLP</b>	<b>96</b>
4.2.1.	CRITÉRIO DE MÍNIMO ERRO QUADRÁTICO PARA UMA MLP	99
4.2.2.	O ALGORITMO DE BACK-PROPAGATION	100
<b>4.3.</b>	<b>REDES DE FUNÇÕES DE BASE RADIAL (RBF)</b>	<b>102</b>
4.3.1.	ESCOLHA DOS CENTROS – CLUSTERIZAÇÃO	103
4.3.2.	ESCOLHA DAS VARIÂNCIAS E DOS PESOS DA CAMADA DE SAÍDA	110
4.3.3.	ALGUNS COMENTÁRIOS SOBRE A ESTRUTURA	110
4.3.4.	REDES RBF E O EQUALIZADOR DE BAYES	111
<b>4.4.</b>	<b>FILTROS FUZZY (NEBULOSOS)</b>	<b>112</b>
4.4.1.	CONJUNTOS E REGRAS	114
4.4.2.	EQUIVALÊNCIA ENTRE O FILTRO FUZZY, A REDE RBF E O EQUALIZADOR DE BAYES	119
<b>4.5.</b>	<b>FILTROS POLINOMIAIS OU DE VOLTERRA</b>	<b>119</b>
4.5.1.	FILTROS DE VOLTERRA E FILTROS LINEARES	123
<b>4.6.</b>	<b>ESTRUTURAS RECURSIVAS</b>	<b>123</b>
<b>4.7.</b>	<b>CONCLUSÕES E PERSPECTIVA HISTÓRICA</b>	<b>125</b>
<b>5.</b>	<b><u>OTIMIZAÇÃO BASEADA EM COMPUTAÇÃO EVOLUTIVA</u></b>	<b>129</b>
<b>5.1.</b>	<b>INTRODUÇÃO</b>	<b>129</b>
<b>5.2.</b>	<b>COMPUTAÇÃO EVOLUTIVA, OTIMIZAÇÃO E FILTRAGEM</b>	<b>131</b>
<b>5.3.</b>	<b>ALGORITMOS GENÉTICOS</b>	<b>132</b>
5.3.1.	CONSIDERAÇÕES SOBRE A FUNÇÃO CUSTO	134
5.3.2.	REPRESENTAÇÃO / CODIFICAÇÃO	134
5.3.3.	POPULAÇÃO INICIAL	135
5.3.4.	SELEÇÃO E REPRODUÇÃO	135
5.3.5.	CROSSOVER	137
5.3.6.	MUTAÇÃO	138
5.3.7.	SELEÇÃO FINAL E ELITISMO	139
5.3.8.	NICHING	140
5.3.9.	SUMÁRIO DE UM ALGORITMO GENÉTICO TÍPICO	141
<b>5.4.</b>	<b>SISTEMAS IMUNOLÓGICOS ARTIFICIAIS</b>	<b>143</b>
5.4.1.	CONCEITOS E MECANISMOS	144
5.4.2.	A REDE IMUNOLÓGICA ARTIFICIAL (OPT-AI <sup>NET</sup> )	146
<b>5.5.</b>	<b>COMPUTAÇÃO EVOLUTIVA COMO UM PARADIGMA PARA FILTRAGEM ÓTIMA</b>	<b>149</b>
<b>6.</b>	<b><u>PRÓLOGO À SEGUNDA PARTE</u></b>	<b>153</b>

<b><u>7. EQUALIZAÇÃO LINEAR NÃO-SUPERVISIONADA E COMPUTAÇÃO EVOLUTIVA</u></b>	<b>155</b>
<b>7.1. PRIMEIRA ABORDAGEM: ALGORITMO GENÉTICO COM NICHING</b>	<b>156</b>
7.1.1. PRIMEIRO ENSAIO	157
7.1.2. SEGUNDO ENSAIO	158
7.1.3. TERCEIRO ENSAIO	158
7.1.4. DISCUSSÃO	159
<b>7.2. SEGUNDA ABORDAGEM: REDE IMUNOLÓGICA</b>	<b>159</b>
7.2.1. PRIMEIRO ENSAIO	160
7.2.2. SEGUNDO ENSAIO	161
7.2.3. TERCEIRO ENSAIO	162
7.2.4. ANÁLISE DA DIVERSIDADE POPULACIONAL	163
7.2.5. DISCUSSÃO	164
<b><u>8. EQUALIZAÇÃO RECURSIVA NÃO-SUPERVISIONADA E COMPUTAÇÃO EVOLUTIVA</u></b>	<b>165</b>
<b>8.1. EQUALIZAÇÃO IIR CEGA</b>	<b>165</b>
8.1.1. PRIMEIRO ENSAIO	166
8.1.2. SEGUNDO ENSAIO	168
8.1.3. TERCEIRO ENSAIO	170
8.1.4. DISCUSSÃO	172
<b>8.2. PROJETO NÃO-SUPERVISIONADO DE UM DFE</b>	<b>173</b>
8.2.1. PRIMEIRO ENSAIO	174
8.2.2. SEGUNDO ENSAIO	176
8.2.3. DISCUSSÃO	177
<b><u>9. NOVAS PROPOSTAS PARA EQUALIZAÇÃO E PRÉ-DISTORÇÃO BASEADAS EM REDES NEURAI</u></b>	<b>179</b>
<b>9.1. EQUALIZAÇÃO SUPERVISIONADA DE UM CANAL NÃO-LINEAR</b>	<b>179</b>
<b>9.2. UMA NOVA PROPOSTA PARA PRÉ-DISTORÇÃO</b>	<b>181</b>
<b>9.3. ANÁLISE DE UMA MLP RECORRENTE</b>	<b>185</b>
9.3.1. PRIMEIRO ENSAIO	186
9.3.2. SEGUNDO ENSAIO	188
<b><u>10. NOVAS PROPOSTAS PARA IDENTIFICAÇÃO DE SISTEMAS BASEADAS EM COMPUTAÇÃO EVOLUTIVA</u></b>	<b>191</b>
<b>10.1. REDE IMUNOLÓGICA E CUMULANT MATCHING</b>	<b>192</b>
10.1.1. PRIMEIRO ENSAIO	193
10.1.2. SEGUNDO ENSAIO	194
10.1.3. CONCLUSÃO	195

<b>10.2. REDE IMUNOLÓGICA E IDENTIFICAÇÃO DE SISTEMAS RECURSIVOS</b>	<b>195</b>
10.2.1. PRIMEIRO ENSAIO	198
10.2.2. SEGUNDO ENSAIO	199
10.2.3. CONCLUSÃO	200
<b><u>11. EMPREGO DE ALGORITMOS GENÉTICOS NO PROJETO NÃO-SUPERVISIONADO DO RECEPTOR DE MÁXIMA VEROSSIMILHANÇA</u></b>	<b>201</b>
11.1. PRIMEIRO ENSAIO	205
11.2. SEGUNDO ENSAIO	206
11.3. DISCUSSÃO	207
<b><u>12. UMA PROPOSTA PARA EQUALIZAÇÃO NÃO-LINEAR CEGA BASEADA NUM FILTRO DE ERRO DE PREDIÇÃO FUZZY</u></b>	<b>209</b>
12.1. ASPECTOS GERAIS DE PROJETO	210
12.2. PRIMEIRO ENSAIO	211
12.3. SEGUNDO ENSAIO	213
12.4. CONCLUSÕES	214
<b><u>13. CONCLUSÃO</u></b>	<b>217</b>
<b><u>14. BIBLIOGRAFIA</u></b>	<b>219</b>
<b><u>A. APÊNDICE: TRABALHOS PUBLICADOS DURANTE O DOUTORADO</u></b>	<b>229</b>
A.1 ARTIGOS PUBLICADOS EM PERIÓDICOS DE CIRCULAÇÃO INTERNACIONAL	229
A.2 ARTIGOS PUBLICADOS EM ANAIS DE CONGRESSOS	229

# 1. Introdução

*“Qu'on ne dise pas que je n'ai rien dit de nouveau: la disposition des matières est nouvelle”<sup>2</sup>*

*Blaise Pascal, Pensamentos, I, 22*

O caráter gregário inerente a boa parte dos seres vivos demanda mecanismos capazes de tornar sua interação um processo organizado e verdadeiramente proveitoso. Para a espécie humana, em que tal caráter é sobremaneira pronunciado, a *comunicação* tem tido um papel de destaque desde tempos imemoriais, conforme atestam exemplos tão distintos quanto a onipresença da linguagem e as pinturas rupestres. Nesses e em muitos outros casos, subjazem refinados esquemas de codificação, transmissão e recepção, verdadeiros prodígios elaborados pela profícua mistura de pendor e necessidade. Junte-se a isso a contínua ampliação de horizontes tecnológicos e teremos uma simbiose que, ao menos parcialmente, explica a miríade de meios de comunicação acessíveis a parte da população mundial.

Embora esse “progresso” seja até certo ponto natural, ele tem seu grau característico de inércia: para adentrar novos territórios, é preciso construir as indispensáveis estradas e pontes conceituais e técnicas, ou seja, faz-se necessário que se pague o devido tributo à engenharia. Devem ser resolvidos incontáveis problemas de relevo para que uma possibilidade se torne uma efetiva aplicação, e foi exatamente esse o caso quando os avanços nas áreas de eletricidade e eletrônica terminaram por desembocar no avanço irresistível das comunicações digitais: sucederam-se novos resultados e teorias e surgiram campos do conhecimento sem os quais um dedo a pulsar jamais poderia ser sucedido por um oscilador de altíssima frequência.

Na evolução da moderna teoria de comunicação digital, sempre mereceram especial atenção os problemas provenientes de um aspecto inelutável: a limitação intrínseca do meio de transmissão. Todo canal por que trafega uma determinada mensagem tende a atuar sobre ela de acordo com suas características próprias, causando uma distorção que pode acarretar uma perda de desempenho muito significativa. Esse fato abre duas perspectivas imediatas ao engenheiro: procurar meios para entender e modelar os efeitos do canal, raiz do *problema de identificação*, e elaborar métodos para anular ou contrabalançar tais efeitos, razão de ser do problema de *equalização*. Tais perspectivas não tardaram a se transfigurar em formulações e abordagens concretas, as quais, em conjunto com um terceiro problema de relevo, o de *predição*, formam a base da *teoria de filtragem ótima*.

---

<sup>2</sup> Aproveitando a tradução de Sérgio Milliet, parte da versão publicada pela Ediouro: “Não se diga que eu nada disse de novo; a disposição da matéria é nova”. Essa frase abre o “Tratado de Semiótica” de Umberto Eco.

Numa formulação suficientemente ampla, o problema genérico de filtragem envolve três aspectos essenciais: escolha da estrutura, do critério de ajuste de seus parâmetros e estabelecimento de um método para busca da solução ótima associada a tal critério. O primeiro aspecto é dotado de uma riqueza que facilmente pode passar despercebida à primeira vista: temos em mãos nada menos que a tarefa de selecionar, dentre infinitas possibilidades, um dispositivo que seja capaz de produzir um mapeamento apropriado entre entrada e saída. Tal grau de generalidade abre um leque vastíssimo de caminhos que transcende o universo das comunicações e abrange campos multidisciplinares como o da inteligência computacional. O segundo aspecto, por sua vez, depende inevitavelmente dos objetivos do problema em mãos, embora possa suscitar formulações dependentes de resultados teóricos de áreas como a estatística ou reconhecimento de padrões. A terceira parte, como a primeira, é uma instância de uma tarefa de validade universal em engenharia: a de otimizar uma função custo. Por conseguinte, trata-se de uma etapa que tem muito a ganhar com uma visita a outras paragens, como, por exemplo, a computação evolutiva.

Essa divisão tripartite mostra de maneira, digamos, combinatória, quão complexo é o problema de filtragem adaptativa: é preciso escolher uma estrutura com suficiente potencial de emulação, um critério coerente e um método de otimização capaz de transformá-lo em real desempenho. Para o projetista, que precisa efetuar essas escolhas à luz de diversas restrições sistêmicas, é imperativo poder transitar com toda a liberdade e segurança por essas possibilidades. Para que isso ocorra, faz-se necessário *o estabelecimento de um paradigma sólido para filtragem ótima, ou seja, para o projeto genérico de um dispositivo processador de informação.*

O principal objetivo deste trabalho é analisar a viabilidade do estabelecimento de um paradigma desse tipo a partir do uso conjunto de critérios e estruturas genéricas e de ferramentas de otimização global baseadas em computação evolutiva. Para atingir tal meta, optamos por dividir nosso esforço em duas partes, intituladas, respectivamente, “Fundamentos” e “Contribuições e Resultados”. Na primeira etapa, partiremos dos conceitos fundamentais de comunicação, passaremos pelos pilares da filtragem adaptativa e pelas principais estruturas não-lineares empregadas e, por fim, chegaremos às ferramentas de busca que formarão o cerne da proposta. Detalhemo-la:

**Capítulo 2 – Sistemas de Comunicação Digital:** No capítulo 2, são expostos e analisados os fundamentos da transmissão digital de informação. Permeou sua feitura a idéia de partir da noção primitiva de intercâmbio de mensagens para, ao final do processo, chegar ao modelo discreto de um sistema de comunicação digital, base dos estudos subseqüentes. Ao longo do caminho, são discutidos conceitos essenciais como modulação, amostragem, interferência intersimbólica, ruído etc.

**Capítulo 3 – Equalização, Identificação e Predição:** O objetivo fundamental desse capítulo é definir as bases do processo de filtragem e de sua análise. Primeiramente, estabelece-se uma descrição sistemática do projeto de um filtro, para que, em seguida, sejam analisadas em detalhe as múltiplas facetas dos problemas de equalização, identificação e predição, bem como dos métodos analíticos a eles associados.

**Capítulo 4 – Estruturas Não-lineares de Filtragem:** No capítulo 4, apresentamos as principais estruturas não-lineares empregadas em problemas de filtragem adaptativa – redes

neurais, filtros *fuzzy* e filtros de Volterra – e analisamos suas características essenciais e os principais aspectos de seu projeto. O processo de exposição é acompanhado de uma avaliação crítica do potencial de cada estrutura e também das diferentes metodologias de treinamento usuais, o que dá ao leitor uma visão suficientemente ampla da inserção dos dispositivos apresentados no cenário estabelecido pelo capítulo 3. Por fim, também estão presentes considerações acerca da relação entre as diferentes estruturas.

**Capítulo 5 – Otimização Baseada em Computação Evolutiva:** O capítulo 5 é devotado a uma cuidadosa exposição das duas técnicas de otimização capazes de empreender de maneira assaz eficiente tarefas de otimização numa ampla gama de cenários e aplicações: algoritmos genéticos e a rede imunológica artificial opt-aiNet. Tanto as ferramentas quanto os conceitos fundamentais a elas relacionados são expostos e discutidos em detalhe, de modo a permitir, ao final do capítulo, uma avaliação da forma pela qual se relacionam as ferramentas de busca ao problema genérico de filtragem exposto no capítulo 3.

Tendo, com a primeira parte, estabelecido uma sólida base teórica, passamos, na segunda parte, à exposição e análise de diversas propostas originais que, além de possuírem um valor inerente ao contexto de suas aplicações específicas, inserem-se no espírito mais amplo do paradigma norteador. Os capítulos de que se constitui a segunda etapa podem ser descritos como a seguir:

**Capítulo 6 – Prólogo à Segunda Parte:** No prólogo à segunda parte, são expostos com clareza os objetivos do paradigma que se insinua ao longo da etapa anterior. Trata-se de uma espécie de “manifesto”, uma “declaração de princípios” que permeará toda a discussão subsequente.

**Capítulo 7 – Equalização Linear Não-supervisionada e Computação Evolutiva:** O capítulo 7 é dedicado à análise do emprego de ferramentas evolutivas ao problema de equalização linear cega. Nele, mostra-se como, a partir do critério do módulo constante e de uma ferramenta de busca global, é possível chegar a um paradigma para obtenção da melhor solução de Wiener sem a necessidade de supervisão. Consideram-se duas propostas: uma baseada num algoritmo genético com nichos, e a outra, numa rede imunológica artificial.

**Capítulo 8 – Equalização Recursiva Não-supervisionada e Computação Evolutiva:** No capítulo 8, aborda-se o problema de equalização não-supervisionada baseada em estruturas recorrentes clássicas: o filtro IIR (*infinite impulse response*) e o *decision-feedback equalizer* (DFE). Em ambos os casos, é adotada uma estratégia de projeto baseada no critério do módulo constante e numa rede imunológica artificial, a qual é avaliada sob diferentes modelos de canal e à luz de limitações teóricas bastante conhecidas.

**Capítulo 9 – Novas Propostas para Equalização e Pré-distorção Baseadas em Redes Neurais:** No capítulo 9, analisam-se três problemas: equalização e pré-distorção baseadas numa rede MLP (*Multilayer Perceptron*) e equalização baseada numa rede neural recorrente. São elaboradas e estudadas propostas baseadas em critérios supervisionados e numa rede imunológica artificial que, conforme mostram os resultados obtidos, engendram

sólidas abordagens do problema de equalização e uma nova metodologia para solução do problema de pré-distorção.

**Capítulo 10 – Novas Propostas para Identificação de Sistemas Baseadas em**

**Computação Evolutiva:** O capítulo 10 é dedicado a propostas relativas a dois problemas essenciais: identificação não-supervisionada via *cumulant matching* e identificação de plantas recursivas. No primeiro caso, avaliamos como o uso de uma ferramenta imunológica pode levar a uma proposta eficiente e robusta à presença de ruído aditivo. Já no caso faltante, mostramos que o uso de tal técnica de otimização pode eliminar um dilema teórico clássico e prover bases para um método sólido e genérico de identificação de plantas lineares e não-lineares na presença de ruído.

**Capítulo 11 – Emprego de Algoritmos Genéticos no Projeto Não-supervisionado do Receptor de Máxima Verossimilhança:**

No capítulo 11, abordamos um problema essencial em comunicações digitais: o de obter a seqüência de dados transmitidos que produz a máxima medida de verossimilhança. Após formular o problema de maneira independente da estimação de canal, mostramos que um algoritmo genético simples pode resolvê-lo de forma eficiente e decisivamente parcimoniosa, provendo, destarte, meios para uma abordagem confiável desse que, de certa maneira, é um “problema genérico de equalização”.

**Capítulo 12 – Uma Proposta para Equalização Não-linear Cega Baseada num Filtro de Erro de Predição Fuzzy:**

No capítulo 12, apresentamos uma proposta ligada a um assunto ainda pouco explorado: equalização não-linear cega. A abordagem, baseada no uso de um filtro de erro de predição *fuzzy* como equalizador, mostra-se bastante eficiente tanto na equalização de canais de fase mínima quanto na de canais de fase não-mínima, o que atesta sua solidez e abre importantes perspectivas conceituais.

À medida que apresentarmos os resultados obtidos, procuraremos situá-los no contexto mais específico de seus domínios de validade, e, ali, trataremos de tecer uma avaliação particular das contribuições expostas. Por fim, ao final do trabalho, de posse de um *corpus* de elementos teóricos e práticos, poderemos ter uma visão mais ampla do edifício que haveremos erguido e, de lá, julgar apropriadamente o paradigma que nos norteia.

# PARTE I

## Fundamentos

*“Quando se quer pôr ordem e método numa discussão, é preciso dar início definindo a coisa de que se debate, para se ter dela uma idéia clara e precisa”.*

*Marco Túlio Cícero, Dos Deveres, II<sup>3</sup>*

---

<sup>3</sup> Da versão publicada pela Editora Martin Claret.



## 2. Sistemas de Comunicação Digital

### 2.1. Introdução

A necessidade de interagir através do intercâmbio de mensagens é um fato humano que se revela, ao longo da História, sob inumeráveis aspectos. Sendo um ser eminentemente gregário, o homem tem buscado, desde a aurora de sua existência como espécie, formas de “codificar” e “transmitir” mensagens de teores diversos. Embora a simplicidade de tais objetivos seja bastante aparente, a riqueza de possíveis abordagens é estonteante, tangendo campos tão ricos quanto a semiótica, a lingüística e a teoria de comunicações. Atendo-nos apenas ao último desses campos, no sentido mais familiar ao engenheiro de sistemas, não poderemos negar quantas possibilidades há de fazer chegar, de um ponto a outro, uma informação qualquer. A escolha adequada dependerá de diversos fatores, tais como a distância entre os pontos, as características do meio físico entre eles, o grau de tecnologia disponível ao projetista, a urgência da mensagem etc.

Conforme nos conta Heródoto [Heródoto, 1952], na iminência do confronto com os persas na célebre batalha de Maratona, os atenienses enviaram um certo Feidípides a Esparta para pedir auxílio militar à poderosa cidade do Peloponeso. O mensageiro então, numa corrida de proporções impressionantes, percorreu a pé os mais de duzentos quilômetros que separam as duas cidades em menos de dois dias, apenas para ouvir uma negativa dos lacedemônios. É lícito supor que tal solução devesse ser a mais adequada que os engenhosos atenienses puderam conceber a partir das restrições práticas impostas pelas circunstâncias. Provavelmente outros povos da época fossem capazes de levar o mesmo pedido de maneira distinta, dependendo de seus costumes e da magnitude de seus domínios. Impérios gigantescos, como o próprio império persa, dispunham de uma notável “rede” de arautos aptos a levar celeremente informações a destinos longínquos por meio de uma alternância que poderíamos compreender hoje como um princípio de repetição. E nos dias atuais, caso a presença física de um embaixador fosse imprescindível, em menos de uma hora a mensagem chegaria à autoridade competente. Se a comunicação não exigisse presença física, bastaria uma linha ou uma conexão segura para que o pedido fosse comunicado em uma questão de, no máximo, segundos.

Esse pequeno exemplo histórico serve para ilustrar diferentes *sistemas de comunicação* concebidos para realizar a mesma tarefa: enviar uma mensagem de um transmissor a um receptor. Cada possível solução contém estágios pelos quais, num certo sentido, a informação trafega sob diversas “formas”: vozes variadas, conexões do cérebro do arauto em movimento, sinais elétricos na rede telefônica moderna etc. A atuação coordenada de todos eles forma um mecanismo complexo que permite que todos nós nos entendamos mutuamente<sup>4</sup>.

A adequada compreensão desse mecanismo passa necessariamente pelo estudo de cada etapa do processo. Assim, é imperativo que todo aquele que deseje propor novas soluções entenda, no mínimo com razoável profundidade, o funcionamento de cada elemento do sistema global. Tendo em vista esse fato, pareceu-nos apropriado iniciar este

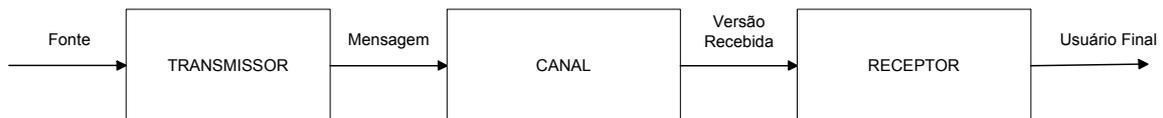
---

<sup>4</sup> Note que a definição acima se estende a outras modalidades de comunicação (máquina-máquina, homem-máquina, homem-animal etc.).

trabalho por um estudo dos principais aspectos dos modernos sistemas de comunicação, com particular ênfase na descrição de efeitos do meio físico no sinal portador da mensagem, efeitos estes que são a principal motivação prática de nosso esforço de pesquisa.

## 2.2. Uma visão geral dos elementos de um sistema de comunicação

Começemos a sistematizar os conceitos que até agora expusemos de maneira genérica e intuitiva. Conforme discutimos, a comunicação se dá através do intercâmbio de mensagens entre pontos distintos, os quais podem ser entendidos como *origem* e *destino* do fluxo<sup>5</sup>. À origem daremos o nome de *fonte*, e, ao destino, o nome de *usuário*, obedecendo assim à nomenclatura empregada em [Haykin, 1994b]. Associado à fonte temos um *transmissor*, cuja função é produzir um sinal que porte a mensagem que se deseja enviar de uma maneira adequada às características do meio que separa origem e destino e às restrições inerentes ao próprio sistema como um todo. Após a ação do transmissor, o sinal resultante é confiado ao *meio* ou *canal de comunicação*, a “estrada” pela qual ele trafegará até que chegue ao *receptor*, dispositivo cuja função é transformar o sinal recebido em uma mensagem inteligível ao usuário final. A Fig. 2.1 traz um esquema simples e coerente com essas definições.



**Figura 2.1:** Esquema Genérico de Comunicação

Para que o leitor se familiarize com esses conceitos básicos, convém que os utilizemos para classificar os componentes de um mecanismo de comunicação extremamente familiar: a fala. Em tal contexto, podemos entender a mente do falante como sendo a fonte de informação, o trato vocal educado em uma certa língua como sendo o transmissor, o ar como sendo o meio de comunicação, o aparelho auditivo como sendo o receptor e, finalmente, a mente do outro indivíduo como sendo o usuário final. Interessantemente, caso o indivíduo receptor fosse um surdo que conhecesse a técnica de leitura labial, teríamos a mesma fonte, os lábios do indivíduo falante fazendo o papel de transmissor, o mesmo meio de comunicação, e o aparelho visual do usuário como receptor. Caso a conversa entre ouvintes se desse por meio de uma linha telefônica, o meio de comunicação passaria a ser o sistema formado pelos aparelhos, centrais, linhas etc. Em qualquer caso, podemos distinguir estágios bem definidos: *geração* da mensagem e do sinal transmitido, sua *propagação* e, finalmente, a *recepção* do sinal e a *extração* da mensagem que ele porta. É importante que, ao analisar um sistema de comunicação, tenhamos a idéia bem clara de onde se situa cada estágio: talvez essa seja a primeira informação que um

---

<sup>5</sup> É interessante considerar que é possível que, em casos muito peculiares, esse fluxo se dê apenas no tempo. Um exemplo seria o de alguém que transmite informação à posteridade através de uma carta guardada num baú enterrado. Mesmo nesse caso, curiosamente, haveria uma deterioração da mensagem!

engenheiro de sistemas busca, mesmo que inconscientemente, num esquema associado a uma aplicação prática.

### **2.3. Modernos sistemas de comunicação**

Tendo construído um arcabouço teórico bastante genérico, que seria útil tanto para um trabalho de engenharia quanto para um estudo lingüístico, é hora de concentrar nossa atenção num campo mais específico: o da *telecomunicação*, ou seja, da comunicação à distância. Como nos mostra o exemplo histórico do início deste capítulo, o conceito de telecomunicação não é novo, mas surge de uma ampliação da rede de relações humanas. Apesar disso, podemos estabelecer dois marcos que assinalam uma mudança substancial em tal sorte de comunicação: o advento da transmissão por ondas de rádio e o avanço no campo da eletrônica ao longo do século passado. A consolidação desse novo paradigma, baseado no uso de fenômenos eletromagnéticos, deu-se com invenções como o telefone, o telégrafo sem fio, o rádio, a televisão etc. A evolução concomitante dos computadores digitais logo levou a uma convergência que modificou substancialmente o *modus vivendi* de boa parte da humanidade em poucas décadas.

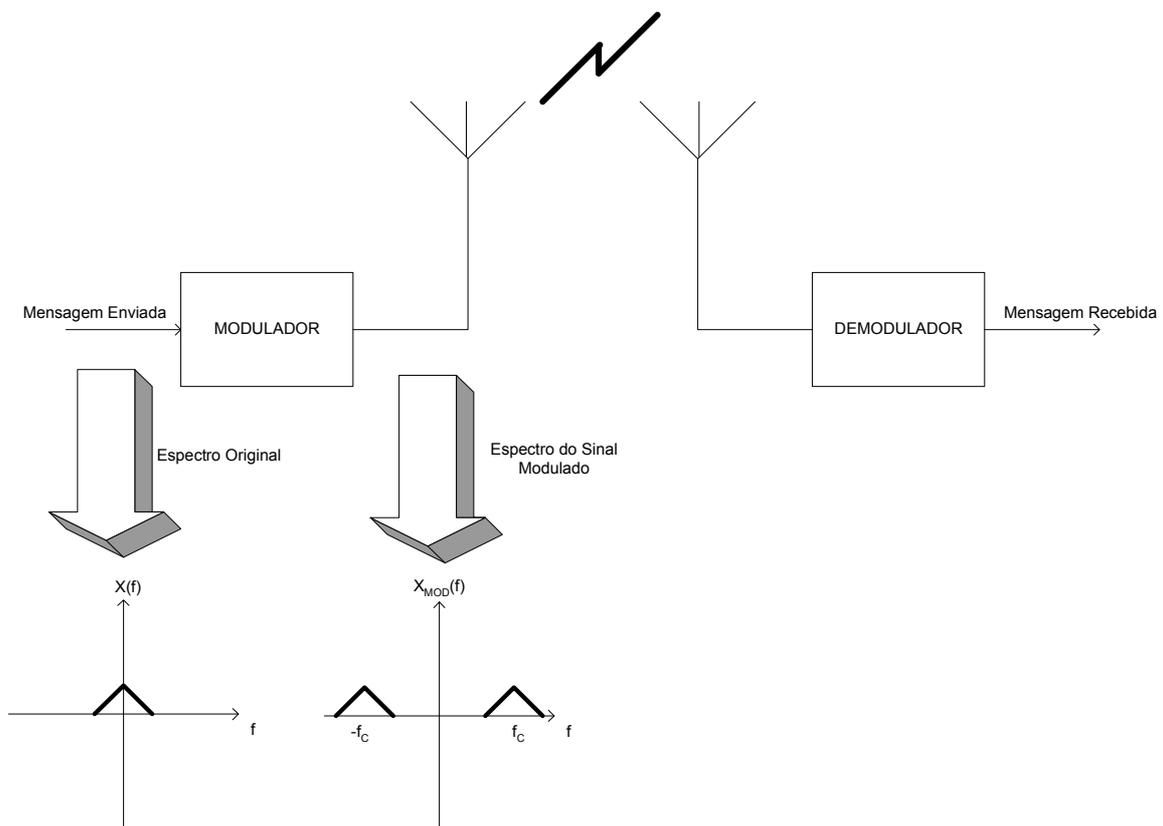
É no contexto desse paradigma reinante que se insere nosso trabalho, concebido a partir de problemas e soluções que têm suas bases na corrida tecnológica que marcou o século XX. Faz-se necessário, portanto, começar a analisar mais detalhadamente como esses sistemas “modernos” se encaixam nas definições amplas da seção anterior, ou seja, importa que aprofundemos nosso conhecimento tendo já como pano de fundo o conjunto de noções próprias que pautaram as propostas que compõem este trabalho. Como ponto de partida, analisemos uma “bipolaridade” que marcou o desenvolvimento do atual paradigma, embora, em verdade, não seja de maneira alguma exclusiva dele: analógico ou digital?

### **2.4. Sistemas Analógicos**

Voltemos a um exemplo que já usamos anteriormente: a comunicação entre duas pessoas através da fala. No entanto, assumamos agora que uma dessas pessoas se encontra em São Paulo e a outra em Campinas. Felizmente, não será preciso designar um arauto para percorrer os cem quilômetros que separam as cidades: as ondas eletromagnéticas podem perfeitamente fazer esse papel em uma fração de segundo imperceptível a um ser humano. Utilizar essas ondas como veículo de transporte da voz humana é o princípio fundamental da comunicação radiofônica.

A mente do locutor é ainda a fonte de informação num sistema de radiodifusão, pois nela serão concebidas as mensagens que chegarão aos ouvintes. Seu trato vocal ainda articula fonemas de forma a produzir frases inteligíveis; as ondas sonoras geradas ainda se propagam, mas não mais diretamente ao aparelho auditivo do usuário: será necessário converter esses sinais em outros mais adequados ao meio de transporte escolhido. O primeiro passo é converter o sinal sonoro em um sinal elétrico, o que se faz por meio de um microfone. O sinal elétrico, como o sinal sonoro, conterà predominantemente frequências entre 300 e 3400 Hz. Tal faixa representa um problema, pois transmitir nela acarretaria perdas inaceitáveis na atmosfera e sérias dificuldades no projeto das antenas utilizadas. Existiria ainda uma outra questão complexa: todas as transmissões simultâneas ocupariam a mesma região do espectro, o que produziria um inaceitável nível de interferência. Por essas

razões, será imperativo submeter nossa mensagem a mais uma metamorfose: uma translação do conteúdo espectral a uma faixa localizada em uma região de frequências bem mais altas, e que, de preferência, esteja alocada exclusivamente para nossa transmissão. Dá-se o nome de *modulação* a esse processo de moldagem do sinal que se deseja enviar às características e restrições sistêmicas. No caso da radiodifusão, o processo tem lugar de duas maneiras principais: embutindo a informação na amplitude de uma onda senoidal com uma frequência  $f_c$ , em torno da qual ficará o espectro resultante (modulação em amplitude, AM), ou na frequência/fase de tal senóide (modulação em frequência, FM, modulação em fase, PM). Finda essa etapa, o sinal está pronto a excitar a antena, que, por sua vez, é o ente responsável pela emissão das ondas eletromagnéticas. Após a propagação pelo ar, as ondas chegam a uma antena receptora que produz um sinal elétrico cujo espectro é devolvido à posição original por meio de um processo de *demodulação*. O sinal demodulado é então convertido em ondas sonoras por um alto-falante. A Fig. 2.2 apresenta um resumo geral do processo.



**Figura 2.2:** Esquema Simplificado de um Sistema com Modulação Analógica

Na terminologia que definimos, o transmissor seria o conjunto formado pelo trato vocal, o ar entre a boca e o microfone, o microfone, o circuito modulador e a antena. O meio físico entre transmissor e antena receptora seria o canal de comunicação, e o conjunto composto de antena receptora, circuito demodulador e alto-falante formaria o receptor. Entretanto, há uma outra forma interessante de enquadrar cada estágio: podemos entender como sinal transmitido o sinal elétrico (assumindo que o microfone é ideal) e atribuir ao canal de comunicação todos os efeitos nocivos e ruídos introduzidos por antenas e circuitos

imperfeitos, os quais, na primeira interpretação, eram parte de transmissor e receptor. Como o leitor perceberá, deveremos adotar preferencialmente o segundo ponto de vista ao longo do texto, pois é vantajoso pensar em um modelo único de todas as distorções impostas ao sinal enviado<sup>6</sup>.

A mais marcante característica da transmissão analógica, que, como vimos, baseia-se essencialmente no processo de modulação, é, sem dúvida, sua simplicidade conceitual. Isso explica seu emprego em meios de comunicação de massa tão importantes quanto rádio, TV e telefones celulares. Apesar disso, esquemas analógicos são preteridos em aplicações de ponta, pois o aperfeiçoamento dos computadores tornou viável o emprego massivo dos robustos e eficientes esquemas digitais, cujos fundamentos começaremos já a analisar.

## 2.5. **Sistemas Digitais: o Teorema da Amostragem**

O sistema analógico é muito simples e elegante, mas a constante busca por um melhor desempenho, que, aliás, pode ser decisivo em determinadas aplicações, motivou os projetistas a procurar métodos que fossem capazes de se ajustar melhor às características e limitações do meio de transmissão. A chave para isso está na representação de um sinal contínuo por meio de um conjunto de valores discretos obtidos através de um processo de *amostragem*. À primeira vista, tal processo parece levar a uma perda irreparável, pois envolve a representação de um *continuum* de valores através de um conjunto discreto. Interessantemente, há um teorema usualmente atribuído a Harry Nyquist<sup>7</sup> que mostra que, sob algumas hipóteses, pode haver uma perfeita equivalência entre ambas as representações i.e. *que é possível que o sinal contínuo seja reconstruído a partir do conjunto de amostras*:

**Teorema da Amostragem:** *Um sinal com conteúdo espectral limitado a frequências menores que  $W$  hertz pode ser perfeitamente descrito por uma seqüência de amostras colhidas com uma frequência não inferior a  $2W$  amostras/s.*

A chave do teorema é a finidade do espectro do sinal que se deseja amostrar: é essa característica que permite a perfeita recuperação, como tentaremos mostrar. Imaginemos que o sinal seja amostrado por meio de um trem de impulsos periódicos definido por:

$$I(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT_A), k \in Z \quad (2.1),$$

sendo  $T_A$  o período de amostragem. A versão amostrada de um sinal  $x(t)$  será:

$$x_A(t) = x(t)I(t) = \sum_{k=-\infty}^{\infty} x(kT_A)\delta(t - kT_A) \quad (2.2)$$

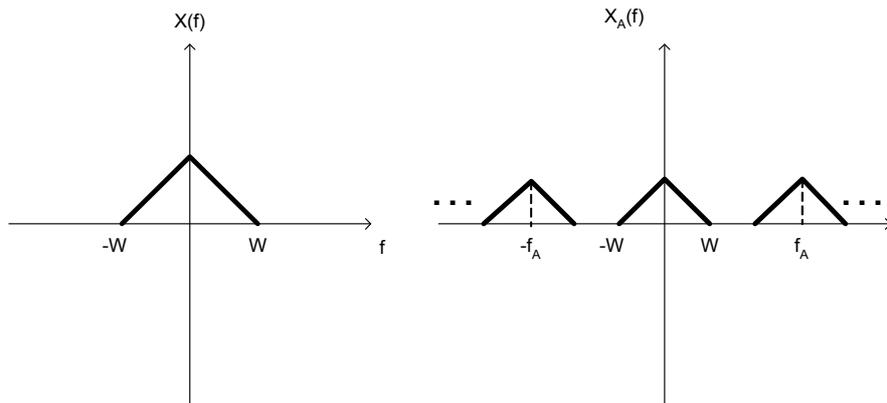
<sup>6</sup> Esse modelo, por força de um abuso consagrado pelo uso, será chamado muitas vezes de canal de comunicação, ou seja, daquilo que busca representar.

<sup>7</sup> Modernamente, também se dá crédito a Shannon e Gabor, conforme discussão apresentada em [Oppenheim et al., 1997].

O espectro de  $x_A(t)$ , por sua vez, será:

$$X_A(f) = f_A \sum_{m=-\infty}^{\infty} X(f - mf_A), m \in Z \quad (2.3)$$

onde  $X(f)$  é o espectro de  $x(t)$  e  $f_A = 1/T_A$ . Note o leitor um ponto crucial:  $X_A(f)$  é uma soma de versões de  $X(f)$  igualmente espaçadas na frequência, como mostra a Fig. 2.3.



**Figura 2.3:** Visão Espectral da Amostragem

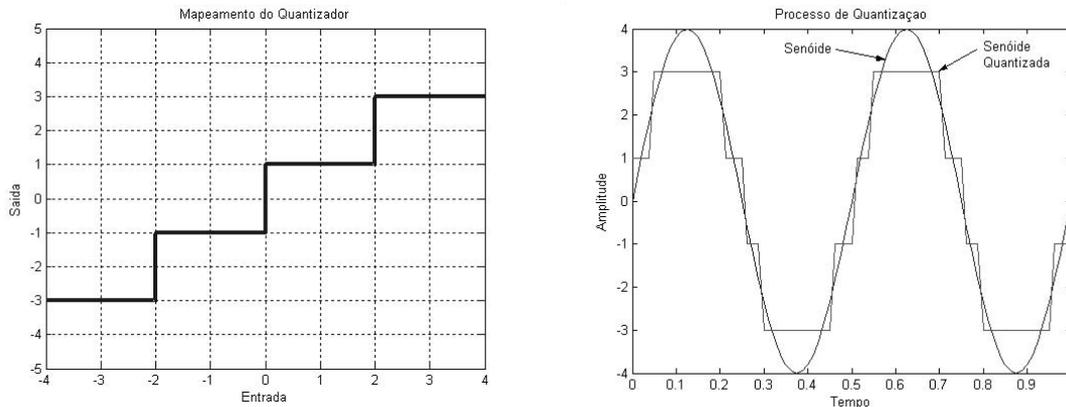
Na Fig. 2.3, notamos que o espectro do sinal amostrado é formado por repetições de  $X(f)$  que *não se sobrepõem*. Isso ocorre porque duas condições foram obedecidas:  $X(f)$  tem faixa limitada por um valor máximo  $W$ , e a taxa de amostragem  $f_A$ , que determina o espaçamento entre as repetições, é maior que  $2W$ . Já é possível entender, portanto, as exigências do teorema de Nyquist. Resta uma última questão: como recuperar o sinal original  $x(t)$  a partir de  $x_A(t)$ ? Observando o espectro  $X_A(f)$ , vemos que um filtro passa-baixas ideal com largura idêntica à da faixa de  $x(t)$  pode eliminar as repetições e fazer com que se mantenha apenas a versão em torno de  $f = 0$ . Dessa forma, a recuperação se dá por meio de uma filtragem do sinal  $x_A(t)$ . Perceba o leitor que, para explicar o teorema, empregamos um método de amostragem abstrato, baseado num conjunto de funções impulso, o que, no entanto, não reduz sua generalidade.

Por fim, caso o sinal  $x(t)$  não tenha um espectro limitado, haverá uma sobreposição de conteúdo espectral, responsável pela introdução de distorções no sinal recuperado. A tal fenômeno se dá o nome de *aliasing*, e é justamente para evitá-lo que se efetua preliminarmente uma filtragem passa-baixas do sinal que se deseja amostrar (filtragem *anti-aliasing*).

## 2.6. Sistemas Digitais: Quantização

O teorema da amostragem propicia uma representação muito interessante de um sinal contínuo por meio de suas amostras. Elas, no entanto, pertencem a um *continuum*, ou seja, podem assumir, em tese, qualquer valor contido num intervalo real. É preciso, portanto, que também limitemos as possibilidades nesse domínio, chegando assim a uma representação discreta tanto no domínio do tempo (o que obtivemos com a amostragem) quanto no domínio da amplitude do sinal. Isso pode ser obtido através da *quantização* i.e.

da representação do sinal por um conjunto finito de níveis. Em termos matemáticos, o processo consiste da filtragem do sinal amostrado por um elemento não-linear sem memória, sendo assim produzidas amostras de saída pertencentes a um repertório finito de possibilidades. A Fig. 2.4 apresenta um exemplo de mapeamento não-linear e também o resultado de sua aplicação a uma senóide.



**Figura 2.4:** Mapeamento Não-linear e Quantização de uma Senóide

Como era de se esperar, o processo introduz distorções, às quais se dá habitualmente o nome de *ruído de quantização*. Esse ruído é tanto menor quanto maior for o número de níveis, fato que estabelece um compromisso entre precisão e número de possibilidades que, como veremos imediatamente, relaciona-se com a “quantidade de informação” necessária para que o sinal seja representado.

Aproveitemos o caso mostrado na Fig. 2.4 para tecer mais algumas considerações relevantes. Nele, há quatro níveis: +3, +1, -1 e -3. Seria possível, dessa forma, representá-los através de um código quaternário pela atribuição de um símbolo a cada nível. Suponhamos que usemos as quatro primeiras letras do alfabeto para representar os níveis, e que adotemos a ordem em que se encontram no início deste parágrafo. Seríamos capazes, então, de representar um sinal amostrado através de uma seqüência de letras como, por exemplo, AABBBBCDDC. Se fosse mais conveniente, poderíamos utilizar uma codificação binária, empregando processos como os que discutiremos na próxima seção. De qualquer maneira, é muito importante atentar para o seguinte: temos em mãos um conjunto de processos que nos permite converter um sinal analógico qualquer em uma seqüência de bits, com um grau de distorção controlável. Trata-se de uma ponte entre o mundo analógico e o mundo digital, ponte esta que nos permite armazenar sinais de áudio e vídeo em nossos computadores ou transmitir sinais utilizando modulações que explorem a limitação da quantidade de possíveis símbolos transmitidos.

## 2.7. Sistemas Digitais: Codificação

Acabamos de indicar o caminho para efetuar a chamada *conversão analógico-digital* (*conversão A/D*), ou seja, representar um sinal analógico através de uma seqüência

de bits<sup>8</sup>. Como vimos, uma importante vantagem dessa representação é que ela é perfeitamente adequada ao processamento e à armazenagem por computadores digitais, pilares da tecnologia moderna. Daqui por diante, o sinal de voz se torna apenas um conjunto de dados idêntico àqueles trocados, por exemplo, entre máquinas nas redes de computadores. Resta conhecer os passos necessários para que esses dados sejam transmitidos adequadamente por um meio qualquer.

Partamos de uma fonte de informação digital, não sendo neste momento relevante se os bits representam um sinal de voz ou dados produzidos por uma máquina. Nos modernos sistemas de comunicação, antes de pensarmos em transmitir bits, faz-se necessário *codificar* a informação de maneira apropriada. O processo global se divide usualmente em duas partes: *codificação de fonte* e *codificação de canal*. O objetivo da codificação de fonte é eliminar redundâncias, ou seja, representar a informação gerada pela fonte de maneira tão parcimoniosa quanto possível. Já a codificação de canal realiza um processo distinto: ela introduz algumas redundâncias controladas que cooperam para mitigar distorções impostas pelo canal. A base dos processos são dois famosos teoremas atribuídos a Claude Shannon [Shannon, 1948], que trazem à baila conceitos importantes como *entropia da fonte* e *capacidade de canal*.

## 2.8. Sistemas Digitais: Modulação

Estando a informação já numa forma compacta e adequada a determinados efeitos nocivos do canal, é preciso moldá-la à transmissão pelo meio físico, ou seja, modulá-la digitalmente. Analisemos algumas maneiras por que o processo pode se dar.

### 2.8.1. Modulação por Amplitude de Pulso (PAM):

A primeira técnica de modulação que abordaremos é a chamada *modulação por amplitude de pulso*, ou seja, *PAM (Pulse Amplitude Modulation)*. Trata-se de um esquema de modulação em banda-base i.e. *que não envolve translação espectral*, sendo, portanto, aplicável a sistemas que não abrangem difusão via rádio, como é o caso, por exemplo, na telefonia fixa. A idéia básica é simples: associar a cada símbolo um pulso de determinada amplitude, sempre escolhida de modo a otimizar parâmetros sistêmicos importantes como a potência média transmitida e a robustez à presença de ruído.

Analisemos, em primeiro lugar, um exemplo binário em que a fonte de informação produz bits que rotularemos como +1 e -1<sup>9</sup>. A essa seqüência damos o nome de  $a_n$ , sendo o sinal PAM produzido igual a

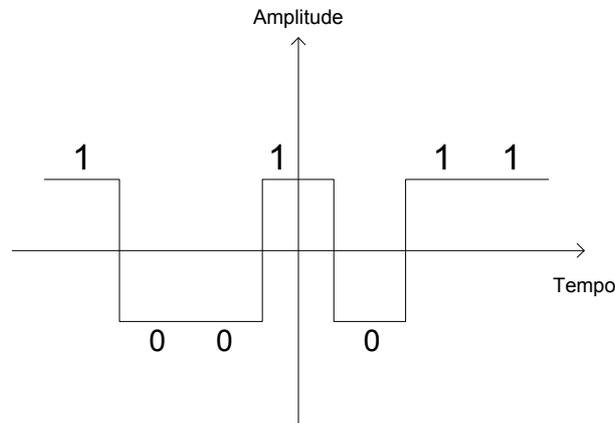
$$s(t) = \sum_{k=-\infty}^{\infty} a_k p(t - kT_A), k \in Z \quad (2.4)$$

---

<sup>8</sup> Embora, por razões práticas, usualmente falemos em *bits*, seria perfeitamente possível basear o processo todo em alfabetos com mais possibilidades.

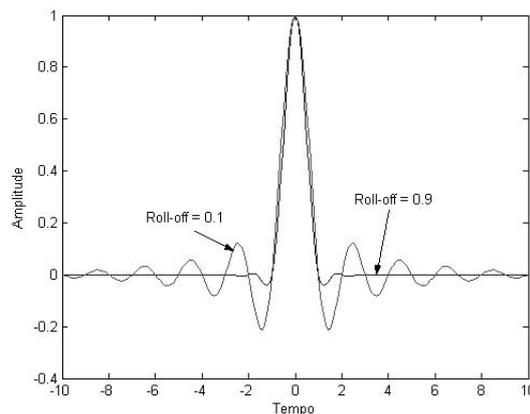
<sup>9</sup> É possível mostrar que a configuração simétrica é ótima do ponto de vista do compromisso entre potência e desempenho [Barry et al., 2004].

onde  $p(t)$  é o formato de pulso escolhido. Na Fig. 2.5, mostramos um sinal PAM produzido quando se adota um pulso retangular.



**Figura 2.5:** Pulsos PAM

Tal formato de pulso tem um valor eminentemente teórico, pois leva idealmente a um espectro com extensão infinita em frequência, não sendo, destarte, particularmente adequado à transmissão por canais limitados em frequência<sup>10</sup>. Para superar essa dificuldade, é necessário adotar um formato de pulso que satisfaça o chamado *critério de Nyquist para transmissão em banda-base*, ou seja, que garanta que o conjunto formado pelo filtro de transmissão que modela o pulso, pelo próprio meio e pelo filtro de recepção não provoque distorções na faixa passante do canal. A solução mais adotada na prática é recorrer à classe de pulsos com espectro do tipo *coseno levantado*, capazes de satisfazer o critério supramencionado. Voltaremos ao tema quando discutirmos os efeitos do canal, sendo suficiente por ora apresentar a Fig. 2.6, em que se encontram alguns pulsos desse tipo (diferenciados por um parâmetro denominado *fator de roll-off*).



**Figura 2.6:** Pulsos de Cosseno Levantado

<sup>10</sup> Canais desse tipo são a regra em comunicações.

### 2.8.2. Modulações em Banda Passante

Quando é preciso transmitir a informação em altas frequências, faz-se necessário recorrer a modulações baseadas em uma portadora senoidal, num processo que guarda similaridades com o esquema de modulação analógica discutido anteriormente. Em vez de pulsos em banda-base, como os que caracterizam a modulação PAM, são utilizados pulsos senoidais, ou seja, com espectro situado em banda passante i.e. *em torno da frequência  $f_c$  da portadora*. Isso nos leva a três possibilidades básicas:

1) Embutir a informação na amplitude da portadora (ASK, *Amplitude-Shift Keying*). Nesse caso, poderíamos representar os bits 0 e 1 por, respectivamente:

$$s_0(t) = 0 \text{ e } s_1(t) = A.\cos(2\pi f_c t) \quad (2.5)$$

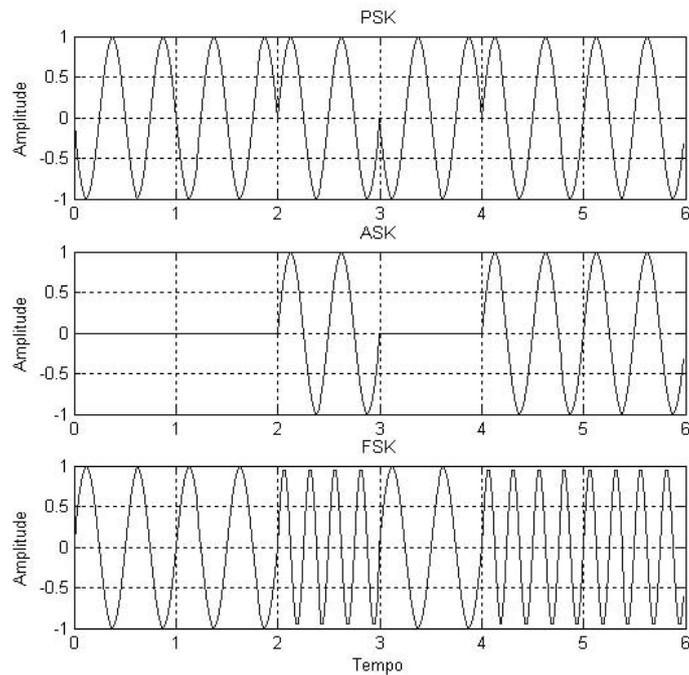
2) Embutir a informação na frequência da portadora (FSK, *Frequency-Shift Keying*). A representação binária tomara então a forma

$$s_0(t) = A.\cos(2\pi f_{c1} t) \text{ e } s_1(t) = A.\cos(2\pi f_{c2} t) \quad (2.6)$$

3) Finalmente, embutir a informação na fase da portadora (PSK, *Phase-Shift Keying*). A representação poderia ser

$$s_0(t) = A.\cos(2\pi f_c t - \phi) \text{ e } s_1(t) = A.\cos(2\pi f_c t + \phi) \quad (2.7)$$

Na Fig. 2.7, mostramos como seria o sinal transmitido correspondente a uma seqüência arbitrária 001011 para cada modulação.



**Figura 2.7:** Exemplos de Modulação Digital

Quatro pontos merecem destaque:

- Vale para esses casos um modelo idêntico ao apresentado em (2.4), sendo apenas necessário assumir que o formato do pulso obedece às representações  $s_0(t)$  e  $s_1(t)$  dos exemplos anteriormente mostrados.
- Persiste a necessidade de efetuar o processo de conformação de pulso para que seja satisfeito o critério de Nyquist, dessa vez adaptado a um canal em torno da frequência da portadora. No exemplo mostrado na Fig. 2.7, note o leitor que a portadora se encontra “empacotada” por um formato retangular de pulso, válido para ilustração, mas, como já discutimos, pouco prático.
- É possível também combinar duas das estratégias apresentadas em uma única modulação: é o caso do esquema QAM (*Quadrature Amplitude Modulation*), no qual a informação se encontra *tanto na amplitude quanto na fase da portadora*.
- Finalmente, pode ser vantajoso embutir informação não nos próprios símbolos, mas na *transição entre eles*. Tal idéia é a base de *modulações diferenciais*, que podem tornar mais simples os processos de detecção (e demodulação).

## **2.9. Sistemas Digitais: Considerações sobre o Canal de Comunicação**

Tendo analisado os processos iniciais de codificação e modulação, chegamos por fim à transmissão propriamente dita. Para que ela ocorra, é fundamental que se disponha de um *meio*, de um *canal* por onde trafegue o sinal modulado. Esse canal terá características próprias, que, por sua vez, se refletirão no sinal recebido, ou seja, *o meio modificará a mensagem que por ele viaja*. É essencial que o projetista conheça os principais efeitos produzidos pelo canal, pois eles definirão limitações sistêmicas e, conseqüentemente, inspirarão as contramedidas necessárias à manutenção de um determinado patamar de qualidade. Dois fenômenos têm particular importância para nós neste trabalho: a *interferência intersimbólica (IIS)* e o *ruído*. Discutiremos brevemente ambos, visando trazer à baila conceitos essenciais que serão aprofundados à medida que assim exigir a necessidade.

### **2.9.1. Interferência Intersimbólica (IIS)**

Como já adiantamos em nossos argumentos acerca da necessidade da obediência ao critério de Nyquist para transmissão, a largura de banda limitada do canal é um fator que deve ser cuidadosamente considerado no projeto de um sistema de comunicação digital. A razão de ser desse fato é que o corte de frequências pertencentes ao espectro do sinal transmitido distorce o formato dos pulsos, dando margem a uma potencial *sobreposição de pulsos espalhados* capaz de ocasionar erros durante a reconstrução da mensagem. A idéia de efetuar um processo de conformação de pulso é, exatamente, reduzir essa interferência intersimbólica (IIS) adotando formatos apropriados às características gerais do meio. É importante frisar que mesmo fenômenos *não-lineares* podem causar interferência intersimbólica ou, como é mais usual, torná-la mais pronunciada. Em cenários desse tipo, a impossibilidade de recorrer ao princípio da superposição torna o domínio da frequência bem menos útil à análise.

### **2.9.2. Ruído**

Intuitivamente, associamos ao conceito de *ruído* algum tipo de perturbação que, por sua atuação sobre o sinal, torna-se responsável por distorcer a informação nele contida. Há várias possíveis fontes de ruído em um sistema de comunicação [Haykin, 1994b]: fenômenos atmosféricos e cósmicos, efeitos inerentes a dispositivos semicondutores, agitação térmica típica de elementos dos circuitos que compõem o sistema etc. Devido à imensa complexidade dessa espécie de fenômenos, os modelos de ruído têm geralmente caráter *estocástico*, ou seja, são de natureza *aleatória*. Tais modelos precisam se adequar tanto às peculiaridades fenomenológicas quanto às observações colhidas na prática, o que engendra uma enorme diversidade de abordagens.

Duas espécies de ruído merecerão especial atenção neste trabalho: o *ruído térmico*, onipresente em sistemas de comunicação, e o chamado *ruído impulsivo*, associado a causas tão diversas quanto fenômenos atmosféricos e submarinos [Kuruoglu et al., 1998a]. Discutiremos cada um separadamente, buscando estabelecer um contraste de importante valor conceitual para o leitor.

## Ruído Térmico

É um resultado conhecido da Termodinâmica que haverá agitação térmica de elétrons de um condutor com temperatura superior ao zero absoluto, ou seja, que existirá uma corrente ocasionada por essas flutuações. Tal fenômeno afeta os circuitos de um sistema de comunicação, distorcendo a mensagem que chega ao usuário final. Essa distorção é normalmente modelada por um processo estocástico que se sobrepõe ao sinal recebido, ou seja, que se soma a ele. O epíteto “*aditivo*” é empregado para descrever ruídos desse tipo, os mais comuns no campo das telecomunicações.

Devido ao número muito elevado de elétrons que se movem, ou seja, da magnitude de efeitos que se sobrepõem, é natural que se opte por um modelo gaussiano de ruído, o que se justifica pelo *teorema central do limite* (*central limit theorem*), que enunciamos da seguinte forma:

**Teorema Central do Limite:** *Seja um conjunto de  $N$  variáveis aleatórias i.i.d. (independentes e identicamente distribuídas)  $X_i$ , com média nula e variância unitária. Definamos a variável  $V_N$ :*

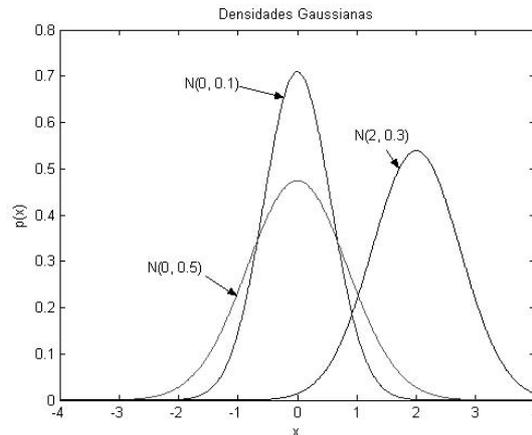
$$V_N = \frac{1}{\sqrt{N}} \sum_{i=1}^N X_i \quad (2.8)$$

*O teorema central do limite garante que o limite de  $V_N$  quando  $N$  tende a infinito será uma variável aleatória gaussiana com média nula e variância unitária [Papoulis, 1991].*

Em termos bem simples, o teorema mostra que a soma de variáveis aleatórias i.i.d. tende a uma gaussiana, o que explica a popularidade dessa densidade em contextos marcados pela convergência de um grande número de eventos individuais. A expressão da densidade gaussiana obedece à fórmula

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left[-\frac{(x - \mu_x)^2}{2\sigma_x^2}\right] \quad (2.9)$$

onde  $\mu_x$  é a média da variável e  $\sigma_x^2$  sua variância. Uma variável aleatória com distribuição igual à mostrada em (2.9) pode ser denominada simplesmente  $N(\mu_x, \sigma_x^2)$ . A Fig. 2.8 traz exemplos de possíveis densidades dessa espécie.



**Figura 2.8:** Densidades Gaussianas

Havendo já justificado a opção por um processo gaussiano, é preciso que agora consideremos um outro aspecto do ruído: suas características espectrais. O conteúdo em frequência de um processo estocástico é comumente expresso através de um ente matemático denominado *densidade espectral de potência* (DEP), que, intuitivamente, representa uma média do valor quadrático absoluto das possíveis transformadas de Fourier de infinitas realizações suas. A definição, válida para um processo estacionário no sentido amplo<sup>11</sup> contínuo, é a seguinte:

$$S_x(f) = \mathfrak{F}[r(\tau)] \quad (2.10)$$

Para um processo discreto, a definição seria:

$$S_x(f) = \mathfrak{F}[r(k)] \quad (2.11)$$

sendo  $\mathfrak{F}(\cdot)$  um operador que representa a transformada de Fourier e  $r(\tau)$  e  $r(k)$ , respectivamente, as funções de autocorrelação contínua e discreta. A autocorrelação é uma grandeza estatística que avalia o grau de “similaridade” ou de “dependência” entre um processo e uma versão sua atrasada de um intervalo. Suas definições são:

$$r_x(\tau) = E[x(t)x(t+\tau)] \quad (2.12)$$

e

$$r_x(k) = E[x(n)x(n+k)] \quad (2.13)$$

sendo  $E(\cdot)$  o operador *esperança matemática*.

Há uma classe muito particular de processos que têm grande valor teórico quando se lida com ruído: a dos *processos brancos*, ou seja, dotados um conteúdo uniforme e de

<sup>11</sup> Um *processo estacionário no sentido amplo* é aquele cuja média não varia com o tempo e cuja autocorrelação depende apenas da separação entre dois instantes temporais  $t_1$  e  $t_2$  quaisquer.

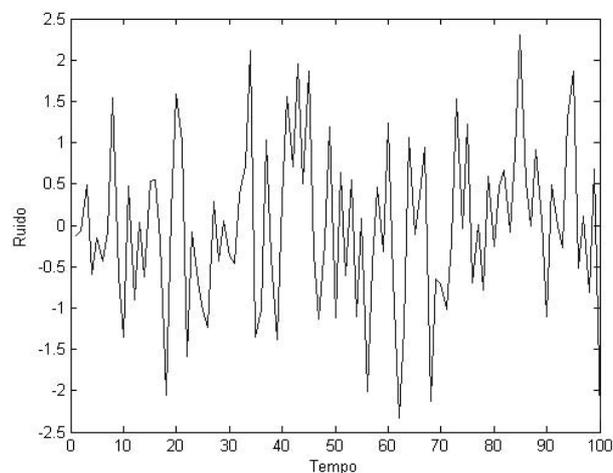
extensão infinita em frequência<sup>12</sup>. A densidade espectral de potência de um processo desse tipo é:

$$S_x(f) = \frac{N_0}{2} \quad (2.14)$$

e, conseqüentemente, sua autocorrelação tem a forma de um impulso:

$$r_x(\tau) = \frac{N_0}{2} \delta(\tau) \quad (2.15)$$

É muito interessante imaginar, do ponto de vista físico, o que significa uma autocorrelação como a mostrada em (2.15): ela indica que o processo *não terá qualquer correlação* com uma versão atrasada de si mesmo, *não importando quão ínfimo seja o atraso*. Em outras palavras, trata-se de um caso extremo de variação temporal, de uma mudança instantânea que, na prática, não se verificará<sup>13</sup>. Entretanto, a simplicidade matemática proporcionada por essa hipótese a torna muito atraente e justifica sua ampla aceitação em estudos na área de telecomunicações. Caso o ruído não tenha densidade espectral plana, será denominado *colorido*, ainda em respeito à analogia cromática apresentada na nota de rodapé. Um ruído branco pode se tornar colorido por meio de um processo de filtragem que elimine o caráter plano de sua DEP. Finalmente, para concluir este item, apresentamos na Fig. 2.9 uma realização de um ruído branco e gaussiano.



**Figura 2.9:** Ruído Branco e Gaussiano

<sup>12</sup> Em analogia com a luz branca, que contém, em proporção equitativa, todas as cores do espectro.

<sup>13</sup> O que nos lembra a frase de Leibniz no prefácio a seus “Novos Ensaio sobre o Entendimento Humano”: “*la nature ne fait jamais des sautes*”, ou seja, na tradução de L. Baraúna, “*a natureza nunca faz saltos*”.

## Ruído Impulsivo

Embora o ruído gaussiano tenha uma importância inegável, há alguns fenômenos cujos efeitos não se encaixam no esquema que traçamos no item anterior por não poderem ser entendidos como uma sobreposição de flutuações *de variância finita*. Em outras palavras, esses fenômenos produzem eventos de caráter mais abrupto i.e. *provocam variações mais pronunciadas em menores intervalos de tempo*, ou seja, *impulsivas*.

Dados experimentais indicam que a densidade de probabilidade associada a processos impulsivos tem uma aparência semelhante à gaussiana, com um formato “de sino”, simétrico e com curvas suaves [Georgiadis, 2000]. Entretanto, há uma característica importante: a cauda assintótica da densidade medida decai “menos rapidamente” que a de uma gaussiana, o que se reflete numa maior chance de ocorrência de grandes picos, a qual, como expusemos, é a mais marcante característica de um ruído impulsivo. Neste trabalho, concentrar-nos-emos numa importante densidade de probabilidade que se adequa ao padrão estabelecido pelos ensaios experimentais, razão pela qual tem atraído nos últimos anos a atenção de vários pesquisadores [Kosko, 1996; Georgiadis, 2000; Kuruoglu et al., 1998a]: a densidade *alfa-estável*.

A densidade alfa-estável pode ser expressa matematicamente da seguinte forma:

$$p_\alpha(x) = \mathfrak{F}^{-1}[\Phi_\alpha(\omega)] \quad (2.16)$$

sendo

$$\Phi_\alpha(\omega) = \exp\left\{j\delta\omega - \gamma|\omega|^\alpha [1 + j\beta \operatorname{sgn}(\omega)\rho(\omega, \alpha)]\right\} \quad (2.17)$$

e

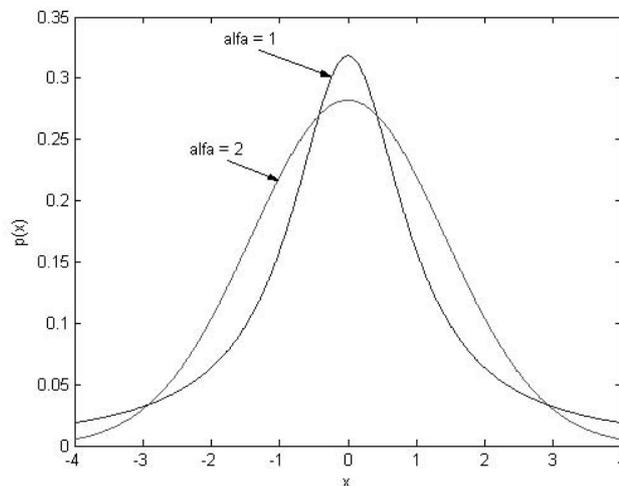
$$\rho(\omega, \alpha) = \begin{cases} \operatorname{tg}\left(\frac{\alpha\pi}{2}\right), & \alpha \neq 1 \\ \frac{2}{\pi} \ln(|\omega|), & \alpha = 1 \end{cases} \quad (2.18)$$

São importantes alguns comentários:

- O parâmetro  $\alpha$  é responsável por controlar o “peso” das caudas da distribuição, ou seja, determinar quão impulsivo será o ruído. São aceitos valores maiores que zero e menores ou iguais a dois, sendo que, no caso extremo superior,  $p_\alpha(x)$  *será uma densidade gaussiana*, ou seja, o caráter impulsivo será mínimo. Ao contrário, valores perto do mínimo corresponderão a uma valorização da ocorrência de picos pronunciados.

- O parâmetro  $\beta$  controla a simetria da densidade, pertencendo ao intervalo  $[-1;1]$ . A configuração simétrica ocorre para  $\beta = 0$ .
- O parâmetro  $\gamma$  é uma dispersão, ou seja, tem um papel análogo ao da variância. Se a distribuição alfa-estável tiver se transformado em uma gaussiana ( $\alpha = 2$ ),  $\sigma^2 = 2\gamma$ .
- O parâmetro  $\delta$  é a média da densidade quando  $\alpha > 1$  ou  $\beta = 0$ . Se  $\alpha$  for menor que  $-1$  e  $\beta$  for diferente de zero, a média da densidade será infinita [Georgiadis, 2000].

Há, portanto, quatro parâmetros essenciais para caracterizar uma função do tipo alfa-estável, sendo de especial valor para nós os casos em que  $\beta = \delta = 0$ , condições usuais em aplicações como as que nos interessam aqui. A Fig. 2.10 traz representações de densidades desse tipo.



**Fig. 2.10:** Densidades para  $\alpha=1$  (Cauchy) e  $\alpha=2$  (Gaussiana)

Passemos à análise dos momentos de uma variável aleatória alfa-estável. Se  $X$  é uma variável desse tipo e  $0 < \alpha < 2$ , é possível afirmar que:

$$E\left[|X|^p\right] = \infty, p \geq \alpha \quad (2.19)$$

e

$$E\left[|X|^p\right] < \infty, p < \alpha \quad (2.20)$$

Se  $\alpha = 2$ , teremos:

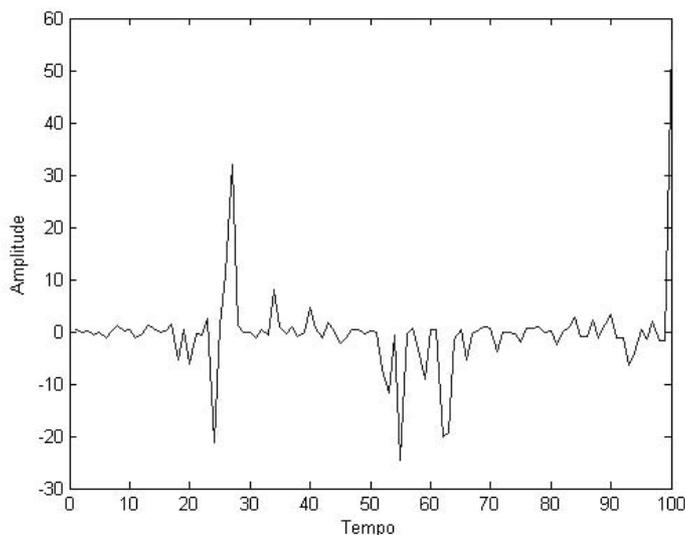
$$E\left[|X|^p\right] < \infty, \forall p > 0 \quad (2.21)$$

O leitor habituado a densidades “bem-comportadas” como a gaussiana poderá se sentir surpreso ao encontrar uma densidade que, via de regra, tem variância infinita, e, além disso, pode ter, em alguns casos, uma média infinita. Tudo isso, frisamos, decorre da “leveza” da cauda de uma densidade alfa-estável.

Tendo exposto as principais características da densidade que escolhemos, consideramos importante justificar nossa opção. A principal base para isso é o *teorema central do limite generalizado*, que pode ser enunciado da seguinte forma:

***Teorema Central do Limite Generalizado:*** *Seja uma soma de variáveis aleatórias como a mostrada em (2.8). É possível mostrar que, quando  $N$  tende a infinito,  $V_N$  só existirá como distribuição limite se tiver a forma de uma distribuição estável. Caso as variáveis  $X_i$  tenham variância finita,  $V_N$  tenderá a uma gaussiana, coerentemente com o que indica o teorema clássico.*

O teorema mostra que, no sentido mais amplo, se houver convergência de uma soma de variáveis aleatórias, a densidade produzida deverá ser estável, o que empresta enorme solidez a essa família de densidades. Para concluir o item, apresentamos na Fig. 2.11 uma realização de um processo gerador de um ruído impulsivo alfa-estável<sup>14</sup>. Uma comparação com a Fig. 2.9 explicita a natureza impulsiva do processo.



**Figura 2.11:** Processo Impulsivo ( $\alpha = 1$ ,  $\beta = 0$ ,  $\gamma = 1$ ,  $\delta = 0$ )

<sup>14</sup> Os usuários de MATLAB podem encontrar um utilíssimo programa que gera processos estáveis em <http://www.econ.ohio-state.edu/jhm/jhm.html>

## 2.10. Sistemas Digitais: Detecção Ótima de Sinais em Ruído Branco

Tendo enviado a informação por um canal qualquer, o próximo passo é recuperá-la no receptor. Como há um repertório finito de possíveis formas de onda, torna-se necessário que haja um mecanismo racional de detecção, ou seja, um esquema que indique a qual dos componentes da modulação digital escolhida corresponde o sinal recebido em um dado período de tempo. Em outras palavras, é preciso reconstruir a seqüência de símbolos enviada a partir de uma seqüência de pulsos ruidosos.

Consideremos que não há interferência intersimbólica (ou que ela foi eliminada), e concentremos nossa atenção em um canal composto exclusivamente por ruído aditivo, gaussiano e branco. Nesse caso, teremos um conjunto de  $N_S$  pulsos de duração  $T$  (o período de símbolo) aos quais se sobrepõe um processo estocástico que não guarda qualquer correlação com eles, mas que termina por distorcê-los. Matematicamente, o sinal recebido terá a seguinte forma [Haykin, 1994b]:

$$x_i(t) = s_i(t) + r_i(t) = \sum_{k=1}^{N_B} s_{ik} \phi_k(t) + r_i(t), \quad 0 \leq t \leq T \text{ e } i = 1, 2, \dots, N_S \quad (2.22)$$

onde  $\phi_j(t)$  representa uma das  $N_B$  funções associadas à modulação (essas funções, juntamente com a limitação temporal expressa pelo período  $T$ , formam o pulso) e  $s_{ij}$  é o conteúdo do  $i$ -ésimo símbolo no domínio de  $\phi_j(t)$ . O primeiro termo do segundo membro da de (2.22) indica o conteúdo de  $x_i(t)$  devido ao sinal:

$$s_i(t) = \sum_{k=1}^{N_B} s_{ik} \phi_k(t), \quad 0 \leq t \leq T \text{ e } i = 1, 2, \dots, N_S \quad (2.23),$$

Esse conteúdo pode ser perfeitamente expresso através do conjunto de valores  $s_{ij}$ . Assumindo que as funções da modulação são ortonormais, ou seja, que valem as expressões

$$\int_0^T \phi_i(t) \phi_j(t) dt = 1, \text{ se } i = j \quad (2.24)$$

e

$$\int_0^T \phi_i(t) \phi_j(t) dt = 0, \text{ se } i \neq j \quad (2.25),$$

chegamos à seguinte expressão para  $s_{ij}$ :

$$s_{ij} = \int_0^T s_i(t) \phi_j(t) dt \quad (2.26)$$

Em termos simples, todo esse formalismo nos mostra que o sinal  $s_i(t)$  pode ser entendido como um ponto em um espaço formado pela base de funções  $\phi_i$ . As coordenadas desse

ponto, ou seja, os valores  $s_{ij}$ , formam um vetor que denominaremos  $s_i$ , o qual, portanto, passa a ser uma representação de um dado pulso do repertório escolhido.

Como (2.26) nos mostra, a integral definida em um período do produto entre um sinal e uma função da base nos revela a coordenada “no eixo” associado a tal função. Em outras palavras, essa *correlação* entre sinais tem um papel análogo ao de um *produto escalar* num espaço vetorial, sendo muito útil para que se decomponha o sinal e se possa, assim, compreender a informação nele embutida. Parece natural, portanto, que empreguemos essa ferramenta para extrair do sinal contaminado  $x(t)$  a informação que nos interessa, ou seja, o vetor  $s_i$ , a partir do qual é possível recuperar a mensagem enviada. A grande questão é entender como o ruído perturbará o processo.

Definamos, nesse espírito, o componente  $x_{ij}$  i.e. o fruto da projeção do sinal recebido na direção de um elemento  $\phi_j$  qualquer da base:

$$x_{ij} = \int_0^T x_i(t)\phi_j(t)dt = \int_0^T [s_i(t) + r_i(t)]\phi_j(t)dt = s_{ij} + \int_0^T r_i(t)\phi_j(t)dt = s_{ij} + r_{ij} \quad (2.27)$$

Percebe-se que há dois componentes bem definidos: a informação desejada  $s_{ij}$  e um termo indesejado  $r_{ij}$ . Já poderemos começar a intuir como a presença de ruído influencia a detecção se imaginarmos que todos os produtos escalares serão afetados por termos dessa espécie, os quais, é lícito imaginar, devem ser tão mais nocivos quanto maior for o patamar associado à densidade espectral de potência plana do ruído. Para testar essas idéias iniciais, convém que recorramos a uma análise estatística de  $x_{ij}$ , ou seja, que estudemos as características essenciais do componente ruidoso. Definamos um processo  $\xi(t)$ :

$$\xi(t) = \int_{-\infty}^{\infty} h(\tau)\chi(t-\tau)d\tau = \int_0^T h(\tau)\chi(t-\tau)d\tau \quad (2.28),$$

que é o resultado da filtragem de um processo gaussiano  $\chi(t)$  por um dispositivo linear com resposta ao impulso limitada entre 0 e T e igual a  $h(t)$ . É possível mostrar [Papoulis, 1991] que  $\xi(t)$  também será um processo gaussiano. Ora, comparando as equações (2.28) e (2.27), vemos que, sendo:

$$r(t) = \chi(-t) \quad (2.29)$$

e

$$h(t) = \phi_j(t), \quad 0 \leq t \leq T, \quad (2.30)$$

teremos que  $r_{ij} = \xi(0)$ , o que demonstra ser  $r_{ij}$  *uma variável aleatória gaussiana*<sup>15</sup>. Sendo  $s_{ij}$ , do ponto de vista dessa análise, uma grandeza *determinística e fixa*, deduzimos que  $x_{ij}$  também será uma variável gaussiana, o que nos poupa muito trabalho, pois poderemos nos ater ao estudo de duas grandezas: média e variância.

Como em comunicações usualmente se assume que o ruído tem média zero, é possível mostrar que a média de  $r_{ij}$  também será zero, pois o resultado da filtragem linear de

<sup>15</sup> Sendo o ruído gaussiano predominante em sistemas de comunicação, não é necessário considerar outros modelos aqui. A detecção de outros ruídos nas aplicações reinantes se dará como se eles fossem gaussianos.

um processo de média nula é um processo com a mesma característica [Papoulis, 1991]. Dessa forma, teremos que

$$E[x_{ij}] = E[s_{ij} + r_{ij}] = s_{ij} + E[r_{ij}] = s_{ij} \quad (2.31)$$

ou seja, que a média de  $x_{ij}$  será exatamente o componente de informação  $s_{ij}$ .

Define-se a variância como:

$$\text{var}[x_{ij}] = E[(x_{ij} - s_{ij})^2] = E[r_{ij}^2] \quad (2.32)$$

É possível mostrar [Haykin, 1994b] que, sendo o processo  $r(t)$  um ruído branco, a variância de  $x_{ij}$  será:

$$\text{var}[x_{ij}] = N_0/2 \quad (2.33)$$

sendo  $N_0/2$  o valor constante da densidade espectral de potência do ruído. Pode-se mostrar ainda que a covariância entre dois componentes distintos  $x_{ij}$  e  $x_{ik}$  será sempre nula, o que revela, no contexto de variáveis gaussianas, que eles são *independentes*.

Acabamos de mostrar que  $x_i(t)$  pode ser expresso por meio de um vetor  $\mathbf{x}_i$  formado pela soma do vetor  $\mathbf{s}_i$ , que contém a informação que se deseja recuperar, e um vetor de ruído  $\mathbf{r}_i$ , com média nula e variância igual a  $N_0/2$ :

$$\mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iN_B} \end{bmatrix} = \mathbf{s}_i + \mathbf{r}_i = \begin{bmatrix} s_{i1} \\ s_{i2} \\ \vdots \\ s_{iN_B} \end{bmatrix} + \begin{bmatrix} r_{i1} \\ r_{i2} \\ \vdots \\ r_{iN_B} \end{bmatrix} \quad (2.34)$$

Num sistema digital, há um número finito  $N_S$  de possíveis símbolos, ou seja, de vetores  $\mathbf{s}_i$ . A tarefa de detecção é, portanto, decidir, a partir de um certo vetor  $\mathbf{x}_i$  recebido, qual foi o vetor  $\mathbf{s}_i$  enviado. Em outras palavras, deseja-se determinar qual a estimativa de  $\mathbf{s}_i$  que reduz a probabilidade de erro de decisão e, conseqüentemente, torna o sistema tão confiável quanto possível. Quando a probabilidade de envio de cada possível símbolo é igual, ou seja, num cenário equiprovável, o critério de *mínima probabilidade de erro*<sup>16</sup> equivale ao *critério de máxima verossimilhança (MV)* [Papoulis, 1991]:

$$\max_a p_{\mathbf{x}}(\mathbf{x}_i / \mathbf{s}_i = \mathbf{s}_a), a = 1, \dots, N_S \quad (2.35)$$

Em termos simples, é preciso escolher o vetor  $\mathbf{s}_a$  que, no papel de  $\mathbf{s}_i$ , maximize a probabilidade condicional de recebimento do vetor  $\mathbf{x}_i$ . Essa maximização responde pela idéia de “mais verossímil”.

<sup>16</sup> O critério de mínima probabilidade de erro é equivalente a um outro critério muito importante e famoso: o critério de *máxima probabilidade a posteriori (MAP)*.

Tendo em vista as características de nosso modelo de ruído, podemos desenvolver (2.35):

$$p_x(\mathbf{x}_i / \mathbf{s}_i = \mathbf{s}_a) = \frac{1}{\sqrt{(\pi N_0)^{N_B}}} \exp\left[-\frac{\|\mathbf{x}_i - \mathbf{s}_a\|^2}{N_0}\right] \quad (2.36)$$

Também podemos tirar proveito do fato de a função logaritmo ser monotonicamente crescente e maximizar a seguinte função  $f_x$  em vez da função  $p_x$ :

$$f_x(\mathbf{x}_i / \mathbf{s}_i = \mathbf{s}_a) = \ln[p_x(\mathbf{x}_i / \mathbf{s}_i = \mathbf{s}_a)] = \frac{-N_B}{2} \ln(\pi N_0) - \frac{1}{N_0} \|\mathbf{x}_i - \mathbf{s}_a\|^2 \quad (2.37)$$

A maximização de (2.37), por sua vez, equivale a uma minimização da distância

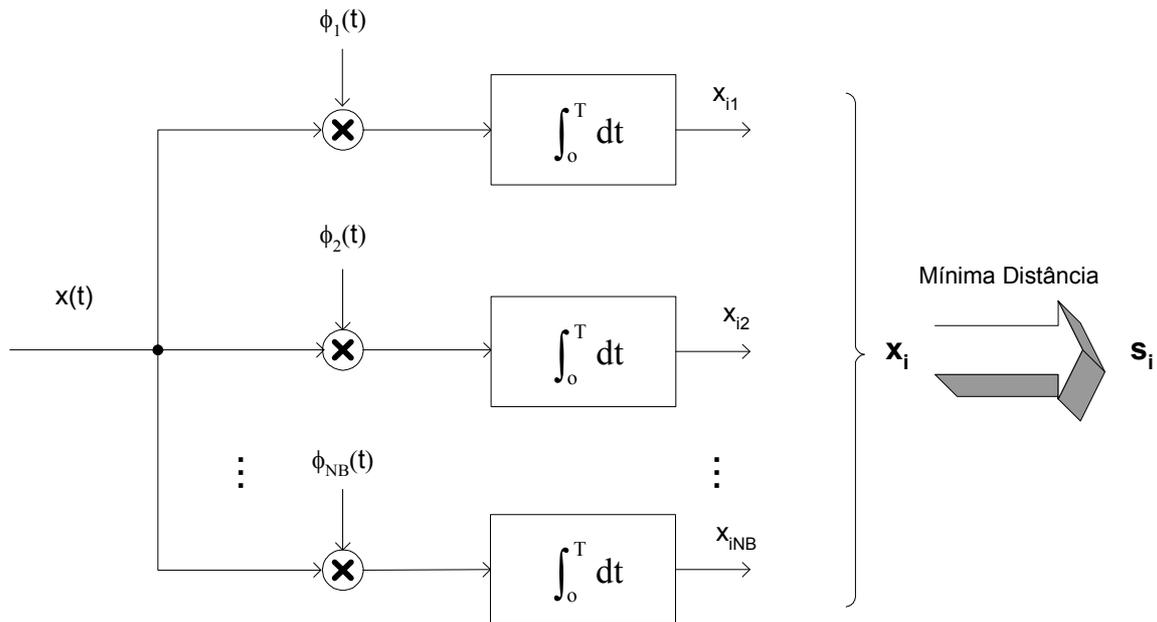
$$d = \|\mathbf{x}_i - \mathbf{s}_a\| \quad (2.38)$$

Com isso, o critério para estimação do símbolo transmitido será:

*Escolha o vetor  $\mathbf{s}_a$  pertencente ao alfabeto da modulação que minimiza a distância*  
 $d = \|\mathbf{x}_i - \mathbf{s}_a\|$

Dessa maneira, a escolha de uma certa modulação digital, com seu conjunto de vetores  $\mathbf{s}_a$ , produz *regiões de decisão* no espaço do sinal i.e. *lugares geométricos cuja ocupação por parte do vetor observado  $\mathbf{x}_i$  leva necessariamente à opção por um dos possíveis símbolos transmitidos.*

Tendo em mãos tal critério, chegamos finalmente a um esquema fechado de receptor, formado por uma etapa centrada no cálculo da correlação para obter o vetor  $\mathbf{x}_i$  e por outra em que ocorre a escolha propriamente dita. A Fig. 2.12 traz um esquema desse processo.



**Figura 2.12:** Esquema do Receptor de Correlação

Note o leitor que, no processo de correlação, implicitamente houve um estágio de *demodulação*, caracterizado pelo produto do sinal recebido pelas funções da base e por uma integração que podemos entender, em termos simples, como uma filtragem. Vale ainda frisar que há outras formas de compreender esse processo de recepção, sendo particularmente relevante aquela baseada na noção de *filtro casado*. Nesse caso, entende-se o processo de cálculo da correlação como uma filtragem por um dispositivo cuja resposta ao impulso é uma versão refletida da função base seguida de uma amostragem precisa. Matematicamente, ambas as concepções são perfeitamente equivalentes, de modo que julgamos apropriado não estender mais essa discussão.

Após esse estágio de reconstrução da seqüência de símbolos, é preciso efetuar os processos de decodificação de canal e de fonte, sendo finalmente obtida uma versão da mensagem original. Isso conclui nossa exposição genérica dos principais elementos de um sistema de comunicação digital. A Fig. 2.13 fornece um sumário das etapas apresentadas.

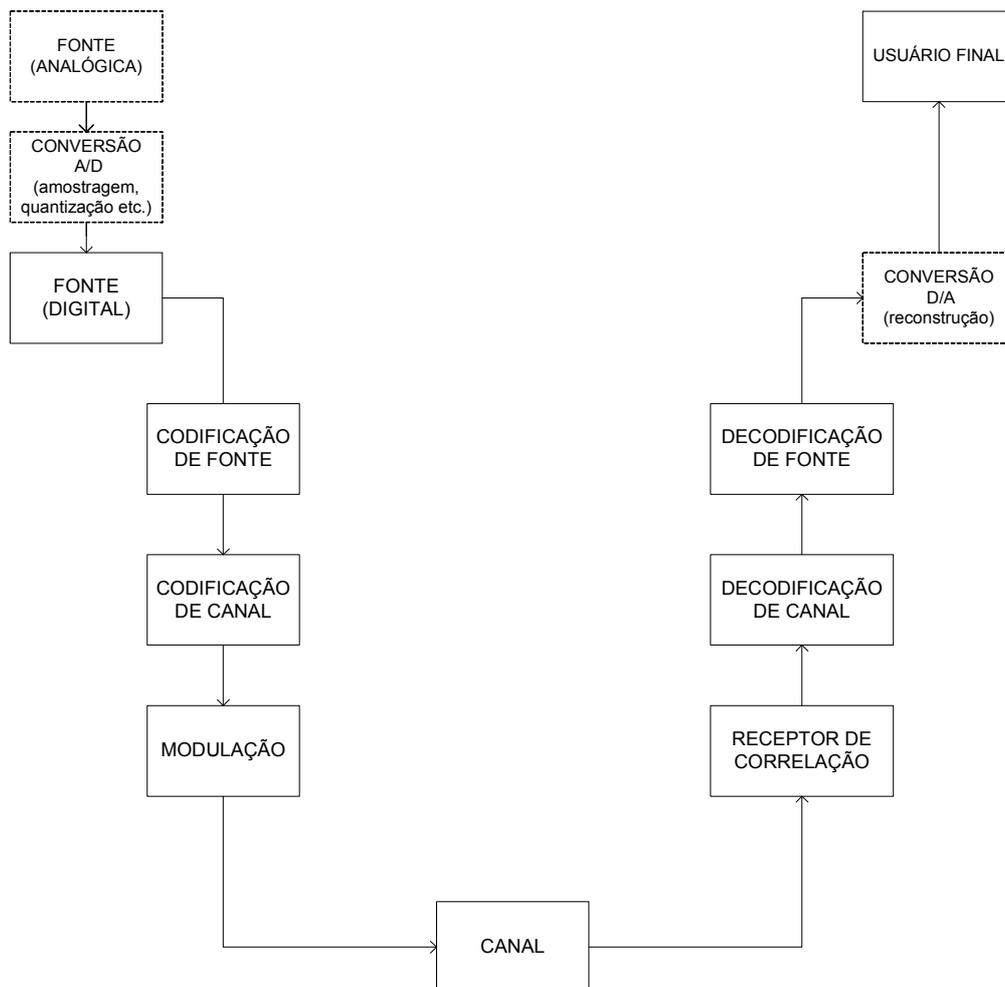


Figura 2.13: Esquema Genérico de um Sistema Digital<sup>17</sup>

## 2.11. Modelo Discreto de Canal

Nas considerações que tecemos acerca do canal, mantivemo-nos no universo de sinais contínuos, tendo sempre em vista que a transmissão real se baseia neles. No entanto, para fins de estudo e análise, é bastante útil tirar proveito do fato de lidarmos com um sistema que transmite informação em períodos de amostragem bem determinados. Uma consequência importante é que se torna possível trabalhar com um *modelo discreto de canal*, ou seja, com uma versão amostrada da resposta ao impulso do meio e também do processo estocástico que modela o ruído. A razão de ser disso é, em última análise, o fato de o receptor efetuar um processo de amostragem (que, como indicamos, está implícito na obtenção do vetor  $\mathbf{x}_i$  no receptor de correlação ou no filtro casado). Se a reconstrução é feita

<sup>17</sup> Os blocos pontilhados indicam elementos de sistemas baseados na conversão de sinais analógicos. Tais estágios não são necessários quando ocorre um intercâmbio de dados digitais “por natureza” (como os trocados por computadores, por exemplo).

a partir das amostras, basta considerar os elementos do sistema que contribuem para a determinação de seu valor, não sendo necessário levar em conta os demais instantes. Isso, em termos mui simples, significa que podemos representar o sinal modulado por uma seqüência e o canal por um sistema discreto com ruído. Caso a modulação empregada utilize a banda passante, ou seja, empregue senóides, será necessário obter uma *representação equivalente em banda-base* para todos elementos do sistema. O motivo dessa estratégia é efetuar uma translação de espectro para fins de análise, o que, eventualmente, origina modelos baseados em números complexos, similares, aliás, à conhecida representação fasorial de circuitos de corrente alternada.

### 2.11.1. Sinal Transmitido

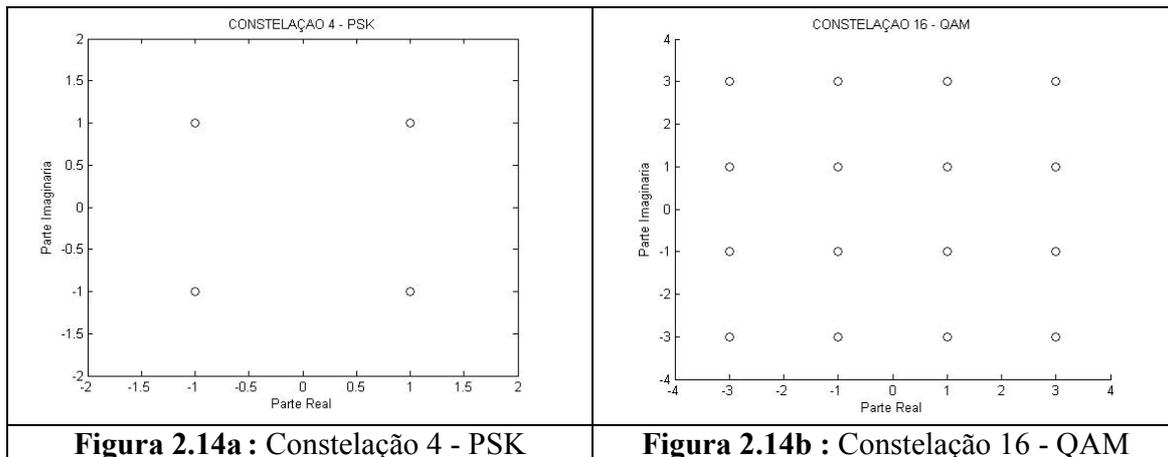
Concentremos nossa atenção, primeiramente, no sinal transmitido. Segundo o que foi dito anteriormente, ele poderá ter por base diversas modulações, tanto em banda-base quanto em banda passante. Analisemos, com a ajuda da Tab. 2.1, como será o modelo discreto do sinal transmitido para um grupo de modulações usuais em sistemas de comunicação digital.

<b>Modulação</b>	<b>Possíveis Símbolos no Modelo Discreto</b>
2 – PAM	$s_1 = -1$ e $s_2 = +1$
2 – PSK	$s_1 = -1$ e $s_2 = +1$
4 – PSK	$s_1 = -1 - j$ , $s_2 = -1 + j$ , $s_3 = 1 - j$ e $s_4 = 1 + j$

**Tabela 2.1:** Exemplos de Símbolos Transmitidos

Note o leitor que, no modelo discreto de canal, uma onda formada por diversos pulsos concatenados é representada por uma seqüência de valores reais ou complexos, dependendo da modulação adotada.

Uma forma muito importante de visualizar o conjunto de símbolos pertencentes a uma modulação é através de uma representação no plano complexo. Os símbolos da modulação 4 – PSK apresentados na Tab. 2.1, por exemplo, dão origem à seguinte configuração:



A representação apresentada na Fig. 2.14a mostra por que o conjunto de símbolos do modelo discreto de uma modulação digital recebe também o nome de *constelação*. Essa analogia fica ainda mais patente quando há mais símbolos, conforme mostra a Fig. 2.14b, que traz os símbolos de uma modulação 16-QAM.

### 2.11.2. Canal

A seqüência transmitida é o sinal de entrada do canal, que, usualmente, inclui em seu bojo os filtros responsáveis pela conformação de pulso. Conforme discutimos na seção 2.9, há dois efeitos que nos interessam em particular: a interferência intersimbólica, cujo modelo tem caráter essencialmente determinístico, e o ruído aditivo, de caráter estocástico. No modelo discreto, o primeiro efeito é modelado através de um processo de filtragem, enquanto o segundo é obtido por meio de um processo discreto que se soma ao sinal filtrado. Matematicamente, o sinal recebido pode ser expresso da seguinte forma:

$$x(n) = F[s(n)] + r(n) \quad (2.39)$$

sendo  $F[.]$  um operador que representa a filtragem e  $r(n)$  o processo estocástico que produz as amostras de ruído. Para que entendamos as propriedades do sistema dinâmico que relaciona entrada e saída, é preciso considerar com especial atenção o operador  $F[.]$ , responsável pela IIS. Como vimos na seção 2.9.1, a interferência intersimbólica é causada pela sobreposição de pulsos associados a instantes de amostragem distintos. Isso quer dizer que um valor atual hipotético do sinal recebido,  $x(n_0)$ , corresponderá não apenas a uma única amostra do sinal transmitido, como esperaríamos caso o canal fosse ideal, mas a uma sobreposição de amostras associadas a instantes de tempo distintos de  $s(n)$ . Matematicamente, isso significa que

$$F[s(n)] = C[s(n), s(n-1), \dots, s(n-N)] \quad (2.40)$$

sendo  $N$  um valor inteiro que representa a *ordem* do operador  $C[.]$ .

Há inúmeras formas por que o operador  $C[.]$  pode combinar as amostras do sinal transmitido para gerar o modelo de IIS, sendo possível enquadrá-las todas em duas classes principais:

Canais Lineares: São aqueles em que o operador  $C[.]$  é *linear*, ou seja, efetua apenas uma soma ponderada dos sinais  $s(n)$ ,  $s(n-1)$  etc.<sup>18</sup>. Alguns exemplos são:

$$C_{L1}(n) = s(n) + 0.5s(n-1)$$

$$C_{L2}(n) = 0.4s(n-2) + s(n-5) + 0.77s(n-20)$$

Canais Não-lineares: São aqueles em que o operador  $C[.]$  é *não-linear*, ou seja, combina os sinais  $s(n)$ ,  $s(n-1)$  etc. de uma maneira tal que o princípio da superposição não seja obedecido. Exemplos:

$$C_{NL1}(n) = s(n) + 0.34s(n)s(n-1) + s(n-2)$$

$$C_{NL2}(n) = s(n) + 0.6s(n-1) + 0.3[s(n) + 0.6s(n-1)]^2 + 0.1[s(n) + 0.6s(n-1)]^3$$

$$C_{NL2}(n) = \zeta[s(n)+0.2s(n-1) + 0.4s(n-3)]$$

sendo  $\zeta[.]$  uma função não-linear sem memória.

### Canais Lineares

Não seria exagero dizer que modelos lineares de canal predominam em boa parte das aplicações presentes no mundo moderno. Esses modelos são capazes de representar adequadamente fenômenos muito importantes, como a propagação por múltiplos percursos em um canal de rádio móvel, a existência de ecos em um par de fios, a sobreposição de sinais provenientes de fontes distintas etc. Além disso, modelos lineares primam pela *simplicidade*, tanto de manipulação quanto de análise, oferecendo ao pesquisador uma ampla palheta de ferramentas e conceitos utilíssimos. Desconsiderando por ora o ruído, podemos expressar matematicamente um modelo linear da seguinte forma:

$$x(n) = -\sum_{k=1}^{N_a} a_k x(n-k) + \sum_{i=0}^{N_b} b_i s(n-i) \quad (2.41)$$

sendo  $s(n)$  o sinal de entrada e  $N_a$  e  $N_b$  índices que denotam a ordem do canal. A equação (2.41) contém dois termos distintos: um somatório de amostras passadas *da saída do canal* e um somatório de amostras *do sinal de entrada*. A primeira parcela indica uma recorrência i.e. aponta a possibilidade de um sistema linear ser dependente de valores passados de sua saída. Já a segunda parcela é responsável por uma IIS mais intuitiva, ou seja, aquela causada por uma sobreposição de amostras do sinal transmitido. A existência dessas duas possibilidades nos permite definir dois tipos de canal:

- *Canais IIR (infinite impulse response)*: são aqueles em que há recursão, ou seja, em que a resposta ao impulso tem duração infinita devido à memória de amostras passadas da saída.

<sup>18</sup> Essa soma pode conter infinitos termos, como veremos quando definirmos sistemas IIR.

- *Canais FIR (finite impulse response)*: são aqueles em que a saída depende apenas de valores passados da entrada, ou seja, em que não há recursão. Nesse caso, a resposta ao impulso desvanece após um determinado número de amostras.

Como o efeito predominante em comunicações é a existência de um espalhamento temporal causado por limitação de faixa ou pela existência de múltiplos caminhos de propagação, não são usuais modelos IIR. Entretanto, sua maior generalidade (eles podem se reduzir a um canal FIR, sendo o contrário impossível) torna-os atraentes quando se deseja testar soluções que melhorem o desempenho de um sistema de comunicação. Além disso, um modelo IIR pode ser útil se o canal tiver uma resposta ao impulso finita mas excessivamente longa.

Concluída essa breve digressão, voltemos a (2.41). Ela indica a atuação do canal no domínio do tempo, das amostras tomadas em seqüência cronológica. Há outras formas de observar o mesmo efeito em outros domínios, domínios transformados em que determinadas características podem ser observadas com mais clareza. Quando lidamos com sistemas discretos, adquire particular importância a *transformada Z*, que pode ser entendida como a contraparte discreta da célebre transformada de Laplace [Oppenheim et al., 1997]. A transformada *Z* de uma seqüência  $u(n)$  qualquer é dada por:

$$U(z) = \sum_{k=-\infty}^{\infty} u(k)z^{-k} \quad (2.42)$$

O estudo dessa transformada poderia ser tema de um capítulo inteiro, ou mesmo de um livro. Sendo nossos objetivos nesse campo bastante modestos, não nos aprofundaremos nos pormenores relacionados a assuntos como a convergência do somatório mostrado em (2.42), a qual definirá um domínio para a variável  $z$ . Como um canal de comunicação deve ser, necessariamente, *causal*, ou seja, *deve ter uma saída que dependa apenas do valor presente e dos valores passados da entrada*, as possibilidades de análise se restringem bastante, sendo possível retomar o assunto apenas pontualmente.

Considerando a linearidade da transformada e a propriedade

$$Z[u(n-1)] = U(z).z^{-1} \quad (2.43)$$

onde  $Z[.]$  é um operador que indica a operação expressa em (2.42), podemos aplicar a transformação à definição genérica apresentada em (2.41):

$$X(z) = -X(z) \sum_{k=1}^{N_a} a_k z^{-k} + S(z) \sum_{i=0}^{N_b} b_i z^{-i} \quad (2.44)$$

Ordenando convenientemente os termos acima, obtemos uma relação mui simples entre a transformada da saída e a da entrada:

$$X(z) = \frac{\sum_{i=0}^{N_b} b_i z^{-i}}{1 + \sum_{k=1}^{N_a} a_k z^{-k}} S(z) = H(z)S(z) \quad (2.45)$$

sendo  $H(z)$  a *função de transferência* do sistema linear e invariante no tempo. Vale lembrar que essa função é a transformada  $Z$  da *resposta ao impulso do canal*, à qual daremos o nome de  $h(n)$ , que é simplesmente a saída do canal quando a entrada é igual à função impulso  $\delta(n)$ . A contraparte temporal da equação (2.45) pode ser expressa como:

$$x(n) = s(n)*h(n) \quad (2.46)$$

sendo ‘\*’ o operador *convolução*, definido no contexto discreto como:

$$x(n) = \sum_{k=-\infty}^{\infty} h(k)s(n-k) = \sum_{k=-\infty}^{\infty} s(k)h(n-k) \quad (2.47)$$

### **Análise dos pólos e zeros de um canal linear e invariante no tempo**

A análise dos pólos e zeros de  $H(z)$  é uma forma clássica de caracterizar um sistema linear discreto. A referência do estudo é usualmente o chamado *círculo de raio unitário (CRU)*<sup>19</sup>, o lugar geométrico no plano complexo em que  $|z| \leq 1$ . Intuitivamente, tal referência tem a ver com o fato de que a soma de uma série exponencial baseada em um número com módulo menor que um converge.

Consideremos em primeiro lugar os pólos de  $H(z)$ . Dada a restrição de causalidade, não é difícil demonstrar que, se todos eles estiverem no círculo unitário, o sistema será *estável*, ou seja, terá uma saída limitada se a entrada for limitada [Oppenheim et al., 1997]. Em termos muito simples, isso quer dizer que, se a entrada não “explodir para o infinito”, a saída também não o fará, pois o sistema não terá mecanismos que causem um “efeito bola de neve”<sup>20</sup>.

Já a posição dos zeros abre três possibilidades:

- *Canais de Fase Mínima*: são aqueles cujos zeros se encontram no CRU
- *Canais de Fase Máxima*: são aqueles cujos zeros se encontram fora do CRU
- *Canais de Fase Mista*: são aqueles que têm zeros tanto no CRU quanto fora dele

A origem dos nomes associados a esses canais não deve preocupar o leitor: ela se liga a uma propriedade relacionada à transformada de Fourier discreta, propriedade esta que não seria conveniente discutir nesse momento. Entretanto, é muito importante definir as três situações acima, pois elas terão um papel decisivo quando estudarmos o problema de equalização.

<sup>19</sup> Quando for conveniente, tomaremos por referência a circunferência de raio unitário.

<sup>20</sup> I.e. uma realimentação positiva.

## **Transformada de Fourier**

Caso a circunferência de raio unitário pertença ao domínio da transformada  $Z$ , como será a regra nas situações aqui estudadas, é bastante simples obter a transformada de Fourier a partir da transformada  $Z$ . Basta que se faça a restrição  $z = \exp(j2\pi f)$ . Dessa forma, podemos voltar a (2.45) e obter uma relação entre as transformadas de Fourier da entrada e da saída de um dado canal de comunicação:

$$X(f) = H(f)S(f) \quad (2.48)$$

i.e. a transformada de Fourier da saída de um canal de comunicação é o produto entre a transformada de Fourier da entrada e a transformada de sua resposta ao impulso, que recebe o nome de *resposta em frequência do canal*. Tal grandeza tem enorme importância, pois indica como o canal responde a um estímulo numa frequência qualquer, o que permite que se entenda como o canal moldaria um conteúdo espectral hipotético.

Sem exagero algum, afirmamos que a análise em frequência é o “pão de cada dia” do engenheiro de comunicações, pois possibilita que ele, que deve sempre lidar com restrições sistêmicas relacionadas a conteúdo espectral, avalie instantaneamente como atuará o canal de comunicação em qualquer faixa ou frequência. Ele será capaz de saber se o canal corta altas frequências, se é plano em uma faixa qualquer, se rejeita determinadas regiões etc., decorrendo tudo isso da relação simples expressa em (2.48), a qual envolve apenas um produto de números complexos, ou seja, produto de amplitudes e soma de fases.

## **Canais Não-lineares**

*Canais não-lineares são aqueles formados por sistemas que não obedecem ao princípio da superposição, ou seja, cuja resposta a uma sobreposição de entradas não é a sobreposição das respostas às entradas isoladas.* Uma definição como essa, negativa por natureza, dá origem a uma classe cujo estudo é bem mais complexo e cujas possibilidades são muito mais amplas.

Nas últimas quatro décadas, idéias inovadoras e instigantes como a constatação da existência de comportamento *caótico* em uma ampla gama de modelos de sistemas reais mostrou a importância de aprofundar as bases do estudo de sistemas não-lineares, uma vez que determinados comportamentos não mais poderiam ser entendidos a partir de *linearizações*. O concomitante avanço em áreas como a neurociência, que depende fortemente de modelos neurais inerentemente não-lineares, deu origem a um novo arcabouço teórico que tornou possível emular e mesmo analisar modelos que, de outra forma, seriam insatisfatoriamente “enquadrados” em pequenos domínios lineares.

Não tardou para que esse novo impulso chegasse à área de comunicações, em duas frentes distintas: modelos mais precisos de canais e técnicas mais eficientes na tarefa de melhorar a qualidade da transmissão digital. Como veremos no capítulo 3, essas duas frentes se relacionam a três problemas centrais: *equalização, identificação e predição*. Por ora, atenhamo-nos exclusivamente a considerações acerca de canais de comunicação, tema deste capítulo. Nesse contexto, a introdução de não-linearidades permitiu que fossem adequadamente levados em conta efeitos de dispositivos eletrônicos como os amplificadores empregados em comunicação via satélite [Ibnkahla e Castanié, 1996], que,

por extrema necessidade de economizar recursos, operam em uma zona de saturação [Benedetto et al., 1979].

Dada a riqueza de comportamentos que podem emergir de um modelo não-linear, seria inapropriado estender excessivamente esse capítulo para buscar ferramentas com um grau de generalidade similar ao daquelas que expusemos no contexto linear. Além disso, tais ferramentas não teriam grande importância para o tipo de análise que faremos neste trabalho. Sendo assim, optamos por tecer comentários pertinentes a cada modelo específico de canal *à medida que ele surgir em nossa discussão*. Por ora, basta que o leitor tenha uma idéia geral de que canais não-lineares são mais difíceis de analisar e produzem efeitos nocivos que exigem soluções mais complexas da parte do engenheiro de sistemas. Soluções desse tipo serão a tônica de nosso esforço.

## 2.12. Conclusões e Modelo Básico de Canal

O grande objetivo deste capítulo foi buscar, juntamente com o leitor, *construir* o modelo discreto em que doravante nos basearemos e, paralelamente, introduzir alguns conceitos-chave que serão úteis à nossa análise posterior. Vimos, numa breve perspectiva histórica, a emergência da comunicação digital e as principais etapas de transmissão, propagação e recepção num sistema desse tipo. Estudamos ainda os fundamentos matemáticos da detecção por máxima verossimilhança, que nos será útil outras vezes ao longo deste trabalho. Por fim, introduzimos as noções de *canal linear* e *canal não-linear*, com alguns comentários e conceitos acessórios pertinentes. Esse riquíssimo contraste será retomado diversas vezes em nossas discussões, para que sejam explicitadas as vantagens e desvantagens de cada possível estratégia.

Para encerrar este capítulo, expomos na Fig. 2.15 o diagrama do modelo básico de canal que empregaremos ao longo da tese. Suporemos, a menos que seja explicitado o contrário, que o transmissor gera uma seqüência de símbolos independentes e identicamente distribuídos, o que é aceitável à luz de processos existentes de *embaralhamento (interleaving)*. Na maioria das vezes assumiremos que a modulação é do tipo 2-PAM (ou 2-PSK em banda-base), um caso simples, mas que permite tirar conclusões com suficiente grau de generalidade. Em determinados casos, poderemos utilizar modulações mais complexas, especialmente quando nos ativermos a modelos de determinadas aplicações de “mundo real” que delas fazem uso. O canal será um sistema linear ou não-linear com ou sem ruído, dependendo do caso em questão. Finalmente, no receptor sempre haverá um dispositivo de decisão (ou decisor) que, como vimos na seção 2.10, adotará uma regra de distância mínima. Além de apresentar o diagrama de blocos sistêmico, a Fig. 2.15 já contém alguns elementos de nossa notação básica.

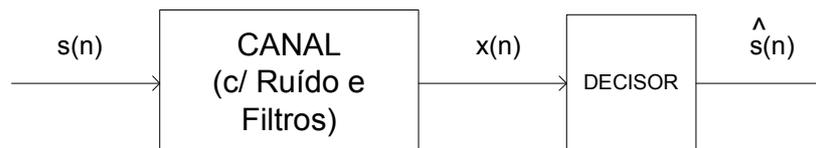


Figura 2.15: Modelo Discreto de Canal



## 3. Equalização, Identificação e Predição

### 3.1. Introdução

No capítulo anterior, apresentamos os principais elementos de um sistema de comunicação, partindo do ente que envia a mensagem (transmissor), e chegando àquele que recebe a informação enviada (receptor). Como já discutimos, a separação física entre ambos torna inevitável a existência de um canal, um meio que atua sobre o sinal transmitido e ocasiona modificações que podem tornar a comunicação impraticável. Tal fato exige do projetista contramedidas que eliminem ou, o que é mais comum, atenuem suficientemente os efeitos nocivos do canal, de modo a restaurar níveis apropriados de qualidade no intercâmbio de mensagens.

Vimos na seção anterior uma possível contramedida, a *codificação de canal*, amplamente utilizada nos sistemas reinantes. Entretanto, é de nosso interesse neste trabalho outra classe de técnicas, baseadas em noções da área de processamento de sinais e, mais particularmente, na idéia de *filtragem*. Sob a égide de tal idéia, deter-nos-emos no estudo de três problemas essenciais que, por assim dizer, formam a base dos esforços que relataremos ao longo deste trabalho: a *equalização*, a *identificação* e a *predição*. Definamos, com muita brevidade, cada um deles:

- *Equalização* é um processo pelo qual os efeitos nocivos do canal são mitigados através da filtragem do sinal recebido por um dispositivo adequadamente projetado para tal fim.
- *Identificação* é um processo em que se busca obter um modelo matemático adequado de uma planta ou, mais especificamente, de um dado canal de comunicação.
- *Predição* é um processo através do qual são obtidas estimativas de valores futuros de um sinal de informação tendo por base, exclusivamente, seu presente e seu passado<sup>21</sup>.

O leitor notará que os três problemas têm motivações bastante distintas, sendo o primeiro baseado diretamente na idéia de *contrabalançar os efeitos do canal*, o segundo na idéia de *entender tais efeitos*, e o terceiro na idéia de, num certo sentido, *antecipá-los*. A forma como os frutos da solução desses problemas serão aproveitados dependerá sempre das peculiaridades da aplicação. Isso significa que, além do mérito próprio dos processos, há a possibilidade de empregá-los de maneira hierárquica. Assim, poderíamos empregar o modelo identificado para equalizar o canal, ou poderíamos mesmo empregar para esse objetivo as características do dispositivo ajustado para predizer as amostras futuras.<sup>22</sup> Não seria inapropriado assumir que o caráter inverso da equalização possa ser empregado para

---

<sup>21</sup> É possível também que se deseje obter estimativas do passado de um sinal a partir de valores presentes ou futuros: trata-se do processo de *retro-predição*.

<sup>22</sup> O que traz à mente um aforismo de Sun Tzu: “*Determina então os planos do inimigo e saberás qual estratégia será bem-sucedida e qual não o será*”, in *A Arte da Guerra*, Editora Paz e Terra, 1997.

identificar o canal, ou que a predição possa servir a fim análogo.<sup>23</sup> Em suma, as possibilidades são imensas.

De qualquer forma, para cumprir qualquer uma das três tarefas apresentadas, será imprescindível efetuar, de alguma maneira, duas escolhas:

- A *escolha da estrutura de filtragem*, ou seja, do filtro que fará o papel de *equalizador*, *modelo* ou *preditor*. Trata-se de um passo muito importante, pois a capacidade de processamento da estrutura é o *limitante essencial* do desempenho que será obtido.
- A *escolha do critério de ajuste* dos parâmetros da estrutura escolhida. Nesse caso, podemos resumir o problema a: como extrair o melhor desempenho possível da estrutura?

Poderíamos, sem dúvida, afirmar que todo o nosso esforço subsequente se dará no terreno tão simplesmente delimitado por essas duas escolhas. Diríamos mesmo que toda a teoria que envolve a noção de *filtragem*, em qualquer de suas formulações, consiste na busca por respostas adequadas a essas duas perguntas. Para complicar o processo, é preciso ainda delimitar bem o significado do epíteto “adequado”, que pode ter acepções bem distintas para, por exemplo, o teórico e o projetista de uma empresa.

Eis os dois pólos por que doravante gravitaremos: conheçamo-los melhor. Analisemos primeiramente algumas classificações gerais das estruturas de filtragem, passando em seguida à delineação de nuances primordiais do problema de ajuste dos parâmetros.

### 3.2. Introdução às Estruturas de Filtragem

Damos o nome de *filtro*<sup>24</sup> a um sistema caracterizado por um mapeamento entre um conjunto de variáveis de entrada e um conjunto de variáveis de saída. Neste trabalho, como lidaremos exclusivamente com sistemas dotados de uma única entrada e uma única saída (*SISO*, *single-input / single-output*), o conjunto de variáveis de saída será unitário, e o conjunto de variáveis de entrada será formado por versões atrasadas de um mesmo sinal. A Fig. 3.1 ilustra essa definição.



**Figura 3.1:** Esquema de um Filtro SISO

<sup>23</sup> O leitor perceberá que essas combinações, aparentemente incidentais, formam a base de técnicas apresentadas posteriormente.

<sup>24</sup> O termo “filtro” tem a ver com a idéia de seletividade em frequência, ou seja, com a capacidade que o dispositivo tem de moldar o espectro do sinal de entrada.

sendo  $x(n)$  o sinal de entrada e  $y(n)$  o sinal de saída. A relação entre essas variáveis pode ser expressa da seguinte forma:

$$y(n) = F[x(n), x(n-1), \dots, x(n-N+1)] \quad (3.1)$$

onde  $F[.]$  é um mapeamento entrada-saída genérico e  $N$  é uma variável que indica a sua memória<sup>25</sup>. A esse mapeamento pode ser associado um conjunto de *parâmetros livres*, quer dizer, de valores ajustáveis cuja escolha estabelece suas características. Isso significa que a elaboração de  $F[.]$  requer que se determinem algumas relações entre as variáveis de entrada e, além disso, *alguns valores que controlam sua intensidade*.

Analogamente ao que vimos na seção 2.11.2, a escolha do mapeamento origina duas classes de particular relevo, delimitadas, em última análise, pela obediência ao princípio da superposição:

- **Estruturas Lineares:** são aquelas cujo mapeamento obedece ao princípio da superposição, ou seja, cuja resposta a uma combinação linear de entradas é a combinação linear das respostas a cada entrada.
- **Estruturas Não-lineares:** são aquelas cujo mapeamento não obedece ao princípio da superposição.

Tendo em mente os comentários feitos acerca de canais lineares na seção 2.11.2, podemos antecipadamente intuir as duas maiores qualidades de um filtro linear: estrutura matemática simples e tratável e enorme aparato analítico a ela associado. Entretanto, a simplicidade pode ser uma “faca de dois gumes”, pois implica restrição à flexibilidade do mapeamento  $F[.]$ . Com os filtros não-lineares, definidos como “tudo o que não é linear”, ocorre o contrário: são flexíveis, mas, por outro lado, matematicamente mais complexos.

Esse cenário estabelece, como é a regra em engenharia, um dilema, dilema com o qual, aliás, deparar-nos-emos continuamente daqui por diante: *complexidade versus desempenho*. É preciso levar em conta o desempenho requerido e o poder de processamento disponível para escolher a estrutura mais adequada ao problema que se deseja resolver. Para tanto, o engenheiro deve conhecer bem o potencial e a limitação de cada candidato, deve saber quão bem cada dispositivo pode cumprir uma tarefa e a que preço. Finalmente, precisa perceber com clareza os recursos que possui para levar a bom termo sua missão. Um primeiro passo nesse sentido é proceder a um exame mais detalhado das duas classes de filtros expostas.

### 3.2.1. Estruturas Lineares

Matematicamente, estruturas lineares são aquelas cujo mapeamento  $F_{LIN}[.]$  obedece à seguinte relação:

---

<sup>25</sup> Na terminologia clássica da área, a grandeza  $N-1$  seria a *ordem* do filtro.

$$y(n) = F_{LIN}[\zeta x_1(n) + \eta x_2(n), \zeta x_1(n-1) + \eta x_2(n-1), \dots, \zeta x_1(n-N+1) + \eta x_2(n-N+1)] = \zeta F_{LIN}[x_1(n), x_1(n-1), \dots, x_1(n-N+1)] + \eta F_{LIN}[x_2(n), x_2(n-1), \dots, x_2(n-N+1)] \quad (3.2)$$

onde  $\zeta$  e  $\eta$  são constantes arbitrárias. Como vimos na seção 2.11.2, sistemas desse tipo obedecem a duas relações essenciais:

$$y(n) = w(n)*x(n), \quad (3.3)$$

sendo  $w(n)$  a resposta ao impulso do filtro e  $x(n)$  o sinal de entrada, e

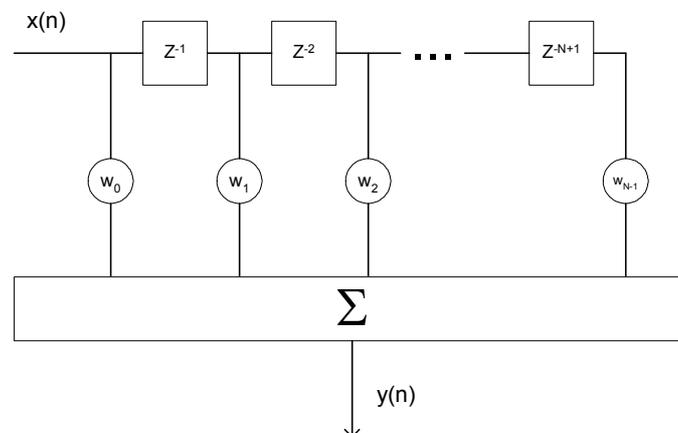
$$Y(z) = W(z)X(z) \quad (3.4)$$

onde  $W(z)$  é a função de transferência do filtro e  $X(z)$  é a transformada  $Z$  do sinal de entrada.

Filtros lineares, como adiantamos no capítulo anterior, podem ser de resposta ao impulso finita (FIR) ou de resposta ao impulso infinita (IIR). Analisemos cada espécie.

### Filtros FIR

O dispositivo FIR emblemático é, sem dúvida, o *filtro linear transversal (FLT)*. A saída de tal filtro é simplesmente uma combinação linear das diferentes versões da entrada, como mostra a Fig. 3.2.



**Figura 3.2:** Filtro Linear e Transversal

Matematicamente, a relação entrada-saída é a seguinte:

$$y(n) = \sum_{k=0}^{N-1} w(k)x(n-k) = \mathbf{w}^T \mathbf{x}(n) \quad (3.5)$$

sendo  $\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-N+1)]^T$  denominado *vetor de entrada* do filtro<sup>26</sup> e  $\mathbf{w}$  seu *vetor de parâmetros*. A função de transferência de um filtro desse tipo é dada pela seguinte expressão:

$$W(z) = \sum_{k=0}^{N-1} w(k)z^{-k} \quad (3.6)$$

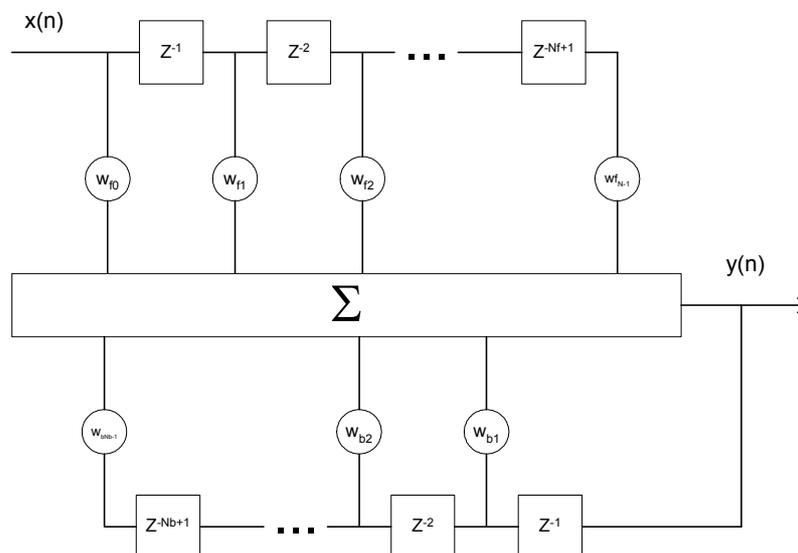
Reescrevendo (3.6) como

$$W(z) = \frac{\sum_{k=0}^{N-1} w(k)z^{N-k-1}}{z^{N-1}}, \quad (3.7)$$

percebemos que a escolha do vetor de parâmetros  $\mathbf{w}$  altera *apenas a posição dos zeros* da função de transferência do filtro, *permanecendo os pólos fixos na origem*. Por esse motivo, é usual dizer que essa estrutura “tem apenas zeros”. Embora essa afirmação não seja rigorosa, ela é bastante fiel às possibilidades práticas de um FLT.

### Filtros IIR

Filtros IIR envolvem *realimentação* i.e. *o valor atual da saída depende de valores atrasados*, conforme mostra a Fig. 3.3.



**Figura 3.3:** Filtro IIR

A relação entre  $y(n)$  e  $x(n)$  para um dispositivo desse tipo pode ser expressa como:

<sup>26</sup> Quando se lida com sinais complexos, é usual substituir a transposição matricial simples pelo operador hermitiano. Por simplicidade, estabeleceremos a notação tendo em mente sinais reais.

$$y(n) = \sum_{k=0}^{N_f-1} w_f(k)x(n-k) - \sum_{p=1}^{N_b-1} w_b(p)y(n-p) = \mathbf{w}_f \mathbf{x}(n) - \mathbf{w}_b \mathbf{y}(n-1) \quad (3.8)$$

onde  $\mathbf{x}(n)$  é o *vetor de entrada* do filtro,  $\mathbf{y}(n-1)$  é o *vetor de saídas*,  $\mathbf{w}_f$  é o vetor de parâmetros *forward* e  $\mathbf{w}_b$  é o vetor de parâmetros *backward*. A função de transferência de um filtro desse tipo tem a seguinte forma:

$$W(z) = \frac{\sum_{k=0}^{N_f-1} w_f(k)z^{-k}}{1 + \sum_{p=1}^{N_b-1} w_b(p)z^{-p}} \quad (3.9)$$

Nesse caso, como mostra (3.9), o ajuste dos parâmetros é capaz de regular a posição tanto de pólos quanto de zeros, devendo ser respeitada a condição de estabilidade<sup>27</sup>. Isso revela que o filtro IIR tem a si associado um mapeamento mais genérico que o de um filtro FIR, fato que estabelece, num certo sentido, uma maior flexibilidade do ponto de vista estrutural. A contrapartida é que o filtro IIR pode assumir uma configuração instável caso os pólos não sejam adequadamente escolhidos, o que limita possibilidades e exige constante atenção caso se deseje alterar os parâmetros. Outro ponto delicado é que a presença do laço de realimentação torna bastante complexa a relação entre a saída e os parâmetros livres devido à memória embutida nas versões atrasadas de  $y(n)$ . Isso, como veremos ao longo deste trabalho, será relevante quando buscarmos critérios de escolha de parâmetros.

### 3.2.2. Estruturas Não-lineares

Estruturas *não-lineares* são definidas simplesmente como aquelas cujo mapeamento não obedece à relação descrita em (3.2). Uma definição desse tipo abre um amplo leque de possibilidades que, se por um lado torna o campo de estudo bastante rico, por outro torna bastante complexa a tarefa de buscar abordagens analíticas genéricas. Além disso, ela torna o esforço de uma descrição geral pouco viável dados os fins de nosso trabalho. Parece-nos mais apropriado trabalhar diretamente com os diferentes tipos de filtros não-lineares, clarificando assim semelhanças e contrastes entre eles. Lidaremos aqui com três tipos básicos de estruturas, propostas originalmente tendo em vista motivações distintas, mas dotadas de diversos pontos em comum: *filtros polinomiais*, *redes neurais* e *filtros fuzzy*.

*Filtros polinomiais* ou *de Volterra* são dispositivos cujo caráter não-linear é produzido, como seu nome indica, por uma não-linearidade do tipo *polynomial* i.e. uma potência qualquer do sinal de entrada ou um termo cruzado entre amostras distintas. Por se basear numa formulação que, como veremos em 4.5.1, tem pontos de contato com o arcabouço teórico de filtros lineares, a não-linearidade polinomial foi a primeira a ser empregada no contexto de comunicações [Benedetto et al., 1976, Benedetto et al., 1979].

<sup>27</sup> Para filtros lineares, causais e invariantes no tempo, como os que nos interessam aqui, a condição básica é que *todos os pólos estejam no círculo de raio unitário*.

Por sua vez, em termos simples, *redes neurais* são estruturas formadas pela conexão, não raro em múltiplas camadas, de unidades de processamento denominadas *neurônios*. Tais estruturas podem ser entendidas como *sistemas de processamento paralelo de informação* cujo ajuste de parâmetros corresponde a um processo de *aprendizado* e, em certos casos, de *auto-organização*. Via de regra, o caráter não-linear de uma rede neural emerge da não-linearidade presente no modelo de neurônio, a qual, aliás, é decisiva para a determinação das características do mapeamento resultante.

Finalmente, *filtros fuzzy* são dispositivos cuja relação entrada-saída se baseia num conjunto de decisões do tipo *SE...ENTÃO* tomadas a partir de uma base de regras *fuzzy* escolhidas por um *expert* humano ou por um mecanismo automático. Isso quer dizer que a informação de entrada precisa ser transportada a um domínio *fuzzy*, processada segundo a base de regras e convertida em um sinal de saída no domínio original (*crisp*). O cerne de todo o processo é, como se depreende dessa breve exposição, a escolha das regras, que pode levar à emergência de um sem-número de possíveis mapeamentos não-lineares. Essa escolha tem, usualmente, uma faceta ligada à idéia de estrutura pela determinação dos tipos de conjunto *fuzzy*, bem como uma nuance ligada à escolha dos parâmetros a eles associados.

No próximo capítulo, teremos a oportunidade de discorrer com mais profundidade sobre cada uma dessas estruturas, bastando por ora que o leitor tenha em mente uma idéia geral de cada uma das classes por nós escolhidas.

### **3.3. Escolha dos Parâmetros**

Uma estrutura de filtragem é, em última análise, uma idéia, uma noção abstrata com enormes potencialidades que, num primeiro momento, *não estão concretizadas*. Para que ela deixe esse mundo ideal multifacetado e se torne uma ferramenta capaz de resolver problemas, é preciso que se dêem valores a seus parâmetros. Tais valores não podem ser quaisquer, mas devem ser escolhidos de maneira a permitir que a estrutura realize adequadamente a tarefa desejada. Em outras palavras, é preciso que haja um *critério* de escolha *adequado ao problema em questão*. Esse critério deverá conter informações relevantes acerca do problema e do modo como a estrutura se insere em contextos representativos do mesmo, verdadeiras bases para a inserção da idéia no “ambiente real”. De posse disso, será necessário, a partir de um processo de análise, determinar uma configuração satisfatória<sup>28</sup> para os parâmetros. Tendo obtido tal configuração, adotamo-la, e o processo se conclui. Pode ainda ser preciso repeti-lo, até mesmo de maneira recorrente, caso as bases do problema variem com a passagem do tempo.

Essa descrição sugere três etapas distintas:

- Escolha de um critério de ajuste e determinação de uma medida condizente com seu espírito.
- Obtenção de um conjunto de parâmetros satisfatório a partir da medida escolhida.
- Atualização do conjunto adotado caso o problema se modifique com o tempo.

---

<sup>28</sup> O grau de adequação deve ser expresso por uma medida associada ao critério, a qual, aliás, é tomada muitas vezes pelo próprio critério, num abuso comum de linguagem.

A primeira etapa envolve uma boa dose de criatividade, pois é preciso determinar um critério de ajuste racional formulado em termos matemáticos, tendo como ponto de partida a conceituação de um problema concreto. A segunda etapa tem um caráter mais sistemático, dependendo de *um processo de busca* i.e. de um método que, num amplo espaço formado por possíveis candidatos, encontre um que satisfaça os padrões exigidos de desempenho. Finalmente, a etapa final trata da capacidade de *adaptação* do receituário escolhido a uma situação em que “a regra do jogo” varie. Passemos a uma avaliação mais detalhada de cada etapa.

### 3.3.1. Etapa 1: Escolha do Critério

O primeiro passo da etapa inicial é buscar compreender a natureza do problema em questão e dos objetivos a atingir. O passo seguinte é expressar matematicamente esses objetivos na forma de um critério que, por necessidade, deverá conter uma medida de qualidade de um candidato genérico a solução. Todos os critérios de filtragem com que nos ocuparemos aqui têm a seguinte forma:

$$\textit{Minimize ou maximize } J(\mathbf{w}) \textit{ com respeito ao vetor de parâmetros } \mathbf{w}, \quad (3.10)$$

sendo, portanto,  $J(\mathbf{w})$  uma medida da qualidade de um vetor de parâmetros qualquer. A função  $J(\cdot)$  irá depender de dois fatores essenciais: *a essência do problema* e *a informação de referência disponível*. O primeiro fator é bastante lógico, pois cada tarefa terá metas diferentes e, portanto, exigirá formulações próprias. O segundo fator traz à baila uma limitação: de que tipo de informação disporemos para elaborar nossa medida? Sem dúvida, como veremos diversas vezes até o fim deste documento, a natureza da resposta a essa pergunta é crucial na elaboração de um critério.

### 3.3.2. Etapa 2: Obtenção dos Parâmetros

Tarefas como a apresentada em (3.10) são exemplos muito claros de problemas de *otimização* i.e. *aqueles cujo objetivo é encontrar a melhor solução em um domínio de possibilidades*. A presente etapa, portanto, resume-se a *otimizar* a função  $J(\mathbf{w})$  com respeito aos parâmetros da estrutura, ou seja, a buscar o vetor *que leve o filtro escolhido ao melhor desempenho possível* na realização da tarefa implícita no critério. Dividimos as técnicas de otimização em três classes que terão para nós particular interesse:

- **Técnicas Clássicas:** são aquelas baseadas nas derivadas da função custo (gradiente, hessiana etc.). São, sem dúvida, as mais empregadas em problemas de filtragem, embora sofram de uma limitação particularmente séria: são capazes de efetuar apenas busca local, não tendo potencial inato de evitar convergência para extremos locais.
- **Técnicas de Busca Global:** são técnicas de inspirações e caráter diversos capazes de efetuar uma varredura mais efetiva do espaço de busca graças a mecanismos que

umentam a possibilidade de obtenção do ótimo global<sup>29</sup>. Incluímos nesse grupo heterogêneo ferramentas como *algoritmos evolutivos*, *simulated annealing*, *busca tabu* etc.

- **Heurísticas / Algoritmos Meméticos:** são técnicas especialmente planejadas para resolver de maneira eficiente *um dado problema* ou *uma classe específica de problemas*, não sendo necessariamente úteis fora desse contexto particular.

Frisamos que não desejamos estabelecer classificações genéricas, mas tão-somente estabelecer uma divisão útil aos problemas de que nos ocuparemos.

Não há uma técnica que seja *a melhor* em toda e qualquer situação: como nos mostram os chamados teoremas “*no free lunch*” [Wolpert e Macready, 1997], a escolha da ferramenta de otimização deve ser feita de acordo com considerações extraídas do caso em mãos. Existem dois aspectos preponderantes, cujo inter-relacionamento produz um dilema fundamental: *complexidade do problema*<sup>30</sup> e *recursos disponíveis*<sup>31</sup> para sua solução. Se for possível obter soluções satisfatórias com técnicas simples de busca local, não será necessário utilizar um esquema mais refinado e custoso computacionalmente. Por outro lado, caso haja necessidade de desempenho ótimo, é aceitável optar por uma técnica mais complexa e apropriada às idiossincrasias do cenário em questão. A escolha final, como de costume, dependerá de uma avaliação de projeto.

### 3.3.3. Etapa 3: Adaptação dos Parâmetros

As duas etapas anteriores *per se* formam um receituário coeso para projeto de um filtro. No entanto, há situações em que a natureza do problema varia com o passar do tempo, o que requer que os parâmetros do filtro se *adaptem* constantemente, ou seja, que o processo de filtragem seja *adaptativo*. Para que entendamos tal necessidade, imaginemos por um momento um indivíduo que trafega, a bordo de um ônibus, portando um celular. Ele faz uma ligação enquanto o ônibus está no centro da cidade, área caracterizada por uma enorme densidade de prédios e na qual o sinal recebido pelo aparelho será formado por muitas versões refletidas. A ligação prossegue enquanto ele passa por uma área rural muito plana, em que não há obstáculos e, portanto, pode existir uma “linha de visada” entre aparelho e a estação mais próxima. Antes do encerramento da conversa, o ônibus passa ao largo de uma montanha e, em seguida, por um túnel. Concentremo-nos no sinal recebido pela antena do aparelho: ele terá características bem distintas em cada uma dessas situações. Dessa forma, caso desejássemos efetuar, de maneira incessantemente efetiva, tarefas de equalização, predição e identificação nesse contexto, seria imperativo contar com um mecanismo de adaptação dos parâmetros de nosso filtro. Matematicamente, tal mecanismo, denominado *algoritmo adaptativo*, é processo iterativo que obedece à seguinte forma geral:

$$w(n+1) \text{ é uma função de } w(n) \text{ e de outros parâmetros e variáveis relevantes} \quad (3.11)$$

---

<sup>29</sup> Note o leitor que não queremos dizer com isso que as técnicas **sempre** levarão ao melhor ótimo!

<sup>30</sup> Que inclui o grau de desempenho desejado.

<sup>31</sup> Os recursos incluem fatores como tempo disponível, poder de processamento etc.

sendo  $w(n)$  o vetor de parâmetros no instante  $n$ . Há uma importante sinergia entre esta etapa e a etapa anterior: como vários métodos de otimização são processos iterativos, ou seja, coerentes com o espírito de (3.11), eles podem fazer o papel de algoritmos adaptativos caso levem em conta, direta ou indiretamente, a variação temporal experimentada pela função custo que se deseja otimizar.

Expusemos os principais elementos de projeto de um filtro de maneira genérica, de modo a sedimentar alguns conceitos que nos acompanharão constantemente. Agora, convém que, de posse de uma base teórica apropriada, analisemos as particularidades dos três problemas de filtragem de que nos ocuparemos: equalização, identificação e predição.

### **3.4. Equalização**

Um costume dos mais rigorosos audiófilos é buscar um ajuste de seus aparelhos de som capaz de garantir a reprodução fiel da experiência sonora *in loco*. Um meio para isso é passar um bom tempo regulando “graves” e “agudos”, ou seja, decidindo o grau de aceitação ou rejeição por parte do equipamento de um conjunto de faixas de frequência pré-determinadas. O perfil resultante, se a fortuna for favorável a nosso aficionado, compensará as imperfeições de gravação e reprodução, proporcionando-lhe, após tanto esforço, uma agradável experiência sonora. Esse hipotético *appréciateur* fez uso de uma conhecida ferramenta: um *equalizador* i.e. *um filtro projetado para contrabalançar efeitos nocivos de um certo sistema portador de informação*. O critério de ajuste do filtro teve por base a satisfação aural do audiófilo, e a otimização dos parâmetros se deu através de um processo de “tentativa-e-erro”. Provavelmente, após alguns minutos nosso personagem se sentirá insatisfeito com alguma imperfeição e dará início a um processo adaptativo de ajuste dos parâmetros.

O exemplo acima ilustra um processo de equalização que, em maior ou menor grau, é familiar a boa parte das pessoas que ouvem música. Nos modernos sistemas de comunicação digital, a tarefa de equalização é análoga: projetar um filtro que, atuando sobre o sinal recebido, contribua para uma mais precisa restauração da mensagem original. Essa definição, propositadamente ampla, dá margem a diversas concepções que, por sua vez, formam a base de distintos paradigmas. A seguir, exporemos, numa ordem que julgamos apropriada do ponto de vista cronológico, as concepções mais relevantes para nosso trabalho, buscando sempre clarificar semelhanças e diferenças úteis à elaboração de uma base conceitual sólida.

#### **3.4.1. Equalização Vista como Inversão do Canal - O Critério *Zero-forcing* (ZF)**

O primeiro ponto de vista que analisaremos, certamente o mais antigo de todos, é aquele baseado na idéia de *inversão* do canal. Trata-se da idéia central subjacente ao procedimento adotado pelo audiófilo do exemplo: procurar um filtro que *atue de maneira oposta à do canal*, tornando a combinação entre ambos um sistema sem distorções. Sendo  $H[.]$  o mapeamento entrada-saída do canal e  $W[.]$  o mapeamento do equalizador, poderíamos expressar matematicamente a condição ideal, nesse caso, como:

$$W\{H[a]\} \approx a \quad (3.12)$$

sendo  $a$  um elemento pertencente ao domínio apropriado e  $\approx$  um sinal indicativo de equivalência sobre algum ponto de vista teórico. Desse ponto de vista, podemos afirmar que:

$$W[.] = H^{-1}[.], \quad (3.13)$$

ou seja, que o mapeamento do equalizador é o inverso daquele efetuado pelo canal.

Assumindo que tanto canal quanto equalizador são filtros lineares e que não há ruído aditivo, teremos, em nossa notação padrão:

$$x(n) = h(n)*s(n) \quad (3.14)$$

sendo  $h(n)$  a resposta ao impulso do canal. Teremos também:

$$y(n) = w(n)*x(n) \quad (3.15)$$

onde  $w(n)$  é a resposta ao impulso do equalizador. A equação (3.14) nos mostra que o sinal recebido  $x(n)$  é formado por uma sobreposição de versões do sinal enviado  $s(n)$ . Como vimos em 2.9.1, isso significa que o sinal está sujeito à interferência intersimbólica (IIS). Nesse caso, a IIS é o efeito nocivo que desejamos contrabalançar, ou até, se possível, eliminar. Relacionemos diretamente a saída do equalizador ao sinal transmitido:

$$y(n) = w(n)*x(n) = [w(n)*h(n)]*s(n) = c(n)*s(n) \quad (3.16)$$

sendo  $c(n) = w(n)*h(n)$  a *resposta conjunta canal + equalizador*. Como, em comunicações, usualmente é tolerável a recuperação de uma versão do sinal transmitido atrasada e amplificada por um fator de escala, a condição de inversão perfeita poderia ser definida como:

$$c(n) = w(n)*h(n) = K\delta(n-d) \quad (3.17)$$

onde  $K$  é um fator de escala e  $d$  é o *atraso de equalização*.

Um equalizador capaz de produzir uma resposta conjunta  $c(n)$  como a mostrada em (3.17) recebe o nome de *zero-forcing*<sup>32</sup> (ZF) [Lucky, 1965], sendo um dispositivo capaz de efetuar a *perfeita inversão* do canal. Aplicando a transformada  $Z$  a ambos os membros de (3.17), obtemos

$$W(z)H(z) = Kz^{-d}, \quad (3.18)$$

ou seja,

---

<sup>32</sup> Note que todos os valores de  $c(n)$  foram “forçados a zero”, exceto aquele localizado no instante temporal  $n-d$ , o que justifica a nomenclatura estabelecida.

$$W(z) = \frac{Kz^{-d}}{H(z)} \quad (3.19)$$

A equação (3.19) expressa de maneira muito direta os objetivos que delineamos, pois mostra que o equalizador *zero-forcing* tem uma função de transferência que é, em essência, a *inversa* da função associada ao canal. Analisando o problema do ponto de vista da frequência, temos a base para uma intuição firmemente estabelecida na mente de qualquer engenheiro de comunicações: a resposta em frequência do equalizador deve ter um perfil oposto ao do canal, o que explica noções como “o equalizador de um canal passa-baixas deve ser passa-alta”. O exemplo 3.1 ilustra esse fato.

### Exemplo 3.1

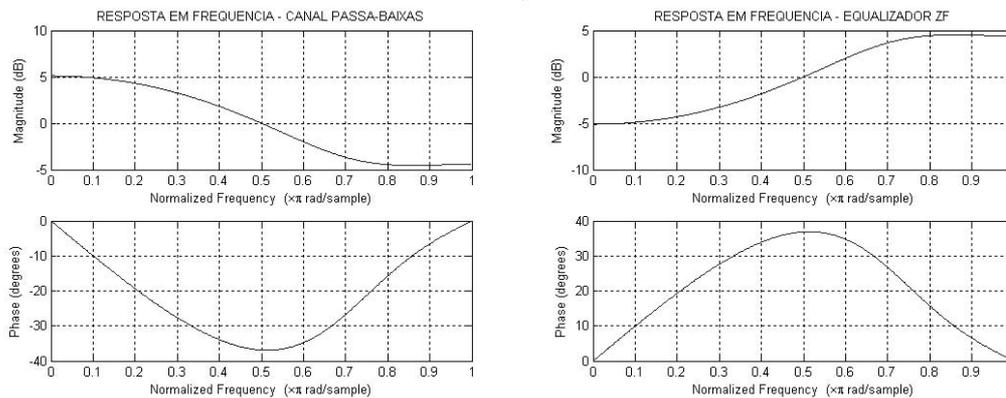
Seja um canal passa-baixas com a seguinte função de transferência:

$$H_{\text{canal}}(z) = 1 + 0.6z^{-1} + 0.2z^{-2} \quad (3.20)$$

O equalizador ZF para atraso zero, neste caso, terá a seguinte função de transferência:

$$H_{\text{eqZF}}(z) = \frac{1}{1 + 0.6z^{-1} + 0.2z^{-2}} \quad (3.21)$$

Na Fig. 3.4, apresentamos as respostas em frequência (magnitude e fase) de ambos os sistemas, e se torna bastante claro o caráter inverso de seus espectros, em perfeita concordância com o espírito do critério.



**Figura 3.4:** Respostas em Frequência: Canal e Equalizador

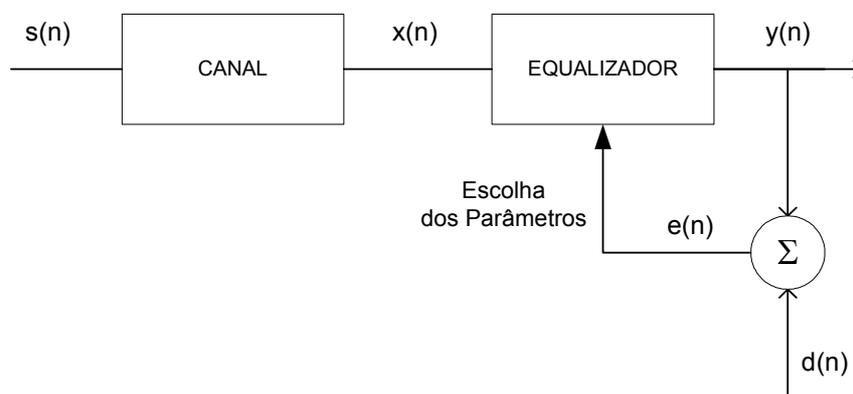
Voltemos à equação (3.19), e assumamos que o canal que desejamos equalizar seja modelado por um filtro FIR, ou, em outras palavras, que tenha apenas zeros em sua função

de transferência<sup>33</sup>. Nesse caso, como mostra o exemplo 3.1,  $W(z)$  deverá ter um conjunto de pólos que seja um contraponto aos zeros do canal, ou seja, *não poderá ser também um filtro FIR*. Isso significa que, caso desejemos nos manter no contexto linear, será necessário empregar um dispositivo com resposta ao impulso infinita para que cumpramos plenamente o objetivo inerente ao critério ZF. Essa limitação fundamental é de grande importância, pois modelos FIR são os mais comuns tanto para canais quanto para equalizadores.

Outro ponto muito importante é que, como frisamos no início da exposição, um equalizador ZF não contém em seu bojo nada que leve em conta a existência de ruído. Isso quer dizer que o critério se restringe à mitigação de IIS, não sendo, portanto, efetivo caso o fenômeno estocástico seja preponderante. Essa limitação pode até levar a cenários em que o ruído chega ter seu efeito realçado pela ação do equalizador, caracterizando uma situação em que há *noise enhancement*.

### 3.4.2. Equalização a Partir de uma Medida de Erro: o Critério de Wiener

O foco do critério ZF, como acabamos de ver, está na idéia de inverter a função de transferência do canal, ou seja, de estabelecer um modelo inverso que compense a distorção associada à interferência intersimbólica. Uma outra maneira de forjar uma função custo coerente com os objetivos da equalização é estabelecer um termo de comparação entre a saída do filtro e aquilo que se esperaria dela num caso favorável. Em outras palavras, a idéia seria medir o *erro* entre o que se tem na saída do filtro e um *signal desejado* que indique o que se gostaria de ter<sup>34</sup>. Na Fig. 3.5, mostramos um esquema simplificado de um sistema de comunicação com a nomenclatura por nós adotada. Vemos que o sinal transmitido  $s(n)$  passa por um canal, chegando ao receptor na forma distorcida  $x(n)$ . Esse sinal distorcido é a “matéria-prima” sobre a qual atuará o equalizador com um único fim: produzir um sinal de saída  $y(n)$  tão próximo quanto possível de  $d(n)$ <sup>35</sup>.



**Figura 3.5:** Modelo de um Esquema de Equalização Supervisionada

<sup>33</sup> Chamamos a atenção do leitor, pela última vez, para esse abuso de linguagem. Na realidade, como já discutimos anteriormente, o canal tem também pólos, os quais, no entanto, não estão associados a nenhum tipo de informação sistêmica relevante para nossos fins.

<sup>34</sup> A existência de um sinal piloto dessa natureza é o que caracteriza uma técnica *supervisionada*.

<sup>35</sup> O sinal desejado está relacionado ao sinal transmitido, como veremos em seguida.

Uma medida muito comum e efetiva de erro é o quadrado da diferença  $e(n)$  entre o sinal de saída  $y(n)$  e o sinal desejado  $d(n)$ . Com isso, temos<sup>36</sup>:

$$\Delta(n) = e^2(n) \quad (3.22)$$

Na equação (3.22), vê-se apenas a situação em um dado momento, nela não havendo nada que indique *um padrão de comportamento* rigoroso e sistemático. Para sanar tal limitação, faz-se uso da *esperança matemática*, uma média entre todas as possibilidades do espaço amostral:

$$J_W = E[\Delta(n)] = E[e^2(n)] \quad (3.23)$$

A função  $J_W$  resultante é a expressão do chamado *critério de Wiener*, sendo seus valores medidos do *erro quadrático médio (EQM)* entre sinal desejado e sinal de saída. Para concluir o projeto da estrutura, é preciso *minimizar a função  $J_W$  com respeito aos parâmetros livres*.

Para aprofundar nossas idéias sobre a formulação, assumamos por um momento que o equalizador que desejamos projetar seja um filtro linear e transversal com vetor de parâmetros  $\mathbf{w}$ . Nesse caso, podemos reescrever  $J_W$  como:

$$J_W = \mathbf{w}^T \mathbf{R} \mathbf{w} - 2\mathbf{w}^T \mathbf{p} + \sigma_d^2 \quad (3.24)$$

A matriz  $\mathbf{R}$  recebe o nome de *matriz de autocorrelação* do sinal de entrada do filtro :

$$\mathbf{R} = E[\mathbf{x}(n)\mathbf{x}(n)^T] = \begin{bmatrix} r_x(0) & r_x(1) & \dots & r_x(N-1) \\ r_x(1) & r_x(0) & \dots & r_x(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_x(N-1) & r_x(N-2) & \dots & r_x(0) \end{bmatrix} \quad (3.25)$$

sendo

$$r_x(m) = E[x(n-q)x(n-q-m)], \quad q \in Z \quad (3.26)$$

a função de autocorrelação discreta<sup>37</sup> definida em 2.9.2. Já o vetor  $\mathbf{p}$ , que relaciona o sinal desejado ao sinal de entrada do filtro, é denominado *vetor de correlação cruzada*:

$$\mathbf{p} = E[\mathbf{x}(n)d(n)] = \begin{bmatrix} E[x(n)d(n)] \\ E[x(n-1)d(n)] \\ \vdots \\ E[x(n-N+1)d(n)] \end{bmatrix} \quad (3.27)$$

<sup>36</sup> Nossas definições, por uma questão de didática, têm sempre por base sinais reais.

<sup>37</sup> Assumimos aqui, como o faremos ao longo do trabalho, que os sinais envolvidos são estacionários no sentido amplo.

Basicamente, a matriz  $\mathbf{R}$  é uma medida da relação entre as diferentes amostras do sinal recebido. Já o vetor de correlação cruzada indica “o quanto tem a ver” com o sinal desejado o sinal recebido. Tal medida é muito importante, pois o filtro não é capaz de produzir um sinal de saída que não tenha correlação com sua entrada, ou seja, o filtro não pode “operar o milagre” de moldar um sinal desejado a partir de uma entrada que nada tenha a ver com ele.

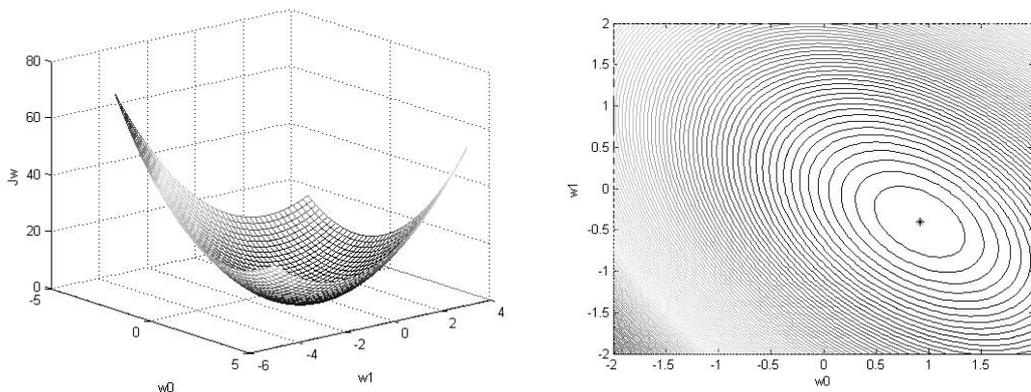
Para minimizar a função  $J_w$ , obtemos o vetor gradiente com relação aos parâmetros:

$$\nabla J_w = 2\mathbf{R}\mathbf{w} - 2\mathbf{p} \quad (3.28)$$

Igualando o vetor gradiente ao vetor nulo, obtemos o ótimo:

$$\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p} \quad (3.29)$$

Uma análise de derivadas de segunda ordem mostra que o ótimo, nesse caso, é sempre um mínimo, que recebe o nome de *solução de Wiener*. Para ilustrar o que discutimos, apresentamos na Fig. 3.6 uma típica função custo de Wiener acompanhada de suas curvas de nível.



**Figura 3.6:** Função Custo de Wiener e Curvas de Nível

Note o leitor que, dado um certo sinal desejado, a função custo tem um único mínimo<sup>38</sup>. Essa propriedade, sem dúvida desejável, cristalizou no imaginário de muitos pesquisadores a idéia de que a aplicação do paradigma de Wiener à equalização de canais levaria a um problema unimodal. Entretanto, em equalização, há potencialmente várias escolhas para o sinal desejado *em um único problema*, o que gera um conjunto de múltiplas soluções possíveis. A chave para isso se encontra *na escolha do atraso de equalização*. Como vimos em (3.17), é perfeitamente tolerável que se recupere uma versão atrasada do sinal transmitido, o que significa que o sinal desejado terá a seguinte forma genérica:

$$d(n) = s(n-d) \quad (3.30)$$

<sup>38</sup> A unicidade da solução de Wiener para esse caso particular decorre diretamente da linearidade do equalizador escolhido. O uso de uma estrutura não-linear, como discutiremos adiante, pode levar a cenários com múltiplas soluções.

sendo  $d$  o *atraso de equalização*. A escolha do atraso não afeta a matriz  $\mathbf{R}$ , mas modifica decisivamente o vetor de correlação cruzada. Portanto, em tese, cada atraso levará a uma solução de Wiener distinta, a qual, por sua vez, terá a si associado um EQM residual próprio. Recomendamos ao leitor que tenha sempre em mente essa visão mais ampla do problema de equalização de Wiener, pois ela é essencial para a compreensão das técnicas não-supervisionadas e da abordagem de classificação. Para sedimentar os conceitos expostos, apresentamos a seguir um exemplo.

### Exemplo 3.2

Seja um canal de fase mínima com a seguinte função de transferência:

$$H_{\text{canal}}(z) = 1 + 0.6z^{-1} \quad (3.31)$$

e tomemos como equalizador um filtro linear com dois coeficientes. Consideremos ainda que o sinal transmitido é formado por amostras binárias i.i.d.  $+1$  e  $-1$ . Essa situação levará a três possibilidades para o sinal desejado:  $s(n)$ ,  $s(n-1)$  e  $s(n-2)$ , pois qualquer outra escolha não teria correlação com o vetor de entrada  $\mathbf{x}(n) = [x(n) \ x(n-1)]^T$ . Na Tab. 3.1, apresentamos os três casos.

Atraso de Equalização	Solução de Wiener	Erro Quadrático Médio
0	$[0.9130 \ -0.4028]^T$	0.0870
1	$[0.1450 \ 0.6713]^T$	0.2417
2	$[-0.2417 \ 0.5478]^T$	0.6713

**Tabela 3.1:** Comparação entre Soluções de Wiener

A tabela mostra que as três soluções são bastante diferentes e que a escolha do atraso leva a problemas bastante distintos, conforme atesta o erro residual da melhor solução em cada caso. Isso revela com nitidez que a determinação do atraso é parte essencial do problema de equalização, parte esta que jamais deve ser dissociada de sua análise. Para ilustrar de maneira mais abrangente a questão, apresentamos na Fig. 3.7 as curvas de nível de uma função de Wiener “amalgamada”<sup>39</sup> [Johnson et al., 1998], *composta pelo valor mínimo entre as funções associadas a cada um dos três atrasos*, que ilustra bem o caráter multimodal do paradigma<sup>40</sup>.

<sup>39</sup> O epíteto “amalgamada” advém do fato de a função ser composta por trechos de diferentes funções.

<sup>40</sup> Note o leitor que a pior solução nem chega a “fazer parte” da função global.

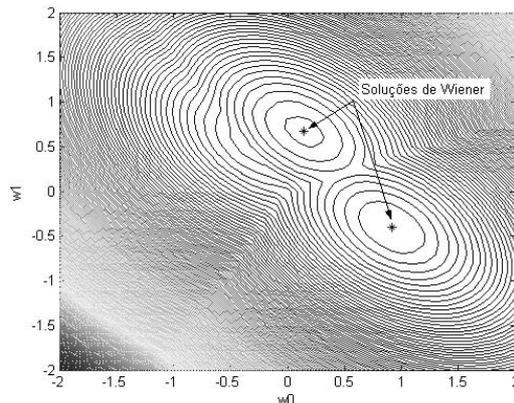


Figura 3.7: Função “Amalgamada” de Wiener

---

### **Estruturas Não-lineares e o Critério de Wiener**

O conhecimento do sinal desejado faz do critério de Wiener uma base sólida também para a equalização não-linear. O uso de uma estrutura mais complexa representa um acréscimo em termos de flexibilidade de mapeamento, que, por sua vez, se refletirá num desempenho potencialmente melhor. O “preço a pagar” por esse possível melhor desempenho é o aumento da complexidade do problema a resolver, tanto em termos de formulação e processamento quanto no que diz respeito à obtenção da solução ótima. A função custo de Wiener pode ser bastante complexa, mesmo para um atraso fixo, caso se empregue uma estrutura não-linear.

Interessantemente, caso o mapeamento não-linear do filtro seja *linear nos parâmetros*, i.e., tenha a forma

$$W(\mathbf{x}) = \mathbf{w}^T F_W(\mathbf{x}), \quad (3.32)$$

sendo  $F_W(\cdot)$  uma função não-linear, o desenvolvimento apresentado nas equações (3.24-3.29) continuará válido<sup>41</sup>, embora, implicitamente, lide-se com estatísticas de ordem superior. Esse caso, como veremos ao longo dos próximos capítulos, representa um importante meio termo, embora permaneça em aberto o problema de escolher um mapeamento  $F_W(\cdot)$  que leve a um efetivo ganho de desempenho.

### **Steepest Descent, LMS e RLS**

Para obter um algoritmo adaptativo a partir do critério de Wiener, a praxe é partir do método de *steepest descent*, cujo *modus operandi* se baseia numa busca iterativa seguindo a direção contrária à de maior crescimento, indicada pelo vetor gradiente. Matematicamente, a adaptação de um vetor de parâmetros  $\mathbf{w}$  obedece à seguinte expressão:

---

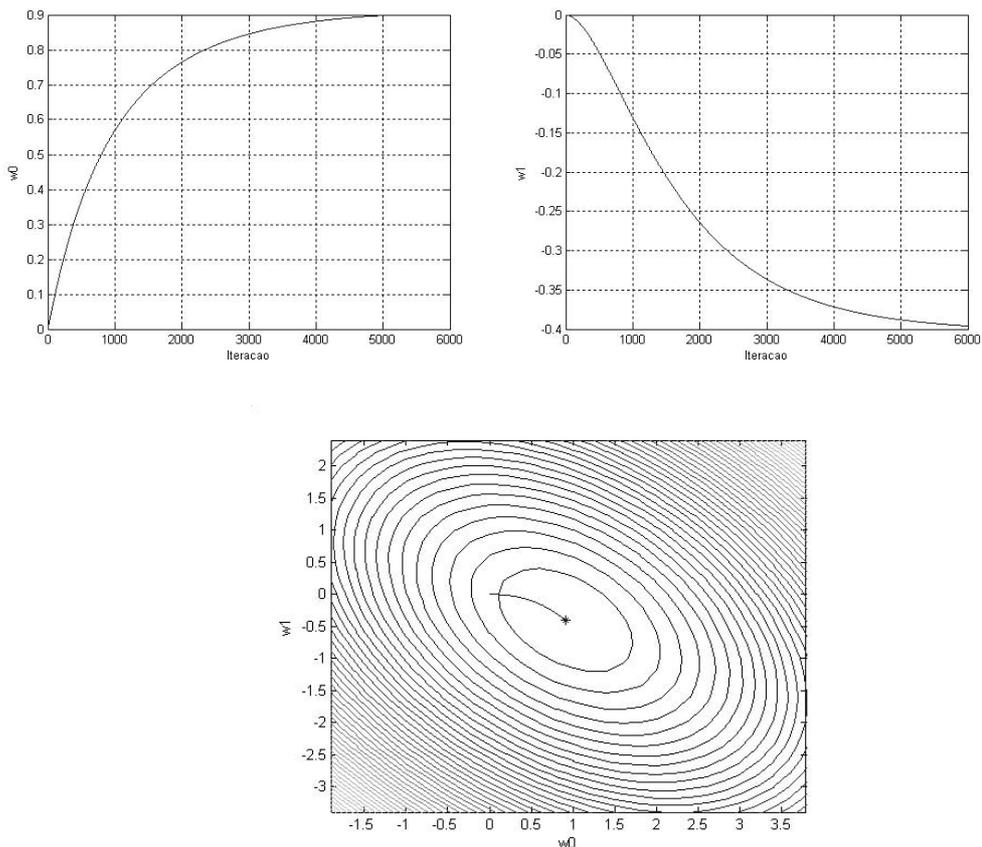
<sup>41</sup> Bem como qualquer outra formulação supervisionada, devendo o leitor estar atento a esse fato no futuro. A discussão sobre os algoritmos LMS e RLS também se aplica a estruturas desse tipo.

$$\mathbf{w}(n+1) = \mathbf{w}(n) - 0.5\mu\nabla J_{\mathbf{w}} \quad (3.33)$$

onde  $\mu$  é o *passo de adaptação*. Substituindo (3.28) em (3.33) e reagrupando alguns termos, chegamos à expressão do algoritmo *steepest descent (ASD)*:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu[\mathbf{p}-\mathbf{R}\mathbf{w}] \quad (3.34)$$

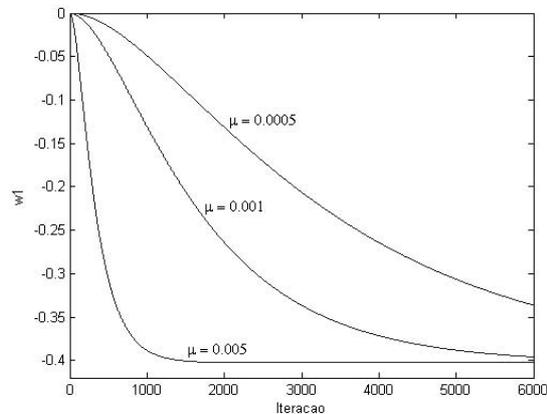
Com o gradual desvanecimento dos transitórios, o vetor de parâmetros tenderá, como é razoável, para a solução de Wiener. O parâmetro  $\mu$  regula a velocidade de convergência, que será tanto maior quanto maior for o passo. O limite superior para esse parâmetro é determinado pela estrutura dos autovalores da matriz  $\mathbf{R}$ , sendo qualquer excesso penalizado com divergência para o infinito. Na Fig. 3.8., ilustramos uma realização do algoritmo *steepest descent* no cenário discutido no exemplo 3.2, com atraso nulo e  $\mu = 0.001$ . Mostramos a evolução temporal de ambos os coeficientes do equalizador, bem como a convergência no espaço de estados, tendo como pano de fundo as curvas de nível da função custo. É bastante nítido o típico comportamento de primeira ordem do sistema em termos do transiente e a estabilização na vizinhança da solução de Wiener.



**Figura 3.8:** Comportamento do ASD

Para ilustrar a influência do passo de adaptação, apresentamos na Fig. 3.9 a evolução temporal do coeficiente  $w_1$  do equalizador para três valores de passo:  $\mu = 0.005$ ,  $\mu$

= 0.001 e  $\mu = 0.0005$ . Notamos com clareza que o aumento do passo de adaptação leva a uma convergência mais rápida para a solução.



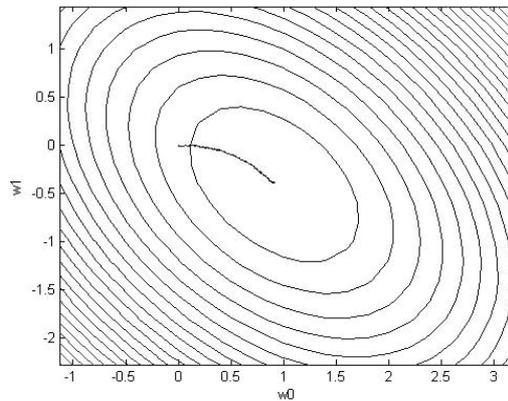
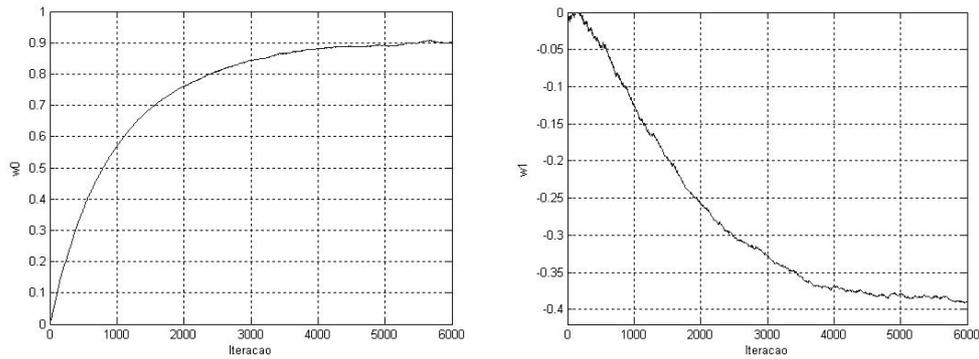
**Figura 3.9:** Influência do Passo de Adaptação no ASD

Voltemo-nos agora para um problema que permanece: numa aplicação real, como obter o valor exato de  $\mathbf{R}$  e  $\mathbf{p}$ ? Em alguns casos, acumular amostras para realizar uma média temporal poderia ser mais dispendioso que o desejável, o que terminou por motivar uma idéia bastante simples e direta [Widrow e Hoff, 1960]: *substituir as esperanças por valores instantâneos*<sup>42</sup>. Com isso, o ASD se transforma no célebre algoritmo LMS (*Least Mean Square*):

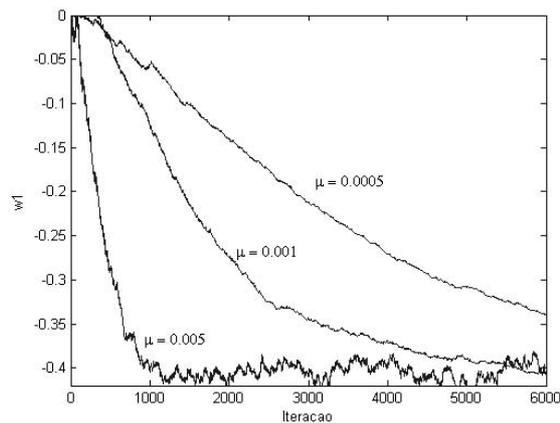
$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e(n)\mathbf{x}(n) \quad (3.35)$$

A aproximação que dá origem ao LMS, tão rudimentar à primeira vista, revela-se, na prática, bastante robusta. Em termos muito simples, as flutuações sucessivas tendem a se compensar, levando o vetor de parâmetros à vizinhança da solução de Wiener [Haykin, 1996]. O passo de adaptação continua sendo uma variável decisiva, havendo agora um compromisso entre a velocidade de convergência e a “amplitude” da flutuação em torno do ponto ótimo, denominada *desajuste* (em inglês, *misadjustment*). O limite superior para o passo depende ainda da estrutura dos autovalores da matriz de autocorrelação, embora de uma maneira mais complexa. Na Fig. 3.10, trazemos situações análogas às mostradas na Fig. 3.9, utilizando desta vez o LMS. Percebemos uma notável semelhança entre a evolução temporal do LMS e a do ASD, existindo, no entanto, um grau de flutuação estocástica tanto no transitório quanto em regime. Na Fig. 3.11, o caráter estocástico se torna ainda mais explícito, pois nela estão realizadas realizações para diferentes valores de passo.

<sup>42</sup> Essa é a célebre *aproximação estocástica*.



**Figura 3.10:** Comportamento do LMS



**Figura 3.11:** Influência do Passo de Adaptação no LMS

A Fig. 3.11 explicita bem o compromisso inerente ao LMS: há um menor grau de flutuação e desajuste para valores menores de passo, sendo o preço a pagar uma menor velocidade de convergência.

A maior diferença entre o ASD e o LMS é que a expressão do último representa um sistema dinâmico *estocástico*. É muito mais difícil extrair conclusões altamente precisas de

um sistema desse tipo, o que torna a análise do LMS mais complexa que a de um sistema linear determinístico. A regra é procurar comportamentos médios tão precisos quanto possível, principalmente no que diz respeito à análise dinâmica, ou seja, de convergência.

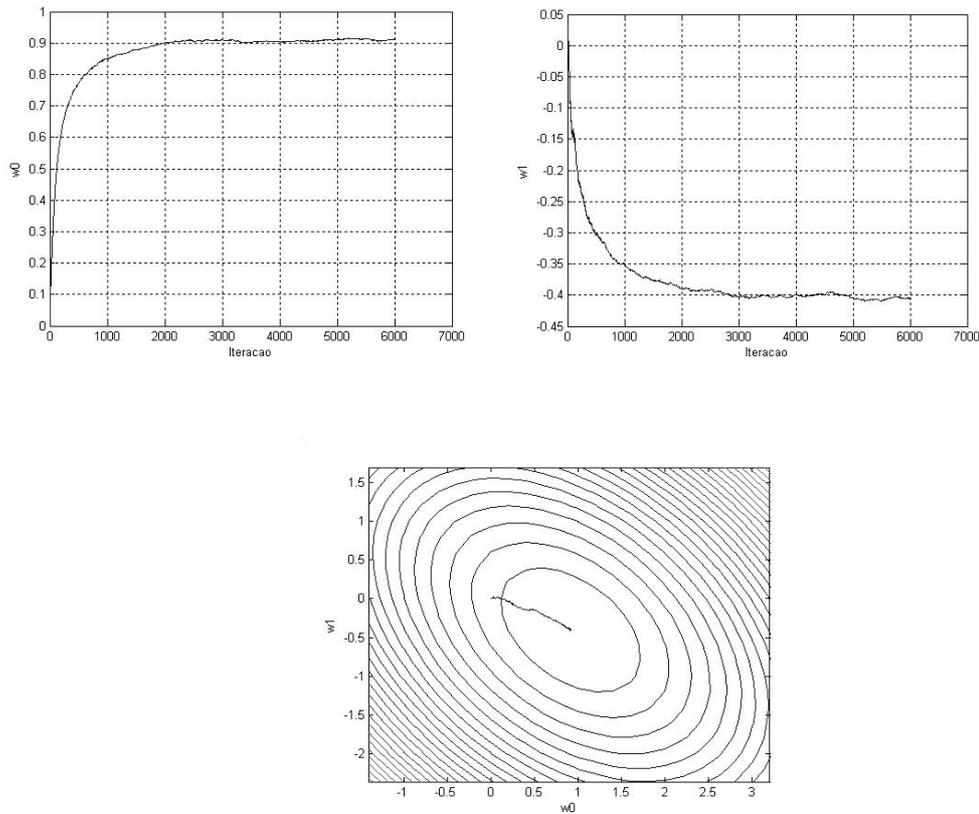
O algoritmo RLS (*Recursive Least-Squares*) é tido na teoria clássica como uma espécie de contrapartida do LMS. Sua motivação principal é buscar uma alternativa mais precisa à aproximação estocástica do LMS que mantenha um caráter iterativo sem se tornar excessivamente complexa. A chave para isso é estimar dinamicamente tanto a matriz de correlação quanto o vetor de correlação cruzada, utilizando para tal fim uma recursão que leve em conta os dados mais atuais e a estimativa mais recente. De posse dessa estimativa, recorre-se à equação (3.29) para que se obtenha uma solução análoga à de Wiener, evitando efetuar diretamente a inversão da matriz por meio de um engenhoso lema matemático<sup>43</sup> [Haykin, 1996]. A concatenação desses passos leva ao algoritmo, exposto de maneira simples na Tab. 3.2.

<p>1. <i>Inicialização Típica</i></p> <ul style="list-style-type: none"> <li>• <math>\mathbf{P}(0) = \delta^{-1}\mathbf{I}</math>, sendo <math>\delta</math> uma constante positiva pequena.</li> <li>• <math>\mathbf{w}(0) = \mathbf{0}</math></li> <li>• Escolha o <i>fator de esquecimento</i> <math>\lambda</math>, que controla o grau de peso do passado nas estimativas das variáveis aleatórias de interesse.</li> </ul> <p>2. Laço Principal – Para <math>n</math> variando de 1 até <math>N_{\text{iterações}}</math>, faça:</p> <ul style="list-style-type: none"> <li>• <math display="block">\mathbf{k}(n) = \frac{\lambda^{-1}\mathbf{P}(n-1)\mathbf{x}(n)}{1 + \lambda^{-1}\mathbf{x}^T(n)\mathbf{P}(n-1)\mathbf{x}(n)}</math></li> <li>• <math>\mathbf{e}(n) = d(n) - \mathbf{w}^T(n-1)\mathbf{x}(n)</math></li> <li>• <math>\mathbf{w}(n) = \mathbf{w}(n-1) + \mathbf{k}(n)\mathbf{e}(n)</math></li> <li>• <math>\mathbf{P}(n) = \lambda^{-1}\mathbf{P}(n-1) - \lambda^{-1}\mathbf{k}(n)\mathbf{x}^T(n)\mathbf{P}(n-1)</math></li> </ul>
---

**Tabela 3.2:** Algoritmo RLS

Em relação ao LMS, o RLS apresenta uma maior velocidade de convergência e um menor desajuste ao preço de uma maior complexidade computacional. Estabelece-se, assim, mais um compromisso que precisa ser levado em conta pelo projetista à luz de suas próprias restrições sistêmicas. Na Fig. 3.12, ilustramos o comportamento do algoritmo RLS no mesmo caso exposto nas figuras anteriores, com  $\lambda = 0.999$ . Notamos que, para o fator de esquecimento escolhido, há um desajuste não superior ao mostrado na Fig. 3.10 com uma velocidade de convergência maior, o que ressalta as vantagens da técnica mais complexa.

<sup>43</sup> Denominado *lema de inversão de matrizes*



**Figura 3.12:** Comportamento do RLS

A exposição das três técnicas adaptativas supervisionadas não esgota o universo de possíveis abordagens, mas é suficiente para nossos fins. A primeira delas, por depender de um perfeito conhecimento das estatísticas envolvidas, não é muito usada na prática, tendo um valor preponderantemente analítico. Já as duas outras técnicas não exigem em nenhum momento que o usuário forneça *explicitamente*  $\mathbf{R}$  e  $\mathbf{p}$ , havendo mecanismos iterativos de estimação dessas grandezas. Além disso, tal propriedade as torna atraentes por um outro motivo: elas são capazes de obter uma estimativa da solução de Wiener *sem que se recorra explicitamente à manipulação de  $J_w$* . Tal característica é útil neste caso e o será também quando lidarmos com funções custo mais complexas, cuja manipulação pode ser inviável e para cujos mínimos pode não haver expressão analítica<sup>44</sup>.

### 3.4.3. Equalização a Partir de Estatísticas de Ordem Superior: Critérios Não-supervisionados

Os paradigmas discutidos anteriormente têm por base informações bastante precisas: as características da função de transferência do canal, no caso do critério ZF, e

<sup>44</sup> Num caso extremo como esse último, até mesmo o ASD pode ser útil, tendo em vista que o cálculo do gradiente pode ser possível mesmo se não houver solução analítica para os pontos de equilíbrio.

uma versão do sinal transmitido, no caso do critério de Wiener. Um grau de conhecimento dessa magnitude é muito bem-vindo, pois torna o problema bastante tratável e, até certo ponto, garante que as soluções produzidas corresponderão a um bom aproveitamento das características da estrutura para os fins de equalização. Entretanto, a natureza prática dos problemas de telecomunicações impõe, em alguns casos, *limites à aplicabilidade de técnicas supervisionadas*. Basicamente, há dois fatores principais que respondem por tal restrição: a escassez de recursos sistêmicos<sup>45</sup>, que pode tornar assaz dispendioso empregá-los para transmitir a seqüência de treinamento, e uma impossibilidade inerente ao sistema de transmitir um “cabeçalho” do tipo requerido. Em qualquer caso, torna-se imperativo buscar formulações alternativas do problema de equalização, formulações capazes de contemplar os objetivos buscados sem o uso de informações de natureza tão precisa quanto as utilizadas nos casos anteriores. A *teoria de equalização não-supervisionada* concentra os esforços teóricos nesse sentido.

A pergunta crucial é: que tipo de critério pode indicar, de maneira precisa, o grau de sucesso na tarefa de equalizar um canal sem que, para isso, empreguem-se informações de difícil conhecimento no receptor? Cada nova resposta a essa pergunta dá à luz uma nova técnica, e a busca por um registro completo das diversas propostas excede o escopo do presente documento. Há, no entanto, dois resultados essenciais que, sem exagero, formam uma base genérica da teoria: os teoremas de Benveniste-Goursat-Ruget (BGR) e de Shalvi-Weinstein (SW).

### **O Teorema de Benveniste-Goursat-Ruget (BGR)**

Sinais de informação, devido a seu caráter incerto, revestem de estatística qualquer tipo de esforço no sentido de tratá-los. Caracterizar um ente desse tipo sem recorrer a possíveis realizações requer, portanto, *o uso de informação estatística*. Uma opção natural é tentar substituir o conhecimento das amostras do sinal pelo conhecimento de sua distribuição<sup>46</sup>.

Trata-se, sem dúvida, de uma grandeza apropriada do ponto de vista prático, pois não é problema que o receptor conheça, *a priori*, a distribuição do sinal, dado que tanto a modulação quanto o tipo de mensagem que se transmite são notórios. Resta apenas saber se informação desse tipo pode ser empregada com êxito para os fins desejados. Um ponto de partida muito interessante é relacionar a densidade do sinal transmitido àquela obtida na saída do equalizador, tendo sempre em mente que a densidade do sinal recebido será, necessariamente, distinta daquela inerente à mensagem original devido aos efeitos nocivos do meio.

Antes de qualquer novo passo, cumpre-nos assumir algumas hipóteses úteis. Em primeiro lugar, para facilitar nossa caminhada pelos inevitáveis meandros estatísticos, assumamos, como, aliás, é a regra, que as amostras são independentes e identicamente distribuídas (i.i.d.). Feito isso, resta-nos encampar mais uma hipótese: que tanto canal quanto equalizador são filtros lineares. O motivo dessa nada desprezível restrição está na enorme flexibilidade de um dispositivo não-linear genérico, que tornaria sobremaneira

---

<sup>45</sup> Tipicamente, a transmissão de uma seqüência de treinamento se reflete num dispêndio de espectro causado pelo envio de dados que não contém informação.

<sup>46</sup> A distinção entre *densidade de probabilidade* e *distribuição de probabilidade* não é particularmente relevante para nossos fins didáticos.

difícil a tarefa de tecer considerações estatísticas genéricas sobre as distribuições ao longo do sistema. Dadas tais restrições, é possível mostrar que vale o seguinte resultado [Benveniste et al., 1980]:

***Teorema de Benveniste-Goursat-Ruget (BGR):*** *Sejam  $p_s(x)$  e  $p_y(x)$ , respectivamente, as densidades de probabilidade do sinal transmitido e do sinal na saída do equalizador. Caso ambas sejam não-gaussianas, elas serão idênticas se, e somente se, a resposta combinada canal + equalizador  $c(n)$  for um impulso<sup>47</sup>, ou seja, se houver equalização perfeita.*

Em outras palavras, o teorema mostra que é possível formular a condição de equalização perfeita como uma *igualdade de densidades*, sem que seja necessário recorrer a qualquer tipo de supervisão. A exigência de não-gaussianidade é muito comum em problemas não-supervisionados, pois a saída de um sistema linear excitado por um processo gaussiano é também um processo gaussiano, o que transforma o critério de igualar as densidades, ou seja, de, em tese, igualar todos os momentos dos sinais, num problema trivial de casar média e variância i.e. de *controle de potência*.

Apesar do sucesso representado pelo teorema, devemos ter em mente que não obtivemos nada além de uma maneira de expressar uma condição ideal. Não há nada no teorema, *in abstracto*, que garanta, na prática, a possibilidade de equalização perfeita. Aliás, é possível saber, de antemão, que a condição indicada no teorema jamais se verificará no mais comum dos casos: aquele em que canal e equalizador são filtros FIR, pois, devido às limitações estruturais já discutidas na seção 3.4.1, uma solução em que a resposta combinada seja igual a um impulso não é atingível. Isso revela que a aplicação do teorema BGR a problemas práticos de caráter geral requer também um critério que indique o grau de desvio em relação à situação ideal, sendo o caso de mínimo desvio a solução desejada.

### **O Teorema de Shalvi-Weinstein (SW)**

O teorema BGR mostra que a idéia de igualar as distribuições é suficiente para caracterizar uma condição ZF no contexto de critérios sem supervisão. Entretanto, em 1994, Shalvi e Weinstein [Shalvi e Weinstein, 1990] mostraram que tal condição pode ser simplificada, ou seja, que é possível garantir equalização perfeita sem que, para isso, sejam necessários *todos os momentos*<sup>48</sup> dos sinais envolvidos. Tendo por base as mesmas condições explicitadas no preâmbulo da discussão anterior, eles propuseram o seguinte teorema:

***Teorema de Shalvi-Weinstein (SW):*** *A curtose e a variância do sinal de saída e do sinal transmitido só serão iguais se, e somente se, a resposta combinada canal + equalizador  $c(n)$  for um impulso, ou seja, se houver equalização perfeita.*

A *curtose* é uma grandeza estatística de quarta ordem definida por:

---

<sup>47</sup> Note o leitor a explícita conexão com o critério ZF.

<sup>48</sup> Implicitamente, assumimos que conhecer uma densidade significa conhecer todos os seus momentos, e vice-versa.

$$\kappa(a) = E[a^4] - 3E^2[a^2] \quad (3.36)$$

O teorema SW garante que não é necessário igualar densidades de probabilidade para que se tenha uma condição ZF: basta que o segundo momento e um momento de ordem superior dos sinais relevantes sejam idênticos. Com isso, os autores reduziram a formulação delineada pelo cenário subjacente ao teorema BGR a um mínimo, explicitando os fatores verdadeiramente essenciais ao processo de equalização não-supervisionada. Frisamos que a restrição de não-gaussianidade contínua, naturalmente, válida, bem como a necessidade de se buscar um critério de distância para que se apliquem as idéias apresentadas em casos práticos.

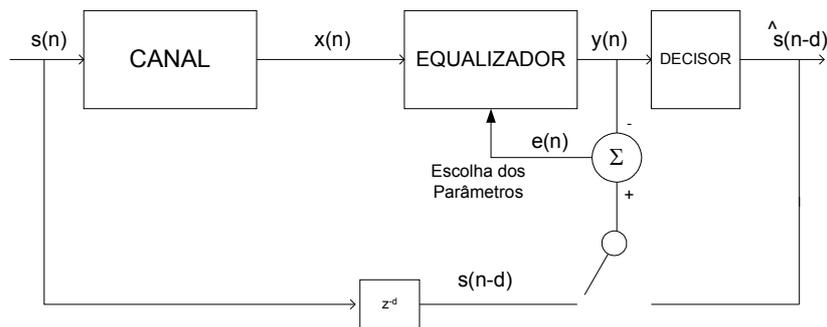
### **Critérios para Equalização Cega**

Os teoremas que apresentamos são os pilares da teoria de equalização não-supervisionada ou *cega*. Ambos, como já mencionamos, indicam novas formas de expressar o que representa *equalizar*. Vimos que a equalização pode ser entendida como uma *igualdade de distribuições* ou, no caso menos restritivo, *de momentos de ordem superior*. De posse desse conhecimento, estamos aptos a analisar um outro ângulo do problema: o dos *critérios para equalização cega*, das construções matemáticas destinadas ao ajuste dos parâmetros do filtro, e não simplesmente à expressão de determinadas condições. Em termos intuitivos, teoremas como os dois que discutimos indicam caminhos, caminhos estes que compete aos critérios trilhar quando se trata de uma aplicação prática. Embora sempre se possa ganhar entendimento recorrendo ao arcabouço conceitual que mana dos teoremas, frisamos, por uma questão de precisão histórica, que nem toda técnica de equalização cega se baseou neles. Os esforços pioneiros, particularmente, fundamentaram-se em noções intuitivas de menor alcance, noções que, embora não fossem suficientes para que delas se erigisse uma teoria, serviam à solução de problemas reais. Curiosamente, determinadas abordagens se revelaram, *a posteriori*, bem mais profundas que se imaginava.

Para que esse exórdio não se afigure deveras hermético, analisaremos brevemente dois dos principais critérios para equalização cega, tecendo, quando possível e conveniente, algumas considerações históricas e analíticas. Como conviria aos antigos, tratemos de dar a primazia aos mais velhos, analisando o critério de decisão direta<sup>49</sup> (DD). Para que o entendamos, imagine o leitor que operamos em um sistema em que é possível ter acesso a um “cabeçalho” de treinamento por um pequeno período de tempo, e que, posteriormente, só nos restaria a opção de recorrer a um ajuste cego, como mostra a Fig. 3.13.

---

<sup>49</sup> A tradução literal de *decision-directed* seria “dirigido por decisão”, em total concordância com a intuição. Entretanto, dado o papel periférico da técnica neste trabalho, optamos pela facilidade didática de manter a nomenclatura mais usual, feita a devida ressalva quanto à sua precisão.



**Figura 3.13:** Esquema Clássico de Treinamento com o Critério DD

Supondo que o treinamento, apesar de sua duração relativamente curta, tenha sido capaz de conduzir o equalizador a uma situação em que sua saída leve a poucos erros no processo de decisão, ou seja, em que valha a condição:

$$\text{dec}[y(n)] = s(n-d), \quad (3.37)$$

seria possível utilizar tal estimativa para construir um critério e um algoritmo “tipo Wiener” i.e. de “caráter” supervisionado. As escolhas naturais seriam:

$$J_{DD} = E\{\text{dec}[y(n)] - y(n)\}^2 \quad (3.38)$$

e

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu\{\text{dec}[y(n)] - y(n)\}\mathbf{x}(n) \quad (3.39)$$

Nosso objetivo não é analisar propriamente o mecanismo de treinamento apresentado na Fig. 3.13, e sim a técnica não-supervisionada em si. Sob esse prisma, toda a primeira etapa se resume a *fornecer uma condição inicial apropriada*. Isso já indica, de imediato, que *deve haver condições iniciais que levam a pontos de equilíbrio ruins*, ou seja, que o critério deve ser *multimodal*. Tais condições devem corresponder a casos em que há muitos erros de decisão, ou seja, em que não corresponde à realidade a equação (3.37). Numa situação “desesperada” como essa, só podemos mesmo esperar que o algoritmo convirja para uma solução insatisfatória.

A conexão com o critério de Wiener já foi implicitamente indicada e, sob nenhuma hipótese, deve passar despercebida. Analisando a equação (3.38), vemos que, *caso se esteja numa condição de “diagrama de olho aberto”, ou seja, em que não há erros de decisão, a função custo DD equivalerá a uma função de Wiener* [Mazo, 1980, Macchi e Eweda, 1984]. A escolha do artigo indefinido “uma” no final da frase não é acidental: *é possível, a partir da função DD, identificar diferentes regiões correspondentes à recuperação do sinal transmitido com diferentes atrasos*. A conclusão é que a função DD deve conter “trechos” de diferentes funções de Wiener, sendo que as funções supervisionadas contempladas devem estar associadas a atrasos de equalização que levam a

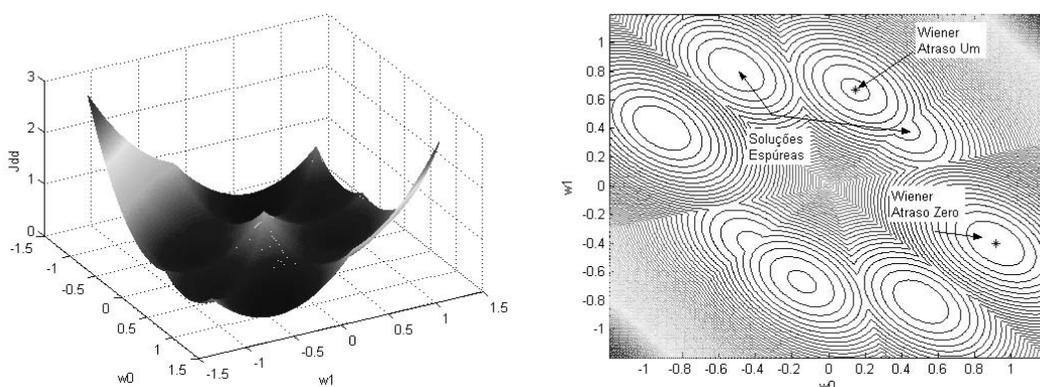
“boas” soluções, ou seja, soluções em que não haja erros de decisão<sup>50</sup>. Para clarificar essa torrente de idéias, recorramos a um exemplo.

### **Exemplo 3.3**

Consideremos o cenário apresentado no exemplo 3.2. O canal tem a seguinte função de transferência:

$$H_{\text{canal}}(z) = 1 + 0.6z^{-1} \quad (3.40)$$

e o equalizador tem dois coeficientes. Na Fig. 3.14, apresentamos a função custo DD e suas curvas de nível para este caso.



**Figura 3.14:** Função Custo DD

Observamos, de imediato, que a função custo é multimodal, ou seja, que contém vários mínimos associados a custos distintos. Note o leitor que, como é a regra em todo critério não-supervisionado, há uma simetria em relação à origem, o que faz com que os mínimos ocorram sempre aos pares.

Em perfeita concordância com o que discutimos, os dois melhores mínimos da função custo coincidem com as duas melhores soluções de Wiener (vide Tab. 3.1), o que revela que condições de olho aberto são a quintessência da relação entre os critérios. É digna de nota a presença de mínimos espúrios em regiões de “olho fechado”, mínimos estes que não guardam relação com o paradigma supervisionado.

Um outro critério particularmente representativo foi proposto por Godard em 1980 [Godard, 1980]. Sua idéia básica foi explorar as características da constelação transmitida para produzir uma medida de dispersão que indicasse, de maneira não-supervisionada, o

<sup>50</sup> Há um pequeno abuso de linguagem aqui. Quando falamos da ausência de erros de decisão, referimo-nos a erros de decisão causados por *IIS residual*, e não por ruído. Ao longo dessa explicação, aliás, recomendamos ao leitor que tenha mais presente em sua mente a IIS, considerando o ruído como uma “fricção” a ser acomodada posteriormente.

grau de proximidade da condição de equalização perfeita. Mais ainda, Godard desejava que sua medida *não dependesse da fase do sinal na saída do equalizador*, de modo a evitar problemas decorrentes de uma má recuperação de portadora na demodulação. A função custo proveniente dessas condições foi:

$$J_G = E[R_p - |y(n)|^p]^2, \quad (3.41)$$

sendo  $R_p = \frac{E[|s(n)|^{2p}]}{E[|s(n)|^p]^2}$ . O caso mais comum é aquele em que  $p = 2$ , dando origem ao *critério do módulo constante*<sup>51</sup> (CM):

$$J_{CM} = E\{R_2 - |y(n)|^2\}^2 \quad (3.42)$$

e ao correspondente *algoritmo CMA*:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu y(n)[R_2 - |y(n)|^2] \mathbf{x}(n) \quad (3.43)$$

O critério CM é uma das mais sólidas técnicas não-supervisionadas, tendo merecido muita atenção por parte dos pesquisadores nos últimos 20 anos [Attux, 2001 e Suyama, 2003]. Os primeiros resultados sobre a função custo mostraram que ela tem como mínimos condições ZF *quando elas são atingíveis*. Posteriormente, constatou-se a existência de mínimos locais em duas circunstâncias distintas, embora essa distinção não tenha sido devidamente explicitada nos trabalhos da época [Ding et al., 1991]:

- Quando o canal é um filtro IIR e o equalizador é FIR, situação em que há soluções que levam à inversão perfeita;
- Quando tanto canal quanto equalizador são filtros FIR, situação em que a inversão perfeita é uma impossibilidade teórica;

A célebre prova de Ding et al. [Ding et al., 1991] concentrou-se no primeiro caso, para o qual mostrou-se que havia, além das condições ideais, mínimos locais associados a um desempenho paupérrimo do ponto de vista da redução de IIS. Essa noção terminou por cristalizar uma idéia: *a de que a multimodalidade do critério CM era uma deficiência de sua construção*. A análise mais detalhada do segundo cenário, no entanto, terminou por indicar um caminho oposto: há uma relação entre os mínimos locais de  $J_{CM}$  e a existência de múltiplas soluções de Wiener para diferentes atrasos<sup>52</sup>.

A questão crítica era a seguinte: como se considerava que o critério de Wiener era unimodal e os critérios cegos, como o CM, multimodais, parecia haver uma barreira intransponível entre ambos os mundos. Entretanto, como apontamos, *o critério de Wiener também é multimodal se considerarmos o atraso um parâmetro livre*. Um critério não-supervisionado, por não depender diretamente de versões atrasadas específicas, deve abranger várias possibilidades. No caso do critério DD, vimos que os “bons mínimos” i.e.

<sup>51</sup> O nome do critério tem a ver com o fato de a função indicar um desvio em relação à condição de módulo constante para constelações com essa propriedade (e.g. PSK).

<sup>52</sup> Relação que, como mostramos, também existe no critério DD.

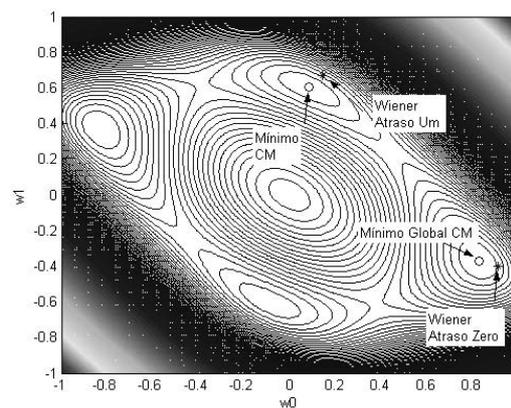
aqueles associados a uma perfeita recuperação do sinal, devem estar contemplados, acompanhados de mínimos espúrios. No caso do critério CM, não há ainda uma prova tão contundente, mas tem se mostrado bastante fundamentada a conjectura de que *todo “bom” mínimo de Wiener deve ter a si associado um mínimo CM, sendo a distância entre eles tão pequena quanto melhor for a condição geral de equalização, ou seja, o número de graus de liberdade da estrutura* [Johnson et al., 1998]. Tal conjectura, como mostraremos no capítulo 8, também tem se sustentado quando se emprega um filtro IIR como equalizador, sendo, destarte, um poderoso elo de ligação entre o critério cego e o sólido paradigma de Wiener. Explorar essa valiosa sinergia será uma das tônicas do presente trabalho.

Há ainda outro aspecto importante relacionado ao critério CM: como mostrado por Li et al. [Li et al., 1995] e, de maneira mais elegante e abrangente, por Regalia [Regalia, 1999], há uma estreita relação entre o critério CM e o critério de Shalvi-Weinstein, totalmente baseado no teorema homônimo há pouco discutido. De fato, os autores mostraram que há uma perfeita equivalência, ou seja, uma correspondência biunívoca entre as soluções decorrentes de ambos os critérios. Em outras palavras, ambos levam, em tese, às mesmas escolhas de parâmetros. Tal resultado fornece uma solidez teórica jamais imaginada por Godard, consolidando definitivamente o *status* do critério CM no *hall* de técnicas cegas. Um exemplo pode melhor ilustrar as relações e características que vimos discutindo.

---

### **Exemplo 3.4**

Consideremos o mesmo contexto delineado pelo exemplo 3.3. Nesse caso, a função custo CM terá a forma apresentada na Fig. 3.15. Há apenas dois pares de mínimos simétricos, dois pares de selas e o máximo na origem que, frisamos, é ubíquo em critérios sem supervisão. Um ponto muito importante é que os mínimos da função custo CM *estão próximos às soluções de Wiener*, em perfeita concordância com a conjectura previamente apresentada.



**Figura 3.15:** Soluções CM e de Wiener

---

## **Discussão**

A discussão das possíveis técnicas de equalização cega e de suas relações seria, sem sombra de dúvida, uma fonte capaz de verter inúmeras páginas plenas de estudos e considerações. A restrição de nosso escopo, no entanto, torna este um local inadequado para o cumprimento dessa hercúlea tarefa, restando-nos apenas tecer alguns comentários finais.

Tivemos aqui a oportunidade de expor os dois mais sólidos pilares da teoria: os teoremas de Benveniste-Goursat-Ruget e de Shalvi-Weinstein. Através deles, mostramos a possibilidade de formular os objetivos da equalização de maneira sólida sem a necessidade de conhecer o canal ou um sinal de treinamento. Paralelamente, apresentamos os que ousamos denominar mais importantes critérios para equalização cega: o critério DD e o critério CM. Ambos foram propostos a partir de intuições sólidas, mas, posteriormente, mostraram-se muito bem fundamentados teoricamente: a função custo DD emula a de Wiener nas situações de maior interesse e os pontos de equilíbrio do CMA, além de estarem ligados ao teorema de Shalvi-Weinstein, parecem guardar uma importante relação de proximidade com as melhores soluções supervisionadas. De qualquer forma, ulula uma indagação provocativa e instigante, que, felizmente, teremos a possibilidade de desenvolver em capítulos futuros: caso se disponha de uma ferramenta de busca global efetiva, não serão os critérios não-supervisionados excelentes plataformas para o projeto de receptores de Wiener?

## **Filtros Não-lineares e Equalização Cega**

Tanto os teoremas quanto os algoritmos expostos pertencem a e foram derivados de um contexto linear. Sua extensão direta para o caso não-linear é um procedimento muito arriscado, pois a flexibilidade de um mapeamento genérico, tão útil no caso supervisionado, em que abunda a informação, pode ser pernicioso quando a base de tudo é puramente estatística. A questão crucial é que *pode haver soluções que satisfaçam os critérios não-supervisionados que apresentamos, mas que não correspondam a uma boa condição de equalização*. Em outras palavras, quebra-se o elo essencial e genérico: a correspondência entre determinados critérios estatísticos cegos e a noção geral de equalização.

---

### **Exemplo 3.5**

Para ilustrar as armadilhas que abundam no caminho para estabelecer um paradigma ótimo para equalização não-linear cega, assumamos o mesmo cenário dos exemplos anteriores. Consideremos ainda que estejamos utilizando o critério CM para ajustar um mapeamento não-linear genérico  $W(\cdot)$ . Consideremos o caso em que  $W[\mathbf{x}(n)] = 1$ : trata-se de um mínimo ideal do ponto de vista do critério CM (pois o módulo constante está recuperado), mas péssimo do ponto de vista supervisionado, pois metade dos bits serão necessariamente perdidos. Esse exemplo de gritante simplicidade ilustra o tipo de dificuldade que um mapeamento muito flexível pode gerar.

---

### 3.4.4. Equalização vista como um problema de classificação de padrões

Sistemas com memória finita, tanto de natureza linear quanto não-linear, são a opção mais comum quando se busca modelar um canal de comunicação. De fato, a regra em comunicações é elaborar um modelo que compreenda um determinado número de atrasos representativos, sejam eles ocasionados pela resposta ao impulso de um dispositivo temporal ou por reflexões producentes de múltiplos percursos.

É também norma, como vimos no capítulo 2, que os possíveis valores das amostras do sinal transmitido assumam valores pertencentes a um conjunto finito de possibilidades, o qual, no modelo padrão em banda-base, recebe o nome de *constelação*. Suponhamos, sem significativa perda de generalidade, que o sinal transmitido possa assumir valores +1 e -1, ou seja, pertença a uma constelação 2-PAM ou 2-PSK. Assumamos ainda que o canal de memória finita tenha a seguinte relação entrada-saída:

$$x(n) = H[s(n), s(n-1), \dots, s(n-M+1)] \quad (3.44)$$

onde  $H(\cdot)$  é um mapeamento não-linear genérico<sup>53</sup> de ordem  $M$ . Consideremos agora que desejemos projetar um equalizador com uma linha de atrasos de ordem  $N-1$ , ou seja, que a saída seja do tipo:

$$y(n) = W[x(n), x(n-1), \dots, x(n-N+1)] \quad (3.45)$$

sendo  $W(\cdot)$  um mapeamento não-linear genérico.

A cada possível vetor de entrada  $\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-N+1)]$  corresponde uma seqüência  $\mathbf{s}(n) = [s(n), s(n-1), \dots, s(n-N-M+2)]$ , de acordo com (3.44). Isso significa que, descontada a possibilidade de repetição, haverá  $2^{N+M-1}$  possíveis vetores ou *estados* representativos do canal. Na ausência de ruído, nada há na entrada do equalizador que não pertença a essa classe: *os estados esgotam as possibilidades devidas à interferência intersimbólica*.

A cada estado, como indicam as equações (3.44) e (3.45), associa-se um conjunto de amostras do sinal transmitido. Isso significa que podemos *rotular* um certo estado de  $N+M-1$  diferentes maneiras, *de acordo com o atraso de interesse*. Como assumimos um caso binário, há dois possíveis rótulos: +1 e -1. Uma vez que tenhamos escolhido um atraso, devemos impô-lo aos rótulos associados a todos estados. Isso cria um cenário com padrões e duas classes distintas, típico de *um problema de classificação*. O exemplo 3.6 ilustrará o processo.

---

#### Exemplo 3.6

Mantenhamos aqui o mesmo cenário dos exemplos anteriores, com o canal apresentado em (3.40) e um equalizador com duas entradas. Nesse caso, os estados e seus

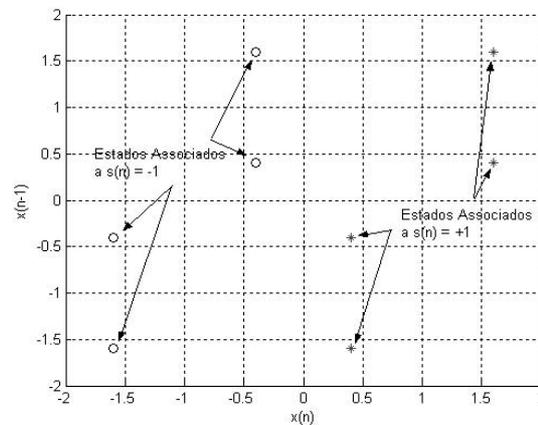
---

<sup>53</sup> Não julgamos necessário esmiuçar excessivamente as restrições que devem ser impostas a esse mapeamento, pois elas não terão um papel relevante aqui.

possíveis rótulos serão como mostra a Tab. 3.3. Na Fig. 3.16, apresentamos o problema resultante para atraso nulo.

$x(n)$	$x(n-1)$	$s(n)$	$s(n-1)$	$s(n-2)$
1.6	1.6	+1	+1	+1
1.6	0.4	+1	+1	-1
0.4	-0.4	+1	-1	+1
0.4	-1.6	+1	-1	-1
-0.4	1.6	-1	+1	+1
-0.4	0.4	-1	+1	-1
-1.6	-0.4	-1	-1	+1
-1.6	-1.6	-1	-1	-1

**Tabela 3.3:** Estados e Possíveis Rótulos

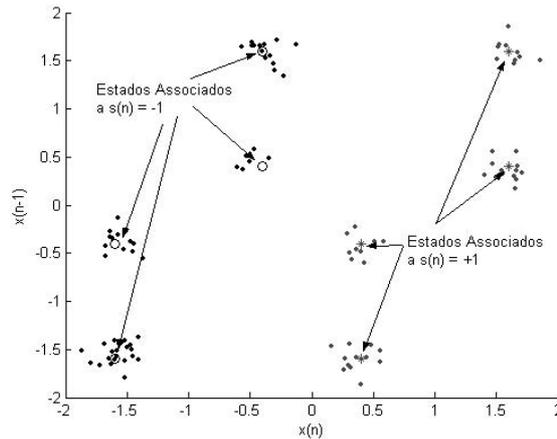


**Figura 3.16:** Problema de Classificação

A configuração apresentada na Fig. 3.16 é linearmente separável, o que é razoável tendo em vista que, como vimos nos exemplos anteriores, um equalizador linear com atraso zero tem um EQM residual associado relativamente pequeno. Entretanto, como veremos, a presença de ruído pode tornar importante o caráter não-linear do separador mesmo que o problema seja tratável linearmente.

O cenário que delineamos corresponde, em última análise, a um problema de equalização, pois buscamos identificar uma versão de  $s(n)$  a partir de vetores de amostras recebidas. No entanto, as restrições de modelo apontadas nos trazem um novo foco: em vez de aproximar sinais ou recuperar estatísticas de ordem superior, desejamos impor alguma “ordem” a alguns padrões rotulados. A natureza dessa “ordem” é a essência do problema de equalização: obter, após o processo de decisão, um sinal reconstruído que corresponda a uma versão fidedigna de  $s(n)$ . No problema de classificação que delineamos, isso seria equivalente a *obter uma superfície de separação entre as classes, de tal forma que sejam agrupados em classes distintas os estados associados a padrões distintos*. Obtida a curva de separação, o processo de decisão se encarrega de produzir, associando a cada padrão recebido o rótulo adequado, a estimativa da mensagem enviada.

Para tornar o problema ainda mais realista, assumamos que o canal, além dos efeitos de IIS indicados em (3.40), introduz também ruído aditivo com amostras i.i.d. e densidade igual a  $p_r(x)$ . Isso faz com que *haja uma “nuvem” de amostras recebidas ruidosas em torno dos estados do canal*, como mostra, para um caso em que ruído é aditivo, branco e gaussiano (AWGN), a Fig. 3.17.



**Figura 3.17:** Problema de Classificação – AWGN

Com isso, estamos aptos a sumarizar o problema que temos em mãos: encontrar uma superfície que separe adequadamente os estados do canal, sendo, portanto, capaz de classificar, *com o menor erro possível*, as amostras recebidas. Tal problema é bastante similar ao que apresentamos na seção 2.10: obter uma fronteira de decisão ótima segundo um critério de mínima probabilidade de erro. A abordagem que empregaremos é a mesma, baseada no critério de máxima verossimilhança<sup>54</sup>. A base de tudo é, portanto, maximizar a seguinte probabilidade condicional:

$$\max_s p[s(n-d) / \mathbf{x}(n)] \quad (3.46)$$

i.e. descobrir qual opção,  $s(n-d) = -1$ , ou  $s(n-d) = +1$ , torna máxima a probabilidade relativa ao vetor de entrada recebido. Podemos definir a seguinte diferença:

$$\mathfrak{J}[\mathbf{x}(n)] = p[s(n-d) = +1 / \mathbf{x}(n)] - p[s(n-d) = -1 / \mathbf{x}(n)], \quad (3.47)$$

que será positiva se a hipótese  $s(n-d) = +1$  for a mais provável e negativa se  $s(n-d) = -1$  for a hipótese a ser escolhida. Tratemos, pois, de desenvolvê-la.

A regra de Bayes [Papoulis, 1991] nos permite escrever:

$$p[s(n-d) = s / \mathbf{x}(n)] = \frac{p[\mathbf{x}(n) / s(n-d) = s] p[s(n-d) = s]}{p[\mathbf{x}(n)]} \quad (3.48)$$

Dadas as hipóteses usuais, temos que:

<sup>54</sup> Assumimos aqui que os símbolos transmitidos +1 e -1 são equiprováveis.

$$p[s(n-d) = +1] = p[s(n-d) = -1] \quad (3.49)$$

Como  $p[\mathbf{x}(n)]$  é a mesma para ambas as hipóteses, percebemos, a partir da análise de (3.48), que a grandeza que determina o sinal de (3.47) é  $p[\mathbf{x}(n) / s(n-d) = s]$ . Com isso, vale a simplificação:

$$\mathfrak{I}[\mathbf{x}(n)] = p[\mathbf{x}(n) / s(n-d) = +1] - p[\mathbf{x}(n) / s(n-d) = -1], \quad (3.50)$$

Denominando  $X^+$  o conjunto de todos os estados do canal associados a  $s(n-d) = +1$  e  $X^-$  o conjunto de todos os estados associados a  $s(n-d) = -1$ , temos, com base na hipótese de independência das amostras transmitidas:

$$p[\mathbf{x}(n) / s(n-d) = +1] = \frac{2}{N_E} \sum_{X^+} p[\mathbf{x}(n) / \mathbf{x}_j] \quad (3.51)$$

sendo  $N_E$  o número de estados e  $\mathbf{x}_j$  o  $j$ -ésimo estado associado ao rótulo +1. Analogamente, é possível escrever:

$$p[\mathbf{x}(n) / s(n-d) = -1] = \frac{2}{N_E} \sum_{X^-} p[\mathbf{x}(n) / \mathbf{x}_i] \quad (3.52)$$

Ora, as probabilidades do tipo  $p[\mathbf{x}(n) / \mathbf{x}_i]$  são determinadas pelo ruído. Se ele tiver uma forma tal que  $p[\mathbf{x}(n) / \mathbf{x}_i] = f(\|\mathbf{x}(n) - \mathbf{x}_i\|)$ , sendo  $f(\cdot)$  a densidade de probabilidade de suas amostras, teremos:

$$\mathfrak{I}[\mathbf{x}(n)] = \frac{2}{N_E} \sum_{X^+} f(\|\mathbf{x}(n) - \mathbf{x}_j\|) - \frac{2}{N_E} \sum_{X^-} f(\|\mathbf{x}(n) - \mathbf{x}_i\|) = \frac{1}{N_E} \sum_{k=1}^{N_E} w_k f(\|\mathbf{x}(n) - \mathbf{x}_k\|) \quad (3.53)$$

onde  $w_k = 1$  se  $\mathbf{x}_k \in X^+$  e  $w_k = -1$  se  $\mathbf{x}_k \in X^-$ . Para que obtenhamos um elo direto dessa expressão final com o processo de filtragem, e, conseqüentemente, com os outros paradigmas de equalização, basta que assumamos que a escolha ditada pelo mapeamento é, no fundo, a atuação natural de um decisor, e que  $\mathfrak{I}[\cdot]$  é a relação entrada-saída de um filtro não-linear, o *equalizador de Bayes* ou *bayesiano*.

O projeto do equalizador bayesiano requer que sejam conhecidos:

- A posição dos estados do canal i.e. os centros das funções  $f(\cdot)$ .
- A forma e os parâmetros das funções  $f(\cdot)$ , que se relacionam com as características do processo estocástico que emula o ruído.
- Os membros das classes  $X^+$  e  $X^-$ , ou seja, os rótulos de cada estado do canal.

Para satisfazer a primeira exigência, tradicionalmente se emprega um processo de *clusterização*, que, em tese, deve ser capaz de obter os estados do canal como centros do conjunto de nuvens de ruído recebidas. A segunda exigência é normalmente satisfeita através da estimação de momentos de uma função que se considera conhecida *a priori*. Por

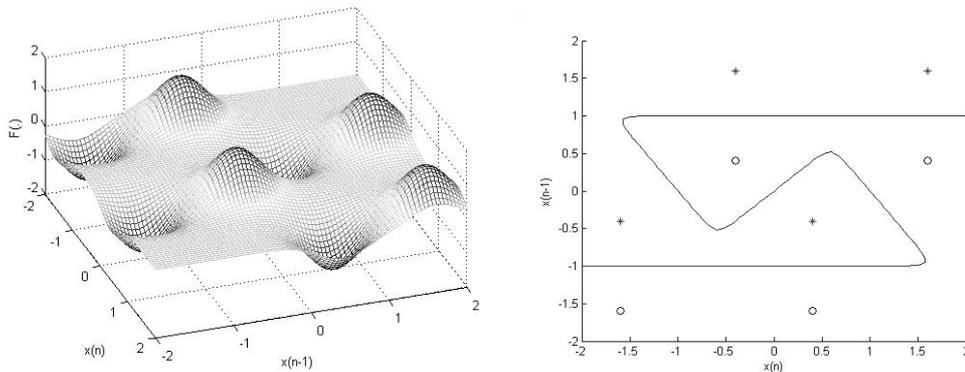
fim, o último requisito torna o equalizador de Bayes um paradigma supervisionado, pois exigirá uma etapa de treinamento com um sinal piloto. Frisamos que as três etapas não precisam ter lugar em sucessão, podendo ser efetuadas conjuntamente caso isso seja interessante do ponto de vista das técnicas de otimização e estimação disponíveis.

O caráter não-linear de  $f(\cdot)$  nos casos de interesse faz com que o equalizador de Bayes não possa ser uma estrutura linear. Isso significa que *a obtenção do equalizador ótimo para estimação símbolo-a-símbolo requer a introdução de não-linearidades no modelo do dispositivo*. Mostra também que paradigmas como o de erro quadrático médio e as estratégias não-supervisionadas não são, ao menos *a priori*, ótimas do ponto de vista de minimização da taxa de erro de símbolo (SER). Ambas as constatações mostram que o uso de estruturas não-lineares de filtragem *traz um ganho significativo em termos de desempenho* mesmo em problemas lineares, elevando assim o patamar de expectativas do projetista à vizinhança das possibilidades teóricas.

Neste trabalho, interessam-nos particularmente dois tipos de ruído, conforme discutimos em 2.9.2: o ruído gaussiano e o ruído impulsivo alfa-estável. Portanto, mostraremos a seguir os receptores de Bayes para ambos os casos em um canal com a mesma característica em termos de IIS. Também aproveitaremos os subsídios para discutir a questão do atraso de equalização.

### Exemplo 3.7

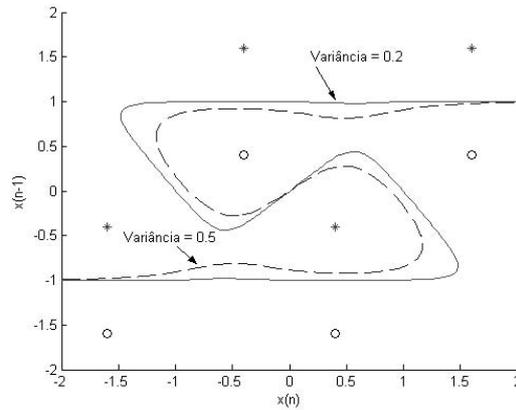
Assumamos o modelo usual de canal e sinal transmitido, mas tomemos desta vez, para ilustrar também o problema do atraso,  $d = 2$ . Na Fig. 3.18, apresentamos a função  $\mathfrak{J}[\cdot]$  e a curva de separação dela decorrente para um ruído gaussiano com  $\sigma^2 = 0.1$ .



**Figura 3.18:** Superfície de Bayes e Solução para a Tarefa de Classificação

Nota-se que essa escolha de atraso leva a um problema insolúvel por uma estrutura linear, o que nos ajuda a entender o mau desempenho da solução de Wiener apresentada na Tab. 3.1 para o caso em questão. Trata-se, destarte, de um problema em que o canal é linear, mas *apenas uma estrutura não-linear* pode resolvê-lo adequadamente. A curva de separação de Bayes, para essa magnitude de ruído, tem quase um caráter linear por partes, estabelecendo um padrão aproximado de distância mínima entre estados. Conforme mostra

a Fig. 3.19, o aumento da potência do ruído leva a uma curva mais suave, reflexo da maior sobreposição das diversas gaussianas multidimensionais.

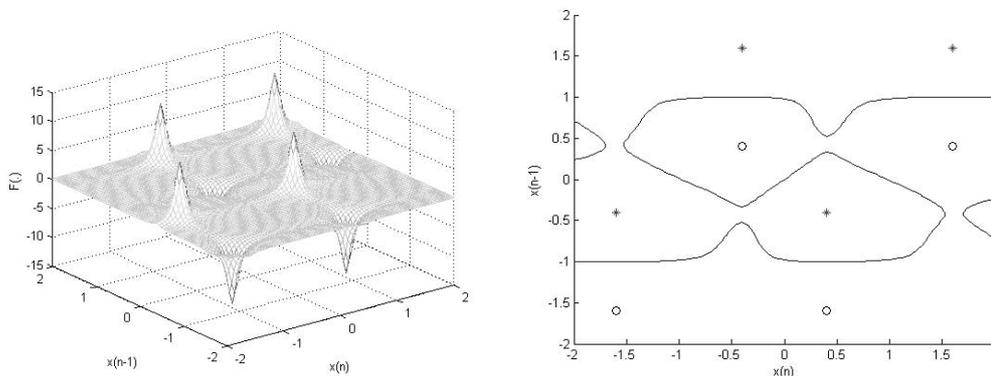


**Figura 3.19:** Influência da Potência do Ruído na Curva de Separação

Passemos agora a uma análise do receptor de Bayes para um ruído impulsivo. Assumamos que a densidade alfa-estável corresponde a uma densidade de Cauchy do tipo:

$$p_r(x) = \frac{1}{\pi(1+x^2)} \quad (3.54)$$

Na Fig. 3.20, apresentamos a superfície  $\mathfrak{I}[\cdot]$  e a curva de separação. Nota-se aqui um padrão menos semelhante ao “bem-comportado” esquema de mínima distância típico do ruído gaussiano. A explicação para isso é a maior complexidade da expressão para a probabilidade  $p[\mathbf{x}(n) / s(n-d) = s]$ , que não admite  $p[\mathbf{x}(n) / \mathbf{x}_i] = f(\|\mathbf{x}(n) - \mathbf{x}_i\|)$ . Fica estabelecido aqui o receptor ótimo em termos de probabilidade de erro de símbolo para ruído impulsivo [Georgiadis, 2000].



**Figura 3.20:** Superfície de Bayes e Solução para Ruído Impulsivo de Cauchy

### ***Interlúdio: Reflexões sobre a Relação entre o paradigma de classificação e as demais abordagens***

Uma das primeiras contribuições obtidas no contexto do presente trabalho foi uma investigação [Attux e Romano, 2002] das relações entre os paradigmas de classificação e as abordagens anteriores (Wiener, CM etc.). A idéia subjacente ao esforço é explorar a finitude do conjunto de possíveis estados do canal na ausência de ruído para representar de maneira mais simples as funções custo associadas aos critérios de equalização. Para tanto, assumamos uma função custo genérica com a seguinte forma:

$$J = E \{k(n) - G[\mathbf{w}, \mathbf{x}(n)]\}^2, \quad (3.55)$$

sendo  $k(n)$  um processo aleatório,  $\mathbf{w}$  o vetor de parâmetros do equalizador e  $\mathbf{x}(n)$  o vetor de entrada. Assumindo que o canal tem memória finita e que não há ruído, a variável aleatória cuja esperança desejamos calcular será *discreta*, permitindo-nos escrever:

$$J = \sum_{i=1}^{N_E} \text{prob}(\mathbf{x}_i) [k_i - G(\mathbf{w}, \mathbf{x}_i)]^2 \quad (3.56)$$

onde  $\text{prob}(a)$  é a probabilidade de ocorrência do evento  $a$ ,  $\mathbf{x}_i$  é o  $i$ -ésimo possível vetor  $\mathbf{x}(n)$  e  $k_i$  é o  $i$ -ésimo valor possível de  $k(n)$ . Assumindo, como de costume, equiprobabilidade, chegamos a:

$$J = \frac{1}{N_E} \sum_{i=1}^{N_E} [k_i - G(\mathbf{w}, \mathbf{x}_i)]^2 \quad (3.57)$$

Analisando a equação (3.57), chegamos a uma importante conclusão: é possível formular a tarefa de equalização, mesmo de equalização cega, como um problema de *mínimos quadrados*. Note o leitor que essa abordagem não guarda qualquer relação com os tradicionais métodos de mínimos quadrados de filtragem adaptativa (como o RLS), estando relacionada apenas ao entendimento *da função custo*, e não de técnicas para otimizá-la. Analisemos agora como os critérios de equalização apresentados se encaixam na expressão apresentada em (3.57):

#### **Wiener**

Comparando as funções (3.57) e (3.23), percebemos que:

$$k_i = s_i \quad (3.58)$$

e

$$G(\mathbf{w}, \mathbf{x}_i) = W(\mathbf{w}, \mathbf{x}_i) \quad (3.59)$$

sendo  $W$  o mapeamento efetuado pelo equalizador.

### DD

Comparando (3.57) e (3.38), percebemos que:

$$k_i = \text{dec}[W(\mathbf{w}, \mathbf{x}_i)] \quad (3.60)$$

e

$$G(\mathbf{w}, \mathbf{x}_i) = W(\mathbf{w}, \mathbf{x}_i) \quad (3.61)$$

### CM

A partir de (3.57) e (3.38), notamos que:

$$k_i = R_2 \quad (3.62)$$

e

$$G(\mathbf{w}, \mathbf{x}_i) = |W(\mathbf{w}, \mathbf{x}_i)|^2 \quad (3.63)$$

O critério de Wiener é, como se espera, a técnica mais identificada com o espírito da idéia de equalização, por buscar a situação em que:

$$W(\mathbf{w}, \mathbf{x}_i) = s_i, \forall i \quad (3.64)$$

Os outros dois critérios, por serem não-supervisionados, procuram obter resultados análogos a partir de escolhas alternativas em que não haja dependência direta de  $s_i$ . A simples análise dessas escolhas nos permite tirar algumas conclusões imediatas importantes:

- A relação entre os critérios DD e de Wiener, como já apontamos, depende essencialmente da relação entre suas escolhas de  $k_i$ , pois, para ambos,  $G(\mathbf{w}, \mathbf{x}_i) = W(\mathbf{w}, \mathbf{x}_i)$ . Assim, o custo associado a essas técnicas será idêntico caso  $\text{dec}[W(\mathbf{w}, \mathbf{x}_i)]$  seja igual a  $s_i$  em todos os possíveis estados.
- A relação entre os critérios CM e Wiener é mais complexa, pois há diferenças tanto na escolha de  $k_i$  quanto na escolha de  $G(\mathbf{w}, \mathbf{x}_i)$ . Além disso, no caso do critério CM, procura-se aproximar  $k_i$  de  $|W(\mathbf{w}, \mathbf{x}_i)|^2$  e não de  $W(\mathbf{w}, \mathbf{x}_i)$ , como nos outros casos. *Isso explica a discrepância entre soluções próximas de Wiener e CM, observada nos casos em que não há inversão perfeita (vide a Fig. 3.15): o critério CM otimiza uma função  $G(\cdot)$  diferente de  $W(\cdot)$ , ou seja, utiliza os parâmetros ótimos de  $G(\cdot)$  no mapeamento  $W(\cdot)$ , para o qual os parâmetros não são, pelo menos em tese, ótimos do ponto de vista de mínimos quadrados!*

Analisemos agora, mais detidamente, como essa abordagem, relacionada ao cenário de classificação, pode nos ajudar a compreender as relações entre os critérios DD e de Wiener<sup>55</sup>. Em primeiro lugar, atenhamo-nos, sem perda de generalidade, ao caso em que as amostras transmitidas são binárias (+1 ou -1) e i.i.d. As funções custo definirão problemas do seguinte tipo: dado o conjunto de pontos no espaço formado pelos estados e pelo rótulo de cada estado, encontre o mapeamento  $W(\cdot)$  “mais próximo” ao conjunto, no sentido clássico do problema de mínimos quadrados.

A condição

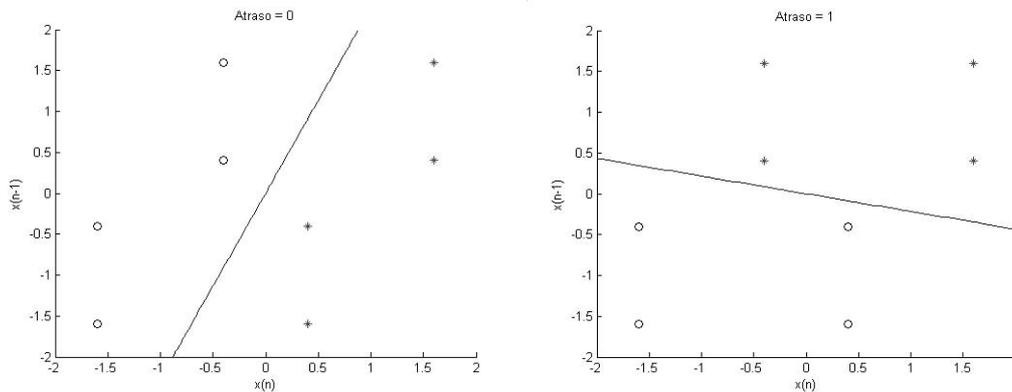
$$W(\mathbf{w}, \mathbf{x}_i) = 0 \quad (3.65)$$

revela a superfície de separação fornecida pelo equalizador, ou seja, indica a forma pela qual ele *classifica* os vetores recebidos. Isso é essencial quando se analisa a função DD, pois ela depende diretamente da saída do decisor.

Para que  $\text{dec}[W(\mathbf{w}, \mathbf{x}_i)] = s_i$ , é preciso que a superfície de separação esteja em uma boa condição de Wiener, ou seja, *que todos os estados pertencentes à classe  $X^+$  estejam devidamente separados daqueles pertencentes à classe  $X^-$* . Nesse caso, teremos que:

$$J_{DD} = J_W \quad (3.66)$$

A Fig. 3.21 traz exemplos de situações em que isso se dá, tendo por pano de fundo o mesmo cenário dos exemplos anteriores.



**Figura 3.21:** Condições de Separação Linear – Atraso nulo e unitário

A coincidência de funções que emerge de uma situação de separação perfeita é a essência da relação entre os critérios DD e de Wiener. Nos trabalhos clássicos que relacionam ambos os critérios, essa condição é entendida como *uma condição de olho aberto*, ou seja, exatamente uma situação em que a análise dos pulsos mostra que há uma boa margem para decisão. Entretanto, há uma diferença essencial que não poderia ter sido explorada à época: *nem toda condição de olho aberto corresponde a uma condição de separação*. Em particular, a correspondência poderá se quebrar quando houver estruturas não-lineares envolvidas, como já indicamos no exemplo 3.5. Isso significa que pode haver

<sup>55</sup> Tarefa que, como o leitor perceberá, irá nos ajudar a entender a relação entre as abordagens mais tradicionais de equalização e a abordagem de classificação.

casos em que haja uma situação de olho aberto, ou seja, em que a saída do equalizador esteja sempre próxima a valores +1 ou -1, sem que a separação efetuada pelo equalizador seja efetiva. Já no caso em que o mapeamento é linear, ou seja, em que vale:

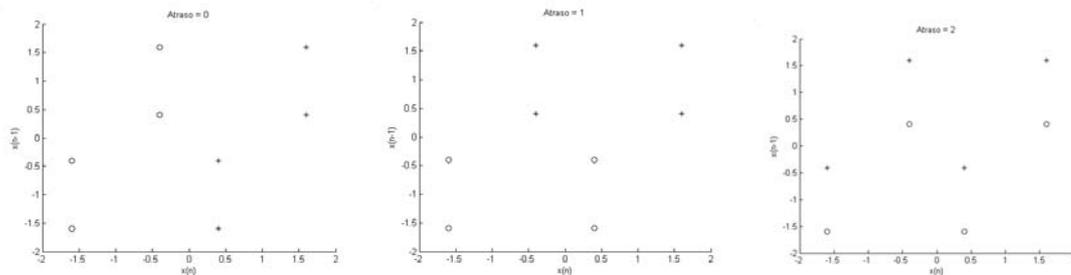
$$W(\mathbf{w}, \mathbf{x}_i) = \mathbf{w}^T \mathbf{x}_i, \quad (3.67)$$

podemos tirar conclusões mais sólidas sobre a relação entre os critérios. Nesse caso, um valor baixo de custo para funções cegas como a do critério DD *corresponderá efetivamente a uma equalização apropriada*. Nesse domínio particular, concluímos que todos os bons mínimos de Wiener, ou seja, aqueles correspondentes a condições de separação linear e abertura de olho, *serão os melhores mínimos de  $J_{DD}$* . Os mínimos de Wiener que não correspondem a condições de separação não deverão ser mínimos do critério DD, pois nessa condição não vale  $\text{dec}[W(\mathbf{w}, \mathbf{x}_i)] = s_i$  (vide Fig. 3.14). Isso conclui nossa exposição.

### A influência do atraso

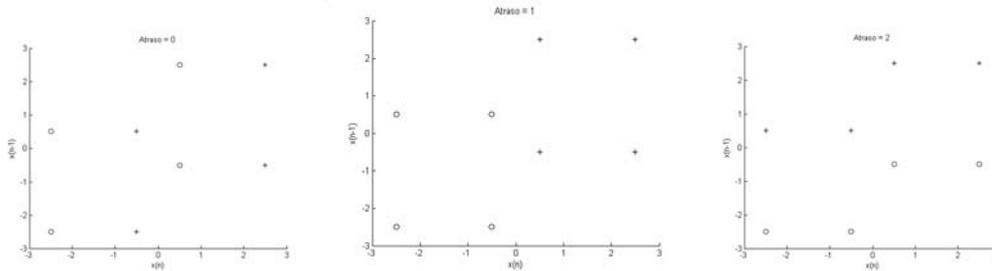
Temos tocado num ponto muito relevante: *a importância da escolha do atraso de equalização*. Como os vários exemplos mostrados indicam, a posição dos estados do canal *depende exclusivamente de sua resposta ao impulso*. Entretanto, os rótulos de cada um desses estados, decisivos para que se caracterize o problema de equalização, dependerão do atraso escolhido.

A escolha do atraso leva a problemas de classificação mais ou menos complexos, de acordo com as características do canal. Suponhamos mais uma vez que nosso canal seja  $h(n) = [1 \ 0.6]$ , que nosso equalizador tenha dois parâmetros, e que o sinal transmitido seja binário e i.i.d. Mostramos na Tab. 3.3 os estados do canal e seus possíveis rótulos e, na Fig. 3.22, apresentamos os três problemas de classificação resultantes.



**Figura 3.22:** Possíveis Problemas de Classificação

Observamos mais uma vez que os atrasos menores levam a problemas solucionáveis através de um separador linear, enquanto que o caso em que  $d = 2$  exigirá uma estrutura de filtragem não-linear. Esse estado de coisas se liga, de maneira não trivial, às características da resposta ao impulso do meio. Para que constatemos isso, mostramos na Fig. 3.23 a situação para o canal  $h = [1 \ 1.5]$ . Ocorre, nesse caso, a situação inversa: o maior atraso é o melhor, e o atraso  $d = 0$  requer um equalizador não-linear.



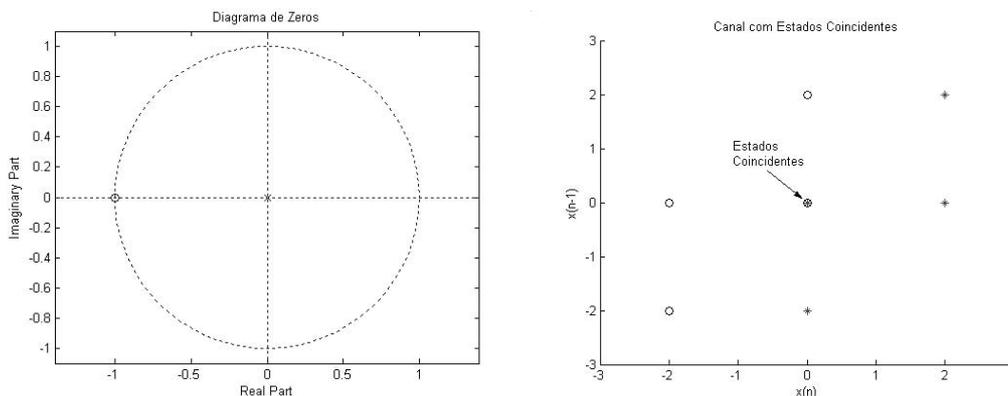
**Figura 3.23:** Possíveis Problemas de Classificação

A dificuldade de escolher, *a priori*, um atraso, sem que se tenha pleno conhecimento do canal, é um outro fator que motiva o uso de estruturas não-lineares, pois usar um filtro desse tipo representa uma segurança de que problemas não-separáveis linearmente podem ser resolvidos.

### Zeros sobre a circunferência de raio unitário

A posição dos zeros do canal de comunicação também pode se refletir em problemas de classificação mais ou menos complexos. Como vimos em 2.11.2, a chave de tudo está no caso limite em que os zeros se aproximam da circunferência de raio unitário, fronteira entre canais de fase mínima e não-mínima. Em termos espectrais, situações desse tipo levam a nulos profundos, cuja “inversão” por meio de um equalizador exigiria picos na resposta em frequência, tanto maiores quanto mais próximo estivessem os zeros.

Do ponto de vista da classificação, a proximidade dos zeros da circunferência de raio unitário pode levar a uma situação em que haja configurações com estados muito próximos, que podem ser de muito difícil separação. É de particular interesse o caso em que há zeros sobre +1 ou -1. Como mostrado em [Montalvão et al., 1998], nesse caso *haverá estados coincidentes*, ou seja, *estados idênticos com rótulos distintos*. Sob tal circunstância, sempre haverá erros de classificação, mesmo que o ruído tenha potência ínfima, fato que explicita uma degradação significativa da informação transmitida. A Fig. 3.24 ilustra um cenário desse tipo.



**Figura 3.24:** Diagrama de Zeros do Canal para Situação com Estados Coincidentes

De nada adiantará aumentar a ordem do equalizador ou atribuir a limitação à flexibilidade da estrutura: nenhum dispositivo FIR será capaz de separar os estados. Entretanto, caso se

introduza um laço de realimentação, o problema de separação poderá ser factível, pois, indiretamente, a informação sobre o passado da saída fornecerá subsídios para diferenciar os estados coincidentes [Silva et al., 2000].

### 3.4.5. Finale: Alguns Comentários sobre o Equalizador de Máxima Verossimilhança

Embora a formulação de *máxima verossimilhança (MV)* seja um tanto distinta da idéia estrutural de equalização que vimos construindo, julgamos essencial apresentá-la de maneira resumida neste capítulo. Sua idéia básica é *buscar a seqüência de símbolos mais verossímil dado o conjunto de dados recebidos*. Trata-se, portanto, de um procedimento que visa recuperar uma *seqüência de símbolos*, ao contrário dos procedimentos de equalização que formam o cerne de nosso trabalho, que operam numa base *símbolo-a-símbolo*. Destarte, podemos entender o processo de MV como *uma extensão das idéias gerais da equalização de Bayes*.

A idéia básica é maximizar a probabilidade:

$$p(\mathbf{x}/\mathbf{s},\mathbf{h}) \quad (3.68)$$

sendo  $\mathbf{x}$  uma seqüência de amostras recebidas,  $\mathbf{s}$  a correspondente seqüência de símbolos transmitidos e  $\mathbf{h}$  um vetor com os elementos da resposta ao impulso do canal, que, por simplicidade, supomos ser linear. Dessa maneira, a base de tudo é procurar a seqüência transmitida que maximiza a probabilidade mostrada em (3.68) para um certo vetor  $\mathbf{x}$  recebido. A relação entre ambas as grandezas depende do ruído e de  $\mathbf{h}$ .

Usualmente, o processo de otimização se dá por meio de uma busca guiada por uma estimativa de  $\mathbf{h}$ , busca esta que pode ser decisivamente simplificada através do uso do célebre algoritmo de Viterbi, resultado que originalmente pertenceu ao campo dos códigos convolucionais [Proakis, 1995]. Caso se procure solucionar o problema sem explicitar a resposta  $\mathbf{h}$ , torna-se importante dispor de um adequado método de busca. Apresentaremos nossa proposta de solução [Attux et al., 2004b] no capítulo 11. Por ora, são suficientes essas noções básicas, que, aliás, concluem nosso debate sobre equalização.

## 3.5. Identificação

Construir modelos de sistemas reais é um procedimento que, em muitos aspectos, se confunde com a própria natureza da investigação científica. Seu caráter geral pode ser entendido, de maneira muito simples, a partir das três etapas descritas na seção 3.1: *escolha a estrutura, o critério e a ferramenta de otimização*. No primeiro passo, é preciso considerar cuidadosamente o gênero de sistema real com que se lida, tendo sempre em mente o potencial e as limitações de dispositivos lineares e não-lineares. No segundo passo, é necessário formular um critério de escolha dos parâmetros do modelo e, no terceiro, uma maneira de encontrá-los.

### 3.5.1. Escolha da Estrutura

Apesar de completo, o cenário acima não esmiúça algumas marcantes idiossincrasias do *problema de identificação*. Um ponto de relevo neste problema é a escolha da estrutura de filtragem utilizada, ou seja, do *modelo em estado bruto*. A opção dependerá de maneira particularmente dramática do grau de conhecimento do sistema e também do tipo de informação que se deseja obter. Em vários sistemas reais, é possível postular *a priori* a forma geral da relação entre as mais importantes variáveis envolvidas i.e. seu caráter estrutural. Destarte, o problema de identificação se resumiria a uma tarefa de *obter parâmetros*, passando ao campo dos fatores conhecidos, em grande extensão, a estrutura do modelo. Quando isso se dá, normalmente é possível atribuir um sentido físico aos parâmetros, ou seja, associar a determinadas grandezas do modelo um significado teórico coerente. Problemas de identificação desse tipo recebem o nome de *paramétricos*.

Uma situação bem diferente tem lugar quando o sistema que desejamos identificar tem para nós o aspecto de uma *caixa-preta*, ou seja, quando há pouca informação *a priori* sobre ele. Nesse caso, a tarefa primordial de identificação passa a ser a elaboração de um modelo que aproxime o mapeamento realizado pelo sistema em um conjunto de padrões de treinamento, e que, espera-se, generalize o comportamento do ente desconhecido quando houver situações não contempladas na fase de ajuste<sup>56</sup>. Nesse cenário de *identificação não-paramétrica*, ganham destaque estruturas não-lineares com capacidade de aproximação universal, como, por exemplo, redes neurais.

### 3.5.2. Identificação Supervisionada

Uma vez que o projetista tenha escolhido uma estrutura, voltamos ao velho problema de ajustar seus parâmetros livres. Assim como na seção 3.4, deparamo-nos com a esfingica questão: qual o grau de informação de que dispomos? Antes de enfrentar o temido enigma, convém que nos indaguemos mais detidamente sobre qual seria a configuração ideal de um esquema de identificação.

Basicamente, desejamos que modelo seja tão próximo quanto possível do sistema (ou *planta*) que ignoramos. Para que sejamos capazes de quantificar essa proximidade, precisamos buscar uma medida que independa da estrutura sistêmica em si, mas que seja compreensível em termos de grandezas mensuráveis. Uma estratégia natural poderia ser excitar tanto a planta quanto o modelo com o mesmo sinal e verificar o grau de semelhança entre seus sinais de saída. Em outras palavras, imporíamos a ambos os modelos uma mesma entrada que, preferencialmente, fosse capaz de estimular diversos aspectos da resposta da planta, e colheríamos uma referência para ajustar os parâmetros do modelo. Caso o sinal de entrada fosse suficientemente significativo<sup>57</sup> e as saídas fossem próximas, poderíamos, com segurança, cancelar o modelo obtido. A Fig. 3.25 traz uma representação de tal esquema.

---

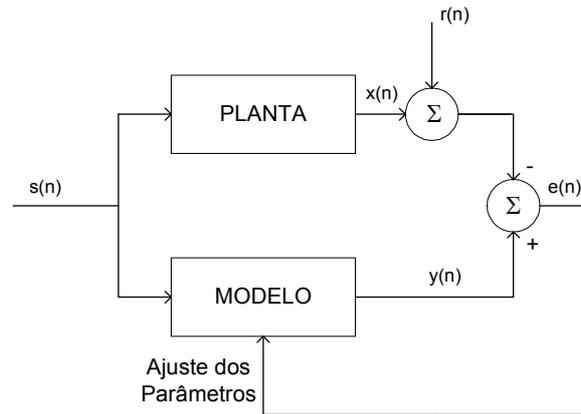
<sup>56</sup> Em alguns casos, pode ser que a obtenção desse modelo acabe por levar a uma formulação paramétrica *a posteriori*.

<sup>57</sup> Classicamente, o sinal é um ruído branco ou com amplo espectro.

Assim como no caso de equalização, é bastante natural que se utilize o erro quadrático médio entre as saídas como uma medida de adequação do modelo. Portanto, temos novamente o critério de Wiener e sua respectiva função custo:

$$J_w = E[x(n) + r(n) - y(n)]^2 \quad (3.69)$$

onde  $x(n)$  é o sinal na saída da planta,  $r(n)$  é um ruído de medição e  $y(n)$  é a saída do modelo.



**Figura 3.25:** Esquema de Identificação Supervisionada

Todo o desenvolvimento mostrado em (3.24-3.29) vale para o caso de um modelo linear nos parâmetros, assim como valem as considerações gerais acerca dos algoritmos LMS e RLS tecidas na seção 3.4.2. Nesse caso, entretanto, a análise das soluções obtidas é bem distinta daquela empreendida anteriormente.

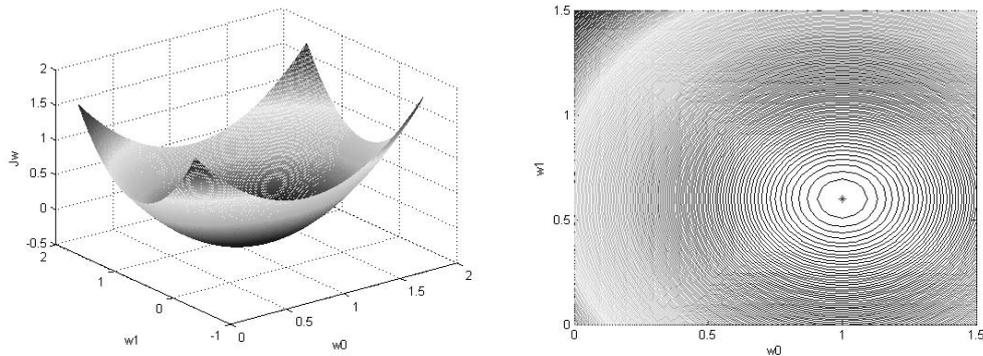
As soluções ideais, em identificação, são aquelas em que a planta é idêntica ao modelo, pelo menos no domínio abrangido pelo conjunto de treinamento e pelas características do sinal de excitação. Caso o problema seja não-paramétrico, o erro normalmente não deverá se anular, mas, se as escolhas de projeto tiverem sido apropriadas, poderá ser arbitrariamente pequeno. A existência de ruído, em tese, não deve modificar a posição do ótimo de Wiener, pois se espera que não haja correlação entre ele e o sinal desejado. Entretanto, caso se utilizem algoritmos estocásticos como o LMS e o RLS, o ruído deverá interferir no transiente do processo. O exemplo 3.8 ilustra essas idéias.

### Exemplo 3.8

Seja um esquema de identificação como o mostrado na Fig. 3.25. Assumamos o modelo usual de sinal transmitido, com amostras binárias e i.i.d., e consideremos que o canal tem a seguinte resposta ao impulso:

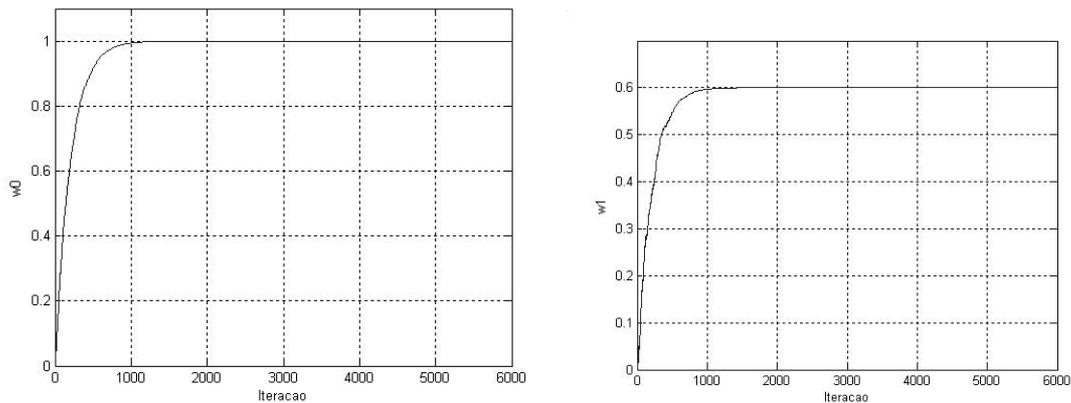
$$H_{\text{canal}}(z) = 1 + 0.6z^{-1} \quad (3.70)$$

Admitamos que o modelo é também um filtro linear e transversal com dois coeficientes, caso em que há a possibilidade de identificação perfeita. Nesse cenário, a função de Wiener será como mostrado na Fig. 3.26.



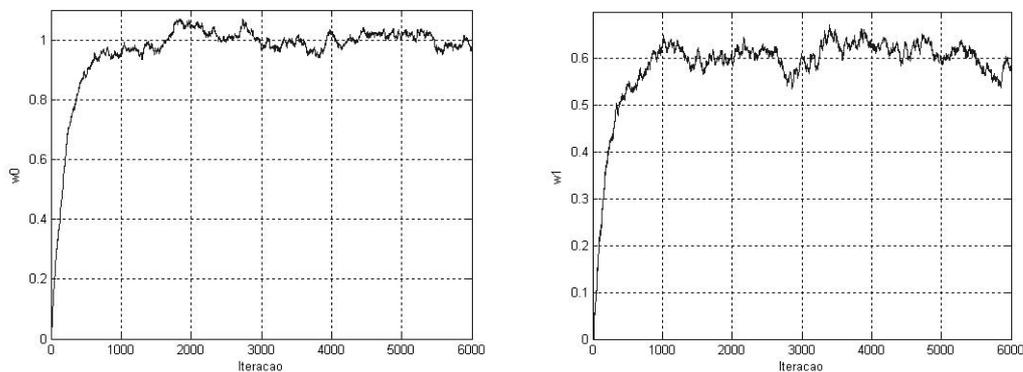
**Figura 3.26:** Função Custo de Wiener e Curvas de Nível

Nota-se que, como era de se esperar, a função é unimodal e o ponto de mínimo corresponde aos coeficientes do canal. Na Fig. 3.27, mostramos a evolução temporal dos coeficientes do modelo para uma realização típica do algoritmo LMS, com  $\mu = 0.001$ .



**Figura 3.27:** Realização do Algoritmo LMS

A influência do ruído na adaptação via LMS é mostrada na Fig. 3.28, que ilustra uma realização sob uma relação sinal-ruído de cerca de 12 dB. É nítida a presença de flutuações estocásticas nocivas, que tornam a estimativa mais imprecisa.



**Figura 3.28:** Realização do Algoritmo LMS – SNR = 12dB

A estrutura da função custo dependerá, como é de praxe, da estrutura escolhida. Se o modelo for linear nos parâmetros, o problema será unimodal. Entretanto, caso o modelo seja não-linear nos parâmetros ou recorrente, haverá a possibilidade concreta de existirem *mínimos locais*. Não nos estenderemos aqui na análise desse problema, pois teremos a oportunidade de considerá-lo mais detidamente quando discutirmos, na seção 10.2, uma contribuição original deste trabalho.

### 3.5.3. Identificação Não-supervisionada e *Cumulant Matching*

Analogamente ao que discutimos no contexto de equalização, dispor de um sinal de excitação comum tanto à planta quanto ao modelo nem sempre é viável. Essa impossibilidade pode ser causada por uma separação física entre ambos ou por algum fator que proíba o acesso à grandeza desejada. No caso de um sistema de comunicação, o problema mais natural é a distância entre os pólos do diálogo, que atrela a disponibilidade de um sinal piloto à existência de um cabeçalho e de sincronismo. Quando isso não for possível, tornar-se-á necessário buscar outras maneiras de formular o problema de identificação.

Assim como no caso da equalização, a base natural para a busca de critérios para identificação cega se encontra na teoria de estatísticas de ordem superior a dois (HOS, *higher-order statistics*). Estatísticas desse tipo são valiosas, pois, contrariamente ao que se dá com a autocorrelação, elas *preservam informação de fase do canal*. São, portanto, candidatas naturais à elaboração de um critério cego. Ilustraremos a essência do *modus operandi* de um identificador não-supervisionado através de uma técnica emblemática: *cumulant matching*. Doravante, suporemos que modelo e planta são *lineares*, sendo tal restrição requerida para que se obtenham resultados estatísticos relevantes.

Em perfeita consonância com a base teórica dos teoremas BGR e SW, a idéia é comparar os cumulantes da saída da planta e do modelo e buscar uma configuração que, dadas as características do sinal de entrada, produza a igualdade entre eles [Tugnait, 1987]. Trata-se de uma modificação direta do esquema apresentado na Fig. 3.25 para que se opere apenas com informação estatística, sem o conhecimento de amostras ou realizações particulares dos sinais. Para que nos aprofundemos do ponto de vista matemático, definamos o cumulante de quarta ordem do sinal recebido:

$$C_{4,x}(\tau_1, \tau_2, \tau_3) = E[x(n)x(n+\tau_1)x(n+\tau_2)x(n+\tau_3)] - C_{2,x}(\tau_1)C_{2,x}(\tau_2 - \tau_3) - C_{2,x}(\tau_2)C_{2,x}(\tau_3 - \tau_1) - C_{2,x}(\tau_3)C_{2,x}(\tau_1 - \tau_2) \quad (3.71)$$

sendo  $C_{2,x}(\cdot)$  a autocorrelação e  $\tau_i$  um atraso discreto. Assumindo que o canal é um filtro FIR e que o ruído é AWGN, é possível mostrar que [Chen et al., 1997]:

$$C_{4,x}(\tau_1, \tau_2, \tau_3) = \gamma_{4,s} \sum_{i=L}^U h_i h_{i+\tau_1} h_{i+\tau_2} h_{i+\tau_3} \quad (3.72)$$

onde  $L = \max\{0, -\tau_1, -\tau_2, -\tau_3\}$ ,  $U = \min\{N, N-\tau_1, N-\tau_2, N-\tau_3\}$  e  $\gamma_{4,s}$  é a curtose do sinal transmitido.

Suponhamos então que, a partir dos dados recebidos, tenhamos sido capazes de estimar os cumulantes de quarta ordem segundo (3.71). De posse dessa informação, podemos, a partir de (3.72), estabelecer um critério que a empregue na busca dos parâmetros do canal. Uma forma simples e natural para relacionar ambas as equações num critério é adotar uma medida de custo de erro quadrático:

$$J_{\text{CUM}}(\mathbf{h}) = [C_{4,x,\text{dados}}(\tau_1, \tau_2, \tau_3) - C_{4,x}(\tau_1, \tau_2, \tau_3)]^2 = \left[ C_{4,x,\text{dados}}(\tau_1, \tau_2, \tau_3) - \gamma_{4,s} \sum_{k=L}^U h_k h_{k+\tau_1} h_{k+\tau_2} h_{k+\tau_3} \right]^2 \quad (3.73)$$

O objetivo do processo de otimização é obter o valor do vetor de parâmetros do canal,  $\mathbf{h}$ , que minimiza a função custo definida. Para tornar o processo mais simples, pode-se efetuar o processo de otimização ao longo de uma “fatia” dos cumulantes definida por  $\tau_1 = \tau_2 = \tau_3$ , o que simplifica o processo sem comprometer a qualidade das estimativas obtidas [Chen et al., 1997].

A função custo obtida, embora conceitualmente simples e coerente, é complexa no que diz respeito à sua relação com os parâmetros livres. Uma consequência dessa inerente complexidade é a existência de múltiplos mínimos associados a diferentes desempenhos, havendo comumente mínimos correspondentes a estimativas bastante pobres. Esse fato abre a perspectiva concreta do uso de ferramentas evolutivas no processo de otimização, o que será por nós explorado na seção 10.1. Por ora, concluiremos a seção com um exemplo em que discutiremos uma aplicação do método de *cumulant matching*.

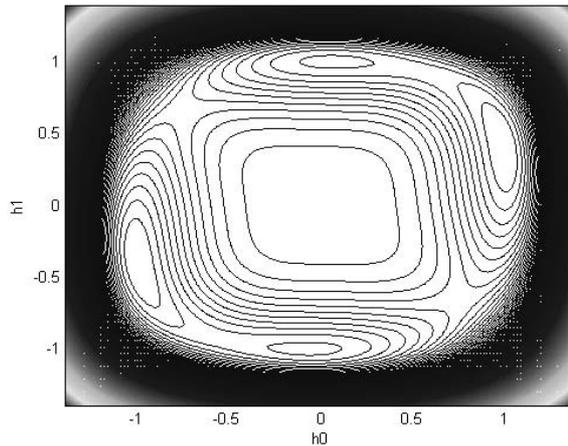
### **Exemplo 3.9**

Suponhamos que o canal que buscamos identificar tenha a seguinte função de transferência:

$$H(z) = 1 + 0.4z^{-1} \quad (3.74)$$

e que o sinal transmitido seja formado por amostras i.i.d.  $+1$  e  $-1$ . Nesse caso, a tarefa de otimização gravitará em torno da obtenção do vetor de parâmetros  $\mathbf{h} = [h_0 \ h_1]$  ótimo.

Na Fig. 3.29, mostramos as curvas de nível da função custo  $J_{\text{CUM}}$  que decorre do modelo adotado.



**Figura 3.29:** Curvas de Nível da Função Custo – Cumulant Matching

São perceptíveis dois pares de mínimos globais, localizados no ponto ótimo  $\mathbf{h} = [1 \ 0.4]$  e em seu simétrico, e dois pares de mínimos locais correspondentes a um modelo que não guarda relação perceptível com a planta a identificar. Isso demonstra que o problema de *cumulant matching* dá origem a um cenário complexo de otimização, existindo a possibilidade concreta de falha caso se use um método de gradiente na busca pela solução ótima.

---

#### 3.5.4. Discussão

Apresentamos nesta seção, de maneira sucinta, as bases do problema de identificação de sistemas. Trata-se de um assunto muito importante em comunicações, seja do ponto de vista investigativo (levantamento de parâmetros de reflexão, estimação da resposta de amplificadores etc.), seja no contexto de uma necessidade prática imediata (e.g. fornecer subsídios para um algoritmo de máxima verossimilhança).

O problema supervisionado se presta a uma abordagem muito elegante e direta baseada no critério de Wiener. Entretanto, se a estrutura de identificação for um filtro não-linear ou recorrente, haverá a possibilidade concreta de que existam mínimos locais na função custo. Esse problema, como já frisamos, merecerá nossa atenção posteriormente.

Quando se aborda o problema linear de maneira não-supervisionada por meio de *cumulant matching*, a obtenção de uma estimativa adequada será possível *desde que se evitem os mínimos locais*. Novamente, a busca por um modelo apropriado pode dar origem a um problema multimodal, cuja solução está condicionada à aplicação de um bom método

de busca. Também teremos a oportunidade de, posteriormente, contribuir para a solução desse problema.

### 3.6. **Predição**

Talvez um dos mais antigos problemas conhecidos pelo homem seja o de estimar o futuro de uma determinada seqüência de informações, ou, em outras palavras, o de *predizer* valores vindouros a partir de amostras passadas. Em termos matemáticos, a tarefa de predição consiste em obter um mapeamento  $F_P(\cdot)$  tal que:

$$y(n) = F_P[x(n-1), x(n-2), \dots, x(n-N)] \cong x(n) \quad (3.75)$$

sendo  $N-1$  um índice de ordem. O filtro caracterizado pelo mapeamento  $F_P(\cdot)$  recebe o nome de *preditor*.

A escolha da estrutura de predição depende do compromisso habitual entre desempenho e custo que vimos discutindo ao longo do capítulo. O fator central a esse dilema jaz nas características da variação temporal do sinal que se deseja predizer. Sinais cuja inter-relação entre amostras é de natureza particularmente complexa requer, usualmente, um preditor não-linear. Tal é o caso, por exemplo, quando se deseja predizer uma variável de estado de um sistema não-linear, ou mesmo de um sistema operando em regime caótico. Se tal caráter complexo comprovadamente não for determinante, a praxe é o emprego de filtros lineares.

Como o problema de predição não depende, propriamente, de supervisão, tendo por base exclusivamente o valor presente do sinal, não há muita diversidade no que diz respeito à sua formulação. Usualmente, mede-se a proximidade entre a estimativa de  $x(n)$  e seu valor real por meio de uma medida de erro quadrático médio, o que, mais uma vez, remete-nos ao arcabouço teórico inerente ao critério de Wiener. A função custo básica de predição é

$$J_{\text{PRED}} = E[x(n) - y(n)]^2 \quad (3.76)$$

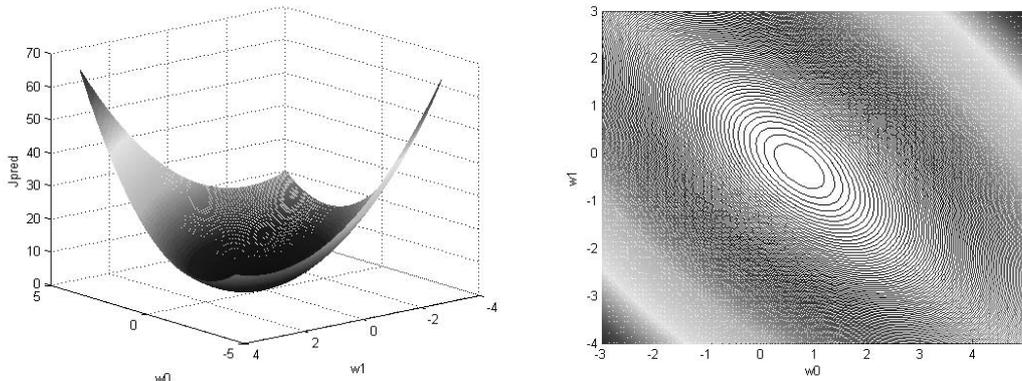
sendo  $x(n)$  o sinal a predizer e  $y(n)$  a saída do preditor. Se a estrutura de predição for linear nos parâmetros, valerá o arcabouço teórico exposto na seção 3.4.2. Se não for esse o caso, será preciso buscar métodos de otimização adequados às características do dispositivo não-linear escolhido. No exemplo 3.10, apresentamos um problema de predição clássico que aborda os pontos expostos.

#### **Exemplo 3.10**

Assumamos que o sinal transmitido é o mesmo com que temos lidado, mas que o canal tem a seguinte função de transferência:

$$H_{\text{canal}}(z) = 1 + 0.6z^{-1} + 0.2z^{-2} \quad (3.77)$$

Tomemos um preditor com dois coeficientes. Na Fig. 3.29, apresentamos a função de Wiener e as correspondentes curvas de nível.



**Figura 3.29:** Função de Wiener e Curvas de Nível - Predição

O comportamento expresso na Fig. 3.29 é exatamente o que se espera de uma função de Wiener num contexto linear. Entretanto, há um ponto curioso: o erro quadrático médio residual é elevado, da ordem de 0.5. Embora isso pareça estranho à primeira vista, não se trata de um fracasso metodológico, e sim, como veremos a seguir, de uma perspectiva muito interessante.

### 3.6.1. Predição e Equalização

Embora não seja patente, há uma estreita relação entre a minimização do *erro de predição* apresentado em (3.76) e o processo de equalização de um canal de comunicação com entradas i.i.d. Para que entendamos por que isso acontece, assumamos que o sinal que desejamos prever,  $x(n)$ , é um sinal recebido que tem a seguinte forma<sup>58</sup>:

$$x(n) = h_0 s(n) + H_{gen}[s(n-1), s(n-2), \dots, s(n-M)] \quad (3.78)$$

onde  $H_{gen}[\cdot]$  é um mapeamento genérico e  $M$  é um índice de memória. Destacamos que essa definição engloba toda a classe de canais lineares. Suponhamos ainda que o preditor tem ordem  $N-1$ , ou seja, que sua relação entrada-saída é

$$y(n) = F_P[x(n-1), x(n-2), \dots, x(n-N+1)] \quad (3.79)$$

Nesse caso, devido a (3.78), teremos que

$$y(n) = \Xi_P[s(n-1), s(n-2), \dots, s(n-N-M+1)] \quad (3.80)$$

onde  $\Xi_P(\cdot)$  é um mapeamento decorrente de  $F_P(\cdot)$ . O erro de predição é:

<sup>58</sup> Note o leitor que tal forma abrange todos os canais lineares e uma classe de canais não-lineares.

$$e_p(n) = x(n) - y(n) = h_0 s(n) + H_{\text{gen}} [s(n-1), s(n-2), \dots, s(n-M)] - \Xi_p [s(n-1), \dots, s(n-N-M+1)] \quad (3.81)$$

Inspecionando com cautela (3.81), percebemos que há um único termo no somatório que não está contemplado no domínio do mapeamento  $\Xi_p[\cdot]$ : justamente a amostra  $s(n)$  do sinal transmitido. Isso mostra que, caso tentemos minimizar o quadrado do erro de predição, *não será possível atuar no sentido de reduzir a influência de  $s(n)$* , ou seja, mesmo se o projeto do preditor for o melhor possível, o erro residual conterá o sinal  $s(n)$ . Portanto, um filtro cuja saída é  $e_p(n)$ , denominado *filtro de erro de predição (FEP)*, pode atuar como um equalizador com atraso de equalização fixo em zero [Cavalcante, 2001]. Isso explica o erro residual alto apresentado no exemplo 3.10.

O fato de o atraso de equalização ser fixo *a priori* é uma limitação bem conhecida da teoria clássica de filtragem [Haykin, 1996]. Em particular, se o preditor for linear, a imutabilidade do atraso acarretará um efeito muito bem conhecido: a impossibilidade de equalizar um canal de fase não-mínima por meio da abordagem de predição. Uma forma de entender a razão de ser dessa restrição está no fato de canais desse tipo poderem produzir, como vimos na seção 3.4.4, configurações não-separáveis linearmente. Essa restrição, no entanto, pode ser contornada por meio do uso de *filtros não-lineares*, como discutido em [Cavalcante, 2001, Ferrari et al., 2003, Ferrari, 2005]. Estruturas desse tipo, devido a sua maior flexibilidade de elaboração do mapeamento, podem minimizar adequadamente o erro de predição mesmo que o problema subjacente não seja solucionável por meio de uma abordagem linear. Destarte, *a abordagem preditiva é uma sólida base sobre a qual se pode erigir uma teoria de equalização não-supervisionada*. Voltaremos a esse problema no capítulo 12.

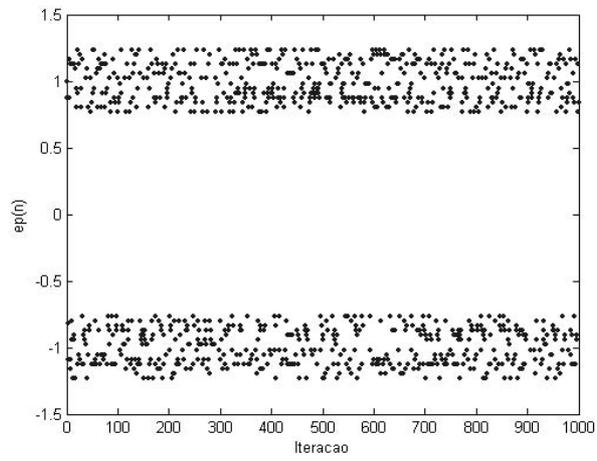
---

### **Exemplo 3.11**

Retomemos o exemplo anterior. Sendo  $\mathbf{w}_{\text{pred}}$  a solução de Wiener correspondente, teremos que o filtro de erro de predição terá a seguinte resposta:

$$e_p(n) = x(n) - \mathbf{w}^T [x(n-1) \ x(n-2)]^T \quad (3.82)$$

Na Fig. 3.30, mostramos a evolução temporal de  $e_p(n)$ . É nítido que o equalizador foi capaz de separar adequadamente os símbolos +1 e -1, sendo que, nesse caso, o filtro de erro de predição opera como um bom equalizador.



**Figura 3.30:** Evolução do Erro de Predição

---

## 4. Estruturas Não-lineares de Filtragem

No capítulo anterior, expusemos os três problemas básicos de que nos ocuparemos neste trabalho, bem como os passos necessários para sua solução. Uma etapa essencial para que as tarefas de equalização, identificação e predição sejam conduzidas de maneira satisfatória é a escolha de uma estrutura de filtragem adequada às dificuldades inerentes à tarefa em mãos e ao nível de desempenho esperado. Se desejarmos uma abordagem permeada pela idéia de otimalidade<sup>59</sup>, precisaremos necessariamente lidar com *estruturas não-lineares de filtragem*.

Para melhor situar a questão do uso desse tipo de estrutura, optamos por dividir os problemas de filtragem em três classes fundamentais<sup>60</sup>:

- *Problemas inerentemente lineares*, ou seja, aqueles cuja solução ótima é linear;
- *Problemas linearmente solúveis*, mas cuja solução ótima *não é linear*;
- *Problemas essencialmente não-lineares*, para os quais o desempenho de uma ferramenta linear é substancialmente inferior ao patamar ótimo.

Não é razoável empregar dispositivos não-lineares em tarefas pertencentes à primeira classe acima, pois basta uma estrutura linear para que se obtenha desempenho ótimo por um custo computacional usualmente menor. Exemplos de problemas desse tipo seriam: identificação supervisionada de uma planta FIR ou equalização supervisionada de um canal trivial (sem IIS e ruído). Já a segunda classe de tarefas abre perspectivas interessantes, *pois há um ganho de desempenho, que pode ser substancial, quando se utiliza um filtro não-linear*. Em casos desse tipo, deve-se ponderar se esse “lucro” compensa a maior complexidade requerida. Um exemplo de problema poderia ser a equalização de um canal com IIS na presença de ruído, pois, mesmo que os estados fossem separáveis linearmente, a solução de Bayes seria, provavelmente, não-linear. No exemplo 4.1 ilustramos esse ponto.

---

### Exemplo 4.1

Consideremos mais uma vez o canal com resposta ao impulso:

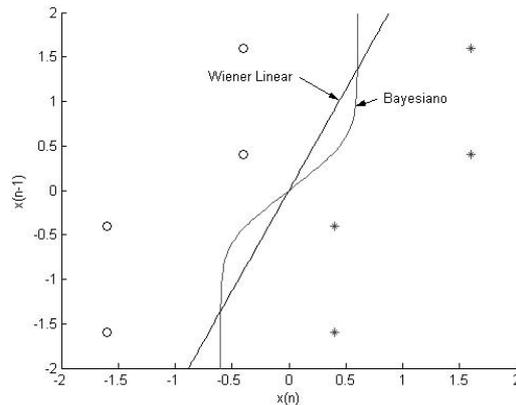
$$H_{\text{canal}}(z) = 1 + 0.6z^{-1} \quad (4.1)$$

e um equalizador com dois coeficientes. Assumamos agora que há ruído com SNR = 15 dB e que o atraso de equalização é nulo. Na Fig. 4.1 apresentamos o problema de classificação resolvido por dois meios: um classificador linear e o classificador bayesiano.

---

<sup>59</sup> Como mostra, por exemplo, a discussão sobre o equalizador de Bayes.

<sup>60</sup> Consideramos aqui apenas os problemas que têm alguma solução.



**Figura 4.1:** Classificadores Linear e de Bayes (SNR = 15dB)

A análise da figura revela uma importante discrepância entre as soluções: ambas cumprem a tarefa, mas é preciso que a curva de separação seja não-linear para que se consiga uma taxa de erro de bit mínima. Há, portanto, uma diferença de desempenho entre dois “especialistas” capazes de abordar o problema.

Por fim, é imperativo o uso de estruturas não-lineares para que problemas da terceira classe sejam resolvidos, pois dispositivos lineares serão, *a priori*, inadequados. Possíveis problemas desse tipo seriam a equalização de um canal linear para um atraso “ruim”<sup>61</sup> (vide o exemplo 3.7), a identificação de um modelo fortemente não-linear, ou a predição de uma série temporal caótica. Discutamos, no exemplo 4.2, a segunda dessas possibilidades.

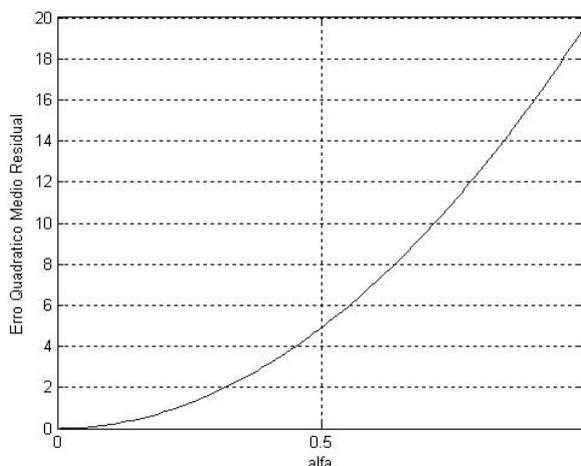
### Exemplo 4.2

Consideremos um canal não-linear com a seguinte resposta entrada-saída:

$$x(n) = [s(n) + 1.5s(n-1)] + \alpha[s(n) + 1.5s(n-1)]^2 \quad (4.2)$$

O caráter não-linear do sistema está associado à constante  $\alpha$ , que pondera um termo polinomial de grau dois. Adotemos como modelo um filtro linear e transversal de dois coeficientes, ou seja, de ordem idêntica à do modelo não-linear. Na Fig. 4.2, apresentamos o valor do erro quadrático médio residual do processo de identificação para diferentes valores de  $\alpha$ . É bastante claro que, mesmo para valores muito pequenos de  $\alpha$ , o modelo linear é bastante insatisfatório. Portanto, um problema desse tipo pode ser considerado, em essência, como um pertencente à terceira classe.

<sup>61</sup> Sem que haja realimentação.



**Figura 4.2:** Relação entre o Erro Quadrático Médio Residual e o Grau de Não-linearidade

Essas considerações que apresentamos têm por fim o estabelecimento de um receituário para a escolha da estrutura de filtragem. Em síntese, a idéia é: se o problema puder ser resolvido de maneira satisfatória por uma estrutura linear, empregue-a. Caso contrário, faça uso de um filtro não-linear apropriado. Na fronteira entre as duas situações, avalie o compromisso entre desempenho e custo computacional<sup>62</sup>. A base para avaliação é sempre formada pelos paradigmas apresentados no capítulo anterior:

- **Equalização:** potencial de inversão (paradigma ZF), patamar de erro quadrático médio (paradigma de Wiener) e taxa de erro de símbolo (paradigma Bayesiano). Ordenamos os três paradigmas em ordem crescente de relevo em sistemas de comunicação digital, permeados pelo espírito de recuperar, com máxima probabilidade, os símbolos enviados.
- **Identificação:** a regra é avaliar o erro quadrático médio num esquema como o mostrado na Fig. 3.25. Eventualmente, pode-se avaliar diretamente se o mapeamento do identificador emula adequadamente o da planta.
- **Predição:** a regra é considerar o erro quadrático médio entre o sinal predito e o sinal real.

Como mostramos há pouco, em equalização, mesmo que o canal seja linear, há usualmente a possibilidade concreta de ganho de desempenho caso se use um filtro não-linear: uma prova disso é o caráter do equalizador de Bayes. Em identificação, a necessidade de um modelo não-linear dependerá da presença desse caráter na planta, e, em predição, da complexidade do sinal que se deseja predizer. Em todo caso, para que o leitor tenha plenas condições de fazer as avaliações requeridas pelo receituário, julgamos conveniente visitar as classes de filtros não-lineares que empregaremos na seqüência deste

<sup>62</sup> Atente sempre o leitor para a *ordem do problema*. Um filtro linear com vinte entradas pode ser mais complexo que uma rede neural com duas entradas.

documento, apontando e discutindo suas principais características. Desejamos que, com isso, fique mais nítida a aplicabilidade desse tipo aos problemas que apresentamos<sup>63</sup>.

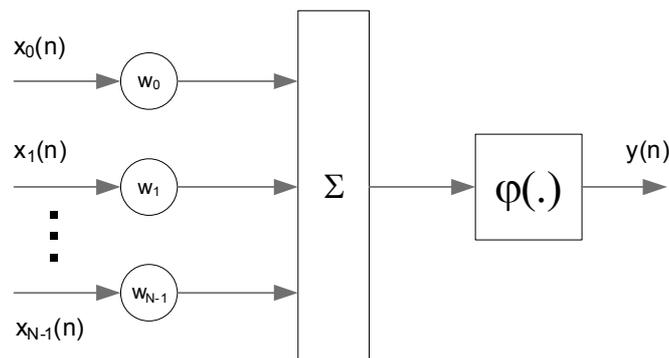
#### 4.1. Redes Neurais

Tendo em mente o enfoque deste trabalho, uma *rede neural (RN)*<sup>64</sup> pode ser definida como um filtro adaptativo formado pela interconexão de diversas unidades de processamento, usualmente não-lineares, num arranjo topológico baseado em camadas. Essa simples definição nos permite explicitar uma série de características relevantes [Haykin, 1994a]:

- Uma RN é um filtro, ou seja, efetua um mapeamento entrada-saída;
- Uma RN tem parâmetros livres que podem ser ajustados de maneira adaptativa;
- Uma RN efetua um processamento distribuído em diferentes unidades;
- Uma RN se baseia fortemente na idéia de conexão;
- Uma RN é, usualmente, não-linear;

Em termos puramente estruturais, há dois aspectos preponderantes na construção de uma RN: *a escolha das unidades básicas de processamento e da forma pela qual elas se interconectam.*

A unidade básica de processamento de uma RN é denominada *neurônio*. Um neurônio pode ser entendido como um dispositivo dotado de conexões e sinapses, ou seja, de entradas e saídas às quais se associam *ponderações* ou *pesos* que determinam seu grau de ativação ou inibição. Biologicamente, muito cedo se constatou que um modelo preciso de neurônio não poderia ser puramente linear, devido à existência de fenômenos relacionados à ocorrência de saturação. Isso levou a uma concepção baseada em duas etapas: *combinação dos estímulos de entrada e passagem por uma não-linearidade sem memória.* A Fig. 4.3 traz um esquema desse modelo.



**Figura 4.3:** Modelo de Neurônio

<sup>63</sup> Optamos por não analisar aqui o *Decision-Feedback Equalizer (DFE)*, devido a seu caráter “quase linear” e específico ao problema de equalização digital. Discuti-lo-emos posteriormente, no contexto específico da contribuição a ele associada.

<sup>64</sup> Como só lidaremos com redes *artificiais*, dispensamos o uso constante desse qualificativo.

Matematicamente, a saída de um neurônio desse tipo é<sup>65</sup>:

$$y(n) = \varphi \left[ \sum_{k=0}^{N-1} w_k x_k(n) \right] \quad (4.3)$$

A função  $\varphi(\cdot)$  recebe o nome de *função de ativação*, sendo decisiva na determinação do potencial de processamento do neurônio. De fato, diferentes escolhas de  $\varphi(\cdot)$  estão no cerne da diferença entre alguns tipos de redes neurais, como veremos logo. Em tese, a escolha da função de ativação deveria ser feita de acordo com a finalidade da rede, sendo, frisamos, decisiva do ponto de vista estrutural. A outra escolha importante é a dos pesos sinápticos, que, em concordância com o espírito da seção 3.3, deverá ser feita de acordo com algum critério sistemático.

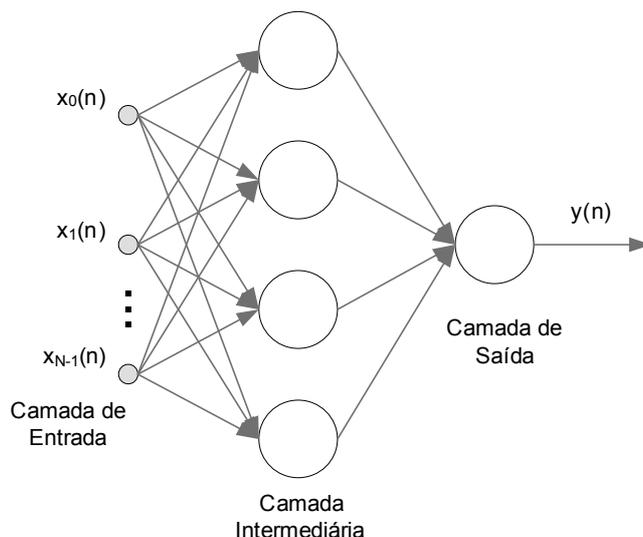
Outro fator estrutural decisivo de uma RN é a forma pela qual seus neurônios se interconectam. A regra é agrupá-los em *camadas* por que a informação trafegará durante o natural percurso entrada-saída, sofrendo as modificações determinadas pelo processamento. Nesse contexto, há três estágios fundamentais:

- Uma *camada de entrada*, formada pelo agrupamento dos estímulos que excitarão os elementos da rede;
- Uma ou mais *camadas intermediárias*, formadas por uma interconexão em cascata de grupos de neurônios;
- Uma *camada de saída*, formada por neurônios que têm por entrada a saída dos neurônios da última camada intermediária e cujas saídas são as variáveis de interesse para o processo de filtragem.

O grau de conexão entre os elementos de cada camada é uma escolha de projeto. Comumente, opta-se por uma completa interconexão entre camadas, em que cada elemento da camada  $C+1$  recebe estímulos provenientes de todas saídas de neurônios da camada  $C$ . Na Fig. 4.4, apresentamos um exemplo de rede desse tipo:

---

<sup>65</sup> Note que o modelo apresentado contempla a possível existência de um termo de polarização (*bias*). Para isso, basta fazer com que uma das entradas seja sempre igual a um.



**Figura 4.4:** Possível Arquitetura de uma Rede Neural

Essas são as considerações fundamentais do ponto de vista estrutural. Do ponto de vista da escolha e da adaptação dos parâmetros, uma RN, como filtro não-linear que é, segue exatamente os passos descritos na seção 3.2.

Neste trabalho, interessam-nos particularmente dois tipos de redes neurais: *redes do tipo perceptron de múltiplas camadas (MLP, multilayer perceptron)* e *redes de funções de base radial (RBF, radial-basis functions)*. Analisemos cada uma delas em separado.

## 4.2. Redes MLP

Podemos traçar a origem da rede MLP ao trabalho pioneiro de McCulloch e Pitts e de Rosenblatt [Haykin, 1994a, Kóvacs, 1997, Príncipe et al., 2000], centrado na noção de neurônios como dispositivos capazes de efetuar tarefas de classificação<sup>66</sup>. A base de seus esforços é o modelo de neurônio apresentado na Fig. 4.3, com uma função de ativação do tipo:

$$\varphi(x) = \begin{cases} +1 & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases} \quad (4.4)$$

Nessas condições, o neurônio pode ser entendido como um discriminante linear, fornecendo um valor de saída igual a +1 quando a combinação linear dos estímulos de entrada for maior que zero e 0 se for menor ou igual a zero.

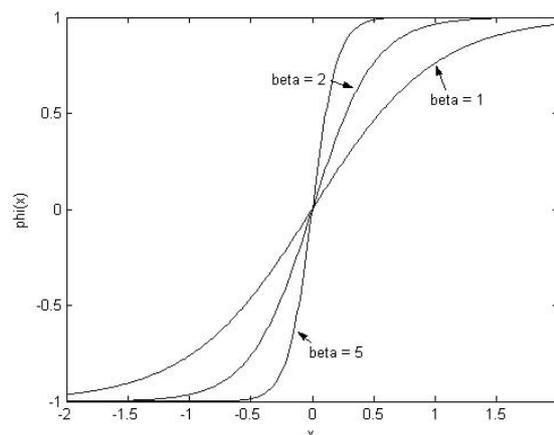
Para construir uma rede MLP, mantém-se a mesma base dos trabalhos pioneiros, a menos de uma modificação essencial: emprega-se uma função de ativação similar à

<sup>66</sup> Um problema que adquiriu notoriedade na aurora do estudo de redes neurais foi o de efetuar o mapeamento lógico do “ou exclusivo (XOR)”.

mostrada em (4.4), mas destituída de seu caráter abrupto<sup>67</sup>. Usualmente, a escolha recai numa função de ativação *sigmoidal* como

$$\varphi(x) = \tanh(\beta x) \quad (4.5)$$

sendo  $\beta$  um parâmetro de controle da transição entre os dois valores assintóticos. Conforme ilustra a Fig. 4.5, o aumento do valor desse parâmetro valoriza o caráter abrupto da não-linearidade.



**Figura 4.5:** Relação entre  $\varphi(x)$  e  $\beta$

De posse desse modelo de neurônio, é simples definir a MLP: *é a rede resultante da interconexão de neurônios com função de ativação sigmoideal numa topologia com pelo menos uma camada intermediária*. Para os fins de nosso trabalho, não será necessário lidar com um modelo tão geral de rede MLP. Faremos, desde já, algumas restrições que teremos como padrão daqui por diante:

- Empregaremos *apenas uma camada intermediária*, formada por neurônios com função de ativação sigmoideal;
- De acordo com o modelo de filtro exposto no capítulo 3, haverá apenas uma saída;
- Os neurônios da camada de saída serão lineares, com função de ativação igual à identidade i.e.  $\varphi(x) = x$ .

Sob essas hipóteses, podemos caracterizar a topologia de uma rede MLP através de duas grandezas: o número de entradas ( $N_{ent}$ ) e o número de neurônios na camada intermediária ( $N_{int}$ )<sup>68</sup>. A relação entrada-saída de nossa rede padrão é:

$$y(n) = \mathbf{w}_{saída} \varphi[\mathbf{W}\mathbf{x}(n) + \mathbf{b}_{int}] + \mathbf{b}_{saída} \quad (4.6)$$

<sup>67</sup> Uma vantagem essencial de usar uma função contínua é a possibilidade de obter, sem maiores problemas, suas derivadas, que, como vimos na seção 3.3.2, podem ser muito importantes para a escolha dos parâmetros do filtro.

<sup>68</sup> Com isso, poderíamos nos referir a uma rede desse tipo como sendo uma MLP( $N_{ent}, N_{int}$ ).

onde  $\mathbf{W}$  é uma matriz cujas linhas são formadas pelos vetores de pesos dos neurônios da camada intermediária,  $\mathbf{b}_{\text{int}}$  é o vetor formado pelas polarizações da camada intermediária,  $\mathbf{w}_{\text{saída}}$  é o vetor de pesos do neurônio linear da camada de saída e  $\mathbf{b}_{\text{saída}}$  é o valor de seu termo de polarização.

A introdução da não-linearidade contida no modelo de neurônio faz da rede MLP uma potente ferramenta para aproximar funções. Conseqüentemente, uma rede desse tipo se afigura uma atraente opção em problemas de equalização, identificação e predição, especialmente quando há necessidade de boa performance num cenário parcamente caracterizado. Em outras palavras, a rede neural se credencia como um adequado dispositivo para operar num esquema do tipo “caixa-preta”, que requer, essencialmente, suficiente flexibilidade e generalidade. É certo que a topologia da MLP que apresentamos nos permite intuir seu potencial de aproximação, mas há uma base mais sólida para a inclusão dessa rede no *hall* de técnicas não-lineares particularmente úteis à solução dos problemas apresentados no capítulo 3: sua capacidade de *aproximação universal*.

Em termos simples, um dispositivo com capacidade de aproximação universal deve ser suficientemente flexível para produzir um grupo bastante representativo de mapeamentos entrada-saída. Dito de outra forma, espera-se que a um filtro com tal capacidade esteja associada uma formulação matemática que defina uma base de funções suficiente para aproximar mapeamentos, digamos, “não-patológicos”. A existência de tal formulação garante, caso as escolhas de estrutura e métodos de otimização tenham sido adequadas, um desempenho satisfatório em virtualmente qualquer tarefa associada à noção de filtragem.

Há alguns anos sabemos que redes MLP são aproximadores universais, graças aos esforços de Cybenko e outros [Haykin, 1994a, Ellacott e Bose, 1996]. Matematicamente, seus trabalhos atestam a validade do seguinte resultado:

***Teorema – Capacidade de Aproximação Universal de uma MLP:*** *Seja uma rede MLP com uma camada intermediária, como a que tomamos por padrão, e  $I_p$  o hipercubo unitário com  $p$  dimensões  $[0,1]^p$ . Para qualquer função contínua definida em  $I_p$ , haverá uma rede neural com um número finito de neurônios que será capaz de aproximá-la em  $I_p$  com um erro máximo  $\varepsilon$ .*

Do ponto de vista de equalização, identificação e predição (vide, novamente, a seção 3.2), esse teorema talvez seja a maior garantia de aplicabilidade de uma MLP: qualquer problema que respeite as condições acima deve poder ser resolvido por uma rede desse tipo, desde que topologia e parâmetros sejam escolhidos apropriadamente.

A escolha do número de neurônios não é uma tarefa trivial, mas depende de algum tipo de informação de que disponha o projetista, a qual é passível de refinamento por um processo de “tentativa e erro” ou por uma estratégia construtiva de aprendizagem<sup>69</sup> [Sweatman, 1995, Von Zuben, 1996]. Na maioria das aplicações em processamento de sinais, opta-se por uma configuração estática, especialmente porque a introdução de mais um processo de ajuste implica aumentar o esforço computacional requerido no período de

---

<sup>69</sup> Uma estratégia desse tipo é capaz de modificar de maneira adaptativa as características topológicas da estrutura.

adaptação. Por outro lado, o uso de uma estratégia construtiva pode significar economia de recursos caso o período de treinamento seja pequeno em relação ao período de teste.

Em suma, o leitor deve estar atento a dois fatos básicos:

- Um número pequeno de neurônios pode levar a uma estrutura com potencial insuficiente de aproximação;
- Um número muito grande de neurônios pode levar a uma *sobreparametrização*, ou seja, uma situação em que a rede opera muito bem para um conjunto-exemplo de amostras, mas tem *excessiva* capacidade de interpolação, o que leva eventualmente a um desempenho pobre para amostras fora do conjunto de treinamento (voltaremos a esse assunto na seção 4.3.3).

#### 4.2.1. Critério de Mínimo Erro Quadrático para uma MLP

Feita a escolha da estrutura, resta a crucial questão de como ajustar seus pesos sinápticos<sup>70</sup>. O processo seguirá, naturalmente, os passos gerais por nós definidos na seção 3.1: escolha de um critério e otimização do mesmo. No caso de uma rede MLP, a base de trabalho é um critério “tipo Wiener”, ou seja, baseado na idéia de *erro quadrático médio*. Como o cálculo das esperanças matemáticas pode ser muito complexo no presente caso, é usual definir uma função custo  $J_{EQM}$  correspondente à média do erro quadrático associado a um conjunto de  $N_T$  padrões de treinamento<sup>71</sup>:

$$J_{EQM} = \frac{1}{N_T} \sum_{k=1}^{N_T} e^2(k) \quad (4.7)$$

Como a relação definida por (4.6) não é linear com respeito aos parâmetros, podemos esperar que a função  $J_{EQM}$  seja bem mais complexa neste caso que a função de Wiener apresentada na seção 3.4.2. De fato, sabe-se que a função de erro quadrático médio associada a uma MLP possui uma estrutura mais rica que o parabolóide típico, sendo dotada de *pontos de sela e mínimos locais* [Haykin, 1994a, Principe et al., 2000]. Um cenário desse tipo enriquece as possibilidades de escolha do método de otimização: passam a ser particularmente interessantes, *mesmo num caso supervisionado*, técnicas de busca global e, eventualmente, heurísticas. Exploraremos essa possibilidade no capítulo 9.

Por ora, convém que cedamos à primazia de quem a possui por direito: métodos baseados nas derivadas da função custo, que, por sua inerente simplicidade, são a regra em otimização de redes neurais. A obtenção do gradiente de  $J_{EQM}$ , no entanto, não é tarefa trivial para uma estrutura não-linear nos parâmetros e, além disso, dotada de múltiplas camadas. A dependência do erro com relação aos pesos da camada intermediária requer uma análise cuidadosa que, aliás, não pode ter lugar imediatamente no desenvolvimento da pesquisa em redes neurais. Para realizá-la, foi preciso introduzir a noção de *propagação do sinal de erro* pelas diversas camadas da estrutura, essência do algoritmo de *back-*

<sup>70</sup> Não nos ocuparemos nesta tese com a possibilidade, considerada em [Von Zuben, 1996], de ajustar das funções de ativação.

<sup>71</sup> A função  $J_w$ , definida na seção 3.4.2, pode ser entendida como um caso limite de  $J_{EQM}$ . Optamos por fazer uma distinção entre ambas nesse ponto para tornar o processo tão claro quanto possível.

*propagation* (BP), o método de gradiente padrão para a MLP [Haykin, 1994a]. Analisemos com mais cuidado esse importante algoritmo.

#### 4.2.2. O Algoritmo de Back-Propagation

É preciso dividir a obtenção do gradiente de  $J_{EQM}$  em duas partes: uma dedicada aos pesos dos neurônios da camada de saída e a outra devotada aos pesos da camada intermediária. Adotaremos a seguinte nomenclatura:  $w_{ji}$  é o  $i$ -ésimo peso do  $j$ -ésimo neurônio da camada intermediária,  $w_{saida,k}$  é o  $k$ -ésimo peso do neurônio linear da saída e  $v_k$  é a saída do  $k$ -ésimo neurônio da camada intermediária.

Para o neurônio da camada de saída, vale a relação:

$$y(n) = \sum_{k=0}^{N_{int}-1} w_{saida,k}(n) v_k(n) + b_{saida}(n) \quad (4.8)$$

Portanto, valerá para seus pesos sinápticos a seguinte regra:

$$w_{saida,k}(n+1) = w_{saida,k}(n) + \mu \frac{1}{N_T} \sum_{k=1}^{N_T} e(k) v_k(k) \quad (4.9)$$

e

$$b_{saida}(n+1) = b_{saida}(n) + \mu \frac{1}{N_T} \sum_{k=1}^{N_T} e(k) \quad (4.10)$$

Analisemos agora o caso mais complicado: a adaptação dos neurônios da camada intermediária. Será útil definir uma grandeza adicional:  $\psi_k$ , o sinal produzido pela parte linear do  $k$ -ésimo neurônio da camada intermediária. Vale, portanto:

$$v_k(n) = \phi[\psi_k(n)] \quad (4.11)$$

Para que cheguemos da função custo aos parâmetros da camada intermediária, será útil recorrer, diversas vezes, à regra da cadeia. Começemos da seguinte forma:

$$\frac{\partial J_{EQM}}{\partial w_{ji}} = \frac{\partial J_{EQM}}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial \psi_j(n)} \frac{\partial \psi_j(n)}{\partial w_{ji}(n)} \quad (4.12)$$

De antemão, sabemos que:

$$\frac{\partial \psi_j(n)}{\partial w_{ji}(n)} = x_i(n) \quad (4.13)$$

Refiramo-nos aos restantes termos de (4.12) como  $\delta_j(n)$ , um *gradiente local* útil a todos os pesos sinápticos do  $j$ -ésimo neurônio:

$$\delta_j(n) = \frac{\partial J_{\text{EQM}}}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial \psi_j(n)} \quad (4.14)$$

Tendo em vista (4.11), explicita-se sem maiores dificuldades o segundo termo do segundo membro de (4.14):

$$\frac{\partial v_j(n)}{\partial \psi_j(n)} = \phi'[\psi_j(n)] \quad (4.15)$$

Por fim, analisemos o primeiro termo de (4.12). Uma análise da topologia mostrada na Fig. 4.4 nos leva a

$$\frac{\partial J_{\text{EQM}}}{\partial v_j(n)} = -\frac{2}{N_T} w_{\text{saida},j}(n) \sum_{k=1}^{N_T} e(k) \quad (4.16)$$

Por conseguinte,

$$w_{ji}(n+1) = w_{ji}(n) + \frac{\mu}{N_T} w_{\text{saida},j}(n) \phi'[\psi_j(n)] x_i(n) \sum_{p=1}^{N_T} e(k) \quad (4.17)$$

Os termos de *bias* seguem a regra acima para  $x_i(n) = 1$ .

Além da escolha do passo de adaptação, há uma outra opção muito relevante para o desempenho do algoritmo: a do número de amostras ( $N_T$ ) do conjunto de treinamento que serão utilizadas para estimar a função custo. Se  $N_T = 1$ , temos uma adaptação do tipo *padrão-a-padrão*, ou seja, (4.17) é visitada a cada nova amostra. Caso o valor de  $N_T$  seja maior que um, podemos considerar que o algoritmo opera num regime de adaptação por *batelada* (*batch*). Ambos os métodos estabelecem um compromisso entre precisão de cálculo do erro *médio* e complexidade computacional.

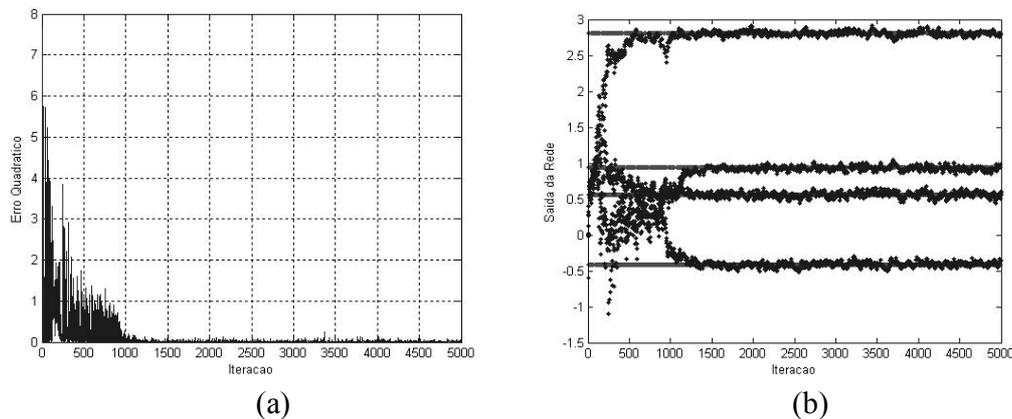
### **Exemplo 4.3**

Neste exemplo, avaliaremos o desempenho de uma rede neural MLP em um problema de identificação não-paramétrica de um modelo que tem uma forma relativamente próxima à de alguns modelos de canais de comunicação via satélite [Benedetto et al., 1979]:

$$x(n) = s(n) + 1.5s(n-1) + 0.3[s(n) + 1.5s(n-1)]^2 - 0.1[s(n) + 1.5s(n-1)]^3 + r(n) \quad (4.18)$$

sendo  $r(n)$  um ruído AWGN (SNR = 23dB) e  $s(n)$  um sinal 2-PAM com amostras i.i.d.

Escolhemos uma MLP bastante “enxuta”, com duas entradas, três neurônios na camada intermediária e  $\beta = 1$ . O esquema de identificação foi o usual, mostrado na seção 3.5.2, e a adaptação da rede foi realizada através de um algoritmo back-propagation operando em regime padrão-a-padrão, com  $\mu = 0.01$  e inicialização aleatória. Na Fig. 4.6(a), mostramos a evolução do erro quadrático de identificação. Após cerca de 1000 iterações, é visível que a rede neural já é capaz de emular bastante satisfatoriamente o modelo não-linear de canal. Podemos atribuir o erro residual à atuação do ruído. Na Fig. 4.6(b), temos uma outra visão do processo: é notável como o padrão de saída da MLP tende a reproduzir com bastante fidelidade as características da saída da parte não-ruidosa do modelo. Concluímos com isso que a rede neural realizou a tarefa com sucesso.



**Figura 4.6:** Evolução do Erro Quadrático e da Saída da MLP

Por fim, ressaltamos duas melhorias importantes no esquema clássico de adaptação: a introdução de um *termo de momento* [Duda et al., 2000, Haykin, 1994a, Von Zuben e de Castro, 2003], capaz de facilitar a convergência em regiões com características de platô, e o uso de métodos de segunda ordem [de Castro, 1998], que podem contribuir imensamente para um melhor desempenho em termos de velocidade de convergência.

### **4.3. Redes de Funções de Base Radial (RBF)**

Uma outra importante classe de estruturas neurais é aquela baseada na idéia de função de base radial (RBF, *radial-basis function*). Para entender o papel de um filtro desse tipo, imaginemos que temos um problema em mãos caracterizado por um conjunto de padrões que desejamos, por exemplo, classificar. Poderia ser interessante conceber algum tipo de “pré-processamento” desses dados que tornasse o problema mais simplesmente solucionável. Tal pré-processamento poderia se dar por meio do revestimento do espaço de entrada por uma base de funções, sobre cujas imagens atuaria um mero combinador linear. Em outras palavras, conceberíamos uma estrutura com um estágio intermediário não-linear de caráter “estático” seguido de um estágio linear em que o problema modificado é, efetivamente, resolvido. Se o conjunto de funções de base for apropriadamente escolhido,

certamente haverá um substancial ganho de desempenho em relação a um dispositivo linear puro [Haykin, 1994a].

Uma solução desse tipo se encaixa perfeitamente na idéia de uma rede neural com múltiplas camadas: há uma camada de entrada, uma camada intermediária com elementos não-lineares e uma camada de saída linear, num esquema idêntico ao mostrado na Fig. 4.4. Os neurônios da camada intermediária, nesse caso, não são elementos lineares seguidos de uma não-linearidade sem memória, mas sim funções de base definidas no domínio dos estímulos de entrada. A relação entrada-saída de uma rede neural desse tipo é:

$$y(n) = \sum_{k=0}^{N_{int}-1} w_k \psi_k [\mathbf{x}(n)] \quad (4.19)$$

sendo o conjunto de funções  $\psi_m(\cdot)$  aquele que forma a base. Há uma importante característica que as funções de base devem possuir para que o epíteto “radial” a elas se aplique:

$$\psi_m(\mathbf{x}) = \varphi_m(\|\mathbf{x} - \mathbf{c}_m\|) \quad (4.20)$$

A definição apresentada em (4.20) introduz um conceito essencial: a idéia de um *centro*  $\mathbf{c}_m$  em torno do qual se organiza a função. Isso significa que a determinação das características da camada intermediária da rede dependerá da escolha de um conjunto desses parâmetros no espaço das entradas. Sem dúvida, podemos intuir que a escolha dos centros será essencial à economia e ao bom desempenho da rede, como, aliás, teremos a oportunidade de verificar posteriormente. Interessantemente, a relação entrada-saída apresentada em (4.19) mostra que as redes RBF também podem ser entendidas como mapeamentos que interpolam pontos formados pelos centros e pelos valores do vetor de parâmetros  $\mathbf{w}$ . Voltaremos logo a essa convergência conceitual.

A escolha mais comum para a base de funções  $\varphi(\cdot)$  é um conjunto de gaussianas com centros  $\mathbf{c}_m$  e variâncias livres  $\sigma_m^2$ :

$$\varphi_m(\|\mathbf{x} - \mathbf{c}_m\|) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}_m\|^2}{2\sigma_m^2}\right) \quad (4.21)$$

Essa opção não é meramente incidental, mas se justifica do ponto de vista do problema de interpolação subjacente [Haykin, 1994a]. Sob tal hipótese, o projeto da camada intermediária de uma rede gaussiana envolve, além da determinação dos centros, a escolha dos valores das variâncias. Temos em mãos, portanto, um problema em múltiplas etapas que nos convém analisar.

#### 4.3.1. Escolha dos Centros - Clusterização

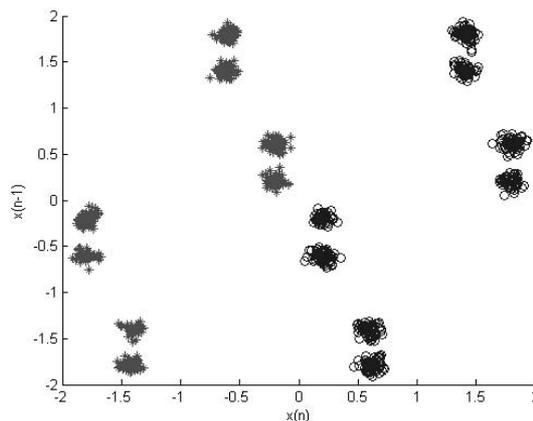
Escolher os centros de uma RBF significa determinar os pontos do espaço de entrada que serão “ocupados” pelas funções de base, ou seja, determinar como o

mapeamento desejado resultará do problema de interpolação delineado. Há basicamente três possibilidades ligadas à realização de tal tarefa [Von Zuben e de Castro, 2003]:

- *Escolher os centros de uma maneira “arbitrária”,* o que pode significar tanto revestir a região mais relevante do espaço com um padrão uniforme de centros quanto com uma configuração aleatória. Em qualquer caso, note o leitor que o processo de escolha é feito *sem que considerem os dados relacionados ao problema.*
- *Escolher os centros conjuntamente com os demais parâmetros da rede,* de uma maneira supervisionada. Nesse caso, utilizar-se-ia um critério de erro quadrático como única referência para o ajuste de todas as grandezas relevantes.
- *Escolher os centros por meio de um processo auto-organizável.*

A primeira opção pode até levar a boas soluções em alguns casos, mas, essencialmente, produz problemas resultantes “artificiais”, até eventualmente sobreparametrizados. A segunda opção é, em tese, efetiva, mas produz um problema bastante não-linear nos parâmetros, o qual, por sua vez, levará a uma função custo de erro quadrático médio *não-convexa*. Por conseguinte, seu emprego só poderia ser considerado seguro caso se dispusesse de informação *a priori* ou de uma poderosa ferramenta de otimização. Por fim, a terceira opção, embora também leve a um problema eventualmente não-convexo, constitui um interessante compromisso entre complexidade e fidelidade ao problema em mãos. Nela, portanto, concentrar-nos-emos.

A idéia da abordagem de *auto-organização* seria buscar, a partir dos dados representados no espaço de entrada, determinadas configurações de centros que definissem *regiões de agrupamento significativo de informação*. Em outras palavras, o processo poderia ser entendido como a busca por centros de determinados grupos ou *clusters* de dados, sendo, por esse motivo, denominado *clusterização*. A Fig. 4.6 mostra um problema em que há dados e regiões com clusters.



**Figura 4.6:** Exemplo de Problema com Clusters

A base mais usual para efetuar tarefas de clusterização é o *aprendizado competitivo*, uma estratégia para ajuste de parâmetros em que diferentes elementos de processamento disputam determinados tipos de padrão, tornando-se tão mais especializados quanto mais

bem sucedidos forem. No presente contexto, podemos conceber a existência de um conjunto inicial de centros que, por meio da competição, torna-se gradualmente mais representativo do conjunto de dados que regula a disputa.

Tal espécie de aprendizado é a essência do chamado *algoritmo K-means*, cujo objetivo é separar um conjunto de  $N_T$  amostras em  $K$  clusters, cada qual identificado por um centro  $\mathbf{c}_i$ . A configuração ótima de centros é, coerentemente com o espírito da tarefa, a que minimiza a seguinte função custo:

$$J_K = \frac{1}{N_T} \sum_{k=1}^{N_T} \|\mathbf{x}_k - \mathbf{c}_k\|^2 \quad (4.22)$$

sendo  $\mathbf{x}_i$  um padrão do conjunto de treinamento e  $\mathbf{c}_i$  o *centro mais próximo a ele*. Portanto, a função  $J_K$  tem por argumento um número arbitrário de centros cuja determinação é a tarefa que o projetista precisará enfrentar. Na Tab. 4.1, apresentamos os passos que compõem o *modus operandi* do algoritmo.

### 1 – Inicialização

- a) Escolha o número de centros ( $N_C$ ), o número de amostras de treinamento ( $N_T$ ) e o passo de adaptação  $\mu$ .
- b) Inicialize os  $N_C$  centros  $\mathbf{c}_i$ .

### 2 - Treinamento

- a) Para  $k = 1$  até  $N_T$ ,
  - a.1) Encontre o centro  $\mathbf{c}_{\text{vencedor}}$  mais próximo do padrão  $\mathbf{x}_k$
  - a.2) Adapte-o segundo a regra:  $\mathbf{c}_i = \mathbf{c}_i + \mu[\mathbf{x}_k - \mathbf{c}_i]$

Fim do laço

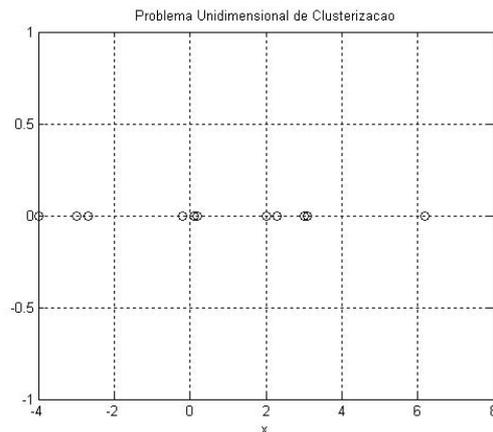
**Tabela 4.1:** Sumário do Algoritmo K-means

Em síntese, o algoritmo K-means pode ser entendido como uma técnica competitiva baseada na aproximação estocástica do gradiente da função custo  $J_K$ . Conforme atesta a literatura [Duda et al., 2000], tal função poderá possuir múltiplos mínimos, o que significa que o algoritmo é vulnerável do ponto de vista de convergência para ótimos locais. O exemplo 4.4 nos ajudará a compreender como isso se dá.

---

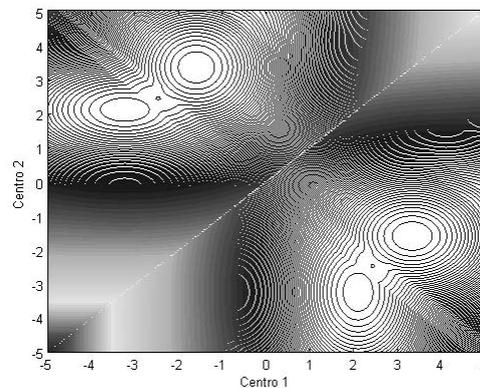
### Exemplo 4.4

Consideremos um problema de clusterização em uma dimensão, em que desejamos agrupar os seguintes padrões em duas classes:



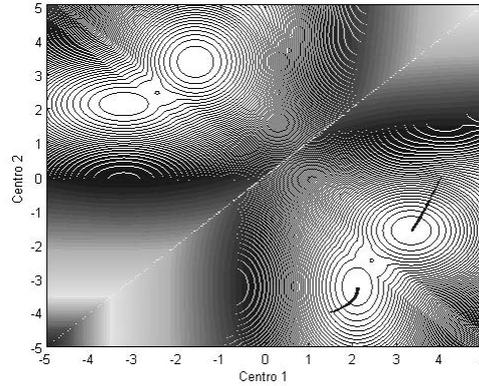
**Figura 4.7:** Padrões a Agrupar

Há dois centros escalares cuja posição precisamos determinar. Para que entendamos a complexidade desse problema, analisemos a função  $J_K$ , cujo perfil de curvas de nível se encontra na Fig. 4.8.



**Figura 4.8:** Curvas de Nível da Função  $J_K$

Observa-se, com nitidez, a existência de dois mínimos, cada qual associado a um custo distinto. Na Fig. 4.9, demonstramos que o algoritmo K-means poderá convergir para qualquer um deles, dependendo da inicialização adotada. Isso ilustra uma importante dificuldade associada ao método.



**Figura 4.9:** Convergência do Algoritmo K-means

Uma alternativa interessante ao algoritmo K-means é o uso de um *mapa auto-organizável de Kohonen*, uma rede neural capaz de, através de um processo de aprendizado competitivo, produzir um mapeamento fiel à topologia dos padrões de entrada [Haykin, 1994a]. No contexto de nosso esforço de clusterização, uma rede de Kohonen pode ser entendida como uma estrutura formada por uma camada de neurônios cujos parâmetros se relacionam aos centros que desejamos encontrar. Para nossos fins, consideramos o processo de aprendizado, em essência, similar ao algoritmo K-means, havendo apenas uma diferença relevante: a possibilidade de introduzir a noção de *vizinhança* no aprendizado, ou seja, de adaptar *não apenas o neurônio vencedor, mas também seus vizinhos*. A intensidade da adaptação será, certamente, menor para os vizinhos, e tanto menor quanto mais distante topologicamente eles estiverem do neurônio vencedor.

Consideraremos aqui apenas uma topologia unidimensional, suficiente para os problemas que nos interessam. O número de neurônios, quando há vizinhança, usualmente é maior que o número de clusters delineados pelos dados de entrada. O aprendizado ocorre através de uma competição entre os neurônios, exatamente como vimos no algoritmo K-means. No entanto, *todos os neurônios* são adaptados ao final da competição segundo a seguinte regra:

$$\mathbf{w}_i(n+1) = \mathbf{w}_i(n) + \mu \Lambda_i(n) [\mathbf{x}(n) - \mathbf{w}_i(n)] \quad (4.22)$$

O fator  $\Lambda_i(n)$  é uma ponderação que controla a intensidade do ajuste efetuado de acordo com a distância entre o  $i$ -ésimo neurônio e o vencedor. Uma possibilidade usual para esse fator é

$$\Lambda_i(n) = \exp\left(\frac{-d_{i,\text{vencedor}}^2}{\xi^2(n)}\right) \quad (4.23)$$

sendo  $d_{i,\text{vencedor}}$  a referida distância. Para vizinhos imediatos, ela vale 1, para os próximos vizinhos, ela vale 2, e assim por diante. O parâmetro  $\xi^2(n)$  é uma espécie de variância, controlando o grau de suavidade da competição: para valores altos, cada competição levará a um incremento substancial nos parâmetros de todos os neurônios, enquanto que, para

valores baixos, o processo tende para uma competição seguida de um único ajuste (como no algoritmo K-means).

Ao final do processo de treinamento, será necessário analisar a configuração geométrica dos neurônios para que, através de seus agrupamentos, sejam obtidos os centros. Um método para realizar tal tarefa é fazer um levantamento da distância entre cada neurônio e seu próximo vizinho topológico. O conjunto dessas distâncias forma a chamada *matriz U* (ou, no caso unidimensional, um *vetor U*). As regiões em que há uma pequena distância entre vizinhos representam agrupamentos, ou seja, indicam a existência de um cluster. O vetor de parâmetros do neurônio com vizinho mais próximo pode então ser tomado como centro desse grupo. Isso conclui o processo.

Poderíamos apresentar várias outras técnicas de clusterização, algumas delas baseadas em melhorias pontuais de algoritmos existentes (e.g. o algoritmo K-means modificado [Chinrungrueng e Sequin, 1995]) ou em idéias distintas de auto-organização [de Castro e Von Zuben, 2000]. Entretanto, julgamos que os subsídios já expostos são suficientes para os fins deste trabalho. Para encerrar a discussão, apresentamos, no exemplo 4.5, uma comparação das técnicas apresentadas num problema de comunicações.

---

#### Exemplo 4.5

Consideremos que um sinal 2-PAM com amostras i.i.d. é transmitido através de um canal com a seguinte função de transferência

$$H_{\text{canal}}(z) = 1 + 0.6z^{-1} \quad (4.24)$$

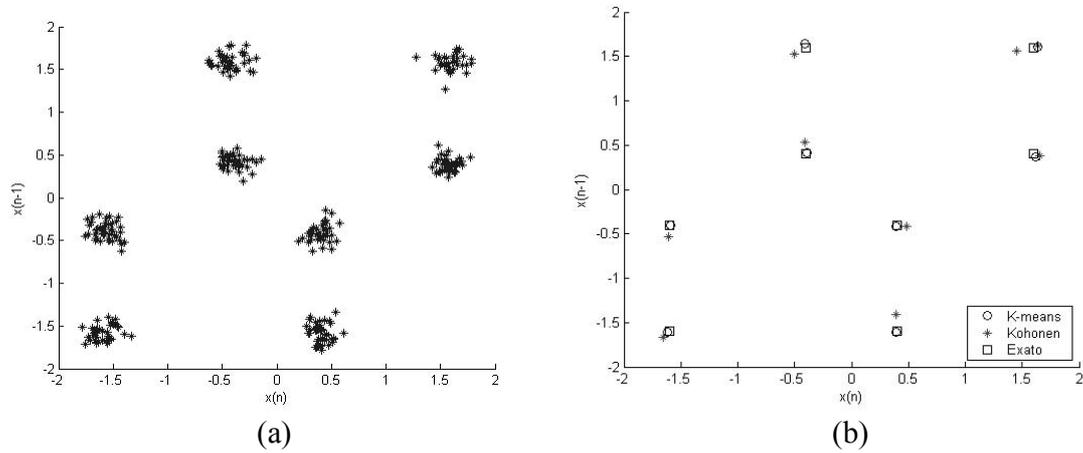
Assumamos ainda que há ruído AWGN com SNR = 21dB. Suponhamos que seja o nosso desejo projetar uma rede neural RBF com duas entradas. Para isso, realizaremos um processo de clusterização por meio de duas técnicas: o algoritmo K-means exposto na Tab. 4.1 e uma rede de Kohonen. Consideremos a existência de 2000 amostras de treinamento, 8 centros para o K-means e 32 neurônios para a rede de Kohonen. Tomemos, ainda, um passo de adaptação variável:

$$\mu = 0.5 \text{ nas } 1500 \text{ primeiras iterações e } \mu = 0.05 \text{ nas demais} \quad (4.25)$$

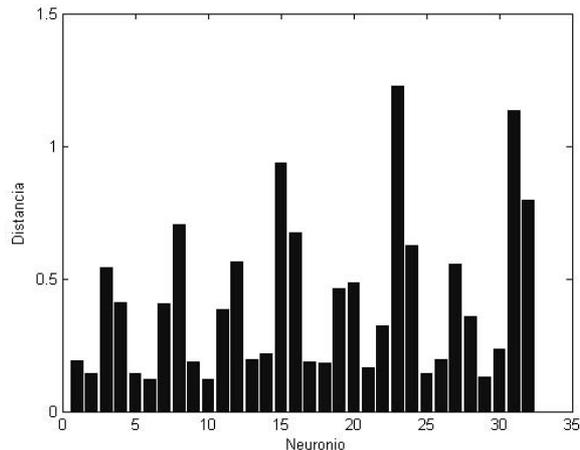
Devido a essa escolha, obtemos uma rápida convergência seguida de um processo de refinamento. Finalmente, tomemos  $\xi = 1$ .

Na Fig. 4.10(a), apresentamos o problema de clusterização decorrente do modelo adotado, enquanto, na Fig. 4.10(b), mostramos os centros obtidos pelas duas técnicas juntamente com o valor ideal. Podemos afirmar que os dois métodos tiveram um bom desempenho, com uma ligeira vantagem para o algoritmo K-means. Consideramos que a razão para isso é a direta associação entre o parâmetro ajustado e o centro que se deseja obter, o que não ocorre na rede com vizinhança, pois, embora sejam obtidos *grupos* de neurônios correspondentes a clusters, a escolha de um deles como ente representativo é uma opção que tem a si associada certa margem de erro. Para ilustrar esse argumento, mostramos na Fig. 4.11 o vetor *U* produzido no processo. Nota-se claramente a presença de 8 vales, correspondentes aos 8 clusters divisíveis na Fig. 4.10(a). Entretanto, será o vetor

de parâmetros do neurônio com menor distância uma estimativa plenamente fiel do centro associado? A qualidade dessa estimativa dependerá de fatores como o próprio número de neurônios e as características da vizinhança escolhida.



**Figura 4.10:** Problema de Clusterização e Soluções Obtidas



**Figura 4.11:** Vetor U

Apesar dessa dificuldade, a rede de Kohonen com vizinhança possui um importante atrativo: seu desempenho não depende tão dramaticamente da escolha do número de entes ajustáveis e do estado inicial. Esses fatores são essenciais no algoritmo K-means, pois, em essência, *cada elemento deve competir por um cluster*. Já no caso da rede de Kohonen, a competição não é necessariamente tão acirrada, o que permite que se obtenham soluções aceitáveis em casos marcados por uma escolha conservadora ou excessiva de neurônios. Um interessante compromisso pode ser refinar a estimativa de Kohonen por meio do algoritmo K-means.

### 4.3.2. Escolha das Variâncias e dos Pesos da Camada de Saída

Feita a escolha dos centros, é necessário que consideremos o problema de determinar as variâncias. Tendo em mente a idéia original de dispor as funções de base, de alguma forma, ao longo do espaço de entrada, chegaremos a uma escolha que não torne a ação da função excessivamente reduzida, mas que, por outro lado, não a torne exagerada, para que não interfira sobremaneira com outros elementos da base. Uma regra útil e comum é ter por ponto de partida a distância entre o centro associado à função que se deseja escolher e o centro vizinho mais próximo. Assumindo que todas as variâncias têm um mesmo valor, e levando em conta o número total de centros, chegamos a uma regra usual [Haykin, 1994a e Von Zuben e de Castro, 2003]:

$$\sigma = \frac{d_{\max}}{\sqrt{2N_c}} \quad (4.26)$$

sendo  $d_{\max}$  a maior distância entre centros e  $N_c$  o número de centros (ou de funções de base). Caso o problema envolva um conjunto de centros com amostras ruidosas a seu redor, pode trazer vantagens levar em conta as características desse ruído na escolha das dispersões. Como veremos logo, esse é o caso ideal em aplicações em equalização [Duda et al., 2000, Mulgrew, 1996].

Supondo que tenhamos projetado, por meio de processos como os que acabamos de descrever, a camada intermediária da rede RBF, restar-nos-á apenas a escolha do vetor de saída  $\mathbf{w}$ . Nesse caso, teremos em mãos um problema de natureza mais familiar, pois a rede RBF, nessas circunstâncias, será *linear nos parâmetros*. Por esse motivo, torna-se concreta a possibilidade de empregar um critério unimodal de erro quadrático médio e fazer uso de todo o arcabouço teórico descrito na seção 3.4.2. Há ainda, como já mencionamos na seção 4.3.1, a possibilidade de unir os projetos da camada de saída e da camada intermediária, o que poderia levar a cenários mais complexos, a menos que haja facilidades de fundir, por exemplo, a informação sobre os centros com as características do sinal de saída desejado num processo de *clusterização supervisionada* [Ferrari, 2005].

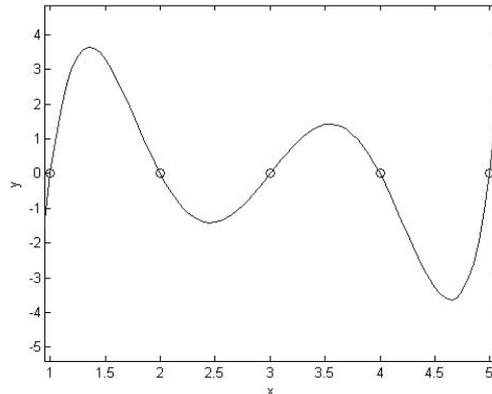
### 4.3.3. Alguns Comentários sobre a Estrutura

Conforme foi discutido anteriormente, um número excessivo de neurônios pode levar a um problema sobreparametrizado. Voltando à idéia de interpolação, compreendemos melhor como isso se dá: para os pontos de referência, o mapeamento será perfeitamente adequado, mas, fora deles, poderá ser excessivamente complexo devido à sobreposição de diversas funções de base. Isso, em problemas “bem-comportados”<sup>72</sup>, não é desejável. A Fig. 4.12 ilustra um caso em que isso se dá, mostrando uma possível

---

<sup>72</sup> A noção de “bom comportamento” se relaciona com a idéia de *regularização*, à qual se associa uma rica teoria [Principe et al., 2000] que, infelizmente, não nos cabe discutir profundamente aqui. No entanto, a base de tudo é impor determinadas restrições ao problema para que a função interpolada não se desvie excessivamente do perfil sugerido pelos pontos disponíveis. Em outras palavras, procura-se inibir a acomodação de excessivos graus de liberdade num perfil demasiadamente complexo.

interpolação de pontos pertencentes à trivial função  $y = 0$ . Vale aqui, como, para alguns, na ética, a regra de ouro do *justo meio-termo*<sup>73</sup>.



**Figura 4.12:** Exemplo de Sobreparametrização

Um outro problema, este típico de redes RBF, é a *maldição da dimensionalidade*. Sua razão de ser jaz no fato de que o aumento da dimensão do espaço de entrada requer um correspondente aumento do número de funções de base que o revestem para que a representatividade do mapeamento se preserve. Esse aumento tem um caráter exponencial, o que pode tornar sobremaneira dispendiosa a manutenção de um grau de cobertura considerado excelente. Surge, dessa forma, um compromisso: número de funções de base disponíveis *versus* dimensão apropriada. Veremos que o problema de equalização ótima *via* redes RBF contém uma interessante manifestação desse aspecto.

Por fim, destacamos que a rede RBF *também tem capacidade de aproximação universal* [Haykin, 1994a], sendo válida para a relação entrada-saída apresentada em (4.19) uma assertiva análoga à que se encontra no cerne do teorema exposto na seção 4.2.

#### 4.3.4. Redes RBF e o Equalizador de Bayes

Apesar de sua capacidade de aproximação universal ser suficiente para credenciar a rede RBF como uma boa estrutura de filtragem em processamento de sinais, há um outro fato que a coloca numa posição muito especial no panteão dos filtros não-lineares: *sua íntima conexão com o equalizador de Bayes*. A similaridade é bastante clara se analisarmos as equações (4.19-4.20) e (3.53), que nos permitem afirmar que haverá equivalência entre ambas as estruturas se e somente se:

- As funções de base escolhidas tiverem um caráter idêntico ao da densidade de probabilidade do ruído aditivo. Isso fornece uma justificativa adicional para o uso de funções gaussianas;
- Os centros dessas funções de base forem idênticos aos estados do canal;

<sup>73</sup> Que poderíamos associar, por exemplo, a idéias de Aristóteles e Buda.

- As dispersões dessas funções forem idênticas às dispersões associadas à densidade do ruído. No caso gaussiano, isso significa que a variância do ruído deve ser idêntica à da função de base;
- Os pesos da camada de saída forem iguais aos rótulos adequados i.e. aos símbolos transmitidos correspondentes.

A primeira etapa não apresenta maiores problemas, pois é aceitável considerar que o projetista conhece as características do canal de comunicação envolvido. A segunda etapa, como adiantamos, exigirá ou o conhecimento do canal, que não é comum, ou um processo de clusterização apropriado. O terceiro passo demandará, em tese, um procedimento de estimação de potência<sup>74</sup>. Por fim, o último passo requer algum tipo de rotulagem supervisionada dos centros obtidos.

O aumento da dimensão do espaço de entrada incrementa exponencialmente o número de estados do canal, podendo mesmo tornar o projeto do equalizador Bayesiano inviável. Isso pode ser uma justificativa para o uso de uma RBF sub-ótima ou mesmo de uma MLP, o que, em muitos casos, pode se dar sem uma dramática perda de desempenho. Mais uma vez, trata-se de uma opção de projeto. Por fim, frisamos que a equivalência entre a rede RBF e o equalizador de Bayes reforça de maneira inequívoca a relevância de estruturas não-lineares em problemas de equalização.

#### **4.4. Filtros Fuzzy (Nebulosos)**

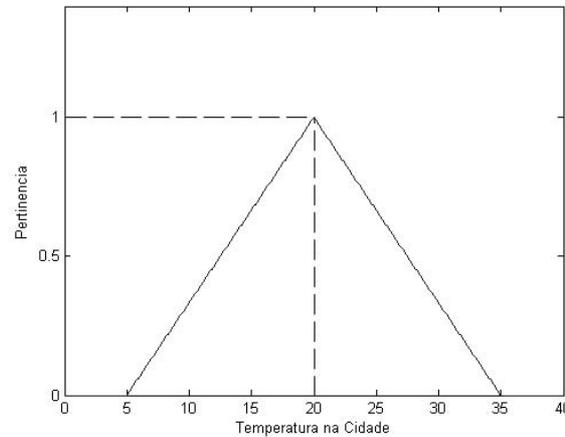
O sistema lógico erigido por Aristóteles em seu *Organon* é uma construção intelectual tão imponente que, por muito tempo, foi considerada o *ne plus ultra* do assunto. Entretanto, tal edifício sempre conteve limitações e dificuldades que, muitas vezes, manifestavam-se na forma de conhecidos paradoxos, alguns dos quais foram analisados e discutidos no colossal trabalho de Bertrand Russell e Alfred North Whitehead<sup>75</sup>. Além disso, ante a riqueza de possibilidades consideradas por todos nós no dia-a-dia, pareceu a muitos que um tratamento binário (verdadeiro ou falso) de proposições seria insuficiente em vários esquemas práticos de decisão. Na vida, muito há que não se encaixa num cenário marcado por transições abruptas: um milímetro a menos torna baixo um homem alto? Um centavo a menos separa o rico do pobre? Um décimo de grau Celsius torna desagradável um ambiente até então aprazível? A questão é: na lógica padrão não há espaço para nuances, para a ambigüidade que reina no cotidiano.

Em 1965, Lofti Zadeh publicou um trabalho seminal [Zadeh, 1965] em que descreveu as bases de um sistema lógico capaz de abrigar e quantificar a *vagueza* de diversos conceitos importantes. A base de tal sistema é a definição de um *conjunto nebuloso* em contraposição a um conjunto tradicional ou *crisp*. Suponhamos, por exemplo, que desejemos estabelecer o conjunto de temperaturas amenas para um habitante de São Paulo. Sentimo-nos seguros ao considerar que uma temperatura de  $-20^{\circ}\text{C}$  está afastada da idéia de amenidade, bem como uma temperatura de  $60^{\circ}\text{C}$ . Mas o que dizer de uma temperatura de  $15^{\circ}\text{C}$ ? E de uma de  $25^{\circ}\text{C}$ ? A própria sensação de dúvida que experimentamos é um indicativo de que tais elementos não se encaixam bem num esquema

<sup>74</sup> Na prática, os resultados da literatura têm mostrado que erros de estimativa nesse caso não são particularmente dramáticos, o que permitiria uma estimação menos pormenorizada [Patra, 1998].

<sup>75</sup> Os célebres *Principia Mathematica*.

rígido. A idéia é, então, definir uma *função de pertinência* capaz de associar a cada elemento um *grau* de pertinência próprio. Uma possibilidade está representada na Fig. 4.13.



**Figura 4.13:** Função de Pertinência Hipotética

Por exemplo, o grau de pertinência de  $T = 30^{\circ}\text{C}$  ao conjunto das temperaturas amenas seria 0.33, enquanto o grau de pertinência de  $T = 50^{\circ}\text{C}$  seria 0, indicando que tal valor está afastado da idéia que define o conjunto.

Formalmente, um *conjunto nebuloso*  $A$  é caracterizado por uma função de pertinência  $P_A(\cdot)$ , que, sem perda de generalidade, normalmente assume valores no intervalo real  $[0,1]$ . Assim, podemos afirmar que um elemento  $a$  não pertence a um conjunto nebuloso se  $P_A(x) = 0$ , mas há diferentes graus de pertinência caso essa igualdade não valha. Quando às operações entre conjuntos, elas abrem múltiplas possibilidades. Escolhas usuais são:

$$P_{\bar{A}}(x) = 1 - P_A(x) \quad (4.27)$$

$$P_{A \cup B}(x) = \max[P_A(x), P_B(x)] \quad (4.28)$$

$$P_{A \cap B}(x) = \min[P_A(x), P_B(x)] \quad (4.29)$$

Em particular, essas definições levam a duas diferenças essenciais em relação ao arcabouço teórico aristotélico:

$$P_{A \cap \bar{A}}(x) \neq 0 \quad (4.30)$$

$$P_{A \cup \bar{A}}(x) \neq 1 \quad (4.31)$$

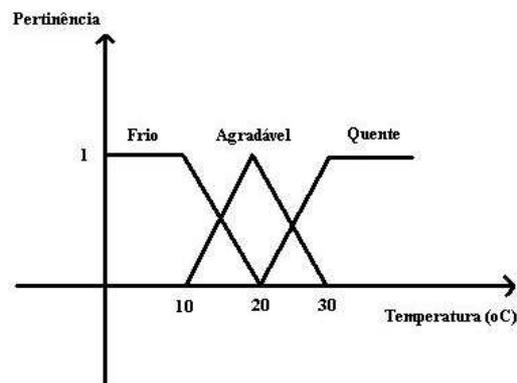
Interessantemente, essas duas possibilidades se relacionam a alguns paradoxos clássicos em que há construções como “A sempre fala a verdade” e “B é um mentiroso” e assim por diante. A chave do problema está na rigidez embutida nas definições tradicionais [Kosko, 1996].

#### 4.4.1. Conjuntos e Regras

Todo processo decisório depende do estabelecimento de regras de atuação. O estabelecimento de regras, por sua vez, envolve *conceitos*. Os conceitos, como discutimos, relacionam-se com definições e conjuntos. Portanto, a natureza desses conjuntos é essencial para a determinação das características gerais do processo decisório.

O código de leis de um país fornece vários exemplos de regras derivadas de uma lógica binária. Por exemplo, pode votar um cidadão que tenha 16 anos no dia do pleito, mas um indivíduo que faça aniversário no dia imediatamente posterior será impedido de efetuar sua escolha. Uma família com renda menor que X salários mínimos poderá ser beneficiada pelo recebimento um determinado auxílio, mas uma outra que ganhe “X + um real” será inexoravelmente privada dele. A tomada de decisão, nesse caso, é muito simples: cada situação estará associada a um único conjunto, e esse conjunto, por sua vez, estará associado também a um único conjunto de saída. Por exemplo: o conjunto dos maiores de 16 anos estará sempre associado ao conjunto dos votantes, sem que haja intersecção com outras possibilidades (por exemplo, que uma parte dos elementos dos menores de 16 possa, eventualmente, votar).

Entretanto, se considerarmos que os conceitos envolvidos na tomada de decisão são definidos por conjuntos nebulosos, haverá um cenário bem diferente. Para que entendamos o porquê, consideremos um problema clássico discutido em [Kosko, 1996]: o controle de um ar condicionado. Em primeiro lugar, precisamos definir os conjuntos associados ao ambiente. Assumamos três conceitos: frio, agradável e quente. As pertinências sugeridas estão na Fig. 4.14.



**Figura 4.14:** Funções de Pertinência (Temperatura)

Consideremos agora as possibilidades para o regime do motor. Também assumiremos três níveis: lento, normal e rápido. As três funções de pertinência estão na Fig. 4.15.

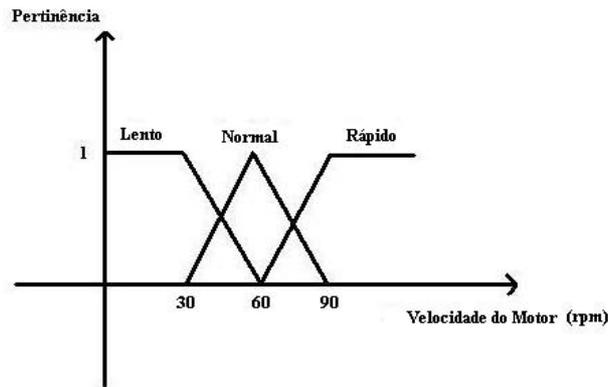


Figura 4.15: Funções de Pertinência (Velocidade)

Agora, tomemos as seguintes regras:

R1: *Se o ambiente estiver frio, então faça o motor operar em regime lento.*

R2: *Se o ambiente estiver agradável, então faça o motor operar em regime normal.*

R3: *Se o ambiente estiver quente, então faça o motor operar em regime rápido.*

Desejamos, a partir dessas regras e dos conjuntos nebulosos, conhecer a velocidade apropriada para cada temperatura. Em outras palavras, trataremos de estabelecer um *mapeamento entrada-saída* coerente com as regras e conceitos estabelecidos pelo usuário. Para tanto, analisemos, na Fig. 4.16, uma representação bidimensional de ambos os lados do processo de inferência: o lado *se* e o lado *então*.

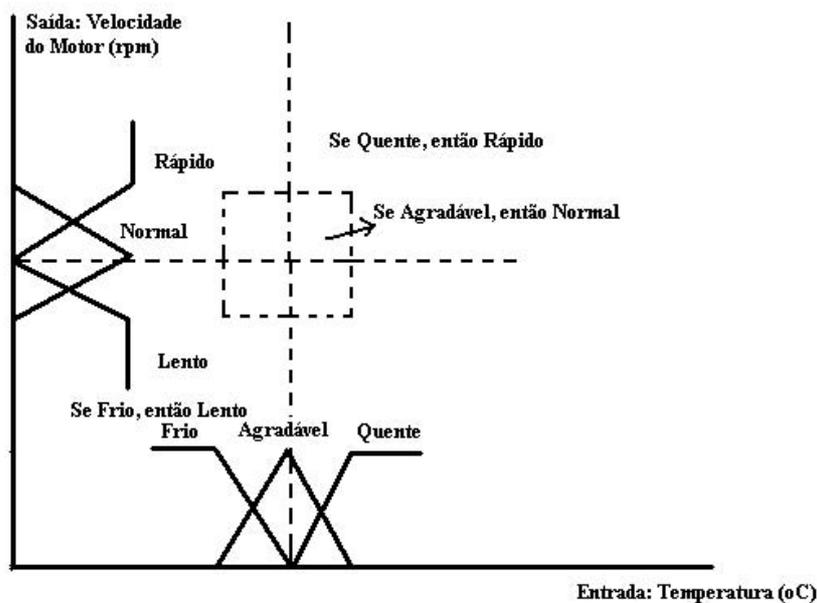
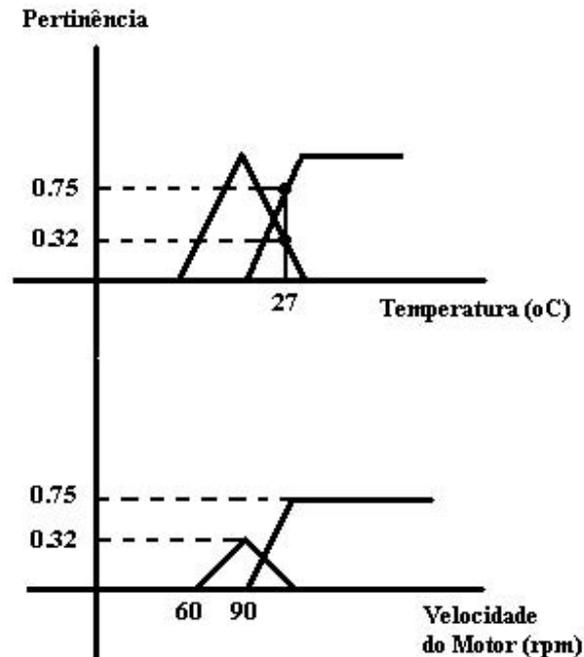


Figura 4.16: Regras Fuzzy e Mapeamento Resultante

O processo tem duas etapas. Em primeiro lugar, dada uma temperatura, é preciso verificar a sua pertinência com respeito a todos os conjuntos. Em seguida, produz-se a saída por meio da sobreposição dos conjuntos relativos à parte *então*, adequadamente ponderados pelas pertinências derivadas dos conjuntos relativos à parte *se*. Ambas as etapas estão representadas na Fig. 4.17.



**Figura 4.17:** Inferência (Parte *Se* e Parte *Então*) para Temperatura de 27°C

Para que transformemos a pertinência da parte *então* num valor de velocidade, é preciso efetuar um processo de *defuzzificação*. Tal processo pode ser entendido como o ato de atribuir um valor real (*crisp*) a uma função de pertinência. Uma possibilidade intuitiva, análoga à noção de *média* em probabilidade, é associar a uma função genérica  $P_{\text{ENTÃO}}(x)$  o valor de seu *centróide* i.e. seu *centro de massa*:

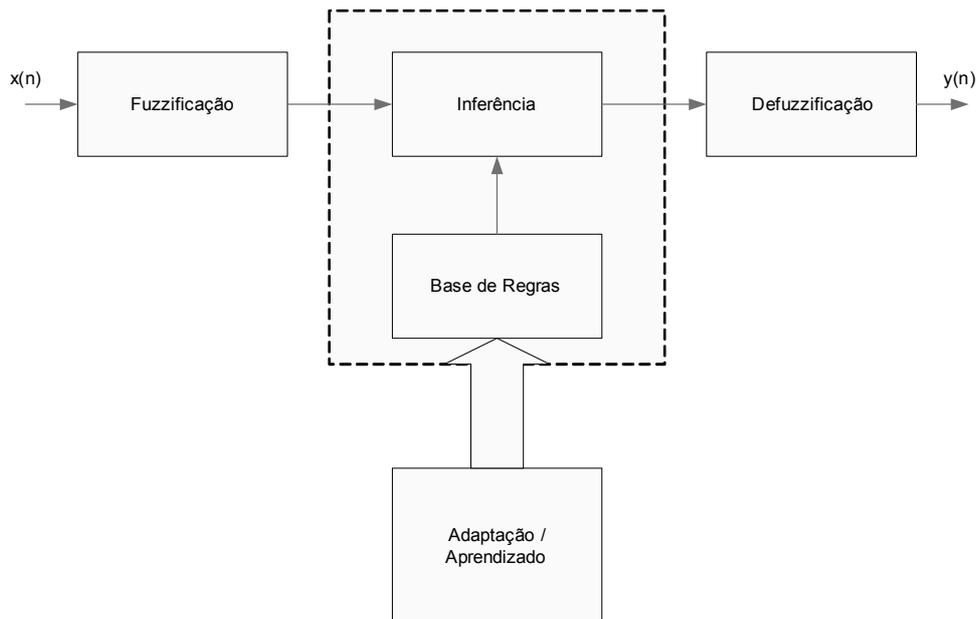
$$\text{Centróide} = \frac{\int_x x P_{\text{ENTÃO}}(x) dx}{\int_x P_{\text{ENTÃO}}(x) dx} \quad (4.32)$$

De posse do centróide, teremos um valor de velocidade, e o processo se encerra.

O receituário que acabamos de delinear indica uma forma interessante de obter um mapeamento entrada-saída (no caso do exemplo, entre temperatura e velocidade): estabelecer regras com antecedentes definidos no espaço de entrada (*se*) e conseqüentes definidos no domínio da saída (*então*)<sup>76</sup>, as quais, por sua vez, processarão valores *crisp*

<sup>76</sup> É importante frisar que há sistemas *fuzzy* em que a parte *então* é formada por conjuntos do tipo *singleton* (passíveis de representação por meio de um simples número real) ou mesmo por funções *crisp* (caso do modelo de Takagi-Sugeno) [Babuska, sem data].

(*não-nebulosos*). Nesse contexto, podemos entender a etapa em que se obtêm as pertinências da parte *se* das regras como um processo de *fuzzificação*, ou seja, de passagem ao “domínio nebuloso”. Lá, o estímulo dispara de maneira variada um conjunto de regras que, por sua vez, leva a um conjunto de saída com função de pertinência formada pela superposição de funções disparadas pela parte *se*. Efetua-se então a defuzzificação desse conjunto, sendo obtido o valor *crisp* de saída. A Fig. 4.18 traz um esquema geral do processo e, conseqüentemente, do dispositivo resultante.



**Figura 4.18:** Filtro *Fuzzy*

Um dispositivo como esse, capaz de processar informação no domínio nebuloso, é um filtro, ao qual doravante nos referiremos como *filtro fuzzy*. Seu projeto depende de duas escolhas fundamentais:

- Das regras, com os respectivos conjuntos *se* e *então*;
- Dos operadores união e intersecção e do processo de defuzzificação;

Não nos aprofundaremos aqui nos meandros da segunda escolha, mas apenas exporemos nossas opções no momento apropriado. Já a primeira escolha é, sem dúvida, deveras relevante do ponto de vista dos problemas de que nos ocupamos, pois definir as regras e os conjuntos significa *determinar a essência do mapeamento* i.e. *a forma pela qual serão revestidos os espaços de entrada e saída*. Consideremo-la, portanto, com atenção.

No exemplo que apresentamos, a escolha das regras foi feita por meio de uma avaliação de especialista, ou seja, de maneira bastante próxima ao denominado *senso comum* acerca do problema<sup>77</sup>. Essa é uma importante característica de um dispositivo

<sup>77</sup> Alertamos que, quando se fala no assunto, é sempre bom recordar Descartes: “Inexiste no mundo coisa mais bem distribuída que o bom senso, visto que cada indivíduo acredita ser tão bem provido dele que mesmo

baseado em lógica *fuzzy*: a possibilidade de introduzir, de maneira intuitiva, informação *a priori* sobre o problema no filtro. Quando houver perspectivas concretas de empregar informação desse tipo, é bastante natural que dela se faça uso. No entanto, nem sempre é esse o caso, pois, em muitas situações, só é possível dispor de um conjunto de dados de treinamento e de um critério de erro quadrático médio, o que nos traz de volta ao nosso problema padrão de filtragem. Nessas condições, o projeto de um filtro *fuzzy* segue, habitualmente, as seguintes etapas, das quais a primeira tem um caráter estrutural e as demais se ligam mais intimamente à fase de ajuste dos parâmetros:

- Definição da forma das regras associadas à parte *se*;
- Definição dos parâmetros dessas regras, incluindo aí “disposição no espaço” e “dispersões”, baseada, se necessário, nos dados de treinamento;
- Definição das regras de saída por meio de treinamento supervisionado<sup>78</sup>;

Assumamos que o vetor de entrada do filtro seja  $\mathbf{x}(n) = [x(n) \ x(n-1) \ \dots \ x(n-N+1)]^T$  e que a mesma parte *se* das regras esteja associada a cada elemento desse vetor. Assumamos ainda que as regras dessa parte sejam gaussianas<sup>79</sup>, que o operador união (ou lógico) seja definido pela soma de funções de pertinência, que o operador intersecção (e lógico) seja definido pelo produto de funções<sup>80</sup> e que a defuzzificação se baseie no cálculo do centróide. Sob tais hipóteses, chegaremos à seguinte relação entrada-saída [Kosko, 1996, Patra, 1998]:

$$y(n) = \frac{\sum_{k=1}^{N_F} w_k \prod_{p=0}^{N-1} \exp\left\{-\frac{[x(n-p) - c_{k,p}]^2}{2\sigma^2}\right\}}{\sum_{k=1}^{N_F} \prod_{p=0}^{N-1} \exp\left\{-\frac{[x(n-p) - c_{k,p}]^2}{2\sigma^2}\right\}} \quad (4.33)$$

sendo  $N-1$  a ordem do filtro,  $N_F$  o número de “funções de base *fuzzy*” formadas pelas regras (e ponderadas por  $w_k$ ) e  $c_{k,p}$  um centro *escalar* genérico. Salientemos dois pontos:

- O processo de defuzzificação por centróide torna o mapeamento independente da *forma* dos conjuntos associados à parte *então*, havendo apenas dependência em relação a seu centróide  $w_i$ .
- O denominador, que pode ser entendido como um *termo de normalização*, origina-se da definição de centróide apresentada em (4.32).

---

os mais difíceis de satisfazer em qualquer outro aspecto não costumam desejar possuí-lo mais do que já possuem”, in *Discurso do Método*, Coleção “Os Pensadores”, Nova Fronteira, 1999.

<sup>78</sup> Como no caso da rede RBF, pode haver situações em que exista uma fusão de passos numa única etapa.

<sup>79</sup> Aqui já começamos a nos concentrar nos problemas centrais de nossa pesquisa.

<sup>80</sup> Note o leitor que tal escolha contraria os exemplos expostos em (4.28) e (4.29), os quais, para nós, tiveram apenas valor didático. A razão de ser dessa escolha é que, em termos simples, os operadores adotados se aplicam melhor ao ambiente de processamento de sinais reais.

#### 4.4.2. Equivalência entre o Filtro Fuzzy, a Rede RBF e o Equalizador de Bayes

A relação entrada-saída mostrada em (4.33) é bastante similar àquela típica de uma rede RBF (vide (4.19)). A diferença fundamental entre as expressões é a presença do termo de normalização das funções de base. Tal diferença, no entanto, não é tão preponderante em problemas de equalização, nos quais a saída do mapeamento sempre é processada pelo dispositivo de decisão. Assim, podemos, para todos efeitos, considerar que há uma espécie de equivalência entre o filtro *fuzzy* gaussiano e uma rede RBF com funções de base de mesmo tipo<sup>81</sup>.

A primeira consequência desse fato é que podemos afirmar que há uma equivalência entre o filtro *fuzzy* que acabamos de expor e o equalizador de Bayes. De fato, podemos afirmar que ambos devem produzir uma mesma curva de separação de estados<sup>82</sup>.

Uma outra consequência é que a análise de resultados e propostas relativas a uma das técnicas também deve frutificar em contribuições à outra, o que estabelece uma útil simbiose entre campos. Em particular, tiraremos proveito desse fato de duas maneiras: consideraremos que os resultados expostos no capítulo 12 serão, de certa forma, também uma contribuição no campo das redes neurais, e optaremos por não incluir aqui uma seção específica sobre treinamento de filtros *fuzzy*, que, em linhas gerais, segue o receituário exposto na seção 4.3.

### 4.5. Filtros Polinomiais ou de Volterra

Iniciamos este capítulo pela apresentação de estruturas não-lineares derivadas do campo da inteligência computacional, ou seja, decorrentes, em maior ou menor escala, da busca por modelos capazes de emular mecanismos cerebrais ou esquemas de tomada de decisão. A estrutura que agora estudaremos tem uma história um tanto diferente. Sua origem é o trabalho de um matemático italiano que muito investigou sistemas dinâmicos: Vito Volterra [Doyle et al., 2002]. Podemos afirmar que a contribuição essencial desse matemático ao processamento de sinais foi seu trabalho com aproximação de funções (1887), o qual levou a uma série que pode ser entendida como uma extensão da *série de Taylor* a mapeamentos dotados de memória [Mathews e Sicuranza, 2000]: a série de Volterra<sup>83</sup>.

Assumamos que  $H[.]$  represente a relação entrada-saída de um sistema não-linear *sem memória*, de tal forma que:

$$y(n) = H[x(n)] \quad (4.34)$$

---

<sup>81</sup> Uma ressalva importante apontada por Patra [Patra, 1998] é que a implementação “tipo *fuzzy*” pode ser mais eficiente do ponto de vista computacional.

<sup>82</sup> A equivalência entre o filtro de Bayes e o filtro nebuloso gaussiano seria ainda mais geral se não houvessemos dispensado o termo  $p[x(n)]$  na passagem da equação (3.48) à equação (3.50).

<sup>83</sup> Também associada a Wiener e Frechet [Mathews e Sicuranza, 2000].

Desde que  $H[.]$  obedeça a algumas restrições, é possível mostrar que sua atuação pode ser perfeitamente expressa através de uma base de polinômios, ou seja, de uma expansão de Taylor<sup>84</sup>[Campello, 2002]:

$$H(x) = H(0) + \sum_{k=1}^{\infty} \frac{1}{k!} \left( \frac{d^k H}{dx^k} \right)_{x=0} x^k \quad (4.35)$$

Consideremos agora que o mapeamento não-linear seja dotado de *memória*. Nesse caso, um raciocínio análogo levaria a uma expansão mais complexa, denominada *série de Volterra* [Kajikawa, 2000, Mathews, 1991]:

$$y(n) = h_0 + \sum_{m_1=-\infty}^{\infty} h_1(m_1)x(n-m_1) + \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} h_2(m_1, m_2)x(n-m_1)x(n-m_2) + \dots \quad (4.36)$$

onde as funções multidimensionais  $h_p(m_1, \dots, m_p)$  são denominadas *kernels de Volterra*. Como a expansão descrita em (4.36) é, *a priori*, dotada de todas as possíveis combinações polinomiais, podemos afirmar que *o conjunto de kernels de Volterra caracteriza perfeitamente o mapeamento  $H[.]$* .

A convergência da série de Volterra é um assunto complexo que não podemos abordar pormenorizadamente aqui. Em termos intuitivos, podemos afirmar as principais dificuldades decorrem da presença de *não-linearidades abruptas* em  $F[.]$ , tais como descontinuidades e transições nas quais não é possível definir suas derivadas [Mathews e Sicuranza, 2000, Campello, 2002]. Portanto, um melhor desempenho deverá ser obtido se  $F[.]$  for “bem-comportado” do ponto de vista da *suavidade*<sup>85</sup>.

A existência de uma classe que podemos considerar ampla do ponto de vista das aplicações que aqui nos interessam permite-nos atribuir à expansão descrita em (4.36) a propriedade de aproximação universal, o que faz dos termos polinomiais uma *base atraente para filtragem*. Podemos definir, portanto, um filtro de Volterra como um dispositivo cuja relação entrada-saída tem a forma geral expressa em (4.36).

Embora seja uma ferramenta muito eficiente, um filtro desse tipo é irrealizável na prática<sup>86</sup>, pois há infinitas grandezas envolvidas. Por esse motivo, costuma-se *truncar* a série em algum ponto, o que dá origem a um dispositivo tratável e passível de implementação [Mathews, 1991]. Voltando aos passos discutidos na seção 3.3, podemos incluir o processo de truncamento na primeira etapa, ou seja, na etapa de escolha da estrutura, com todas suas incertezas características.

A escolha da ordem do modelo de Volterra deve levar em conta o caráter “explosivo” do crescimento do número de possíveis *kernels* com a introdução de polinômios de maior grau e com o aumento do número de entradas. Além das considerações sistêmicas habituais, algumas possibilidades podem nortear a escolha:

<sup>84</sup> No caso, como adotamos uma expansão em torno de zero, temos um caso particular: a *série de Maclaurin* [Boyer, 1991].

<sup>85</sup> Uma discussão mais aprofundada sobre o assunto pode ser encontrada em [Mathews e Sicuranza, 2000].

<sup>86</sup> Note o leitor que, por ora, não consideramos uma estrutura recursiva.

- Explorar a simetria dos *kernels*, ou seja, projetar de maneira unificada todos os parâmetros que multiplicam termos idênticos<sup>87</sup>;
- Caso seja possível contar com algum conhecimento prévio, eventualmente pode ser possível dispensar o uso de alguns termos não-lineares do modelo completo, o que torna o sistema resultante mais simples e tratável;

Devido ao aumento significativo de complexidade com a ordem do modelo, é bastante usual o emprego de modelos *quadráticos* ou *cúbicos*. A primeira espécie de modelo tem a seguinte relação entrada-saída:

$$y(n) = h_0 + \sum_{m_1=-\infty}^{\infty} h_1(m_1)x(n - m_1) + \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} h_2(m_1, m_2)x(n - m_1)x(n - m_2) \quad (4.37)$$

Já os filtros cúbicos, por sua vez, obedecerão à seguinte relação geral:

$$y(n) = h_0 + \sum_{m_1=-\infty}^{\infty} h_1(m_1)x(n - m_1) + \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} h_2(m_1, m_2)x(n - m_1)x(n - m_2) + \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} \sum_{m_3=-\infty}^{\infty} h_3(m_1, m_2, m_3)x(n - m_1)x(n - m_2)x(n - m_3) \quad (4.38)$$

A escolha do modelo de Volterra a ser adotado deve ser feita com muito cuidado, tendo sempre em vista as particularidades do problema a resolver. O fato de um filtro polinomial, ao contrário de uma rede neural, ser formado por unidades de processamento com linearidades bastante distintas cria algumas “armadilhas” relacionadas a circunstâncias específicas. O exemplo 4.6 mostra um caso em que um filtro quadrático não é capaz de ter um comportamento superior ao de um filtro linear com o mesmo número de entradas.

### **Exemplo 4.6**

Consideremos a equalização de um canal linear genérico excitado por um sinal de entrada  $s(n)$  2-PAM com amostras i.i.d.:

$$x(n) = \sum_{k=0}^{N_C-1} h(k)s(n - k) \quad (4.39)$$

onde  $N_C-1$  é a ordem do canal. No projeto de um equalizador quadrático ótimo, seria necessário calcular a correlação cruzada entre os termos polinomiais e o sinal desejado. Por exemplo, supondo que o atraso de equalização fosse nulo, precisaríamos calcular:

$$E[x^2(n)s(n)] \quad (4.40)$$

<sup>87</sup> Por exemplo,  $h_2(0,1)$  e  $h_2(1,0)$ .

ou

$$E[x(n)x(n-1)s(n)] \quad (4.41)$$

No entanto, conforme ilustrado na seção 3.4.4, há uma simetria inerente a um canal linear FIR: o valor de  $x(n)$  para um conjunto  $s(n), s(n-1), \dots, s(n-N_C+1)$  será o oposto do valor para  $-s(n), -s(n-1), \dots, -s(n-N_C+1)$ . Essa simetria fará com que termos como os mostrados em (4.40) e (4.41) *se anulem*, o que, por sua vez, *reduzirá o filtro de Volterra a um mero filtro linear*. Em outras palavras, uma não-linearidade quadrática não é apropriada a um problema de equalização com essa característica.

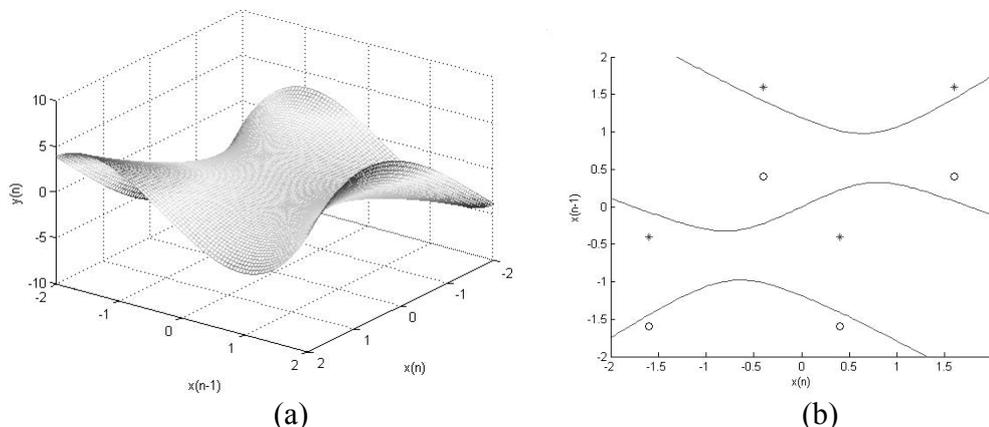
---

Por fim, uma análise de (4.36) mostra uma propriedade fundamental do modelo de Volterra: ele é *linear nos parâmetros*. Isso significa que a escolha dos *kernels* pode se dar dentro do arcabouço conceitual inerente ao paradigma de Wiener. Significa também que os algoritmos LMS e RLS podem ser utilizados exatamente como na descrição presente na seção 3.4.2. O exemplo 4.7 ilustra a aplicação do paradigma de Wiener a um problema de equalização.

---

#### Exemplo 4.7

Consideremos uma vez mais o canal apresentado em (4.24) e um equalizador com dois coeficientes. Tomemos por base um atraso de duas amostras, que, como vimos no exemplo 3.7, produz um problema de impossível solução linear. Como discutido no exemplo anterior, um filtro quadrático não é uma boa escolha para esse problema de equalização, tendo recaído nossa escolha sobre um filtro cúbico sem parte quadrática. Na Fig. 4.19(a), apresentamos o mapeamento entrada-saída do filtro de Volterra ótimo do ponto de vista do paradigma de Wiener e, na Fig. 4.19(b), apresentamos o ponto de vista de classificação de padrões. A figura revela que o filtro é capaz de equalizar o canal adequadamente, sendo, destarte, uma estrutura dotada de um adequado grau de flexibilidade.



**Figura 4.19:** Mapeamento e Curva de Separação do Equalizador Cúbico de Wiener

### 4.5.1. Filtros de Volterra e Filtros Lineares

Não nos causa espanto o pioneirismo dos filtros de Volterra em aplicações ligadas à comunicação digital [Benedetto e Biglieri, 1983]: uma não-linearidade desse tipo pode, em analogia com o conceito de polinômio, ser considerada uma *extensão direta de um filtro linear*<sup>88</sup>. Em primeiro lugar, um filtro linear e transversal pode ser entendido como um filtro de Volterra truncado. Além disso, a relação entrada-saída de um filtro de Volterra pode ser formulada em termos de *convoluções*. Para que entendamos esse ponto, consideremos o seguinte termo da expansão:

$$\sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} h_2(m_1, m_2)x(n - m_1)x(n - m_2) \quad (4.42)$$

Construamos o seguinte sinal de entrada:

$$v(n_1, n_2) = x(n_1)x(n_2) \quad (4.43)$$

Definamos agora a seguinte convolução multidimensional:

$$\zeta(n_1, n_2) = \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} h_2(m_1, m_2)v(n_1 - m_1, n_2 - m_2) \quad (4.44)$$

Comparando (4.44) e (4.42), observamos que a convolução multidimensional apresentada em (4.44) é idêntica ao termo de Volterra caso  $n_1 = n_2 = n$ . De posse dessa propriedade, é possível estabelecer uma análise em frequência para filtros de Volterra [Doyle et al., 2002, Bendat, 1998] baseada em transformadas multidimensionais<sup>89</sup>. Essa relação íntima com o arcabouço teórico ligado a filtros lineares torna possível o emprego de cumulantes na identificação de modelos de Volterra [Mathews e Sicuranza, 2000]<sup>90</sup> e o estabelecimento de abordagens para equalização cega fora do contexto SISO [Giannakis e Sarpedin, 1997].

## 4.6. Estruturas Recursivas

Definidos os aspectos básicos das estruturas de filtragem apresentadas, não é problema definir suas versões recursivas: basta que haja na camada de entrada versões atrasadas do sinal de saída da estrutura<sup>91</sup>. Como no caso do filtro IIR<sup>92</sup>, tal processo de realimentação engendra uma dinâmica subjacente à evolução temporal do estado do

<sup>88</sup> Da mesma maneira como um polinômio de grau dois ou três se relaciona com um polinômio de grau um.

<sup>89</sup> Muito comuns, por sinal, em processamento de imagens.

<sup>90</sup> Conforme já adiantáramos na seção 3.5.3.

<sup>91</sup> No caso de uma estrutura com múltiplas camadas, pode também haver realimentação no âmbito das camadas intermediárias.

<sup>92</sup> As considerações sobre estabilidade do filtro IIR (seção 3.2.1) são fruto de uma análise dinâmica.

dispositivo, que, no caso não-linear, pode ser significativamente mais rica [Haykin, 1994a]. Essa riqueza de comportamentos tem duas faces: é muito útil quando se lida, por exemplo, com a identificação de sistemas reais complexos, mas, por outro lado, significa que pode haver um grande número de comportamentos indesejados passíveis de ocorrência no caso de uma escolha mal-sucedida de parâmetros. Além disso, a natureza da relação matemática entrada-saída associada a um dispositivo desse tipo se torna bastante intrincada. Permanece válido, portanto, o dilema que já discutíramos na seção 3.2: ganho de desempenho<sup>93</sup> *versus* complexidade.

Em tal cenário, mesmo que nos atenhamos a um critério supervisionado, surgirão alguns problemas no processo de escolha dos parâmetros:

- A introdução de um laço de realimentação pode tornar a função custo de Wiener multimodal mesmo para filtros lineares [Regalia, 1995].
- A obtenção das derivadas da função custo, essenciais nos mais comuns métodos de otimização empregados, torna-se uma tarefa proibitiva em muitos casos. O recurso a aproximações é a alternativa usual a esse problema [Regalia, 1995].
- Caso se use um algoritmo clássico baseado em derivadas, haverá sempre o risco de que o vetor de parâmetros assuma uma configuração instável, comprometendo definitivamente o processo.
- A própria expressão iterativa do algoritmo adaptativo será um sistema dinâmico altamente complexo, o que pode acarretar problemas de convergência.

Adiantamos que esses quatro itens indicam uma direção que exploraremos em capítulos posteriores: a possibilidade de empregar métodos de otimização mais eficientes que aqueles baseados nas derivadas da função custo [Attux et al., 2004a].

O emprego de filtros não-lineares recursivos tem se mostrado útil nos três problemas de filtragem que nos interessam primordialmente:

- Em problemas de equalização, o uso de estruturas desse tipo tem se mostrado bastante efetivo caso se resolva apropriadamente a questão da escolha dos parâmetros [Siu et al., 1989, Park e Jeong, 1999], merecendo especial atenção o uso de arranjos do tipo DFE [Chen et al., 1992, Mulgrew, 1996], ou seja, dotados de um laço de *realimentação de decisões* [Proakis, 1995]. A natureza desse ganho de desempenho pode ser entendida à da discussão por nós empreendida acerca do problema de sobreposição de estados (seção 3.4.4). Depreende-se disso que filtros recorrentes deverão ter uma relação desempenho-complexidade bastante atraente em problemas complexos de equalização, tomando por base a comparação com um filtro *feed-forward* de Bayes [Duarte, 2004].
- Em problemas de identificação, modelos recursivos podem ser bastante úteis por dois motivos: são capazes de produzir modelos não-paramétricos mais parcimoniosos, ou seja, dotados de poucos parâmetros em relação a um modelo *feed-forward* equivalente, e também podem ser essenciais quando a planta que se deseja modelar tem uma dinâmica não-linear complexa. A segunda razão, em particular, explica o interesse despertado por redes neurais recorrentes quando se

---

<sup>93</sup> Vale frisar que, em muitos casos, a presença de um laço de realimentação pode ser decisiva, como vimos na seção 3.4.4.

trata de analisar processos complexos de memória ou cognição [Von Zuben e de Castro, 2003].

- Problemas de predição sempre constituíram um importante nicho para a aplicação de estruturas recorrentes [Duda et al., 2001]. A razão desse sucesso é, sem dúvida, a enorme capacidade que um dispositivo desse tipo tem de, em termos simples, *se ajustar à dinâmica da série temporal de interesse*.

#### **4.7. Conclusões e Perspectiva Histórica**

Havendo, pois, apresentado de forma sucinta as mais relevantes estruturas para filtragem não-linear em sistemas de comunicações, é oportuno que tenhamos algumas considerações gerais. Primeiramente, reafirmamos que a escolha de uma estrutura de filtragem está longe de ser um problema trivial. Quando se considera a amplidão do espectro de possibilidades abertas pelo uso racional da não-linearidade, percebe-se de imediato quão complexa é a opção que cabe ao projetista. Além disso, teoremas do tipo “*no free lunch*” [Duda et al., 2001] indicam claramente que não há um candidato que seja o mais eficiente em *todas as possíveis aplicações*. Isso reforça a importância de conhecer o problema em mãos e a palheta de possibilidades, fato que justifica o teor deste capítulo. Apesar disso, nada nos impede de resumir algumas conclusões e características das diferentes estruturas nos problemas que mais particularmente nos interessam aqui.

Quando se estuda o problema de equalização, jamais podemos nos furtar à necessidade de ter em conta o paradigma de Bayes, epítome do processo de decisão símbolo-a-símbolo [Duda et al., 2001]. Isso põe em destaque as estruturas capazes de emular o filtro ótimo, como as redes RBF [Mulgrew, 1996] e os filtros *fuzzy* [Patra, 1998]. Aliás, desde o final da década de 80, época em que tomou corpo a possibilidade de aplicar estruturas baseadas em inteligência computacional a problemas de comunicações, o emprego dessas estruturas tem sido comumente justificado nessa base.

Apesar dessa notável concordância, tais estruturas, em nenhum momento, inibiram o recurso à outra rede neural por nós discutida: a MLP. De fato, o prestígio de que goza esse dispositivo no campo da inteligência computacional fez com que ele fosse uma opção muito considerada desde o princípio [Widrow e Winter, 1988, Gibson et al., 1991]. Além disso, nem sempre é vantajoso do ponto de vista prático buscar a perfeita emulação do filtro de Bayes. Podemos apontar pelo menos dois motivos para isso:

- O projeto do equalizador ótimo requer o ajuste de parâmetros de natureza bastante distinta (centros, variâncias, pesos de saída etc.), o que pode tornar o processo de treinamento significativamente complexo tanto do ponto de vista teórico quanto do ponto de vista computacional.
- O aumento da ordem do canal ou do equalizador cria uma “explosão de dimensionalidade” que precisa ser acompanhada de um aumento equivalente do número de neurônios da camada intermediária (do ponto de vista *fuzzy*, do número de regras). Com isso, pode ser muito interessante empregar uma estrutura de caráter distinto, buscando sempre uma situação na qual uma pequena perda de desempenho seja o preço a pagar por uma vultosa redução de complexidade.

Motivos como esses mantêm plenamente acesa a aplicabilidade de estruturas não-lineares “sub-ótimas” em problemas de equalização, sendo, portanto, muito útil que o projetista tenha um conhecimento tal que o permita transitar com desenvoltura nesse emaranhado de compromissos.

Os filtros de Volterra são um caso especial, pois, nos anos 70 e 80<sup>94</sup>, eram considerados praticamente sinônimos de dispositivos não-lineares em problemas da área de comunicações [Benedetto et al., 1976, Benedetto e Biglieri, 1983]. Por conseguinte, foi através desses filtros que se consolidou a idéia de que era possível obter benefícios através da ampliação do leque de possibilidades estruturais práticas. Além disso, filtros polinomiais também foram decisivos para o desenvolvimento de sistemas compostos por elementos de caráter não-linear, como os sistemas de comunicação via satélite. No estado atual da teoria de equalização, essa classe de estruturas pode ser entendida como uma alternativa interessante ao paradigma de Bayes, de maneira análoga à discutida no contexto da MLP.

Em problemas de identificação de canal, também há uma nítida dominância de técnicas polinomiais nas décadas de 70 e 80 [Benedetto et al., 1979], sendo, nessa época, uma aplicação particularmente considerada a análise de amplificadores operantes no regime de saturação. Com a popularização das estruturas neurais e *fuzzy*, os dispositivos de Volterra se tornaram mais uma opção para o projetista, uma opção, aliás, ainda muito atraente por sua simplicidade conceitual [Kajikawa, 2000, Mathews e Sicuranza, 2000].

Em problemas genéricos de predição, redes neurais, filtros *fuzzy* e filtros de Volterra há muito conquistaram o *status* de estruturas assaz eficientes numa ampla gama de problemas. Já em comunicações, pela novidade do uso de predição não-linear para equalização [Cavalcante, 2001, Ferrari et al., 2003], a aplicabilidade dessas estruturas é bem mais restrita. Não obstante, teremos a oportunidade de expor uma contribuição nessa área no capítulo 12, local em que discutiremos com mais profundidade as características do problema.

Concluído o exórdio, tratemos de conduzir ao fim este capítulo através de algumas reflexões de cunho mais comparativo. Conforme temos insistido, a capacidade de emular o paradigma de Bayes põe a rede RBF e os filtros *fuzzy* numa posição especial em problemas de equalização. Apesar disso, a rede MLP ainda é uma opção muito interessante por dois motivos essenciais: seu treinamento é simples e voltado para parâmetros de um único tipo e sua estrutura tem um caráter que, em tese, é menos susceptível aos problemas de dimensionalidade que rondam o equalizador bayesiano.

O filtro de Volterra, por sua vez, ainda é uma opção atraente devido à sua simplicidade e à sua proximidade em relação ao bem-estabelecido arcabouço linear de filtragem. Há, no entanto, um ponto que merece atenção: o número de parâmetros de uma estrutura desse tipo também cresce de maneira notável com o número de entradas e o grau dos polinômios utilizados [Mathews e Sicuranza, 2000]. Além disso, julgamos que o filtro de Volterra possui um complicador adicional: ao contrário das demais estruturas, que têm unidades de processamento *pré-determinadas*, o filtro de Volterra, para que tenha uma garantida capacidade de aproximação, necessita de *diversos termos não-lineares distintos*. Isso torna o problema de truncamento mais complexo que no caso de uma rede neural: além de decidir o número de unidades, o projetista precisará definir *quais* não-linearidades polinomiais manterá no filtro.

---

<sup>94</sup> Uma era anterior ao *boom* de estudo e utilização de ferramentas baseadas em inteligência computacional.

Em identificação, todas as estruturas se encontram, *a priori*, em pé de igualdade, sendo sempre necessário analisar as características da planta a identificar e a disponibilidade de recursos computacionais. Apesar disso, num problema do tipo “caixa preta”, valem até certo ponto as considerações de projeto do parágrafo anterior.

Por fim, em predição, também é fundamental apreciar com cautela as características do problema e os recursos disponíveis, bem como os elementos de comparação. Há, no entanto, um aspecto importante: quando se tratar de projetar um *equalizador baseado em predição*, pode ser interessante levar em conta, outrossim, a conexão com o paradigma de Bayes [Ferrari, 2005].



## 5. Otimização Baseada em Computação Evolutiva

### 5.1. Introdução

Até o início do século XIX, foi soberana nos meios científicos a hipótese de que as espécies eram fixas e imutáveis [Smith, 1975], ou seja, de que não havia nenhum tipo de dinâmica biológica no que diz respeito às espécies animais e vegetais. Em termos simples, pensava-se que os frutos da reprodução dos seres vivos se mantinham estritamente dentro dos moldes aplicáveis a seus pais, sem que desse processo pudesse surgir algum tipo de “novidade”. Um ente qualquer pertencia a uma das gradações de uma grande escala imutável<sup>95</sup> de seres que manifestava a ordem de tudo o que sempre existiu no universo.

Um paradigma desse tipo, bastante coerente e chancelado por longos anos e por nomes ilustres, só pôde ser decisivamente questionado na esteira do enorme desenvolvimento da ciência empírica iniciado no Renascimento. Esse impulso irresistível trouxe uma torrente de novos subsídios provenientes da observação, sobre os quais foram erigidas novas hipóteses que revolucionaram diversos campos do conhecimento humano. Para o estudo das espécies vivas, foram relevantes as observações sobre a geologia da Terra, que sobremaneira contribuíram para que o início da história de nosso planeta fosse decisivamente deslocado para o passado, o estudo dos fósseis, e a análise comparativa de espécimes provenientes de várias partes do planeta. Foi tomando forma, com todos esses fermentos, uma nova hipótese: a de que *novas formas de vida poderiam se originar das existentes* [Futuyma, 2003].

O trabalho de Jean-Baptiste de Lamarck pode ser considerado um primeiro esforço que já contém o gérmen fundamental dessa nova corrente. Em sua *Philosophie Zoologique*, ele afirma que uma mudança de condições experimentada por determinados indivíduos é capaz de fazer com que seu processo reprodutivo possa, com o passar do tempo, conduzir a uma *nova espécie distinta da original* [Smith, 1975]. Em síntese, para Lamarck, o ambiente poderia levar a modificações na morfologia dos indivíduos, os quais, por sua vez, transmiti-las-iam a seus sucessores, numa corrente que seria capaz de criar uma escala de complexidade crescente. O exemplo clássico desse mecanismo é o fornecido pela girafa, cujo corpo, na opinião de Lamarck, sofreu modificações para melhorar seu acesso às folhas das árvores. Em suma, o cientista francês acreditava numa evolução guiada pela *herança de caracteres adquiridos* i.e. que a atrofia ou o desenvolvimento experimentado pelo indivíduo deveria, de alguma forma, influenciar os órgãos reprodutores de maneira a comunicá-lo à prole. É identificável, portanto, uma espécie de “unidade de ação corporal” [Smith, 1975].

O trabalho de Lamarck foi muito combatido por seus contemporâneos, principalmente pela idéia de evolução por ele apregoada, o que, ironicamente, terminou por contribuir para a popularização do assunto. Apesar disso, ainda se tratava apenas de um esforço corajoso contendo o gérmen de uma idéia ainda por desenvolver.

---

<sup>95</sup> Conhecida por *scala naturae*.

Não obstante tais dificuldades, a idéia vinha ganhando força, carecendo apenas de mais subsídios para frutificar<sup>96</sup>. Subsídios não faltaram a um certo jovem, Charles Robert Darwin, que, em 1831, embarcou no H. M. S. Beagle para uma viagem de quase cinco anos pelo hemisfério sul. Tal viagem foi uma fonte substancial de informações biológicas, geológicas e antropológicas, servindo também para apurar a metodologia do jovem pesquisador<sup>97</sup>. As vivências dessa jornada foram sendo ruminadas por Darwin após seu retorno, e levaram-no a aceitar a idéia de evolução. Suas ponderações, juntamente com a leitura do célebre *Essay on the Principle of Population* de Thomas Malthus, fizeram-no chegar a um princípio capaz de engendrar o processo evolutivo: a *seleção natural*, baseada na idéia da *luta pela existência*<sup>98</sup> travada a todo instante pelos organismos vivos. Em síntese, ele sustentava que a disputa pela sobrevivência era um mecanismo que conduzia à preservação de variações favoráveis e à eliminação de variações desfavoráveis surgidas no curso da história biológica [Futuyma, 2003].

Embora Darwin tivesse feito um ensaio sobre a seleção natural na primeira metade da década de 1840, ele publicou *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* apenas em 1859, após haver tido conhecimento de um trabalho paralelo de Alfred Wallace, cujo manuscrito lhe chegou às mãos em 1858<sup>99</sup>. A hora da nova idéia finalmente chegara.

É possível resumir a contribuição de Darwin através de dois itens fundamentais:

- As espécies surgem como modificações de ancestrais comuns;
- O principal agente da modificação é a *seleção natural*;

Essas noções já delineiam um processo evolutivo mais coerente que o concebido por Lamarck, uma hipótese dotada de muitas evidências favoráveis, como as discutidas por Darwin em seu trabalho seminal. Entretanto, ainda existe uma questão um tanto obscura: *como são transmitidas as modificações surgidas?* Em outras palavras, como se dá a modificação dos indivíduos e, mais ainda, o que a impede de desaparecer na geração seguinte?

Uma resposta convincente a essas perguntas só pôde aparecer com o advento da moderna genética, conseqüência da redescoberta dos trabalhos de Mendel no início do século XX. Isso nos remete à moderna síntese da *teoria da evolução*, fruto dos esforços de homens como Theodosius Dobzhansky, Ernest Mayr, George Simpson etc., cujos pontos para nós mais importantes apresentamos a seguir [Futuyma, 2003]:

- As populações contêm variação genética produzida por *mutações aleatórias* e *recombinação*;
- As populações evoluem através de mudanças nas frequências gênicas trazidas por determinados fenômenos como deriva e fluxo gênico, e, *principalmente, pela seleção natural*;

---

<sup>96</sup> O que traz à mente uma citação de Victor Hugo: “*On ne résiste pas à l’invasion des idées*”. Numa tradução livre: “Não se pode resistir à invasão das idéias”.

<sup>97</sup> Vide a interessante nota biográfica de *A Origem das Espécies – Esboço de 1842*, Clássicos Econômicos Newton, 1996.

<sup>98</sup> Em inglês, “*struggle for life*”.

<sup>99</sup> Elegantemente, Darwin encorajou Wallace e, a seu lado, apresentou formalmente a idéia de seleção natural em 1858.

As mudanças causadas pelos dois fatores essenciais, *mutação e recombinação*, são responsáveis por mudanças fenotípicas usualmente pequenas, as quais, com a passagem de gerações sucessivas e a especiação, terminam por originar alterações em grande escala, verificáveis em diferenças de níveis taxonômicos superiores (como gêneros e famílias) [Futuyma, 2003]. Em linhas gerais, esse é o cerne do estado atual da teoria, posta à prova, aliás, em inúmeras instâncias, fato que lhe granjeou o *status* de “teoria padrão” na biologia moderna.

## **5.2. Computação Evolutiva, Otimização e Filtragem**

A moderna síntese que acabamos de apresentar nos leva a uma conclusão interessante: o processo evolutivo é capaz de dar origem a boas características em populações de indivíduos por meio da atuação de dois elementos básicos: os processos de mutação e recombinação e a seleção natural. Tal sinergia, numa linguagem técnica, tende a incorporar novas características benéficas e a eliminar más características, tendo sempre como referência a compatibilidade entre indivíduo e todos os elementos do ambiente que o cerca. Desse prisma, é bastante natural encarar o processo como uma enorme *tarefa de otimização*<sup>100</sup>.

A idéia de aplicar tais conceitos a problemas de engenharia originou o campo da *computação evolutiva*, cujas raízes remontam à década de 50 [de Castro e Von Zuben, 2002a]. O campo se desenvolveu por múltiplas vias, o que acarretou a existência de paradigmas distintos, como *estratégias evolutivas* e *programação evolutiva*. Entretanto, talvez o germen da futura popularização da área tenha sido a publicação do livro *Adaptation in Natural and Artificial Systems*, escrito por John Holland, em 1975. Nesse trabalho, Holland se propôs a estudar diversos tipos de *sistemas adaptativos*, sendo um deles um *algoritmo genético* [Holland, 1992]. A proposta de uma técnica desse tipo<sup>101</sup> foi um impulso definitivo para a área, vindo a dar frutos numa verdadeira miríade de aplicações [Goldberg, 1989]. O sucesso na solução de problemas dos mais diversos tipos foi suficiente para obter um lugar para as técnicas evolutivas na palheta de muitos engenheiros que lidam com problemas de otimização.

Como vimos no capítulo 3, problemas de otimização são parte fundamental do projeto de equalizadores, identificadores e preditores. Além disso, temos nos deparado nos capítulos 3 e 4 com limitações inerentes aos métodos clássicos de otimização, limitações estas que podem levar a resultados sub-ótimos ou mesmo projetos absolutamente equivocados. Basicamente, os problemas são os seguintes:

- Os critérios de equalização cega, mesmo no caso de uma estrutura linear, podem ser multimodais.
- O emprego de estruturas não-lineares, mesmo em paradigmas supervisionados, produz comumente funções custo com múltiplos ótimos locais.

---

<sup>100</sup> Frisamos que as idéias de bom, ótimo, ruim etc. se referem *apenas* às condições do ambiente, não sendo nossa intenção aqui discutir os aspectos filosóficos da questão.

<sup>101</sup> É importante ressaltar desde já que falamos de métodos de otimização multimodal baseados numa população de candidatos a solução.

- O uso de estruturas recursivas, lineares e não-lineares, leva a problemas multimodais mesmo sob paradigmas supervisionados.
- O uso de técnicas baseadas em gradiente na adaptação de filtros recursivos comumente requer aproximações e, além disso, está sob constante ameaça de instabilidade.

Portanto, há três desafios a superar:

- Realizar a busca de uma maneira mais eficiente do ponto de vista de convergência global.
- Usar uma técnica que não requeira o cálculo do gradiente, evitando assim manipulações proibitivas e aproximações.
- Empregar uma técnica que seja robusta à existência temporária de configurações instáveis.

A superação dessas três barreiras nos levaria a um *paradigma ótimo de filtragem*, pois poderíamos dispor livremente de uma estrutura genérica operando no limite de sua capacidade e projetada, quando possível, sem a necessidade de supervisão.

Como buscaremos mostrar no restante do capítulo, técnicas baseadas em computação evolutiva constituem uma opção muito adequada a essas necessidades. Apresentaremos, em primeiro lugar, os consagrados *algoritmos genéticos*, base inicial de nossa pesquisa. Em seguida, discutiremos uma técnica fundada num processo evolutivo destinado à defesa de organismos superiores: uma *rede imunológica artificial*. Ao final do capítulo, voltaremos aos três desafios para avaliar se dispomos ou não de um paradigma promissor.

### **5.3. Algoritmos Genéticos**

Retornemos, por um instante, aos elementos fundamentais da moderna síntese da teoria da evolução. Consideremos a essência do cenário que ela nos oferece: há um conjunto de indivíduos dotados de diversas características que decorrem de seus genótipos, ou seja, de seu material genético, vivendo e se reproduzindo num determinado ambiente marcado pela existência de adversidades. Esse ambiente influencia a duração de suas vidas e, conseqüentemente, sua capacidade de gerar descendentes. Dessa maneira, o meio fornece *mecanismos de seleção* que influem decisivamente no destino e no caráter geral desse conjunto. Diversos indivíduos, por sua vez, são capazes de trocar material genético entre si por meio da reprodução sexuada, que podemos entender como uma espécie de *recombinação gênica*. Há ainda a presença do “imponderável”, ou seja, de mutações capazes de modificar o genótipo de um indivíduo qualquer sem que haja nisso um padrão aparente.

A interação de todos esses fatores produz gradualmente uma modificação dos indivíduos que, com a passagem do tempo, termina por originar um resultado macroscópico notável: a magnífica diversidade de que qualquer ecossistema é prova cabal: aves que são prodígios aerodinâmicos, peixes cujos corpos oferecem pouquíssima resistência ao fluxo da água, lagartos capazes de enganar seus predadores mudando a cor de suas peles etc. Por trás de tudo, jaz a “ordem no caos”: as *características do ambiente* e o *cenário marcado pela*

*seleção natural, pela tendência à perpetuação da espécie e pelo “acaso”*. Esses são os mecanismos essenciais de que faz uso a teoria para explicar a evolução: parecem muito simples, mas, do ponto de vista conceitual, podem ser associados a riquíssimas construções filosóficas: a “*vontade*” de que fala Schopenhauer, o “*élan vital*” de Bergson, o “*ponto ômega*” para onde converge a evolução segundo Teilhard de Chardin, para ficar em alguns poucos. Trata-se, enfim, de um problema essencial e de abissal profundidade, que podemos apenas discutir superficialmente aqui. Convém, no entanto, deixar no ar uma imagem da teoria evolutiva que nos será útil: a extrema complexidade dos seres vivos e sua notável adaptação às vicissitudes da vida se explicam pelo fato de a evolução ser “guiada” pela “mão invisível”<sup>102</sup> do ambiente, a qual determina o grau de aptidão de cada um, e pelo ímpeto reprodutor do que é vivo, que fornece novas gerações que poderão conter características genéticas favoráveis ao “melhoramento da espécie”.

“Melhoramento da espécie”, aliás, é exatamente o tipo de expressão emblemática que oferece perigo... No entanto, por outro lado, há nela uma analogia que nos interessa muito: num certo sentido, como já dissemos, a evolução<sup>103</sup> pode ser entendida como uma espécie de “otimização”. A base para tal analogia está no fato de o processo evolutivo *tender* a produzir espécies mais adaptadas ao meio, ou seja, “melhores”. Do ponto de vista da engenharia, qualquer processo de otimização pressupõe, ao menos, três elementos essenciais: *uma função objetivo*, que é exatamente a medida que se deseja otimizar, *possíveis soluções*, que são as configurações candidatas, e um *método de busca* qualquer. O processo evolutivo por nós descrito se encaixa nesses moldes da seguinte forma:

- A *função custo* ou *função objetivo* é algum tipo de medida que expresse o *grau de adaptação (fitness)* de um indivíduo qualquer. Essa função será a “mão” que tenderá a moldar os indivíduos com o passar do tempo.
- As possíveis soluções são os seres vivos, podendo cada um ser representado, num cenário simples, pelo seu *material genético*. Nesse caso, as diferentes configurações dos genes seriam responsáveis pela diversidade de possíveis indivíduos.
- Por fim, o processo de busca delineado pela teoria é basicamente uma combinação de mecanismos de seleção e *operadores* que contemplam a reprodução sexuada e a mutação espúria.

A analogia que acabamos de descrever é a essência da *computação evolutiva* como um todo, pois indica claramente uma ponte entre uma interpretação de um processo natural e uma gama de problemas de engenharia. Mais particularmente, essa lista com três itens é uma síntese da fundamentação teórica de um *algoritmo genético (AG)*, que, fundamentalmente, é a versão computacional do processo evolutivo descrito.

Para utilizar um AG num problema de otimização, precisamos seguir os passos da lista tríplex apresentada: *explicitar a função custo, codificar os parâmetros numa forma apropriada e executar algumas vezes o receituário descrito no item final*. Cada uma dessas etapas contém peculiaridades e suscita problemas que analisaremos com profundidade conveniente. Tratemos, pois, de tentar entendê-las.

---

<sup>102</sup> Esperamos que a analogia com o conceito introduzido por Adam Smith em sua obra-prima faça mais bem que mal.

<sup>103</sup> Ressaltamos que, por uma questão de simplicidade, poderemos eventualmente nos referir à “moderna teoria da evolução” simplesmente como “evolução”, sem que haja, com isso, nenhuma adesão implícita.

### 5.3.1. Considerações sobre a Função Custo

Pouco há que se possa arbitrar quando se trata da função custo, pois, via de regra, ela nos é imposta pelas características do problema em mãos. Entretanto, são úteis duas considerações:

- Em algumas etapas posteriores, como na de seleção, pode ser bastante útil que, por exemplo, a função custo não assuma valores negativos. Por esse motivo, o projetista deve analisar a conveniência de submeter a função real a uma transformação linear conveniente [Goldberg, 1989].
- A idéia de evolução sugere um problema de *maximização* de *fitness*. Portanto, pode ser necessário aplicar uma transformação que converta o problema de minimização num problema de maximização. Uma possibilidade comum, que teremos por regra, é fazer:

$$J_{FIT} = \frac{1}{1 + J_{CUSTO}} \quad (5.1)$$

sendo  $J_{CUSTO}$  a função a minimizar<sup>104</sup>.

### 5.3.2. Representação / Codificação

Tendo obtido a função custo na forma apropriada, deparamo-nos com uma nova questão: como representar o conjunto de parâmetros que desejamos otimizar? Em primeiro lugar, esses parâmetros podem ter naturezas diversas: números reais, grandezas binárias, inteiras etc., o que significa que o projetista deverá ponderar a melhor maneira de *codificar* essa informação. Voltando à nossa analogia evolutiva, percebemos que, essencialmente, cada vetor de parâmetros faz o papel de “material genético” de um indivíduo. Por uma questão de simplicidade, é-nos lícito imaginar que nosso vetor corresponde a um determinado *cromossomo*, cujos parâmetros são *genes*. Esse material genético tem a si associado um determinado custo que, como já discutimos, representa o grau de adaptação do indivíduo dotado do fenótipo que dele emerge.

No cenário mais clássico, cada solução é representada por um cromossomo de tamanho fixo e formado por genes binários. Em outras palavras, cada solução é codificada segundo um alfabeto binário. Essa possibilidade apresenta algumas vantagens principais [Michalewicz, 1996]:

- Conforme discutido em [Goldberg, 1989], é possível mostrar que a forma binária é a mais eficiente codificação quando se considera o papel dos possíveis trechos de cromossomos na geração de potenciais características desejáveis<sup>105</sup>.

---

<sup>104</sup> O mapeamento proposto tem duas características interessantes: é bijetor e tem um denominador que não se anula no domínio de  $J_{CUSTO}$ .

<sup>105</sup> O leitor já familiarizado com a área reconhecerá a noção subjacente de *schema*.

- Essa codificação dá origem a operadores genéticos particularmente simples e elegantes.
- A codificação binária se adequa bem ao espírito da computação digital.

Apesar desses pontos positivos, quando se trabalha com problemas de otimização real, torna-se bastante imediata a idéia de empregar um cromossomo que seja exatamente o vetor de parâmetros. Nesse caso, embora haja um afastamento da analogia natural (pois há um número finito de bases que compõem os ácidos nucléicos), certamente se obtém um problema de otimização mais simples e direto. Em [Michalewicz, 1996], o autor compara ambos os esquemas de codificação e chega à conclusão de que a codificação real pode apresentar vantagens significativas do ponto de vista de, por exemplo, precisão e velocidade de convergência. De qualquer forma, “no frigidus dos ovos”, a escolha cabe mesmo ao projetista, cuja sensibilidade às características do problema a resolver deve ser sempre aguçada<sup>106</sup>.

### 5.3.3. População Inicial

Tendo obtido a função de *fitness* e decidido o esquema de codificação, estamos prontos a seguir rumo à etapa final: disparar nosso processo evolutivo. O primeiro passo nesse sentido é gerar uma população inicial de indivíduos, ou seja, o conjunto a partir do qual se dará o trabalho da evolução. Basicamente, a idéia é produzir uma população inicial *aleatória* e espalhada tão uniformemente quanto possível pelo espaço de busca. Eventualmente, caso se disponha de conhecimento *a priori* sobre o problema, ele pode ser explorado para gerar uma inicialização mais eficiente.

A escolha do tamanho da população deve ser feita tendo em mente a complexidade do problema no espaço de busca e as limitações de tempo e recursos. Trata-se de um compromisso que precisa ser ponderado pelo projetista ou contemplado por algum mecanismo de “inserção e poda” presente na técnica evolutiva. A última solução é particularmente desejável em cenários variantes ou em contextos em que a introdução do mecanismo seja menos onerosa que o excesso populacional que ela evita. Veremos que a rede imunológica artificial que apresentaremos adiante possui um interessante controle desse tipo.

### 5.3.4. Seleção e Reprodução

Uma vez que já dispomos da matéria bruta, é preciso que comecemos a buscar os instrumentos que a lapidarão. É hora de fazer nossos indivíduos experimentarem dois processos onipresentes: a *seleção* e o *acasalamento*. Nessa etapa de nosso AG, tais mecanismos se relacionam intimamente: consideramos que existe um elo entre aptidão individual e capacidade reprodutiva. Um argumento simples e direto a favor dessa tese é que uma criatura mais apta tem mais chance de viver até a fase madura, bem como de

---

<sup>106</sup> Teoricamente, para cada problema existe uma codificação mais apropriada. Entretanto, na prática, a regra é buscar uma codificação factível que, mesmo sem ser ideal, seja superior a outras opções estabelecidas. Note o leitor que a escolha da codificação termina por determinar as características da função de *fitness*, o que exige uma dose extra de cautela para que sejam evitadas características nocivas à busca (e.g. mudanças abruptas no valor do custo).

conquistar o direito de combinar seu material com o de seu semelhante<sup>107</sup>. Tal cenário sugere que a seleção dos indivíduos que se agruparão para trocar material genético deve estar relacionada à medida de *fitness*.

Uma maneira bastante direta de relacionar *fitness* a seleção é supor que a probabilidade de que um indivíduo qualquer seja escolhido para se reproduzir é *proporcional ao seu valor de custo*. Tal suposição dá origem ao mais comum dos métodos de seleção, o *algoritmo da roleta*<sup>108</sup> (*roulette-wheel*). Para ilustrar sua aplicação, mostramos uma população fictícia na Tab. 5.1.

<b>Indivíduo</b>	<b><i>Fitness</i></b>	<b>Participação Relativa</b>
0001	4	6.25%
0010	8	12.5%
1111	20	31.25%
1110	32	50%

**Tabela 5.1:** População Hipotética

O primeiro indivíduo contribui com 6.25% do total de *fitness* da população, sendo, portanto, esse mesmo percentual correspondente à sua probabilidade de seleção. Aplicando essa mesma regra para os demais indivíduos, chegamos a um cenário em que, a título de curiosidade, o melhor indivíduo tem oito vezes mais chance de ser selecionado que o pior. Isso revela que há uma significativa valorização das características dos melhores membros da população e uma certa desvalorização da contribuição que o pior poderia dar.

Embora esse estado de coisas seja, até certo ponto, natural no processo, há um risco subjacente que convém desde já explicitar: caso haja um predomínio excessivo de determinados indivíduos, muito cedo a população pode tender a uma significativa uniformização, o que significa que existirá uma *perda de diversidade populacional*. Poderia acontecer o seguinte: surge um indivíduo, que nem sequer está próximo ao ótimo global, mas que, num dado momento, tem um *fitness* muito melhor que o dos demais. Esse indivíduo poderá dominar o processo reprodutivo, fazendo com que suas características predominem cada vez mais na população. Se a polarização for excessiva, logo haverá uma decisiva concentração e uma conseqüente perda da capacidade de busca global do algoritmo<sup>109</sup>. Do ponto de vista de otimização, isso pode ser equivalente a uma estagnação em torno de um mínimo local. Do ponto de vista teórico, podemos dizer que o operador possui uma *forte pressão seletiva*. É preciso que o projetista esteja sempre atento a essa possibilidade.

Um método alternativo que se nos afigura interessante é a *seleção por torneio*. Nesse caso, o indivíduo selecionado é aquele que tiver obtido o maior número de vitórias

<sup>107</sup> Um exemplo disso é a disputa entre machos de um grupo por uma fêmea.

<sup>108</sup> O nome do algoritmo advém do fato de que a probabilidade de que ocorra um determinado número na roleta é proporcional à área por ele ocupada.

<sup>109</sup> É muito interessante notar que a idéia de manutenção de diversidade é a tônica da seguinte crítica feita pelo campeão da defesa da representatividade das minorias, John Stuart Mill, a seus contemporâneos: “É contra a individualidade que batalhamos: julgaríamos ter feito maravilhas se nos houvéssemos tornado semelhantes, todos, uns aos outros, olvidando que a dessemelhança das pessoas é geralmente o que mais fixa a atenção de cada uma imperfeição do próprio tipo e na superioridade do outro – *ou na possibilidade de, combinando as vantagens de ambos, produzir algo melhor que qualquer um dos dois.*” (itálico nosso) in *Sobre a Liberdade*, Editora Vozes, 1991.

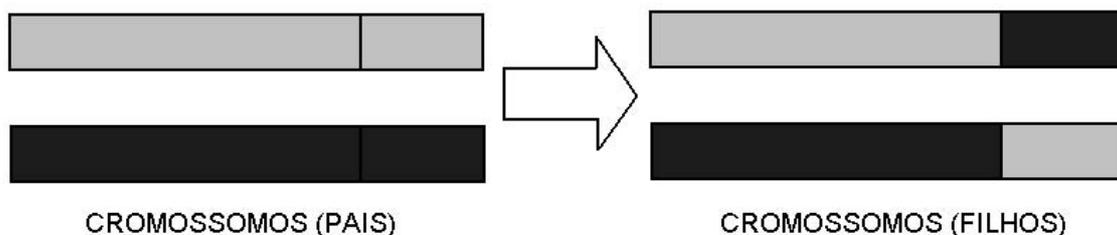
num conjunto de  $N$  “competições” contra  $q$  indivíduos escolhidos aleatoriamente. A competição consiste simplesmente de uma comparação entre os valores da função de *fitness*. Tipicamente, um valor de  $q$  muito alto leva a uma forte pressão seletiva (tendência à perda de diversidade), enquanto um valor entre 3 e 5 leva a uma pressão considerada fraca [de Castro e Von Zuben, 2002a]. Quando  $q = 2$ , tem lugar um *torneio binário*. A grande vantagem de um mecanismo de seleção desse tipo é exatamente a possibilidade de controlar mais efetivamente a tendência à uniformização.

Uma vez que tenhamos escolhido, por um processo qualquer de seleção, um par ou conjunto de pares, estamos aptos a iniciar a reprodução sexuada. Em nosso algoritmo, submeteremos os pares à atuação de um operador de *crossover* ou *recombinação*. Tratemos, pois, de defini-lo.

### 5.3.5. Crossover

Para o processo evolutivo, a reprodução sexuada tem um papel muito importante: ela permite, em tese, que *boas características adquiridas*, ou seja, que determinadas combinações gênicas que produzem manifestações fenotípicas desejáveis sejam intercambiadas e possam se espalhar pela população, gerando, potencialmente, novas combinações que poderão ser ainda melhores. De certa forma, o mecanismo transmite uma idéia de *refinamento*, de explorar ao máximo o patrimônio genético de uma população.

O mecanismo clássico de *crossover* é muito simples. Dados dois indivíduos “pais”, P1 e P2, escolhidos por um processo de seleção qualquer, existirá uma probabilidade  $p_c$  de que seus cromossomos sofram uma cisão num certo ponto e sejam produzidos dois indivíduos a partir do intercâmbio do material genético presente na região que vai do ponto de corte até o final do cromossomo. A Fig. 5.2 ilustra o processo básico de troca de informação.



**Figura 5.2:** Esquema Básico de *Crossover*

São destarte gerados dois filhos, F1 e F2, que podem ter vários destinos: ser incorporados a uma sub-população provisória com seus pais, tomar o lugar de seus pais na população, competir pela existência com eles etc.

Diversas variações podem incrementar o mecanismo de *crossover*: introduzir múltiplos pontos de cisão, determinar novos métodos de escolha dos pontos, criar uma máscara de genes que serão trocados em vez de ter por base trechos do cromossomo (*crossover uniforme*)<sup>110</sup> etc. Quando se trabalha com codificação real, há ainda algumas

<sup>110</sup> No caso do método de *crossover* mostrado na Fig. 5.2, os elementos do cromossomo têm, de acordo com sua posição, uma maior ou menor probabilidade de pertencer ao conjunto trocado. Caso essa distinção não

possibilidades adicionais importantes, dentre as quais destacamos o *crossover aritmético*, que gera os filhos da seguinte forma:

$$F1 = \alpha P1 + (1-\alpha)P2 \quad (5.2)$$

$$F2 = (1-\alpha)P1 + \alpha P2 \quad (5.3)$$

sendo P1 e P2 os pais e  $\alpha$  um parâmetro de controle<sup>111</sup>. Poderíamos expor outras variantes, mas não julgamos que tal procedimento seja necessário aos fins deste trabalho, sendo suficiente que aconselhemos o projetista a conhecer em profundidade os caminhos promissores de seu problema e o repertório de soluções disponível. Isso conclui nossa discussão sobre o primeiro operador.

### 5.3.6. Mutação

Embora talvez viesse a ser um eficiente processo de busca local, um algoritmo genético dotado apenas de um operador de *crossover* seria muito pouco efetivo ante as possibilidades abertas pelo estudo do processo evolutivo. A reprodução sexuada justifica muito bem como podemos cruzar diferentes espécies de soja para obter uma espécie mais resistente a uma praga, mas nos parece assaz insuficiente quando imaginamos que a evolução precisa explicar de um modo convincente a forma pela qual um homem surgiu a partir de seres unicelulares. É imperativo que exista um mecanismo capaz de *introduzir novos elementos* no patrimônio genético da população, de suscitar configurações inovadoras que a chancela da seleção natural cuidará de manter e o mecanismo de *crossover* tenderá a explorar e disseminar. Esse mecanismo é a *mutação* i.e. a *modificação espúria do material genético de um indivíduo*.

Causa estranheza imaginar que a teoria sustente que a imensa complexidade dos organismos vivos provenha, em parte, de um mecanismo “cego”, puramente aleatório. Não nos esqueçamos, no entanto, de que a teoria considera a existência da inexorável<sup>112</sup> “mão invisível” representada pela medida de *fitness*, que exerce uma influência decisiva no processo de seleção, e, conseqüentemente, gera coerência<sup>113</sup> a partir de um movimento randômico. Para ser mais justo, no cenário que vimos esboçando, essa coerência emerge da sinergia entre *crossover* e *mutação*. Do ponto de vista de otimização, essa é uma sinergia entre um fator de *busca mais localizada* e um fator que introduz um enorme potencial de varredura do espaço de busca, ou seja, de *busca global*.

---

seja condizente com a estrutura do problema a resolver, tal operador pode ser pouco representativo. O *crossover* uniforme, por se basear numa “lista” de genes escolhidos aleatoriamente para troca, evita esse problema.

<sup>111</sup> Embora seja comum restringir o domínio de validade do parâmetro  $\alpha$  ao intervalo [0,1], pode ser interessante em alguns casos aumentar a capacidade de varredura do mecanismo através da inclusão de valores negativos.

<sup>112</sup> A inexorabilidade não significa aqui uma imutabilidade do ambiente. Por exemplo, na era glacial, a idéia de adequação era bastante diferente da que poderíamos conceber hoje na selva amazônica.

<sup>113</sup> Julgamos coerente que um morcego disponha de um preciso sonar e que um habitante das regiões abissais seja translúcido. Isso só manifesta que há uma correlação entre nossa impressão acerca da medida de *fitness* e a idéia abstrata de um grau de ajuste inerente à natureza. Novamente, frisamos que não é nossa vontade adentrar esse terreno de discussão filosófica.

Num algoritmo clássico com codificação binária, existe uma probabilidade  $p_m$  de que um gene da população mude de valor<sup>114</sup>. Esse é o mecanismo básico de mutação, que pode ser incrementado pela introdução de um processo de diminuição gradual da probabilidade  $p_m$ , o que dividiria a busca numa etapa mais global e noutra de refinamento, ou mesmo criando uma proporcionalidade entre taxa de mutação e *fitness*. Já no caso real, quando um indivíduo  $I$  sofre mutação, vale usualmente a seguinte regra:

$$I \rightarrow I + \Delta I \quad (5.4)$$

sendo  $\Delta I$  usualmente um vetor de variáveis aleatórias gaussianas ou uniformes de média zero. Os parâmetros dessa variável podem variar com o tempo ou podem depender do valor de *fitness* de  $I$ <sup>115</sup>.

### 5.3.7. Seleção Final e Elitismo

O algoritmo genético mais clássico termina exatamente com o processo de mutação [de Castro e Von Zuben, 2002a]. Tal abordagem, embora coerente, pode levar a uma situação curiosa: *o melhor indivíduo pode se perder no processo*<sup>116</sup>. É inegável que tal possibilidade é factível num ecossistema qualquer: podemos conceber a morte prematura de um indivíduo excelente devido a algum fator fortuito. Entretanto, não somos obrigados a nos ater fielmente a um determinado cenário biológico, especialmente se a introdução de novas idéias for vantajosa para o método de busca. Tendo em vista tais considerações, podemos chegar de imediato a uma possibilidade: ao final de uma geração, se o melhor indivíduo tiver sido eliminado, devemos inseri-lo *a posteriori* no conjunto antes do início de uma nova etapa. Podemos resumir a base desse mecanismo, denominado *elitismo*<sup>117</sup>, de maneira simples: incomoda-nos perder o que já conquistamos, mesmo havendo a possibilidade de que essa perda seja apenas temporária.

Não desejamos, com o início de nossa argumentação, dar a idéia de que o processo só tem valor como um “conforto” para o projetista. A verdade é que esse mecanismo pode trazer ganhos reais em vários sentidos, como, por exemplo, em termos de velocidade de convergência. Mas ele também traz um questionamento de muito interesse: a manutenção forçada do melhor indivíduo não conduz o algoritmo a uma espécie de “situação polarizada”? Em outras palavras, um mecanismo elitista não pode levar a uma perda mais pronunciada de diversidade populacional? Isso nos remete a um problema mais amplo, pois é preciso que consideremos esse fator juntamente com o mecanismo de seleção adotado (discutido na seção 5.3.4). Uma idéia interessante, a qual leva a um algoritmo genético modificado, é gerar, no processo de reprodução, *um conjunto de sub-populações formadas por pais e descendentes já submetidos aos operadores*. Introduce-se, outrossim, uma *fase final de seleção*, em que essa população aumentada é filtrada para que seja obtido o conjunto de soluções que dará início à próxima geração. A escolha do critério de seleção

---

<sup>114</sup> Quando há um alfabeto finito com mais de duas possibilidades, é preciso criar uma árvore de possíveis transições entre os símbolos.

<sup>115</sup> Um processo de mutação desse tipo será parte essencial da opt-aiNet, como veremos mais adiante.

<sup>116</sup> Isso é particularmente marcante da seleção via roleta. A perda também pode se dar, por exemplo, devido a uma mutação.

<sup>117</sup> Alguns autores preferem dar o nome de “salvacionismo” a esse processo.

final pode favorecer a manutenção de diversidade caso sejam preservados indivíduos ruins e medianos<sup>118</sup>.

### 5.3.8. Nicheing

Não é possível imaginar a humanidade surgindo e florescendo nas regiões abissais do oceano. Tampouco é razoável imaginar um tubarão à vontade na savana ou um tucano no pólo norte. Embora residam no domínio da extrapolação, raciocínios desse tipo ilustram bem o fato de que há diversos fatores que influenciam a existência de uma espécie. Tais fatores delimitam os possíveis ambientes em que ela pode sobreviver, ou seja, estabelecem seu *nicho ecológico* [Futuyma, 2003].

No contexto de otimização, podemos imediatamente associar a idéia de nicho a uma determinada região do espaço de busca associada a um grupo de configurações dos parâmetros procurados. Em outras palavras, podemos promover a evolução populacional respeitando os limites estabelecidos pelos nichos, ou seja, efetuando a competição de acordo com o que eles delimitam. Essa idéia garante, de imediato, que a busca contemplará diversas regiões do espaço, contribuindo decisivamente para aumentar o potencial de varredura da ferramenta de otimização. Por esse motivo, a introdução do conceito de *nicho ecológico* no esquema que temos debatido pode ser muito importante para o aumento de sua capacidade de busca global e para seu potencial de manutenção de diversidade. Tal empreitada deu origem às técnicas de *niching*.

Para definir, numa população qualquer de indivíduos, um nicho ecológico, é preciso que se disponha de um critério que possa estabelecer o grau de pertinência de um indivíduo ao nicho. Usualmente, a regra é elaborar uma medida de distância para determinar se dois indivíduos quaisquer compartilham o mesmo nicho. Para explorar essa medida, uma possibilidade é construir uma função que indique, para um dado indivíduo, o *grau de compartilhamento* (*sharing*) de todos os indivíduos da população [Mahfoud, 2000]. Tipicamente, essa função vale 1 para uma distância euclidiana nula e satura em zero para uma distância superior a um dado limiar. Podemos então calcular a soma de valores de compartilhamento para cada indivíduo, levando em conta todos os indivíduos da população. Supondo que essa soma seja  $S_i$ , calcula-se uma nova medida  $f_s$  de *fitness* para o  $i$ -ésimo indivíduo:

$$f_s(x_i) = \frac{f(x_i)}{\sum_{k=1}^{N_i} \eta[d(x_i, x_j)]} \quad (5.5)$$

sendo  $f(x_i)$  o valor original de *fitness*,  $d(\cdot)$  a distância euclidiana e  $\eta$  uma função de compartilhamento que, geralmente, vale:

$$\eta(d) = 1 - (d/\sigma_{SH}), \text{ se } d < \sigma_{SH}, \text{ e } \eta(d) = 0 \text{ para todos os demais casos} \quad (5.6)$$

onde  $\sigma_{SH}$  é um parâmetro de controle do decaimento. A lógica desse método, conhecido como *fitness sharing* [Goldberg, 1989], é penalizar uma situação em que haja uma

---

<sup>118</sup> É importante que o leitor tenha sempre em mente que um indivíduo ruim pode conter trechos valiosos de informação em seus genes, informação esta que pode florescer caso os operadores ajam favoravelmente a isso. Daí a importância de sua preservação.

concentração excessiva de indivíduos, que levaria a um denominador elevado em (5.5), e, conseqüentemente, a um baixo valor ponderado de *fitness*. Com isso, a tendência é que sejam formados nichos em torno das diversas regiões de *fitness* máximo, o que significa que tenderá a haver um satisfatório grau de diversidade e, conseqüentemente, um potencial expressivo de busca global. O principal problema do método é: *como estabelecer, sem conhecer demasiadamente o problema a priori, a medida de sharing?*

Uma outra possibilidade de formar nichos é, simplesmente, anular o *fitness* de um determinado número de indivíduos pertencentes a um dado nicho, sendo o critério de escolha a própria medida de *fitness*. Essa abordagem, denominada *método de clearing* [Sareni e Krähenbühl, 1998], mostra-se eficiente caso seja possível definir satisfatoriamente as fronteiras entre os diversos nichos<sup>119</sup>.

### 5.3.9. Sumário de um Algoritmo Genético Típico

As seções precedentes tangeram os principais elementos de um algoritmo genético, o que nos credencia a buscar uma visão geral do processo. Na Tab. 5.2, apresentamos um diagrama que mostra de maneira resumida e sistemática o que vimos discutindo.

#### 1 – Inicialização

- Escolha a codificação apropriada e efetue as transformações convenientes da função custo.
- Escolha os operadores e decida se existirá algum mecanismo de *niching*.
- Escolha o tamanho da população inicial e gere de maneira aleatória os indivíduos. Caso haja conhecimento *a priori* sobre o problema, ele pode ser usado para gerar parte da população.

#### 2 – Processo Iterativo

Até que um critério de parada<sup>120</sup> seja satisfeito,

- Selecione um conjunto de pais e empregue o operador de *crossover* adotado para gerar os descendentes. Introduza-os juntamente com os pais numa população intermediária ou estabeleça algum processo de competição entre eles.
- Aplique o operador de mutação adotado à população proveniente da etapa anterior.
- Se necessário, forme a população final a partir da população intermediária. Para isso, estabeleça uma composição que pode conter indivíduos bons, ruins e medianos. Aplique, se necessário, uma política elitista.

**Tabela 5.2:** Resumo de um AG típico

Para concluir a seção, apresentamos um exemplo de aplicação de um AG.

<sup>119</sup> O escopo desta tese não torna conveniente uma discussão mais ampla. Recomendamos ao leitor interessado nesses métodos a leitura da referência citada.

<sup>120</sup> O critério de parada é uma condição qualquer que, quando atingida, sinaliza que a técnica empregada alcançou seu objetivo ou já atuou por um período suficiente.

---

### **Exemplo 5.1**

Seja a seguinte função custo, cujo máximo global no intervalo [0,1] desejamos encontrar [de Castro e Von Zuben, 2002a]:

$$f(x) = 2^{-2\left[\frac{x-1}{0.9}\right]^2} \text{sen}^6(5\pi x) \quad (5.7)$$

Para otimizá-la, adotamos um AG clássico com codificação binária, seleção por meio de roleta e elitismo. Os parâmetros adotados se encontram na Tab. 5.3. Por se tratar de um exemplo didático, adotamos como critério de parada a obtenção do valor de *fitness* ótimo.

<b>Parâmetro</b>	<b>Valor</b>
Comprimento do Cromossomo <sup>121</sup>	22
Número de Indivíduos da População	16
Probabilidade de <i>Crossover</i>	0.9
Probabilidade de Mutação	0.005

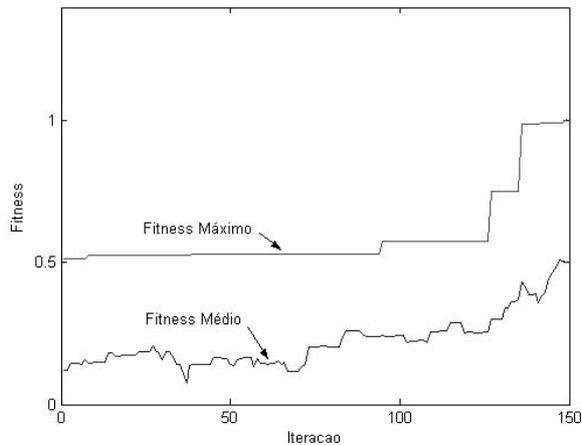
**Tabela 5.3:** Parâmetros do Algoritmo Genético

Em todos os ensaios realizados, o AG foi capaz de encontrar o ponto ótimo. Para ilustrar o comportamento típico do algoritmo, apresentamos na Fig. 5.3 a evolução do maior valor de *fitness* da população e do valor médio. É perceptível o caráter um tanto abrupto dos avanços do valor de *fitness*: isso se deve tanto à característica discreta da codificação quanto aos efeitos de mutações e *crossovers* bem-sucedidos. A curva de *fitness* médio mostra que o AG se comportou bem, pois segue a tendência de evolução do melhor indivíduo *sem que haja uma proximidade excessiva entre as curvas*, fato que indicaria um baixo nível de diversidade.

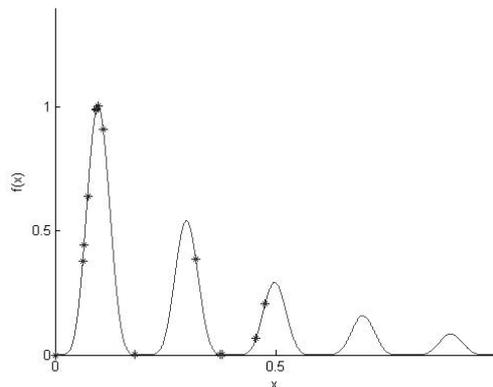
Um outro fato que confirma a existência de diversidade é a análise da população final do algoritmo. Mostramo-la na Fig. 5.4, tendo como pano de fundo a própria função custo. Percebe-se com clareza que há indivíduos em torno não apenas do ótimo global, mas também junto aos dois melhores ótimos locais, e isso tudo sem que haja mecanismos *ad hoc* como *niching*. Isso reafirma a boa atuação do AG.

---

<sup>121</sup> Neste caso, o comprimento do cromossomo se relaciona com a precisão da codificação binária dos valores reais pertencentes ao intervalo escolhido.



**Figura 5.3:** Evolução dos Valores de *Fitness*



**Figura 5.4:** Função Custo e População Final

## 5.4. Sistemas Imunológicos Artificiais

Uma lição que todo ser vivo conhece ou intui é que a manutenção da vida requer um sem-número de cuidados. Esse fato se reflete numa série de mecanismos de admirável precisão presentes em todas as espécies. Dentre os processos essenciais à preservação do indivíduo, merecem destaque aqueles relacionados à defesa de seu organismo contra invasores, processos estes que formam o *sistema imunológico*.

As primeiras evidências substanciais no sentido da existência de tal sistema remontam aos esforços de Edward Jenner para combater a varíola, os quais, em 1796, deram origem à primeira vacina [de Castro e Timmis, 2002a]. Posteriormente, foi demonstrado por Koch que havia *agentes causadores* de doenças infecciosas, o que permitiu que o campo de estudo avançasse a passos cada vez mais largos no sentido de produção de novas vacinas e de aprofundamento da base teórica. No final do século XIX, já se conheciam células brancas responsáveis pela destruição de invasores, e, no século seguinte, os mecanismos de identificação de corpos estranhos foram enquadrados em

sólidas teorias [de Castro, 2001]. Procuraremos, ao longo desta seção, apresentar uma visão bastante simples das modernas teorias que buscam explicar o funcionamento do sistema, com particular ênfase em conceitos que nos serão especialmente valiosos.

#### 5.4.1. Conceitos e Mecanismos

Constantemente, nosso corpo é invadido por agentes causadores de doenças (*patógenos*), bem como por outros elementos estranhos capazes de influir negativamente no funcionamento do organismo. Esses invasores, para complicar ainda mais o quadro, tem naturezas diversas, *o que desafia qualquer possibilidade de pleno conhecimento a priori*.

É desejável, para a manutenção da vida, que tais agressores sejam postos fora de ação ou destruídos. Surge aí, desde já, um cenário complexo: será travado algum tipo de combate, e será travado *ante os muros da cidadela a defender*. Nesse caso, deve ser descartada *a priori* a possibilidade de que o processo de defesa envolva uma simples destruição em massa. O organismo precisará *identificar* os patógenos e combatê-los, mas sempre evitando *que também sejam afetadas partes suas*. Não basta a “força bruta”: há uma necessidade premente de *reconhecimento dos invasores*.

Os dois mecanismos complementares de *inteligência e destruição*<sup>122</sup> integram, de acordo com o estado atual da teoria, um sistema imunológico formado pela cooperação entre uma parte *inata* e uma parte *adaptativa* [Abbas et al., 1997]. De modo geral, a parte inata é capaz de reagir a um amplo conjunto de microorganismos, enquanto a parte adaptativa contém mecanismos de defesa mais refinados, *capazes de sofrer modificações* a partir da exposição a novas substâncias alheias (*antígenos*)<sup>123</sup>. Portanto, temos uma linha de defesa que responde de maneira geral a praticamente qualquer tipo de invasor e uma linha especializada que responde diferentemente a invasores distintos. Essa divisão em duas linhas de defesa é interessante, por exemplo, quando há a invasão de um antígeno desconhecido, pois a primeira linha fornece uma fonte de proteção imediata até que os elementos da segunda tenham conhecido suficientemente os invasores.

Interessa-nos particularmente aqui a segunda linha de defesa, pois nela se encontra a *capacidade de aprendizado* que buscamos em qualquer técnica de computação. Em primeiro lugar, precisamos definir um pouco melhor o que entendemos por *reconhecimento de um antígeno*. Na corrente sanguínea circulam inúmeros linfócitos com características diferentes (essa diversidade decorre do processo de produção dessas células na medula), ou seja, com *receptores* distintos. Dizemos que uma dessas células *reconhece* o antígeno quando há uma compatibilidade entre esses receptores e as características da molécula antigênica. Um reconhecimento positivo dispara uma série de processos dos quais faz parte a secreção de anticorpos, substâncias capazes de neutralizar e sinalizar o patógeno. O anticorpo, naturalmente, é efetivo, pois se relaciona com a estrutura do receptor eficiente.

Um primeiro mecanismo relevante para o aprendizado é a *seleção clonal*. Sua essência é a seguinte: quando há um reconhecimento positivo do antígeno, a célula bem-sucedida *passa a produzir clones de si mesma*, ou seja, dá início a um processo de

---

<sup>122</sup> Aqueles que desejarem traçar analogias entre a guerra e o funcionamento do sistema imune não deixarão de notar a importância dada à sinergia entre esses dois elementos por célebres generais como Sun Tzu e Clausewitz.

<sup>123</sup> Interessantemente, o conhecimento adquirido pela parte adaptativa contribui para o refinamento dos mecanismos gerais da parte inata.

replicação que gera um conjunto de cópias capazes de levar à produção de anticorpos adequados ao antígeno identificado. Esse mecanismo tem duas conseqüências:

- Cria um mecanismo análogo ao de *seleção natural*, pois apenas as boas células de defesa tendem a sofrer replicação.
- Leva, em conjunto com outros mecanismos que não discutiremos aqui [de Castro, 2001], à produção de um repertório de *células de memória* capazes de armazenar a informação obtida acerca do antígeno, fazendo com que a resposta do organismo seja mais imediata no caso de uma segunda invasão pelo mesmo ente<sup>124</sup>.

Entretanto, a simples replicação em nada contribui para que o repertório inicial de linfócitos presente no organismo se desenvolva e passe a ser capaz de reconhecer com mais precisão os antígenos: faltam mecanismos que levem à produção de novas características. A análise dos fenômenos ocorridos durante o processo de geração dos descendentes mostra que podemos entender parte deles como uma *mutação* das células de defesa, tão mais intensa quando menos efetiva a célula for. Em outras palavras, podemos considerar a existência de um mecanismo de mutação proporcional à *afinidade* entre receptor e antígeno. Em conjunto com esse mecanismo principal, atuam ainda mecanismos capazes de selecionar apenas os melhores produtos dessas mutações e um engenhoso esquema capaz de introduzir material significativamente novo na população de células de defesa (*edição de receptores*). Para nossos fins, esse conjunto de processos constitui uma primeira frente de atuação do sistema imunológico.

Devemos a Niels Jerne [de Castro e Timmis, 2002a] uma teoria sobre outro aspecto do funcionamento do sistema imunológico: seu caráter de *rede regulada*, ou seja, de conjunto de células e moléculas *que se reconhecem mesmo na ausência de antígenos* [de Castro, 2001]. Trata-se de um comportamento diametralmente oposto àquele descrito pela teoria da seleção clonal, voltada para fenômenos decorrentes do contato com invasores. O que Jerne postula é que pode haver atividade mesmo sem que ela esteja voltada para o reconhecimento imediato de um antígeno. Na realidade, o conjunto de células de defesa presentes no organismo seria dotado de interação dinâmica, dela decorrendo mecanismos de *ativação e supressão*. Veremos que tais mecanismos serão muito importantes para o processo de otimização subjacente.<sup>125</sup>

Tendo apresentado os elementos e mecanismos essenciais a nossos esforços, estamos prontos a analisar a proposta de uma ferramenta de otimização, uma *rede imunológica artificial* deles decorrente. Trataremos disso imediatamente.

---

<sup>124</sup> Esse mecanismo de aprimoramento das defesas por sucessivos encontros com o “inimigo” traz à mente um artigo da antiga legislação espartana sobre a guerra: “A third ordinance (...) was, that they should not make war often, or long, with the same enemy, lest that they should train and instruct them in war, by habituating them to defend themselves.” in Plutarco, *The Life of Lycurgus*, Britannica Great Books, 1952.

<sup>125</sup> A explanação mostra que ocorre em nossos corpos uma espécie de *processo evolutivo em pequena escala*. Com isso, tudo o que discutimos sobre algoritmos genéticos é útil para a compreensão do sistema imunológico e da rede artificial que apresentaremos a seguir.

#### 5.4.2. A Rede Imunológica Artificial (opt-aiNet)

A rede imunológica artificial denominada *opt-aiNet*, proposta por de Castro e Timmis [de Castro e Timmis, 2002b], é uma ferramenta de otimização baseada na sinergia entre o princípio da seleção clonal e as idéias sobre redes imunológicas de Jerne. O emprego conjunto dessas idéias produz um método evolutivo capaz de realizar eficazmente busca global e busca local, sendo ainda dotado de mecanismos de controle do tamanho da população e de manutenção de diversidade. Nessa metáfora biológica, o conjunto de células de defesa corresponde à população e a função custo passa a fazer o papel de medida de afinidade entre antígeno e anticorpo<sup>126</sup>. Essa exatamente é a grandeza que deve ser maximizada para que haja uma atuação satisfatória do sistema imunológico, sendo o processo de otimização baseado nas idéias de clonagem, hipermutação, seleção e relação entre a rede de células vistas na seção anterior<sup>127</sup>. Adota-se uma codificação real<sup>128</sup>, em concordância com o discutido na seção 5.3.2.

A fase de inicialização, correspondente à geração de um conjunto de vetores reais aleatórios, contempla a etapa de produção celular na medula. O *ensemble* de anticorpos juvenis adentra então o laço principal, em que sofre, primeiramente, a ação do princípio da seleção clonal através de processos de clonagem, mutação proporcional à medida de *fitness* e seleção. É possível regular o número de repetições necessárias desse pequeno laço que emula a ação do princípio da seleção clonal medindo o grau de influência do processo sobre a variação da qualidade das soluções. Entra então em cena a noção de rede, por meio da análise de afinidade entre as células da população e a eliminação de indivíduos redundantes. Finalmente, introduz-se material novo num passo que pode ser visto como uma idéia inspirada pela edição de receptores. A Tab. 5.3 apresenta uma visão mais estruturada do processo.

Do ponto de vista de otimização, os cinco primeiros passos dessa tabela correspondem a um processo de *busca local*, o sexto e o sétimo passos formam uma etapa em que se avalia a convergência dessa busca, o oitavo passo tem por objetivo a eliminação de redundância e o décimo passo é responsável pela introdução de diversidade populacional. O critério de parada usual é a repetição do número de células de memória de uma iteração na seguinte, o que implica que a população já ocupou os pontos de ótimo local do espaço de busca. Em problemas complexos, pode ser interessante associar esse critério a um limite máximo de iterações, que evitaria uma demora excessiva e eventualmente desnecessária, ou mesmo a um patamar de *fitness* desejado<sup>129</sup>.

Comparando essa técnica aos esquemas genéticos apresentados anteriormente, vemos que a *opt-aiNet* é uma ferramenta evolutiva de busca dotada de um método refinado de busca local e de procedimentos capazes de alterar dinamicamente o tamanho da população e de manter em patamares apropriados o grau de diversidade, tendo,

---

<sup>126</sup> São válidas, portanto, as considerações sobre *fitness* presentes na seção 5.3.1. Comumente, seguindo a analogia com um AG, usaremos o termo "*fitness*" quando nos referirmos à afinidade.

<sup>127</sup> E análogas às noções apresentadas na seção 5.3.

<sup>128</sup> Note o leitor que, nesse caso, os vetores de parâmetros correspondem, em tese, à estrutura do receptor que caracteriza a célula de defesa.

<sup>129</sup> Considerações mais detalhadas sobre custo computacional e sensibilidade aos parâmetros envolvidos podem ser encontradas em [de Castro e Von Zuben, 2002b] e [de Castro e Von Zuben, 2000].

conseqüentemente, um potencial de busca global. O exemplo 5.2 ilustra o funcionamento desse algoritmo num problema de otimização multimodal.

### 1 – Inicialização

- Escolha os parâmetros do algoritmo e crie uma população inicial aleatória com um número pequeno de indivíduos.

### 2 – Processo Iterativo

Enquanto um critério de parada não for satisfeito,

- Calcule o *fitness* de todos os indivíduos da população.
- Produza um número  $N_c$  de cópias (clones) de cada indivíduo.
- Aplique um processo de mutação a cada um desses clones, mantendo, no entanto, o indivíduo original inalterado. Adote um esquema de mutação proporcional:

$$c' = c + \alpha N(0,1)$$

$$\alpha = (1/\beta)\exp(-f)$$

sendo  $c'$  o indivíduo modificado,  $N(0,1)$  uma variável aleatória gaussiana padrão com média nula e variância unitária,  $\beta$  um parâmetro de controle e  $f$  o valor de *fitness*, normalizado para que esteja no intervalo  $[0,1]$ . A mutação só será aceita caso  $c'$  esteja no domínio de validade das soluções.

- Determine o valor de *fitness* de todos os novos indivíduos.
- De cada grupo formado pelos clones e pelo indivíduo original, mantenha apenas o melhor indivíduo.
- Calcule o *fitness* médio da população.
- Se o *fitness* médio da população não for significativamente distinto daquele verificado na iteração anterior, então vá em frente. Caso contrário, retorne ao início do processo iterativo.
- Determine a afinidade (grau de similaridade quantificado pela distância Euclidiana) de todos os indivíduos da população. Mantenha apenas um dos indivíduos com afinidade abaixo de um limiar de supressão pré-determinado.
- Verifique o número de indivíduos restantes, que representam as células de memória.
- Introduza um número de novos indivíduos proporcional ao tamanho da população restante.

**Tabela 5.4:** Sumário do Funcionamento da opt-aiNet

---

### Exemplo 5.2

Seja a seguinte função custo que desejamos maximizar no espaço de busca formado pelo produto cartesiano entre dois intervalos unitários:

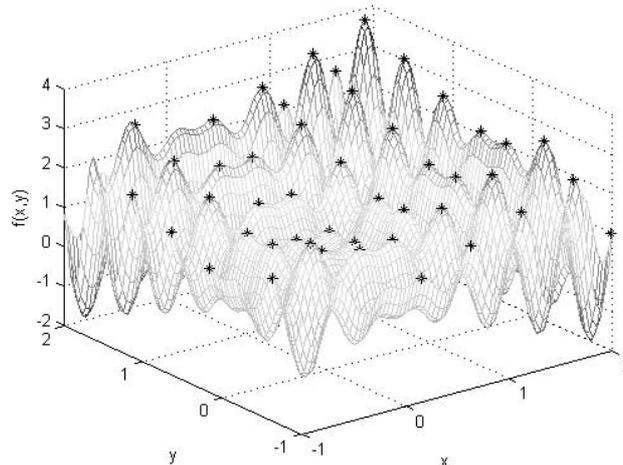
$$f(x, y) = x\sin(4\pi x) - y\sin(4\pi y + \pi) + 1 \quad (5.8)$$

Para atingir tal meta, adotemos a rede opt-aiNet descrita anteriormente, com os seguintes parâmetros:

Parâmetro	Valor
Número Inicial de Indivíduos	20
Número de Clones por Célula	10
$\beta$	100
Limiar de Supressão	0.1
Número Máximo de Iterações	5000

**Tabela 5.5:** Parâmetros da opt-aiNet

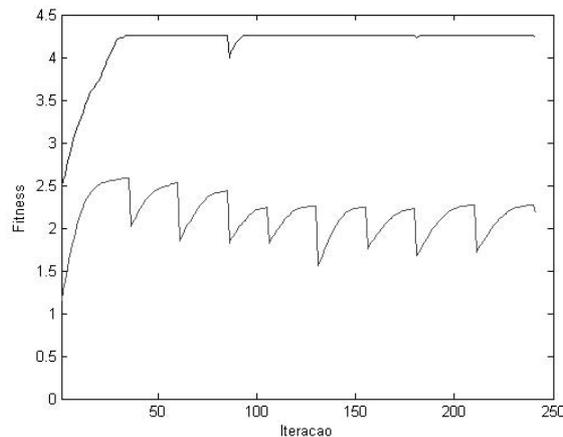
Em todas as realizações, a rede teve um desempenho excelente, tendendo a produzir uma população em que cada indivíduo se associa a um extremo local da função, ou seja, um caso em que a exploração é ideal e o emprego de recursos computacionais é mantido num patamar mínimo. A Fig. 5.5 mostra uma típica população final sobre o plano de fundo da função custo.



**Figura 5.5:** Função Custo e População Final

Para ilustrar o comportamento dinâmico da rede, apresentamos na Fig. 5.6 as curvas de *fitness* médio e *fitness* máximo da população. Verificamos que o patamar ótimo de *fitness* é atingido logo, havendo apenas uma pequena oscilação devida, provavelmente, à supressão de um dos indivíduos próximos ao ótimo global. O comportamento do *fitness* médio, por

sua vez, é muito ilustrativo: seu caráter oscilante mostra claramente a atuação do mecanismo de “poda” e “inserção” de indivíduos da rede imunológica, enquanto sua distância em relação à outra curva reflete a manutenção de diversidade já preconizada pela Fig. 5.5. Concluímos que o comportamento da rede é deveras satisfatório nesse complicado problema de otimização.



**Figura 5.6:** Evolução do Valor de *Fitness*

---

Não deve causar espanto a similaridade entre os principais elementos presentes na opt-aiNet e aqueles discutidos no âmbito da discussão acerca de algoritmos genéticos, pois a técnica que acabamos de apresentar emula um processo evolutivo em pequena escala. Uma discussão sobre a relação entre a opt-aiNet, AGs e outras ferramentas de otimização pode ser encontrada em [Attux et al., 2003a]. Vale ressaltar que há, no entanto, duas diferenças que tornam a opt-aiNet particularmente atraente em relação a um AG convencional: o operador de mutação proporcional e o processo de controle do tamanho da população. A existência desses dois mecanismos faz da rede imunológica um paradigma de busca baseada em computação evolutiva dotada tanto de robustez quando de um *modus operandi* que valoriza a parcimônia.

### **5.5. Computação Evolutiva como um Paradigma para Filtragem Ótima**

No início do capítulo, apresentamos três problemas essenciais no projeto de um filtro adaptativo de caráter genérico. Para que vejamos como as técnicas apresentadas neste capítulo os abordam, tratemos, em primeiro lugar, de recapitulá-los e, em seguida, de mostrar como as técnicas evolutivas trazem um ganho substancial em relação a um *framework* clássico:

- Realizar a busca de maneira eficiente do ponto de vista de convergência global.
- Usar uma técnica que não requeira o cálculo do gradiente, evitando assim manipulações proibitivas e aproximações.

- Empregar uma técnica que seja robusta à existência temporária de configurações instáveis.

O primeiro problema é contemplado pelo paradigma evolutivo, pois os esquemas de busca populacional, desde que haja mecanismos para manutenção de diversidade, são capazes de efetuar eficientemente a tarefa de busca global. No caso dos AGs, pode convir que seja empregada a idéia de nichos ou um esquema de seleção mais elaborado, enquanto que no caso da opt-aiNet o influxo de novos indivíduos mesclado aos esquemas de controle da população contribui decisivamente para a manutenção de diversidade.

O segundo problema é totalmente alheio aos esquemas evolutivos, pois, conceitualmente, não há qualquer base para a utilização do cálculo diferencial. Basta que seja possível calcular o custo associado às soluções que emergirem durante o processo de busca para que haja uma operação conveniente da técnica.

Por fim, o terceiro problema também não acarreta grandes problemas, pois soluções instáveis deverão ter um *fitness* muito baixo e, conseqüentemente, tenderão a desaparecer da população sem que interfiram na qualidade das boas soluções.

Esses argumentos nos levam a propor o uso de ferramentas evolutivas como parte de um paradigma para filtragem ótima, aliando generalidade na escolha da estrutura e do critério a uma elevada expectativa de desempenho. A partir do próximo capítulo, apresentaremos uma série de contribuições que, além de seus objetivos individuais, terão, quando tomadas em conjunto, um valor de plataforma de testes dessa concepção. Pedimos ao leitor que tenha em mente esses dois níveis de análise ao longo das páginas vindouras.

## PARTE II

### Contribuições e Resultados

*“Foi dito muitas vezes que era preciso procurar a verdade, mesmo quando nisso não se via utilidade; a utilidade, com efeito, pode não ser imediata e aparecer quando menos se espera”.*

*Arthur Schopenhauer, O Mundo como Vontade e Representação, I, 8<sup>130</sup>*

---

<sup>130</sup> Da versão publicada pela Editora Contraponto.



## 6. Prólogo à Segunda Parte

Na primeira parte, percorremos um caminho que teve por origem a idéia de comunicação e desembocou nas técnicas de otimização baseadas em computação evolutiva, bastante genéricas e dotadas de uma notável capacidade de busca. Ao longo do processo, delineou-se uma idéia fundamental: a possibilidade de considerar o emprego conjunto de uma estrutura de filtragem tão flexível quanto se deseje, de critérios cegos e supervisionados e de um método de busca global como um *paradigma eficiente para filtragem adaptativa*. Sob um paradigma desse tipo, o projetista teria total liberdade para lançar as bases estruturais do processamento, havendo uma chance significativa de que o método de busca escolhesse um conjunto de parâmetros que levasse o filtro adotado ao seu limite de desempenho. Além disso, a eficiência da ferramenta de otimização poderia tornar ainda mais concreta a possibilidade de dispensar qualquer tipo de informação sobre o sinal piloto, lidando com as dificuldades decorrentes de um critério cego.

Norteados por essa idéia, chegamos a um conjunto de abordagens originais dos três problemas centrais de nossa pesquisa: equalização, identificação e predição. Todas as propostas que buscamos construir têm por objetivo tornar mais eficientes os processos de filtragem e reconstrução de informação, abrindo novas perspectivas concretas para o uso de estruturas não-lineares, estruturas recursivas e critérios não-supervisionados. Naturalmente, o preço dessa melhoria de desempenho é um significativo aumento no custo computacional do processo de filtragem adaptativa, causado tanto pelo maior requinte estrutural quanto pela complexidade dos métodos de busca. Dilemas dessa sorte, como discutimos ao longo dos capítulos anteriores, são inevitáveis: importa-nos sempre buscar novos territórios no que diz respeito à eficiência e ao pleno cumprimento dos objetivos do processo de filtragem, cabendo a um eventual passo futuro a tarefa de melhorar a eficiência da metodologia exposta ante os recursos de implementação disponíveis. Mais uma vez, é papel do projetista optar: tornaremos sua decisão mais representativa aumentando o alcance de seu repertório de soluções eficientes.

Nos próximos capítulos, apresentaremos um conjunto de contribuições que abrangem problemas de equalização, identificação, predição, busca da seqüência de máxima verossimilhança e pré-distorção. Vistas separadamente, as diferentes propostas que apresentaremos nesta segunda parte são esforços no sentido de aprimorar ou abrir novos caminhos em diversos processos de filtragem e reconstrução de sinais. Vistas como um todo, as abordagens formam um *corpus* de resultados e análises que atestam a relevância do paradigma geral de filtragem que permeou toda a feitura deste documento.



## 7. Equalização Linear Não-supervisionada e Computação Evolutiva

O problema de equalização linear cega, conforme discutimos no capítulo 3, é, ao mesmo tempo, relevante e dotado de uma complexidade que, do ponto de vista da teoria de sistemas dinâmicos, é bastante desafiadora [Attux e Romano, 2003, Frater et al., 1995]. No cerne das características que tornam a teoria de equalização cega tão diferente da teoria clássica, jaz a presença de estatísticas de ordem superior, cuja estrutura intrínseca tende a gerar funções custo dotadas de múltiplos pontos de gradiente nulo e, pensando em técnicas de otimização baseadas em derivadas, algoritmos adaptativos que podem convergir para ótimos locais ou mesmo operar em regimes mais inusitados (ciclos-limite, caos etc.) [Attux, 2001].

Em conexão com as idéias apresentadas na seção 3.4.3, é possível afirmar que o problema de equalização linear cega apresenta uma característica essencial: independentemente do critério escolhido, *a função custo terá múltiplos mínimos*. Conseqüentemente, o emprego de um algoritmo como os de gradiente levará, dependendo da inicialização adotada, o estado do filtro a se deter numa configuração que não é a melhor possível, o que nos remete a uma conclusão imediata: a abordagem clássica não oferece *a priori* nenhum tipo de garantia quanto à meticulosa exploração das potencialidades da estrutura empregada.

Apesar do cenário que acabamos de delinear, não devemos chegar à tradicional conclusão de que os critérios cegos são “males necessários”: como mostramos na seção 3.4.2, o critério de Wiener também será “multimodal” se o atraso de equalização for uma variável livre. Mais ainda, de acordo com a conjectura exposta na seção 3.4.3, o critério CM é capaz de “preservar os bons atrasos”, ou seja, de ter como mínimos apenas pontos próximos a soluções de Wiener de “olho aberto”<sup>131</sup>. Curiosamente, chegamos a uma situação em que o critério cego *pode ser uma versão mais enxuta e eficiente de um critério supervisionado*. Em outras palavras, nas situações em que valer a conjectura, o critério CM dará origem a uma função custo que, de certa maneira, *escolhe eficientemente o atraso de equalização*. Caso dispuséssemos de um satisfatório método de busca global, *poderíamos obter a melhor solução de Wiener existente sem estimar ou conhecer o sinal desejado*. Sob essa ótica, é inegável que a união do critério CM a uma ferramenta eficiente de otimização dá origem a um paradigma muito sólido de filtragem linear: em tese, num caso ideal, teríamos em mãos um método para obter a melhor configuração possível *sem que houvesse necessidade de dispor de um sinal piloto*. É exatamente essa tarefa que procuraremos cumprir neste capítulo.

Escolhemos duas técnicas evolutivas para o papel de método de busca global: um *algoritmo genético com niching* e a *rede imunológica opt-aiNet*. Seguindo uma ordem natural, discutiremos, em primeiro lugar, os resultados obtidos com o AG, passando, em seguida, àqueles obtidos por meio da rede. Uma breve discussão concluirá o capítulo.

---

<sup>131</sup> Assumindo, como de costume, que o canal é um filtro FIR.

## 7.1. Primeira Abordagem: Algoritmo Genético com Niching

Tendo em vista o desafio de minimizar a função CM, apresentada em (3.42), optamos primeiramente por uma solução evolutiva mais usual: um algoritmo genético [Costa et al., 2002]. Além de nossa avaliação do potencial dessa ferramenta, tínhamos como fatores de motivação trabalhos que indicavam possíveis ganhos de desempenho por parte de um AG em problemas de filtragem supervisionada, como [Neubauer, 1995], e um trabalho que indicava a possibilidade de escape de mínimos locais da função CM através de um AG simples com codificação binária [Cardoso et al., 2000].

Analisando as características do problema, bem como a extensão e generalidade de nossas metas, optamos por um AG com codificação real, seleção por meio do algoritmo da roleta, *crossover* aritmético e mutação real. Buscando um compromisso entre simplicidade e desempenho, optamos por realizar, em cada operação de *crossover*, um número fixo de cruzamentos entre indivíduos de acordo com (5.2-5.3), havendo, em cada caso, uma competição entre pais e filhos pelo direito de permanecer na população. Visando tornar o processo mais flexível, decidimos sortear o parâmetro de *crossover* aritmético  $\alpha$  a partir de uma distribuição uniforme entre  $-0.3$  e  $1.3$ . A mutação real, por sua vez, teve por base uma variável uniforme entre  $-0.4$  e  $0.4$ .

Nossos ensaios preliminares mostraram que era possível que houvesse perda de diversidade nesse cenário, fato que nos levou a adotar uma competição em nichos ecológicos baseada no método de *clearing* [Sareni e Krähenbühl, 1998]. A regra de formação dos nichos foi muito simples: tomar por base a *posição do coeficiente de maior magnitude do vetor de parâmetros*<sup>132</sup>. Adotamos uma estratégia elitista, preservando apenas o melhor indivíduo de cada nicho ecológico. Como consequência direta da definição dos nichos, seu número total foi sempre igual ao número de coeficientes do vetor de parâmetros.

Para testar o AG, assumimos que as amostras do sinal transmitido são i.i.d. e que pertencem a uma constelação binária 2-PAM ou 2-PSK de módulo unitário. Adotamos dois canais de teste, ambos de fase não-mínima, e supomos que não há ruído aditivo. A justificativa de nossa opção é que importa inicialmente concentrar nossos esforços analíticos no problema da busca num cenário de múltiplos mínimos, cenário este devido à presença de IIS e não de ruído, que consideraremos apenas como um fator de perturbação. As funções de transferência dos canais escolhidos são:

---

<sup>132</sup> A idéia de associar essa característica à bacia de atração dos diferentes mínimos do critério CM pode ser relacionada à abordagem geométrica de análise adotada em [Li et al., 1995] (vide, por exemplo, [Attux, 2001]). Entretanto, também faz sentido considerar um caso em que o equalizador tenha uma ordem muito elevada, para a qual seja possível atingir uma solução próxima à solução ZF: nesse caso, o deslocamento entre soluções correspondentes a atrasos  $n_d$  e  $n_d+1$  será, *aproximadamente*, um deslocamento do tipo *shift right* no vetor de parâmetros do equalizador. Tal deslocamento também se dará na posição do coeficiente de maior magnitude, o que justifica nossa escolha.

$$\begin{cases} \text{C1: } H_1(z) = 1 + 0.4z^{-1} + 0.9z^{-2} + 1.4z^{-3} \\ \text{C2: } H_2(z) = 1 + 1.2z^{-1} - 0.3z^{-2} + 0.8z^{-3} \end{cases} \quad (7.1)$$

Conforme podemos intuir da discussão preliminar, a função de *fitness* adotada é uma estimativa da função CM<sup>133</sup> transformada segundo (5.1). Devido à diferença usual entre as soluções CM e as de Wiener, deve o leitor saber que, em todos os ensaios deste capítulo, o melhor indivíduo obtido foi sempre refinado por meio de umas poucas iterações do algoritmo de decisão-direta (vide seção 3.4.3).

Tendo já descrito o processo inteiro, apresentamos na Tab. 7.1 os parâmetros comuns aos ensaios envolvendo o AG. Isso conclui a parte geral de nossa exposição e nos permite passar aos ensaios.

Parâmetro	Valor
Tamanho da População	30
Número de <i>Crossovers</i> por Iteração	10
Probabilidade de Mutação	0.1
Critério de Parada	2000 iterações

**Tabela 7.1:** Parâmetros Gerais do AG

### 7.1.1. Primeiro Ensaio

No primeiro ensaio, escolhemos o canal C1 e um equalizador FIR com cinco coeficientes. Efetuamos 50 realizações do AG e adotamos, como mostrado na Tab. 7.1, um critério de parada baseado no número de iterações. Em todos os casos, a melhor solução obtida<sup>134</sup> foi justamente a melhor solução de Wiener, conforme mostra a Tab. 7.2<sup>135</sup>.

Solução	EQM Residual	Frequência
$w_{opt}$	0.1751	100%
<b>EQM Residual Médio: 0.1751</b>		

**Tabela 7.2:** Resultados do Primeiro Ensaio

O erro quadrático médio residual da melhor solução pode ser considerado relativamente alto, o que mostra que, embora o olho tenha sido aberto, a ordem do equalizador não propicia uma grande margem em relação às necessidades impostas pelo canal. Por esse motivo, ela pode ser considerada baixa.

<sup>133</sup> Essa estimativa será onipresente, pois as funções custo dos diferentes critérios de equalização dependem da aplicação do operador esperança, que, na prática, precisa ser substituído por uma média temporal. Quanto mais refinada for a estimativa, mais custoso será o método de busca dela dependente.

<sup>134</sup> Para prevenir quaisquer mal-entendidos, frisamos que nosso objetivo é *obter a melhor solução de Wiener através da otimização do critério CM*.

<sup>135</sup> Em todas as tabelas de resultado deste capítulo, adotaremos a notação  $w_k$  para designar a k-ésima melhor solução de Wiener.

### 7.1.2. Segundo Ensaio

Devido à significativa parcimônia estrutural presente no primeiro ensaio, optamos por manter, em nosso segundo esforço, o mesmo canal, tratando, no entanto, de aumentar o número de coeficientes do equalizador para oito. Essa medida deverá ter duas conseqüências: o melhor mínimo tenderá a ter um EQM residual bem menor, e, além disso, os melhores ótimos locais tenderão a ter um desempenho mais próximo ao da melhor solução. Isso significa que deverá existir uma menor discrepância entre o desempenho dos mínimos, o que torna a tarefa do AG mais difícil em termos de convergência global.

Mais uma vez, efetuamos 50 realizações do AG, sendo a freqüência de obtenção das diversas soluções de Wiener mostrada na Tab. 7.3.

<b>Solução</b>	<b>EQM Residual</b>	<b>Freqüência</b>
$w_{opt}$	0.1293	48%
$w_2$	0.1397	22%
$w_3$	0.1445	12%
$w_4$	0.1533	10%
$w_5$	0.1890	4%
$w_6$	0.1951	4%
<b>EQM Residual Médio: 0.1408</b>		

**Tabela 7.3:** Resultados do Segundo Ensaio

Verificamos que a melhor solução de Wiener foi obtida em, aproximadamente, metade dos casos. É importante notar que o mínimo local mais freqüente possui um EQM residual muito semelhante ao da melhor solução, tendo os outros dois mínimos mais freqüentes um desempenho próximo ao ótimo. O EQM residual médio das 50 soluções obtidas foi 0.1408, o que mostra que o desempenho do AG é, na média, bastante adequado.

### 7.1.3. Terceiro Ensaio

No ensaio final, adotamos o canal C2 e um equalizador com sete coeficientes. As freqüências relativas às 50 realizações se encontram na Tab. 7.4.

<b>Solução</b>	<b>EQM Residual</b>	<b>Freqüência</b>
$w_{opt}$	0.0312	48%
$w_1$	0.0458	40%
$w_2$	0.0917	8%
$w_3$	0.0918	2%
$w_4$	0.1022	2%
<b>EQM Residual Médio: 0.0445</b>		

**Tabela 7.4:** Resultados do Terceiro Ensaio

Nesse caso, temos um cenário intermediário: há uma semelhança de desempenho entre mínimos menor que a do segundo ensaio, mas maior que aquela exposta no primeiro ensaio. O melhor mínimo é atingido em aproximadamente metade dos casos e o segundo melhor mínimo, cujo desempenho é muito próximo ao do ótimo global, prevalece em 40% das realizações. As demais soluções, de desempenho mais distante do ótimo, têm uma frequência significativamente menor, o que atesta a coerência do processo de busca e da interpretação que demos a seu desempenho. O EQM médio do ensaio foi de 0.0445, um patamar muito próximo ao do mínimo global. Com isso, concluímos que o desempenho do AG foi bastante satisfatório em termos de comportamento médio e significativo no que diz respeito à taxa de convergência global.

#### 7.1.4. Discussão

Nesse primeiro conjunto de testes, demonstrou-se o potencial da abordagem evolutiva em três cenários distintos, todos vinculados a canais de fase não-mínima<sup>136</sup>. No primeiro ensaio, o AG levou a um resultado notável: em todas as realizações foi obtido o melhor mínimo da função CM, o que significa que foi possível alcançar a melhor solução de Wiener em todos os casos sem que houvesse necessidade de supervisão. No segundo ensaio, não foi possível obter 100% de convergência global, fato que atribuímos a um fator principal: a semelhança entre os erros residuais dos melhores mínimos. De qualquer forma, o desempenho do AG foi bastante coerente, nitidamente privilegiando as melhores soluções, o que muito provavelmente não se verificaria, por exemplo, num método de inicialização aleatória do CMA. A solidez do processo de busca também é atestada pelo erro residual médio. O comportamento verificado no terceiro ensaio guardou semelhanças com aquele mostrado no segundo, estando as estimativas concentradas no par de melhores soluções.

Concluímos que o emprego do AG com *niching* atesta a validade do uso de uma técnica evolutiva na obtenção da melhor solução de Wiener de maneira não-supervisionada, representando uma primeira evidência em favor da viabilidade do paradigma delineado no prólogo a esta segunda parte. Para concluir, frisamos que é bem provável que os resultados possam ser melhorados caso se testem diferentes soluções para os múltiplos aspectos de projeto de um algoritmo genético (vide seção 5.3). Entretanto, julgamo-nos satisfeitos com o compromisso entre complexidade e desempenho estabelecido pelo AG adotado: parecemos mais conveniente voltar nossa atenção para uma outra estratégia de busca.

## 7.2. Segunda Abordagem: Rede Imunológica

Embora a ferramenta genética tenha se comportado bem no teste, julgamos muito importante que também fosse avaliado o desempenho da outra técnica de busca considerada neste trabalho: a rede imunológica artificial opt-aiNet, apresentada na seção 5.4.2. A idéia é avaliar o desempenho de tal algoritmo, especialmente em termos de capacidade de busca global, em cenários semelhantes àqueles mostrados na seção anterior. Isso nos permitirá

---

<sup>136</sup> Conforme discutimos na seção 3.6, canais de fase não-mínima não podem ser equalizados por um filtro de erro de predição linear, o que faz deles bases interessantes para teste de técnicas cegas.

avaliar o desempenho da ferramenta imunológica e buscar elementos de uma comparação entre tal técnica e o AG com *niching*. A existência do termo de comparação com o AG também é importante por ser esta a primeira aplicação de que temos notícia de uma técnica imunológica a um problema de filtragem pertencente ao escopo delimitado no capítulo 3 [Attux et al., 2003a].

Consideramos novamente o problema de obter a melhor solução de Wiener a partir da função CM estimada e transformada por (5.1). Assumimos, mais uma vez, que o sinal transmitido é formado por amostras i.i.d. extraídas de uma constelação 2-PAM ou 2-PSK de módulo unitário, bem como que o equalizador é um filtro FIR. Consideraremos agora três modelos de canal FIR, sendo, mais uma vez, descartada a existência de ruído aditivo:

$$\begin{cases} \text{C1} : H_1(z) = 1 + 0.4z^{-1} + 0.9z^{-2} + 1.4z^{-3} \\ \text{C2} : H_2(z) = 1 + 1.2z^{-1} - 0.3z^{-2} + 0.8z^{-3} \\ \text{C3} : H_3(z) = 1 + 0.6z^{-1} - 0.7z^{-2} + 2.5z^{-3} \end{cases} \quad (7.2)$$

Propositadamente, os dois primeiros canais mostrados em (7.2) são exatamente os dois canais que serviram de base para os três ensaios da seção anterior.

Também nesse caso, a grandeza que mais nos interessa é a taxa de convergência global. Buscamos uma configuração que tendesse a levar a população a um patamar parcimonioso, mas não nos deteremos demasiadamente na análise desse aspecto do processo de busca<sup>137</sup>. Os parâmetros adotados como padrão nesse estudo se encontram na Tab. 7.5.

Parâmetro	Valor
População Inicial	5
$\sigma_s$	0.35
Número de Clones ( $N_C$ )	10
$\beta$	50
Número Máximo de Iterações	1000
Número de Realizações por Ensaio	100

**Tabela 7.5:** Parâmetros Gerais da opt-aiNet

### 7.2.1. Primeiro Ensaio

No primeiro ensaio, escolhemos o canal C1 e um equalizador com oito coeficientes, o que nos fornece uma possibilidade concreta de estabelecer paralelos com os resultados mostrados na seção 7.1.2. Na Tab. 7.6, apresentamos as frequências de convergência para a melhor solução obtidas pela rede imunológica juntamente com as fornecidas pelo AG.

<sup>137</sup> Também julgamos interessante, nesse caso, fazer o mínimo de modificações na transição entre ensaios.

Solução	EQM	Freq. (GA + <i>niching</i> )	Freq. (opt-aiNet)
$w_{opt}$	0.1293	48 %	82%
$w_2$	0.1397	22 %	17%
$w_3$	0.1445	12 %	1 %
$w_4$	0.1533	10 %	-
$w_5$	0.1890	4 %	-
$w_6$	0.1951	4 %	-

**Tabela 7.6:** Resultados do Primeiro Ensaio

É nítido que a opt-aiNet produz um significativo ganho de desempenho em relação ao AG, o que confirma seu potencial de busca: ocorre um aumento notável na taxa de convergência global, e, além disso, verifica-se uma maior concentração de ocorrências nos dois melhores mínimos. Esse salto de desempenho também pode ser percebido se compararmos o erro residual médio da opt-aiNet, que foi de 0.1312, com o erro residual médio associado ao AG, que foi de 0.1408.

### 7.2.2. Segundo Ensaio

No segundo ensaio, escolhemos a combinação entre o canal C2 e um equalizador com sete coeficientes. Tais escolhas nos remetem ao ensaio da seção 7.1.3. Na Tab. 7.7, apresentamos as frequências obtidas através das duas técnicas consideradas.

Solução	EQM	Freq. (GA+ <i>niching</i> )	Freq. (opt-aiNet)
$w_{opt}$	0.0312	48 %	100 %
$w_2$	0.0458	40 %	-
$w_3$	0.0917	8 %	-
$w_4$	0.0918	2 %	-
$w_5$	0.1022	2 %	-

**Tabela 7.7:** Resultados do Segundo Ensaio

Nesse caso, o desempenho da rede imunológica foi irrepreensível, o que atesta sua capacidade de busca global superior à do AG. Em relação à tabela anterior, atribuímos uma vez mais a melhoria de desempenho à maior diferença de qualidade entre os mínimos, que favorece uma mais nítida diferenciação de desempenho por parte do algoritmo de busca. A destacada atuação da rede imunológica também pode ser confirmada através da

comparação entre o erro quadrático residual associado a ela, que foi de 0.0312, com o erro apresentado na Tab. 7.4, que foi de 0.0445.

### 7.2.3. Terceiro Ensaio

No terceiro ensaio, não buscamos comparações com o AG mostrado anteriormente, sendo nossa intenção submeter a rede a um teste interessante: uma configuração que leve a uma proximidade de desempenho ainda maior entre os melhores mínimos. Para atingir tal objetivo, escolhemos o canal C3 e um equalizador com doze coeficientes. O tamanho do equalizador tende a gerar uma condição bastante favorável do ponto de vista da capacidade de inversão do canal (paradigma ZF), fato que deve se refletir na existência de diversos mínimos com desempenho semelhante. Na Tab. 7.8, apresentamos as frequências obtidas em tal cenário.

Solução	EQM	Freq. (opt-aiNet)
$w_{opt}$	0.0071	66 %
$w_2$	0.0075	32 %
$w_3$	0.0104	2 %

**Tabela 7.8:** Resultados do Terceiro Ensaio

Em concordância com nossas expectativas, há uma marcante semelhança entre os erros residuais das três melhores soluções, o que explica o fato de a taxa de convergência global ser menor aqui que nos demais ensaios. Apesar disso, o erro residual médio de 0.0073 mostra que o resultado pode ser considerado bastante satisfatório.

Assumindo que a diminuição da taxa de convergência global se deveu à proximidade entre os melhores mínimos, optamos por observar o comportamento da rede com  $\beta = 100$ , ou seja, numa situação em que a busca local é mais precisa devido à menor amplitude das mutações. Para essa configuração, foram obtidas as frequências mostradas na Tab. 7.9.

Solução	EQM	Freq.
$w_{opt}$	0.0071	84 %
$w_1$	0.0075	16 %

**Tabela 7.9:** Resultados do Terceiro Ensaio ( $\beta = 100$ )

Conforme esperávamos, a medida surtiu efeito: houve um significativo aumento na taxa de convergência global, bem como na frequência de obtenção dos melhores mínimos. O erro residual médio, nesse caso, foi de 0.0072.

### 7.2.4. Análise da Diversidade Populacional

Para buscar entender por que a rede imunológica levou a resultados tão significativos em termos de capacidade de busca global, analisemos uma realização do algoritmo sob o canal C1 e um modelo de equalizador com oito coeficientes. Na Tab. 7.10, apresentamos a população final da rede, juntamente com uma coluna que mostra a solução de Wiener mais próxima de cada indivíduo.

Indivíduo	Próximo a
[0.1740 -0.1297 -0.0852 -0.1882 0.4516 0.1137 -0.1007 0.0860]	$W_{opt}$
[-0.1805 0.1303 0.0967 0.1862 -0.4559 -0.1021 0.0949 -0.0781]	$W_{opt}$
[0.1113 -0.0377 0.0297 -0.2622 0.2195 0.0414 0.2687 -0.5315]	$W_2$
[-0.1041 0.0394 -0.0205 0.2460 -0.2160 -0.0475 -0.2449 0.5218]	$W_2$
[-0.0094 0.2060 -0.1512 -0.0991 -0.2261 0.4798 0.1023 -0.0547]	$W_3$
[0.0045 -0.2133 0.1477 0.1002 0.2389 -0.4794 -0.1108 0.0527]	$W_3$
[0.0025 0.0025 -0.2134 0.1518 0.0773 0.2414 -0.4720 -0.0815]	$W_4$
[-0.0019 0.0036 0.2117 -0.1416 -0.0747 -0.2416 0.4692 0.0835]	$W_4$
[-0.0795 -0.0877 -0.1181 0.3541 0.1515 -0.1279 0.1087 -0.0761]	$W_6$
[0.0824 0.0959 0.1175 -0.3742 -0.1458 0.1282 -0.1022 0.0508]	$W_6$
[-0.1197 -0.1265 0.3463 0.1777 -0.1494 0.1260 -0.1174 0.0538]	$W_5$
[0.1355 0.1127 -0.3420 -0.1769 0.1485 -0.1198 0.1248 -0.0601]	$W_5$
[-0.1768 0.3269 0.1402 -0.0967 0.1024 -0.1249 0.0497 0.0252]	$W_7$

**Tabela 7.10:** Análise da População Final

Uma breve inspeção nos mostra que a população contém indivíduos na vizinhança de *sete soluções de Wiener*, o que atesta que a diversidade populacional se manteve num patamar

bastante significativo. Tal estado de coisas tem como reflexo natural as excelentes taxas de convergência global por nós verificadas em todos os ensaios.

### 7.2.5. Discussão

A idéia que é a razão de ser desse capítulo, a saber, buscar a melhor solução de Wiener por meio do critério do módulo constante, teve sua validade confirmada pelos ensaios realizados com os dois métodos de busca. Tanto o AG quanto a rede imunológica se mostraram ferramentas de otimização dotadas de significativa capacidade de busca, fato atestado tanto pelas taxas de convergência global quanto pela média do erro quadrático residual.

Numa breve comparação, verificamos que a rede imunológica apresentou um desempenho melhor que o obtido através do algoritmo genético. No entanto, alertamos o leitor para que afaste de si qualquer conclusão do tipo “a opt-aiNet é uma técnica superior a um AG”: só podemos esboçar um paralelo entre *a AG por nós elaborado e a rede imunológica com a configuração escolhida*. Mesmo assim, não desejamos nos deter nesse assunto mais que o necessário, pois não é nossa intenção estabelecer uma comparação rigorosa entre técnicas, bastando para nossos fins apontar algumas tendências de comportamento nos problemas de interesse. Notamos que a principal dificuldade encontrada por ambas as técnicas foi produzir taxas de convergência global na vizinhança de 100% em cenários com múltiplos mínimos de desempenho equivalente. Há duas razões adicionais para isso:

- Em primeiro lugar, o critério de parada adotado se baseou num número máximo de iterações. Isso significa que, eventualmente, o processo de busca poderia ser interrompido antes da plena utilização do potencial da técnica e da configuração dos parâmetros. Em outras palavras, com mais tempo para operar, diminui a chance de que uma cadeia de “ligeiros infortúnios” aleatórios termine por deixar incompleta a varredura da vizinhança de um determinado máximo local. Mostramos na seção 7.2.3 como a modificação da taxa de mutação foi capaz de atenuar esse problema.
- A função custo, como mencionamos anteriormente, é sempre *estimada*. Isso quer dizer que há um determinado nível de incerteza na avaliação do *fitness* dos indivíduos, nível este tanto menor quanto maior for a janela de dados empregada. Estabelece-se, portanto, um compromisso entre precisão e custo computacional, uma vez que o cálculo de *fitness* é um fator que influencia a quantidade de operações de cada iteração do processo.

Pensando em termos dos elementos do paradigma apresentado no prólogo, vimos como o emprego da ferramenta evolutiva leva a um significativo grau de aproveitamento do potencial de filtragem da estrutura FIR escolhida. Nesse caso, a eficiência do método de busca, aliada à solidez do critério CM, permitem que o pleno aproveitamento se dê sem que seja necessário contar com um sinal piloto.

## 8. Equalização Recursiva Não-supervisionada e Computação Evolutiva

No capítulo anterior, verificamos que o uso de uma técnica evolutiva pode trazer um significativo ganho de desempenho no projeto não-supervisionado de um equalizador linear, por ser capaz de lidar de maneira efetiva com o problema da existência de múltiplos mínimos na função custo CM. Neste capítulo, voltaremos a considerar o problema de equalização cega, porém, com uma fonte de complexidade adicional: a presença de realimentação na estrutura de filtragem, ou seja, de uma *recursão*. Numa situação desse tipo, conforme discutimos nas seções 3.2.1 e 4.6 e no capítulo 6, além do problema da existência de mínimos locais, surgem dificuldades de manipulação da função custo e também emerge a ameaça da instabilidade.

Esse cenário bastante desafiador é um campo excelente para a aplicação de técnicas de busca como as discutidas no capítulo 5: além de possuírem um excelente potencial de busca, patente, aliás, no capítulo anterior, elas não requerem nenhum tipo de operação sobre a medida de *fitness* e, devido ao emprego de uma população de soluções, são mais robustas à degradação de desempenho produzida pela existência de configurações instáveis. Por conseguinte, valoriza-se enormemente a sinergia entre estrutura, critério e técnica de busca, confluindo todos elementos para a confecção de um dispositivo eficiente e plenamente explorado.

Apresentaremos neste capítulo duas propostas diretamente ligadas a esse espírito, voltadas à aplicação de duas estruturas recorrentes de filtragem, o filtro IIR e o DFE, ao complexo problema de equalização cega. Buscaremos mostrar que o uso da rede opt-aiNet pode produzir um significativo ganho de desempenho em relação às técnicas clássicas comumente empregadas. A comprovação de tal fato seria, ousaríamos dizer, uma evidência de quão vantajoso pode ser uma mudança de paradigma no treinamento de filtros recursivos.

### 8.1. Equalização IIR Cega

Conforme discutimos na seção 3.2.1, um filtro IIR tem uma diferença essencial em relação ao equalizador linear FIR clássico: sua resposta também pode ser moldada pelo posicionamento de pólos, ou seja, pelo controle de estímulos provenientes de um laço de realimentação. Há duas vantagens principais que decorrem dessa característica:

- Em tese, um filtro IIR é capaz de inverter um canal FIR.
- Mesmo quando não for possível inverter o canal, o filtro IIR pode atingir um patamar de desempenho equivalente ao de um equalizador FIR com um número menor de parâmetros.

Além dessas duas vantagens, há um fato adicional que justifica o emprego da estrutura recursiva em problemas de equalização: *o filtro IIR é uma síntese das potencialidades dos sistemas lineares como um todo* [Oppenheim et al., 1997].

Tendo por base idéias modernas sobre o critério do módulo constante [Johnson et al., 1998] e também na experiência adquirida com seu emprego [Attux, 2001], optamos por combinar a estrutura IIR, o critério CM e a rede opt-aiNet descrita na seção 5.4.2 numa série de testes que revelará se esses entes formam, juntos, uma abordagem inovadora e robusta. No que diz respeito ao critério, conjecturamos aqui, em consonância com a hipótese de vanguarda na literatura [Johnson et al., 1998, Suyama et al., 2003a, Zeng et al., 1999], que, assim como no caso da aplicação de filtros FIR, haverá uma estreita relação entre bons mínimos da função custo CM e boas soluções de Wiener (vide seção 3.4.3). Essa conjectura é o elo de que precisamos para relacionar o projeto do equalizador cego ao máximo nível de desempenho que se pode esperar da estrutura da filtragem [Attux et al., 2003b].

Para testar a validade da proposta, empregaremos três canais com características distintas e, em todos os casos, adotaremos um sinal transmitido formado por amostras i.i.d. pertencentes a uma constelação 2-PAM (ou 2-PSK) de módulo unitário. Empregaremos um critério de parada baseado num nível de *fitness* a ser atingido pelo melhor indivíduo da população, escolha que, por diferir daquela apresentada no capítulo anterior, serve para ampliar nosso espectro de análises da ferramenta evolutiva. Finalmente, os dois vetores de parâmetros do filtro IIR, correspondentes aos elementos de seu numerador e de seu denominador, foram codificados num único grande vetor a ser escolhido pelo algoritmo.

### 8.1.1. Primeiro Ensaio

No primeiro ensaio, escolhemos o seguinte modelo de canal, cujos pólos e zeros estão no círculo unitário, e assumimos que não existe ruído aditivo:

$$H_1(z) = \frac{1 + 0.6z^{-1} + 0.18z^{-2} + 0.504z^{-3} + 0.1856z^{-4}}{1 - 0.6z^{-1} + 0.25z^{-2}} \quad (8.1)$$

O fato de o canal ser de fase mínima garante que o equalizador pode invertê-lo perfeitamente, pois, para isso, precisará apenas de pólos contidos no CRU. Para que essa condição ideal possa ser satisfeita, escolhemos um filtro com uma característica exatamente inversa à do canal: cinco parâmetros no denominador de sua função de transferência e três no numerador. A existência de uma solução ideal é muito útil do ponto de vista analítico, pois ela será a solução ótima de qualquer critério coerente, seja ele supervisionado ou cego.

Assim como no capítulo 7, a função custo escolhida foi uma estimativa da função CM mostrada em (3.42), devidamente transformada segundo (5.1). Os parâmetros da opt-aiNet, escolhidos heurísticamente, encontram-se na Tab. 8.1<sup>138</sup>.

---

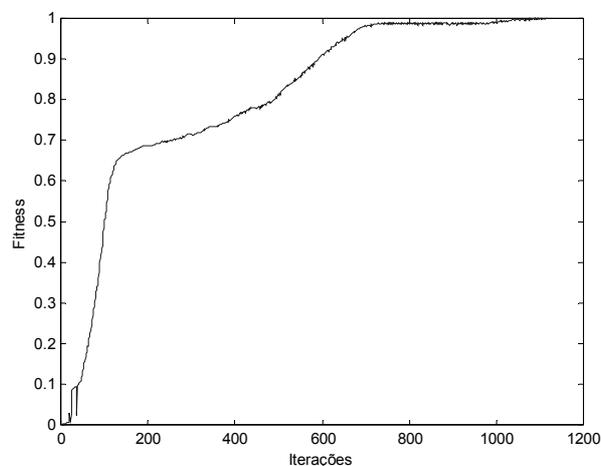
<sup>138</sup> Nesse caso, devido à existência de uma configuração ideal, não julgamos necessário adotar qualquer tipo de refinamento baseado no algoritmo DD.

Parâmetro	Valor
População Inicial	10
$\sigma_s$	2.5
Número de Clones ( $N_C$ )	10
$\beta$	100
<i>Fitness</i> Exigido para Parada	0.9995
Número de Realizações por Ensaio	12

**Tabela 8.1:** Parâmetros – Primeiro Ensaio

Em todas as realizações, a melhor solução obtida correspondeu à solução ideal, ou seja, à configuração inversa ao modelo de canal. Essa é exatamente a situação que buscávamos, conforme atesta a escolha do valor de *fitness* que compõe o critério de parada, muito próximo ao ideal (=1). O número médio de iterações para que fosse atingido o critério foi de 1172, o que mostra que a exigência de desempenho ótimo não leva a um tempo de convergência excessivamente elevado, mesmo para os padrões clássicos.

Para ilustrar o comportamento dinâmico da rede, apresentamos na Fig. 8.1 a evolução do valor do *fitness* associado ao melhor indivíduo da população para uma realização típica<sup>139</sup>.



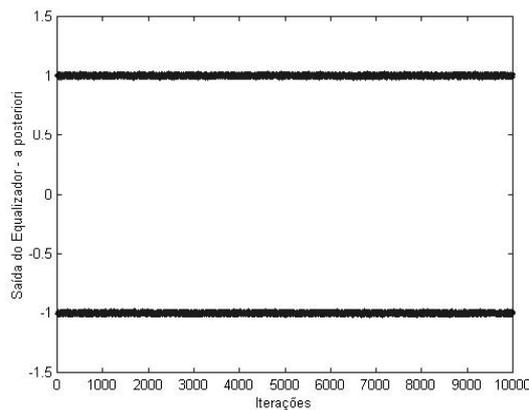
**Figura 8.1:** Evolução Temporal do *Fitness* Máximo – Primeiro Ensaio

Verificamos que a convergência se dá num ritmo variável, sendo bem nítidas as fases iniciais de exploração (rápido crescimento) e refinamento (variações mais lentas). Tal comportamento é bastante normal e coerente com o que se espera do processo de busca. A melhor solução encontrada ao final desse processo foi:

<sup>139</sup> Para nós, um ensaio típico tem um número de iterações próximo ao número médio verificado para o conjunto.

$$W(z) = \frac{1.0021 - 0.6209z^{-1} + 0.2523z^{-2}}{1 + 0.5805z^{-1} + 0.1614z^{-2} + 0.4932z^{-3} + 0.1786z^{-4}} \quad (8.2)$$

A proximidade dessa solução em relação ao inverso da função de transferência do canal ilustra o bom desempenho do método de busca. Para que confirmemos tal fato de outra maneira, mostramos como seria o sinal filtrado pelo melhor indivíduo na Fig. 8.2. É notável como o padrão se aproximada daquele que se espera da constelação em si mesma.



**Figura 8.2:** Saída do Melhor Indivíduo – Primeiro Ensaio

### 8.1.2. Segundo Ensaio

No segundo ensaio, utilizamos um canal mais simples e também passível de inversão perfeita:

$$H(z) = \frac{1 + 0.2z^{-1} - 0.24z^{-2}}{1 - 0.6z^{-1} + 0.25z^{-2}} \quad (8.3)$$

Entretanto, nesse caso, assumimos a existência de ruído AWGN com SNR = 22dB. Teremos como variável adicional de análise, portanto, a influência desse fator estocástico no processo de busca. Para que tal influência seja nítida, adotamos, novamente, um equalizador com característica exatamente inversa à do canal, o que nos permite saber *a priori* a posição do ótimo global. Os parâmetros da rede imunológica se encontram na Tab. 8.2.

Parâmetro	Valor
População Inicial	10
$\sigma_s$	2.5
Número de Clones ( $N_C$ )	10
$\beta$	80
<i>Fitness</i> Exigido para Parada	0.931
Número de Realizações por Ensaio	12

**Tabela 8.2:** Parâmetros – Segundo Ensaio

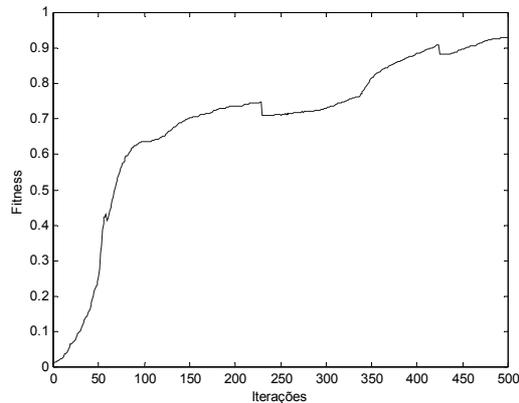
A principal diferença entre a Tab. 8.2 e a Tab. 8.1 é a diminuição do patamar de custo exigido para que o processo pare. A razão de ser dessa mudança é exatamente a presença do ruído, cuja atuação não mais permite a perfeita recuperação do sinal transmitido, reduzindo assim o valor máximo da função custo. Um outro efeito do ruído é deslocar o ponto ótimo da configuração inversa à função de transferência do canal, devido ao fato de o critério CM buscar reduzir, *conjuntamente*, ruído e IIS.

Em todas as realizações, obteve-se uma solução próxima à ideal, tendo sido o número médio de iterações igual a 597. Tal número é menor que aquele obtido no ensaio anterior, fato que se justifica por dois motivos básicos: o espaço de busca tem dimensão menor e o critério de parada é menos exigente. Nossa conclusão é que a rede teve um excelente desempenho.

Na Fig. 8.3, apresentamos a evolução temporal do *fitness* associado ao melhor indivíduo da população para uma realização típica. Verificamos que há um padrão de convergência similar ao mostrado na Fig. 8.1, com uma diferença preponderante: a presença do ruído causa flutuações na estimativa da função custo, que, por sua vez, levam aos pequenos “saltos” mostrados na figura. Esse tipo de efeito é a principal consequência da existência de ruído para o processo de otimização baseado nas ferramentas propostas, pois a estimação da função custo, dependente de uma média temporal, tende a ser mais imprecisa quando há ruído aditivo. Felizmente, o grau de imprecisão pode ser controlado através de uma variável livre: o número de amostras que compõe a janela de estimação. Por fim, apresentamos a melhor solução obtida na realização mostrada:

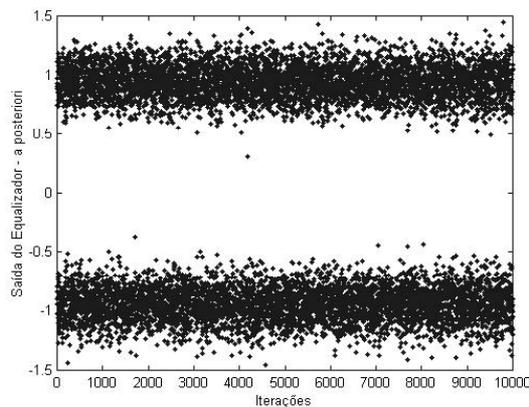
$$W(z) = \frac{0.9466 - 0.6486z^{-1} + 0.2649z^{-2}}{1 + 0.0878z^{-1} - 0.2491z^{-2}} \quad (8.4)$$

Tal solução ainda se aproxima da inversa do canal, embora haja um desvio significativo causado pela presença do processo aditivo gaussiano.



**Figura 8.3:** Evolução Temporal do *Fitness* Máximo – Segundo Ensaio

Na Fig. 8.4, mostramos o padrão correspondente ao sinal de saída numa hipotética fase de teste da solução obtida. Verificamos que o equalizador é capaz de abrir o olho, mas não de eliminar completamente a influência do ruído, o que, aliás, já era esperado devido à extensão do conteúdo em frequência de um ruído AWGN. De qualquer maneira, consideramos que uma situação como a mostrada na figura é deveras satisfatória para um filtro de ordem reduzida como o por nós escolhido.



**Figura 8.4:** Saída do Melhor Indivíduo – Segundo Ensaio

### 8.1.3. Terceiro Ensaio

No terceiro ensaio, adentramos um terreno bem mais incerto, pois consideramos um canal que não é de fase mínima, ou seja, que só pode ser invertido, em tese, por um filtro IIR com *pólos fora do CRU* i.e. *instável*. Numa situação desse tipo, é certo que a melhor solução obtível não será a inversa do canal, mas uma configuração estável que mais adequadamente recupere a característica de módulo constante.

Para que possamos nos concentrar apenas nas implicações dessa interessante configuração, desconsideramos a existência de ruído, sendo as distorções produzidas responsabilidade exclusiva da seguinte função de transferência:

$$H_3(z) = \frac{1 - 0.6z^{-1} - 0.72z^{-2}}{1 - 0.6z^{-1} + 0.25z^{-2}} \quad (8.5)$$

Ante tal modelo, optamos por um equalizador com cinco coeficientes no numerador e mais cinco no denominador de sua função de transferência, opção esta que, neste caso, podemos considerar parcimoniosa. Para otimizar a função CM resultante, escolhemos os parâmetros da rede imunológica conforme mostrado na Tab. 8.3.

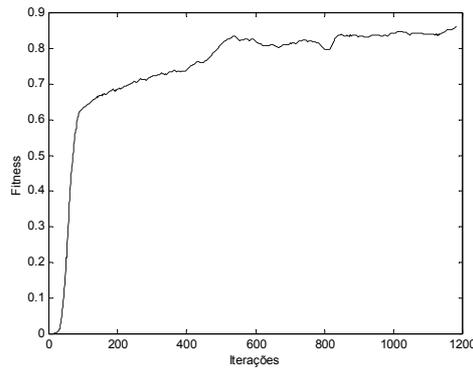
Parâmetro	Valor
População Inicial	10
$\sigma_s$	2.5
Número de Clones ( $N_C$ )	15
$\beta$	70
<i>Fitness</i> Exigido para Parada	0.86
Número de Realizações por Ensaio	12

**Tabela 8.3:** Parâmetros – Terceiro Ensaio

Frisamos que todos os valores, inclusive o custo que regula a parada do processo, foram escolhidos de maneira heurística a partir de ensaios preliminares.

Em 91% das realizações, houve convergência para uma mesma solução, correspondente a um atraso de equalização de cinco amostras. Embora seja bastante complexo analisar a função custo CM para um contexto de filtragem IIR [Regalia, 2001], há um forte indício de que se trate da solução ótima, pois tal seria exatamente o atraso ideal caso fosse usado um equalizador FIR de cinco coeficientes. Dessa forma, tomaremos, com alguma segurança, a frequência dessa solução como taxa de convergência global. Em média, foram necessárias 1107 iterações para que o critério de parada fosse atendido, um valor similar ao obtido no primeiro ensaio, em que o espaço de busca tinha uma ordem semelhante.

Na Fig. 8.5, mostramos a evolução temporal do *fitness* associado ao melhor indivíduo para uma realização típica. Verificamos que o padrão de aprendizado é bastante semelhante ao verificado nos casos anteriores, existindo algumas flutuações devidas, provavelmente, a imprecisões de estimação da função custo.

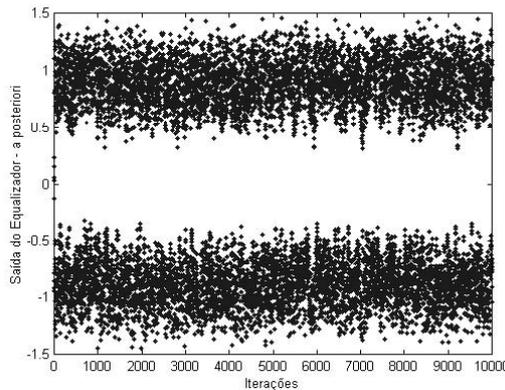


**Figura 8.5:** Evolução Temporal do *Fitness* Máximo – Terceiro Ensaio

O melhor indivíduo para essa realização foi:

$$W(z) = \frac{-0.1279 - 0.2559z^{-1} - 0.3912z^{-2} - 0.5486z^{-3} - 0.7254z^{-4}}{1 + 1.2556z^{-1} + 0.5814z^{-2} + 0.1129z^{-3} - 0.0082z^{-4}} \quad (8.6)$$

e a saída dessa solução para um conjunto de teste hipotético seria como mostrado na Fig. 8.6.



**Figura 8.6:** Saída do Melhor Indivíduo – Terceiro Ensaio

É perceptível que, apesar de se tratar de uma boa solução, ou seja, capaz de “abrir o olho”, há um perceptível grau de interferência intersimbólica que permanece devido à impossibilidade de inversão perfeita.

#### 8.1.4. Discussão

Podemos afirmar que o emprego da rede imunológica artificial no projeto de um equalizador recursivo não-supervisionado teve sucesso, tanto do ponto de vista da convergência global quanto dos outros pontos de vista que nos concernem: inexistência de aproximações e manipulações e ameaça de instabilidade.

Sabíamos *a priori* que o método de busca evolutivo não teria problemas quanto ao aspecto de manipulação da função objetivo, pois tal necessidade é completamente alheia ao processo. Também tínhamos uma forte convicção de que não existiria qualquer problema no que se refere à questão da estabilidade, embora aí pudesse haver, do ponto de vista estatístico, a possibilidade de que a população inteira fosse composta de soluções instáveis. Acreditamos que um caso desse tipo, com que, diga-se de passagem, não nos deparamos jamais, levaria apenas a um mau comportamento transitório, pois a introdução de indivíduos aleatórios tenderia a produzir indivíduos estáveis que logo haveriam de trazer o foco do processo de volta à região “permitida” do espaço de busca. Com isso, consideramos que a robustez do processo de busca a essa questão foi suficientemente atestada.

Do ponto de vista da convergência global, o desempenho da rede imunológica foi bastante satisfatório, pois se atingiu o patamar ideal em dois ensaios e sua vizinhança no ensaio restante. É importante frisar que isso se deu num número médio de iterações perfeitamente compatível com os patamares usuais dos métodos de gradiente.

Essas considerações nos permitem concluir que o paradigma de equalização IIR cega proposto teve um desempenho excelente por aliar uma enorme capacidade de busca global a características complementares às principais dificuldades verificadas no paradigma clássico. Concluimos que o emprego de uma ferramenta de busca como a opt-aiNet representa um avanço no processo de otimização dos parâmetros de um filtro IIR.

## **8.2. Projeto Não-supervisionado de um DFE**

Para submeter essas conclusões otimistas a uma nova prova representativa, analisaremos agora uma proposta distinta: empregar a rede imunológica no projeto de uma estrutura recorrente que é um dos pilares da teoria de equalização digital: o *equalizador com realimentação de decisões* (DFE, *decision-feedback equalizer*) [Proakis, 1995]. Em termos muito simples, o DFE pode ser entendido como um filtro IIR cujo resultado da filtragem *forward* (relativa aos zeros) é realimentada *via decisor* (o que dá à estrutura um caráter não-linear). O projeto cego de uma estrutura desse tipo, embora possa levar a um construto teórico coerente, produz necessariamente uma função custo com múltiplos mínimos. A análise do projeto de um DFE através do critério CM é o tema de alguns trabalhos muito interessantes, como aqueles associados aos esforços de pesquisa de Casas e outros [Casas, 1996, Casas et al., 1998]. Tais esforços levaram a uma conclusão muito interessante: apesar de ser, em tese, possível obter um projeto efetivo a partir do critério do módulo constante, há uma classe de canais para os quais a inicialização usual, que se baseia na noção de anular todos os parâmetros, conduzirá necessariamente à convergência para um mínimo local espúrio. Em outras palavras, existem situações concretas em que o paradigma clássico de equalização terminará por produzir soluções pouco adequadas [Casas, 1996].

Nos demais aspectos pertinentes ao problema de equalização com estruturas recursivas, há uma diferença fundamental entre o DFE e o filtro IIR: *o decisor evita que haja qualquer tipo de divergência para o infinito*. Por outro lado, a existência de um dispositivo não-linear de decisão torna ainda mais complexo qualquer tipo de manipulação

da função custo<sup>140</sup>. Com isso, permanece uma motivação fundamental além daquela delineada pela existência de múltiplos mínimos.

Para testar essa abordagem do problema de equalização com realimentação de decisões, projetaremos DFEs sob diferentes modelos de canal e buscaremos avaliar o desempenho do duo formado pelo critério CM e pela rede imunológica opt-aiNet. Para evitar que a presença de uma não-linearidade na essência do DFE produza soluções incoerentes na função custo cega, restringiremos nossos esforços ao projeto de dispositivos dotados exclusivamente de parâmetros livres na parte *backward*<sup>141</sup>. O problema de otimização que se nos apresenta é, portanto, o de otimizar a função custo CM estimada e transformada segundo (5.1) com respeito ao vetor de parâmetros da parte recursiva do filtro, sendo o filtro *forward* um dispositivo trivial cuja resposta entrada-saída é a identidade.

Consideraremos, como de praxe, que o sinal transmitido tem amostras i.i.d. extraídas de uma constelação 2-PAM ou 2-PSK de módulo unitário. Tomaremos como base de testes dois modelos de canal<sup>142</sup>:

$$\begin{cases} \text{C1 : } H_1(z) = 1 + 0.9z^{-1} - 0.8z^{-2} \\ \text{C2 : } H_2(z) = 1 + 0.1z^{-1} - 1.8z^{-2} - 0.9z^{-3} \end{cases} \quad (8.7)$$

e desconsideraremos a existência de ruído aditivo.

### 8.2.1. Primeiro Ensaio

No primeiro ensaio, escolhemos o modelo de canal  $H_1(z)$  e um DFE com dois parâmetros na parte *backward*, ou seja, com resposta entrada-saída igual a:

$$y(n) = s(n) - b_1 \text{sign}[y(n-1)] - b_2 \text{sign}[y(n-2)] \quad (8.8)$$

sendo  $\text{sign}(\cdot)$  o mapeamento do decisor e o vetor  $\mathbf{f}_b$  formado por  $b_1$  e  $b_2$  o vetor de parâmetros a ser otimizado pelo algoritmo evolutivo. Nesse caso, haverá dois mínimos: a solução ideal, correspondente à configuração  $\mathbf{f}_b = [-0.9 \ 0.8]^T$ , e uma solução sub-ótima localizada em  $\mathbf{f}_b \cong [-0.15 \ -0.35]^T$ . Conforme mostram Casas e outros [Casas et al., 1998], a segunda dessas configurações corresponde a um filtro ineficiente do ponto de vista da redução de IIS, sendo, portanto, um mínimo espúrio.

<sup>140</sup> Embora não seja possível que um DFE tenha uma saída ilimitada, há determinados padrões de comportamento complexo tão nocivos quanto a instabilidade. Um desses padrões é o fenômeno de *propagação de erro*, marcado por um regime no qual a saída do decisor não reproduz o sinal desejado, ou seja, são realimentados erros de decisão. Ao longo dos ensaios que logo discutiremos, deparamo-nos com esse fenômeno na forma de um comportamento transiente durante a estimação do custo CM para algumas soluções, o que nos levou a desprezar um trecho inicial da janela por uma questão de segurança.

<sup>141</sup> Dessa forma, mantemos o cenário estudado por Casas e outros [Casas et al., 1998]. Recomendamos ao leitor interessado no projeto de um DFE completo que analise as restrições discutidas em [Silva, 2005].

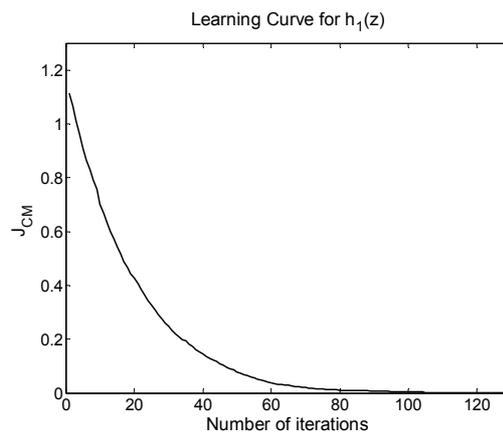
<sup>142</sup> O primeiro canal é um dos estudados em [Casas et al., 1998] e o segundo é um canal de fase não-mínima. Embora o problema resultante não pareça enormemente complexo, consideramos que suas características são suficientes para nosso objetivo: avaliar o potencial de uma ferramenta evolutiva empregada no projeto de estruturas de filtragem.

Na busca de uma configuração ideal, escolhemos os parâmetros da rede segundo a Tab. 8.4.

Parâmetro	Valor
População Inicial	10
$\sigma_s$	0.5
Número de Clones ( $N_C$ )	5
$\beta$	50
<i>Fitness</i> Exigido para Parada	0.999
Número de Realizações por Ensaio	100

**Tabela 8.4:** Parâmetros – Primeiro Ensaio

Após a realização de 100 ensaios, verificamos que, em todos eles, o melhor mínimo foi obtido, ou seja, que a taxa de convergência global obtida foi de 100%. Na Fig. 8.7, mostramos a evolução temporal do custo  $CM^{143}$  associado ao melhor indivíduo para uma simulação típica.



**Figura 8.7:** Evolução Temporal do Custo Associado ao Melhor Indivíduo – Primeiro Ensaio

Novamente, é possível divisar uma etapa de aprendizado mais rápido, marcada pelo rápido aprimoramento de uma população inicial aleatória, seguido por uma etapa de refinamento. É bastante interessante notar que o número de iterações requerido para convergência é pequeno em relação aos padrões de algoritmos clássicos de gradiente, o que, em parte, explica-se se considerarmos que o espaço de busca tem apenas duas dimensões. De qualquer maneira, é patente que o desempenho da ferramenta de busca foi bastante satisfatório nesse caso.

<sup>143</sup> Note o leitor que, nesse caso, é mostrada a evolução do custo  $CM$ , não do *fitness*.

### 8.2.2. Segundo Ensaio

No segundo ensaio, escolhemos o canal com função de transferência  $H_2(z)$  e um DFE com três parâmetros livres na parte *backward* e resposta entrada-saída igual a:

$$y(n) = s(n) - b_1 \text{sign}[y(n-1)] - b_2 \text{sign}[y(n-2)] - b_3 \text{sign}[y(n-3)] \quad (8.9)$$

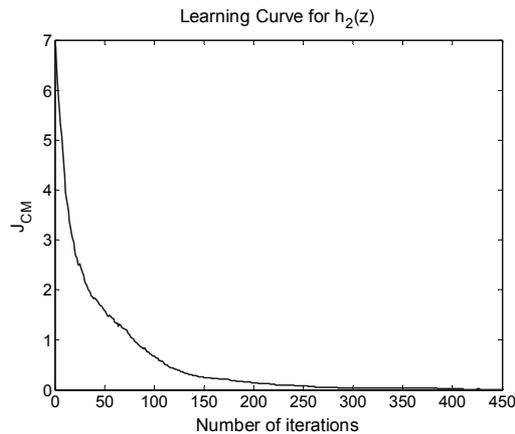
Nesse caso, é possível inverter perfeitamente o canal, o que garante que a solução  $\mathbf{f}_b = [-0.1 \ 1.8 \ 0.9]^T$  será um mínimo global da função CM. Embora não seja possível visualizar a função custo resultante, sabe-se que ela também contém mínimos locais e, devido ao aumento do espaço de busca, deve gerar um cenário de maior complexidade que a daquele mostrado na Fig. 8.7. Para resolver o problema, optamos pelos parâmetros mostrados na Tab. 8.5.

Parâmetro	Valor
População Inicial	10
$\sigma_s$	0.7
Número de Clones ( $N_C$ )	7
$\beta$	40
<i>Fitness</i> Exigido para Parada	0.999
Número de Realizações por Ensaio	100

**Tabela 8.5:** Parâmetros – Segundo Ensaio

Nesse caso, a ferramenta evolutiva também foi capaz de obter a melhor solução em todas as realizações, ou seja, produziu uma taxa de convergência global de 100%. Para que tal nível de desempenho fosse atingido, foi preciso contar com um número médio de 300 iterações, valor que consideramos muito bom em comparação com as clássicas técnicas de gradiente.

Na Fig. 8.8, apresentamos a evolução temporal do custo CM associado ao melhor indivíduo para uma realização típica. Notamos um padrão de convergência bastante similar ao mostrado na Fig. 8.7, e, como naquele caso, bastante satisfatório.



**Figura 8.8:** Evolução Temporal do Custo Associado ao Melhor Indivíduo – Segundo Ensaio

### 8.2.3. Discussão

Uma vez mais, pudemos mostrar que o emprego de uma estrutura flexível de filtragem aliada a um critério coerente e a uma eficiente ferramenta de busca global leva a um paradigma de significativa capacidade. Do ponto de vista da obtenção da melhor solução, a rede imunológica mostrou que é capaz de explorar adequadamente os subsídios fornecidos pelo critério CM, levando a um ajuste ideal do DFE nos dois casos analisados. Isso significa que o uso de um equalizador com realimentação de decisões projetado segundo o critério CM pode formar uma plataforma viável mesmo sob os “maus canais” estudados por Casas e outros. É importante salientar que a ferramenta evolutiva, por não requerer qualquer tipo de manipulação da função custo, evita as aproximações usuais e as recorrentes considerações acerca da qualidade das decisões realimentadas. Isso abre caminho para um método de escolha de parâmetros capaz, efetivamente, de atingir o cerne do potencial associado à estrutura.

Tendo chegado ao final do capítulo, podemos concluir que o uso do critério CM aliado à rede imunológica dá origem a uma abordagem efetiva do problema de equalização por meio de estruturas recursivas, cujas vantagens principais podem ser sucintamente expostas através de uma lista tríplice:

- O potencial de busca da técnica permite o uso seguro de um critério de equalização não-supervisionado.
- O fato de a ferramenta evolutiva demandar apenas o cálculo da função custo permite que o processo de busca se dê sem o auxílio de considerações simplificadoras que, via de regra, são uma fonte significativa de imprecisão.
- Finalmente, a existência de uma população de soluções, ou seja, de uma diversidade de candidatos, reduz substancialmente a chance de que o processo de busca seja infrutífero devido à ameaça da instabilidade.

Tais vantagens contribuem para que se estabeleça a viabilidade das propostas mostradas nesse capítulo e, além disso, são evidências adicionais em favor do paradigma norteador discutido no prólogo à segunda parte deste trabalho.



## 9. Novas Propostas para Equalização e Pré-distorção Baseadas em Redes Neurais

Nos problemas de equalização supervisionada com atraso fixo, como vimos ao longo do capítulo 3, há duas fontes principais de dificuldades de projeto: a *existência de laços de realimentação* e de *um caráter estrutural não-linear*. O primeiro desses problemas esteve no cerne do capítulo anterior, enquanto o presente abordará a questão faltante: a adaptação de um dispositivo de filtragem altamente flexível.

O cerne da questão é que uma dependência não-linear com respeito aos parâmetros livres produz, via de regra, uma função custo de Wiener dotada de múltiplos mínimos e mesmo de pontos de equilíbrio menos considerados, como pontos de sela [Haykin, 1994a]. Por esse motivo, torna-se atraente aliar o potencial de um filtro não-linear dotado de significativa capacidade a uma ferramenta de busca capaz de escolher seus parâmetros de maneira efetiva. Tal é exatamente o primeiro objetivo do presente capítulo: propor o uso da rede imunológica opt-aiNet no projeto de um equalizador baseado numa rede MLP<sup>144</sup>, dispositivo que, sabidamente, é dotado de potencial de aproximação universal (vide seção 4.2). Com isso, esperamos que se estabeleça uma abordagem capaz de, em tese, extrair da própria idéia de minimizar o erro quadrático médio o seu máximo potencial.

Mostraremos também neste capítulo que a sinergia entre estrutura, critério e algoritmo de otimização pode levar, outrossim, ao estabelecimento de uma nova técnica para pré-distorção de canais de satélite. A base de tal empreitada tem dois pilares fundamentais: o potencial de busca da opt-aiNet e a inexistência da necessidade de manipulação da função custo.

Finalmente, apresentaremos uma aplicação pioneira da rede imunológica ao ajuste dos parâmetros de uma rede recorrente destinada à equalização. Num problema desse tipo, são complicadores adicionais os desafios próprios da adaptação de estruturas recursivas, os quais foram discutidos em detalhe no capítulo 8. Tal cenário é, portanto, de uma generalidade que se harmoniza perfeitamente com os objetivos que permeiam o prólogo à segunda parte.

### 9.1. Equalização Supervisionada de um Canal Não-linear

Tratemos, pois, de submeter a proposta a um primeiro teste. Nossa idéia é projetar um equalizador MLP através de uma estimativa da função custo de Wiener a partir de médias temporais<sup>145</sup>, tendo como base a rede imunológica opt-aiNet. Para que o potencial de aproximação da estrutura seja significativamente explorado, assumiremos que o canal de comunicação é não-linear, bem como que existe ruído AWGN. A relação entrada-saída do canal é:

---

<sup>144</sup> Lembramos ao leitor que, ao longo de todo o trabalho, só lidamos com redes MLP dotadas de uma única camada intermediária e de neurônio de saída linear.

<sup>145</sup> Como em todos os problemas de minimização presentes neste trabalho, é preciso submeter a função ao mapeamento apresentado na equação (5.1).

$$x(n) = [0.5547s(n) + 0.8320s(n-1)] + 0.3[0.5547s(n) + 0.8320s(n-1)]^2 + r(n) \quad (9.1)$$

onde  $r(n)$  é o ruído. Consideramos que o sinal transmitido é formado por amostras i.i.d. de uma constelação 2-PAM ou 2-PSK de potência unitária. Como termo de comparação, escolhemos o dispositivo ótimo do ponto de vista da probabilidade de erro de símbolo, ou seja, o equalizador de Bayes. Embora um paralelo entre a abordagem proposta e a bayesiana pareça, *a priori*, infundado devido à incompatibilidade entre os critérios de otimização, na prática é muito relevante confrontar os dois paradigmas para que se tenha uma idéia do verdadeiro grau de discrepância entre eles. Além disso, como uma probabilidade de erro mínima é, “no frigidus ovos”, a meta de um projetista envolvido em comunicação digital, vale a pena medir nossa proposta através dessa referência.

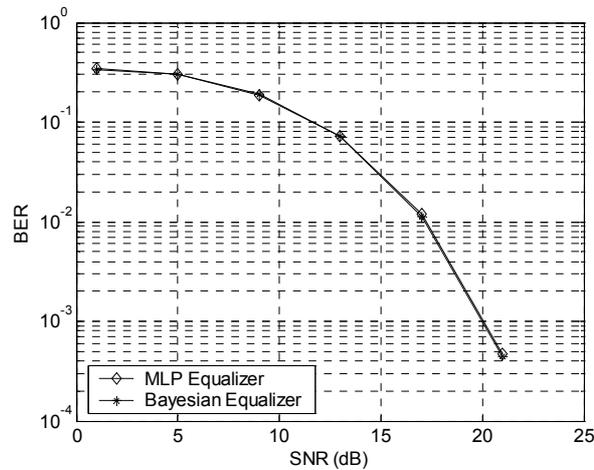
O esquema de codificação adotado é bastante simples: o vetor de parâmetros que serve de base ao algoritmo é formado pelos vetores de todos os neurônios da camada intermediária e pelo vetor do neurônio de saída. Optamos por um atraso de equalização nulo e pela configuração mostrada na Tab. 9.1.

Parâmetro	Valor
População Inicial	5
$\sigma_s$	5
$N_C$	5
$\beta$	30
Critério de Parada	3000 iterações

**Tabela 9.1:** Parâmetros da Rede Imunológica – Equalização com MLP

Tendo em vista a constante busca por uma estrutura tão parcimoniosa quanto possível, escolhemos uma rede formada por duas entradas e seis neurônios na camada intermediária. Adotamos um critério de parada baseado num número máximo de 3000 iterações, após as quais foi armazenada a solução obtida. Esta foi então utilizada no processo de estimação da taxa de erro de bit, caracterizado pela transmissão de dez milhões de amostras por SNR analisada<sup>146</sup> e por uma estimação final baseada em dez repetições do processo global. O resultado de todo esse esforço se encontra na Fig. 9.1, que também contém os dados relativos ao filtro de Bayes.

<sup>146</sup> Para SNRs mais baixas, marcadas pela ocorrência de mais erros, comumente o processo foi interrompido após um limiar de 3000 erros, considerado suficiente para nossos fins.



**Figura 9.1:** Curvas de Erro – MLP e Bayes

É nítido que o equalizador MLP teve um desempenho praticamente idêntico ao do dispositivo bayesiano para todos os valores de relação sinal-ruído, o que nos permite assumir que a rede operou em seu limite estrutural em todos os ensaios pertinentes ao levantamento do gráfico. Dessa forma, ainda que indiretamente, é-nos lícito atestar o sucesso da abordagem em termos de busca global.

Embora tal análise seja coerente, ela não esgota o alcance das conclusões que devem ser tiradas do sucesso dessa proposta. Em particular, podemos destacar mais dois pontos que se ligam a diversos aspectos relevantes abordados ao longo dos capítulos 3 e 4:

- Uma estrutura MLP parcimoniosa tem flexibilidade suficiente para rivalizar com um dispositivo do tipo RBF ótimo (bayesiano) em casos representativos.
- O critério de mínimo erro quadrático médio aliado a uma estrutura flexível pode, em determinados casos, levar a taxas de erro de símbolo *compatíveis com o limite teórico*.

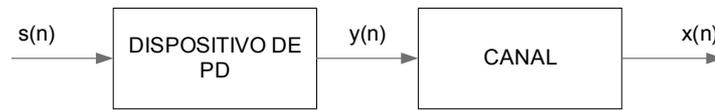
Em poucas palavras, demonstramos que o uso de uma estrutura MLP e do critério de Wiener é capaz de produzir um desempenho comparável ao de um dispositivo ótimo, desde que o método de busca seja eficiente na escolha dos parâmetros. Isso indica que a proposta pode ser uma alternativa eficiente e menos vulnerável, em tese, a problemas de “explosão de dimensionalidade” ou a dificuldades decorrentes de um treinamento em múltiplas etapas.

Por fim, destacamos que se trata da primeira aplicação da rede opt-aiNet ao ajuste de parâmetros de uma rede neural, fato que situa a contribuição exposta como uma opção de projeto no sem-número de problemas solucionáveis através de uma estrutura desse tipo.

## 9.2. Uma Nova Proposta para Pré-distorção

O problema de pré-distorção (PD) pode ser considerado uma “versão invertida” do problema de equalização estudado no capítulo 3. Sua razão de ser é, como no caso da

equalização, *reduzir* ou *eliminar as distorções impostas pelo canal*<sup>147</sup>. A principal diferença está na localização do dispositivo responsável pelo cumprimento da tarefa: enquanto no processo mais usual ele se encontra no receptor, na pré-distorção, *o dispositivo é parte do transmissor*. A Fig. 9.2 traz um diagrama que ilustra o caminho percorrido pela informação num processo de PD típico.



**Figura 9.2:** Esquema de Pré-distorção

Para que compreendamos a lógica da configuração mostrada na Fig. 9.2, imaginemos um sistema de comunicação dotado de um transmissor e de múltiplos receptores. Sob a égide de um mecanismo de equalização, seriam necessários tantos dispositivos de compensação quanto receptores. Tendo por base a idéia de pré-distorção, bastar-nos-ia um filtro no transmissor. Por outro lado, não devemos assumir que se trate de uma panacéia: cada receptor tem a si associado um canal de comunicação distinto e, possivelmente, variável, o que torna muito pouco plausível que um único dispositivo situado no transmissor possa pré-processar o sinal transmitido de maneira a compensar *todos os efeitos nocivos de todos os canais*. Isso nos obriga a reduzir nossas ambições: um dispositivo de pré-distorção (DPD) não pode ser uma solução única. No entanto, isso não quer dizer que ele seja inútil: se houver distorções *impostas pelo próprio transmissor*, resolvê-las nesse mesmo âmbito será, sem dúvida, o ideal.

Uma situação desse tipo emerge comumente em sistemas de comunicação via satélite. É bastante comum que, para que se economize energia, faça-se o amplificador de transmissão do satélite operar num regime próximo ao de saturação, o que termina por potencializar o surgimento de não-linearidade na relação entre o sinal efetivamente enviado e aquele gerado para transmissão. Do ponto de vista sistêmico, esse desvio em relação à linearidade provoca uma modificação no módulo e na fase dos símbolos da constelação adotada, fato que, se não for adequadamente levado em conta, termina por produzir uma degradação de desempenho. É exatamente para evitar essa piora que a idéia de pré-distorção vem sendo considerada por diversos pesquisadores [Ibnkahla e Castanié, 1996, Aschbacher e Rupp, 2003].

Usualmente, o projeto de um DPD é precedido pelo levantamento do modelo do amplificador, subsídio essencial para que se determine a forma pela qual a distorção controlada deverá atuar. Um *modus operandi* desse tipo se nos afigura similar à noção de *zero-forcing*, um tanto incômoda por requerer uma estimação prévia, fator externo aos mecanismos automáticos derivados da informação contida nos sinais. Parece-nos, portanto, bastante atraente contemplar o problema do prisma da abordagem de mínimo erro quadrático, tão prolífico no campo da equalização adaptativa. Analisando a Fig. 9.2, torna-se claro que a idéia seria minimizar o seguinte critério:

$$J_{PD} = E[s(n) - x(n)]^2 \quad (9.2)$$

<sup>147</sup> Usamos aqui o termo canal num sentido que deve abranger todas as fontes de distorção.

Considerando, de imediato, uma análise dos pontos de gradiente nulo dessa função, deparamo-nos com um problema essencial: para que “retro-propaguemos” o erro até os parâmetros do DPD, *será necessário conhecer o modelo de canal*. Como aquele que percorre um labirinto, chegamos a um beco sem saída: eis que surge novamente a necessidade que tanto buscamos evitar.

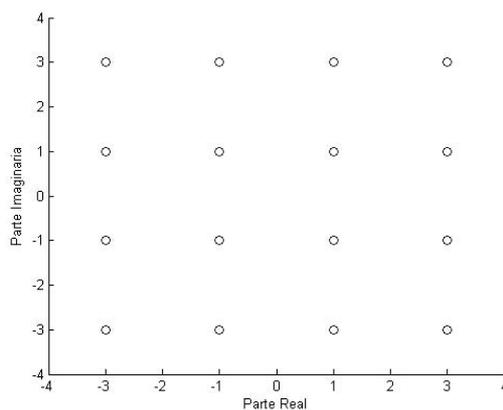
Tratemos, pois, de buscar o fio que nos conduzirá à saída. Analisemos, por um instante, a causa de nosso infortúnio: não foi o critério de erro quadrático médio em si que ocasionou a rediviva necessidade de um modelo, *mas sim a idéia de aplicar o operador gradiente à função custo*. Portanto, se estimássemos o erro quadrático médio com o auxílio de uma média temporal e aplicássemos uma técnica de busca independente de qualquer manipulação, *poderíamos, em tese, chegar a uma abordagem livre de qualquer estimação explícita*. Assim, o novelo salvador pode ser a técnica que vimos empregando, ou seja, a rede imunológica.

Para que ponhamos essa proposta à prova, optamos por um modelo simples de amplificador, baseado no consagrado modelo de Saleh [Ibnkahla, 2000, Aschbacher e Rupp, 2003b]. Em tal modelo, a distorção é expressa em termos de módulo e fase e os parâmetros dependem das características do sinal de entrada  $s(n)$ :

$$\begin{cases} A(|s(n)|) = \frac{\alpha_A |s(n)|}{1 + \beta_A |s(n)|^2} \\ \Phi(|s(n)|) = \frac{\alpha_\Phi |s(n)|}{1 + \beta_\Phi |s(n)|^2} \end{cases} \quad (9.3)$$

sendo  $\alpha_A = 11.534$ ,  $\beta_A = 1.6242$ ,  $\alpha_\Phi = 11.431$  e  $\beta_\Phi = 39.071$ .

Tendo em vista que uma distorção significativa dependerá do uso de uma modulação com múltiplos níveis, decidimos testar a proposta com uma constelação 16-QAM, que, para facilitar o trabalho analítico do leitor, mostramos na Fig. 9.3.



**Figura 9.3:** Constelação 16-QAM

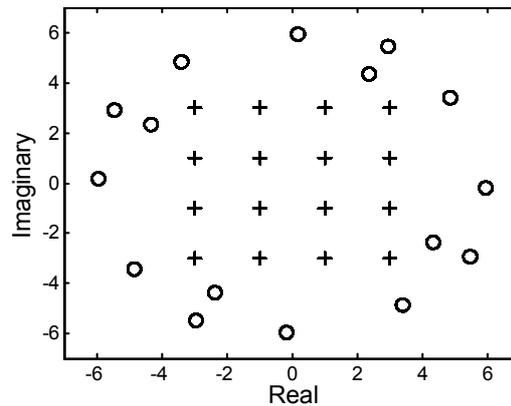
Escolhemos uma rede MLP complexa com uma entrada e 5 neurônios na camada intermediária para o papel de DPD. Conforme discutimos, confiamos a tarefa de otimizar a função estimada de Wiener à rede imunológica artificial opt-aiNet. Adotamos um esquema

de otimização similar ao da seção anterior e dividimos os parâmetros complexos em parte real e parte imaginária. A configuração geral do processo de busca pode ser encontrada na Tab. 9.2.

Parâmetro	Valor
População Inicial	20
$\sigma_s$	3
$N_C$	25
$\beta$	80
Critério de Parada ( <i>Fitness</i> Máximo)	0.99

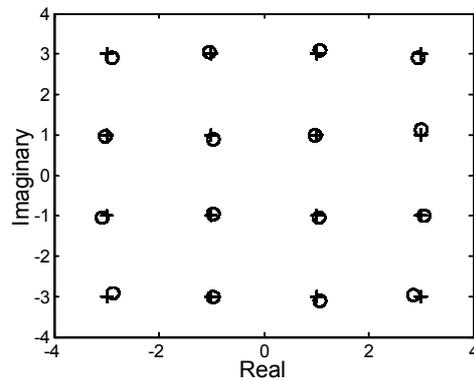
**Tabela 9.2:** Parâmetros da Rede Imunológica – Pré-distorção

Na Fig. 9.4, mostramos a constelação distorcida de acordo com (9.3), ou seja, o sinal na saída do amplificador. Essa figura revela nitidamente que a não-linearidade provoca uma considerável distorção. Esta, por sua vez, agregada aos restantes efeitos nocivos do canal, serviria para incrementar de maneira indesejável a taxa de erro de símbolo.



**Figura 9.4:** Constelação Distorcida ('o') e Constelação Original ('+')

Aplicando o DPD proposto, obtemos um cenário bastante semelhante ao ideal, como nos mostra a Fig. 9.5. Isso revela que a rede neural é capaz de compensar a atuação nociva do amplificador *sem recorrer a um modelo explícito*, fato que atesta a viabilidade do paradigma proposto.



**Figura 9.5:** Constelação “Recuperada” (‘o’) e Constelação Original (‘+’)

A eficiência do processo de pré-distorção nos permite conjecturar que a melhor solução foi obtida, o que, aliás, é perfeitamente plausível à luz da discussão anterior. Isso reafirma a marcante sinergia entre estrutura, critério e ferramenta de busca.

De um ponto de vista voltado ao problema em si, a maior vantagem da nova proposta é dispensar a identificação de canal, baseando-se exclusivamente na função custo elaborada. Interessantemente, essa possibilidade também pode abrir perspectivas de relevo em problemas de outras áreas (como, por exemplo, controle adaptativo). A maior desvantagem teórica da abordagem é, sem dúvida, a necessidade de utilizar o amplificador para a obtenção da estimativa de erro quadrático, fato que, caso se opte por uma janela muito grande, pode onerar excessivamente o sistema.

### 9.3. *Análise de uma MLP Recorrente*

Até agora, percorremos vias que cortam o território das estruturas não-lineares para analisar novas propostas e sua relação com o paradigma que nos norteia. Apenas uma região ainda nos é desconhecida: a dos filtros *não-lineares e recorrentes*. Para os fins que delimitamos no capítulo 3, essa é, de certa forma, a fronteira faltante, o ápice da generalidade estrutural: aliar a flexibilidade de um mapeamento alheio ao princípio da sobreposição às complexas possibilidades abertas pela introdução de *feedback*. Como digna exploradora desse potencial, a rede imunológica que tem dado seguidas provas de valor no campo de testes.

Nossa proposta para esta seção é otimizar os parâmetros de um equalizador neural recorrente através de uma rede imunológica artificial, tendo por norte analítico o binômio formado pelo potencial inerente à estrutura e seu pleno aproveitamento por parte da ferramenta de otimização. Do ponto de vista estrutural, adotamos uma rede MLP com uma topologia interessante, em que a saída de cada neurônio da camada intermediária é realimentada para seus pares e para si mesmo [Williams e Zipser, 1989], e, além disso, decisões passadas da rede são realimentadas para a camada de entrada. Isso cria uma espécie de DFE-Neural-Recorrente, uma estrutura genérica e complexa justificada pelo bom desempenho verificado na prática [Duarte, 2004]. O refinamento de uma estrutura

desse tipo também se encaixa perfeitamente com a nossa idéia de buscar sempre dispositivos tão genéricos quanto possível.

Para testar tal proposta, escolhemos dois canais lineares com as seguintes funções de transferência:

$$\begin{cases} \text{C1 : } H_1(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2} \\ \text{C2 : } H_2(z) = 1 + 2z^{-1} + 1 \end{cases} \quad (9.4)$$

e supomos que existe ruído AWGN. O primeiro dos canais é um canal de fase mista, e o segundo, o que se poderia denominar um canal “patológico”, pois tem dois zeros sobre -1 (vide seção 3.4.4). Para fazer o papel de equalizador, adotamos a rede descrita anteriormente numa versão bastante simples, com três neurônios na camada intermediária, duas entradas e uma realimentação de decisão. Analisemos os resultados.

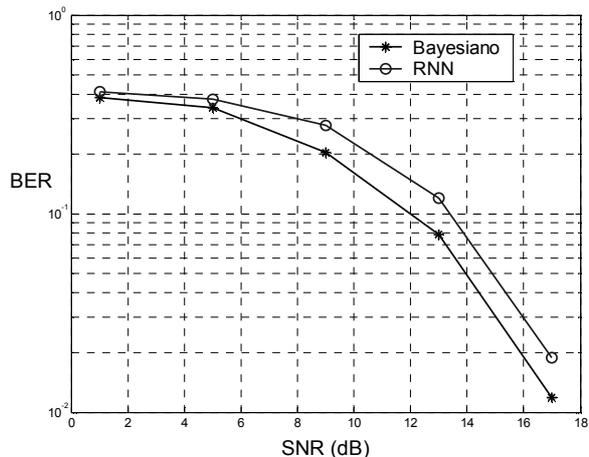
### 9.3.1. Primeiro Ensaio

Para o primeiro ensaio, escolhemos o canal C1 e um sinal transmitido formado por amostras i.i.d. extraídas de uma modulação 2-PAM ou 2-PSK. Adotamos um atraso de equalização nulo, o que, como mostrado em [Duarte, 2004], leva a um problema que só pode ser adequadamente resolvido por meio de um filtro não-linear. Na Tab. 9.3, mostramos os parâmetros do método de otimização escolhido.

Parâmetro	Valor
População Inicial	5
$\sigma_s$	10
$N_C$	5
$\beta$	30
Critério de Parada	3000 iterações

**Tabela 9.3:** Parâmetros da Rede Imunológica – Equalização com Rede Recorrente

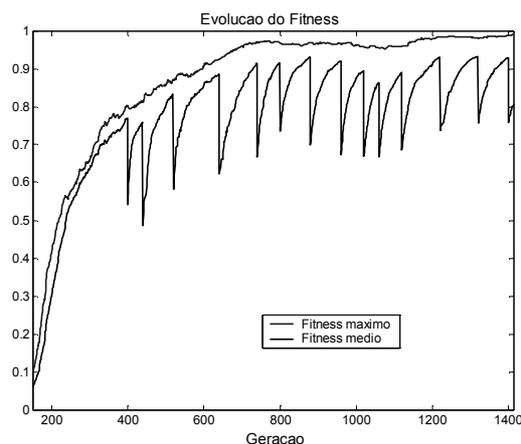
Para testar a rede neural, optamos por levantar a curva de erro a ela associada e depois compará-la à curva ótima produzida pelo equalizador de Bayes, num procedimento idêntico ao que adotamos na seção 9.1. O resultado desse esforço pode ser visto na Fig. 9.6.



**Figura 9.6:** Curvas de Erro – Primeiro Ensaio

Notamos que a rede neural teve um desempenho ligeiramente inferior ao do filtro bayesiano, havendo uma perda de cerca de 1 dB para altas SNRs. À primeira vista, esse resultado pode parecer contraditório devido à presença de laços de realimentação na estrutura adotada. No entanto, não podemos nos esquecer de que o equalizador de Bayes tem 16 “neurônios” na camada intermediária, enquanto nossa rede tem apenas três. Na junção de vantagens e desvantagens, a balança pendeu para o lado do filtro *feedforward* ótimo. Apesar da desvantagem em relação ao filtro de Bayes, o patamar de desempenho atingido pela rede mostra que o método de busca obteve sucesso, projetando um filtro eficiente e, provavelmente, próximo ao limite estrutural. Essa hipótese é confirmada pela coerência do formato da curva, que não revela mudanças abruptas causadas por eventuais fracassos no processo de otimização.

A fim de ilustrar a atuação da *opt-aiNet* numa realização típica, mostramos, na Fig. 9.7, as curvas de evolução temporal do *fitness* máximo e do *fitness* médio da população.



**Figura 9.7:** Evolução Temporal do *Fitness* – Primeiro Ensaio

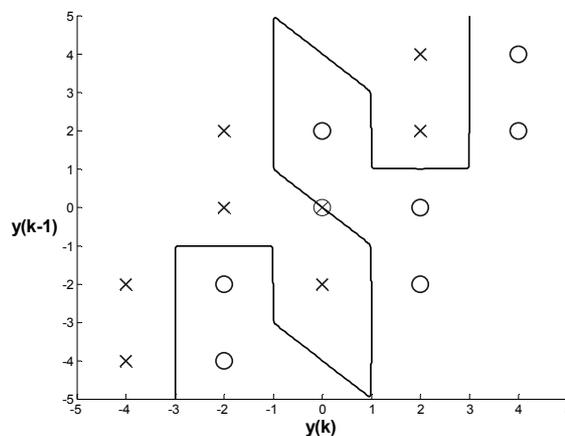
No que diz respeito à curva de *fitness* máximo, percebemos, mais uma vez, um nítido padrão de convergência em duas etapas que leva o nível de desempenho à vizinhança do

ideal. A curva de *fitness* médio, por sua vez, tem um padrão semelhante ao esperado, acompanhando o comportamento da curva ótima à distância e oscilando devido ao processo de poda inerente à rede imunológica.

### 9.3.2. Segundo Ensaio

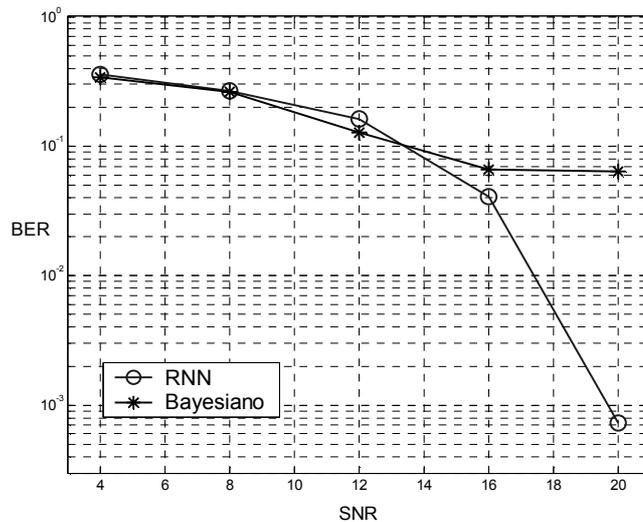
No segundo ensaio, mantivemos o modelo de sinal e escolhemos o canal C2. O motivo dessa escolha não é fortuito: conforme discutido na seção 3.4.3, a existência de zeros sobre +1 ou -1 torna impossível classificar perfeitamente os estados do canal, pois, com certeza, haverá coincidência entre eles. Nesse caso, o equalizador de Bayes terá a si associado um limiar persistente de erro de decisão causado não pelo ruído, mas pela impossibilidade de classificar um estado coincidente. Isso implica a existência de uma significativa degradação de desempenho.

Uma forma de evitar esse erro persistente é introduzir um laço de realimentação na estrutura de filtragem, o qual, em termos muito simples, introduz informação decisiva para que sejam adequadamente discriminados os estados coincidentes. É exatamente nesse cenário que se insere este ensaio: verificaremos como se comportará a rede recorrente em relação ao filtro de Bayes num contexto que, de antemão, desfavorece o dispositivo ótimo. Para que possamos compreender exatamente a situação, mostramos, na Fig. 9.8, os estados do canal e a curva de separação bayesiana. Verificamos a existência de dois estados situados sobre a origem, cada qual associado a um rótulo distinto, o que, mesmo na ausência de ruído, levará a uma taxa de erro de  $1/16 = 0.0625$ .



**Figura 9.8:** Estados e Curva de Separação de Bayes para uma SNR alta

Mantivemos a mesma rede neural, bem como os parâmetros mostrados na Tab. 9.3. Mais uma vez, optamos por aferir o desempenho da proposta através do levantamento de uma curva de erro e de uma subsequente comparação com aquela produzida pelo filtro ótimo. As curvas em questão se encontram na Fig. 9.9.



**Figura 9.9:** Curvas de Erro – Segundo Ensaio

Para valores baixos de SNR, o desempenho da rede neural é levemente inferior àquele verificado para o bayesiano. No entanto, quando a potência do ruído diminui, a rede neural, por ser capaz de classificar o estado coincidente, continua a reduzir a taxa de erro, enquanto o filtro de Bayes, pelas razões expostas, não consegue descer abaixo do limiar de 0.0625. Os resultados, portanto, confirmam que:

- Na presença de estados coincidentes, uma estrutura dotada de realimentação tem um desempenho significativamente superior ao de uma estrutura *feed-forward* ótima.
- Mais uma vez, a rede imunológica levou a estrutura a níveis de desempenho bastante satisfatórios, independentemente da potência do ruído.
- A rede imunológica se presta muito bem ao ajuste de estruturas recorrentes, pois não dá origem a problemas de nenhuma espécie no que se refere à estabilidade e à manipulação da função custo.

Com mais uma série de resultados satisfatórios, concluímos a seção e o capítulo, certos de haver exposto propostas viáveis e dado mais suporte à idéia de aliar estruturas genéricas a ferramentas evolutivas.



## 10. Novas Propostas para Identificação de Sistemas Baseadas em Computação Evolutiva

O problema de identificação emerge em diversas instâncias relacionadas à engenharia por um motivo muito simples: é extremamente valioso levantar informações sobre sistemas com os quais se lida, e isso pode ser feito através da confecção de modelos capazes de reproduzir suas relações entrada-saída. A obtenção desses êmulos, como vimos na seção 3.1, depende da escolha de uma estrutura que sirva de modelo, de um critério de otimização e de uma técnica de busca. Estamos, portanto, em terreno bem conhecido, restando-nos analisar em detalhe o que resulta das possibilidades geradas pelos três itens fundamentais.

A aplicação de um modelo FIR e do critério de Wiener leva a um cenário muito simples e eficiente: a função custo tem um único mínimo e, como técnica de busca, pode ser usado um algoritmo adaptativo como o LMS ou o RLS (vide seção 3.5.2). Para os fins deste trabalho, trata-se de um problema já muito bem resolvido. Restam-nos dois nichos fundamentais para investigação: o uso de uma estrutura distinta de um filtro linear FIR ou de um critério não-supervisionado.

Primeiramente, mantenhamos a estrutura linear mais simples e popular como modelo do sistema ignorado e consideremos o problema de identificá-lo de maneira não-supervisionada. Conforme argumentamos na seção 3.5.3, uma opção natural para resolvê-lo é a abordagem de *cumulant matching*, baseada na idéia de encontrar um modelo capaz de reproduzir o comportamento estatístico da planta investigada. A função custo associada a esse *modus operandi* leva, devido à indispensável existência de estatísticas de ordem superior, a uma dependência não-linear com respeito aos coeficientes do modelo, fato que, por sua vez, abre concretas perspectivas de existência de mínimos locais e outros pontos de equilíbrio [Tugnait, 1987]. Dessa maneira, queda estabelecido um cenário em que adquire relevo a possibilidade de buscar uma melhoria no desempenho através do emprego de uma ferramenta de busca mais eficiente que aquelas baseadas na diferenciação do custo, conforme atestam, aliás, esforços envolvendo a utilização de algoritmos genéticos [Chen et al., 1997]. Inspirados por esses subsídios, investigaremos a aplicabilidade da rede imunológica que vimos adotando com sucesso, e buscaremos, além de uma análise de desempenho do algoritmo, estabelecer alguns elementos de comparação com a abordagem baseada em AGs.

Havendo, pois, preparado o terreno para a primeira contribuição, convém que voltemos nossa atenção à restante possibilidade: o uso de uma estrutura de filtragem distinta de um dispositivo FIR. Saltam aos olhos, de imediato, duas opções: escolher um filtro não-linear, como uma rede neural, e avaliar o problema resultante; ou considerar a existência de um laço de realimentação no modelo. A primeira opção é muito importante, sem dúvida, mas, a partir de uma análise do capítulo anterior, vemos que o paradigma MLP+opt-aiNet de equalização já pode contemplá-la, bastando para isso uma ligeira mudança na formulação do critério de Wiener. Destarte, julgamos que o problema já foi, embora de maneira indireta, abordado. O leitor argumentará que o mesmo vale para o caso de uma estrutura recursiva, valendo como prova o capítulo 8. Até certo ponto isso é correto, mas, neste caso, há algumas sutilezas teóricas importantes que diferenciam o problema de equalização do problema de identificação. Além disso, teremos por base de testes, além de

um filtro IIR, uma estrutura *não-linear* que, até agora, não empregamos. Em suma, na segunda parte deste capítulo analisaremos uma proposta para identificação supervisionada de sistemas recursivos, usando, para isso, uma técnica de otimização evolutiva. O conjunto de testes e análises presente nesses capítulos ampliará ainda mais o escopo do estudo que, implicitamente, vimos empreendendo acerca da viabilidade de um paradigma genérico de filtragem.

### 10.1. Rede Imunológica e Cumulant Matching

Na seção 3.5.3, analisamos em detalhe o problema de *cumulant matching* e chegamos a uma conclusão essencial: é possível identificar, de maneira não-supervisionada, um modelo de canal FIR na presença de ruído AWGN, havendo um único preço a pagar: *a função custo resultante conterá múltiplos ótimos*. A principal consequência disso é que o uso de uma técnica de busca baseada nas derivadas do custo deverá produzir um modelo sub-ótimo para um conjunto significativo de possíveis inicializações.

A existência de dilemas desse tipo abre a perspectiva concreta de introduzir métodos de busca global mais refinados. De fato, uma primeira abordagem desse tipo levou ao uso de algoritmos genéticos, o que, aliás, deu-se com marcante sucesso [Chen et al., 1997]. Por meio desses esforços pioneiros, provou-se ser possível levar o problema de identificação não-supervisionada a patamares de desempenho e confiabilidade mais elevados por meio do uso de uma técnica de busca mais eficiente. Conseqüentemente, abriram-se caminhos para novas propostas e análises, o que despertou em nós o interesse de avaliar de maneira sistemática, tomando como referência a segura medida de *convergência global*, o desempenho de uma proposta baseada na rede imunológica artificial opt-aiNet. Frisamos que não é nosso desejo buscar, com isso, uma base rigorosa de comparação entre a ferramenta escolhida e um AG, o que, aliás, poderia ser feito de maneira mais proveitosa em outras paragens. Apenas consideramos a rede imunológica uma ferramenta com mérito e coerência próprios e julgamos relevante submetê-la a testes voltados a nossos fins e ao arcabouço teórico que serve de base a nossas experimentações.

Trataremos, pois, de resolver o problema de *cumulant matching* otimizando a função custo exposta em (3.73)<sup>148</sup> por meio da rede imunológica que nos serve de padrão. Para testar a nova abordagem, escolhemos dois canais:

$$\begin{cases} C1 : H_1(z) = 1 + 2.1z^{-1} + 0.9z^{-2} \\ C2 : H_1(z) = -0.21 - 0.5z^{-1} + 0.72z^{-2} + 0.36z^{-3} + 0.21z^{-4} \end{cases} \quad (10.1)$$

sendo considerada a existência de ruído AWGN. O sinal transmitido, como de costume, é formado por amostras i.i.d. extraídas de uma constelação 2-PAM ou 2-PSK de módulo unitário.

<sup>148</sup> E, como é de praxe, transformada pela equação (5.1).

### 10.1.1. Primeiro Ensaio

No primeiro teste, escolhemos o canal C1, desconsideramos o ruído e tomamos um modelo com o mesmo número de parâmetros da planta. Tais escolhas produzem um cenário mais complexo que o exposto e analisado no exemplo 3.9, havendo, sem dúvida, mais que dois pares de mínimos. Escolhemos o conjunto de parâmetros mostrado na Tab. 10.1 para o algoritmo de otimização.

Parâmetro	Valor
População Inicial	5
$\sigma_s$	0.35
$N_C$	10
$\beta$	50
Critério de Parada	1000 iterações

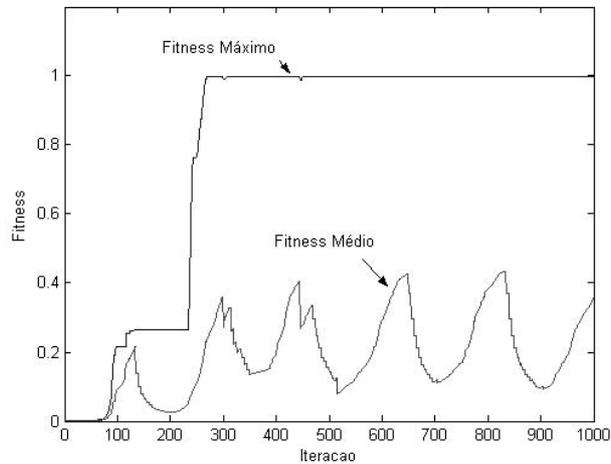
**Tabela 10.1:** Parâmetros da Rede Imunológica – Primeiro Ensaio

Em todos as vinte realizações, o modelo obtido correspondeu exatamente à planta, ou seja, podemos considerar que houve 100% de convergência global. Para avaliar de maneira mais precisa o desempenho da ferramenta, definimos a seguinte medida de desvio em relação ao caso ideal:

$$D = \frac{\|\mathbf{h}_{\text{obtido}} - \mathbf{h}\|}{\|\mathbf{h}\|} \quad (10.2)$$

sendo  $\mathbf{h}_{\text{obtido}}$  a estimativa do vetor de parâmetros e  $\mathbf{h}$  seu valor exato. Nesse primeiro ensaio, o valor médio de D foi 0.00034, o que revela um grau de precisão bastante satisfatório.

Na Fig. 10.1, mostramos a evolução dos custos máximo e médio para uma das realizações. Observamos um padrão de aprendizado bastante rico na curva ótima: há uma fase inicial de aprendizado, marcada pelo refinamento regular numa região promissora, que é interrompida por uma transição abrupta, a qual, provavelmente, deve-se ao surgimento de um indivíduo melhor em outra região do espaço. Esse novo primaz é então lapidado celeremente, sendo o patamar ideal de desempenho atingido bem antes das 1000 iterações previstas. O *fitness* médio, por sua vez, mostra uma ligeira tendência de crescimento acompanhada de oscilações produzidas pelos processos de inserção e poda inerentes à opt-aiNet. A distância entre as curvas revela que a diversidade populacional é significativa, fato que suporta nossa interpretação da mudança abrupta verificada na curva máxima. Concluimos que o desempenho da abordagem proposta foi muito bom em todos os sentidos analisados.



**Figura 10.1:** Evolução Temporal do *Fitness* – Primeiro Ensaio

### 10.1.2. Segundo Ensaio

No segundo ensaio, optamos por um modelo de canal mais complexo, C2, pela introdução de ruído com SNR = 20dB e por um modelo também capaz de engendrar uma condição ideal. Os parâmetros da ferramenta de busca se encontram na Tab. 10.2.

Parâmetro	Valor
População Inicial	20
$\sigma_s$	0.5
$N_C$	10
$\beta$	50
Critério de Parada	1000 iterações

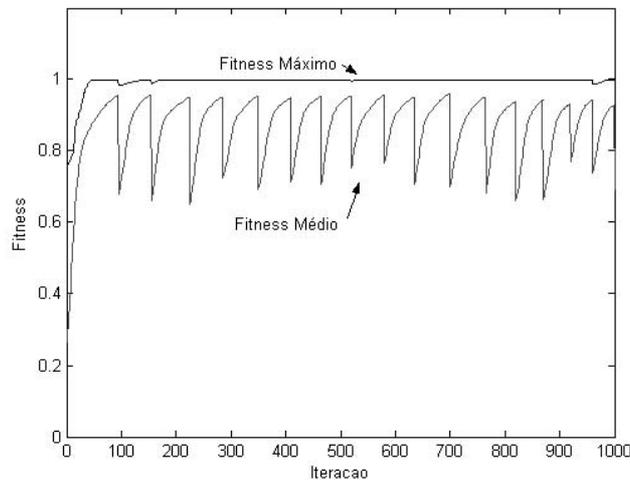
**Tabela 10.2:** Parâmetros da Rede Imunológica – Segundo Ensaio

Em todas as vinte realizações, foi obtida a melhor solução, ou seja, um vetor de parâmetros próximo ao modelo de canal. O valor de D para este segundo ensaio foi 0.0813, um patamar que também consideramos muito satisfatório, especialmente na presença de ruído<sup>149</sup>. Na Fig. 10.2, apresentamos a evolução temporal do *fitness* para uma das realizações. Notamos que o custo máximo rapidamente atinge o limiar ideal, lá permanecendo até o final das 1000 iterações<sup>150</sup>. O custo médio, neste segundo ensaio, manteve uma distância menor em relação ao máximo, fato que indica um menor nível de diversidade. As oscilações, como de costume, devem-se ao processo de inserção e poda da

<sup>149</sup> Em tese, uma precisão maior poderia ser obtida por meio de uma mudança no valor de  $\beta$  ou de um método de refinamento baseado no gradiente.

<sup>150</sup> Nesse caso, as pequenas flutuações em “regime permanente” se devem, provavelmente, à atuação do processo de poda.

rede imunológica. A conjunção desses subsídios analíticos nos faz concluir que, novamente, o desempenho da abordagem proposta foi bastante satisfatório.



**Figura 10.2:** Evolução Temporal do *Fitness* – Segundo Ensaio

### 10.1.3. Conclusão

Os dois ensaios apresentados confirmam que o uso de uma ferramenta de busca evolutiva pode levar a uma metodologia de identificação robusta<sup>151</sup> e parcimoniosa do ponto de vista do emprego de informação acerca do sinal transmitido. A rede opt-aiNet se mostrou uma técnica de busca muito eficiente, tendo atingido o patamar ideal de convergência em ambos os ensaios, sempre com uma gerência racional da população de possíveis soluções. De posse dessas constatações, ousamos incluir a abordagem proposta no *hall* de possíveis técnicas para identificação não-supervisionada.

Alguns esforços de nossa parte no sentido de buscar elementos de comparação com AGs, reportados em [Suyama et al., 2003b], mostram que a rede, no mínimo, tem um desempenho equivalente ao da ferramenta evolutiva mais tradicional. Isso, além de suportar a solidez da proposta, reafirma a importância de que haja um repertório de soluções distintas e eficientes, pois, em engenharia, não se pode escapar à constante avaliação de compromissos.

## 10.2. Rede Imunológica e Identificação de Sistemas Recursivos

A segunda proposta que compõe este capítulo tem uma motivação bastante distinta: identificar, de maneira supervisionada, uma planta recursiva, seja ela linear ou não-linear.

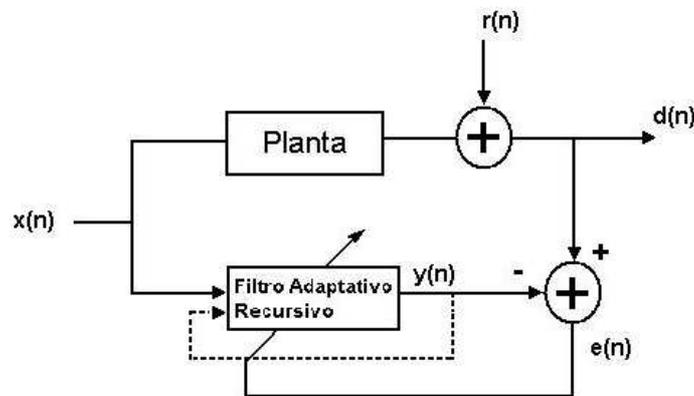
<sup>151</sup> O epíteto “robusta” se refere principalmente à presença de ruído.

Não seria exagero afirmar que a solução desse problema seria, de certa maneira, a solução do problema genérico de identificação supervisionada.

Voltemo-nos para a questão que julgamos essencial: propor uma abordagem que permita o uso de uma estrutura recursiva genérica. Primeiramente, por que damos tanta ênfase à existência de um laço de realimentação? A resposta se encontra, em linhas gerais, nos comentários tecidos ao longo dos capítulos 3, 6 e 8: a tríade “dificuldade de manipulação, instabilidade latente e existência de múltiplos ótimos”. Antes que pareçamos excessivamente redundantes, convém analisar como os investigadores “clássicos” do problema de identificação buscaram lidar com esse problema.

A raiz dos “males” que assombram a filtragem recursiva supervisionada é o processo, em tese infundável, que permeia a idéia de realimentação. A saída de um filtro recursivo depende, segundo um mapeamento regulado por um *ensemble* de parâmetros, de um conjunto de valores de entrada e de suas amostras passadas, que também dependem de entradas e saída passadas segundo um *ensemble*, e isso *ad nauseam*. A sobreposição de todos esses sinais termina por criar uma relação entrada-saída complexa, a qual, por sua vez, reflete-se na tríade de problemas que listamos. Com isso, o esquema de identificação mostrado na Fig. 3.25 sofreria dos mesmos problemas que inspiraram o capítulo 8.

À primeira vista, a única forma de eliminar esses problemas *seria eliminar a própria realimentação*. Embora isso soe como uma declaração pessimista ante a inexorabilidade do destino, há nela um caminho a explorar: permitirá o esquema clássico de identificação alguma brecha que propicie, *ainda que artificialmente*, a remoção do laço? Partamos de uma breve recapitulação: na Fig. 10.3<sup>152</sup>, ilustramos como seria a forma natural de realizar o processo de identificação de uma planta recursiva na presença de um ruído AWGN  $r(n)$ .



**Figura 10.3:** Esquema Clássico de Identificação

Como de costume, o critério de Wiener depende do erro entre a saída da planta e a saída do modelo. O esquema geral é robusto ao ruído devido à ausência de correlação entre os sinais que alimentam o modelo e o processo estocástico  $r(n)$ . Caso o modelo tenha ordem suficiente, a função custo resultante terá um mínimo global na configuração ideal e um conjunto de ótimos locais. Até aqui, estamos em terreno conhecido.

De que maneira seria possível, então, eliminar a realimentação dos parâmetros sem anular a realimentação dos dados? Consideremos a Fig. 10.3 por um instante. Nela, o leitor

<sup>152</sup> Que é uma versão detalhada da Fig. 3.25.

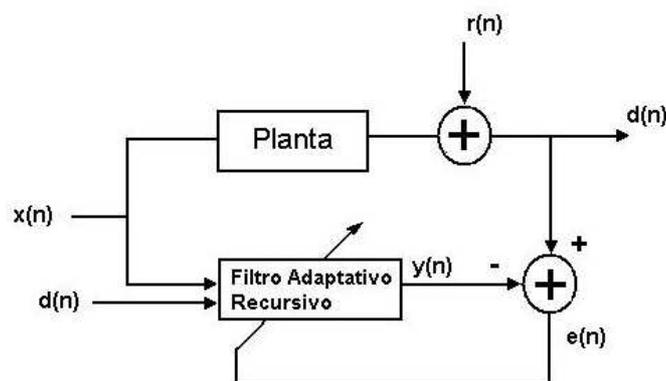
perceberá que há duas estruturas - uma planta e um modelo - que tendem a se igualar no melhor caso. Em outras palavras, há uma *tendência de atração entre o estado do modelo e as características da planta*. É importante ressaltar que, embora o modelo dependa totalmente da planta, o contrário não ocorre: a planta não depende do processo de adaptação do modelo. Surge então uma idéia: por que não *realimentar no modelo o sinal de saída da planta*? Em outras palavras, supondo que o modelo tenha a seguinte relação genérica entrada-saída:

$$y(n) = F[x(n), x(n-1), \dots] - B[y(n-1), y(n-2), \dots] \quad (10.3)$$

sendo  $F[.]$  e  $B[.]$  mapeamentos quaisquer, por que não assumir, de maneira “fictícia”, que sua relação é:

$$y(n) = F[x(n), x(n-1), \dots] - B[d(n-1), d(n-2), \dots] \quad (10.4)$$

Uma análise da equação (10.4) revela que *não há realimentação de parâmetros*, ou seja, que a relação de saída do filtro recursivo se transforma numa “pseudo-relação tipo *feed-forward*”<sup>153</sup>. A Fig. 10.4 mostra o esquema de identificação resultante dessa idéia:



**Figura 10.4:** Esquema Alternativo de Identificação

Sob tal cenário, rompemos as dificuldades da triade e chegamos a uma função custo unimodal e a uma relação simples entre o erro quadrático e os parâmetros do modelo. Há, no entanto, um preço: embora, na ausência de ruído, a função custo resultante tenha um mínimo ideal, a presença de ruído gerará uma polarização, ou seja, um deslocamento da solução obtida em relação ao ótimo. Em outras palavras, como o processo mostrado na Fig. 10.4 *realimenta o ruído no modelo*, haverá correlação entre  $r(n)$  e  $y(n)$ , e, conseqüentemente, a estimativa produzida não mais será idêntica à planta.

As duas formulações apresentadas recebem os nomes, respectivamente, de *output-error formulation (OEF)* e *equation-error formulation (EEF)*, sendo seus prós e contras parte de um dilema clássico em identificação supervisionada [Shynk, 1989, Mathews, 1991]. Tal dilema, no entanto, não é parte da essência do processo, é parte apenas do

<sup>153</sup> Isso ocorre porque não há dependência de  $d(n)$  em relação aos parâmetros do modelo.

cenário clássico de filtragem, marcado pela limitação das técnicas de busca baseadas em gradiente. Nossos esforços têm revelado que o uso de uma ferramenta de otimização com capacidade de busca global pode romper esse “nó górdio” e unir a precisão do critério de Wiener utilizado na OEF a uma taxa de convergência global semelhante à verificada sob a EEF [Attux et al., 2004a]. Essa é a proposta de que nos ocuparemos no restante do capítulo: aplicar a rede opt-aiNet ao problema ilustrado na Fig. 10.3, buscando, com isso, estabelecer um novo paradigma para identificação de sistemas recursivos.

### 10.2.1. Primeiro Ensaio

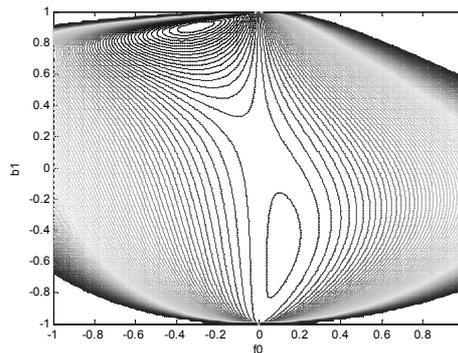
Começemos a testar essa proposta pelo caso mais simples: o de uma planta linear IIR. Para reproduzir um exemplo discutido num importante artigo de J. J. Shynk sobre o assunto [Shynk, 1989], optamos pela seguinte função de transferência para a planta:

$$H(z) = \frac{0.05 - 0.4z^{-1}}{1 - 1.1314z^{-1} + 0.25z^{-2}} \quad (10.5)$$

O modelo escolhido, dotado de apenas dois parâmetros para permitir a visualização da função objetivo, é

$$M(z) = \frac{f_0}{1 - b_1 z^{-1}} \quad (10.6)$$

O sinal de entrada, por sua vez, é formado por amostras i.i.d. de um processo gaussiano de variância unitária e média nula. Para ilustrar as características do problema resultante, mostramos na Fig. 10.5 as curvas de nível associadas à função custo de Wiener para esse problema de identificação.



**Figura 10.5:** Curvas de Nível da Função Custo

Há dois mínimos nessa função: o ótimo global localizado em  $[-0.311 \ 0.906]^T$  e um ótimo local situado em  $[0.114 \ -0.519]^T$ .

Para otimizar a função custo estimada de Wiener<sup>154</sup>, escolhemos a rede opt-aiNet configurada segundo a Tab. 10.3. Desconsideraremos, neste primeiro esforço, a existência de ruído.

Parâmetro	Valor
População Inicial	3
$\sigma_s$	0.15
$N_C$	2
$\beta$	90
Critério de Parada ( <i>Fitness</i> )	0.85

**Tabela 10.3:** Parâmetros da Rede Imunológica – Planta IIR sem Ruído

Em todas as trinta realizações, obteve-se uma solução próxima ao ótimo global. Para avaliar de maneira mais detalhada o desempenho da técnica, empregamos mais uma vez a medida exposta em (10.2). O valor médio dessa grandeza foi de 0.0173, valor este que consideramos satisfatório. Uma solução típica foi  $[-0.3267 \ 0.9007]^T$ .

Consideremos agora a introdução de ruído AWGN com variância igual a 0.25. Os parâmetros para esse caso estão na Tab. 10.4.

Parâmetro	Valor
População Inicial	3
$\sigma_s$	0.4
$N_C$	2
$\beta$	90
Critério de Parada ( <i>Fitness</i> )	0.73

**Tabela 10.4:** Parâmetros da Rede Imunológica – Planta IIR com Ruído

Uma vez mais, foi obtida uma taxa de convergência global de 100%. O valor médio da medida de divergência foi 0.022, e uma solução típica,  $[-0.3308 \ 0.9099]^T$ . Esses resultados mostram que, conforme esperado, a introdução de ruído não afeta significativamente o desempenho do método.

### 10.2.2. Segundo Ensaio

Consideremos agora um caso mais complexo, em que adotamos uma planta polinomial inspirada num exemplo discutido em [Mathews, 1991]:

$$y(n) = ay(n-1) + by(n-2)x(n-1) + cx(n-1) \quad (10.7)$$

A variância do sinal de entrada, um processo gaussiano e branco, é igual a 0.2, e  $[a \ b \ c] = [0.9 \ -0.7 \ 0.5]$ . Assumimos que o ruído de medida tem variância igual a 0.1. Para facilitar o trabalho de análise, supomos que o modelo é idêntico à planta, fato que nos permite conhecer de antemão a solução ideal. Os parâmetros escolhidos para a rede imunológica se encontram na Tab. 10.5.

<sup>154</sup> E, como se trata de um problema de minimização, transformada segundo a equação (5.1).

Parâmetro	Valor
População Inicial	3
$\sigma_s$	0.4
$N_C$	2
$\beta$	90
Critério de Parada ( <i>Fitness</i> )	0.73

**Tabela 10.5:** Parâmetros da Rede Imunológica – Planta Polinomial

Uma vez mais, foi obtida a melhor solução possível em todas as realizações. O valor médio de  $D$  foi 0.0185 e uma estimativa típica,  $[0.9075 \ -0.6989 \ 0.4780]^T$ . Esses resultados mostram que o paradigma proposto é efetivo também quando se adota um modelo não-linear, o que, em verdade, reflete o fato de que a ferramenta de busca tem um *modus operandi* transparente às características da estrutura, dependendo exclusivamente da existência de uma função de *fitness* capaz de avaliar uma solução qualquer. A complexidade adicional trazida pela não-linearidade do modelo polinomial não causou problemas à metodologia proposta.

### 10.2.3. Conclusão

Os resultados obtidos mostram que o uso de uma ferramenta de busca como a opt-aiNet é capaz de produzir um paradigma de identificação supervisionada apto a transcender o dilema essencial estabelecido pelos métodos de *output-error* e *equation-error*. Superando a limitada capacidade de busca das técnicas de gradiente, a rede imunológica propicia a elaboração de uma técnica que alia as melhores características das duas formulações existentes, ilustrando de forma marcante que o paradigma por nós delineado ao longo deste trabalho também fornece alternativas enriquecedoras *do ponto de vista conceitual*. Os resultados, que tangem os dois tipos de planta recursiva mais usuais na literatura, confirmam essa assertiva.

## 11. Emprego de Algoritmos Genéticos no Projeto Não-supervisionado do Receptor de Máxima Verossimilhança

Até agora, vimos estudando o problema de recuperar a informação contida num sinal distorcido sob o clássico prisma de equalização i.e. através de um dispositivo de filtragem destinado a reconstruir a mensagem transmitida símbolo-a-símbolo. Conforme discutimos na seção 3.4.5, há uma abordagem um tanto distinta desse problema, baseada na recuperação de uma *seqüência de dados enviados* a partir de um conjunto de dados recebidos. A chave para o processo é, a partir do sinal corrompido, buscar o *ensemble* de símbolos mais verossímil, ou seja, aquele que, considerado o vetor recebido, constitui a hipótese condicional de transmissão mais provável. Supondo que haja uma equiprobabilidade de transmissão de todas as possíveis seqüências, esse critério será ótimo do ponto de vista da mínima probabilidade de erro.

Um tal nível de desempenho, por si só, seria suficiente para que se considerasse a relevância do *critério de máxima verossimilhança (CMV)*, mas há ainda um fator adicional: não se exige conhecimento algum das amostras enviadas, ou seja, não ocorre treinamento supervisionado. Trata-se de um ponto importantíssimo, pois o CMV garante “o melhor dos dois mundos”: performance irrepreensível e formulação cega. Não seria exagero afirmar que o potencial dessa abordagem, se plenamente explorado, levaria a um paradigma modelar para equalização cega [Attux et al., 2004b].

Há, no entanto, um aspecto que não pode ser desconsiderado: para garantir um nível de desempenho que justificasse plenamente o paradigma, seria preciso testar todas as possíveis seqüências<sup>155</sup>, sendo a dimensão dessa empreitada exponencialmente aumentada com um aumento do tamanho da seqüência. Isso revela quão desafiadora será a tarefa de um processo de busca e, conseqüentemente, explicita a principal limitação do paradigma. Ante tal dificuldade, surgem propostas interessantes, dentre as quais se destaca o *algoritmo de Viterbi*, capaz de reduzir a complexidade do processo tendo em mãos uma estimativa do canal [Proakis, 1995]. Mesmo assim, permanece a “dependência explosiva” do tamanho do canal e, mais uma vez, do tamanho da seqüência adotada. Além disso, introduz-se um novo fator: o processo de identificação.

Em problemas dessa espécie, cuja solução exaustiva se torna excessivamente onerosa ou mesmo inviável, é bastante útil poder contar com ferramentas de busca que aliem uma significativa capacidade de busca a um *modus operandi* genérico e flexível. Como vimos no capítulo 5, ferramentas baseadas em computação evolutiva se encaixam perfeitamente nesse modelo, sendo, portanto, candidatas naturais à tarefa de buscar a seqüência de máxima verossimilhança. Tal idéia é suportada por esforços como o de Chen e Wu [Chen e Wu, 1998], que utilizam um AG para efetuar a *estimação conjunta da seqüência de MV e do canal*.

Neste capítulo, apresentaremos uma proposta diferente: concentrar-nos-emos apenas na busca da seqüência de MV, dispensando qualquer estimação explícita do canal. Isso nos permitirá tornar o problema de busca mais simples no que diz respeito à sua formulação,

---

<sup>155</sup> Na verdade, basta testar metade das seqüências, pois se trata de um critério cego, ou seja, “insensível” a uma inversão completa de símbolos.

pois haverá um único espaço de parâmetros: o formado pelo conjunto de possíveis seqüências. Para atingir tal meta, faz-se necessário que formulemos o critério MV de forma a dispensar qualquer dependência explícita em relação ao canal. Vejamos, pois, como isso pode ser feito.

Primeiramente, suponhamos, sem perda de generalidade, que o sinal transmitido é binário e formado por amostras i.i.d. +1 e -1. Considerando que o canal é um filtro FIR e que existe AWGN, teremos um sinal recebido igual a:

$$\mathbf{x}(n) = \sum_{k=0}^M \mathbf{h}(k)s(n-k) + r(n) \quad (11.1)$$

onde  $\mathbf{h}(n)$  é a resposta ao impulso do canal e  $r(n)$  é um processo AWGN. Seja  $p(\mathbf{x}/\mathbf{s}, \mathbf{h})$  a função de verossimilhança (FV), ou seja, a probabilidade de recebimento de um vetor de entrada  $\mathbf{x}$  dado que foi transmitida a correspondente seqüência  $\mathbf{s}$  por um canal com vetor de coeficientes  $\mathbf{h}$ . Sendo as amostras do ruído independentes, é possível mostrar que a obtenção da seqüência ótima tem por base a maximização de [Attux et al., 2004b]

$$J_{MV} = \ln[p(\mathbf{x} / \hat{\mathbf{s}}, \hat{\mathbf{h}})] = -\|\mathbf{x} - \hat{\mathbf{x}}\|^2 + C \quad (11.2)$$

sendo  $C$  uma constante,

$$\hat{\mathbf{x}}(n) = \sum_{k=0}^M \hat{\mathbf{h}}(k)\hat{s}(n-k) \quad (11.3)$$

e  $\hat{\mathbf{h}}(n)$  e  $\hat{s}(n)$  as estimativas de  $\mathbf{h}(n)$  e  $s(n)$ , respectivamente. Em notação matricial, (11.3) se torna

$$\hat{\mathbf{x}} = \hat{\mathbf{h}}\hat{\mathbf{S}} \quad (11.4)$$

com

$$\hat{\mathbf{S}} = \begin{bmatrix} \hat{s}_1 & \hat{s}_2 & \cdots & \hat{s}_k & \cdots & 0 \\ 0 & \hat{s}_1 & \cdots & \hat{s}_{k-1} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \hat{s}_{k-M} & \cdots & \hat{s}_N \end{bmatrix} \quad (11.5)$$

Analisando a equação (11.2), vemos que a maximização da função custo mostrada corresponde à minimização de

$$J_{MV2} = \|\mathbf{x} - \hat{\mathbf{x}}\|^2 \quad (11.6)$$

ou seja, de uma medida de distância<sup>156</sup>. Juntando (11.6) e (11.4), podemos reescrever  $J_{MV2}$  como

$$J_{MV2} = \left( \|\mathbf{x}\|^2 - \mathbf{x}\hat{\mathbf{S}}^T (\hat{\mathbf{S}}\hat{\mathbf{S}}^T)^{-1} \hat{\mathbf{S}}\mathbf{x}^T \right) \quad (11.7)$$

A minimização de  $J_{MV2}$ , por sua vez, corresponde à *maximização* de nossa função objetivo final:

$$J_{MVfinal} = \mathbf{x}\hat{\mathbf{S}}^T (\hat{\mathbf{S}}\hat{\mathbf{S}}^T)^{-1} \hat{\mathbf{S}}\mathbf{x}^T \quad (11.8)$$

Nesse contexto, a meta do processo de busca é encontrar a seqüência de símbolos que leva ao maior valor de  $J_{MVfinal}$ . Para realizar tal tarefa, escolhemos um AG com codificação binária, seleção baseada em torneio, *crossover* uniforme e mutação convencional. A razão de ser de nossa escolha é aliar o bom potencial de manutenção de diversidade associado ao método de seleção à significativa capacidade de busca do par *crossover* / mutação. Como há alguns mecanismos no AG, como a geração de uma população intermediária, que convém expor em detalhe, apresentamos um sumário da técnica na Tab. 11.1. A opção pelo estabelecimento de uma população intermediária tem uma causa fundamental: a manutenção de diversidade. Nossa ênfase nesse ponto será facilmente compreendida se relembrarmos a amplitude do espaço de busca envolvido, o qual, sem sombra de dúvida, deverá ser explorado de maneira eficiente. O mecanismo descrito no passo 2.4 da Tab. 11.1 é uma nítida manifestação desse ponto: a valorização dos indivíduos medianos e ruins expressa uma preocupação com a representatividade do conjunto de soluções. Por fim, à luz de tais precauções, a política elitista mostrada no passo 2.5 deverá ter um efeito eminentemente positivo.

Para testar a proposta, escolhemos dois canais, C1 e C2, cujos parâmetros mostramos em (11.9).

$$\begin{cases} \text{C1 : } H_1(z) = 0.8575 + 0.5145z^{-1} \\ \text{C2 : } H_2(z) = 0.5 + 0.7z^{-1} + 0.5z^{-2} \end{cases} \quad (11.9)$$

O primeiro é um canal de fase mínima similar a outros por nós empregados anteriormente. O segundo, por sua vez, é um sistema bem mais complexo, dotado de dois zeros sobre a circunferência de raio unitário [Proakis, 1995]. Para que possamos saber *a priori* as características da solução ótima, optamos por considerar que não existe ruído aditivo nos ensaios. Assumiremos que o tamanho da seqüência a estimar é, em ambos os casos, igual a 150 bits, o que gera um espaço do busca com  $2^{(150-1)} \approx 7.136 \times 10^{44}$  possibilidades, um valor que torna inviável o uso de um método exaustivo.

---

<sup>156</sup> Assim como na seção 2.10, o critério MV nos leva a uma medida de distância mínima.

## ***1. Inicialização***

- Crie aleatoriamente os  $N_{ind}$  indivíduos da população inicial. Cada indivíduo consiste de uma seqüência candidata.

## ***2. Enquanto um Critério de Parada Não For Satisfeito,***

### *2.1 – Crie uma População Intermediária com $2N_{ind}$ Indivíduos*

- Selecione  $1.8N_{ind}$  indivíduos através de um torneio binário.
- Insira  $0.2N_{ind}$  indivíduos gerados aleatoriamente.

### *2.2 – Crossover Uniforme*

- Divida os indivíduos da população intermediária em pares de genitores.
- Escolha aleatoriamente  $N/2$  posições do cromossomo ( $N$  é o tamanho da seqüência).
- Faça um intercâmbio dos bits associados às posições escolhidas, produzindo, dessa maneira, dois filhos.
- Faça os filhos ocuparem os lugares dos pais.

### *2.3 – Mutação*

- Modifique, com probabilidade  $p_m$ , bits pertencentes a indivíduos da população intermediária.

### *2.4 – Seleção Final*

- Selecione metade dos indivíduos pertencentes ao grupo formado pelos melhores 10% da população intermediária.
- Selecione metade dos indivíduos pertencentes ao grupo formado pelos piores 40% da população intermediária.
- Selecione metade dos indivíduos pertencentes ao grupo formado pelos restantes 50% da população intermediária.

### *2.5 – Elitismo*

- Reponha o melhor indivíduo da população se ele tiver sido perdido.

## ***Fim do Laço***

**Tabela 11.1:** Sumário do AG Escolhido

### 11.1. Primeiro Ensaio

No primeiro ensaio, escolhemos como base de testes o canal C1. Os parâmetros do processo de busca se encontram na Tab. 11.2.

Parâmetro	Valor
Número de Indivíduos ( $N_{ind}$ )	180
Probabilidade de Mutação ( $p_m$ )	0.002
Número de Indivíduos por Torneio <sup>157</sup>	2

Tabela 11.2: Parâmetros do AG – Primeiro Ensaio

Em todas as 20 realizações, obteve-se, ao final do processo, a seqüência transmitida, havendo sido, portanto, atingida uma taxa de convergência global de 100%. O número médio de iterações foi de 118.9, e o número máximo, de 269, valores plenamente compatíveis com o desempenho de uma técnica clássica de adaptação. O número médio de avaliações do custo de indivíduos foi de 42984, um valor irrisório ante o número de possíveis soluções,  $7.136.10^{44}$ .

Para ilustrar o comportamento dinâmico da técnica de busca, apresentamos na Fig. 11.1 a evolução temporal dos custos máximo e médio para uma realização típica. Observamos que, em cerca de 80 iterações, atinge-se um nível de desempenho próximo ao ótimo, sendo o restante das iterações devotado a uma sorte de refinamento. O comportamento da curva média acompanha as tendências da curva máxima, fazendo-o, entretanto, à distância, fato que indica um bom nível de diversidade populacional. De posse de todos esses dados, só podemos concluir que a metodologia proposta teve a si associado um excelente desempenho, correspondente, sem sombra de dúvida, a nossos objetivos de projeto e expectativas de desempenho.

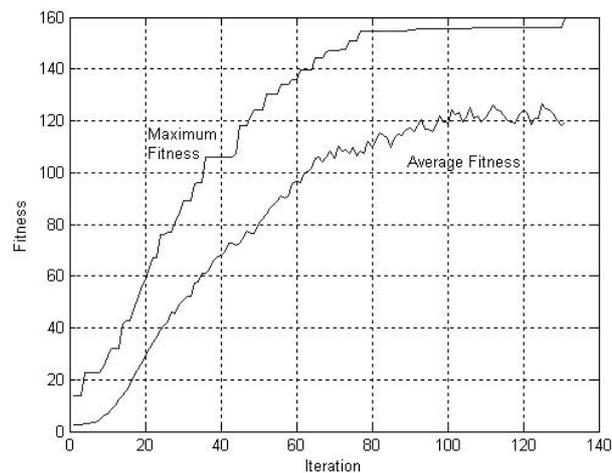


Figura 11.1: Evolução Temporal do *Fitness* – Primeiro Ensaio

<sup>157</sup> Vide seção 5.3.4.

## 11.2. Segundo Ensaio

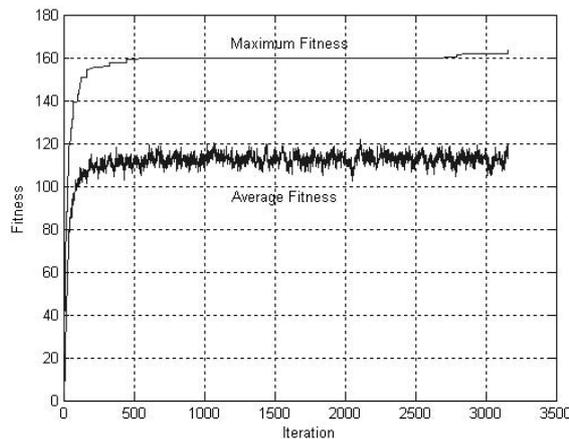
No segundo ensaio, lidamos com o canal C2, o qual origina um cenário bem mais complexo que o anterior. Os parâmetros do AG se encontram na Tab. 11.3.

Parâmetro	Valor
Número de Indivíduos	180
Probabilidade de Mutação	0.005
Número de Indivíduos por Torneio	2

**Tabela 11.3:** Parâmetros do AG – Segundo Ensaio

Novamente, obtivemos a seqüência ótima em todas as iterações. Para tanto, foi necessário um número médio de 1802.8 iterações e um número máximo de 5098 iterações. Uma vez mais, as cifras indicam uma flutuação satisfatória, restrita a uma mesma ordem de grandeza. Nesse cenário mais complexo, fez-se uso, em média, de 649188 cálculos da função custo, valor novamente irrisório ante as  $7.136.10^{44}$  possibilidades. O aumento das cifras em relação ao ensaio anterior se justifica pela maior complexidade do canal C2, que tem por reflexo uma função custo mais rica e plena de soluções mais parelhas.

Na Fig. 11.2, mostramos o comportamento dinâmico do custo máximo e do custo médio associados à população. Novamente, verificamos que a evolução se dá de forma bastante acelerada, sendo nítida a distinção entre as duas fases principais. A curva média acompanha a tendência da curva máxima de uma distância apropriada, o que, como dissemos há pouco, indica que a diversidade populacional é mantida. Isso nos leva a concluir que, também neste segundo ensaio, o AG levou a um desempenho muito bom.



**Figura 11.2:** Evolução Temporal do *Fitness* – Segundo Ensaio

### 11.3. Discussão

Neste capítulo, analisamos uma nova proposta para a obtenção da seqüência de máxima verossimilhança *sem a necessidade de estimar o canal*, usando, para isso, um algoritmo genético. Uma proposta desse tipo leva, de maneira indireta, a um *esquema de equalização cega ótimo do ponto de vista da minimização da taxa de erro de bit*, meta que, sem dúvida, sumaria as mais altas aspirações de um projetista ante tal problema.

Para fazer frente a tão amplas expectativas, escolhemos um algoritmo genético confeccionado com o fim de aliar capacidade de busca local, manutenção de diversidade e alta velocidade de convergência<sup>158</sup>. Escolhemos dois cenários de teste e, em ambos os casos, a técnica proposta teve um desempenho bastante satisfatório. Podemos apontar ao menos três razões subjacentes a esse parecer favorável:

- Foi atingida uma taxa de convergência global de 100% em ambos os cenários.
- O número de cálculos da função custo requeridos para atingir tal desempenho foi várias ordens de grandeza inferior àquele associado a um método exaustivo, fato que explicita a “inteligência” e a eficiência do processo de busca.
- O número médio de iterações requeridas para convergência foi plenamente compatível com os valores manifestos num contexto clássico de filtragem.

Com isso, afirmamos que, seguramente, o uso de uma ferramenta evolutiva de busca *pode trazer um significativo ganho de desempenho ao processo de obtenção da seqüência de máxima verossimilhança*, especialmente em cenários nos quais o tamanho da seqüência ou do canal eleva a complexidade do problema resultante a patamares proibitivos. Do ponto de vista do paradigma geral esboçado no prólogo à segunda parte, a generalidade do problema abordado é expressão direta do valor dos resultados expostos neste capítulo.

Finalmente, há duas extensões imediatas e necessárias deste trabalho: abordar o caso em que há ruído AWGN e considerar cenários com ordens de canal e seqüência mais elevadas [Attux et al., 2004b].

---

<sup>158</sup> Frisamos que velocidade de convergência, de certa forma, significa *parcimônia no uso dos recursos computacionais*.



## 12. Uma Proposta para Equalização Não-linear Cega Baseada num Filtro de Erro de Predição Fuzzy

O problema de equalização não-linear cega, apresentado na seção 3.4.3, ainda é, fundamentalmente, uma questão em aberto. A razão disso é que os mais sólidos pilares da teoria de desconvolução autodidata – os teoremas de Benveniste-Goursat-Ruget e de Shalvi-Weinstein e os critérios de Bussgang [Haykin, 1996] – têm sua validade restrita a um domínio em que tanto canal quanto equalizador são filtros lineares. Uma extensão direta desses resultados para um domínio estrutural mais genérico produz uma situação em que lógica e incoerência coexistem sem que haja a possibilidade de distingui-las (vide exemplo 3.5).

Dificuldades dessa magnitude sugerem fortemente que seria interessante buscar uma abordagem alheia à formulação diretamente baseada em estatísticas de ordem superior. Analisando as possibilidades, não nos é possível desconsiderar que a resposta pode estar num critério que não tolera outro adjetivo senão “clássico”: o *critério de erro de predição*. Conforme vimos na seção 3.6.1, a idéia seria usar um filtro de erro de predição (FEP) como equalizador, o que permitiria a recuperação de um sinal com mínimo ou com máximo atraso<sup>159</sup>.

São bem conhecidas as limitações dessa abordagem, ao menos no contexto linear clássico: *o filtro de erro de predição irá branquear o sinal*, e, genericamente, *branquear e equalizar não são sinônimos*. Por exemplo, um FEP que tentasse compensar os efeitos de um canal de fase mista terminaria por resolver apenas o problema do espectro de amplitude, permanecendo, destarte, uma distorção de fase. A conclusão usual é: restrinja-se a aplicação desse dispositivo a canais de fase mínima, pois qualquer outra opção deverá nos conduzir ao universo dos critérios baseados em estatísticas de ordem superior. Essa constatação, no entanto, omite um ponto crucial: assume-se, tacitamente, *que uma estrutura linear é a única opção possível*. O questionamento de tal assunção abre uma possibilidade instigante: limitações atribuídas *ao critério de erro de predição* podem dizer respeito, em verdade, *às características do dispositivo adotado*. Para ilustrar esse ponto, voltemos novamente nosso olhar à seção 3.6.1: nela, mostramos que o critério de erro de predição leva a um equalizador com atraso nulo. Ora, quando discutimos a equalização como um problema de classificação, vimos que há canais que, sob tal atraso, produzem *problemas linearmente insolúveis* (vide o exemplo 3.7). Conseqüentemente, a impossibilidade de equalizar módulo e fase atribuída ao FEP linear é, nesse caso, o simples reflexo de que *é inviável separar linearmente os estados para o atraso requerido*. Daí o mérito atribuído aos critérios de Bussgang: como vimos em 3.4.3, eles são capazes de “considerar outros atrasos”, desembocando, eventualmente, numa configuração linearmente separável.

Após a análise desse estado de coisas, emerge uma questão natural: por que não tomar o outro caminho? Por que, em vez de afastar o critério de predição, não adotamos *uma estrutura de filtragem mais flexível, capaz de separar os estados do canal mesmo se o atraso nulo levar a um problema complexo*? Essa é exatamente a proposta de Cavalcante

---

<sup>159</sup> Caso se use um preditor ou um retro-preditor. Doravante trataremos, sem perda de generalidade, apenas da primeira alternativa.

[Cavalcante, 2001], que mostra que o critério de erro de predição, ao contrário do que ocorre com os critérios de Bussgang, mantém sua coerência mesmo quando se usa uma estrutura do tipo MLP. Esse resultado pioneiro abriu um importante caminho, e é exatamente nele que se insere a contribuição que relataremos neste capítulo: mantendo a idéia de usar o critério de erro de predição, procuraremos estabelecer um paradigma para equalização não-linear cega baseado numa outra estrutura, um filtro *fuzzy*<sup>160</sup>, que, por sua capacidade de incorporar as características do ruído, deve levar a um desempenho bastante sólido em termos de probabilidade de erro. Em suma, nossa proposta é projetar um preditor *fuzzy* e analisar o desempenho do equalizador nele baseado sob diferentes canais e na presença de ruído. Tomaremos por base fundamental de comparação o desempenho do filtro ótimo de Bayes e de um FEP linear, que, de certa forma, são dois extremos entre os quais deve se situar a nova abordagem.

### 12.1. Aspectos Gerais de Projeto

Basicamente, o problema que teremos de abordar é o de projetar um filtro *fuzzy* com a forma mostrada em (4.33) segundo o critério de erro de predição mostrado em (3.76). Norteados pelas características gerais de treinamento esboçadas na seção 4.4.1, adotaremos um procedimento de três etapas: na primeira, obteremos os centros das regras da parte *se*; na segunda, escolheremos os centros das regras da parte *então*; e, na terceira, submeteremos os parâmetros obtidos na segunda etapa a um processo de refinamento via RLS [Wang e Mendel, 1993].

A base do primeiro estágio é o uso de um algoritmo K-means modificado [Chinrungrueng e Sequin, 1995]<sup>161</sup> na estimação dos estados escalares do canal<sup>162</sup>, partindo de uma inicialização aleatória no intervalo  $[-1,+1]$ . Nesse passo inicial, são também arbitrados valores heurísticamente determinados para a variância das funções de pertinência gaussianas.

Na segunda etapa, cria-se efetivamente a base de regras de nosso filtro. Para atingir tal meta, usamos o processo de geração de regras descrito em [Wang e Mendel, 1992], que pode ser descrito da seguinte maneira:

- Dado um vetor de entrada do conjunto de treinamento do preditor,  $\mathbf{x}(n-1) = [x(n-1) \ x(n-2) \ \dots \ x(n-m+1)]^T$ , calcule a pertinência de cada elemento  $x(n-i)$ ,  $i = 1 \dots m-1$ , segundo cada um dos conjuntos gaussianos de entrada. Associe ao  $i$ -ésimo elemento o conjunto que produzir a maior medida de pertinência. À parte *se* assim produzida, junte a parte *então* formada pelo valor faltante  $x(n)$ <sup>163</sup>, que corresponde ao peso  $w_i$  mostrado em (4.33). Repita o processo até que sejam visitadas todas as amostras do conjunto de treinamento.

<sup>160</sup> Tendo em mente a correspondência entre um filtro *fuzzy* e uma rede RBF, não seria despropositado afirmar que, mesmo que de maneira indireta, este capítulo contempla a aplicação de ambas estruturas.

<sup>161</sup> Tal algoritmo foi mencionado na seção 4.3.1. Sua principal diferença em relação ao K-means clássico é a existência de mecanismos para melhoria de convergência.

<sup>162</sup> Essa opção de trabalhar com os estados escalares é motivada pela idéia de atribuir a *cada um dos elementos do vetor de entrada* um conjunto de regras, o que mantém a “estrutura conceitual” do dispositivo (discutida na seção 4.4.1).

<sup>163</sup> Aqui o projeto se caracteriza como o de um preditor.

- Na base de regras gerada no passo anterior, haverá regras conflitantes, ou seja, regras com a mesma parte *se* e partes *então* diferentes<sup>164</sup>. Impasses desse tipo são resolvidos fundindo as regras conflitantes numa única regra com parte *então*  $w_i$  igual à média aritmética das várias partes componentes.

Por fim, faz-se uso de uma terceira e última etapa de refinamento dos centros, variâncias e parâmetros da parte *então* através de um processo de aprendizado conduzido por um algoritmo RLS clássico (vide seção 3.4.2) baseado no critério de erro de predição. Com esse estágio chega ao fim o processo de treinamento. O próximo passo de nosso estudo é testar a proposta em cenários representativos.

## 12.2. Primeiro Ensaio

No primeiro cenário de teste escolhido, como no resto do capítulo, assumiremos que o sinal transmitido é formado por amostras i.i.d. extraídas de uma constelação 2-PAM ou 2-PSK em banda-base. Sempre consideraremos que as distorções impostas pelo meio podem ser adequadamente modeladas por um sistema linear com ruído AWGN. Por ora, escolhemos o seguinte modelo de canal de fase mínima:

$$C1: H_1(z) = 0.8354 + 0.5012z^{-1} + 0.2256z^{-2} \quad (12.1)$$

A razão para essa escolha não é fortuita: sob um modelo de fase mínima, o esquema de equalização baseada em predição, conforme vimos na seção 3.6.1, tende a ter seu melhor desempenho. Nesse bom caso, emerge um problema linearmente solúvel, o que nos leva a intuir duas coisas: o filtro de erro de predição linear deve ter um desempenho razoável, mesmo na presença de ruído, e o filtro de erro de predição *fuzzy* por nós proposto deve ter um comportamento próximo ao do filtro de Bayes. A primeira dessas conclusões deve estar clara ao leitor, pois se funda explicitamente numa discussão empreendida na primeira parte desta tese. A segunda conclusão, no entanto, talvez não esteja tão clara, e é nosso dever explicá-la em mais detalhe. Supondo que o filtro de erro de predição tenha ordem três, ele terá uma relação entrada-saída como a seguinte:

$$y(n) = x(n) - \mathfrak{F}[x(n-1), x(n-2), x(n-3)] \quad (12.2)$$

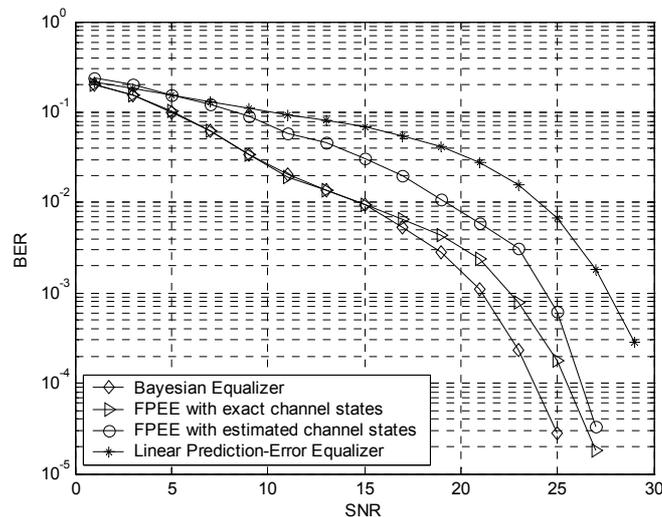
sendo  $\mathfrak{F}[\cdot]$  um mapeamento não-linear genérico. Uma inspeção de (12.2) nos leva a uma conclusão muito importante: embora o filtro conte com quatro entradas, *apenas três delas são processadas pelo mapeamento não-linear*. Em outras palavras, embora o filtro seja não-linear e de ordem três, ele tem uma estrutura híbrida, com uma entrada que se relaciona linearmente com a saída e outras três que, efetivamente, experimentam o caráter não-linear do filtro. Por conseguinte, um FEP *fuzzy* de ordem três deve ter um desempenho inferior a

---

<sup>164</sup> Uma breve inspeção da Tab. 3.3 ilustra a inexorabilidade desse fato: note que, por exemplo, se tomarmos  $x(n-1) = 1.6$ , teremos duas possibilidades para  $x(n)$ : 1.6 e -0.4. O cenário se mantém inalterado caso aumentemos a ordem do canal ou do preditor.

um filtro de Bayes de mesma ordem<sup>165</sup>. O caso em que essa diferença tende a ser menos pronunciada é exatamente aquele em que o canal é de fase mínima, pois é o caso em que o problema de separação subjacente ao processo de predição é mais simples.

Para avaliar o desempenho de nossa proposta, trataremos de compará-la a outras duas estruturas representativas: um FEP linear e o equalizador de Bayes. A base de toda a avaliação será uma medida de taxa de erro de bit, levantada após o processo de treinamento a partir de um máximo de  $10^6$  amostras de teste. Assumiremos que todos os dispositivos têm três entradas e que o projeto do FEP *fuzzy* segue o esquema tríplice exposto no final da seção passada, sendo alocadas 500 amostras para os dois primeiros passos e 500 para o passo de refinamento. Na Fig. 12.1, apresentamos quatro curvas SNR x BER: uma para o FEP linear, uma para o Bayesiano, uma para o FEP *fuzzy* projetado segundo o receituário discutido na seção 12.1 e uma para uma versão ideal do FEP *fuzzy*, na qual todos os parâmetros assumem seus melhores valores. O motivo de termos levantado duas curvas para o FEP proposto é que tal procedimento nos permite ter uma clara percepção do desempenho inerente à estrutura e da influência do esquema de treinamento.



**Figura 12.1:** Desempenho dos Equalizadores – Primeiro Cenário

A Fig. 12.1 nos mostra que o desempenho do FEP *fuzzy*, como era de se esperar, se situa entre o do limitante de Bayes e o do FEP linear. Em sua versão ideal, o FEP proposto atinge patamares de erro próximos àqueles produzidos pelo filtro ótimo, em perfeita consonância com o que prevíamos. O que mais nos chama a atenção é a perda de desempenho acarretada pelo processo de treinamento, o que nos leva a duas conclusões naturais:

- O processo de treinamento a partir de amostras ruidosas tem um significativo e inerente grau de dificuldade.

<sup>165</sup> Julgamos conveniente desconsiderar, por um momento, a diferença inerente entre os critérios de erro quadrático médio e de Bayes.

- O esquema proposto pode ser melhorado em todas as suas esferas: processos de clusterização, geração de regras e refinamento. Alguns esforços nesse sentido podem ser encontrados em [Ferrari, 2005].

Finalmente, o desempenho do FEP linear foi satisfatório, especialmente sob altas SNRs, o que também confirma nossas previsões.

Concluimos que o primeiro ensaio revela de maneira convincente o potencial da proposta, mostrando que ela pode atingir patamares semelhantes aos do filtro de Bayes sem a necessidade de uma seqüência de treinamento supervisionado. Por outro lado, o ensaio também mostra que o processo de treinamento exposto na seção anterior não possibilitou a plena exploração desse potencial, embora o desempenho a ele associado não tenha sido, em nenhum momento, decepcionante.

### 12.3. Segundo Ensaio

O segundo ensaio por nós conduzido tem por base o seguinte canal

$$C2: H_1(z) = 0.5632 - 0.7322z^{-1} - 0.3830z^{-2}, \quad (12.3)$$

sendo as demais condições mantidas. Esse canal, propositadamente, é de fase não-mínima, fato que nos coloca numa situação em que duas previsões imediatas podem ser feitas:

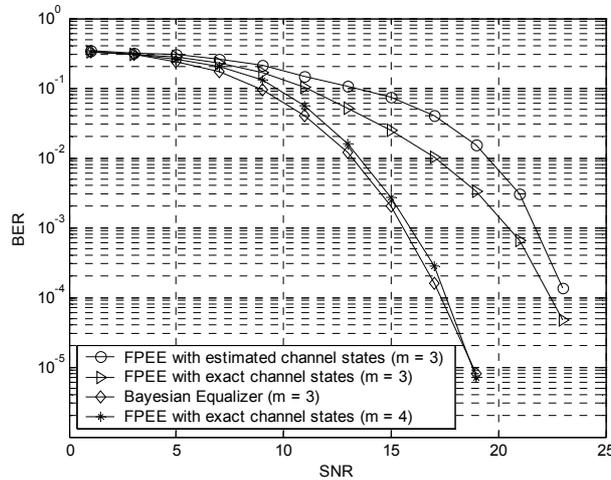
- O FEP linear deverá ter um desempenho altamente insatisfatório, o que, aliás, já é bem conhecido na teoria clássica.
- O FEP *fuzzy* deverá ter um bom desempenho, mas, devido à maior complexidade do problema e ao dilema subjacente a (12.2), existirá uma maior separação entre a curva a ele associada e aquela associada a um filtro de Bayes de mesma ordem.

A segunda conclusão nos levou a uma idéia: incluir uma avaliação de desempenho do FEP *fuzzy* com ordem imediatamente superior ( $m=4$ ), de forma a buscar uma situação em que sua parte não-linear tenha a mesma ordem do filtro de Bayes. Os resultados, obtidos segundo o mesmo receituário da seção anterior, estão na Fig. 12.2<sup>166</sup>. Primeiramente, observamos que, neste caso, a diferença entre o FEP *fuzzy* ideal e o treinado não foi tão significativa, o que pode indicar uma maior uniformidade entre as diversas realizações que originam a curva ou mesmo refletir a maior dificuldade do problema em si, que pode ter contribuído para um “nivelamento por baixo”. De qualquer maneira, trata-se de um resultado favorável ao método de treinamento, um ponto positivo que deve ser considerado juntamente com os pontos a melhorar que o ensaio anterior revela. A distância do FEP *fuzzy* ideal (com  $m=3$ ) em relação ao filtro de Bayes é mais uma evidência no sentido da fundamentação que vimos construindo (vide a segunda conclusão exposta no início da presente seção). O aumento da ordem desse dispositivo levou seu desempenho a patamares muito próximos aos do filtro bayesiano, o que também suporta nossas observações. Isso

---

<sup>166</sup> O leitor notará imediatamente a ausência de uma curva associada ao FEP linear. Como esperávamos, o desempenho desse dispositivo foi muito ruim para todas SNRs, o que tornou um tanto supérfluo o detalhamento de seu desempenho.

atesta a solidez e a coerência do paradigma de predição, mostrando também uma interessante diretiva que tem enorme valor para o projetista.



**Figura 12.2:** Desempenho dos Equalizadores – Segundo Cenário

Podemos concluir que o ensaio conduzido sob os auspícios de um modelo de fase não-mínima nos forneceu os subsídios que faltavam para que pudéssemos ter uma clara visão do potencial da abordagem não-linear de predição e das fronteiras que o delimitam. Isso nos leva a crer que tenhamos cumprido a tarefa inicial de expor e discutir a nova proposta.

## 12.4. Conclusões

Neste capítulo, apresentamos e testamos uma nova proposta para a solução de um problema crucial e, fundamentalmente, em aberto: o da equalização não-linear cega. Partindo de uma idéia originalmente concebida por Cavalcante [Cavalcante, 2001], chegamos a uma proposta baseada num dispositivo capaz de incorporar adequadamente as características do ruído<sup>167</sup>: um filtro *fuzzy*. Por meio de dois ensaios conduzidos sob modelos distintos, pudemos explicitar as principais características da nova proposta. De maneira sintética, demonstramos que o paradigma de predição é sólido e viável, e que um FEP *fuzzy* adequadamente treinado pode, sob certas circunstâncias, atingir um nível de desempenho comparável ao das melhores formulações supervisionadas. Também verificamos que o esquema de treinamento adotado, apesar de eficiente, ainda pode ser melhorado em todas as suas etapas, e que permanece deveras útil buscar maneiras ainda mais eficientes de lidar com o caráter ruidoso das amostras disponíveis.

A contribuição aqui apresentada se encaixa muito bem no espírito do paradigma delineado no prólogo à segunda parte, pois atesta tanto a possibilidade de que seja usado um critério cego em equalização não-linear quanto a adequação de uma estrutura formada por funções de base gaussianas ao problema dele resultante. Com isso, são decisivamente

<sup>167</sup> Essa informação está contida no formato das funções de pertinência.

ampliados os horizontes teóricos do projetista de um sistema de comunicação, ponto que reflete nitidamente o espírito que norteia este trabalho.

Por fim, frisamos que os resultados aqui discutidos são um esforço inicial que, aliás, foi o gérmen do importante trabalho de Ferrari [Ferrari, 2005], o qual, de maneira muito elegante, sistematiza, estende e fundamenta com mais rigor as conclusões que acabamos de tecer. Em particular, seu trabalho mostra que o preditor *fuzzy* é ótimo do ponto de vista do erro quadrático médio, um resultado que leva o preditor proposto do *status* de abordagem promissora ao de solução ótima.



## 13. Conclusão

*“You shall not look through my eyes either,  
nor take things from me,  
You shall listen to all sides and filter them for  
yourself”*

*Walt Whitman, Song of Myself, 2*

Neste trabalho, buscamos investigar de forma tão completa e sistemática quanto possível o problema de filtragem adaptativa no contexto de três tarefas fundamentais: *equalização, identificação e predição*. Tendo por ponto de partida a idéia de comunicação, empreendemos uma jornada ao longo da qual discutimos em detalhe a base teórica subjacente às três tarefas e a diversos meios para sua solução: estruturas, critérios e algoritmos para otimização e adaptação. Caminhando por terrenos mais ou menos áridos, chegamos a um destino inevitável: a constatação de que é possível sustentar uma formulação genérica do problema de filtragem desde que se disponha de *filtros suficientemente flexíveis, critérios coerentes e técnicas para otimização global*. Sob os auspícios dessa tríade, torna-se viável que um projetista seja capaz de obter uma solução apropriada a partir do mínimo de informação possível e de uma estrutura parcimoniosa. Exatamente nesse espírito concluímos a primeira parte do documento, que podemos comparar a um “diário de viandante”, repleto de descrições e ponderações acerca de rumos e da meta em busca da qual direcionamos o olhar ao horizonte.

Após uma “declaração de princípios” no prólogo, passamos, na segunda parte, à discussão de diversas contribuições permeadas pelo espírito do ideário exposto. Apresentamos diversas formulações inovadoras de problemas de equalização, identificação e predição e expusemos, em cada capítulo, as principais conclusões a que chegamos ao final de cada ensaio. Com isso, julgamos ter explorado suficientemente os meandros de cada proposta particular, sendo oportuno, neste momento, tecer um par de considerações gerais:

- Técnicas baseadas em computação evolutiva, particularmente a rede imunológica artificial opt-aiNet, formam uma base viável e eficiente para o projeto de um filtro adaptativo genérico. Tais técnicas tiveram um desempenho notável em problemas lineares, não-lineares, supervisionados e não-supervisionados, o que revela uma característica essencial para o paradigma universal que buscamos: uma imensa flexibilidade aliada a um enorme potencial de busca.
- Não seria exagero afirmar que o uso de uma ferramenta evolutiva acarreta um verdadeiro salto qualitativo no projeto de filtros digitais, o que atestam, por exemplo, a eficiência demonstrada nos delicados problemas de equalização e

identificação recursivas, a abordagem original do problema de pré-distorção e o notável ganho de desempenho no problema de busca da seqüência de máxima verossimilhança. Em todos os casos, não existe apenas um ganho de desempenho, mas há uma efetiva abertura de horizontes inusitados.

- Finalmente, no caso da contribuição específica ao problema de equalização não-linear cega (capítulo 12), chegamos também a uma contribuição que se insere no contexto do paradigma norteador, pois mostramos a efetiva possibilidade de realizar, de maneira coerente e segura, a tarefa genérica de equalização de canais *sem a necessidade de supervisão*.

Antes de passarmos às linhas finais, cumpre-nos reafirmar que o preço da generalidade proporcionada pelo algoritmo não é pequeno. O custo computacional de uma técnica evolutiva, como o leitor pode averiguar pela inspeção, por exemplo, das Tabs. 5.2 e 5.4, é bem maior que o de um algoritmo de gradiente clássico. Esse quadro é agravado pela necessidade prática de, na maioria das aplicações expostas, estimar a função custo a partir de uma média temporal baseada numa janela que, tipicamente, conteve uma ou poucas centenas de amostras. Como temos insistido ao longo deste trabalho, é importante que o projetista tenha conhecimento desse fato e o leve em conta no momento em que for tomar suas decisões de projeto, mas optamos, conscientemente, por não apresentar aqui esforços no sentido de esmiuçar a questão de complexidade computacional. Há inúmeros fatores estocásticos no cerne das técnicas evolutivas, fatores, aliás, dependentes de parâmetros heurísticamente selecionados. Tendo em vista a complexidade desse cenário e o escopo de nosso trabalho, julgamos ser mais proveitoso discutir o problema em termos simples e intuitivos.

Finalmente, gostaríamos de concluir este trabalho com a apresentação de algumas perspectivas de esforços futuros:

- Avaliar o desempenho de técnicas evolutivas bastante apropriadas à implementação, como *micro-AGs*, em problemas de equalização e identificação [Chen e McLaughlin, 1997];
- Buscar novas estratégias de estimação da função custo, sendo uma possibilidade interessante a exploração da finidade do número de estados do canal, de acordo com o espírito da seção 3.4.4.
- Incluir o problema de separação de fontes no paradigma genérico de filtragem. Para atingir tal meta, serão necessários esforços de natureza teórica, como a busca por uma formulação conjunta dos problemas de separação e equalização, e esforços de natureza prática, como a avaliação detalhada da aplicabilidade de estruturas não-lineares e ferramentas evolutivas ao problema.

## 14. Bibliografia

[Abbas et al., 1997] A. K. Abbas, A. H. Lichtman, J. S. Pober, *Cellular and Molecular Immunology*, 3<sup>rd</sup> Edition, W. B. Saunders Company, 1997.

[Aschbacher e Rupp, 2003a] E. Aschbacher, M. Rupp, “Identification of a Nonlinear Power-Amplifier LNL Structure for Pre-distortion Purposes”, Proceedings of the 37<sup>th</sup> Asilomar Conference on Signals, Systems and Computers, Pacific Grove, EUA, 2003.

[Aschbacher e Rupp, 2003b] E. Aschbacher, M. Rupp, “Modeling and Identification of a Nonlinear Power-Amplifier with Memory for Nonlinear Adaptive Pre-Distortion”, Proceedings of the SPAWC Workshop, Roma, Itália, 2003.

[Attux, 2001] R. R. F. Attux, *Sobre Dinâmica Caótica e Convergência em Algoritmos para Equalização Autodidata*, Tese de Mestrado, Unicamp, 2001.

[Attux e Romano, 2002] R. R. F. Attux, J. M. T. Romano, “Adaptive Criteria Optimization as a Least Squares Problem”, Proceedings of the IEEE International Telecommunications Symposium, Natal, Brasil, 2002.

[Attux et al., 2003a] R. R. F. Attux, M. B. Loiola, R. Suyama, L. N. de Castro, F. J. Von Zuben, J. M. T. Romano, “Blind Search for Optimal Wiener Equalizers Using an Artificial Immune Network Model”, EURASIP Journal on Applied Signal Processing, Special Issue on Genetic and Evolutionary Computation for Signal Processing and Image Analysis, Vol. 8, pp. 740-747, 2003.

[Attux et al., 2003b] R. R. F. Attux, L. N. de Castro, F. J. Von Zuben, J. M. T. Romano, “A Paradigm for Blind IIR Equalization Using the Constant Modulus Criterion and an Artificial Immune Network”, Proceedings of the IEEE Workshop on Neural Networks for Signal Processing (NNSP), Toulouse, França, 2003.

[Attux e Romano, 2003] R. R. F. Attux, J. M. T. Romano, “Chaotic Phenomena in Adaptive Blind Equalisers”, Proceedings of the IEE: Vision, Image and Signal Processing, Vol. 150, No. 6, pp. 360-364, 2003.

[Attux et al., 2004a] R. R. F. Attux, L. T. Duarte, L. N. de Castro, F. J. Von Zuben, J. M. T. Romano, “Adaptive Identification of Feedback Systems Using an Artificial Immune Network”, Proceedings of the International Conference on Telecommunications (ICT), Fortaleza, Brasil, 2004.

[Attux et al., 2004b] R. R. F. Attux, R. R. Lopes, L. N. de Castro, F. J. Von Zuben, J. M. T. Romano, “Genetic Algorithms for Blind Maximum Likelihood Receivers”, Proceedings of the IEEE Workshop on Machine Learning for Signal Processing (MLSP), São Luís, 2004.

[Babuska, sem data] R. Babuska, *Fuzzy Systems, Modeling and Identification*, Material de Apoio, Delft University of Technology.

[Barry et al., 2004] J. R. Barry, E. A. Lee, D. G. Messerschmitt, *Digital Communication*, 3<sup>rd</sup> Edition, Kluwer Academic Publishers, 2004.

[Bellafemina e Benedetto, 1985] M. Bellafemina e S. Benedetto, “Identification and Equalization of Nonlinear Channels for Digital Transmission”, Proceedings of the IEEE International Symposium on Circuits and Systems, pp. 1477-1480, Japão, 1985.

[Bendat, 1998] J. S. Bendat, *Nonlinear System Techniques and Applications*, Wiley-Interscience, 1998.

[Benedetto et al., 1976] S. Benedetto, E. Biglieri, R. Daffara, “Performance of Multilevel Baseband Digital Systems in Nonlinear Environments”, IEEE Trans. on Communications, Vol. COM-24, pp. 1166-1175, 1976.

[Benedetto et al., 1979] S. Benedetto, E. Biglieri, R. Daffara, “Modeling and Performance Evaluation of Nonlinear Satellite Links – a Volterra Series Approach”, IEEE Trans. on Aerospace and Electronic Systems, Vol. AES-15, pp. 494-507, 1979.

[Benedetto e Biglieri, 1983] S. Benedetto, E. Biglieri, “Nonlinear Equalization of Digital Satellite Channels”, IEEE Journal of Selected Areas on Communications, Vol. SAC-1, pp.57-62, 1983.

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

[Boyer, 1991] C. B. Boyer, *A History of Mathematics*, Wiley, 1991.

[Campello, 2002] R. J. G. B. Campello, *Arquiteturas e Metodologias para Modelagem e Controle de Sistemas Complexos Utilizando Ferramentas Clássicas e Modernas*, Tese de Doutorado, Unicamp, 2002.

[Cardoso et al., 2000] F. A. C. M. Cardoso, D. S. Arantes, J. M. T. Romano, J. B. Destro Filho, “Uma Versão Evolutiva do Algoritmo de Godard”, XVIII Simpósio Brasileiro de Telecomunicações, Gramado, Brasil, 2000.

[Casas, 1996] R. A. Casas, *Blind Adaptive Decision Feedback Equalization: a Class of Bad Channels*, Tese de Mestrado, Cornell University, 1996.

[Casas et al., 1998] R. A. Casas, C. R. Johnson, R. A. Kennedy, Z. Ding, R. Malamut, “Blind Adaptive Decision Feedback Equalization: a Class of Channels Resulting in Ill-convergence from a Zero Initialization”, International Journal on Adaptive Control and Signal Processing, Vol. 12, No. 2, pp.173-193, 1998.

[Castanié e Roviras, 2002] F. Castanié, D. Roviras, “Neural Networks in Space Communications”, 14th International Conference on Digital Signal Processing, Vol. 1, pp. 3-7, 2002.

[de Castro, 1998] L. N. de Castro, *Análise e Síntese de Estratégias de Aprendizado para Redes Neurais Artificiais*, Tese de Mestrado, Unicamp, 1998.

[de Castro e Von Zuben, 2000] L. N. de Castro, F. J. Von Zuben, “An Evolutionary Immune Network for Data Clustering”, Anais do Simpósio Brasileiro de Redes Neurais (SBRN '00), pp. 84-89, 2000.

[de Castro, 2001] L. N. de Castro, *Engenharia Imunológica: Desenvolvimento e Aplicação de Ferramentas Computacionais Inspiradas em Sistemas Imunológicos Artificiais*, Tese de Doutorado, Unicamp, 2001.

[de Castro e Timmis, 2002a] L. N. de Castro, J. Timmis, *Artificial Immune Systems: a New Computational Intelligence Approach*, Springer, 2002.

[de Castro e Timmis, 2002b] L. N. de Castro, J. Timmis, “An Artificial Immune System for Multimodal Function Optimization”, Proceedings of the IEEE Congress on Evolutionary Computation, pp. 669-674, 2002.

[de Castro e Von Zuben, 2002a] L. N. de Castro, F. J. Von Zuben, *Notas de Aula do Curso de IA707 – Computação Evolutiva*, Unicamp, 2002.

[de Castro e Von Zuben, 2002b] L. N. de Castro, F. J. Von Zuben, “Learning and Optimization Using the Clonal Selection Principle”, IEEE Transactions on Evolutionary Computation, Vol. 6, No. 3, pp. 239-251, 2002.

[Cavalcante, 2001] C. C. Cavalcante, *Predição Neural e Estimação de Função de Densidade de Probabilidade Aplicadas à Equalização Cega*, Tese de Mestrado, Universidade Federal do Ceará, 2001.

[Chang e Ye, 1994] P. Chang, B. Yeh, “Nonlinear Communication Channel Equalization using Wavelet Neural Networks”, Proceedings of the IEEE International Conference on Neural Networks, Vol. 6, pp. 3605-3610, 1994.

[Chen et al., 1992] S. Chen, B. Mulgrew, S. McLaughlin, “Adaptive Bayesian Decision Feedback Equaliser Based on a Radial Basis Function Network”, Proceedings of the International Conference on Communications (ICC), Vol. 3, pp. 1267-1271, 1992.

[Chen e McLaughlin, 1997] Blind Channel Identification Based on Higher-order Cumulant Fitting Using Genetic Algorithms”, Proceedings of the IEEE Signal Processing Workshop on Higher-order Statistics, pp. 184-188, 1997.

- [Chen et al., 1997] S. Chen, Y. Wu, S. McLaughlin, “Genetic Algorithm Optimization for Blind Channel Identification with Higher-order Cumulant Fitting”, *IEEE Transactions on Evolutionary Computation*, Vol. 1, No. 4, pp. 259-265, 1997.
- [Chen e Wu, 1998] S. Chen, Y. Wu, “Maximum Likelihood Joint Channel and Data Estimation Using Genetic Algorithms”, *IEEE Transactions on Signal Processing*, Vol. 46, No. 5, pp. 1469-1475, 1998.
- [Chinrungrueng e Sequin, 1995] C. Chinrungrueng, C. H. Sequin, “Optimal Adaptive K-means Algorithm with Dynamic Adjustment of Learning Rate”, *IEEE Transactions on Neural Networks*, Vol. 6, pp. 157-169, 1995.
- [Costa et al., 2002] A. M. Costa, R. R. F. Attux, J. M. T. Romano, “A New Method for Blind Channel Identification with Genetic Algorithms”, *Proceedings of the IEEE International Telecommunications Symposium (ITS)*, Natal, Brasil, 2002.
- [Ding et al., 1991] Z. Ding, R. A. Kennedy, B. D. O. Anderson, C. R. Johnson, “Ill-convergence of Godard Blind Equalizers in Data Communication Systems”, *IEEE Transactions on Signal Processing*, vol. 39, pp. 1313-1327, 1991.
- [Doyle et al., 2002] F. J. Doyle III, R. K. Pearson, B. A. Ogunnaike, *Identification and Control Using Volterra Models*, Springer-Verlag, 2002.
- [Duarte, 2004] L. T. Duarte, *Redes Neurais para Equalização de Canais de Comunicação*, Relatório Final de Iniciação Científica, Unicamp, 2004.
- [Duda et al., 2000] R. O. Duda, P. E. Hart, D. G. Store, *Pattern Classification*, Wiley-Interscience, 2a. Edição, 2000.
- [Ellacott e Bose, 1996] S. Ellacott, D. Bose, *Neural Networks: Deterministic Methods of Analysis*, International Thomson Computer Press, 1996.
- [Ferrari et al., 2003] R. Ferrari, C. M. Panazio, R. R. F. Attux, C. C. Cavalcante, L. N. de Castro, F. J. Von Zuben, J. M. T. Romano, “Unsupervised Channel Equalization Using Fuzzy Prediction-Error Filters”, *Proceedings of the IEEE Workshop on Neural Networks for Signal Processing*, Toulouse, França, 2003.
- [Ferrari, 2005] R. Ferrari, *Equalização de Canais de Comunicação Baseada em Filtros Fuzzy*, Tese de Mestrado, Unicamp, 2005.
- [Frater et al., 1995] 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.
- [Futuyma, 2003] D. J. Futuyma, *Biologia Evolutiva*, FUNPEC – Editora, 2003.

- [Georgiadis, 2000] A. T. Georgiadis, *Adaptive Equalisation for Impulsive Noise Environments*, Tese de Doutorado, Universidade de Edimburgo, 2000.
- [Giannakis e Serpedin, 1997] G. Giannakis, E. Serpedin, “Linear Multichannel Blind Equalizers of Nonlinear FIR Volterra Channels”, *IEEE Trans. on Signal Processing*, Vol. 45, No. 1, pp. 67-81, 1997.
- [Gibson et al., 1991] G. J. Gibson, S. Siu, C. F. Cowan, “The Application of Nonlinear Structures to the Reconstruction of Binary Signals”, *IEEE Trans. on Signal Processing*, Vol. 39, No. 8, pp. 1877-1884, 1991.
- [Goldberg, 1989] D. E. Goldberg, *Genetic Algorithms in Machine Learning and Optimization*, Addison-Wesley, 1989.
- [Haykin, 1994a] S. Haykin, *Neural Networks: A Comprehensive Foundation*, MacMillan, 1994.
- [Haykin, 1994b] S. Haykin, *Communication Systems*, Wiley, 1994.
- [Haykin, 1996] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 1996.
- [Heródoto, 1952] Heródoto, *The History of Herodotus*, Britannica Great Books, 1952.
- [Holland, 1992] J. H. Holland, *Adaptation in Natural and Artificial Systems*, MIT Press, 1992.
- [Ibnkahla e Castanié, 1996] M. Ibnkahla, F. Castanié, “Neural Network Identification of Digital Satellite Channels: the Adaptive Nonlinear Enhancer”, *IEEE International Conference on Neural Networks*, Vol. 3, pp. 3 – 6, 1996.
- [Ibnkahla, 2000] M. Ibnkahla, “Neural Network Pre-distortion Technique for Digital Satellite Communications”, *Proceedings of the ICASSP*, Vol. 6, pp. 5-9, 2000.
- [Johnson et al., 1998] C. R. Johnson, P. Schniter, T. J. Endres, J. 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.
- [Kajikawa, 2000] Y. Kajikawa, “The Adaptive Volterra Filter: its Present and Future”, *Electronics and Communications in Japan, Part 3*, Vol. 83, No. 12, pp. 51-61, 2000.
- [Kohonen et al., 1990] T. Kohonen, K. Raivio, O. Simula, O. Venta, J. Henriksson, “Combining Linear Equalization and Self-Organizing Adaptation in Dynamic Discrete-signal Detection”, *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN)*, Vol. 1, pp. 223-228, 1990.
- [Kosko, 1996] B. Kosko, *Fuzzy Engineering*, Prentice Hall, 1996.

- [Kóvacs, 1997] Z. L. Kóvacs, *O Cérebro e a sua Mente: uma Introdução à Neurociência Computacional*, Edição Acadêmica, 1997.
- [Kwan et al., 1997] H. Kwan, I. Sungbin, E. J. Powers, E. E. Swartzlander, “Parallel Implementation of a Fast Third-order Volterra Filter”, Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS), Vol. 4, pp. 2473-2476, 1997.
- [Kuruoglu et al., 1998a] E. E. Kuruoglu, P. J. W. Rayner, W. J. Fitzgerald, “Least  $l_p$ -norm Impulsive Noise Cancellation with Polynomial Filters”, *Signal Processing*, No. 69, pp. 1-14, 1998.
- [Kuruoglu et al., 1998b] E. E. Kuruoglu, W. J. Fitzgerald, P. J. W. Rayner, “Near Optimal Detection of Signals in Impulsive Noise Modeled with a Symmetric  $\alpha$ -Stable Distribution”, *IEEE Communications Letters*, Vol. 2, No. 10, pp. 282-284, 1998.
- [Li et al., 1995] 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.
- [Lin e Wei, 1994] J. Lin e C. Wei, “Adaptive Volterra-DFE and Timing Recovery in Digital Magnetic Recording System”, Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS), Vol. 3, pp. 41-44, 1994.
- [Lucky, 1965] R. W. Lucky, “Automatic Equalization for Digital Communication”, *Bell System Technical Journal*, Vol. 44, pp. 547-588.
- [Mahfoud, 2000] S. W. Mahfoud, “Niching Methods”, in *T. Bäck et al. (editors) - Evolutionary Computation 2: Advanced Algorithms and Operators*, Institute of Physics Publishing, 2000.
- [Mathews, 1991] V. J. Mathews, “Adaptive Polynomial Filters”, *IEEE Signal Processing Magazine*, Vol. 8, No. 3, pp. 10-26, 1991.
- [Mathews e Sicuranza, 2000] V. J. Mathews e G. L. Sicuranza, *Polynomial Signal Processing*, Wiley-Interscience, 2000.
- [Michalewicz, 1996] Z. Michalewicz, *Genetic Algorithm + Data Structures = Evolution Programs*, 3rd Edition, Springer-Verlag, 1996.
- [Montalvão et al., 1998] J. R. Montalvão Filho, B. Dorizzi, J. C. M. Mota, “Some Theoretical Limits of Efficiency of Linear and Nonlinear Equalizers”, *Revista da Sociedade Brasileira de Telecomunicações*, Vol. 14, pp. 85-92, 1998.
- [Mulgrew, 1996] B. Mulgrew, “Applying Radial Basis Functions”, *IEEE Signal Processing Magazine*, Vol. 13, No. 2, pp. 50-65, 1996.

[Neubauer, 1995] A. Neubauer, “Non-linear Adaptive Filters Based on Genetic Algorithms with Applications to Signal Processing”, Proceedings of the IEEE International Conference on Evolutionary Computation, Vol. 2, pp. 572-532, 1995.

[Oppenheim et al., 1997] A. V. Oppenheim, A. S. Willsky, S. H. Nawab, *Signals and Systems*, 2<sup>nd</sup> Edition, Prentice Hall, 1997.

[Papoulis, 1991] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 1991.

[Park e Jeong, 1999] D. Park, K. J. Jeong, “Complex Bilinear Recurrent Neural Network for Equalization of a Digital Satellite Channel”, Proceedings of the International Joint Conference on Neural Networks (IJCNN), Vol. 3, pp. 1485-1490, 1999.

[Park e Jeong, 2002] D. Park, K. J. Jeong, “Complex-bilinear Recurrent Neural Network of Equalization of a Digital Satellite Channel”, IEEE Trans. on Neural Networks, Vol. 13, No. 3, pp. 711-725, 2002.

[Patra, 1998] S. K. Patra, *Development of Fuzzy System Based Channel Equalisers*, Tese de Doutorado, Universidade de Edimburgo, 1998.

[Patra e Mulgrew, 1998] S. K. Patra, B. Mulgrew, “Fuzzy Implementation of a Bayesian Equalizer in the Presence of Intersymbol and Co-channel Interference”, Proceedings of the IEE: Communications, Vol. 145, No. 5, pp. 323-330, 1998.

[Principe et al., 2000] J. C. Principe, N. R. Euliano, W. C. Lefebvre, *Neural and Adaptive Systems*, Wiley, 2000.

[Proakis, 1995] J. Proakis, *Digital Communications*, 3<sup>rd</sup> Edition, McGraw-Hill, 1995.

[Regalia, 1995] P. A. Regalia, *Adaptive IIR Filtering in Signal Processing and Control*, Marcel Dekker, 1995.

[Regalia, 1999] P. A. Regalia, “On the Equivalence between the Godard and Shalvi-Weinstein Schemes of Blind Equalization”, Signal Processing, No. 73, pp. 185-190, 1999.

[Regalia, 2001] P. A. Regalia, “On the Convergent Points of Blind Adaptive IIR Equalizers”, Proceedings of the IEEE/EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP), Baltimore, EUA, 2001.

[Sareni e Krähenbühl, 1998] B. Sareni, L. Krähenbühl, “Fitness Sharing and Niching Methods Revisited”, IEEE Transactions on Evolutionary Computation, Vol. 2, No. 3, pp. 97-106, 1998.

[Shalvi e Weinstein, 1990] 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.

- [Shannon, 1948] C. E. Shannon, "A Mathematical Theory of Communication", Bell System Technical Journal, No. 27, pp. 379-423 / 623-656, 1948.
- [Shynk, 1989] J. J. Shynk, "Adaptive IIR Filtering", IEEE ASSP Magazine, April, pp. 4-21, 1989.
- [Silva e Gerken, 2000] M. T. M. Silva, M. Gerken, "Um Algoritmo Acelerador para Treinamento de Redes Neurais tipo MLP", Anais do XVIII Simpósio Brasileiro de Telecomunicações, 2000.
- [Silva et al., 2000] M. T. M. Silva, M. Gerken, M. D. Miranda, "Estruturas Não-lineares Aplicadas à Recuperação de Sinais Binários", Anais do XVIII Simpósio Brasileiro de Telecomunicações, Gramado, Brasil, 2000.
- [Silva, 2005] M. T. M. Silva, *Um Estudo sobre Técnicas de Equalização Autodidata*, Tese de Doutorado, USP, 2005.
- [Siu et al., 1989] S. Siu, G. J. Gibson, C. F. Cowan, "Decision Feedback Equalization using Neural Network Structures", Proceedings of the First IEE Conference on Artificial Neural Networks, pp. 125-128, 1989.
- [Smith, 1975] C. U. M. Smith, *El problema de la vida*, Alianza Editorial, 1975.
- [Suyama et al., 2003a] R. Suyama, R. R. F. Attux, J. M. T. Romano, M. Bellanger, "Relations Entre les Critères du Module Constante et de Wiener", 19<sup>e</sup> Colloque GRETSI, Paris, França, 2003.
- [Suyama et al., 2003b] R. Suyama, M. S. de Castro, R. R. F. Attux, L. N. de Castro, F. J. Von Zuben, J. M. T. Romano, "Identificação Não-supervisionada de Canais de Comunicação Baseada em Sistemas Imunológicos Artificiais", XX Simpósio Brasileiro de Telecomunicações (SBrT), Rio de Janeiro, Brasil, 2003.
- [Sweatman, 1995] C. Z. W. H. Sweatman, "Constructive Neural Network Design for the Solution of Two-state Classification Problems with Application to Channel Equalization", Proceedings of the IEEE Workshop on Neural Networks for Signal Processing (NNSP), pp. 551-560, 1995.
- [Thomas, 1971] E. J. Thomas, "Some Considerations on the Application of the Volterra Representation of Nonlinear Networks to Adaptive Echo Cancellation", Bell Syst. Technical Journal, Vol. 50, pp. 2979-2805, 1971.
- [Tugnait, 1987] J. Tugnait, "Identification of Linear Stochastic Systems via Second- and Fourth-order Cumulant Matching", IEEE Transactions on Information Theory, Vol. IT-33, No. 3, pp. 393-407, 1987.

[Von Zuben, 1996] F. J. Von Zuben, *Modelos Paramétricos e Não-paramétricos de Redes Neurais Artificiais e Aplicações*, Tese de Doutorado, Unicamp, 1996.

[Von Zuben e de Castro, 2003] F. J. Von Zuben, L. N. de Castro, *Notas de Aula do Curso de IA353 – Redes Neurais*, Unicamp, 2003.

[Wang e Mendel, 1992] L. Wang, J. M. Mendel, “Generating Fuzzy Rules by Learning from Examples”, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 22, No. 6, pp. 1414-1427, 1992.

[Wang e Mendel, 1993] L. Wang, J. M. Mendel, “Fuzzy Adaptive Filters, with Application to Nonlinear Channel Equalization”, *IEEE Transactions on Fuzzy Systems*, Vol. 1, No. 3, pp. 161-170, 1993.

[Widrow e Hoff, 1960] B. Widrow, M. E. Hoff Jr., “Adaptive Switching Circuits”, *IRE WESCON Conv. Rec.*, Pt. 4, pp. 96-104, 1960.

[Widrow e Winter, 1988] B. Widrow e R. Winter, “Neural Nets for Adaptive Filtering and Adaptive Pattern Recognition”, *Computer*, Vol. 21, No. 3, pp. 25-39, 1988.

[Williams e Zipser, 1989] R. J. Williams, D. Zipser, “A Learning Algorithm for Continually Running Fully Recurrent Neural Networks”, *Neural Computation*, vol. 1, pp. 270-280, 1989.

[Wolpert e Macready, 1997] D. H. Wolpert, W. C. Macready, “No Free Lunch Theorems for Optimization”, *IEEE Transactions on Evolutionary Computation*, Vol. 1, No. 1, pp. 67-82, 1997.

[Zadeh, 1965] L. A. Zadeh, “Fuzzy Sets”, *Information and Control*, No. 8, pp. 338-353, 1965.

[Zeng et al., 1999] 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.



## **A. Apêndice: Trabalhos Publicados Durante o Doutorado**

### ***A.1 Artigos Publicados em Periódicos de Circulação Internacional***

**R. R. F. Attux**, J. M. T. Romano, “Chaotic Phenomena in Adaptive Blind Equalisers”, IEE Proceedings – Vision, Image and Signal Processing, vol. 150, pp. 360-364, 2003.

**R. R. F. Attux**, M. B. Loiola, R. Suyama, L. N. de Castro, F. J. Von Zuben, J. M. T. Romano, “Blind Search for Optimal Wiener Equalizers Using an Artificial Immune Network Model”, EURASIP Journal on Applied Signal Processing – Special Issue on Genetic and Evolutionary Computation for Signal Processing and Image Analysis, vol. 2003, No. 8, pp. 740-747, 2003.

### ***A.2 Artigos Publicados em Anais de Congressos***

C. M. Panazio, **R. R. F. Attux**, “A  $4^N$ -QAM Adaptive Decision Device to Mitigate I/Q Imbalance and Impairments Caused by Time-varying Flat Fading Channels”, IEEE Workshop on Machine Learning for Signal Processing (MLSP), São Luís, Brasil, 2004.

**R. R. F. Attux**, R. R. Lopes, L. N. de Castro, F. J. Von Zuben, J. M. T. Romano, “Genetic Algorithms for Blind Maximum-likelihood Receivers”, IEEE Workshop on Machine Learning for Signal Processing (MLSP), São Luís, Brasil, 2004.

**R. R. F. Attux**, L. T. Duarte, L. N. de Castro, F. J. Von Zuben, J. M. T. Romano, “Adaptive Identification of Feedback Systems Using an Artificial Immune Network”, International Conference on Telecommunications (ICT), Fortaleza, Brasil, 2004.

**R. R. F. Attux**, L. N. de Castro, F. J. Von Zuben, J. M. T. Romano, “A Paradigm for Blind IIR Equalization Using the Constant Modulus Criterion and an Artificial Immune Network”, IEEE Workshop on Neural Networks for Signal Processing (NNSP), Toulouse, França, 2003.

R. Ferrari, C. M. Panazio, **R. R. F. Attux**, C. C. Cavalcante, L. N. de Castro, F. J. Von Zuben, J. M. T. Romano, “Unsupervised Channel Equalization Using Fuzzy Prediction-error Filters”, IEEE Workshop on Neural Networks for Signal Processing (NNSP), Toulouse, França, 2003.

R. Suyama, **R. R. F. Attux**, J. M. T. Romano, M. Bellanger, “On the Relationship between Least Squares and Constant Modulus Criteria for Adaptive Filtering”, Invited Paper, Asilomar Conference on Signals, Systems and Computers, Pacific Grove, EUA, 2003.

R. Suyama, **R. R. F. Attux**, J. M. T. Romano, M. Bellanger, “Relations entre les Critères du Module Constant et de Wiener”, 19<sup>e</sup> Colloque GRETSI, Paris, França, 2003.

R. Suyama, M. S. de Castro, **R. R. F. Attux**, L. N. de Castro, F. J. Von Zuben, J. M. T. Romano, “Identificação Não-supervisionada de Canais de Comunicação Baseada em Sistemas Imunológicos Artificiais”, XX Simpósio Brasileiro de Telecomunicações, Rio de Janeiro, Brasil, 2003.

A. L. T. Romano, **R. R. F. Attux**, L. S. Baptista Filho, F. J. Von Zuben, “Aplicação de Autômatos Celulares no Estudo da Dinâmica Oceanográfica - Estudo de Caso: Previsão de Espalhamento em uma Baía Fictícia”, III Congresso Brasileiro de Computação e I Workshop de Tecnologia de Informação Aplicada ao Meio Ambiente, Itajaí, Brasil, 2003.

**R. R. F. Attux**, J. M. T. Romano, “Adaptive Criteria Optimization as a Least Squares Problem”, Proceedings of the IEEE International Telecommunications Symposium (ITS), Natal, Brasil, 2002.

A. M. Costa, **R. R. F. Attux**, J. M. T. Romano, “A New Method for Channel Identification with Genetic Algorithms”, Proceedings of the IEEE International Telecommunications Symposium (ITS), Natal, Brasil, 2002.