

UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
CURSO DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Predição Neural e Estimação de Função de Densidade de Probabilidade Aplicadas à Equalização Cega

Autor: Charles Casimiro Cavalcante

Orientador: João Cesar Moura Mota
Co-Orientador: Jugurta Rosa Montalvão Filho

*Dissertação apresentada à Coordenação do Curso
de Pós-Graduação em Engenharia Elétrica da
Universidade Federal do Ceará como parte dos
requisitos exigidos para obtenção do grau de **Mestre
em Engenharia Elétrica***

Fevereiro-2001

Esta dissertação foi submetida como parte dos requisitos necessários à obtenção do Grau de Mestre em Engenharia Elétrica, outorgado pela Universidade Federal do Ceará, e encontra-se à disposição dos interessados na Biblioteca da referida Universidade.

A citação de qualquer trecho desta Dissertação é permitida, desde que feita na conformidade com as normas da ética científica.

Charles Casimiro Cavalcante

DISSERTAÇÃO APROVADA EM 07/02/2001

Prof. Dr. João Cesar Moura Mota
Universidade Federal do Ceará - UFC

Prof. Dr. Francisco Rodrigo Porto Cavalcanti
Universidade Federal do Ceará - UFC

Prof. Dr. Jugurta Rosa Montalvão Filho
Universidade Tiradentes - UNIT

Prof. Dr. João Marcos Travassos Romano
Universidade Estadual de Campinas - UNICAMP

*Aos meus pais,
Casimiro e Ivoneide,
pelo amor incondicional.*

AGRADECIMENTOS

Em uma das bibliografias que consultei, o autor iniciou os agradecimentos com a seguinte frase: *“Agradecer é tarefa muito difícil... agradecer pouco significa não reconhecer a importância das pessoas que foram cúmplices, agradecer muito coloca todo o peso do trabalho nos ombros de outrem. Tentarei então fazer na medida justa”*. Também seguirei o exemplo para tentar agradecer à todos aqueles que foram parte dessa caminhada.

Aos meus pais, **Casimiro e Ivoneide**, cuja tarefa é a mais difícil, ver a importância de um trabalho que me obrigou, várias vezes, a me tornar ausente e distante tendo causado dor tanto para mim quanto para eles. Certamente palavras são insuficientes para agradecer aqueles sem os quais nada disso teria sido possível ou teria valido a pena e cujo amor sempre me inspirou e incentivou.

Gostaria de agradecer profundamente a meu orientador **João Cesar Moura Mota**, pela orientação segura e precisa e também pela amizade que, certamente, ultrapassou os limites da relação profissional. Não só por isso, mas também pelo profundo incentivo e esforço para com minha carreira acadêmica refletidos nas inúmeras oportunidades que tive durante minha formação.

Ao meu co-orientador **Jugurta R. Montalvão Filho**, pelo apoio, incentivo e preocupação que demonstrou com o trabalho, além disso tenho muito orgulho de dizer que mais que um colega de trabalho, ganhei um amigo muito querido. Sua enorme contribuição faz, no meu entendimento, o trabalho tão dele quanto meu.

A Prof. **Bernadette Dorizzi** por ter orientado meus estudos durante a temporada no EPH do Institut National des Télécommunications (INT) em Evry, França. Não somente por seus conhecimentos, que são sempre uma referência para mim, mas também pela amizade e carinho dispensados. Também não posso esquecer de agradecer sua enorme paciência durante o período de enormes limitações com a língua francesa. Aproveito para, em seu nome, agradecer fortemente ao INT pela estrutura oferecida.

Ao Prof. **Rodrigo Cavalcanti**, pelas discussões técnicas bastante proveitosas e também pelo apoio computacional do laboratório da Cooperação Ericsson/UFC.

Ao Prof. **João Marcos Travassos Romano**, pelas sugestões técnicas bastante pertinentes que contribuíram bastante para a forma final do trabalho.

Aos professores e colaboradores do curso de mestrado em Engenharia Elétrica da UFC que souberam não somente transmitir o conhecimento técnico, mas também permitir a realização desse trabalho. Agradeço de maneira especial aqueles que além de professores e mestres se tornaram amigos e cúmplices, entre eles: Paulo César Cortez, Mário Fiallos, Elvio Giraudo e Menotti Machado.

Aos (vários) colegas e companheiros brasileiros ou não que tornaram a estada na França uma período muito agradável e promovendo também proveitosas discussões técnicas, culturais e sociais, entre eles: Jugurta, Marta, Mário Igor, Fábio Laurino (Fabão), Adriano (Blau), Frandberto, Walmir (Cocão), André Caetano (Zeca), Belen (Argentina), Eleftherios (grego-brasileiro), Ricardo Wigman, Menotti (e família), Ricardo Rodrigues (e família), Waltemar entre tantos que não consigo lembrar os nomes.

A Guga e Marta que mais que grandes amigos foram minha família na França. Além de muito carinho tenho para com eles uma profunda dívida de gratidão que espero algum dia poder retribuir.

A Erika, pelo companheirismo, incentivo e por tudo que temos passado juntos durante toda a duração deste trabalho.

Aos familiares e amigos, parte integrante tão importante em todo o processo, pelo apoio e torcida. Em especial àqueles que demonstraram interesse pelo trabalho e também contribuíram de alguma forma: Guilherme, Robson e Josué.

À Fundação Cearense de Amparo à Pesquisa (FUNCAP) e ao programa TCE-M da União Européia pelo apoio financeiro.

A todos os funcionários do corpo administrativo da UFC que foram de fundamental importância durante todo o decorrer do curso entre eles: Fernanda, Mário Sérgio e Francisca. E àqueles que propiciaram uma estadia na França mais que agradável nesse aspecto: Mme Lafont, Mme Courcier e M. Claude Pèrpere pelo suporte computacional.

Aos meus irmãos Kleber, Rogério e Marcelo que, à sua maneira mais silenciosa, sempre souberam mostrar seu amor entre nós e incentivos tão necessários.

A Deus, a quem tudo devo.

Charles Casimiro Cavalcante

ÍNDICE

Agradecimentos	i
Resumo	vii
Abstract	ix
Lista de Figuras	xi
Lista de Tabelas	xv
Notações e Abreviações	xvii
1 Introdução	1
1.1 Pequena história da evolução da comunicação	2
1.2 Comunicação digital e desconvolução	3
1.3 Contaminação do sinal	6
1.4 Equalização adaptativa: técnicas clássicas e novas tendências	6
1.4.1 Estratégias preditivas de equalização autodidata	7
1.4.2 Redes neurais artificiais e equalização autodidata	8
1.4.3 Critérios estatísticos para equalização autodidata	9
1.5 Organização, estrutura e contribuições	9
I Conceitos Básicos	11
2 Modelo do Sistema de Comunicação Digital	13
2.1 Cadeia de comunicação digital	13
2.2 Modelo em banda básica equivalente	15
2.3 Fonte de informação	16

2.4	Modelo do canal de comunicação	17
2.4.1	Ruído gaussiano	17
2.4.2	Função de transferência	18
2.4.3	Tipos de canais	19
2.4.4	Interferência entre símbolos	21
2.5	Medidas de desempenho	23
2.5.1	Relação sinal-ruído	23
2.5.2	Taxa de símbolos errados	24
2.5.3	Erro quadrático médio	24
2.5.4	Relação interferência entrada-saída	24
2.5.5	Métodos de Monte Carlo	25
2.6	Síntese	28
3	Modelos Baseados em Teoria da Informação	29
3.1	Informação e incerteza	30
3.1.1	Propriedades da quantidade de informação	31
3.2	Entropia	31
3.2.1	Taxa de informação	34
3.2.2	Informação mútua	34
3.2.3	Capacidade de Canal	36
3.3	Divergência de Kullback-Leibler	37
3.4	Estimação de Função de Densidade de Probabilidade	38
3.4.1	Estimação paramétrica	38
3.4.2	Estimação não paramétrica	43
3.5	Síntese	44
4	Redes Neurais Artificiais	45
4.1	O neurônio biológico	46
4.2	O neurônio artificial	48
4.2.1	Neurônio de MacCulloch & Pitts	49
4.2.2	Neurônio de Rosenblatt - Perceptron	50
4.2.3	Função de ativação	51
4.3	Redes neurais artificiais	52
4.3.1	Redes neurais multicamadas	52
4.3.2	Redes neurais de função de base radial	54
4.3.3	Redes neurais recorrentes	56
4.4	Aprendizado das redes neuronais	56
4.4.1	Aprendizado supervisionado	57
4.4.2	Aprendizado não-supervisionado	60
4.4.3	Aprendizado hebbiano	61
4.5	Síntese e conclusões	63
5	Equalização e Classificação: Recuperação da Informação	65
5.1	Abordagem clássica	66
5.2	Principais técnicas de equalização adaptativa	68
5.2.1	Equalização adaptativa supervisionada	68
5.2.2	Equalização adaptativa não-supervisionada ou autodidata	69

5.2.3	Técnicas de Bussgang	69
5.3	Classificação	72
5.3.1	Influência da ordem: opção (a)	74
5.3.2	Influência do atraso: opção (b)	77
5.4	Síntese e conclusões	78
II	Proposição	79
6	Predição Não Linear para Equalização Autodidata	81
6.1	Predição linear	82
6.1.1	Desenvolvimento matemático	83
6.1.2	Influência do ruído e atraso	85
6.2	Preditores não lineares	85
6.2.1	Equalizador de magnitude e fase	86
6.2.2	Critério “Predictive Constant Modulus” (PCM)	87
6.3	DFE como preditor não linear	88
6.4	Predição × Interpolação	89
6.5	Preditor neural	91
6.5.1	Aprendizado da NPS	93
6.5.2	Limites e complexidade	101
6.6	Simulações computacionais	102
6.6.1	Canais de fase mínima	103
6.6.2	Canais de fase não mínima	109
6.7	Síntese e conclusões	114
7	Critério Estatístico Baseado na Estimação da Função de Densidade de Probabilidade	117
7.1	Características do sinal na saída do equalizador	118
7.1.1	fdp do sinal na saída do equalizador	118
7.2	Critério estatístico para equalização	120
7.2.1	Crítérios cegos: abordagem de classificação	123
7.3	Simulações computacionais	125
7.4	Síntese e conclusões	131
8	Conclusões e Perspectivas	133
III	Apêndices	137
A	Derivação de Funções Escalares em Relação a um Vetor	139
A.1	Definições básicas	139
A.2	Alguns casos particulares	140
A.3	Relação: vetor gradiente - derivada em relação a um vetor	143

B	Sumário de Algoritmos Propostos na Dissertação	145
B.1	“Self Organized for Finding Valleys Algorithm” (SOFVA)	145
B.2	Adaptação NPS-MPE	146
B.3	Adaptação NPS-CMC	146
B.4	“Fitting pdf Algorithm” (FPA)	147
C	Artigos Publicados	149
	Referências	161
	Índice Remissivo	167
	Índice de Autores	171

RESUMO

Esta dissertação é dedicada ao estudo de estratégias de equalização cega baseadas em filtro de predição não linear e em critério de similaridade entre as funções de densidade de probabilidade (fdp) dos sinais de entrada do canal linear e de saída de um equalizador. Estes dois estudos são feitos sob a óptica da teoria da classificação aplicada à recuperação da informação.

A estratégia de equalização cega neuro-preditiva estudada nesta dissertação baseia-se numa estrutura de filtragem neural com uma única entrada e uma camada escondida no ramo preditivo. Um critério de otimização inédito baseado na divergência de Kullback-Leibler é utilizado para o ajuste dos parâmetros de polarização dos neurônios via algoritmo LMS. Os pesos de saída dos neurônios foram ajustados de forma supervisionada, cuja referência é o próprio sinal recebido. O algoritmo LMS foi utilizado no ajuste dos pesos, baseado nos critérios do erro quadrático médio (EQM) e módulo constante (CM).

Esta estratégia, além de realizar satisfatoriamente a equalização, permitiu identificar o número de neurônios necessários para combater a distorção provocada pelo canal.

Outra estratégia de equalização cega inédita estudada nesta dissertação objetivou obter a fdp dos dados transmitidos na saída do equalizador, a partir dos conceitos de similaridade entre funções contida na medida de divergência de Kullback-Leibler. O modelo de fdp baseado na mistura de gaussianas, com controle de variância média sobre o alfabeto de dados transmitidos, foi adotado como fdp de referência. O equalizador linear transversal cego teve seus coeficientes ajustados por um algoritmo LMS baseado no critério de similaridade citado. Seu bom desempenho foi constatado a partir de comparações com outras estratégias cegas clássicas.

Palavras-chave: Equalização cega, predição não linear, redes neurais,

classificação/equalização, estimação da fdp.

ABSTRACT

THIS dissertation is dedicated to the study of blind equalization strategies based on nonlinear prediction filter and a criterion of similarity between probability density functions (pdf) of the linear channel input signals and of the output signals of an equalizer. These two studies are done by means of the classification theory applied to the information recovering.

The neuro-predictive blind strategy studied in this dissertation is based on a neural filtering structure with only one input and one hidden layer in the predictive branch. A novel optimization criterion based on the Kullback-Leibler divergence is used for the adjustment of the neurons polarization parameters through an LMS algorithm. The output taps of the neurons were updated in a supervised way, for which the reference is the received signal. The LMS algorithm was used in the taps adjustment, based on the criteria of the mean squared error (MSE) and constant modulus (CM).

This strategy, besides satisfactorily providing equalization, has permitted the identification of the number of neurons necessary to combat the distortion provided by the channel.

Another novel blind equalization strategy studied in this dissertation has aimed to obtain the pdf of transmitted data on the equalizer output, using the concepts of similarity of functions contained in the Kullback-Leibler divergence. The pdf model, based on the Gaussians mixture, with variance and mean control over the transmitted data alphabet, was adopted as reference. The blind transversal linear equalizer had its coefficients updated by means of an LMS algorithm based on the cited similarity criterion. Its good performance was observed through comparisons with other classical blind strategies.

Keywords: Blind equalization, nonlinear prediction, neural networks,

classification/equalization, pdf estimation.

LISTA DE FIGURAS

1.1	Rede de comunicação de serviços integrados.	2
1.2	Síntese do problema de desconvolução.	4
1.3	Rede de dados multiponto.	5
1.4	Multipercurso em comunicações móveis.	6
2.1	Sistema de comunicação digital genérico.	14
2.2	Representação em banda básica equivalente de um sistema de transmissão digital.	15
2.3	Modelo discreto no tempo do problema de equalização em um sistema de comunicação digital.	15
2.4	Três tipos de constelação.	16
2.5	Modelo do canal de comunicação.	18
2.6	Canal de fase mínima.	21
2.7	Canal de fase máxima.	21
2.8	Canal de fase não mínima.	22
2.9	Diagrama de olho para sinais binários.	23
2.10	Gráfico de RSI de saída versus entrada.	25
2.11	Simulações de Monte Carlo sobre um experimento.	28
3.1	Quantidade de informação na ocorrência de um evento.	32
3.2	Entropia de uma fonte binária em função da probabilidade de ocorrência p_1	33
3.3	Transmissão de informação através de um canal de comunicação.	34
3.4	Relação entre informação mútua $I(\mathcal{A}; \mathcal{X})$ e entropias $\mathcal{H}(\mathcal{A})$ e $\mathcal{H}(\mathcal{X})$	36
3.5	Funções densidade de probabilidade gaussianas.	39
3.6	Diferenças entre a utilização de um modelo adequado e inadequado na estimação paramétrica.	40
4.1	Representação de um neurônio biológico.	46
4.2	Sinapse entre dois nerônios biológicos.	47

4.3	Funcionamento da <i>bomba de sódio</i> no neurônio.	48
4.4	Processamento espacial-temporal dos estímulos para o neurônio biológico. .	48
4.5	Neurônio de McCulloch-Pitts.	49
4.6	Modelo do neurônio artificial <i>perceptron</i>	50
4.7	Relação entre a frequência de pulsos em um neurônio e intensidade da despolarização da membrana.	51
4.8	Algumas funções de ativação de neurônios artificiais.	53
4.9	Arquitetura de uma rede de perceptrons multicamada com duas camadas escondidas.	54
4.10	Rede de função de base radial.	55
4.11	Redes neurais artificiais recorrentes.	56
4.12	Modelo da máquina de aprendizado.	57
4.13	Diagrama de blocos do aprendizado supervisionado.	58
4.14	Diagrama de blocos do aprendizado não supervisionado.	60
4.15	Estrutura da rede de Kohonen.	62
5.1	Filtro linear transversal.	66
5.2	Estimador não linear.	69
5.3	Superfície (função) de separação unidimensional.	73
5.4	Representação unidimensional dos estados e superfície de separação.	74
5.5	Representação bidimensional dos estados, superfícies de separação e projeções. .	76
5.6	Representação dos estados ruidosos com variação do atraso de decisão. . . .	78
6.1	Diagrama de uma estrutura de predição.	82
6.2	Equalizador de magnitude e fase.	86
6.3	Estrutura ARMA composta de BPEF e FPEF.	87
6.4	Diagrama do DFE.	88
6.5	Predição vista como um problema de interpolação - exemplo de um caso bidimensional.	89
6.6	Superfície de interpolação não linear para equalização de um canal de FNM. .	92
6.7	Esquema representando a arquitetura da “Neural Predictor Structure” (NPS). .	92
6.8	Função densidade de probabilidade do sinal na saída de um canal linear para uma modulação BPSK.	94
6.9	Função módulo.	95
6.10	Esquema do processamento para encontrar os <i>vales</i> na fdp do sinal recebido. .	96
6.11	Funcional J_1 e fdp do sinal recebido $p(x)$	96
6.12	Função $\ln(N(x, d))$	97
6.13	Funcional J_2 e fdp do sinal recebido $p(x)$	98
6.14	Esquema do aprendizado para a NPS - auto-organizado e supervisionado. . .	99
6.15	Evolução do DSE para o canal \mathbf{f}_1 para uma SNR = 40 dB com modulação BPSK.	103
6.16	Evolução do DSE para o canal \mathbf{f}_1 para uma SNR = 20 dB com modulação BPSK.	104
6.17	Evolução do DSE para o canal \mathbf{f}_1 para uma SNR = 40 dB com modulação BPSK - comparação entre preditor linear, NPS-MPE e NPS-MPE-LS. . . .	105
6.18	Superfícies de interpolação para as estruturas NPS - canal \mathbf{f}_1 , SNR = 40 dB e modulação BPSK.	106

6.19	Evolução temporal dos parâmetros θ_i - (—) valores ótimos e (-) valores obtidos.	107
6.20	Evolução do DSE para o canal \mathbf{f}_2 para uma SNR = 40 dB com modulação BPSK.	107
6.21	Evolução do DSE para o canal \mathbf{f}_2 para uma SNR = 20 dB com modulação BPSK.	109
6.22	Evolução do DSE para o canal \mathbf{f}_2 para uma SNR = 40 dB com modulação BPSK - comparação entre preditor linear, NPS-MPE e NPS-MPE-LS. . . .	109
6.23	Superfícies de interpolação para as estruturas NPS - canal \mathbf{f}_2 , SNR = 40 dB e modulação BPSK.	110
6.24	Evolução do DSE para o canal \mathbf{f}_3 para uma SNR = 40 dB com modulação BPSK.	112
6.25	Evolução do DSE para o canal \mathbf{f}_3 para uma SNR = 30 dB com modulação BPSK.	112
6.26	Superfícies de interpolação para as estruturas NPS - canal \mathbf{f}_3 e modulação BPSK.	113
6.27	Evolução do DSE para o canal \mathbf{f}_4 para uma SNR = 40 dB com modulação BPSK.	114
7.1	Modelo em banda básica equivalente do sistema digital.	118
7.2	Funções de contraste.	121
7.3	Funcional $J_{\text{FPC}}(\mathbf{h})$ em função dos parâmetros h_1 e h_2 do equalizador. . . .	123
7.4	Curvas de nível para o funcional $J_{\text{FPC}}(\mathbf{h})$ em função dos parâmetros h_1 e h_2 do equalizador.	123
7.5	Função $\ln(\Phi(y))$ para alguns tipos de modulação.	124
7.6	Comparação entre funções alvos para os critérios de Sato, Godard e FPC. . .	125
7.7	Esquema de recuperação de fase em um sistema com equalizador.	126
7.8	Características do canal \mathbf{f}_5 - resposta em frequência e diagrama de zeros. .	126
7.9	Evolução do DSE para o canal \mathbf{f}_5 - modulação BPSK e SNR = 30 dB. . . .	127
7.10	Características do canal \mathbf{f}_6 - resposta em frequência e diagrama de zeros. .	128
7.11	Evolução do DSE para o canal \mathbf{f}_6 - modulação BPSK e SNR = 30 dB. . . .	128
7.12	Curvas de taxa de símbolos errados para os algoritmos CMA e FPA - canal \mathbf{f}_6	129
7.13	Características do canal \mathbf{f}_7 - resposta em frequência e diagrama de zeros. .	130
7.14	Evolução do DSE para o canal \mathbf{f}_7 - modulação 4QAM e SNR = 30 dB. . .	131

LISTA DE TABELAS

5.1	Principais estimadores em algoritmos de Bussgang.	72
5.2	Equivalências entre alguns termos de equalização e classificação.	77
6.1	Resumo do algoritmo auto-organizado SOFVA.	99
6.2	Resumo dos algoritmos supervisionados para a NPS.	100
6.3	Sumário da simulação para o canal \mathbf{f}_1	108
6.4	Sumário da simulação para o canal \mathbf{f}_2	111
6.5	Sumário da simulação para o canal \mathbf{f}_3	114
7.1	Sumário da simulação para o canal \mathbf{f}_5	127
7.2	Sumário da simulação para o canal \mathbf{f}_6	129
7.3	Sumário da simulação para o canal \mathbf{f}_7	131

NOTAÇÕES E ABREVIACÕES

Símbolos e notação

\mathbf{x}	: vetor \mathbf{x} (letras minúsculas em negrito)
\mathbf{X}	: matriz \mathbf{X} (letras maiúsculas em negrito)
\mathbf{X}^T	: transposta de \mathbf{X}
\mathbf{x}^*	: conjugado de \mathbf{x}
\mathbf{X}^H	: Hessiana (transposto conjugado) de \mathbf{X}
$ \bullet $: módulo (valor absoluto)
\mathcal{A}	: alfabeto de símbolos
δ	: atraso de decisão
\mathbf{h}	: vetor de coeficientes do filtro
$\mathbf{a}(n)$: seqüência de v.a. de símbolos transmitidos (janela)
$\mathbf{a}_i(n)$: seqüência determinística de símbolos transmitidos
$\mathbf{b}(n)$: seqüência de v.a. de ruído AWGN (janela)
\mathbf{f}	: resposta impulsiva do canal
σ_a^2	: variância do alfabeto de símbolos
σ_b^2	: variância do ruído AWGN
σ_y^2	: variância da saída do equalizador
$\Pr(\mathbf{A})$: probabilidade do evento \mathbf{A}
$p(\cdot)$: função densidade de probabilidade (fdp)
$p(\mathbf{A} \mathbf{B})$: função densidade de probabilidade de \mathbf{A} dado \mathbf{B}
$\mathcal{H}(\cdot)$: entropia
$\mathcal{H}(\mathbf{A} \mathbf{B})$: entropia condicional de \mathbf{A} dado \mathbf{B}
$J(\cdot)$: função custo (funcional)
$y(n)$: saída do equalizador
$x(n)$: entrada do equalizador
$\mathbb{E}\{\cdot\}$: operador esperança
\mathbb{Z}	: conjunto dos números inteiros
\mathbb{R}	: conjunto dos números reais

$\varphi(\cdot)$: função de ativação do neurônio artificial
$\text{sgn}(\cdot)$: função signum em que $\text{sgn}(a) = \begin{cases} 1, & a \geq 0 \\ -1, & a < 0 \end{cases}$
θ_k	: bias do neurônio k
$\boldsymbol{\theta}$: vetor de parâmetros θ_i
w_{kM}	: peso sináptico entre a entrada M e o neurônio k
$U[\mathbf{a}, \mathbf{b}]$: distribuição uniforme no intervalo entre \mathbf{a} e \mathbf{b}
$N(\mathbf{a}, \mathbf{b})$: distribuição gaussiana (normal) de média \mathbf{a} e variância \mathbf{b}
$\text{LS}[\mathbf{a}, \mathbf{b}]$: elementos linearmente espaçados entre \mathbf{a} e \mathbf{b}
$\nabla J(\alpha)$: gradiente de J em relação ao parâmetro α $\left(\frac{\partial J(\alpha)}{\partial \alpha}\right)$
β_i	: coeficientes da saída do neurônio i
$\boldsymbol{\beta}$: vetor de parâmetros β_i
λ	: taxa de aprendizagem
μ	: fator de passo (convergência)
j	: complexo $j = \sqrt{-1}$
$D_{\text{KL}}[\cdot]$: divergência de Kullback-Leibler
$I(\cdot)$: quantidade de informação
$I(\mathbf{A}; \mathbf{B})$: informação mútua entre \mathbf{A} e \mathbf{B}
\mathbf{F}	: matriz de convolução do canal

Abreviações

fdp :	Função Densidade de Probabilidade
i.i.d. :	independentes e identicamente distribuídas
v.a. :	variável aleatória
ARMA :	“Auto Regressive and Moving Average”
BPA :	“Back-Propagation Algorithm”
BPSK :	“Binary Phase Shift Keying”
CAG :	Controle Automático de Ganho
CM :	“Constant Modulus”
CMA :	“Constant Modulus Algorithm”
CRU :	Circunferência de Raio Unitário
DD :	Decisão Direta
DFE :	“Decision Feedback Equalizer”
EQM :	Erro Quadrático Médio
FPA :	“Fitting pdf Algorithm”
FPC :	“Fitting pdf Criterion”
FPEF :	“Forward Prediction Error Filter”
BPEF :	“Backward Prediction Error Filter”
FMáx :	Fase Máxima (canal)
FMín :	Fase Mínima (canal)
FNM :	Fase Não Mínima (canal)
FPC :	“Fitting Probability Criterion”
FT :	Função de Transferência

IES:	Interferência Entre Símbolos
IS:	Interferência do Sinal
KLD:	“Kullback-Leibler Divergence”
LMS:	“Least Mean Square”
LS:	“Linearly Spaced”
MBBE:	Modelo em Banda Básica Equivalente
MD:	Máxima Distorção
ML:	“Maximum Likelihood”
MLE:	“Maximum Likelihood Estimator”
MLP:	“Multi Layer Perceptron”
MOS:	Momentos de Ordem Superior
NM&P:	Neurônio de MacCulloch & Pitts
NPS:	“Neural Predictive Structure”
NPS-CMA:	NPS minimizada a partir do “Constant Modulus Criterion”
NPS-MPE:	NPS minizada a partir da “Minimization of Prediction Error”
NPS-MPE-LS:	NPS-MPE utilizando a inicialização dos θ_i “Linearly Spaced”
PCMA:	“Predictive Constant Modulus Algorithm”
QAM:	“Quadrature Amplitude Modulation”
RBF:	“Radial Basis Function”
RI :	Resposta ao Impulso
RNA:	Redes Neurais (Neuronais) Artificiais
RSI:	Relação Sinal-Interferência
SCD:	Sistema de Comunicação Digital
SOFVA:	“Self-Organized for Finding Valleys Algorithm”
TSE:	Taxa de Símbolos Errados

INTRODUÇÃO

DESDE o uso de sinais de fumaça, à cerca de 1.100 anos a.c., a comunicação à distância entre os seres humanos evoluiu de maneira espantosa. A invenção do telefone em 1876 por Alexander Graham Bell tornou-se um marco devido à possibilidade da comunicação pessoal privada mesmo em pontos geográficos distantes.

O desenvolvimento do chaveamento automático nas décadas de 30 e 40 possibilitou o surgimento e a popularização das redes de comunicação, inicialmente de voz e posteriormente de dados, e mais recentemente as redes de comunicação em *banda larga*, que buscam uma integração dos diversos tipos de mídia: voz, dados e imagens. A Figura 1.1 mostra um esquema de uma rede de comunicação de serviços integrados.

Com o crescente número de usuários, o uso de técnicas de acesso por multiplexação no tempo (TDMA¹), por exemplo, vem sendo proposto como solução para aproveitamento do mesmo meio para vários usuários, enquanto a complexidade e a sensibilidade ao ruído inerentes ao processamento analógico do sinal motivaram a conversão dos sistemas telefônicos para o uso de transmissão digital.

De fato, a introdução dos sistemas digitais trouxe várias vantagens adicionais além de uma menor complexidade no projeto com componentes eletrônicos: a manutenção da qualidade do sinal através de longas distâncias tornou-se mais fácil, o aumento na capacidade de transmissão e a possibilidade de inserção de novos serviços de maneira mais flexível.

Em contrapartida, traz consigo novos problemas, entre eles a interferência entre símbolos (IES). A IES ocorre quando a duração de um pulso em um intervalo de tempo de

¹Do inglês “Time Division Multiplex Access”.

transmissão se altera causando uma contaminação ou perturbação dos pulsos adjacentes sobre o pulso desejado na sequência transmitida. Para tratar tal efeito, utilizam-se *equalizadores*².

Esta dissertação versa sobre o processamento adaptativo de sinais para equalização de canais de comunicação digital. Técnicas de predição não linear síncrona baseadas em redes neurais e técnicas de estimação de funções de densidade de probabilidade do sinal no receptor são elementos que fundamentam a análise feita neste trabalho.

Este capítulo introdutório apresenta o problema sobre o qual discorre esta dissertação bem como sua organização estrutura e contribuições. Na seção 1.1 um pequeno histórico sobre a evolução dos sistemas de comunicação é apresentado. A seção 1.2 faz uma introdução à comunicação de dados e suas principais características. Na seção 1.3 são abordados os conceitos básicos sobre a degradação do sinal no processo de transmissão digital, seção 1.4 introduz-se algumas técnicas de equalização adaptativa existentes bem como a técnica proposta neste trabalho. Finalmente, na seção 1.5 são apresentadas a organização e estrutura do restante da dissertação bem como suas contribuições.

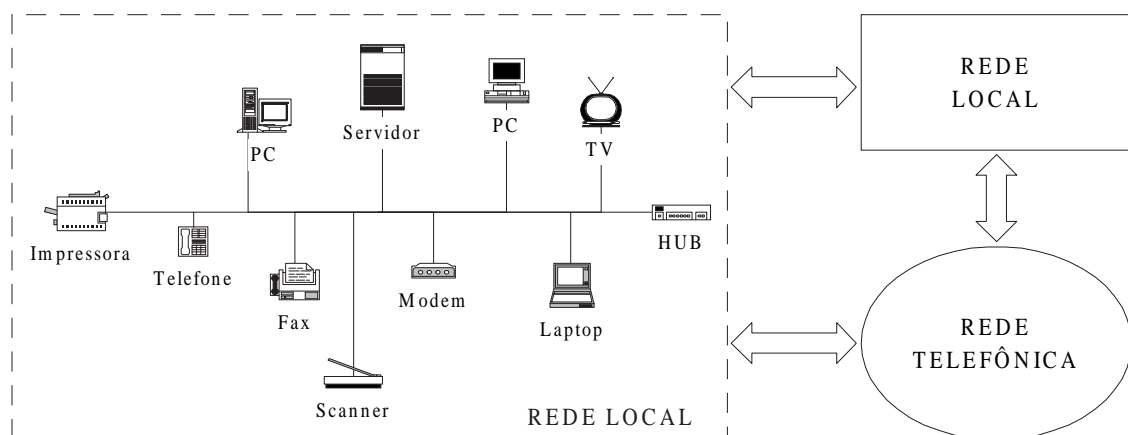


Figura 1.1: Rede de comunicação de serviços integrados.

1.1 Pequena história da evolução da comunicação

Alguns fatos e datas importantes na evolução das comunicações:

1876 - Alexander Graham Bell patenteia o telefone.

1880 - Começam a ser estabelecidas linhas telefônicas de longa distância.

1894 - O italiano Guglielmo Marconi inventa o rádio.

1900 - O brasileiro Roberto Landell de Moura, o Padre Landell, realiza a primeira transmissão de voz humana através de um sinal de rádio. A distância entre o transmissor

²O termo provém do inglês “equalizer” (usado nos EUA) ou “equaliser” (usado na Inglaterra), mas encontra-se incorporado ao vocabulário da língua portuguesa como sinônimo do dispositivo capaz de corrigir alguma distorção do sinal.

e o receptor foi de 8 km.

1901 - Marconi faz a primeira transmissão transcontinental de rádio, utilizando código Morse.

1927 - É lançado o primeiro aparelho telefônico que tem o transmissor e receptor em uma única peça. No mesmo ano é feita a primeira demonstração de transmissão de televisão à cabo.

1944 - Nos Estados Unidos, começa a funcionar o Mark 1, o primeiro computador a entrar em operação.

1946 - Entra em operação o primeiro sistema de telefonia móvel, ligando veículos ao sistema convencional de telefonia através de rádio. Neste mesmo ano começam a ser utilizados comercialmente os primeiros cabos coaxiais.

1947 - Começa a ser utilizada a transmissão por microondas para ligações de longa distância.

1961 - A IBM lança o teleprocessamento, sistema de transmissão de dados de um computador para o outro através de linhas telefônicas.

1962 - Lançado o Telstar 1, o primeiro satélite internacional de comunicações.

1969 - O Departamento de Defesa Americano começa a organizar a rede de computadores que daria origem à Internet.

1974 - É proposto o protocolo TCP/IP.

1977 - Começam a ser testados os primeiros telefones celulares.

1981 - A IBM lança o “Personal Computer” (PC).

1984 - A Motorola vende o primeiro aparelho de telefone celular.

A partir da década de 80, o intenso desenvolvimento na informática marcou a comunicação de dados. A Internet caracterizou-se como sendo a maior rede de troca de informações do mundo, forçando o avanço de toda uma tecnologia de integração de sistemas.

Outro grande marco também nas comunicações foi a explosão nas comunicações móveis pessoais. O número exponencialmente crescente de acessos na telefonia móvel abre um novo campo, a “wireless communication”.

A nova era de desenvolvimento tem se caracterizado como uma busca na melhoria da qualidade de serviços (QoS³) e desenvolvimento de sistemas cada vez mais robustos.

1.2 Comunicação digital e desconvolução

A operação de um sistema de comunicação digital (SCD) consiste, de uma maneira bastante simplista, na transmissão através de um meio físico de propagação de uma informação. Deste ponto de vista, a restrição que o meio físico, também chamado de *canal de comunicação* (ou simplesmente canal) impõe à taxa de transmissão é a inserção, na informação transmitida, de perturbações não desejadas. Estas perturbações estão associadas à limitação em faixa do canal.

³Do inglês “Quality of Service”.

De fato, considerando que o canal é modelado por um filtro linear⁴, é conhecido, da teoria de sistemas lineares, que a saída de um sistema linear com memória é a convolução de sua resposta impulsiva com o sinal de entrada.

A recuperação da informação deve ser capaz de extrair a partir dos dados recebidos (saída do sistema) a informação transmitida (entrada do sistema). Para este tratamento nos dados recebidos dá-se o nome de *equalização*. Desta maneira o problema de equalização é, na sua essência, um problema de *desconvolução* [Haykin, 1994a].

Utilizando um processamento adaptativo para a desconvolução, pois sabe-se que os filtros adaptativos podem fornecer um desempenho bem melhor que os filtros fixos além de serem adequados à utilização em ambientes de estatísticas desconhecidas [Qureshi, 1985], podemos então classificar a equalização adaptativa de duas formas:

- *Equalização Supervisionada* - onde a existência de uma seqüência de treinamento numa fase de aquisição se faz necessária para o ajuste do filtro.
- *Equalização Não Supervisionada ou Cega ou Autodidata*⁵ - aqui a adaptação do filtro é feita apenas com base no conhecimento *a priori* de certas propriedades estatísticas dos sinais de entrada do sistema.

A Figura 1.2 apresenta uma representação simplificada do problema de desconvolução cega.

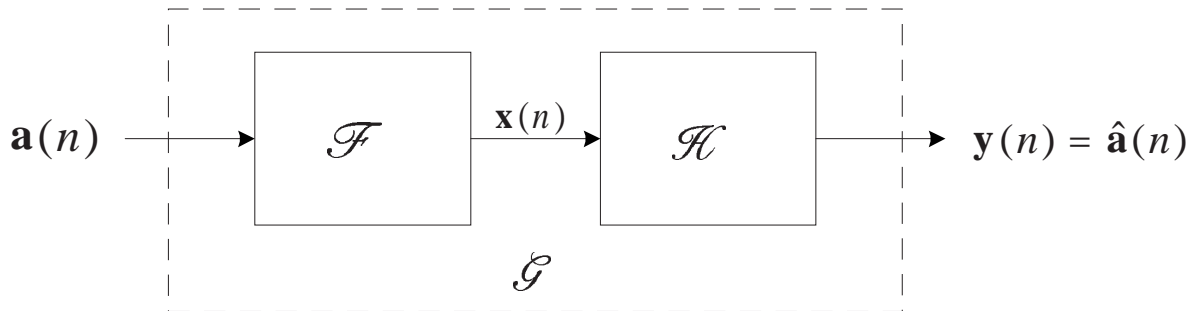


Figura 1.2: Síntese do problema de desconvolução.

O filtro que implementa o equalizador, representado por \mathcal{K} na Figura 1.2, deve então fazer uma operação tal que, a resposta global do sistema \mathcal{G} , a menos de um atraso, seja igual à unidade. Desta maneira, desconsiderando o ruído⁶ e supondo que \mathcal{F} não apresente nulos espectrais [Montalvão et al., 1999], tem-se a chamada *identificação inversa* onde $\mathcal{K} = \mathcal{F}^{-1}$. Ao considerar o ruído não se tem uma inversão perfeita e para a recuperação da potência de saída o uso de um quantizador é imperativo [Proakis, 1995].

⁴A justificativa de considerar-se um modelo linear para o canal será feita no capítulo 2.

⁵Os termos, cego e autodidata são sinônimos na literatura, embora alguns autores considerem o termo cego de conotação pejorativa. Neste trabalho ambos serão usados indistintamente.

⁶A consideração do ruído na cadeia do SCD será realizada oportunamente.

As estatísticas *a priori* da saída do sistema não sendo conhecidas, o uso de filtros adaptativos capazes de rastrear as características da interferência do canal faz-se necessário. Desta maneira a adaptação do filtro deve ser feita segundo alguma estratégia que implemente um algoritmo de adaptação.

A disponibilidade de uma seqüência de dados para o treinamento determina a forma de ajuste através do erro verdadeiro, ou seja, o sinal desejado menos a saída do filtro. Se há alguma indisponibilidade de tal seqüência, a estratégia a ser adotada pode conter, por exemplo, uma estimativa do sinal desejado. Algumas técnicas desta estimativa serão abordadas posteriormente na seção 1.4.

No caso da desconvolução autodidata, a indisponibilidade ou inexistência de uma réplica dos dados transmitidos (seqüência de treinamento) no receptor está associada à própria concepção do sistema, como, por exemplo, as redes de transmissão de dados multiponto (descrita na Figura 1.3). Neste caso, seria necessário o treinamento de todo o conjunto de terminais remotos a cada nova inclusão de um terminal, inviabilizando um bom desempenho das redes. Outro exemplo onde a réplica dos dados pode ter seu efeito comprometido é oferecido pelos sistemas de comunicação móvel, onde a característica variante no tempo do canal exigiria um treinamento freqüente, diminuindo assim a transmissão efetiva de dados em ambos os casos.

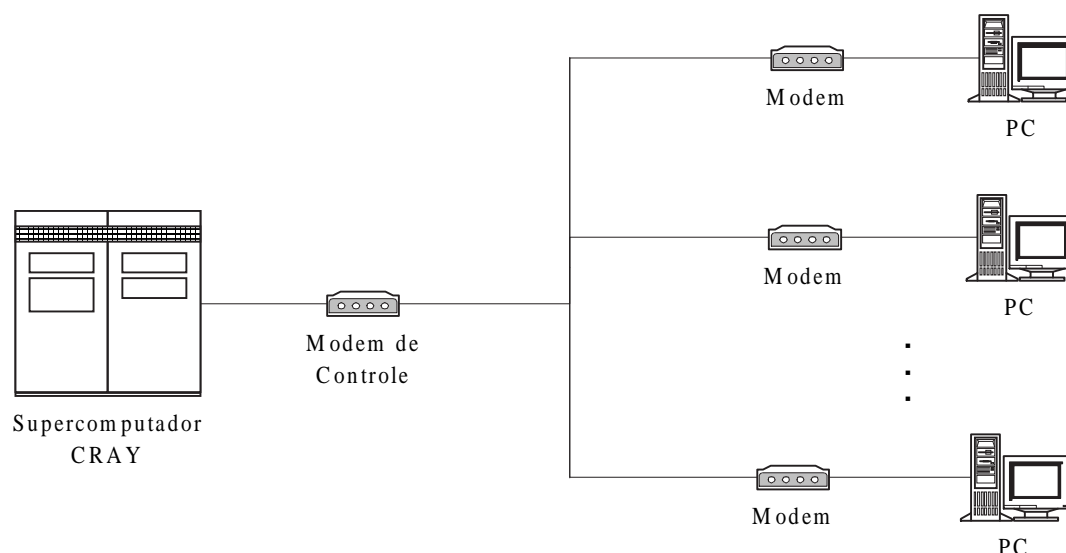


Figura 1.3: Rede de dados multiponto.

Com isso, o uso de técnicas autodidatas vêm sendo bastante investigado por se apresentar como uma alternativa plausível para o problema de equalização.

1.3 Contaminação do sinal

Nos SCD modernos, as taxas de transmissão alcançam valores de grande ordem. Nestes casos, a proximidade de um pulso a um outro durante a transmissão pode torná-los praticamente idênticos, surgindo assim a IES.

Como já mencionando anteriormente, a IES é causada devido a duração de um pulso por um período de tempo maior do que o período de amostragem T . A propagação por multipercursos (ex: sistemas de comunicações móveis) e a limitação de faixa (ex: canais telefônicos), são alguns fatores responsáveis pelo surgimento da IES [Rocha, 1996]. A Figura 1.4 exemplifica a idéia de multipercurso em uma transmissão.

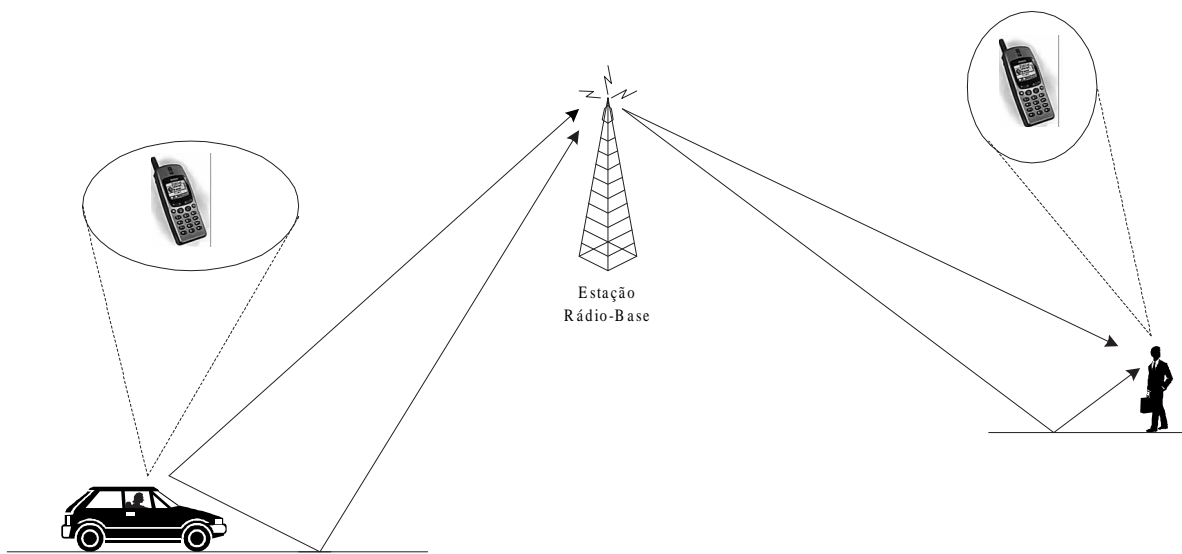


Figura 1.4: Multipercurso em comunicações móveis.

A dificuldade de supressão (ou diminuição) da IES a partir dos sinais disponíveis no receptor impõe-se a partir do fato do não conhecimento *a priori* e em tempo real das estatísticas do sinal recebido, motivando assim o uso de estratégias adaptativas.

A estratégia de como o equalizador realiza o tratamento da IES é baseada em três aspectos: estrutura de filtragem, critério e algoritmo de adaptação. A seção 1.4 discute algumas das técnicas clássicas de equalização adaptativa.

1.4 Equalização adaptativa: técnicas clássicas e novas tendências

O histórico da equalização tem seu marco inicial dentro do uso da teoria de filtragem linear. Embora os conceitos da filtragem linear datem do século XVIII com Gauss, tudo

indica que Norbert Wiener foi o primeiro a fazer aplicações em engenharia [Paulraj et al., 1997].

A teoria de filtragem desenvolvida por Wiener possibilitou a implementação da filtragem inversa, já descrita na seção 1.2. A solução encontrada para as equações de Wiener-Hopf [Haykin, 1996] é mostrada como sendo a melhor solução para o uso de estruturas lineares [Mota, 1992].

Uma dificuldade presente no uso da solução de Wiener é que necessita do conhecimento *a priori* de médias estatísticas de sinais do sistema. Desta maneira em situações práticas, onde o emprego em tempo real da solução de Wiener é praticamente impossível, faz-se necessário a utilização de uma estratégia adaptativa no tempo. Portanto, a busca pela identificação inversa é feita através de um algoritmo de adaptação no tempo dos coeficientes do filtro.

A proposição do algoritmo LMS⁷ por Widrow e Hoff em 1960, entre outros, tornou possível a implementação dos primeiros sistemas ajustáveis, ou adaptativos, que tornaram peça fundamental no desenvolvimento das comunicações digitais através dos “modems”⁸.

Algumas estratégias de equalização adaptativa com treinamento fazem uso do algoritmo LMS para a adaptação dos coeficientes do filtro. Nestes casos, o erro considerado é calculado através de uma referência, dada por uma seqüência de treinamento em um período de aquisição. Na indisponibilidade de um símbolo para o cálculo do erro, o mesmo deve ser estimado para construção do algoritmo de adaptação.

Algumas das mais importantes técnicas de estimação do símbolo transmitido são as técnicas de Bussgang. Estas técnicas tomam por base uma certa função de distribuição de probabilidade (fdp) dos sinais transmitidos para construção dos estimadores. Tais técnicas serão abordadas em mais detalhes no capítulo 5.

1.4.1 Estratégias preditivas de equalização autodidata

Outra estratégia de equalização cega foi proposta por Shalvi e Weinstein [Shalvi & Weinstein, 1990] onde há a hipótese dos dados transmitidos formarem uma seqüência de variáveis aleatórias independentes. Esta estratégia apresenta uma velocidade de convergência baixa.

A hipótese de trabalhar-se com seqüências de dados formadas por variáveis aleatórias decorrelacionadas foi também explorada em [Macchi & Hachicha, 1986] e posteriormente em [Macchi & Gu, 1987], mostrando que, se esta hipótese é satisfeita, o problema de equalização pode ser abordado por predição.

As técnicas de predição apresentam uma diferença básica em relação às técnicas de Bussgang. As técnicas de Bussgang utilizam uma estimativa de máxima verossimilhança,

⁷Do inglês “Least Mean Square”.

⁸Do inglês “Modulator/Demodulator”.

enquanto às técnicas preditivas se baseiam no branqueamento da sequência de sinais na saída do equalizador. Outra diferença entre as estratégias é no uso das estruturas de filtragem. Enquanto nas técnicas de Bussgang faz-se normalmente uso dos filtros transversais, as técnicas preditivas usam filtros de predição e filtro de retropredição.

O trabalho de Macchi e Hachicha realizou apenas a equalização de amplitude de um canal de comunicação, mas iniciou uma série de investigações em torno da idéia do uso de equalizadores preditivos para a equalização de amplitude e fase de canais de comunicação [Rocha & Macchi, 1994; Cavalcanti, 1996; Macchi, 1998; Labat et al., 1998]. Estes trabalhos propunham modificações na estrutura preditiva e uma função custo que levava em conta os momentos de ordem superior (MOS).

1.4.2 Redes neurais artificiais e equalização autodidata

O surgimento das redes neuronais⁹ artificiais (RNA) expandiu o horizonte do processamento inteligente de sinais. A formulação de uma ferramenta capaz de *aprender* alguma informação que lhe seja apresentada, possibilitou também a aplicação em diversos campos da engenharia.

Na área de processamento de sinais, as RNA ganharam bastante espaço no processamento de voz, processamento de imagens, codificação entre outros [Svarer, 1991].

Algumas das características das RNA, permitiu a concepção de filtros não lineares [Mulgrew, 1996], fertilizou a imaginação dos pesquisadores ao tentarem buscar uma alternativa para fazer frente ao aumento da potência de ruído existente quando do uso de estruturas lineares transversais [Mota, 1992].

A partir da década de 90 surgiram vários trabalhos explorando as RNA como filtros não lineares nos problemas sobre equalização adaptativa para canais de comunicação de SCD. Alguns trabalhos [Balay, 1995] mostraram uma análise comparativa das RNA com as estruturas transversais clássicas.

O uso de RNA para predição de séries temporais com bons resultados também já é bastante vasto na literatura [Barreto, 1998; Boné, 2000; Czernichow, 1996]. Neste aspecto o uso de uma estrutura preditiva para a equalização de dados parece uma estratégia plausível.

A proposta principal deste trabalho é a de investigar a utilização de uma estrutura neuronal como filtro de erro de predição para equalização de canais de sistemas de comunicação digital.

⁹Não existe, na gramática portuguesa, um consenso sobre um adjetivo para descrever a relação aos neurônios. Embora a tradução direta do inglês “neural” pareça mais freqüente na literatura técnica, nesta dissertação será usado também o termo *neuronal*.

1.4.3 Critérios estatísticos para equalização autodidata

Os critérios estatísticos são a base da teoria de equalização autodidata. Isto torna-se evidente da técnica de Bussgang, onde a suposição de uma determinada fdp da saída do equalizador é imperativa para a realização da MLE da saída do equalizador com a fdp da sequência de dados de transmissão.

O trabalho [Sala-Alvarez & Vázquez-Grau, 1997], que faz uma estimação da fdp da saída do equalizador para construção de um funcional, motivou nesta dissertação a investigação sobre funcionais de critério estatístico com base na suposição da fdp do sinal de saída do equalizador ser constituído de uma composição de gaussianas. Estes critérios incorporam necessariamente uma medida de similaridade entre fdps, levando ao critério de minimização da entropia através da divergência de Kullback-Leibler [Haykin, 1998].

Esta dissertação propõe também uma nova estratégia para a equalização adaptativa, utilizando estruturas de filtragem transversal com um critério cujo funcional é composto de fdps gaussianas e o processo de otimização é baseado na máxima verossimilhança da fdp do sinal na saída do equalizador.

1.5 Organização, estrutura e contribuições

O restante desta dissertação se organiza da seguinte maneira:

- **Capítulo 2 - Modelo do Sistema de Comunicação Digital**
Apresenta o modelo do sistema que será utilizado no restante do trabalho. A revisão de alguns conceitos sobre sistemas de transmissão digital necessários para a compreensão da proposição de uma nova estrutura de equalizador é também contemplada.
- **Capítulo 3 - Modelos Baseados em Teoria da Informação**
Neste capítulo são explorados e desenvolvidos alguns conceitos matemáticos e definições da teoria da informação com ênfase na estimação de função densidade de probabilidade. Este capítulo fornece conhecimento fundamental para o desenvolvimento dos algoritmos propostos.
- **Capítulo 4 - Redes Neurais Artificiais**
Os conceitos básicos sobre redes neurais e suas principais arquiteturas são abordados neste capítulo. Uma abordagem do aprendizado das redes neurais artificiais é também objetivo deste capítulo.
- **Capítulo 5 - Equalização e Classificação: Recuperação da Informação**
Este capítulo descreve as técnicas de equalização clássicas e discorre sobre a abordagem da classificação, ambas objetivando a recuperação da informação. A descrição e compreensão da abordagem da classificação ganha importância devido à proposição de estratégias que levam em conta estes aspectos para uma proposta.

- **Capítulo 6** - *Predição Não Linear para Equalização Autodidata*

A proposição de um equalizador não linear baseado nos conceitos de predição, através da abordagem da classificação, e através do uso de redes neurais é apresentada nesse capítulo. A elaboração de um novo algoritmo de adaptação para os parâmetros da rede neuronal é desenvolvido.

- **Capítulo 7** - *Critério Estatístico Baseado na Estimação da Função de Densidade de Probabilidade*

Este capítulo também apresenta uma proposta de inovação na utilização de estruturas lineares para equalização de canais de comunicação digital. Um novo funcional baseado na estatística dos dados na saída do equalizador é desenvolvido. Comparações com outros critérios também são contempladas.

- **Capítulo 8** - *Conclusões e Perspectivas*

- **Apêndice A** - *Derivação de uma Função Escalar em Relação a um Vetor*

- **Apêndice B** - *Sumário de Algoritmos Propostos na Dissertação*

- **Apêndice C** - *Artigos Publicados*

PARTE **I**

CONCEITOS BÁSICOS

MODELO DO SISTEMA DE COMUNICAÇÃO DIGITAL

A definição do modelo de um sistema de comunicação digital é um aspecto de extrema importância para a compreensão do funcionamento da cadeia de comunicação. Todo e qualquer dispositivo a ser inserido no sistema, não importando sua função, tem como base de concepção tal modelo.

Este capítulo mostra alguns conceitos fundamentais necessários para a compreensão de um sistema de comunicação digital, bem como esclarece e determina o modelo a ser usado no restante da dissertação.

A Seção 2.1 mostra os conceitos genéricos de um sistema de comunicação digital. A representação do sistema em banda básica equivalente é contemplada na Seção 2.2, e considerações sobre a fonte de informação na Seção 2.3. O modelo a ser adotado para o canal é desenvolvido na Seção 2.4 e na Seção 2.5 são apresentadas as medidas de desempenho a serem utilizadas.

2.1 Cadeia de comunicação digital

Por volta de 150 a.c. o Império Romano possuía uma rede de *telégrafos* a sinais de fumaça com mais de 4.500 quilômetros de extensão. Após mais de 2.000 anos, os SCD guardam os mesmos princípios básicos.

A codificação da informação é feita através dos elementos de um conjunto finito,

denominado *alfabeto*. Cada um dos elementos do alfabeto é então denominado *símbolo* e cada um dos símbolos será usado a cada T segundos, que por sua vez define o intervalo de amostragem do sistema [Proakis, 1995; Benedetto et al., 1987].

O processo de transmissão e recepção da informação exige um tratamento para cada etapa da informação na cadeia de comunicação. A Figura 2.1 mostra um sistema de comunicação genérico [Haykin, 1994b].

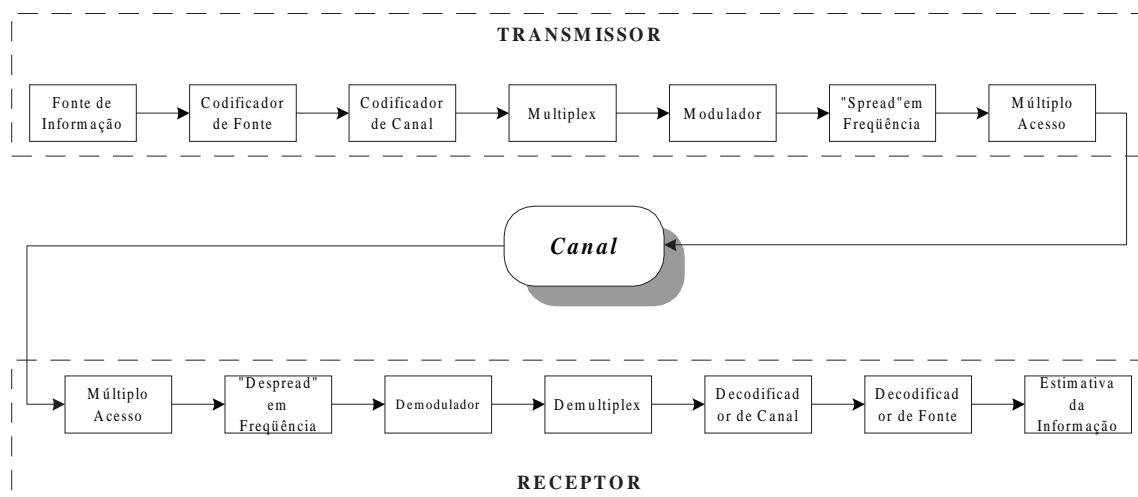


Figura 2.1: Sistema de comunicação digital genérico.

O codificador de fonte é responsável por formar a partir da sequência a ser transmitida a menor sequência possível contendo as mesmas informações. O codificador de canal insere redundâncias nesta sequência a fim de detectar e corrigir erros durante a transmissão. A multiplexação possibilita a utilização do mesmo meio de transmissão por outras fontes. Neste ponto, é importante salientar que a informação será convertida em formas de onda para a transmissão e logicamente esta informação ocupará uma determinada faixa de frequência que pode-se alterar através do espalhamento em frequência, representado na Figura 2.1 pelo bloco “*Spread*” em *Frequência* [Papadias, 1995].

A descrição acima corresponde ao procedimento realizado pelo *transmissor*. No receptor o processamento se faz na ordem inversa para se recuperar a informação na forma original, ou fazer uma estimativa da mesma.

No receptor, onde as perturbações geradas pelo canal podem provocar uma má estimativa da informação, deve existir um dispositivo, um equalizador, capaz de retirar (ou minimizar) as perturbações inseridas pelo canal.

O estudo desta dissertação será focalizado, então, no receptor de um sistema de transmissão. Entretanto, na sequência deste trabalho, o modelo de sistema que será utilizado será o modelo em *banda básica equivalente*, descrito na Seção 2.2. Nesta seção serão melhor descritos os principais componentes do SCD a serem explorados nesta dissertação.

2.2 Modelo em banda básica equivalente

Em um SCD, pode-se modelar o conjunto modulador, canal e demodulador como um único filtro analógico [Proakis, 1995]. No caso da transmissão digital, onde a entrada e a saída deste bloco são discretas no tempo, a inserção do amostrador nesta cadeia (não representado na Figura 2.1) possibilita a troca deste filtro por seu equivalente discreto [Oppenheim & Schaffer, 1989]. A este modelo, representado na Figura 2.2, dá-se o nome de *modelo em banda básica equivalente* (MBBE).

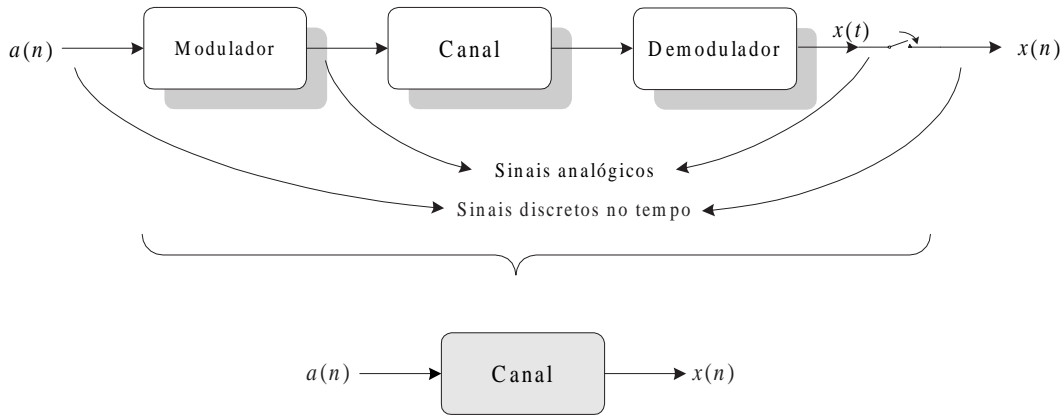


Figura 2.2: Representação em banda básica equivalente de um sistema de transmissão digital.

O uso do MBBE simplifica a formulação do problema de equalização¹. Neste caso, o equalizador fica responsável pelo tratamento discreto no tempo da distorção inserida pelo canal. Assim, o problema de equalização em um SCD pode ser simplificado para o esquema representado na Figura 2.3.

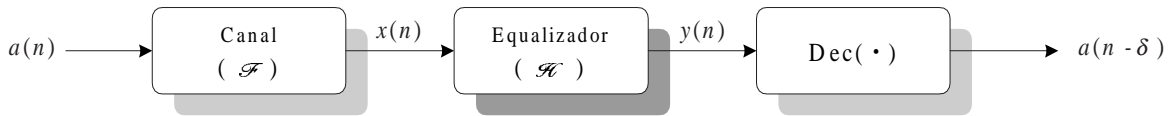


Figura 2.3: Modelo discreto no tempo do problema de equalização em um sistema de comunicação digital.

O dispositivo de decisão ($\text{Dec}(\cdot)$) ou quantizador é necessário para restituir à saída do equalizador o mesmo valor de potência dos símbolos transmitidos, bem como projetar o sinal no espaço de representação finito da informação.

¹Em todo o decorrer desta dissertação o MBBE será utilizado nos desenvolvimentos teóricos e nas simulações.

2.3 Fonte de informação

Como já visto na seção anterior, a fonte do SCD, usando o MBBE, é a saída do modulador, sendo que a definição de como se produz tais sinais é de significativa importância.

Considera-se que a fonte do sistema de transmissão digital em banda básica equivalente fornece símbolos $a(n)$, escolhidos aleatoriamente a cada instante de tempo, com igual probabilidade, de um conjunto ou alfabeto finito de símbolos $\mathcal{A} = \{a_s : 1 \leq s \leq S\}$, produzindo uma seqüência no tempo de variáveis aleatórias (v.a.) discretas, $\{a(n)\}$, $n \in \mathbb{Z}$, independentes e identicamente distribuídas (i.i.d.).

O alfabeto também é denominado de *constelação* por apresentar semelhança do conjunto discreto com um grupo de estrelas [Rocha, 1996]. A Figura 2.4 ilustra três tipos de constelações para modulação digital.

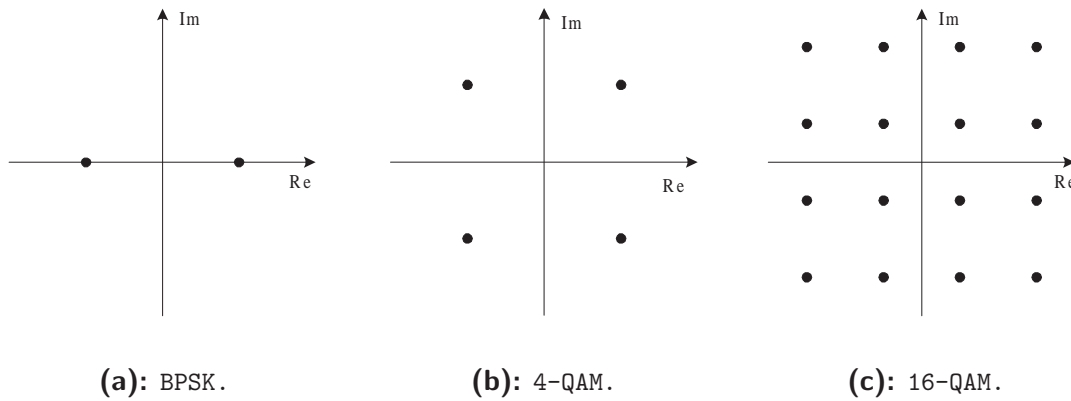


Figura 2.4: Três tipos de constelação.

A representação de uma seqüência temporal dos dados de comprimento finito na forma vetorial será freqüentemente preferida:

$$\mathbf{a}(n) = \begin{bmatrix} a(n) & a(n-1) & \cdots & a(n-k+1) \end{bmatrix}^T,$$

em que k é a ordem do vetor. No caso de necessitar-se explicitar deterministicamente os valores assumidos pelas v.a. na seqüência, será utilizada a seguinte notação:

$$\mathbf{a}_i(n) = \begin{bmatrix} a_0 & a_1 & \cdots & a_{k-1} \end{bmatrix}^T,$$

em que o subscrito i indica uma realização determinada e os valores a_0, a_1, \dots, a_{k-1} são definidos em \mathcal{A} . Os valores de tal realização serão então descritos explicitamente.

Neste trabalho a modulação empregada nos casos em simulação ou desenvolvimento teórico será claramente assinalada.

2.4 Modelo do canal de comunicação

O canal de comunicação desempenha um papel fundamental no sistema de comunicação. Suas características e restrições determinam os limitantes para um sistema. Naturalmente, seu modelamento determina, da mesma maneira, aspecto de extrema importância.

Na grande maioria dos casos de comunicação, o modelo do canal é assumido *linear*. Esta aproximação é boa se considerar que a propagação da energia se dá através de uma perturbação física imposta ao meio e se a perturbação, que depende das características do meio, é limitada a uma certa faixa de potência [Montalvão, 2000].

Outro aspecto que favorece considerar o modelo linear deve-se ao extenso conhecimento na literatura sobre sistemas lineares, permitindo, assim, uma melhor análise.

Por outro lado, ao considerar-se os dispositivos eletrônicos, que possuem, em geral, características altamente não lineares como, por exemplo, os amplificadores tipo AB, o teorema da superposição linear não mais pode ser aplicado em toda a faixa dinâmica do dispositivo. Estes tipos de situações são freqüentes nos casos dos canais de sistemas satelitários.

Outra hipótese largamente empregada diz respeito à invariância no tempo do canal. Levando-se em consideração sistemas onde os meios de propagação são fixos, por exemplo sistemas telefônicos fixos, esta consideração é bastante aceitável, já que as variações que podem ocorrer devido a fatores externos (umidade, calor, etc.) apresentam-se de maneira muito lenta.

Entretanto, ao observar sistemas móveis, observa-se que as características sofrem mudanças muito mais rapidamente. Neste caso, a consideração que permite tratar o canal como invariante é considerar a taxa de transmissão muito maior que a taxa de variação dos parâmetros. Tal consideração pode ser utilizada em sistemas móveis tipo GSM (“Global System for Mobile Communications”) que utiliza taxas de transmissão elevadas.

Nesta dissertação será usado o modelo de canal linear invariante no tempo. Alguma exceção será claramente sinalizada.

2.4.1 Ruído gaussiano

As perturbações físicas inerentes nos SCD, tais como ruído térmico, têm um tratamento geralmente simplificado. Nestes casos, representa-se genericamente tais perturbações por um termo ruidoso no modelo do sinal.

O ruído é modelado como sendo um sinal representado por variáveis aleatórias com distribuição gaussiana de média nula, variância σ_b^2 e espectro de potência branco. Este modelo justifica-se através do Teorema Central do Limite [Papoulis, 1991] usando-se uma consideração de que as diversas fontes de ruído são independentes.

Assim, o modelo do canal ao considerar o ruído pode ser representado na Figura 2.5.

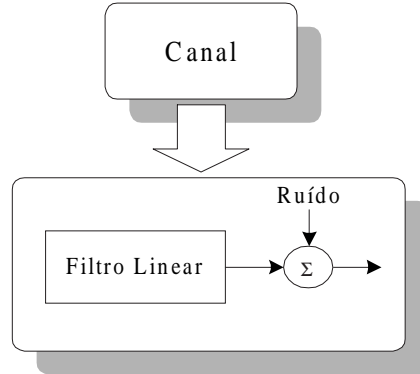


Figura 2.5: Modelo do canal de comunicação.

Um caso particular deste modelo, representado na Figura 2.5, considera o filtro linear sem provocar distorção. Neste caso, o canal é conhecido como *canal AWGN*².

No decorrer deste trabalho, o ruído será representado pela v.a. $b(n)$, e uma seqüência temporal finita da v.a. do ruído será representada através da notação vetorial:

$$\mathbf{b}(n) = \begin{bmatrix} b(n) & b(n-1) & \cdots & b(n-k+1) \end{bmatrix}^T,$$

em que k é a ordem do vetor.

2.4.2 Função de transferência

A representação do canal através de sua função de transferência (FT) decorre normalmente a partir de sua linearidade. A FT discreta do canal pode ser obtida a partir da sua FT contínua através da Equação (2.1):

$$f_{discreta}(t) = f(t) \cdot \sum_{n=-\infty}^{\infty} \delta(t - nT), \quad (2.1)$$

em que T é o intervalo de amostragem, $f(t)$ é a resposta impulsiva do canal e $\delta(\cdot)$ é a função delta de Dirac [Proakis, 1995]. Outra representação é através da transformada de Fourier da função $f_{discreta}(t)$, que devido à amostragem regular leva à transformada \mathcal{Z} [Oppenheim & Schaffer, 1989], da seqüência discreta $\{f(nT)\}$. Assim, a representação da FT do canal discreto é:

$$F(z) = \sum_{n=-\infty}^{\infty} f_n z^{-n}, \quad (2.2)$$

²Do inglês “Additive White Gaussian Noise”.

em que $z = \rho e^{jwT}$, w é a frequência digital e $f_n = f(nT)$. Desta maneira a FT do canal será dada por um polinômio em z .

Uma consideração que também será feita é sobre o comprimento da resposta ao impulso. Será considerado um tamanho *finito* para a FT do canal, resultando desta maneira num filtro de resposta impulsiva finita (FIR³). Pode-se então reescrever a FT do canal como:

$$F(z) = \sum_{n=0}^{N-1} f_n z^{-n}, \quad (2.3)$$

em que N é o tamanho da resposta impulsiva do canal, também chamada de *ordem do canal*.

A representação vetorial \mathbf{f} da resposta ao impulso (RI) do canal será bastante utilizada durante este trabalho:

$$\mathbf{f} = [f_0 \ f_1 \ \cdots \ f_{N-1}]^T.$$

Através desta representação pode-se então escrever a operação de convolução discreta dos dados provenientes do modulador e resposta ao impulso do canal para produzir a saída do canal como:

$$x(n) = \sum_{i=0}^{N-1} f_i a(n-i) + b(n). \quad (2.4)$$

As v.a. $x(n)$ serão chamadas de observações, ou ainda saídas do canal. As observações não-ruídas serão chamadas *estados do canal* e representadas por $\bar{x}(n)$, onde $\bar{x}(n) = \sum_{i=0}^{N-1} f_i a(n-i)$, e que também podem ser representadas pela operação matemática dada por $\bar{x}(n) = \mathbb{E}\{x(n)|\mathbf{a}(n)\}$, em que $\mathbb{E}\{\cdot\}$ é o operador esperança matemática.

Em resumo, a representação vetorial para o sistema em banda básica equivalente será:

$$\begin{aligned} \mathbf{a}(n) &= [a(n) \ a(n-1) \ \cdots \ a(n-N-M+2)]^T && \text{(seq. transmitida aleatória)} \\ \mathbf{a}_i(n) &= [a_0 \ a_1 \ \cdots \ a_{N+M-2}]^T && \text{(seq. determinística de dados)} \\ \mathbf{f} &= [f_0 \ f_1 \ \cdots \ f_{N-1}]^T && \text{(RI do canal)} \\ \mathbf{b}(n) &= [b(n) \ b(n-1) \ \cdots \ b(n-M+1)]^T && \text{(seq. de ruído)} \\ \mathbf{x}(n) &= [x(n) \ x(n-1) \ \cdots \ x(n-M+1)]^T && \text{(seq. de saídas do canal)} \\ \bar{\mathbf{x}}(n) &= [\bar{x}(n) \ \bar{x}(n-1) \ \cdots \ \bar{x}(n-M+1)]^T && \text{(seq. de estados do canal).} \end{aligned}$$

2.4.3 Tipos de canais

Ao analisar-se a posição dos zeros da FT de um canal sem zeros sobre a circunferência de raio unitário (CRU) pode-se fazer a seguinte classificação: canais de fase mínima (FMín),

³Do inglês “Finite Impulse Response”.

fase máxima (FMáx) ou fase não mínima (FNM).

É fato também conhecido que um canal de fase não mínima genérico poder ser representado por uma composição de suas fase mínima e fase máxima [Mota, 1992]. Assim, sendo a FT em z dada pela Equação (2.2), pode-se escrevê-la como:

$$F(z) = f \cdot F_{Min}(z) \cdot F_{Max}(z), \quad (2.5)$$

em que

$$F_{Min}(z) = \prod_{i=1}^N (1 - a_i z^{-1}) \quad |a_i| < 1, \quad (2.6a)$$

$$F_{Max}(z) = \prod_{i=1}^N (z^{-1} - b_i^{-1}) \quad |b_i| > 1, \quad (2.6b)$$

e f é uma constante (fator) de ganho, que pode ser complexa ou não.

Destaca-se ainda o fato de que a proximidade dos zeros da CRU determina uma maior ou menor distorção na amplitude e/ou fase da resposta em frequência [Mota, 1992].

2.4.3.1 Fase mínima

Em geral, os canais de fase mínima caracterizam-se por possuir uma frente de descida, onde a duração desta frente de descida depende da ordem do canal. Desta maneira, os zeros da FT dos canais de fase mínima encontram-se todos *dentro* da CRU. Isto se reflete na caracterização da FT em possuir o primeiro coeficiente como o mais forte [Mota, 1992]. A Figura 2.6 exemplifica as características de um canal de fase mínima.

2.4.3.2 Fase máxima

Em geral, os canais de fase máxima caracterizam-se por possuir uma frente de subida, sendo que a duração desta frente de subida depende da ordem do canal. Desta maneira, os zeros da FT dos canais de fase máxima encontram-se todos *fora* da CRU. Isto se reflete na caracterização da FT em possuir o último coeficiente como o mais forte [Mota, 1992]. A Figura 2.7 exemplifica as características de um canal de fase máxima.

2.4.3.3 Fase não mínima

Em geral, os canais de fase não mínima caracterizam-se por possuir uma frente de descida após uma frente de subida, sendo que a duração destas frentes de subida e descida

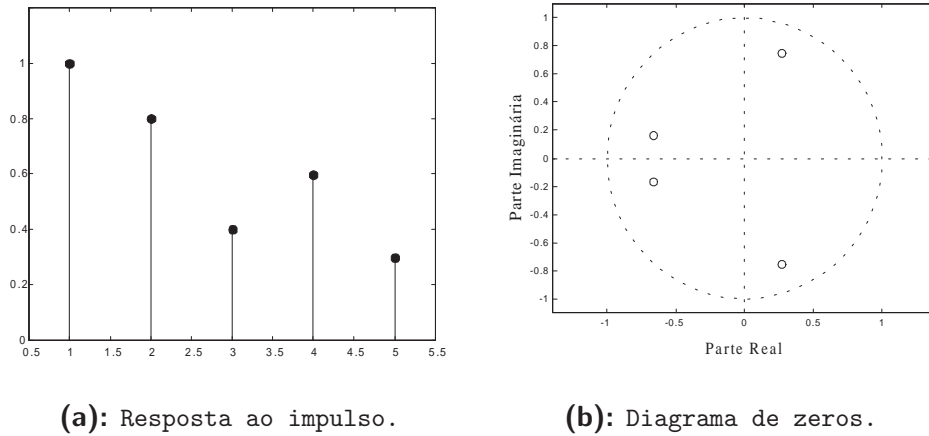


Figura 2.6: Canal de fase mínima.

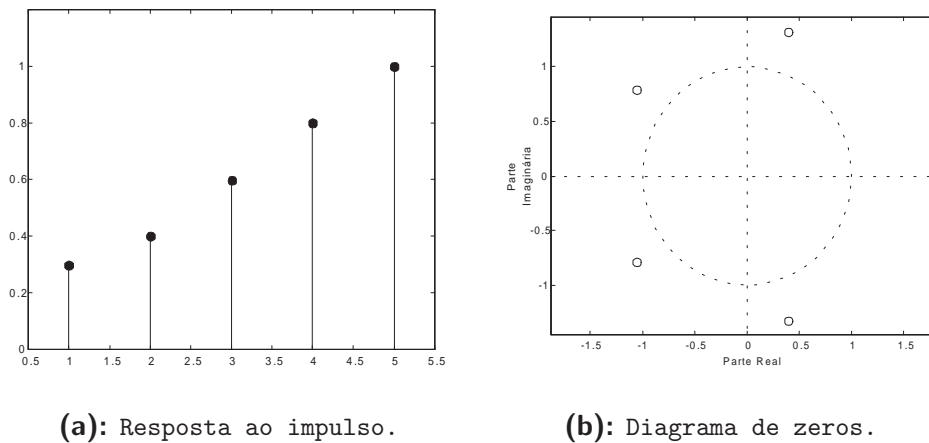


Figura 2.7: Canal de fase máxima.

dependem da ordem do canal. Desta maneira, os zeros da FT dos canais de fase máxima encontram-se *dentro* e *fora* da CRU. Isto se reflete na caracterização da FT em possuir um coeficiente intermediário como o mais forte [Mota, 1992]. A Figura 2.8 exemplifica as características de um canal de fase não mínima.

2.4.4 Interferência entre símbolos

A interferência entre símbolos já mencionada anteriormente, havia sido definida apenas conceitualmente, neste ítem desenvolve-se a formulação matemática da IES.

Considere a saída de um canal dada por:

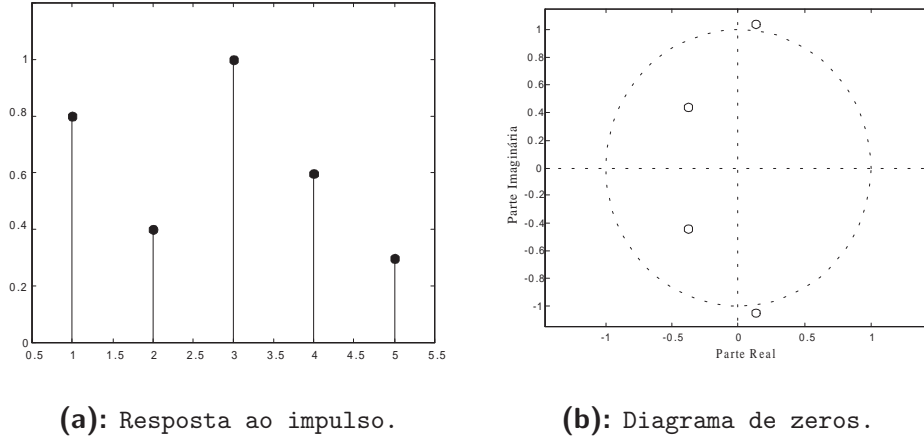


Figura 2.8: Canal de fase não mínima.

$$x(n) = \mathbf{f}^T \mathbf{a}(n) + b(n),$$

pode-se então representar na forma:

$$x(n) = \sum_{k=0}^{N-1} f_k a(n-k) + b(n). \quad (2.7)$$

Desta maneira, reescrevendo a Equação (2.7) e isolando-se o termo correspondente à amostra atrasada de δ , tem-se:

$$x(n) = a(n-\delta)f_\delta + \underbrace{\sum_{\substack{k=0 \\ k \neq \delta}}^{N-1} a(n-k)f_k}_{\text{IES}} + b(n). \quad (2.8)$$

A IES compõe uma das principais fontes de degradação do sinal, sendo dependente da resposta impulsiva do canal. Uma maneira de avaliar a IES em um sinal é através do *diagrama de olho*. O diagrama de olho consiste no traçado do sinal analógico com uma taxa de repetição igual à taxa de amostragem de símbolos $\frac{1}{T}$ [Proakis, 1995; Brossier, 1997]. A Figura 2.9 ilustra o diagrama de olho para sinais binários nas condições de *olho aberto* e *olho fechado*.

A abertura do olho em algum ponto indica o instante de amostragem favorável (atraso de decisão) a uma decisão correta dos dados recebidos. A não existência de um ponto de amostragem favorável indica a necessidade do uso de algum dispositivo capaz de *abrir o olho*. No caso dos SCD, onde não se conhece as características *a priori* do sistema, o dispositivo capaz de realizar tal tarefa é um equalizador adaptativo.

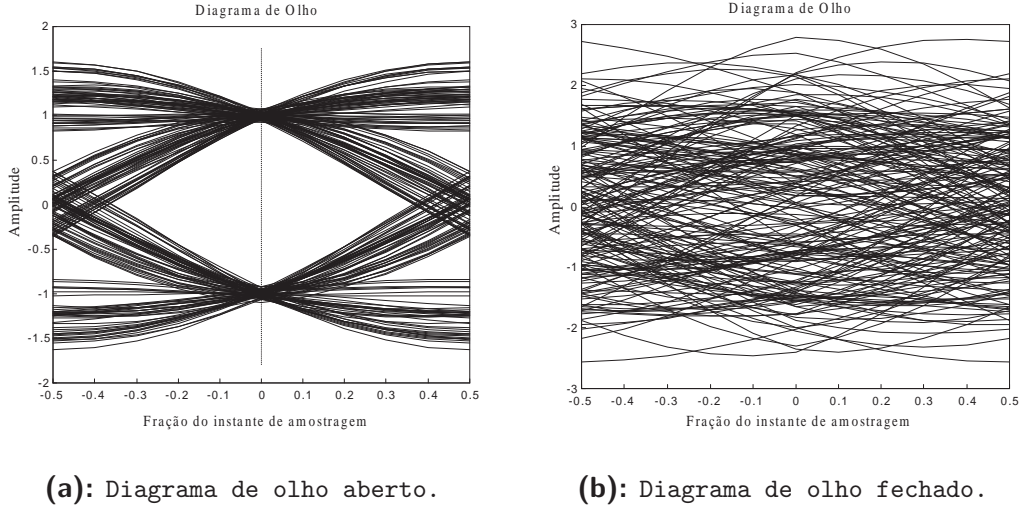


Figura 2.9: Diagrama de olho para sinais binários.

2.5 Medidas de desempenho

A seguir são descritas algumas medidas de desempenho que serão usadas no decorrer desta dissertação.

2.5.1 Relação sinal-ruído

A influência do ruído no sinal transmitido é medida através da Relação Sinal-Ruído (SNR⁴). Esta relação (avaliada geralmente em decibéis) é medida no receptor e leva em consideração a potência do sinal recebido bem como a potência do ruído. Define-se então a SNR em dB como:

$$\text{SNR} = 10 \cdot \log_{10} \left(\frac{\sigma_a^2 \cdot \sum_{i=0}^{N-1} |f_i|^2 + \sigma_b^2}{\sigma_b^2} \right), \quad (2.9)$$

em que σ_a^2 e σ_b^2 são as variâncias do sinal transmitido e do ruído, respectivamente e f_i é o i -ésimo coeficiente da FT em z do canal [Montalvão, 2000].

⁴Do inglês "Signal-to-Noise Ratio".

2.5.2 Taxa de símbolos errados

A taxa de símbolos errados (TSE) mede a quantidade de símbolos que são perdidos após o tratamento no receptor. O cálculo da TSE é realizado segundo a equação abaixo:

$$\text{TSE} = \frac{N_e}{N_t}, \quad (2.10)$$

em que N_e é o número de decisões erradas e N_t é o número total de símbolos transmitidos.

Os limites teóricos para a TSE são dependentes do tipo de modulação e características da FT do canal [Benedetto et al., 1987; Proakis, 1995].

2.5.3 Erro quadrático médio

A evolução no tempo do erro quadrático médio (EQM) tem utilidade na análise da convergência do equalizador.

A definição do erro depende da estratégia adotada:

- erro verdadeiro: neste caso tem-se *a priori* o conhecimento do sinal desejado.
- erro estimado: onde o sinal desejado é substituído pela sua estimativa.
- erro de decisão:⁵ a medida é feita através da diferença entre os sinais na saída e na entrada do quantizador (decisor).

É comum colocar-se o EQM mínimo oriundo da solução da equação de Wiener-Hopf como referência para a solução adaptativa.

Nesta dissertação o uso de um sinais de erro descritos acima será claramente mencionado.

2.5.4 Relação interferência entrada-saída

Uma outra medida de desempenho pode ser explorada em termos do ganho do sistema sobre a interferência do sinal.

Definindo-se a relação sinal-interferência (RSI) como a medida da razão entre as potências do sinal recebido e a potência da interferência na entrada (saída) do equalizador, tem-se:

⁵Sabe-se que o erro de decisão é também uma estimativa do erro verdadeiro, entretanto a sua classificação em separado deve-se ao fato de ser largamente utilizado no decorrer deste trabalho.

$$\text{RSI}_{\text{entrada}} = \frac{\sigma_a^2 + \sigma_{\text{IS}_{\text{entrada}}}^2}{\sigma_{\text{IS}_{\text{entrada}}}^2} \quad (2.11a)$$

$$\text{RSI}_{\text{saída}} = \frac{\sigma_a^2 + \sigma_{\text{IS}_{\text{saída}}}^2}{\sigma_{\text{IS}_{\text{saída}}}^2}, \quad (2.11b)$$

Isto porque considera-se que:

$$\begin{aligned} x(n) &= a(n) + \text{IS}_{\text{entrada}} \\ y(n) &= a(n - \delta) + \text{IS}_{\text{saída}}, \end{aligned} \quad (2.12)$$

Com isto, pode-se calcular a variâncias das interferências dos sinais como $\sigma_{\text{IS}_{\text{entrada}}}^2 = \sigma_x^2 - \sigma_a^2$ e $\sigma_{\text{IS}_{\text{saída}}}^2 = \sigma_y^2 - \sigma_a^2$. Então, uma melhoria na relação sinal-interferência da saída do equalizador em relação à sua entrada representa um *ganho* provocado pelo processamento. Geralmente as RSI serão calculadas em dB: $\text{RSI}(\text{db}) = 10 \cdot \log_{10}(\text{RSI})$. A Figura 2.10 indica as regiões do gráfico $\text{RSI}_{\text{entrada}} \times \text{RSI}_{\text{saída}}$.

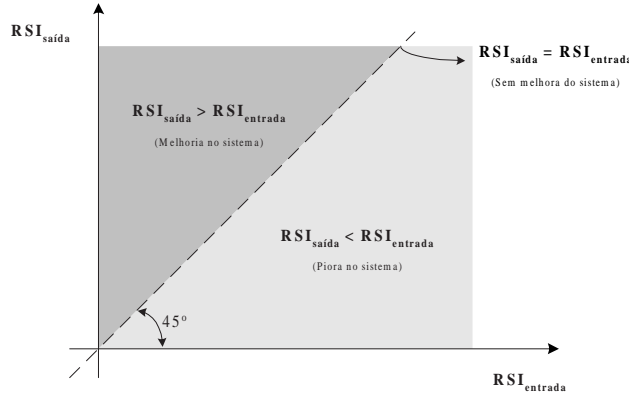


Figura 2.10: Gráfico de RSI de saída versus entrada.

2.5.5 Métodos de Monte Carlo

Os métodos de Monte Carlo, também conhecidos como simulações de Monte Carlo, são amplamente utilizados em simulações computacionais de processos que fazem uso de sinais aleatórios.

De maneira formal, as simulações de Monte Carlo são definidas para a resolução, de um ou ambos, dos seguintes problemas:

Problema (1): gerar amostras $\{x^{(r)}\}_{r=1}^R$ a partir de uma determinada função de densidade de probabilidade $p(\mathbf{x})$.

Problema (2): estimar expectativas de funções sob esta distribuição.

A grosso modo, pode-se entender o procedimento como necessário para fornecer uma resposta estatística, retirando desta maneira a influência de uma seqüência dada em uma determinada experiência.

No caso particular de SCD, a modelagem do ruído (componente do modelo do canal) como uma v.a. gaussiana, implica na geração do mesmo através de uma distribuição que assemelhe-se à dele (gaussiana).

Neste ponto, remete-se para a geração de v.a. em computadores. A construção de geradores pseudo-aleatórios torna a quantidade de amostras, fator determinante para uma representação de $p(\mathbf{x})$ através das amostras geradas. Na sua grande maioria, os geradores são construídos a partir de outros pré-existentes (uniforme na maioria dos casos) e com isto nem sempre esta estimativa é satisfatória.

Para o primeiro problema, as simulações de Monte Carlo fornecem ferramentas para a geração de v.a. cujas fdp não são facilmente encontradas através de outras.

O segundo problema é justamente o cenário para o qual se propõe a execução de simulações através dos métodos de Monte Carlo em SCD.

Numa determinada realização, a influência da seqüência de v.a. gerada *pode favorecer ou prejudicar* o desempenho de alguns dos componentes do SCD, em especial o equalizador. Assim, uma realização de uma série de execuções independentes tenta tirar a influência de uma (ou mais) execução favorável (prejudicial).

Formalizando matematicamente tem-se:

Deseja-se estimar a expectativa de funções, por exemplo:

$$\Phi = \mathbb{E}\{\phi(\mathbf{x})\} = \int p(\mathbf{x})\phi(\mathbf{x})d\mathbf{x}, \quad (2.13)$$

em que \mathbf{x} é uma variável N -dimensional e $p(\mathbf{x})$ é chamada de *distribuição alvo*, pois ela é conhecida e deseja-se que $\phi(\mathbf{x})$ seja igual à ela.

Sendo uma quantidade t que depende de \mathbf{x} , é de interesse encontrar os momentos de ordem 1 e 2: $\phi_1(\mathbf{x}) = t(\mathbf{x})$ e $\phi_2(\mathbf{x}) = (t(\mathbf{x}))^2$.

Então, usando a definição do estimador da Equação (2.13) tem-se:

$$\Phi_1 \equiv \mathbb{E}\{\phi_1(\mathbf{x})\} \quad \text{e} \quad \Phi_2 \equiv \mathbb{E}\{\phi_2(\mathbf{x})\}.$$

Constrói-se um estimador através das amostras como sendo:

$$\hat{\Phi} \equiv \frac{1}{R} \sum_r \phi(\mathbf{x}^{(r)}), \quad (2.14)$$

em que R é o número de realizações do processo.

Claramente, se os vetores $\mathbf{x}^{(r)}$ forem gerados a partir de $p(x)$ a estimação $\hat{\Phi}$ terá média igual à Φ . Senão, para um aumento de R , o valor da variância de $\hat{\Phi}$ decai de $\frac{\sigma^2}{R}$ em que σ^2 é a variância de $\phi(\mathbf{x})$.

Este fato também reflete uma das mais importantes propriedades dos métodos de Monte Carlo:

A precisão do estimador de Monte Carlo, Equação (2.14), é independente da dimensão do espaço de amostras. Ou seja, o decaimento da variância de $\hat{\Phi}$ com $\frac{\sigma^2}{R}$ não tem dependência da dimensão do espaço amostral.

Isso também ressalta a importância do número de realizações para uma melhor precisão através dos métodos de Monte Carlo. Um maior número de realizações (R) promove um decréscimo da variância de $\hat{\Phi}$.

A Figura 2.11 exemplifica um caso no qual a realização de várias experiências (execuções ou rodadas ou realizações) fornece, estatisticamente, uma resposta mais adequada. O sentido de adequação deve-se ao comportamento mais próximo da função original ($\exp(-x)$).

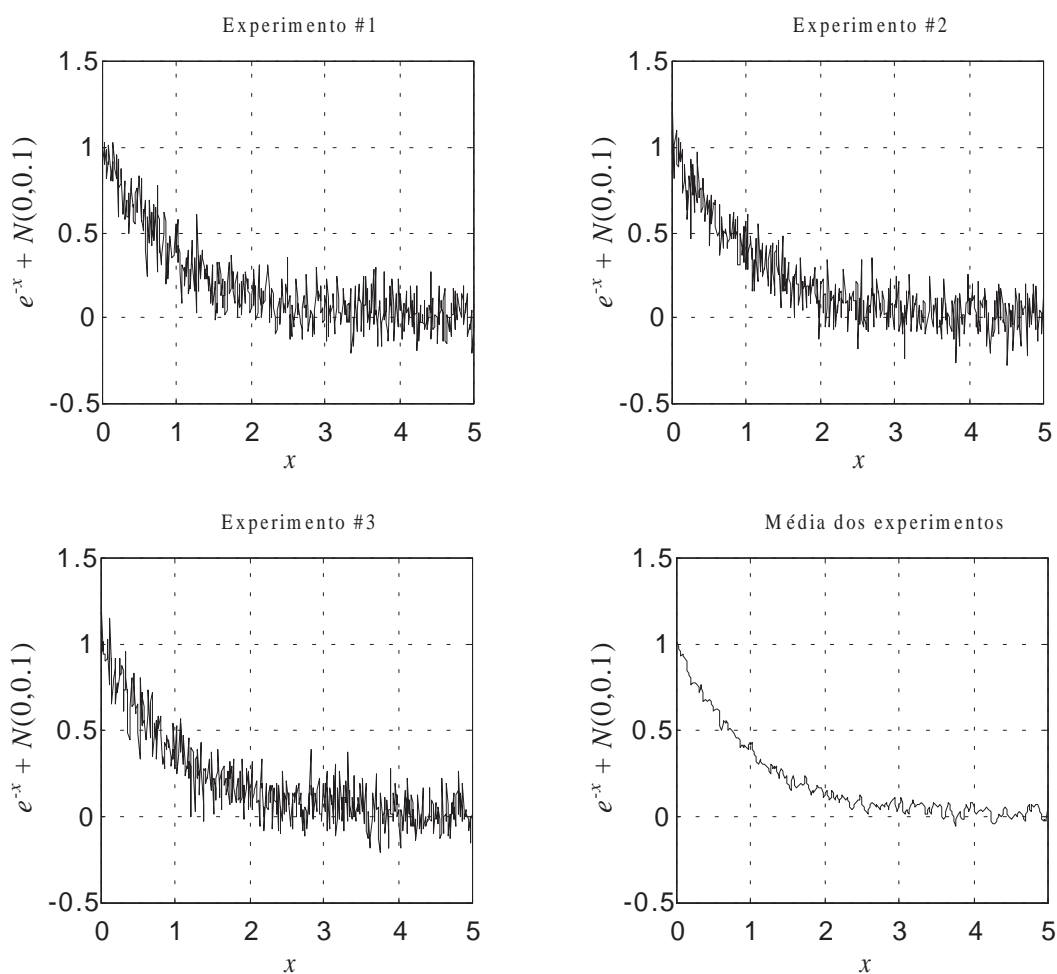


Figura 2.11: Simulações de Monte Carlo sobre um experimento.

2.6 Síntese

O modelo do sistema de comunicação ilustrado neste capítulo serve de base para o restante das considerações nesta dissertação.

O modelo em banda básica equivalente e as características da fonte de informação e do modelo do canal constituem a parte essencial deste capítulo, sendo sua compreensão um ponto de fundamental importância para as proposições descritas nos capítulos 6 e 7. Além disso, todas as simulações serão realizadas considerando o modelo em banda básica equivalente.

Uma outra abordagem que complementa a base teórica da análise a ser realizada nos capítulos posteriores é relacionada com os aspectos de teoria da informação que serão abordados no capítulo seguinte.

MODELOS BASEADOS EM TEORIA DA INFORMAÇÃO

O trabalho de Claude Shannon (1948) marca o início de um novo campo de estudos nos sistemas de comunicação, a *teoria da informação*¹. Sua formalização possibilitou contribuições não só no campo de comunicação, mas também em computação, física estatística, inferência estatística e em probabilidade e estatística [Haykin, 1994b].

No contexto das comunicações, a teoria da informação tem preocupação maior com a análise e o modelamento matemático de um sistema de comunicação em detrimento de fontes e canais físicos. A análise de um sistema de comunicações pode, inicialmente, responder a duas perguntas fundamentais:

- (1) qual a complexidade irreduzível abaixo da qual um sinal não pode ser comprimido?
- (2) qual a maior taxa de transmissão para um sistema de comunicação confiável utilizando um canal ruidoso?

Outro aspecto de extrema importância abordado por Shannon foi a definição matemática de *informação* que até então era apenas intuitiva. Tais conceitos e definições possibilitaram a análise teórica mais elaborada de um sistema de comunicação para determinação de seus limites de desempenho.

¹Curiosidade: embora o trabalho [Shannon, 1948] tenha determinado as bases do estudo da teoria da informação, não aparece citado neste trabalho a frase *teoria da informação*. A primeira citação ocorreu antes, no trabalho “A Mathematical Theory of Cryptography” do próprio Shannon em 1945 [Haykin, 1994b].

Este capítulo aborda alguns modelos da teoria da informação e descreve algumas de suas formulações utilizadas no tratamento da informação para a sua recuperação. Embora nem todos os conceitos apresentados neste capítulo sejam necessários para a compreensão da proposição desse trabalho, julga-se importante um esboço dos principais conceitos do campo da teoria da informação.

A organização deste capítulo segue a seguinte ordem. Na seção 3.1 é descrito o conceito de informação. Na seção 3.2 são apresentados os conceitos de entropia e na seção 3.3 os conceitos da divergência de Kullback-Leibler ou entropia diferencial. Na seção 3.4 são desenvolvidos os princípios de estimação (paramétrica e não paramétrica) de funções de densidade de probabilidade. A seção 3.5 sintetiza a abordagem do capítulo.

3.1 Informação e incerteza

O conceito matemático de informação desenvolvido por Shannon (1948) leva em consideração a quantidade de *incerteza* produzida pela fonte.

Como citado no capítulo 2, a fonte de informação de um sistema de comunicação pode ser modelada como um gerador, cuja saída é, a cada instante de observação, uma realização da variável aleatória discreta (a_s) pertencente a um determinado conjunto finito \mathcal{A} chamado alfabeto. Cada um dos elementos (símbolos) possui uma probabilidade de ocorrência e, antes da ocorrência de um evento descrito por uma variável aleatória, há uma quantidade de incerteza quanto ao valor que esta variável assumirá.

A incerteza caracteriza o ganho de informação que a ocorrência de um evento pode promover. A incerteza de uma informação pode então, ser traduzida em termos estatísticos através da probabilidade de ocorrência de seu evento.

Um evento cuja ocorrência é certa não traz nenhum acréscimo de informação, pois toda a informação já está contida na sua certeza de ocorrência. Desta maneira pode-se dizer que a determinação da quantidade de informação fornecida por uma ocorrência de símbolo é determinada pela quantidade de “surpresa” que essa ocorrência traz.

A definição matemática da quantidade de informação contida em um sinal aleatório, representada por $I(\cdot)$, e produzido por uma fonte que produz símbolos a_s pertencentes a um alfabeto $\mathcal{A} = \{a_s : 1 \leq s \leq S\}$, é escrita como:

$$I(a_s) = \log \left(\frac{1}{p_s} \right) = -\log(p_s), \quad (3.1)$$

em que $p_s = \Pr(s = a_s)$.

A base do logaritmo é arbitrária e, no caso de utilizar-se a base 2, tem-se como unidade de informação *bit*² e no caso do logaritmo natural tem-se então *nats* [Gibson, 1993; Haykin,

²Uma contração do inglês “binary digit”.

1998]. O uso da base 2 em comunicações digitais, por razões óbvias, é mais freqüente. Quando tem-se uma fonte que produz símbolos binários equiprováveis ($p_s = \frac{1}{2}$) obtém-se:

$$I(a_s) = -\log_2(p_s)$$

$$I(a_s) = -\log_2\left(\frac{1}{2}\right) = 1 \text{ bit.}$$

Então, *um bit é a quantidade de informação ganha quando um evento, de dois possíveis igualmente prováveis, ocorre.*

3.1.1 Propriedades da quantidade de informação

A definição de informação da Equação 3.1 apresenta as seguintes propriedades:

- (1) $I(a_s) = 0$ para $p_s = 1$,
ou seja, não há informação num evento de ocorrência garantida;
- (2) $I(a_s) \geq 0$ para $0 \leq p_s \leq 1$,
ou seja, a ocorrência de um evento a_s fornece alguma ou nenhuma informação, mas nunca resulta na perda de informação;
- (3) $I(a_s) > I(a_i)$ para $p_s < p_i$,
ou seja, quão menor a probabilidade de um evento, mais informação sua ocorrência fornece;
- (4) $I(a_s a_i) = I(a_s) + I(a_i)$ se a_s e a_i são independentes,
ou seja, as informações de eventos de ocorrência conjunta somam-se.

A Figura 3.1 ilustra a quantidade de informação produzida pela ocorrência de um evento qualquer.

Observa-se que a quantidade de informação $I(\cdot)$ diz respeito a *cada* símbolo do alfabeto da fonte, entretanto não apresenta uma medida para caracterizar a quantidade de informação da própria fonte. A grandeza que permite esta caracterização é estudada na seção seguinte.

3.2 Entropia

Uma maneira de avaliar a quantidade de informação de uma fonte consiste em obter uma *média* sobre as quantidades de informação dos eventos gerados pela fonte. Portanto, sendo a quantidade de informação uma grandeza discreta com a probabilidade p_k , pode-se representar a esperança estatística da quantidade de informação de uma fonte por:

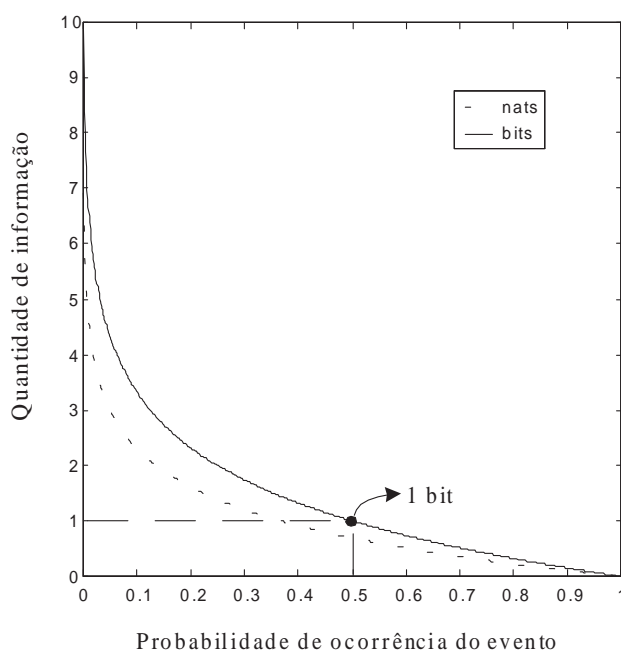


Figura 3.1: Quantidade de informação na ocorrência de um evento.

$$\begin{aligned}
 \mathcal{H}(\mathcal{A}) &= \mathbb{E} \{I(\mathcal{A})\} \\
 &= \sum_{k=1}^S p_k I(a_k) \\
 &= - \sum_{k=1}^S p_k \log(p_k).
 \end{aligned} \tag{3.2}$$

A grandeza $\mathcal{H}(\mathcal{A})$ é chamada de *entropia da fonte*³. A entropia é uma medida da quantidade média de informação contida em uma mensagem.

³O termo entropia em teoria da informação surgiu de uma analogia com o homônimo utilizado na termodinâmica. Neste caso, entropia é definida como:

$$\mathcal{H} = -k_B \sum_{\alpha} p_{\alpha} \log p_{\alpha},$$

em que k_B é a constante de Boltzmann e p_{α} é a probabilidade do sistema estar no estado α . Exceto pela constante k_B as duas expressões matemáticas de \mathcal{H} (teoria da informação e termodinâmica) têm a mesma forma [Haykin, 1998; Bishop, 1995].

Informação adicional:

Existem ainda, definições menos conhecidas para entropia que a fornecida por [Shannon, 1948], como por exemplo a definição para variáveis contínuas de Renyi [Renyi, 1987]:

$$\mathcal{H}_{R\alpha}(\mathcal{A}) = \frac{1}{1-\alpha} \log \int p_{\mathcal{A}}^{\alpha}(\xi) d\xi \quad \alpha \geq 0, \alpha \neq 1,$$

Entretanto, apenas aquela fornecida pela Equação (3.2) será abordada neste trabalho.

De acordo com a Equação 3.2, a entropia apresenta os seguintes limites:

$$0 \leq \mathcal{H}(\mathcal{A}) \leq \log(S), \quad (3.3)$$

em que (S) é o número de níveis discretos, ou seja, o número de símbolos no alfabeto \mathcal{A} .

Os limites da entropia de uma fonte são determinados então por:

- (1) $\mathcal{H}(\mathcal{A}) = 0$ se e somente se a probabilidade de ocorrência p_s de um certo evento a_s for $p_s = 1$ e todas as demais forem iguais à zero. Neste ponto não existe nenhuma *incerteza* e conseqüentemente a entropia é mínima.
- (2) $\mathcal{H}(\mathcal{A}) = \log(S)$ se e somente se as probabilidades de todos os eventos a_s forem iguais, ou seja, os eventos forem equiprováveis ($p_k = \frac{1}{S}$).

A Figura 3.2 esboça o comportamento da entropia, de uma fonte binária, em função da probabilidade p_1 de ocorrência de um dos símbolos. Neste caso, a entropia será dada por:

$$\begin{aligned} \mathcal{H}(\mathcal{A}) &= - \sum_{s=1}^2 p_s \cdot \log_2(p_s) \\ &= -p_1 \log_2(p_1) - (1-p_1) \log_2(1-p_1), \end{aligned}$$

Neste exemplo, pode-se verificar as duas propriedades da entropia citadas anteriormente. Uma outra verificação deve-se ao fato de que quão maior a incerteza de ocorrência dos eventos (símbolos) de uma fonte, maior a entropia.

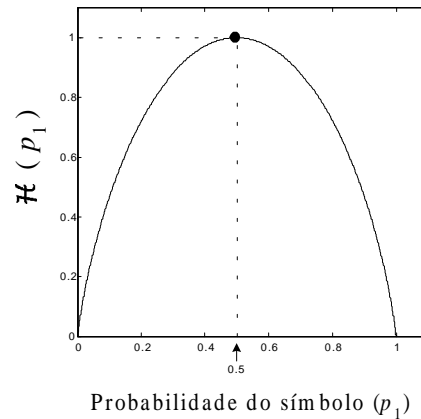


Figura 3.2: Entropia de uma fonte binária em função da probabilidade de ocorrência p_1 .

Uma concepção um pouco mais intuitiva de entropia pode ser enunciada como:

*A entropia de uma fonte é uma medida da **desorganização** com que ela emite seus símbolos. Quão menos certo for a ocorrência de um símbolo, maior a sua desorganização e conseqüentemente maior a entropia.*

3.2.1 Taxa de informação

A taxa de informação de uma fonte é determinada a partir de suas entropia e taxa de transmissão.

Sendo uma fonte de informação \mathcal{S} que transmite r símbolos a cada segundo, cujos símbolos são variáveis aleatórias a_s pertencentes a um alfabeto $\mathcal{A} = \{a_s : 1 \leq s \leq S\}$ com entropia $\mathcal{H}(\mathcal{A})$, define-se a taxa de informação R como sendo determinada pela equação:

$$R = r \cdot \mathcal{H}(\mathcal{A}). \quad (3.4)$$

3.2.2 Informação mútua

Uma importante medida de informação é aquela que calcula a quantidade de informação fornecida na saída de um sistema, a partir de uma dada quantidade de informação na entrada. No caso de considerar-se uma medida *média*, isto corresponde a calcular a entropia da saída $\mathcal{H}(\mathcal{X})$ dada uma determinada entropia de entrada $\mathcal{H}(\mathcal{A})$. Esta medida denomina-se *informação mútua* e tem aplicação no projeto de sistemas auto-organizáveis, onde deseja-se que o sistema produza uma relação entrada-saída baseado somente nas entradas [Haykin, 1998]. A Figura 3.3 ilustra o caso para a transmissão de informação através de um canal de comunicações, nesse caso deseja-se saber quanto de informação da fonte carrega o sinal recebido.

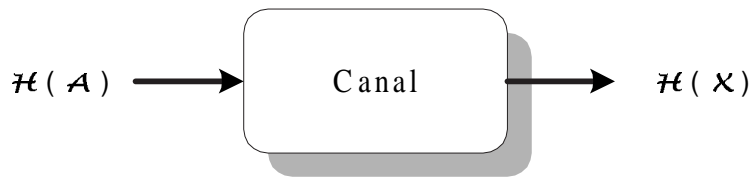


Figura 3.3: Transmissão de informação através de um canal de comunicação.

Utilizando-se então a definição de quantidade de informação na Equação (3.1) e substituindo-se a probabilidade de ocorrência de um evento isolado p_s pela probabilidade de ocorrência conjunta, $P_{\mathcal{X};\mathcal{A}}(x, a)$, em que $P_{\mathcal{A}}(a_j)$ é a probabilidade de ocorrência de a_j no alfabeto \mathcal{A} , define-se a informação mútua como [Haykin, 1998]:

$$I_{\mathcal{X};\mathcal{A}}(x_k; a_j) = P_{\mathcal{X};\mathcal{A}}(x_k; a_j) \log \left(\frac{P_{\mathcal{X};\mathcal{A}}(x_k, a_j)}{P_{\mathcal{X}}(x_k) \cdot P_{\mathcal{A}}(a_j)} \right). \quad (3.5)$$

Fazendo-se uso das propriedades de probabilidade conjunta⁴ de modo a obter-se as probabilidades condicionais em que $P_{\mathcal{X}|\mathcal{A}}(x_k|a_j)$ é a probabilidade condicional de obter-se o evento x_k dada a ocorrência de a_j , pode-se encontrar uma expressão da informação mútua em função das entropias das variáveis consideradas.

Ao tomar-se o valor médio da informação mútua sobre os conjuntos $(\mathbb{E}_{\mathcal{X};\mathcal{A}}\{I(\mathcal{X};\mathcal{A})\})$, tem-se então:

$$\begin{aligned} I(\mathcal{X}; \mathcal{A}) &= \sum_{k=1}^X \sum_{j=1}^S P_{\mathcal{X}|\mathcal{A}}(x_k|a_j) \cdot P_{\mathcal{A}}(a_j) \cdot \log \left(\frac{P_{\mathcal{X}|\mathcal{A}}(x_k|a_j)}{P_{\mathcal{X}}(x_k)} \right) \\ I(\mathcal{X}; \mathcal{A}) &= \sum_{k=1}^X \sum_{j=1}^S P_{\mathcal{X}|\mathcal{A}}(x_k|a_j) \cdot P_{\mathcal{A}}(a_j) (\log (P_{\mathcal{X}|\mathcal{A}}(x_k|a_j)) - \log (P_{\mathcal{X}}(x_k))) \\ I(\mathcal{X}; \mathcal{A}) &= \sum_{k=1}^X \sum_{j=1}^S P_{\mathcal{A}}(a_j) \cdot P_{\mathcal{X}|\mathcal{A}}(x_k|a_j) \cdot \log (P_{\mathcal{X}|\mathcal{A}}(x_k|a_j)) \\ &\quad - P_{\mathcal{A}}(a_j) \cdot P_{\mathcal{X}|\mathcal{A}}(x_k|a_j) \cdot \log (P_{\mathcal{X}}(x_k)). \end{aligned}$$

Então, pode-se facilmente observar que:

$$I(\mathcal{X}; \mathcal{A}) = \mathcal{H}(\mathcal{X}) - \mathcal{H}(\mathcal{X}|\mathcal{A}), \quad (3.6)$$

em que $\mathcal{H}(\mathcal{X}|\mathcal{A})$ é chamada de *entropia condicional*. Vê-se facilmente da Equação (3.6) que a entropia é um caso especial da informação mútua média, $\mathcal{H}(\mathcal{X}) = I(\mathcal{X}; \mathcal{X})$ [Haykin, 1998]. A Figura 3.4 mostra um diagrama associado à interpretação da Equação (3.6).

3.2.2.1 Propriedades da informação mútua

A informação mútua possui as seguintes propriedades:

- (1) a informação mútua média entre \mathcal{A} e \mathcal{X} é simétrica: $I(\mathcal{A}; \mathcal{X}) = I(\mathcal{X}; \mathcal{A})$.
- (2) a informação mútua média entre \mathcal{A} e \mathcal{X} é sempre não negativa: $I(\mathcal{A}; \mathcal{X}) \geq 0$.
- (3) a informação mútua média entre \mathcal{A} e \mathcal{X} pode ser expressa como uma função da entropia de \mathcal{X} : $I(\mathcal{A}; \mathcal{X}) = \mathcal{H}(\mathcal{X}) - \mathcal{H}(\mathcal{X}|\mathcal{A})$.
- (4) a informação mútua de \mathcal{A} e \mathcal{X} pode ser expressa como uma relação entre as entropias de \mathcal{A} e \mathcal{X} e sua entropia conjunta: $I(\mathcal{A}; \mathcal{X}) = H(\mathcal{A}) + H(\mathcal{X}) - H(\mathcal{A}|\mathcal{X})$, em que $H(X; A) = - \sum_k \sum_j P(x_k, a_j) \log (P(x_k|a_j))$.

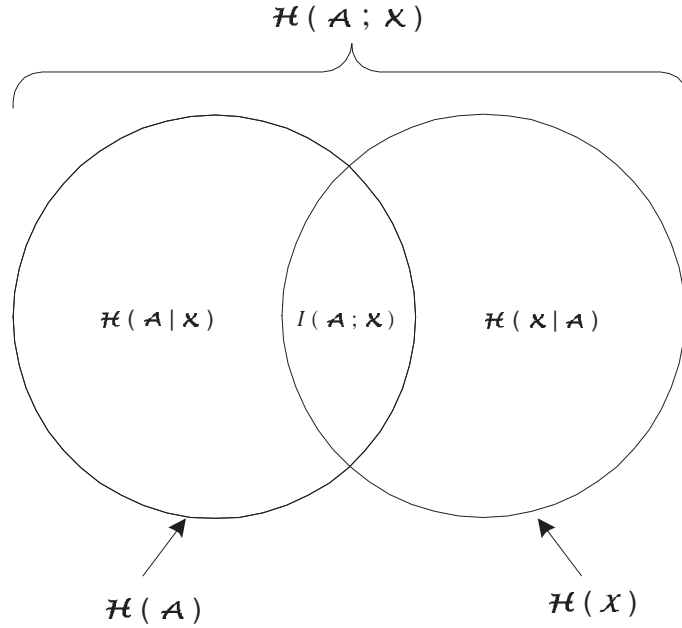


Figura 3.4: Relação entre informação mútua $I(\mathcal{A}; \mathcal{X})$ e entropias $\mathcal{H}(\mathcal{A})$ e $\mathcal{H}(\mathcal{X})$.

Diagramaticamente, a assimilação da dependência entre as entropias e informação mútua fica mais fácil. Note que a informação mútua é comum tanto à entropia da saída como da entrada de um sistema (como já descrito nas propriedades de informação mútua). Este conceito será particularmente importante na determinação (minimização ou maximização) da entropia mútua em um SCD.

3.2.3 Capacidade de Canal

Considere um canal em que é transmitido um símbolo a_j e o símbolo x_k é recebido, a informação mútua fornecida pela saída \mathcal{X} sobre a entrada \mathcal{A} é:

$$I(\mathcal{X}; \mathcal{A}) = \sum_{j=1}^S \sum_{k=1}^X P(a_j) P(x_k|a_j) \log \left(\frac{P(x_k|a_j)}{P(x_k)} \right).$$

As características do canal determinam então as probabilidades $P(x_k|a_j)$, mas as probabilidades dos símbolos de entrada são determinadas pelo codificador de canal. Logo, pode-se maximizar a informação transportada através de um canal pela maximização da informação mútua sobre a probabilidade dos símbolos de entrada. Daí define-se a capacidade $C(\cdot)$ de um canal de comunicação \mathcal{F} como sendo:

⁴ $P_{\mathcal{X}; \mathcal{A}}(x_k, a_j) = P_{\mathcal{X}|\mathcal{A}}(x_k|a_j) \cdot P_{\mathcal{A}}(a_j).$

$$C(\mathcal{F}) = \max_{P_{\mathcal{A}}} I(\mathcal{X}; \mathcal{A}). \quad (3.7)$$

A distribuição $P_{\mathcal{A}}$ que maximiza a Equação (3.7) é chamada de *distribuição de entrada ótima*, podendo haver mais de uma distribuição ótima para o mesmo valor $I(\mathcal{X}; \mathcal{A})$.

3.3 Divergência de Kullback-Leibler

Um problema freqüentemente encontrado em inferência estatística é a definição da medida de similaridade entre funções. A divergência de Kullback-Leibler (KLD⁵) é uma medida de similaridade entre duas funções estritamente positivas. É também chamada de *entropia diferencial* ou ainda *entropia cruzada* [Deco & Obradovic, 1996] .

Matematicamente, a definição da KLD pode ser explicitada como:

$$D_{p(x)||g(x)} = \int_{-\infty}^{\infty} p(x) \cdot \ln \left(\frac{p(x)}{g(x)} \right) dx, \quad (3.8)$$

em que $p(x)$ e $g(x)$ são duas funções estritamente positivas. Para o caso discreto tem-se:

$$D_{p(x)||g(x)} = \sum_{x \in \mathbb{N}} p(x) \cdot \ln \left(\frac{p(x)}{g(x)} \right), \quad (3.9)$$

em que \mathbb{N} é o conjunto domínio da variável x .

É bastante usual a utilização da KLD para a comparação entre duas *funções densidade de probabilidade* (fdp). Neste caso, a fdp $g(x)$ realiza o papel de uma *função de referência*.

Deve-se observar que $D_{p(x)||g(x)} \neq D_{g(x)||p(x)}$, significando que $D_{\bullet||\circ}$ não é uma medida de distância⁶, embora seja comum na literatura ser referenciada como distância de Kullback-Leibler ou ainda divergência assimétrica .

KLD tem algumas importantes propriedades [Haykin, 1998]:

1. é sempre de valor positivo ou zero; KLD é zero para o caso específico de $p(x) = g(x)$.
2. é invariante com relação as seguintes mudanças nas componentes do vetor \mathbf{x} ;

⁵Do inglês “Kullback-Leibler Divergence”.

⁶Não é uma medida de distância no espaço euclidiano, mas pode ser apresentada no contexto de geometria diferencial como uma métrica de Riemman no espaço das distribuições [Deco & Obradovic, 1996].

- permutação de ordem
- escalonamento de amplitude
- transformação monotônica não linear

Utilizando-se duas funções estritamente positivas, preserva-se ainda a propriedade da KLD possuir apenas valores positivos ou zero.

Esta abordagem é de extremo interesse para a construção de funções de otimização, pois proporciona a elaboração de uma classe de funcionais que têm apelo em teoria da informação, por utilizar o princípio de mínima entropia através da comparação entre funções.

3.4 Estimação de Função de Densidade de Probabilidade

O modelamento de funções de densidade de probabilidade consiste em sua estimação a partir de um conjunto de dados gerados da mesma fdp. Pode-se, a partir deste modelamento, construir sistemas de classificação por separação de classes através da estimação da fdp e do teorema de Bayes. Tais sistemas podem ser então empregados para a equalização de dados em sistemas de comunicação digitais, como será visto a seguir no capítulo 5.

As abordagens para a estimação de fdp podem ser de dois tipos: *paramétrica* e *não paramétrica*. A escolha para utilização de cada um dos métodos depende não somente da aplicação mas também do compromisso entre a disponibilidade de dados e performance. As duas abordagens são discutidas a seguir.

3.4.1 Estimação paramétrica

Uma das técnicas mais difundidas em estimação de fdp é a *estimação paramétrica*, na qual utiliza-se um funcional que possui um certo número de parâmetros ajustáveis a partir dos dados. Assim, os parâmetros são ajustados de maneira a melhor adequar os dados ao funcional idealizado. Este método é bastante apropriado para os casos em que se conhece *a priori* a forma da distribuição dos dados e deseja-se então estimar apenas os parâmetros.

A distribuição *normal* ou *gaussiana* é a mais utilizada para estimação paramétrica devido a sua simplicidade e ao grande número de convenientes propriedades analíticas e estatísticas. A fdp normal é escrita, para o caso unidimensional, da seguinte forma:

$$p(w) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(w - \mu)^2}{2\sigma^2}\right), \quad (3.10)$$

em que μ e σ são, respectivamente, média e desvio padrão. O fator multiplicativo $\frac{1}{\sqrt{2\pi}\sigma^2}$ assegura que $\int_{-\infty}^{\infty} p(w)dw = 1$.

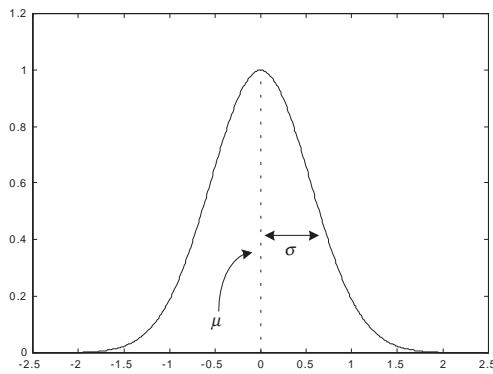
Para o caso multidimensional (dimensão d), temos:

$$p(\mathbf{w}) = \frac{1}{\sqrt{(2\pi)^d} [\det(\Sigma)]^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{w} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{w} - \boldsymbol{\mu})\right), \quad (3.11)$$

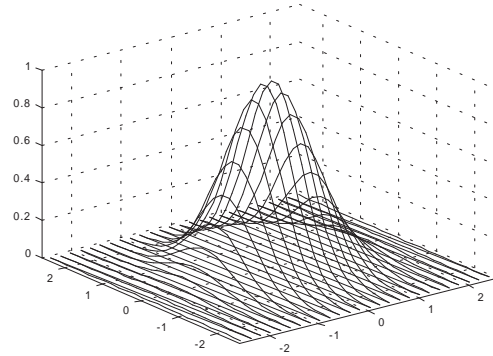
em que $\Sigma = E\{(\mathbf{w} - \boldsymbol{\mu})(\mathbf{w} - \boldsymbol{\mu})^T\}$ é a matriz de covariância (de dimensão $d \times d$), $\det(\cdot)$ é o determinante da matriz e $\boldsymbol{\mu}$ é o vetor médio da distribuição.

A expressão $\Delta^2 = (\mathbf{w} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{w} - \boldsymbol{\mu})$ que aparece na exponencial da Equação (3.11) é chamada de distância de Mahalanobis de \mathbf{w} para $\boldsymbol{\mu}$ [Bishop, 1995; Duda & Hart, 1973; Montalvão et al., 1998].

A Figura 3.5 ilustra a fdp gaussiana para os casos uni e multidimensional ($d = 2$).



(a): fdp gaussiana unidimensional.



(b): fdp gaussiana multidimensional ($d = 2$).

Figura 3.5: Funções densidade de probabilidade gaussianas.

Algumas das propriedades da distribuição gaussiana que tornam bastante interessante sua utilização são [Bishop, 1995]:

1. propriedades analíticas simples possibilitam a obtenção de vários resultados explicitamente. Por exemplo, qualquer momento da distribuição pode ser expresso como função de $\boldsymbol{\mu}$ e Σ .
2. o teorema central do limite permite a aproximação da fdp da soma de um certo número de v.a. como uma distribuição gaussiana.
3. sob qualquer transformação não-singular do sistema de coordenadas, a distância de Mahalanobis mantém sua forma quadrática e permanece definida positiva. Além

disso, após tal transformação a distribuição é ainda gaussiana, de média e covariância diferentes.

4. são normais as densidades marginais e condicionais de uma densidade normal.
5. para dados valores da média e matriz de covariância, a função de densidade normal maximiza a entropia.

Embora tais propriedades sejam favoráveis, a principal razão para a escolha de tal distribuição é a sua simplicidade analítica.

Ainda que o modelo paramétrico, particularmente para o caso de densidades gaussianas, apresente várias vantagens, também possui uma desvantagem que é devido à escolha do número de parâmetros. O número de parâmetros escolhido pode não ser suficiente para fornecer uma boa representação da fdp dos dados e, neste caso, o modelo assumido para a fdp forneceria uma representação incompleta.

A Figura 3.6 ilustra esses casos, para a estimação de uma fdp que é uma composição de gaussianas.

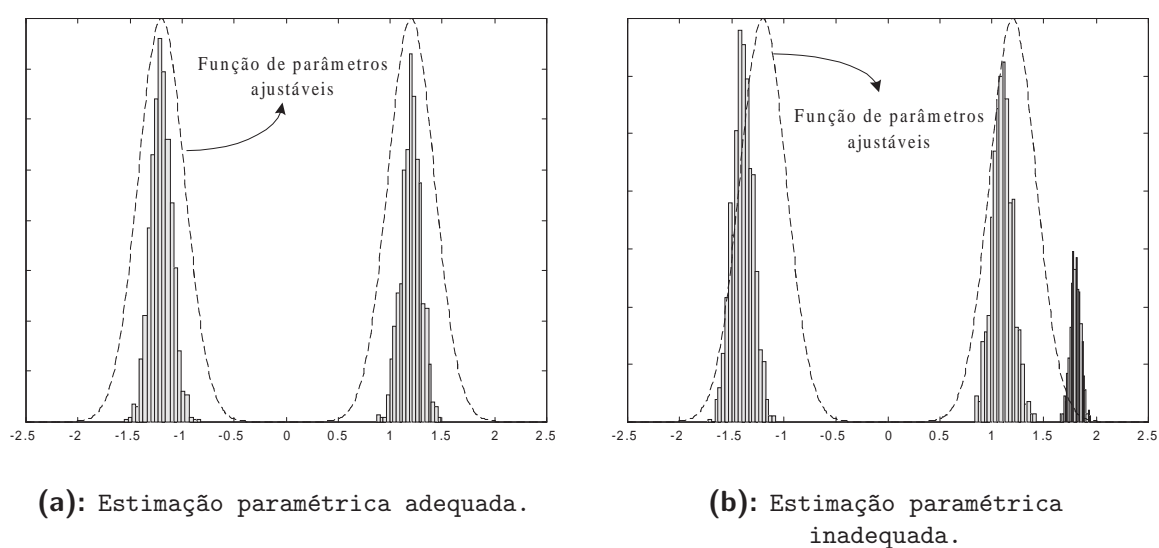


Figura 3.6: Diferenças entre a utilização de um modelo adequado e inadequado na estimação paramétrica.

Observe na Figura 3.6(b) que o estimador não foi capaz de discriminar as três gaussianas, diferentemente da Figura 3.6(a).

Os métodos paramétricos de estimação de fdp são basicamente fundamentados sobre duas técnicas, máxima verossimilhança e inferência Bayesiana, descritas brevemente a seguir.

3.4.1.1 Máxima verossimilhança

O critério de máxima verossimilhança (ML⁷) consiste na maximização de uma função de verossimilhança derivada dos dados. Desta maneira, dada a escolha do modelo, os parâmetros são ajustados para melhor representar os dados cuja função de verossimilhança foi construída.

Considerando-se um vetor de parâmetros $\boldsymbol{\eta}$ para serem ajustados e definindo-se a função de densidade:

$$p_{w;\boldsymbol{\eta}}(\mathbf{w}; \boldsymbol{\eta}), \quad (3.12)$$

para um dado conjunto de dados emitidos $\mathbf{w}^{(0)}$, a estimativa de ML dos parâmetros, denotado por $\hat{\boldsymbol{\eta}}_{ML}$ é definido como o valor de $\boldsymbol{\eta}$ que maximiza $p_{w;\boldsymbol{\eta}}(\mathbf{w}^{(0)}; \boldsymbol{\eta})$. Formalizando esta definição:

$$\hat{\boldsymbol{\eta}}_{ML}(\mathbf{w}) = \arg \max_{\boldsymbol{\eta}} p_{w;\boldsymbol{\eta}}(\mathbf{w}; \boldsymbol{\eta}). \quad (3.13)$$

Tal notação mostra que $\hat{\boldsymbol{\eta}}_{ML}$ depende do vetor de observações (dados) $\mathbf{w}^{(0)}$ sendo então ele mesmo uma v.a..

Para o caso particular da função gaussiana, onde os parâmetros a serem ajustados são média e covariância, o critério ML fornece as seguintes expressões numéricas para cálculo iterativo da estimativa dos parâmetros [Therrien, 1992; Bishop, 1995]

$$\hat{\boldsymbol{\mu}} = \frac{1}{N} \sum_{n=1}^N \mathbf{w}^{(n)} \quad (3.14a)$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N} \sum_{n=1}^N (\mathbf{w}^{(n)} - \hat{\boldsymbol{\mu}})(\mathbf{w}^{(n)} - \hat{\boldsymbol{\mu}})^T, \quad (3.14b)$$

em que N é o número de amostras. Note-se que a estimativa ML corresponde à *média dos dados*.

Embora o critério ML apresente-se como razoável, ele possui certas limitações. Considerando-se o caso unidimensional (por motivos de simplicidade de demonstração), a esperança matemática da estimativa da variância é $\mathbb{E}\{\hat{\sigma}^2\} = \frac{N-1}{N}\sigma^2$. É obvio que para $N \rightarrow \infty$ a esperança é igual para ambas ($\hat{\sigma}^2 = \sigma^2$), mas para valores finitos a igualdade não acontece. Isto deve-se ao fato de que a estimativa de σ^2 depende da estimativa $\hat{\mu}$ para a média.

3.4.1.2 Inferência bayesiana

Como os dados recebidos são descritos por uma função densidade *anterior*, a observação dos dados recebidos permite, através do uso do teorema de Bayes [Papoulis, 1991], encontrar a função densidade *posterior*.

⁷Do inglês “Maximum Likelihood”.

Nesta técnica o vetor de parâmetros ajustáveis deve ser adaptado tendo apenas o conhecimento da distribuição *posterior* dos dados. Pode-se então escrever a função densidade de \mathbf{w} dado um vetor de dados recebidos \mathcal{W} como:

$$p(\mathbf{w}|\mathcal{W}) = \int p(\mathbf{w}, \boldsymbol{\eta}|\mathcal{W}) d\boldsymbol{\eta}, \quad (3.15)$$

em que a distribuição conjunta é escrita $p(\mathbf{w}, \boldsymbol{\eta}|\mathcal{W}) = p(\mathbf{w}|\boldsymbol{\eta}, \mathcal{W})p(\boldsymbol{\eta}|\mathcal{W})$.

Uma vez que $p(\mathbf{w}|\boldsymbol{\eta}, \mathcal{W})$ é a função densidade assumida, ela não depende dos dados observados, porque uma vez escolhidos os valores de $\boldsymbol{\eta}$ a fdp não depende dos valores de \mathcal{W} , então pode-se reescrever a Equação (3.15)

$$p(\mathbf{w}|\mathcal{W}) = \int p(\mathbf{w}|\boldsymbol{\eta})p(\boldsymbol{\eta}|\mathcal{W})d\boldsymbol{\eta}. \quad (3.16)$$

Aqui vê-se a diferença entre as técnicas de máxima verossimilhança e inferência bayesiana. Ao invés de escolher um valor para $\boldsymbol{\eta}$, é realizada uma ponderação sobre todos os valores de $\boldsymbol{\eta}$. O fator de ponderação $p(\boldsymbol{\eta}|\mathcal{W})$, que é a distribuição *posterior* de $\boldsymbol{\eta}$, é determinada através da inicialização de alguma distribuição *anterior* $p(\boldsymbol{\eta})$ e adaptada através do teorema de Bayes, considerando o conjunto de dados \mathcal{W} .

Considerando-se que os dados sejam independentes, pode-se escrever a função de distribuição de $\boldsymbol{\eta}$ *posterior* na seguinte forma

$$p(\boldsymbol{\eta}|\mathcal{W}) = \frac{p(\boldsymbol{\eta})}{p(\mathcal{W})} \underbrace{\prod_{n=1}^N p(\mathbf{w}^{(n)}|\boldsymbol{\eta})}_{p(\mathcal{W}|\boldsymbol{\eta})}. \quad (3.17)$$

Para exemplificar, considera-se a estimação dos parâmetros média (μ) e variância (σ^2) de uma função densidade gaussiana, a partir de uma seqüência de dados dessa distribuição.

Considerando a função densidade *anterior* da v.a. μ como sendo uma gaussiana da forma

$$p_0(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(\mu - \mu_0)^2}{2\sigma_0^2}\right). \quad (3.18)$$

Esta expressão expressa o conhecimento *anterior* da média μ , e através dos dados, pode-se calcular a função densidade *posterior* $p(\mu|\mathcal{W})$ usando o teorema de Bayes.

Usando também hipótese da independência a estimativa da média e variância é feita através da seguinte equação [Bishop, 1995]

$$\mu_N = \frac{N\sigma_0^2}{N\sigma_0^2 + \sigma^2} \bar{x} + \frac{\sigma^2}{N\sigma_0^2 + \sigma^2} \mu_0 \quad (3.19a)$$

$$\frac{1}{\sigma_N^2} = \frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}, \quad (3.19b)$$

em que N é o número de amostras (dados) e $\bar{x} = \frac{1}{N} \sum_{n=1}^N w^{(n)}$.

3.4.2 Estimação não paramétrica

Contrastando com a estimação paramétrica, a estimação não paramétrica não faz uso de um funcional pré-estabelecido para fazer a estimação da fdp, pois neste caso a fdp é estimada apenas a partir dos dados recebidos.

Claramente essa estimação é mais delicada, já que sua precisão depende da confiabilidade dos dados recebidos, constituindo uma estatística suficiente para representar a sua verdadeira fdp. Do ponto de vista geométrico, a estimação não paramétrica deve encontrar valores, cuja distribuição dos dados seja uma estimativa suavizada da distribuição em histograma, por exemplo.

A própria representação em histograma apresenta parâmetros, como a largura dos intervalos da distribuição, que podem modificar a distribuição. Assim, a estimação não paramétrica deve buscar a otimização de tais parâmetros para que a estimativa seja o mais próximo possível da original. Na busca por essa estimação, remete-se ao conceito de fdp.

A probabilidade de que um vetor de dados \mathbf{w} gerado a partir de uma fdp $p(\mathbf{w})$, esteja dentro de uma região \mathcal{R} do espaço \mathbf{w} -dimensional é, por definição

$$P = \int_{\mathcal{R}} p(\mathbf{w}') d\mathbf{w}'. \quad (3.20)$$

Se há N dados gerados independentes com uma fdp $p(\mathbf{w})$, a probabilidade de que K deles estejam dentro da região \mathcal{R} é dada pela lei binomial

$$\Pr(K) = \frac{N!}{K!(N-K)!} P^K (1-P)^{N-K}. \quad (3.21)$$

Assim, a fração média de pontos dessas regiões é dada por $\mathbb{E}\{\frac{K}{N}\} = P$ e a variância em torno dessa média é dada por $\mathbb{E}\{(\frac{K}{N} - P)^2\} = \frac{P(1-P)}{N}$ [Bishop, 1995]. Com isto, a distribuição apresenta um pico quando $N \rightarrow \infty$. Entretanto, espera-se que uma boa estimativa da probabilidade P possa ser obtida da fração média de pontos na região \mathcal{R} tal que

$$P \simeq \frac{K}{N}. \quad (3.22)$$

Se for assumido que $p(\mathbf{w})$ é contínua e tem pequena variação na região \mathcal{R} então é possível aproximar a Equação (3.20) por

$$P = \int_{\mathcal{R}} p(\mathbf{w}') d\mathbf{w}' \simeq p(\mathbf{w})V, \quad (3.23)$$

em que V é o volume de \mathcal{R} e \mathbf{w} é algum ponto interno a \mathcal{R} . Então a partir das Equações (3.22) e (3.23), obtém-se o resultado

$$p(\mathbf{w}) \simeq \frac{K}{NV}, \quad (3.24)$$

Deve-se notar que a obtenção deste resultado é feita através de duas hipóteses sobre a região \mathcal{R} . A primeira, para a Equação (3.22) ser precisa é requisito que \mathcal{R} seja relativamente grande. Entretanto, a Equação (3.23) é mais precisa quando a região \mathcal{R} é relativamente pequena (para não apresentar variações consideráveis na região). Vê-se então que existe um parâmetro ótimo de \mathcal{R} para fornecer a melhor estimativa de $p(\mathbf{w})$.

Se for aplicada a Equação (3.24) pode-se escolher duas abordagens cujos desenvolvimentos não serão efetuados. A primeira consiste em fixar o valor de K e determinar o volume V correspondente. Essa técnica é conhecida como “ K -nearest neighbors”. A outra alternativa é fixar o volume V e determinar K a partir dos dados. A essa técnica dá-se o nome de “kernel-based” [Duda & Hart, 1973].

Todas estas técnicas procuram, obviamente, fazer uma estimativa da fdp. Vale ressaltar que para mensurar a diferença entre a fdp original e aquela estimada, a KLD é uma das métricas mais usadas pelas razões já explicadas na seção 3.3.

3.5 Síntese

As técnicas e ferramentas decorrentes da teoria da informação descritas nesse capítulo reúnem um material básico para qualquer análise mais detalhada sobre o processo de comunicação. Neste trabalho, elas servirão para fundamentar algumas das análises realizadas bem como situar o leitor no campo (muito extenso) da teoria da informação.

Algumas das técnicas encontram-se expressas apenas de maneira descritiva não fornecendo assim uma análise mais detalhada dos resultados a serem utilizados. Para tal fim, sugere-se a consulta às referências [Duda & Hart, 1973; Bishop, 1995; Haykin, 1998; Deco & Obradovic, 1996] para maiores demonstrações e explicações sobre o assunto.

De maneira mais pontual, os conceitos que serão efetivamente utilizados a partir deste capítulo são: a noção de entropia e a divergência de Kullback-Leibler. Um ponto chave também será a estimação da fdp que será novamente abordada no capítulo 7. Aspectos mais comuns de análise em teoria da informação, como capacidade de canal e taxa de informação, serão restritos à breve descrição feita no texto.

Vale também ressaltar que não foram discutidas, nem mesmo descritas, as técnicas de estimação semiparamétricas de fdp. Essas técnicas buscam resolver os problemas inerentes às estimações paramétricas e não paramétricas de uso de um funcional bem diferente da verdadeira fdp (paramétrica) e crescimento da complexidade do modelo da verdadeira fdp com o aumento do número de dados (não paramétrica). Uma excelente fonte bibliográfica para a consulta destas técnicas é [Bishop, 1995].

CAPÍTULO 4

REDES NEURAIS ARTIFICIAIS

A teoria de redes neurais artificiais (RNA) tem-se desenvolvido bastante a partir da década de 50, suportada pela melhor compreensão do funcionamento do neurônio biológico e com a proposição de um modelo matemático para o mesmo. O desejo de tentar mimetizar o comportamento e processamento do cérebro humano, para a realização automática de tarefas cujo aprendizado das mesmas fosse satisfatoriamente inteligente, deixava de ser apenas uma idealização.

O grande avanço na área de modelagem das RNA, determinou uma grande evolução na utilização das mesmas. Várias áreas de pesquisa, tais como classificação e reconhecimento de padrões, processamento de imagens e separação de fontes têm experimentado grandes avanços devido ao uso das técnicas de RNA.

Paralelamente ao modelamento das RNA, o aprendizado das mesmas também ganhou grande projeção no cenário de pesquisas em inteligência artificial. A execução do aprendizado, tornou-se aspecto de grande importância devido à necessidade de se promover melhorias sensíveis nos algoritmos de treinamento para que as RNA fossem capazes de realizar as mais diversas tarefas.

Este capítulo busca apresentar os conceitos das RNA e seus modelos mais importantes, bem como uma discussão sobre as principais técnicas de aprendizado. Este desenvolvimento teórico é fundamental para a compreensão da proposição realizada no capítulo 6.

A organização deste capítulo tem a seguinte forma: na seção 4.1 são mostrados os conceitos sobre o modelo do neurônio biológico e suas características; a seção 4.2 discute um pouco os principais modelos de neurônios artificiais; a seção 4.3 mostra os principais modelos de RNA; na seção 4.4 são abordados alguns princípios e leis de aprendizado em

RNA; a seção 4.5 sintetiza a apresentação do capítulo com suas conclusões.

4.1 O neurônio biológico

A célula nervosa, ou neurônio, foi identificada anatômicamente e descrita em detalhes pela primeira vez no século XIX pelo médico espanhol Ramón y Cajal. Nessa célula, a membrana celular possui, além de sua função biológica normal, determinadas propriedades essenciais ao funcionamento elétrico da célula nervosa [Kovács, 1996].

O neurônio pode ser dividido em três partes: corpo celular (ou soma) que é o responsável pelos processos metabólicos, axônio e dendritos que são extensões filamentosas a partir da soma. A Figura 4.1 ilustra um neurônio biológico e suas partes.

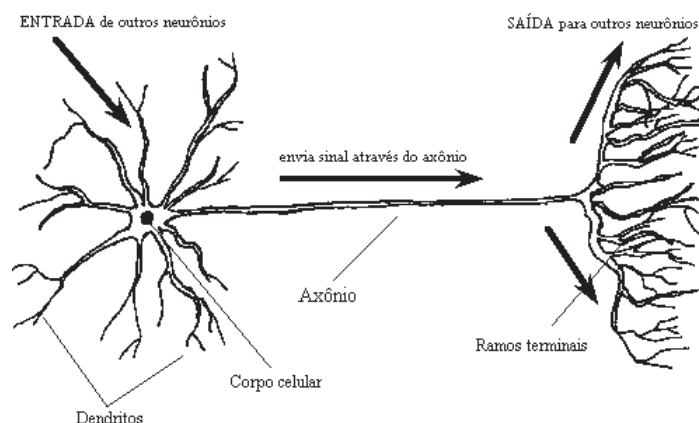


Figura 4.1: Representação de um neurônio biológico.

Os dendritos, em geral, cobrem um volume muitas vezes maior que o próprio corpo celular formando uma *árvore dendrítica*. O axônio, também conhecido como fibra nervosa é a parte que serve para conectar um neurônio a outro(s). Geralmente o neurônio possui apenas um axônio, mas este pode apresentar ramificações. O fluxo de um sinal elétrico através do neurônio segue o caminho: dendritos, soma e axônio, como mostrado na Figura 4.1. Na maioria das células nervosas que constituem o sistema nervoso central dos vertebrados, há uma camada de mielina que serve para acelerar a transmissão do sinal pelo axônio através do *impulso saltatório* [Kovács, 1996].

Em relação à transmissão de informações através dos neurônios, estas são executadas através de impulsos elétricos que são transmitidos pela célula nervosa. As manifestações elétricas de neurônios biológicos foram primeiramente observadas no século XIX por Dubois Reymond com o auxílio de galvanômetros [Kovács, 1996]. Posteriormente, com a invenção do tubo de raios catódicos este processo pôde ser melhor compreendido.

A partir daí, o neurônio passou a ser conceituado como o elemento básico do sistema nervoso, sendo constituído por muitas entradas e uma saída. As entradas ocorrem através

das *conexões sinápticas* (ou sinapses) que ligam a árvore dendrítica ao axônio de outras células nervosas. A Figura 4.2 ilustra a sinapse entre dois neurônios. Nestes pontos a informação é constituída por pulsos elétricos, ou potenciais de ação ou ainda impulsos nervosos [Harvey, 1994; Hassoun, 1995].

A ativação destes pulsos é feita através de um processo de difusão assimétrica entre íons de potássio e sódio através da membrana celular, conhecido como *bomba de sódio*. Os potenciais elétricos gerados por este processo possuem funções específicas e podem ser assim divididos: o potencial de repouso é aquele quando o interior da membrana está eletronegativo de algumas dezenas de volts em relação ao exterior; diz-se que a membrana está despolarizada quando está menos eletronegativa que o seu potencial de repouso; o potencial de ação é uma onda de despolarização propagada ao longo da membrana [Kovács, 1996].

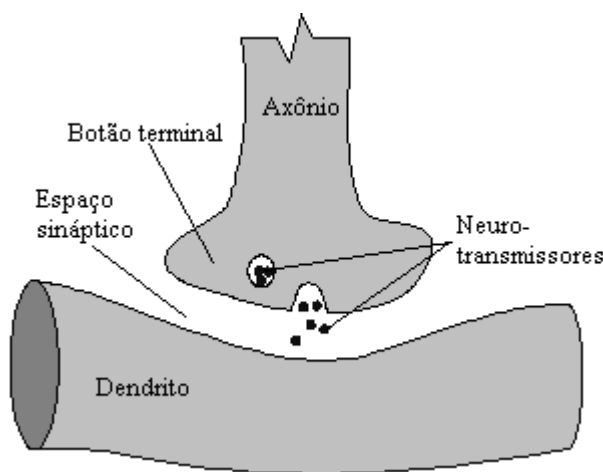


Figura 4.2: Sinapse entre dois neurônios biológicos.

Resumidamente, o funcionamento de um neurônio biológico é dirigido pelo potencial de ação. Este potencial possui várias características e intervalos próprios, como mostrado na Figura 4.3, que promovem o funcionamento e limite de tempo entre potenciais de ação consecutivos. Há ainda a característica do neurônio possuir um *limiar de disparo*, que é um determinado potencial que deve ser ultrapassado para o neurônio fornecer uma saída [Kovács, 1996, 1997].

Também importante, é saber que cada uma das entradas em cada um dos dendritos no neurônio é uma das componentes que compõe a entrada total da unidade e que essas podem ser temporalmente deslocadas. Desta maneira, o processamento da informação no neurônio é um *processamento espacial-temporal* [Kovács, 1996]. A Figura 4.4 ilustra o processamento espacial-temporal do neurônio, onde as entradas A, B e C, cada uma apresentando um potencial de ação em tempos diferentes (t_1, t_2 e t_3), produzindo uma saída que depende da composição das entradas. O sinal de saída do neurônio depende também da função do neurônio que leva em consideração o limiar de disparo, a qual é denominada *função de ativação* do neurônio [Haykin, 1998].

Ainda na Figura 4.4 deve-se notar que as diversas entradas através das sinapses

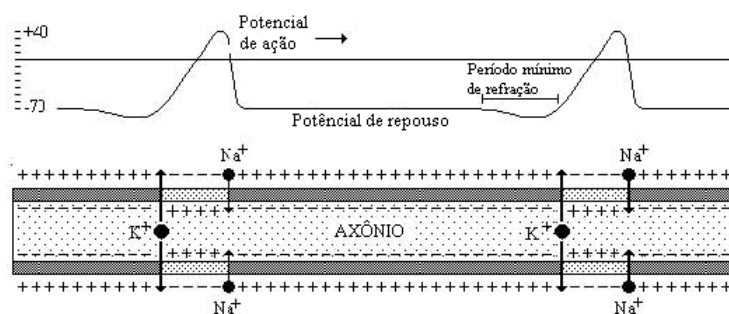


Figura 4.3: Funcionamento da *bomba de sódio* no neurônio.

apresentam ganhos diferentes, denominados ganhos (ou pesos) sinápticos, promovendo assim uma ponderação sobre cada uma das entradas. Os sinais \oplus e \ominus indicam que as entradas são para o neurônio, respectivamente, *excitatórias* ou *inibitórias*. Uma entrada excitatória é aquela que promove um acréscimo no potencial de ação total e inibitória para um decréscimo do mesmo.

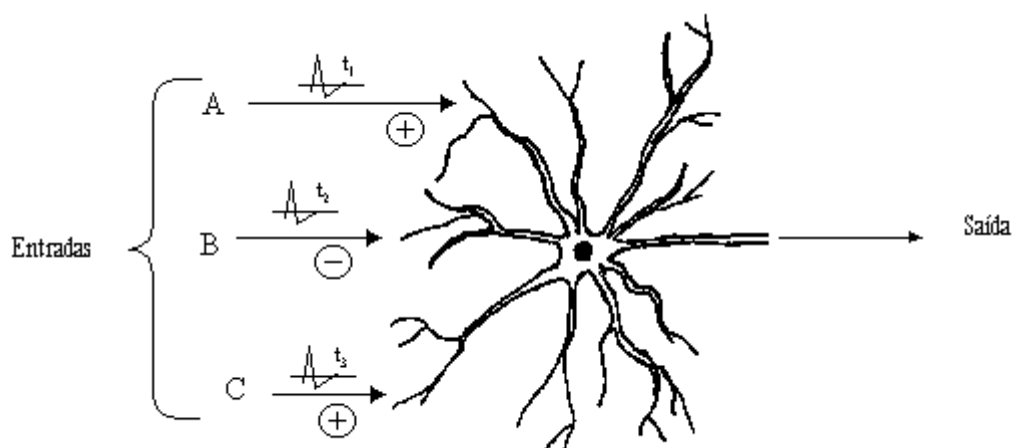


Figura 4.4: Processamento espacial-temporal dos estímulos para o neurônio biológico.

Este modelo de processamento permite então a elaboração para um neurônio artificial que reproduza o comportamento do neurônio biológico e é discutido na seção seguinte.

4.2 O neurônio artificial

Com o conhecimento sobre o comportamento elétrico do neurônio biológico, o desenvolvimento de modelos para os neurônios artificiais iniciou seu crescimento por volta da década de 40.

Nesta época, os avanços na determinação (experimental) das propriedades eletrofisiológicas do neurônio biológico e os resultados divulgados por Alan Turing e John von Neumann que indicavam que a natureza essencial da inteligência seria booleana, resultante (possivelmente) da atividade neuronal, surgiram os primeiros modelos [Harvey, 1994; Kovács, 1996].

4.2.1 Neurônio de MacCulloch & Pitts

Sem dúvida, o primeiro trabalho sobre redes neurais artificiais foi da autoria de McCulloch & Pitts (1943). Nele, era proposto um modelo para o neurônio artificial baseado num dispositivo binário.

O modelo proposto era bastante simples a despeito do conhecimento sobre o comportamento elétrico do neurônio biológico adquirido até aquela época. O neurônio era um dispositivo de saída binária, com entradas apresentando ganhos sinápticos diversos podendo ainda ser excitatórias ou inibitórias. Uma representação do neurônio de McCulloch & Pitts (NM&P) encontra-se na Figura 4.5.

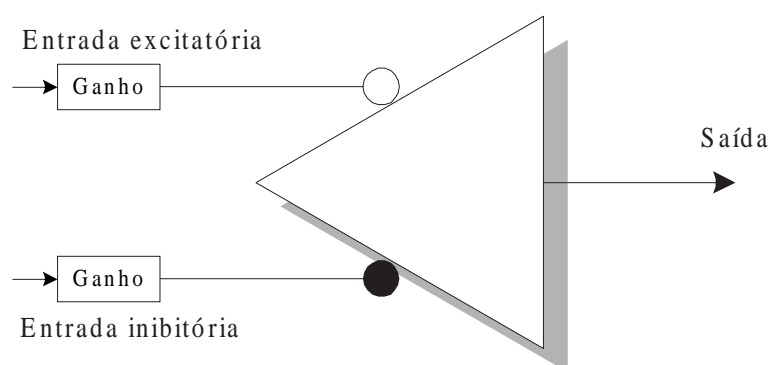


Figura 4.5: Neurônio de McCulloch-Pitts.

O funcionamento era bastante simples. Calculada a soma ponderada das entradas (excitatórias e inibitórias), se este resultado fosse maior que um certo limiar a saída seria “A”, caso contrário seria “não A”, onde “A” corresponde a um certo valor lógico. Desta maneira o NM&P consiste de uma função não linear tipo decisor, ou seja, era um *discriminador linear* para o caso particular de entradas binárias [Kovács, 1996].

O modelo desse neurônio permitiu a implementação de várias funções lógicas. Isso era possível porque o modelo implementava uma função linear capaz de separar em duas classes (“A” e “não A”), ou seja, separava o espaço n -dimensional do \mathbb{R}^n em duas regiões. Com isso o NM&P era capaz de resolver os problemas linearmente separáveis.

Entretanto, embora a maioria das funções lógicas de duas entradas seja linearmente separáveis, com o aumento do número de entradas a maioria dos problemas não é linearmente separável. Essa limitação foi ponto de partida para outras pesquisas em RNA.

A essência da proposta em [McCulloch & Pitts, 1943] foi a seguinte [Kovács, 1996]:

“A inteligência é equivalente ao cálculo de predicados que por sua vez pode ser implementado por funções booleanas. Por outro lado, o sistemas nervoso é composto de redes de neurônios, que com as devidas simplificações, tem a capacidade básica de implementar estas funções booleanas. Conclusão: a ligação entre inteligência e atividade nervosa fica estabelecida de forma científica.”

4.2.2 Neurônio de Rosenblatt - Perceptron

No final da década de 50, Frank Rosenblatt tomando por base o modelo do NM&P formulou um modelo para o neurônio que ele chamou de *perceptron*.

O modelo estendeu o do NM&P para os casos de variáveis não booleanas, consistindo de um combinador linear das entradas seguido de um decisor ($\text{sgn}(\cdot)$). A Figura 4.6 ilustra o modelo do perceptron.

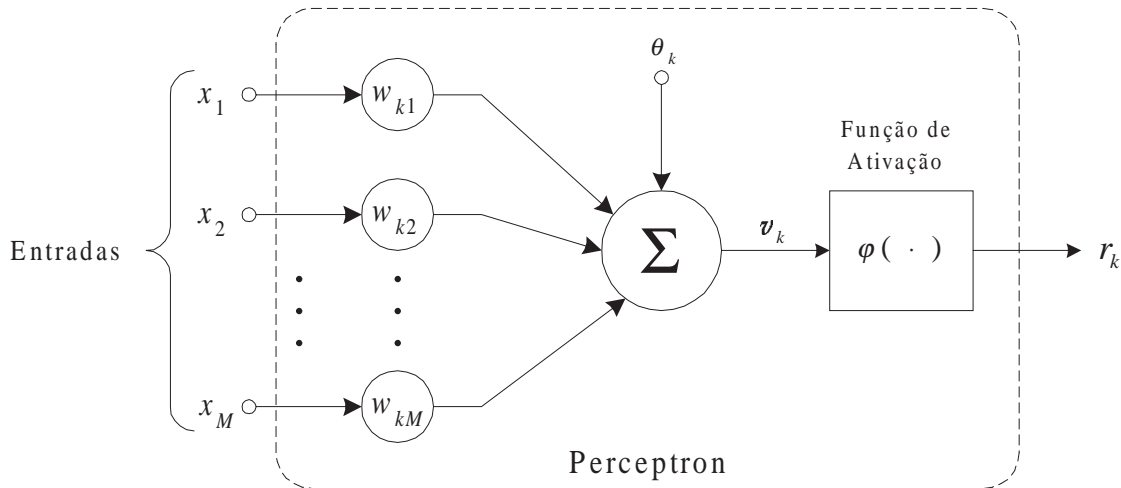


Figura 4.6: Modelo do neurônio artificial *perceptron*.

em que os valores w_{ki} , $i \in \mathbb{N}^*$, são chamados de *pesos sinápticos*, ou simplesmente pesos, do neurônio k que ponderam as entradas i e θ_k é a polarização ou *bias*¹ do neurônio k .

Assim, a saída de perceptron k será escrita por:

$$r_k = \varphi \left(\sum_{i=1}^M w_{ki} x_i + \theta_k \right), \quad (4.1)$$

em que $\varphi(\cdot)$ é a chamada função de ativação, que no caso do modelo proposto por Rosenblatt: $\varphi(\cdot) = \text{sgn}(\cdot)$. A próxima seção aborda em maiores detalhes a função de ativação do neurônio artificial.

¹Embora a palavra tenha origem na língua inglesa, será preferida nessa dissertação.

Uma outra grande contribuição do trabalho de Rosenblatt foi a proposição de um método para escolher os valores de w_{ki} que solucionariam um determinado problema de discriminação linear. A descrição desse tópico será realizada na seção 4.4.

O perceptron é sem dúvida um dos modelos mais utilizados em RNA, entretanto como o modelo originalmente proposto implementa um discriminador linear, os problemas não linearmente separáveis não são contemplados.

4.2.3 Função de ativação

A função de ativação do neurônio artificial corresponde a implementação do valor do limiar do neurônio biológico.

Esse limiar deve inferir sobre a composição (ponderada) das entradas (processamento espacial-temporal) do neurônio para fornecer uma determinada saída. Obviamente, a tarefa de escolha de tal função não é tão simples.

No neurônio biológico, essa função é dependente da despolarização da membrana da célula nervosa. Essa despolarização deve ser capaz de suportar uma certa frequência de pulsos de potencial de ação. Mas a frequência de pulsos é limitada bem como a capacidade de acúmulo de cargas na membrana, então a sua despolarização apresenta um comportamento de saturação [Hassoun, 1995; Kovács, 1996], como mostrado na Figura 4.7.

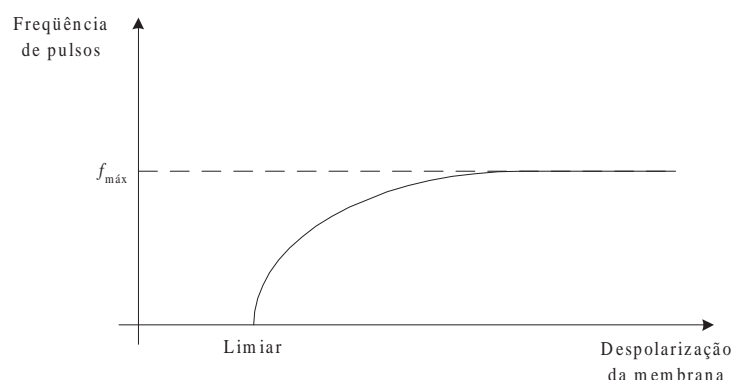


Figura 4.7: Relação entre a frequência de pulsos em um neurônio e intensidade da despolarização da membrana.

Trazendo a análise para o caso do neurônio artificial, a função de ativação deve então tentar reproduzir o comportamento de apresentar uma região de saturação.

Os trabalhos que surgiram inicialmente utilizaram a função $\text{sgn}(\cdot)$, como nos casos de MacCulloch & Pitts e Rosenblatt. Entretanto a função mais simples é a linear proposta no *Adaline* de Widrow [Kovács, 1996], na qual apenas a monotonicidade é preservada.

Siebert, ao modelar os neurônios do sistema auditivo dos vertebrados, propôs uma

função que preserva as características de monotonicidade e saturação na faixa dinâmica, permitindo ajustar o ganho médio através de um parâmetro k , essa função é do tipo [Kovács, 1996]:

$$\varphi(x) = \frac{x^k}{x^k + \alpha^k}, \quad (4.2)$$

em que α é uma constante apropriada.

Algumas funções matemáticas vêm sendo bastante utilizadas nos trabalhos mais recentes sobre RNA. São elas: função sigmóide e função tangente hiperbólica, representadas nas Equações (4.3) e (4.4). Essas funções apresentam tanto as características de monotonicidade e saturação quanto possuem todas as derivadas contínuas.

$$\varphi(x) = \frac{1}{1+\exp(-\beta x)} \quad \text{Sigmóide} \quad (4.3)$$

$$\varphi(x) = \frac{1-\exp(-2\beta x)}{1+\exp(-2\beta x)} \quad \text{Tang. hiperbólica.} \quad (4.4)$$

A Figura 4.8 representa as funções de ativação descritas anteriormente e que são as mais empregadas nos trabalhos sobre RNA.

4.3 Redes neurais artificiais

Com os modelos de neurônios artificiais, as RNA foram formadas por conjuntos de neurônios dispostos em uma determinada arquitetura.

Essas arquiteturas buscam uma disposição dos neurônios de maneira a solucionar os problemas que isoladamente os neurônios não são capazes, tais como os problemas não linearmente separáveis.

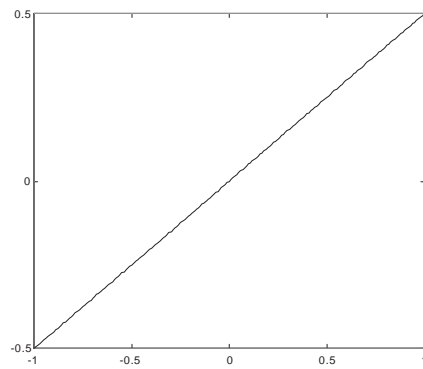
As RNA e suas arquiteturas são classificadas em vários tipos, tais como: RNA multicamadas, RNA de função de base radial, RNA recorrentes, RNA auto-organizáveis, dentre outras.

4.3.1 Redes neurais multicamadas

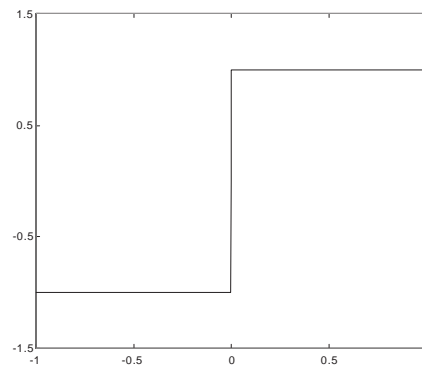
As redes neurais multicamadas (MLP²) é uma rede de perceptrons combinados em camadas como mostrado na Figura 4.9.

As camadas, por sua vez, são classificadas em três tipos: camada de entrada, camadas escondidas e camada de saída. A camada de entrada corresponde aos sinais de estímulo da RNA, as camadas escondidas são responsáveis pelo mapeamento das entradas em algum

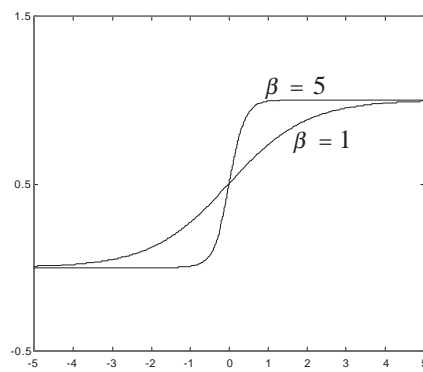
²Do inglês “Multi Layer Perceptron”.



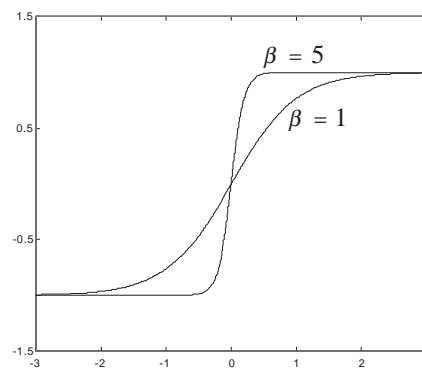
(a): Função linear
 $\varphi(x) = \alpha x$.



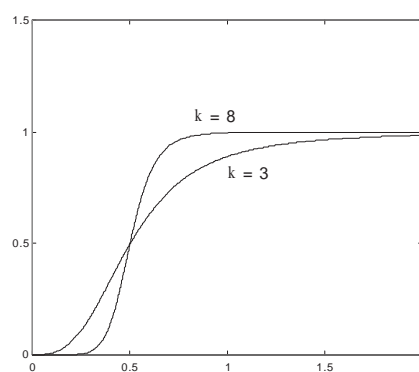
(b): Função degrau
 $\varphi(x) = \text{sgn}(x)$.



(c): Função sigmóide
 $\varphi(x) = \frac{1}{1+\exp -\beta x}$.



(d): Função tangente
 hiperbólica
 $\varphi(x) = \frac{1-\exp -2\beta x}{1+\exp -2\beta x}$.



(e): Função de Siebert
 $\varphi(x) = \frac{x^k}{x^k + 0.5^k}$.

Figura 4.8: Algumas funções de ativação de neurônios artificiais.

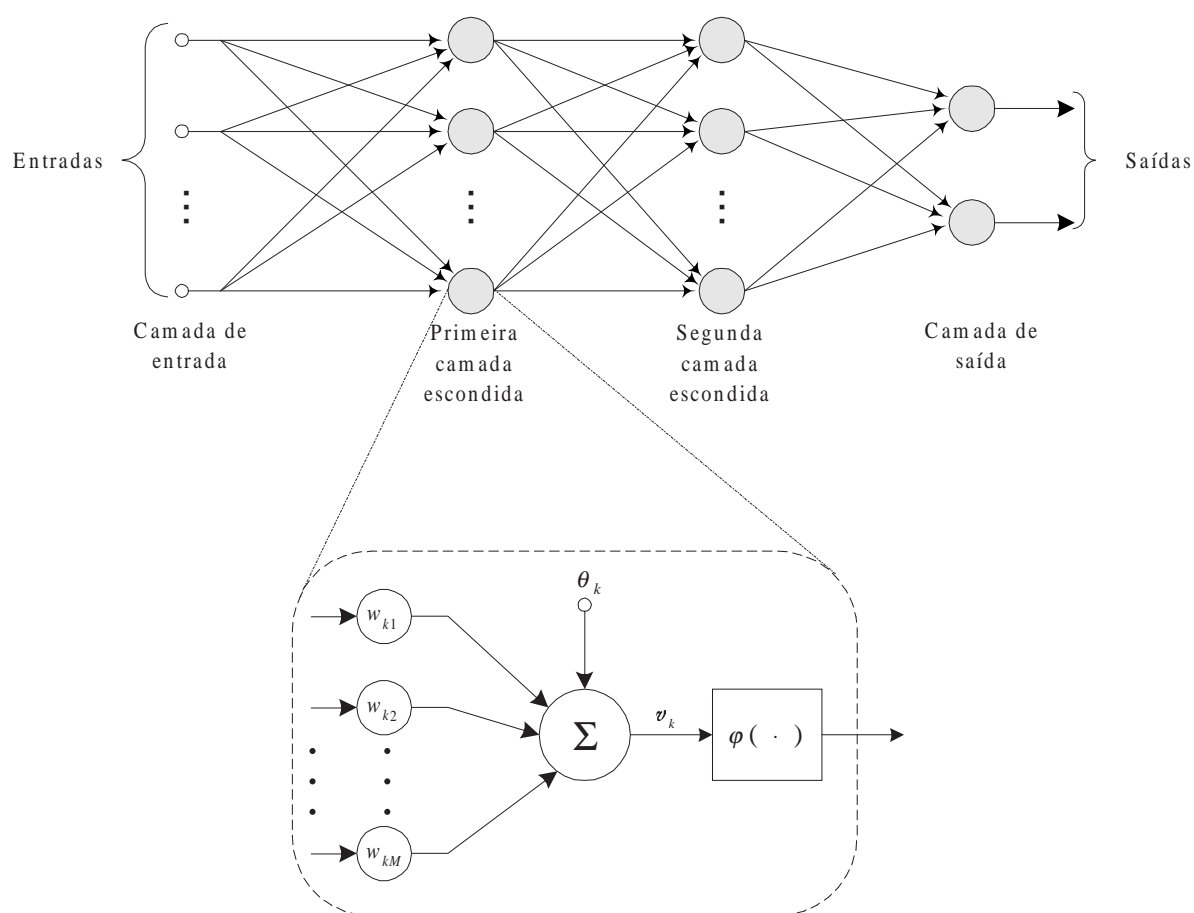


Figura 4.9: Arquitetura de uma rede de perceptrons multicamada com duas camadas escondidas.

espaço de projeção e a camadas de saída é aquela que apresenta os neurônios que fornecem a resposta total da RNA.

O cálculo da saída é então feito através da propagação direta dos sinais de entrada através da rede.

4.3.2 Redes neurais de função de base radial

Analisando-se uma RNA como uma aproximador de funções no espaço de dimensão igual ao número de entradas, a função da RNA corresponde a encontrar uma superfície no espaço multidimensional que seja capaz de melhor “adequar” os dados de entrada, baseado sobre algum critério estatístico.

Essa abordagem motiva o uso das funções de base radial (RFB³) que fornecem uma interpolação estrita no espaço multidimensional já que as camadas escondidas formam um

³Do inglês “Radial-Basis Function”.

tipo de *base* para os padrões de entrada no espaço escondido [Haykin, 1998].

As redes RBF são na sua forma mais simples uma rede de três camadas com *apenas uma* camada escondida, como exemplificado na Figura 4.10. Dessa maneira, a camada escondida aplica uma transformação não linear nos padrões (vetores) de entrada e a camada de saída executa uma combinação linear das saídas dos neurônios.

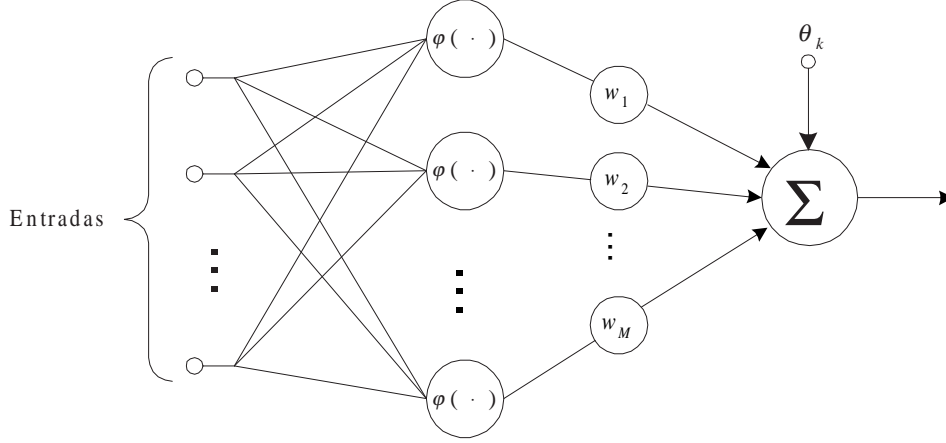


Figura 4.10: Rede de função de base radial.

A função de ativação para as redes RBF também é um pouco diferente daquelas descritas na seção 4.2.3. Nesse caso a função de ativação deve utilizar uma métrica de distância radial, necessitando assim implementar uma função (não linear) que penalize os padrões fora dessa métrica. Essa métrica deriva do teorema de Micchelli que prova a necessidade da utilização de uma matriz de interpolação cujos elementos são não singulares [Haykin, 1998].

As principais funções para as redes RBF, satisfazendo o teorema de Micchelli, são:

⇒ Multiquadrática:

$$\varphi(x) = \sqrt{x^2 + c^2} \quad c > 0 \quad \text{e} \quad x \in \mathbb{R}. \quad (4.5)$$

⇒ Multiquadrática inversa:

$$\varphi(x) = \frac{1}{\sqrt{x^2 + c^2}} \quad c > 0 \quad \text{e} \quad x \in \mathbb{R}. \quad (4.6)$$

⇒ Gaussiana:

$$\varphi(x) = \exp\left(-\frac{(x - c)^2}{2\sigma^2}\right) \quad \sigma > 0, \quad c > 0 \quad \text{e} \quad x \in \mathbb{R}. \quad (4.7)$$

As redes RBF, como as redes MLP, são também redes de propagação direta (“feedforward”) mas com diferentes funções de cada uma das camadas.

As constantes c que aparecem nas Equações (4.5) à (4.7) são chamadas de *centros* das RBFs e são calculados a partir dos dados. Em alguns casos esses centros podem ser determinados pela natureza do problema, mas a estimativa dos mesmos através dos dados é mais comum [Haykin, 1996, 1998].

4.3.3 Redes neurais recorrentes

As redes recorrentes diferem das redes MLP pelo fato de que elas possuem *pelo menos* um laço de realimentação. As redes recorrentes podem ou não apresentar camadas escondidas, como exemplificado na Figura 4.11. Dessa maneira, o processamento dos sinais da rede não é somente direto.

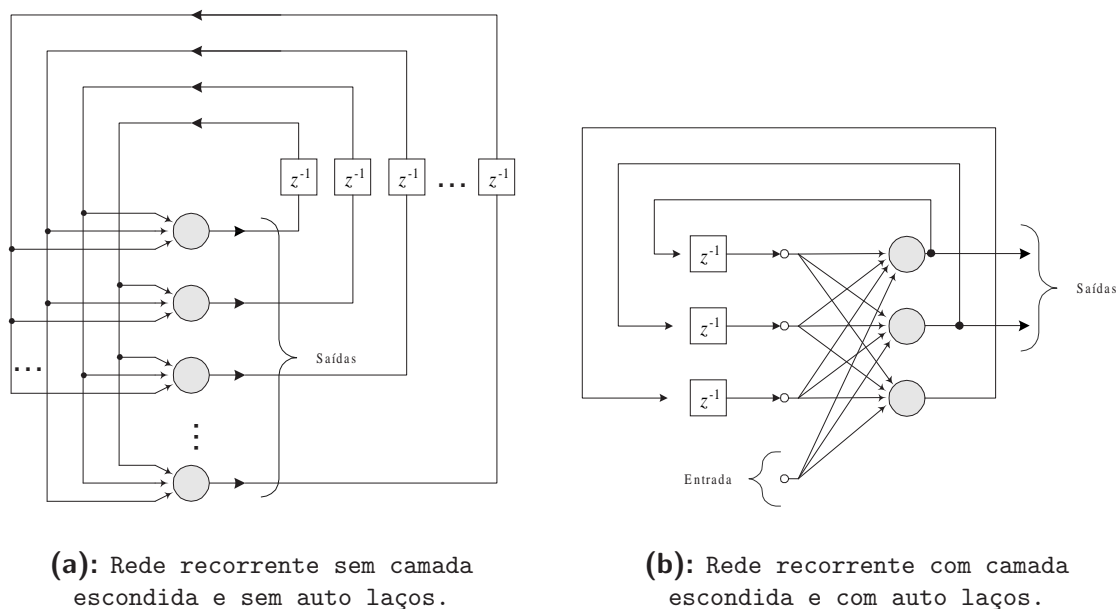


Figura 4.11: Redes neurais artificiais recorrentes.

Uma das mais utilizadas estruturas de redes recorrentes é a *rede de Hopfield* [Kovács, 1996; Haykin, 1998], exemplificada na Figura 4.11(a).

4.4 Aprendizado das redes neuronais

Aprendizado: *S. m.* 1. Ato ou efeito de aprender uma tarefa. 2. *P. ext.* O tempo que dura tal aprendizagem.

Para o cérebro humano aplicar a definição de aprendizado é imediato. A capacidade de executar ou não determinada tarefa ou função determina se o indivíduo aprendeu ou

não. Tal efeito é suposto ser provocado através de um complexo processo de modificação nas conexões sinápticas entre os neurônios que possibilitam a execução da nova tarefa.

No caso das RNA, “aprender” significa encontrar um conjunto de valores para os pesos sinápticos que tornem a rede capaz de realizar uma tarefa. Esses valores são então adaptados através de algum critério (normalmente estatístico) capaz de mensurar algum progresso na capacidade de execução da tarefa em consideração. A Figura 4.12 ilustra um modelo simples de aprendizado de uma máquina de aprendizado.

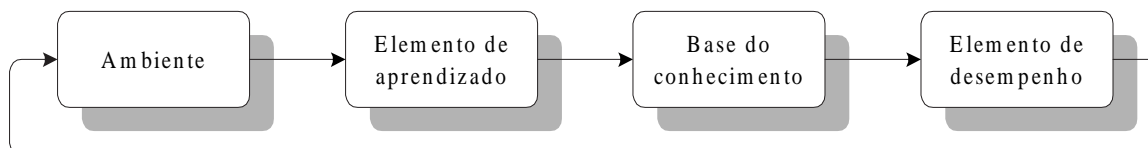


Figura 4.12: Modelo da máquina de aprendizado.

Conjuntamente com a proposição do modelo do perceptron, Rosenblatt propôs uma regra de adaptação para encontrar os valores dos pesos sinápticos, baseado no princípio hebbiano (discutido a seguir) para um discriminador linear. Esse foi um dos primeiros algoritmos de aprendizado surgidos em RNA.

O aprendizado das RNA pode ser subdividido em três etapas considerando os aspectos da convergência algorítmica dos pesos sinápticos e o espaço amostral dos resultados das experiências:

- *Treinamento*: a RNA apenas adapta os pesos sinápticos de acordo com algum critério para algum conjunto de dados.
- *Teste*: a convergência da RNA é testada para um novo conjunto de dados.
- *Validação*: o modelo é validado observando a convergência algorítmica para os conjuntos de dados utilizados na etapa de treinamento e de teste.

Pode-se ainda classificar o aprendizado em duas classes considerando a disponibilidade ou não de dados de referência: aprendizado supervisionado e não supervisionado.

4.4.1 Aprendizado supervisionado

No caso supervisionado existe um agente externo, *um professor*, que é usado como parâmetro de referência, capaz de fornecer a saída desejada, para medir a evolução do aprendizado. A Figura 4.13 ilustra a idéia do aprendizado supervisionado.

Nesse caso o critério de aprendizado é a minimização de uma medida de erro de dados real, ou seja, a diferença entre a resposta desejada, que é aquela fornecida pelo *professor*,

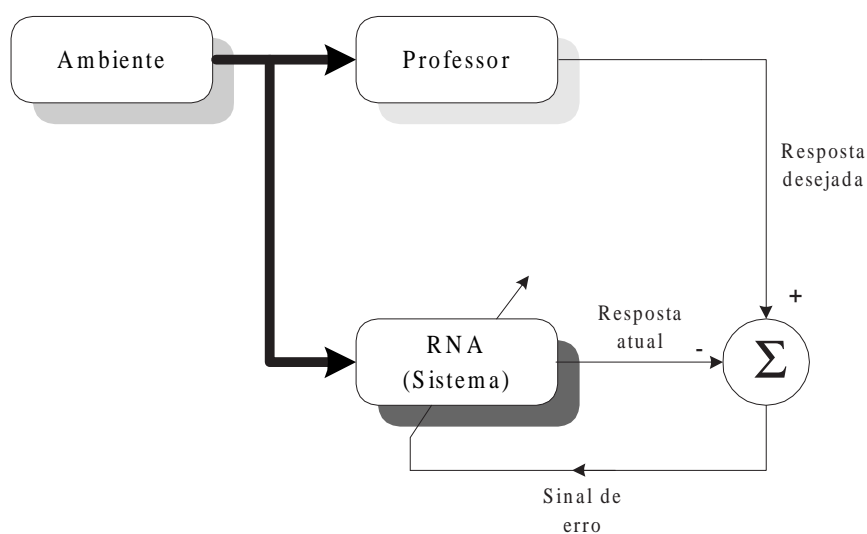


Figura 4.13: Diagrama de blocos do aprendizado supervisionado.

e a resposta fornecida pela rede. Entretanto, em uma ambiente real essa diferença mesmo muito pequena não é zero, faz-se então necessária um limiar para o qual o erro é aceitável.

Mas a adaptação dos pesos sinápticos é bastante delicada visto que a saída é a composição de um processamento não linear através de várias camadas da rede. Um dos mais importantes algoritmos supervisionados para adaptação de RNA é descrito a seguir.

4.4.1.1 Algoritmo back-propagation

O algoritmo back-propagation (BPA⁴) é, provavelmente, o mais conhecido e utilizado algoritmo para adaptação de RNA, foi desenvolvido por Rumelhart, Hinton e Williams em 1986 e consiste numa retropropagação do erro através das diversas camadas da RNA.

Esse algoritmo toma a propagação do erro no sentido inverso ao da propagação dos sinais de entrada. Esse algoritmo tomou como base a *regra delta* proposta por Widrow [Kovács, 1996], ou seja, propunha um processo de iteração local.

A seguir é apresentado um procedimento simplificado de dedução do BPA.

⁴Do inglês “Back-Propagation Algorithm”.

• Dedução do BPA:

Sendo o erro no instante n em um neurônio de saída k , dado por:

$$e_k(n) = d_k(n) - r_k(n), \quad (4.8)$$

em que $d_k(n)$ é a resposta desejada (fornecida pelo *professor*) e $r_k(n)$ é a resposta do neurônio.

Para considerar-se os erros quadráticos de todos os neurônios de saída da rede, tem-se:

$$\mathcal{E}(n) = \frac{1}{2} \sum_{k \in C} e_k^2(n), \quad (4.9)$$

em que o conjunto C inclui todos os neurônios da camada de saída.

Então o funcional a ser minimizado é aquele na Equação (4.9) que representará uma medida de aprendizado. De acordo com a Figura 4.6, tem-se:

$$v_k(n) = \sum_{i=1}^M w_{ki} x_i(n) + \theta_k(n), \quad (4.10)$$

em que M é o total de entradas aplicadas ao neurônio k . E também:

$$y_k(n) = \varphi_k(v_k(n)), \quad (4.11)$$

O processo de busca dos parâmetros é similar ao do algoritmo LMS, ou seja, calculando-se as derivadas parciais $\frac{\partial \mathcal{E}(n)}{\partial w_{ki}(n)}$. E aplicando a regra da cadeia:

$$\frac{\partial \mathcal{E}(n)}{\partial w_{ki}(n)} = \frac{\partial \mathcal{E}(n)}{\partial e_k(n)} \cdot \frac{\partial e_k(n)}{\partial r_k(n)} \cdot \frac{\partial r_k(n)}{\partial v_k(n)} \cdot \frac{\partial v_k(n)}{\partial w_{ki}(n)}, \quad (4.12)$$

em que as derivadas parciais de cada um dos termos:

$$\frac{\partial \mathcal{E}(n)}{\partial e_k(n)} = e_k(n) \quad (4.13)$$

$$\frac{\partial e_k(n)}{\partial r_k(n)} = -1 \quad (4.14)$$

$$\frac{\partial r_k(n)}{\partial v_k(n)} = \varphi'_k(v_k(n))^a \quad (4.15)$$

$$\frac{\partial v_k(n)}{\partial w_{ki}(n)} = x_i(n). \quad (4.16)$$

Logo, a regra de adaptação será:

$$\frac{\partial \mathcal{E}(n)}{\partial w_{ki}(n)} = -\lambda \underbrace{e_k(n) \cdot \varphi'_k(v_k(n))}_{\zeta_k(n)} \cdot x_i(n), \quad (4.17)$$

em que λ é a *taxa de aprendizagem* e $\zeta_k(n)$ é o *gradiente local*.

^a $\varphi'_k(v_k(n)) = \frac{\partial(\varphi_k(v_k(n)))}{\partial v_k(n)}$

As equações de adaptação para os *bias* dos neurônios são análogas às equações para adaptação dos pesos sinápticos. Dessa maneira, o algoritmo para adaptação dos θ_k obedece também ao gradiente descendente do funcional.

As equações de correção dos bias podem ser escritas:

$$\begin{aligned}\Delta\theta_k(n) &= -\lambda \frac{\partial \mathcal{E}(n)}{\partial \theta_k(n)} \\ \Delta\theta_k(n) &= -\lambda \cdot \zeta_k(n).\end{aligned}\tag{4.18}$$

Existem ainda algumas versões modificadas do BPA tentando promover uma aceleração de convergência, tais como: BPA *momentum* e regra “delta-bar-delta” [Haykin, 1998]. Alguns trabalhos têm proposto modificações na regra de treinamento utilizando o BPA para aceleração do processo visando equalização supervisionada [Silva & Gerken, 2000].

A versão complexa do BPA tem dedução análoga, tendo-se que considerar com mais ênfase a funções de ativação pois as mesmas devem ser limitadas, possuir derivadas limitadas e também não linear nas componentes real e imaginária [Haykin, 1996] .

4.4.2 Aprendizado não-supervisionado

Contrário a idéia do aprendizado supervisionado, o aprendizado não supervisionado (ou auto-organizado) não apresenta um agente externo (*professor*) que realiza o critério de medida estatística da evolução da RNA. Nesse caso apenas o ambiente fornece as informações para adaptação dos pesos sinápticos, como mostrado na Figura 4.14.

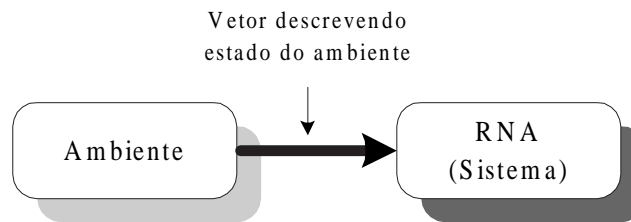


Figura 4.14: Diagrama de blocos do aprendizado não supervisionado.

Como a figura sugere, alguma regra de aprendizado deve então inferir a partir dos dados de excitação (entradas) para que a resposta seja satisfatória. Alguns desses algoritmos são descritos a seguir.

4.4.2.1 Algoritmos auto-organizáveis

- Algoritmo de Hopfield

Um dos algoritmos auto organizáveis mais importantes é o algoritmo de Hopfield, onde a rede de Hopfield já havia sido descrita esquematicamente na Figura 4.11(a). Nesse algoritmo, a determinação dos pesos sinápticos pode ser dada de forma explícita, não há processo de treinamento. Os pesos sinápticos são armazenados através da regra do produto externo dos vetores de dados. Nesse caso tem-se uma matriz de pesos simétrica. Essa fase é chamada de *fase de armazenagem*.

Uma outra fase, chamada de *fase de execução*, é responsável pela atualização dos pesos sinápticos para que a rede forneça *estados estáveis*, ou seja, $r_k(n+1) = r_k(n)$.

Da característica da matriz simétrica de pesos, a rede pode ser então descrita pela função de energia de Lyapunov, que é uma função decrescente monotônica [Haykin, 1998].

- Mapas de Kohonen

Esses sistemas, inspirados nos sistemas de percepção como a visão, têm como principal propriedade poder codificar nos neurônios vizinhos os sinais de entrada que possuem características muito próximas.

As regras de aprendizado propostas por Kohonen têm base sobre observações biológicas. Dessa maneira, a rede apresenta neurônios que respondem de maneira específica a um tipo de estímulo, ou seja, é uma rede de respostas localizadas. Os problemas de classificação e reconhecimento de padrões encontram uma ferramenta bastante adequada nas redes de Kohonen [Balay, 1995].

Para implementar um funcionamento como o descrito acima, utiliza-se uma função que possui as seguintes características:

- dentro de uma vizinhança lateral próxima, os neurônios conectados à ele têm ação excitatória;
- numa vizinhança mais distante as ações são inibitórias.

Este tipo de aprendizado é dito *competitivo*, pois apenas alguns (ou apenas um) neurônios recebem atualização dos seus pesos através da regra de aprendizado.

A Figura 4.15 ilustra uma rede de Kohonen.

4.4.3 Aprendizado hebbiano

O postulado sobre aprendizado de Hebb, um neurofisiologista, enunciado em 1949, é o mais antigo e também mais famoso de todas as regras de aprendizado. O enunciado diz o seguinte [Haykin, 1998]:

“Quando um axônio de uma célula A está próximo o suficiente para excitar uma célula B e repetidamente ou persistentemente toma parte na sua ativação, algum processo de crescimento ou mudanças metabólicas tomam lugar em uma delas ou em ambas as células de tal maneira que a eficiência de A como uma das células ativadoras de B é aumentada.”

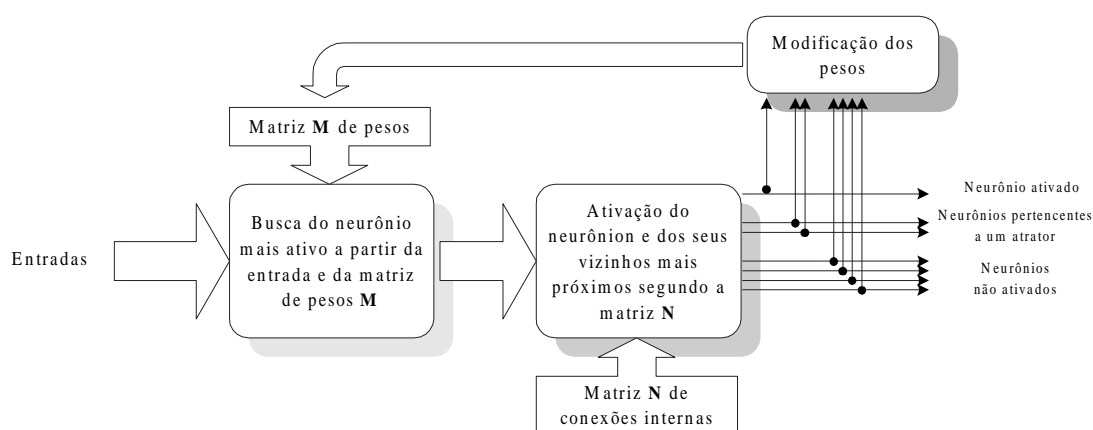


Figura 4.15: Estrutura da rede de Kohonen.

Essa modificação foi proposta por Hebb como base do aprendizado associativo à nível celular, e foi enunciada no contexto neurobiológico. Uma expansão do aprendizado hebbiano pode ser escrita em duas partes [Haykin, 1998]:

1. Se dois ou mais neurônios na sinapse são ativados sincronamente a potência da sinapse é então aumentada.
2. Se dois ou mais neurônios na sinapse são ativados assincronamente a potência da sinapse é então diminuída.

A sinapse que observa o comportamento descrito anteriormente é então chamada de *sinapse hebbiana*⁵.

Através da definição da sinapse de Hebb, pode-se então, generalizar o conceito de aprendizado hebbiano como sendo um processo que fortalece (aumenta o ganho) de uma sinapse que produz atividades sinápticas (entradas e saídas) positivamente correlacionadas.

Pode-se então dividir o princípio de aprendizado em três categorias:

- Lei hebbiana
Nesse tipo de sinapse, são reforçadas positivamente as sinapses cujas entradas e saídas apresentem correlações positivas.
- Lei anti-hebbiana
Contrapondo a lei hebbiana, nesse caso são reforçadas positivamente as sinapses cujas entradas e saídas apresentem correlações negativas.
- Lei não hebbiana
As sinapses são reforçadas através de uma regra que não obedece ao postulado de Hebb.

⁵Embora o nome de Hebb tenha sido atribuído a esse tipo de sinapse o postulado original de Hebb contém apenas a primeira característica.

Matematicamente, isso corresponde a uma regra do tipo:

$$\Delta w_{ki}(n) = \lambda \cdot \underbrace{r_k(n)}_{\text{Saída}} \cdot \underbrace{x_i(n)}_{\text{Entrada}}. \quad (4.19)$$

em que λ é a taxa de aprendizado.

O tratamento sobre o aprendizado de Hebb será importante para a compreensão da proposição de um algoritmo de aprendizado no capítulo 6.

4.5 Síntese e conclusões

O material apresentado nesse capítulo é a base do estudo da teoria das RNA. A descrição do modelo do neurônio biológico, bem como dos modelos de neurônios artificiais servem para fazer um encadeamento de como evoluiu a ciência cognitiva e a inteligência artificial. Provavelmente não é necessário uma descrição como a feita nesse capítulo para permitir uma compreensão da proposição do trabalho (capítulo 6), mas julga-se importante a abordagem aqui descrita.

A inserção de uma seção de discussão sobre o neurônio biológico é de importância somente histórica para notar as bases dos modelos de neurônio artificial.

A descrição do neurônio de MacCulloch & Pitts é apresentada por sua grande importância no campo da pesquisa em RNA. Embora o seu modelo seja limitado, as suas idéias serviram de base para os trabalhos seguintes e inspiraram vários pesquisadores nas mais diversas áreas tais como: Wiener (cibernética), von Neumann (ciência da computação) e Minsky (inteligência artificial).

O modelo do perceptron possibilitou o desenvolvimento de várias arquiteturas de RNA, sendo hoje ainda bastante utilizado.

As redes de Hopfield bem como os mapas de Kohonen foram um marco na teoria de RNA pois marcaram o fim do obscurantismo nesse campo, que se instalou entre a proposição do NM&P até a proposição de uma estrutura de memória associativa.

Entretanto, em processamento de sinais, as redes RBF têm encontrado bastante espaço devido à característica de mapeamento não linear do espaço das entradas, permitindo com isso uma aplicação através da abordagem de classificação [Mulgrew, 1996, 1998; Montalvão et al., 1998; Montalvão, 2000].

Uma maior atenção sobre as características do aprendizado em redes neurais é justificada pela proposição, no capítulo 6 desta dissertação, de um modelo híbrido de aprendizado e que requer alguns conceitos nesse campo.

Por fim, as referências citadas nesse capítulo fornecem uma ampla e não exaustiva discussão sobre as RNA que não são discutidas nesta dissertação. No entanto, servem como sugestão para aqueles que desejem de maiores informações nessa área.

EQUALIZAÇÃO E CLASSIFICAÇÃO: RECUPERAÇÃO DA INFORMAÇÃO

EQUALIZAÇÃO adaptativa tem sido durante muito tempo abordada pelo ponto de vista da recuperação do sinal através da estimativa temporal do sinal transmitido.

Os métodos de estimação de canais e de sinais têm sido amplamente investigados durante os últimos anos e os grandes avanços na área de equalização adaptativa foram devido a tais desenvolvimentos.

Uma nova abordagem para o problema de equalização é feita através de técnicas baseadas na teoria da *classificação*, bastante sedimentadas no campo do reconhecimento de formas. Sob este enfoque, a recuperação da informação ganha então um aspecto de organização de sinais feita através de estruturas (equalizadores) capazes de formular uma superfície ou função de separação entre classes diferentes de sinais geradas, a partir dos símbolos transmitidos.

Este capítulo aborda algumas das técnicas de classificação aplicadas em processamento de sinais. Algumas técnicas clássicas de equalização, algoritmos e estruturas também encontram-se descritas. Embora sejam conhecidas as relações entre filtragem linear e classificação para equalização [Montalvão, 2000], o rigor matemático das definições não será objetivo deste capítulo.

O restante do capítulo é organizado da seguinte forma: a seção 5.1 descreve as principais técnicas de filtragem aplicadas à equalização; a seção 5.2 descreve algumas das técnicas de filtragem adaptativa utilizadas na equalização bem como alguns dos principais algoritmos autodidatas; a seção 5.3 ilustra os principais conceitos sobre classificação para recuperação da informação e finalmente na seção 5.4 são apresentadas a síntese e as conclusões sobre

o capítulo.

5.1 Abordagem clássica

Como anteriormente mencionado, as técnicas clássicas de equalização adaptativa são baseadas nas técnicas de filtragem linear. É então necessário fazer uso de uma estrutura linear que possua parâmetros adaptáveis.

Uma das estruturas lineares mais utilizadas é o *filtro transversal*, que encontra-se esquematizado na Figura 5.1. Isso é justificado tanto pela sua simplicidade algorítmica quanto pela característica de estabilidade numérica [Montalvão, 2000].

Neste tipo de estrutura, também conhecida como “tapped-delay line”¹, o sinal de saída corresponde a uma combinação linear dos sinais de entrada deslocados no tempo com os coeficientes² do filtro.

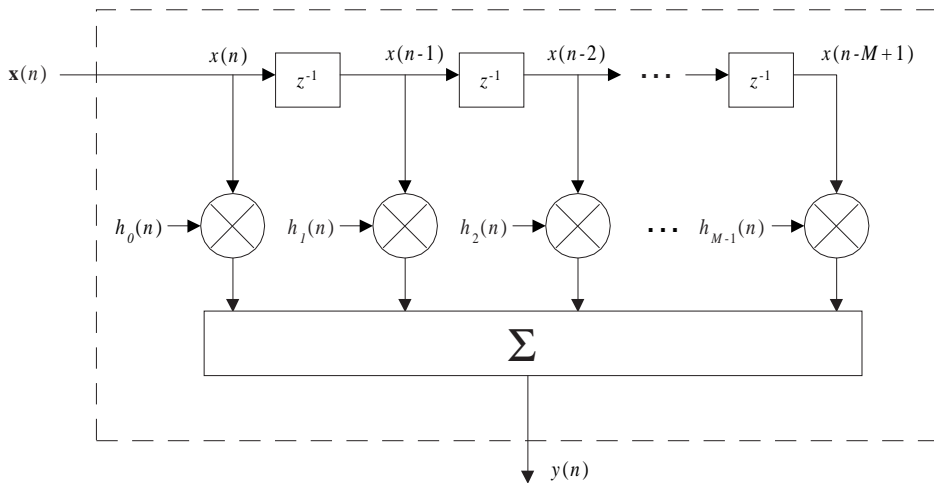


Figura 5.1: Filtro linear transversal.

Em termos de recuperação da informação o filtro equalizador deve ser capaz de retirar a interferência provida pelo canal ao sinal transmitido através da compensação dos efeitos distorsivos do canal. Assim, de acordo com o critério de seleção do equalizador pode-se classificar a equalização síncrona como [Brossier, 1997]:

¹Linha de coeficientes deslocados temporalmente.

²É muito comum na literatura referir-se aos coeficientes adaptáveis do filtro linear transversal através de seu equivalente em inglês “taps”.

- **Abordagem ótima**

Do ponto de vista da transmissão da informação, a informação no receptor deve ser restituída com uma taxa de erro mínima possível. E considerando-se que a RI do canal não excede alguns intervalos de símbolos, é factível utilizar um estimador sub-ótimo em relação a taxa de erro. O algoritmo de Viterbi [Forney, 1973; Proakis, 1995] é bastante utilizado devido à sua capacidade de minimizar a taxa de erro com uma complexidade de cálculo relativamente aceitável.

- **Abordagem sub-ótima**

Nos casos em que a duração da RI do canal ultrapassa alguns poucos intervalos de símbolo, a utilização do algoritmo de Viterbi para o processamento em tempo real pode ter uma complexidade computacional extremamente alta. Faz-se então necessário a utilização de um outro critério que não corresponde a minimização da taxa de erro. Alguns destes critérios são:

- **Filtragem de Wiener síncrona**

O critério consiste em minimizar a potência do erro de estimação, ou seja minimizar o EQM. Desde que a presença de ruído não permite uma anulação completa da IES, pode ocorrer uma amplificação importante do ruído [Mota, 1992]. Assim, é interessante obter uma solução de compromisso que permite uma boa redução da IES e uma atenuação do ruído. Logo, minimizando o critério $J(\mathbf{h}) = \mathbb{E} \{|y(n) - a(n - \delta)|^2\}$ obtém-se a solução [Haykin, 1996]:

$$\mathbf{h}_W = \mathbf{R}_x^{-1} \mathbf{p}, \quad (5.1)$$

em que $\mathbf{R}_x = \mathbb{E}\{\mathbf{x}^*(n)\mathbf{x}^T(n)\}$ (matriz de auto correlação) e $\mathbf{p} = \mathbb{E}\{\mathbf{x}^*(n)a(n - \delta)\}$ (vetor de correlação cruzada). Como a solução da filtragem de Wiener sem considerar o ruído fornece a melhor resposta para uma estrutura linear [Mota, 1992], é interessante estabelecer como referência o EQM mínimo da solução de Wiener sem ruído, fornecido pela Equação (5.2):

$$J_{\min} = \sigma_a^2 - \mathbf{p}^H \mathbf{R}_{\bar{x}}^{-1} \mathbf{p}, \quad (5.2)$$

em que $\mathbf{R}_{\bar{x}}$ é a matriz de auto correlação sem presença do ruído.

- **Filtragem inversa síncrona**

Este critério, também conhecido como *forçagem a zero* (ZF³) ou *identificação inversa*, visa anular as componentes da RI global do sistema, salvo uma única componente dentro de um suporte finito, sem contemplar a presença do ruído. Nesse ponto é suficiente a recuperação da informação através de um filtro causal e inversível. Esse método permite recuperar a informação da mesma forma que a filtragem de Wiener, considerando o ruído nulo, e o controle sobre a IES residual. A solução do critério ZF é então dada por [Mota, 1992; Haykin, 1996; Montalvão, 2000]:

$$\mathbf{h}_{ZF} = \mathbf{F}^\dagger \mathbf{d}_\delta, \quad (5.3)$$

em que \mathbf{F}^\dagger é a *pseudo-inversa* [Haykin, 1996] da matriz de convolução do canal [Montalvão et al., 1998] e $\mathbf{d}_\delta = [0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0]^T$.

↑
linha δ

³Do inglês “Zero Forcing”.

Vale ressaltar que o critério ZF tenta minimizar de alguma maneira o critério de **máxima distorção** (MD) [Brossier, 1997; Montalvão, 2000] definido como

$$\text{MD}(\mathbf{g}) = \frac{\sum_{k=0}^{N+M-2} |g_k| - \max_k |g_k|}{\max_k |g_k|}, \quad (5.4)$$

em que $\mathbf{g} = \mathbf{F}\mathbf{h}$ é a *resposta global* do sistema. O critério MD corresponde à condição de *olho aberto* para os casos em que $\text{MD}(\mathbf{g}) < 1$ e à condição de *olho fechado* para $\text{MD}(\mathbf{g}) \geq 1$.

Na ausência ou indisponibilidade do conhecimento sobre as estatísticas pertinentes aos critérios empregados, faz-se necessário o uso da *equalização adaptativa*, para buscar o conhecimento da informação durante o processo. Esta abordagem, que permite a atuação algorítmica em tempo real, é discutida na seção seguinte.

5.2 Principais técnicas de equalização adaptativa

As principais técnicas de equalização adaptativa são classificadas em relação à estratégia de adaptação dos parâmetros variáveis do filtro de maneira a recuperar a informação transmitida.

Neste ponto várias estratégias podem ser listadas e classificadas de acordo com o critério de estimação empregado para a restauração do sinal transmitido. Pode-se então classificar as técnicas de equalização adaptativa em duas grandes categorias.

5.2.1 Equalização adaptativa supervisionada

Na equalização supervisionada, também comumente denominada *treinada*, a exemplo do aprendizado supervisionado nas redes neurais, o erro de estimação é computado através do conhecimento da informação transmitida. Dessa maneira, o *professor*, será o próprio receptor que durante breves períodos entre períodos regulares de transmissão de dados disporá localmente de uma determinada sequência sincronizada a uma sequência réplica transmitida.

Neste tipo de processamento descrito, é muito comum utilizar-se a minimização do EQM verdadeiro, ou seja, $\mathbb{E} \left\{ \left| y(n) - \hat{d}(n) \right|^2 \right\}$, em que $\hat{d}(n) = a(n - \delta)$. É também conhecido que a solução da equalização supervisionada através da minimização do EQM, utilizando-se o algoritmo LMS, converge em média para a solução de Wiener [Mota, 1992; Cavalcanti, 1996].

Logicamente, como já mencionado acima, existe nesta estratégia um período reservado para a adaptação dos coeficientes do equalizador, chamado *período de treinamento*, onde não há transmissão efetiva de informação.

Nos casos já citados anteriormente, em que não é viável ou mesmo possível a disponibilização de tal período para adaptação, faz-se necessário o uso de outras estratégias para adaptação do equalizador.

5.2.2 Equalização adaptativa não-supervisionada ou autodidata

A equalização não-supervisionada ou autodidata ou cega é uma estratégia de adaptação onde não se faz uso de uma sequência de treinamento. O símbolo de referência é estimado segundo algum critério estatístico. A Figura 5.2 ilustra a idéia da estimação do sinal de referência.

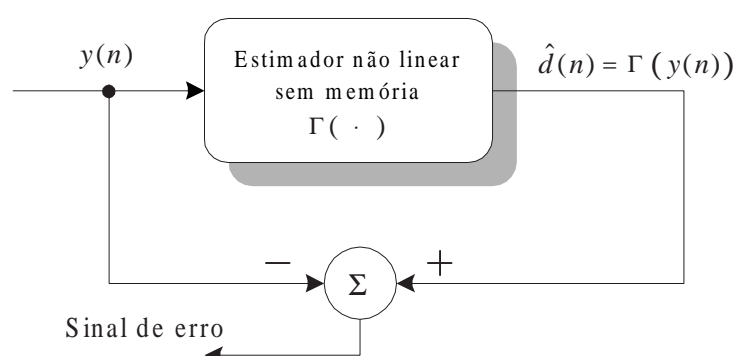


Figura 5.2: Estimador não linear.

Como o erro será então uma função não linear dos coeficientes do filtro, a minimização será de uma função não quadrática (ordem maior que 2).

Várias técnicas de equalização autodidata são possíveis de serem usadas, tais como as baseada em *cumulantes*, *poliespectrais* e as que exploram as técnicas *cicloestacionárias* [Porat & Friedlander, 1991; Rocha, 1996]. Entretanto, somente as técnicas de Bussgang serão discutidas neste trabalho.

5.2.3 Técnicas de Bussgang

Dentre as principais técnicas autodidatas as técnicas de Bussgang são as mais utilizadas. As técnicas de Bussgang usam o conhecimento da fdp não gaussiana da sequência de dados transmitidos e da sequência de sinais na saída do equalizador. O erro é então construído através de uma estimativa de máxima verossimilhança (MLE⁴) da sequência de transmissão, baseada em uma suposta fdp dos sinais na saída do equalizador [Haykin, 1994a, 1996]. Alguns dos principais algoritmos que fazem uso das técnicas de Bussgang para equalização cega [Hilal, 1993; Rocha, 1996; Haykin, 1996] são descritos a seguir.

⁴Do inglês “Maximum Likelihood Estimation”.

5.2.3.1 Algoritmo de Decisão Dirigida

A primeira estratégia de equalização autodidata utilizou o algoritmo LMS para atualizar os parâmetros do filtro, onde o erro considerado foi o erro de decisão $\varepsilon_{DD}(n)$ (Equação (5.5)). Tal estratégia funciona para os casos de baixa IES. Isso porque, com as decisões corretas, a convergência para o ponto ótimo de operação é assegurada [Haykin, 1996]. Na Equação (5.5) é mostrado o algoritmo DD-LMS.

$$\mathbf{h}(n+1) = \mathbf{h}(n) - \underbrace{\mu (y(n) - \text{Dec}(y(n)))}_{\varepsilon_{DD}(n)} \mathbf{x}(n) \quad \mu > 0. \quad (5.5)$$

5.2.3.2 Algoritmo de Sato

O trabalho de Sato [Sato, 1975] iniciou um período de busca de novas estratégias de equalização cega onde a condição de baixa IES não fosse restrigente. O algoritmo de Sato utiliza uma ponderação da função $\text{sgn}(\cdot)$ como estimador (Equação (5.6)). Apesar de apresentar bons resultados para sistemas com modulações mais robustas, o algoritmo tem atuação limitada diante de canais com forte distorção. No entanto, a partir desse trabalho, o surgimento de pesquisas em critérios utilizando funções não convexas provocou um grande avanço na teoria de equalização cega. A seguir, a Equação (5.6) apresenta a adaptação para o algoritmo em que $\varepsilon_{\text{Sato}}(n)$ é o erro de Sato e $[\gamma \text{sgn}(y(n))]$ é a estimativa do sinal desejado.

$$\mathbf{h}(n+1) = \mathbf{h}(n) - \underbrace{\mu (y(n) - \gamma \text{sgn}(y(n)))}_{\varepsilon_{\text{Sato}}(n)} \mathbf{x}(n) \quad \mu, \gamma > 0. \quad (5.6)$$

5.2.3.3 Algoritmo de Godard

O trabalho de Godard [Godard, 1980] começou a ir além dos limites de robustez deixados pela estratégia de Sato. Neste caso, um conjunto de funções custo não convexas penalizam o desvio do módulo do sinal na saída do equalizador de um módulo constante chamado *raio de equalização* R_p . Atualmente, o algoritmo mais utilizado para equalização cega, o *Constant Modulus Algorithm* (CMA), é um caso particular do algoritmo de Godard⁵ e conhecido como *critério CM*. As boas características de convergência e estabilidade do CMA têm sido exaustivamente investigadas, tornando-o uma das referências em algoritmos para equalização autodidata [Papadias, 1995; Johnson et al., 1998]. A seguir, na Equação (5.7) encontra-se o algoritmo de Godard, em que o erro de Godard apresenta uma forma distinta daquelas dos erros de decisão dirigida e de Sato.

⁵O CMA foi desenvolvido independentemente do trabalho de Godard [Treichler & Agee, 1983].

$$\mathbf{h}(n+1) = \mathbf{h}(n) - \underbrace{\mu \left(y(n) \cdot |y^{p-2}(n)| \cdot (|y(n)|^p - R_p) \right)}_{\varepsilon_{\text{God}}(n)} \mathbf{x}(n) \quad \mu, p > 0 \quad \text{e} \quad R_p = \frac{\mathbb{E}\{a^{2p}\}}{\mathbb{E}\{a^p\}}. \quad (5.7)$$

5.2.3.4 Algoritmo de Benveniste-Goursat

O algoritmo para equalizadores cegos proposto por Benveniste e Goursat [Benveniste & Goursat, 1984]⁶ realiza um chaveamento funcional automático no algoritmo LMS do modo de rastreo, em que o termo de ajuste predominante é associado ao erro de decisão dirigida (DD), para o modo de aquisição, em que o termo de ajuste predominante é associado ao erro de Sato, quando ocorrem mudanças abruptas na característica do canal. Caso contrário, o chaveamento é invertido. A intenção neste caso é buscar uma aceleração da convergência dos algoritmos anteriores. Esta aceleração é alcançada utilizando-se uma função erro que é uma composição ponderada dos erros DD e de Sato. Os erros influem diferentemente nas etapas de adaptação provocando um ganho de velocidade de convergência, conforme é indicado pela Equação (5.8).

$$\mathbf{h}(n+1) = \mathbf{h}(n) - \underbrace{\mu (\kappa_1 \cdot \varepsilon_{\text{DD}}(n) + \kappa_2 \cdot \varepsilon_{\text{Sato}}(n))}_{\varepsilon_{\text{BG}}(n)} \mathbf{x}(n) \quad \mu, \kappa_1, \kappa_2 > 0. \quad (5.8)$$

5.2.3.5 Algoritmo “Stop-and-Go”

O algoritmo Stop-and-Go (SAG) [Picchi & Prati, 1987] também apresenta uma melhoria na velocidade de convergência. Dentro do mesmo espírito de chaveamento do algoritmo Benveniste-Goursat, o SAG utiliza os erros DD e de Sato indiretamente na construção da equação de adaptação, cuja expressão é mostrada na Equação (5.9).

$$\mathbf{h}(n+1) = \mathbf{h}(n) - \underbrace{\mu f \cdot \varepsilon_{\text{DD}}(n)}_{\varepsilon_{\text{SAG}}(n)} \mathbf{x}(n) \quad \mu > 0 \quad (5.9)$$

$$f = \begin{cases} 1, & \text{sgn}(\varepsilon_{\text{DD}}(n)) = \text{sgn}(\varepsilon_{\text{Sato}}(n)) \rightarrow \text{Adapta} \\ 0, & \text{sgn}(\varepsilon_{\text{DD}}(n)) \neq \text{sgn}(\varepsilon_{\text{Sato}}(n)) \rightarrow \text{Não adapta.} \end{cases}$$

Neste caso, o algoritmo SAG utiliza um critério binário baseado na confiabilidade da decisão do sinal de saída para realizar ou não uma adaptação no algoritmo.

Outras variações destes algoritmos também buscam uma aceleração na velocidade de convergência [Glentis et al., 1999].

⁶Estes pesquisadores foram, provavelmente, os primeiros a adotar o termo *cego* para a equalização autodidata.

A Tabela 5.1 mostra o estimador do símbolo desejado para as principais estratégias de Bussgang.

Algoritmo	Estimador $\Gamma(\cdot)$
Decisão Dirigida	$\text{Dec}(y)$
Sato	$\gamma \cdot \text{sgn}(y)$
Godard (p=2)	$y - y \cdot (y - R_2)$
Benveniste-Goursat	$y - (\kappa_1 \cdot \varepsilon_{\text{DD}} + \kappa_2 \cdot \varepsilon_{\text{DD}} \cdot \varepsilon_{\text{Sato}})$
Stop-and-Go	$y - f \cdot (y - \text{Dec}(y))$

Tabela 5.1: Principais estimadores em algoritmos de Bussgang.

5.3 Classificação

As técnicas de classificação de padrões são bastante conhecidas no campo do reconhecimento de formas e identificação. Recentemente, alguns trabalhos têm abordado o problema de equalização como um problema de classificação para recuperação da informação [Mulgrew, 1996; Montalvão et al., 1998].

Como visto anteriormente na abordagem clássica, a equalização consiste na utilização de um critério de estimação da informação transmitida para minimizar a diferença entre a informação e sua estimativa através da adaptação dos parâmetros do filtro.

Na abordagem de classificação, a equalização consiste em encontrar uma superfície no espaço n -dimensional das entradas do equalizador, capaz de separar as classes dos símbolos. Isto é, a superfície deve fornecer uma separação de tal maneira a representar os símbolos do alfabeto da fonte, considerando a projeção das entradas sobre ela. Neste caso, o número de classes será igual ao número de símbolos no alfabeto da fonte de informação.

Uma outra consideração é que, no problema em questão, os sinais no receptor (saídas do canal) são rotulados de acordo com o símbolo de transmissão com um determinado atraso dentro do suporte do canal.

A maneira mais evidente de entendimento destes conceitos é através da abordagem geométrica, como mostrado a seguir.

EXEMPLO 5.1:

Para um equalizador de *apenas uma* entrada, deve-se encontrar um mapeamento unidimensional cuja resposta seja capaz de separar as classes distintas. Inicialmente, será desconsiderado o ruído nas análises.

Considere-se um canal cuja RI é representada por: $\mathbf{f} = [0.2 \ 0.8 \ 0.4]^T$. Para efeito de simplificação visual, será considerada a modulação BPSK.

Deve-se procurar os estados do canal e colocá-los no espaço de entradas do equalizador.

$$\underbrace{\begin{bmatrix} -1 & -1 & -1 \\ -1 & -1 & +1 \\ -1 & +1 & -1 \\ -1 & +1 & +1 \\ +1 & -1 & -1 \\ +1 & -1 & +1 \\ +1 & +1 & -1 \\ +1 & +1 & +1 \end{bmatrix}}_{\mathbf{B}} \underbrace{\begin{bmatrix} 0.2 \\ 0.8 \\ 0.4 \end{bmatrix}}_{\mathbf{f}} = \underbrace{\begin{bmatrix} -1.4 \\ -0.6 \\ +0.2 \\ +1 \\ -1 \\ -0.2 \\ +0.6 \\ +1.4 \end{bmatrix}}_{\bar{\mathbf{x}}} \quad (5.10)$$

em que \mathbf{B} é a matriz com todas as possíveis combinações dos símbolos na entrada do equalizador e $\bar{\mathbf{x}}$ é o vetor com todas as possíveis saídas do canal.

O rótulo de cada um dos elementos do vetor de estados $\bar{\mathbf{x}}$ corresponde ao atraso de decisão δ , escolhido nas colunas de \mathbf{B} , e observando-se os valores de cada uma das linhas associada a coluna de \mathbf{B} referente ao atraso δ . Tomando, por exemplo, $\delta = 0$, tem-se:

- Rótulo -1 (representado por \times): $-1.4, -0.6, +0.2$ e $+1$.
- Rótulo $+1$ (representado por \circ): $-1, -0.2, +0.6$ e $+1.4$.

Assim, pode-se representar na Figura 5.3 as classes rotuladas e a superfície de separação.

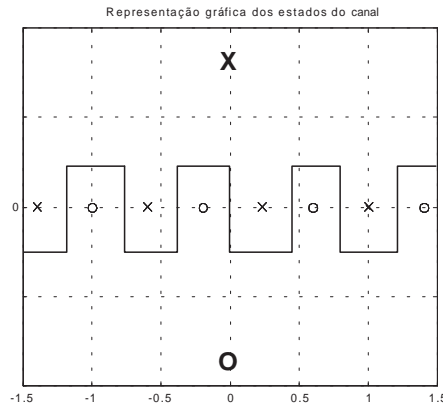


Figura 5.3: Superfície (função) de separação unidimensional.

Observa-se na Figura 5.3 que os estados (\times) estão na região X do plano separados dos estados (\circ), os quais encontram-se na região O , convexa à região X .

Pode-se então verificar que a superfície de separação neste caso é uma função não linear. Sendo as estruturas de equalização em sua grande maioria lineares, como é o caso do filtro

transversal, torna-se impossível a realização da separação, pois a função implementada pelo filtro transversal é linear no espaço de dimensão igual ao número de entradas. No entanto, buscando realizar a separação (ou discriminação) linear, pode-se escolher entre duas opções para utilizar uma estrutura linear

- (a) aumento da ordem de entradas do filtro transversal, o que significa aumentara dimensão do espaço de classificação, procurando desta maneira um espaço n -dimensional em que haja uma projeção hiperplana que separe as classes;
- (b) mudança no atraso, significando mudar a referência dos rótulos favorecendo a uma possível separação linear de classes.

5.3.1 Influência da ordem: opção (a)

O exemplo seguinte ilustra a opção (a).

EXEMPLO 5.2:

Sendo um canal dado cuja RI é: $\mathbf{f} = [0.8 \ 0.6 \ 0.4]^T$, tem-se a seguinte representação dos estados do canal no caso unidimensional:

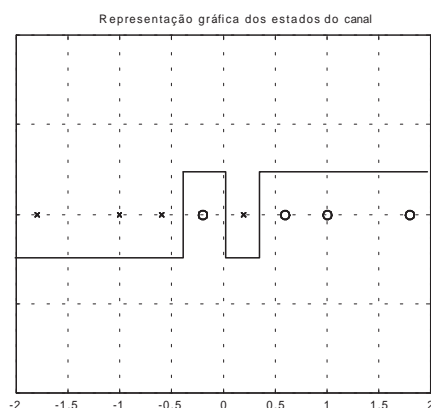


Figura 5.4: Representação unidimensional dos estados e superfície de separação.

Observa-se claramente que uma estrutura linear não é capaz de implementar a fronteira (superfície) de separação. Um aumento da ordem é implementado, no caso de considerar-se agora um equalizador de ordem 2 (duas entradas) e novamente (por motivo de simplicidade de

cálculo) utilizando-se modulação BPSK, tem-se:

$$\begin{bmatrix} \bar{x}_0 \\ \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \\ \bar{x}_4 \\ \bar{x}_5 \\ \bar{x}_6 \\ \bar{x}_7 \end{bmatrix} = \underbrace{\begin{bmatrix} a(n-1) & a(n-2) & a(n-3) \\ \downarrow & \downarrow & \downarrow \\ -1 & -1 & -1 \\ -1 & -1 & +1 \\ -1 & +1 & -1 \\ -1 & +1 & +1 \\ +1 & -1 & -1 \\ +1 & -1 & +1 \\ +1 & +1 & -1 \\ +1 & +1 & +1 \end{bmatrix}}_{\mathbf{B}(n-1)} \begin{bmatrix} 0.8 \\ 0.6 \\ 0.4 \end{bmatrix} = \begin{bmatrix} -1.8 \\ -1 \\ -0.6 \\ +0.2 \\ -0.2 \\ +0.6 \\ +1 \\ +1.8 \end{bmatrix}. \quad (5.11)$$

Então, como considera-se duas dimensões a representação de estados do canal deve ser feita no mapeamento $\bar{x}(n-1) \times \bar{x}(n)$. Isso implica que cada símbolo no instante $n-1$ estará associado a $S^{\Delta n}$ em que S é o número de símbolos no alfabeto da fonte e Δn é o valor da variação temporal da representação, no caso de $n-1$ e n , $\Delta n = n - (n-1) = 1$. Vale notar que Δn é o número de entradas (M) do equalizador.

As relações dos estados nos diferentes instantes de tempo são determinadas por essas variações. No caso da modulação BPSK, essas variações são apenas os símbolos $+1$ ou -1 . O preenchimento do vetor $\mathbf{a}(n)$ ocorre da maneira seguinte:

$$\begin{aligned} \mathbf{a}(n-1) &= [+1 \quad -1 \quad -1]^T, \text{ para o instante seguinte:} \\ \mathbf{a}(n-1) &\rightarrow \mathbf{a}(n) = [+1 \quad \curvearrowright \quad -1 \quad \curvearrowright \quad -1 \quad \curvearrowright]^T \\ \text{resultando em } \mathbf{a}(n) &= [a_n \quad +1 \quad -1]^T. \end{aligned}$$

em que a_n é o símbolo ocorrido no instante n .

Assim, as possibilidades para a matriz \mathbf{B} no instante n gerando os respectivos estados são:

$$\underbrace{\begin{bmatrix} a(n) & a(n-1) & a(n-2) \\ -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & +1 \\ -1 & -1 & +1 \\ -1 & +1 & -1 \\ -1 & +1 & -1 \\ -1 & +1 & +1 \\ -1 & +1 & +1 \end{bmatrix}}_{\mathbf{B}(n)} \mathbf{f} = \underbrace{\begin{bmatrix} \bar{x}_0 \\ \bar{x}_0 \\ \bar{x}_1 \\ \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_2 \\ \bar{x}_3 \\ \bar{x}_3 \end{bmatrix}}_{\bar{\mathbf{x}}(n)} \quad \text{e} \quad \underbrace{\begin{bmatrix} a(n) & a(n-1) & a(n-2) \\ +1 & -1 & -1 \\ +1 & -1 & -1 \\ +1 & -1 & +1 \\ +1 & -1 & +1 \\ +1 & +1 & -1 \\ +1 & +1 & -1 \\ +1 & +1 & +1 \\ +1 & +1 & +1 \end{bmatrix}}_{\mathbf{B}(n)} \mathbf{f} = \underbrace{\begin{bmatrix} \bar{x}_4 \\ \bar{x}_4 \\ \bar{x}_5 \\ \bar{x}_5 \\ \bar{x}_6 \\ \bar{x}_6 \\ \bar{x}_7 \\ \bar{x}_7 \end{bmatrix}}_{\bar{\mathbf{x}}(n)}. \quad (5.12)$$

Logo, tem-se:

$$\underbrace{\begin{bmatrix} \bar{x}_0 \\ \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \\ \bar{x}_4 \\ \bar{x}_5 \\ \bar{x}_6 \\ \bar{x}_7 \end{bmatrix}}_{\bar{x}(n-1)} \xrightarrow{a(n)=-1} \underbrace{\begin{bmatrix} \bar{x}_0 \\ \bar{x}_1 \\ \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_2 \\ \bar{x}_3 \\ \bar{x}_3 \end{bmatrix}}_{\bar{x}(n)} \quad \text{e} \quad \underbrace{\begin{bmatrix} \bar{x}_0 \\ \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \\ \bar{x}_4 \\ \bar{x}_5 \\ \bar{x}_6 \\ \bar{x}_7 \end{bmatrix}}_{\bar{x}(n-1)} \xrightarrow{a(n)=+1} \underbrace{\begin{bmatrix} \bar{x}_4 \\ \bar{x}_4 \\ \bar{x}_5 \\ \bar{x}_5 \\ \bar{x}_6 \\ \bar{x}_6 \\ \bar{x}_7 \\ \bar{x}_7 \end{bmatrix}}_{\bar{x}(n)}. \quad (5.13)$$

Dessa maneira, representando-se os estados do canal no espaço bidimensional, obtém-se a representação da Figura 5.5.

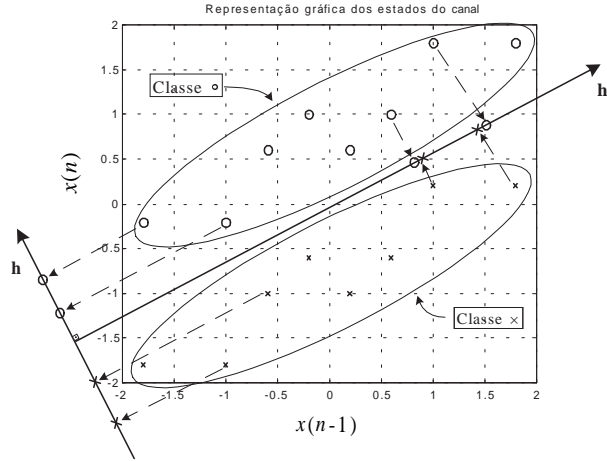


Figura 5.5: Representação bidimensional dos estados, superfícies de separação e projeções.

Observe, que nesse caso uma estrutura linear é suficiente para prover a separação das classes no espaço de projeção \mathbf{h} formado pelo vetor de coeficientes do equalizador, ao contrário da projeção \mathbf{h}' que não possibilita uma separação linear. É conhecido da literatura que para o caso sem ruído, a solução de Wiener e a solução do critério ZF correspondem àquela obtida pelo discriminante linear de Fisher, ou seja, a superfície linear de separação [Montalvão, 2000].

É importante ressaltar que o aumento da ordem de classificação implica também no aumento da potência de ruído, o que pode vir a prejudicar o desempenho do sistema. Com isso, a opção de aumentar a ordem do espaço de classificação nem sempre promove um ganho no processamento.

Finalmente, é sabido que o critério de discriminação linear, exemplificado acima, corresponde ao critério de Wiener [Montalvão et al., 1998; Montalvão, 2000].

A Tabela 5.2 resume algumas correspondências entre os parâmetros típicos de equalização linear e a classificação.

Equalização	Classificação
Ordem do equalizador	Espaço de classificação
Atraso de decisão	Classes de estados (saídas do canal)
Saída do equalizador	Projeção ortogonal dos estados
RI do equalizador	Vetor de projeção

Tabela 5.2: Equivalências entre alguns termos de equalização e classificação.

5.3.2 Influência do atraso: opção (b)

Nas técnicas de filtragem, uma modificação do atraso corresponde a uma modificação no instante de amostragem, para buscar um instante de tempo no qual o *olho esteja aberto* ou o menos fechado possível.

Na abordagem de classificação, o atraso corresponde a fixação do rótulo das classes de acordo com os símbolos transmitidos no instante correspondente de tempo. De tal maneira que a representação das saídas do canal sofre variações no seu agrupamento para permitir uma eventual separação linear.

Assim uma escolha do atraso δ mais adequado pode, sem alteração da ordem do filtro, favorecer a uma separação linear, ou seja, à possibilidade de equalização através do uso de estruturas lineares.

A seguir, é mostrado num exemplo que ilustra a opção (b) da separação linear.

EXEMPLO 5.3:

Considere um canal de fase não mínima de RI dada por $\mathbf{f} = [0.35 \quad 1 \quad 0.45]^T$, a modulação BPSK e um ruído gaussiano de variância $\sigma_b^2 = 0.05$.

A Figura 5.6(a) mostra a representação dos estados no plano bidimensional para $\delta = 0$ e a Figura 5.6(b) mostra a mesma representação para $\delta = 1$.

Neste caso, é fácil visualizar que para o atraso igual a 1, é possível encontrar um equalizador linear de ordem 2 que separa as classes.

Pode-se observar então que, como esperado, o atraso de decisão exerce um papel fundamental na abordagem de recuperação da informação por classificação.

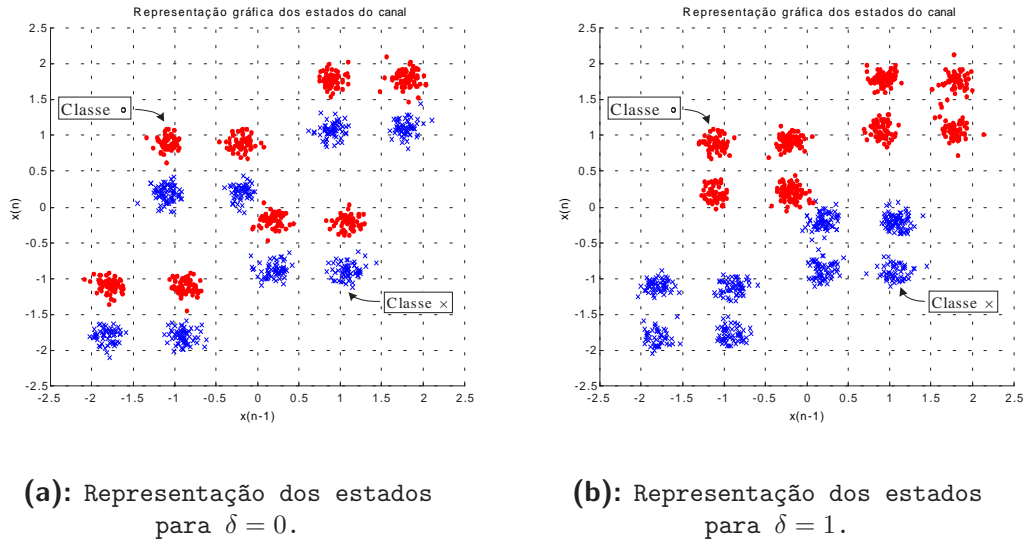


Figura 5.6: Representação dos estados ruidosos com variação do atraso de decisão.

5.4 Síntese e conclusões

Este capítulo descreve a abordagem da equalização através da classificação. Como mencionado no início, este texto não tem preocupação com a formalização matemática entre a filtragem linear e análise discriminante, e provavelmente alguns conceitos necessitem de um maior aprofundamento para completa compreensão. Vale também ressaltar que a necessidade da inclusão do capítulo no trabalho deve-se à questão de normalização da linguagem a ser utilizada.

As técnicas clássicas de equalização descritas nas seções 5.1 e 5.2 são discutidas brevemente apenas para explicitar um pouco a idéia das estratégias existentes e que têm desenvolvido-se bastante nos últimos anos. Entretanto, o critério ZF será utilizado para justificar uma das proposições dessa dissertação no capítulo 7.

Os conceitos de classificação são descritos baseados em sua compreensão geométrica. Este é o motivo principal por ser mantida a utilização de exemplos uni e bidimensionais para efeitos de compreensão e visualização. Este será o mesmo apelo quando da utilização dos conceitos de classificação nos capítulos seguintes.

Uma sugestão bibliográfica completa e aprofundada nos aspectos matemáticos sobre as relações entre equalização e análise discriminante linear pode ser encontrada em [Montalvão, 2000].

PARTE **II**

PROPOSIÇÃO

CAPÍTULO 6

PREDIÇÃO NÃO LINEAR PARA EQUALIZAÇÃO AUTODIDATA

TÉCNICAS preditivas para equalização vêm sendo bastante estudadas desde a metade da década de 80. A partir do trabalho de Macchi e Hachicha (1986), várias investigações em torno das estruturas preditivas foram desenvolvidas.

A maioria das estratégias preditivas utilizam um *processo de branqueamento* do sinal na saída do preditor como critério para a adaptação do filtro. Isto é possível devido ao fato de que os símbolos transmitidos formam uma seqüência de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) e conseqüentemente branca.

A proposta deste capítulo da dissertação, neste capítulo, é a utilização de uma estrutura preditiva não linear baseada em uma RNA. Para isso, a predição é estudada através da abordagem da classificação. Uma estratégia de aprendizado híbrido (auto-organizado + supervisionado) é empregada para permitir tal abordagem, resultando também na proposição de um novo algoritmo auto-organizado para o aprendizado da estrutura proposta.

O restante deste capítulo é organizado da seguinte maneira: a seção 6.1 aborda os conceitos da predição linear revisitando alguns aspectos da teoria; a seção 6.2 descreve algumas das estratégias preditivas que utilizam estruturas não lineares; na seção 6.3 é ilustrada a abordagem do equalizador realimentado por decisão (DFE¹ como uma estrutura preditiva não linear simples, servindo como comparação para a proposição; na seção 6.4 é discutido, segundo a abordagem da classificação, o problema da predição e suas características particulares; na seção 6.5 é descrita a proposta, descrevendo os aspectos

¹Do inglês “Decision Feedback Equalize”.

teóricos bem como os algoritmos desenvolvidos e aspectos de limitação e complexidade; a seção 6.6 é dedicada às simulações computacionais que verificam o desempenho da proposição através de comparações com outras estruturas de equalização; finalmente na seção 6.7 são apresentadas as conclusões pertinentes ao capítulo.

6.1 Predição linear

É conhecido que a minimização do EQM de predição entre um sinal ($x(n)$) e suas amostras passadas ($\mathbf{x}(n-1)$) corresponde em aplicar o *princípio da ortogonalidade* entre o erro de estimação e as amostras passadas. Com isso o erro de predição tende a ser um sinal descorrelacionado [Papoulis, 1991; Rocha, 1996]. Como, no caso dos SCD, os símbolos do alfabeto transmitido \mathcal{A} são descorrelacionados e nenhuma memória é inserida no sistema exceto aquela produzida pelo canal de comunicação. Assim a seqüência descorrelacionada do erro de predição deve ser igual àquela transmitida, a menos de um fator de ganho e de uma rotação de fase. Esse processo também corresponde a um *processo de branqueamento* do sinal de entrada do preditor.

Pode-se representar o erro de predição no instante n como:

$$e_p(n) = x(n) - \underbrace{T(\mathbf{x}(n-1))}_{\hat{x}(n)}, \quad (6.1)$$

em que $T(\cdot)$ é uma transformação sobre o vetor de dados $\mathbf{x}(n-1)$, o índice p indica que a predição é *direta*² e o símbolo $\hat{x}(n)$ é a estimação do sinal $x(n)$.

Na utilização de técnicas preditivas para equalização é muito comum o uso de estruturas lineares, tais como o filtro transversal, e então nestes casos a transformação $T(\cdot)$ será uma transformação linear. A Figura 6.1 ilustra uma estrutura de predição.

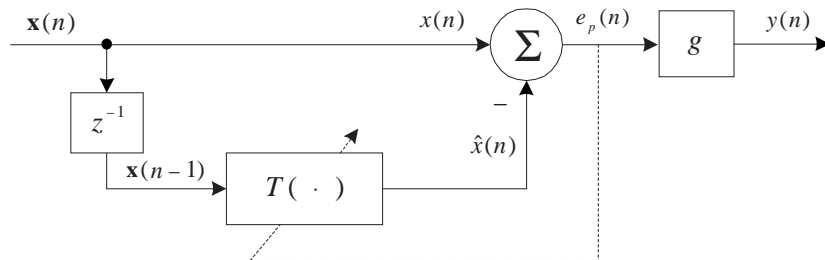


Figura 6.1: Diagrama de uma estrutura de predição.

Na Figura 6.1, g é um controle automático de ganho (CAG), responsável por colocar a potência do sinal de saída $y(n)$ no mesmo nível de potência dos símbolos transmitidos

²Nesta dissertação os termos predição direta e predição reversa serão substituídos por *predição* e *retropredição*, respectivamente.

[Rocha & Macchi, 1994; Macchi, 1995]. Observa-se ainda que a estrutura preditiva favorece a adoção da equalização cega, pois não há dependência com os sinais não-observáveis.

No caso da retropredição, o erro é dado por:

$$e_r(n) = x(n - M) - \underbrace{T(\mathbf{x}(n))}_{\hat{x}(n-M)}, \quad (6.2)$$

em que M é o tamanho do vetor de dados $\mathbf{x}(n)$.

Em ambos os casos citados anteriormente, ao utilizar-se estruturas lineares (filtros transversais) para equalização, o erro de predição (retropredição) tende a ser descorrelacionado a medida que aumenta-se a ordem do filtro. É fato também conhecido que os filtros de erro de predição e retropredição somente são capazes de realizar a equalização cega de canais de fase mínima ou fase máxima, através da minimização de estatísticas de 2ª ordem ($\mathbb{E}\{|e_p(n)|^2\}$ ou $\mathbb{E}\{|e_r(n)|^2\}$), respectivamente [Scharf, 1991; Haykin, 1996]. Uma maneira de justificar esta propriedade é dada a seguir.

6.1.1 Desenvolvimento matemático

O desenvolvimento a seguir utiliza o modelo de predição, esquematizado na Figura 6.1. O desenvolvimento para o caso de retropredição é análogo.

Para um canal de RI dada por

$$\mathbf{f} = [f_0 \ f_1 \ \cdots \ f_{N-1}]^T,$$

e a seqüência de dados transmitidos, representada na forma vetorial

$$\mathbf{a}(n) = [a(n) \ a(n-1) \ \cdots \ a(n-N+1)]^T.$$

tem-se que as saídas do canal podem ser escritas

$$\begin{aligned} x(n) &= a(n)f_0 + \cdots + a(n-N+1)f_{N-1} + b(n) \\ x(n-1) &= a(n-1)f_0 + \cdots + a(n-N+2)f_{N-1} + b(n-1) \\ x(n-2) &= a(n-2)f_0 + \cdots + a(n-N+3)f_{N-1} + b(n-2) \\ &\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \\ x(n-M) &= a(n-M)f_0 + \cdots + a(n-M-N+1)f_{N-1} + b(n-M), \end{aligned} \quad (6.3)$$

em que M é o comprimento da RI do filtro.

Tomando o filtro que implementa o preditor como tendo uma RI dada por $\mathbf{p} = [p_1 \ p_2 \ \cdots \ p_M]^T$, tem-se que a estimativa do sinal $x(n)$ é dada por:

$$\hat{x}(n) = \sum_{i=1}^M p_i \cdot x(n-i), \quad (6.4)$$

Então, expandindo-se a Equação (6.1) utilizando as Equações (6.3) e (6.4), obtém-se

$$e_p(n) = \underbrace{a(n)f_0 + a(n-1)f_1 + \cdots + a(n-N+1)f_{N-1} + b(n)}_{x(n)} - \underbrace{x(n-1)p_1 + x(n-2)p_2 + \cdots + x(n-M)p_M}_{\hat{x}(n)}. \quad (6.5)$$

Por motivos de simplificação, utiliza-se o caso particular em que o canal e o filtro linear são considerados com memória 2. Reescrevendo a Equação (6.5) utilizando a Equação (6.3) e o fato de $(N+1) = M = 2$

$$\begin{aligned} e_p(n) &= \underbrace{a(n)f_0 + a(n-1)f_1 + a(n-2)f_2 + b(n)}_{x(n)} \\ &- \underbrace{[a(n-1)f_0p_1 + a(n-2)f_1p_1 + a(n-3)f_2p_1 + b(n-1)p_1]}_{x(n-1)p_1} \\ &- \underbrace{[a(n-2)f_0p_2 + a(n-3)f_1p_2 + a(n-4)f_2p_2 + b(n-2)p_1]}_{x(n-2)p_2}. \end{aligned} \quad (6.6)$$

Como a informação é transmitida através dos símbolos, a saída do filtro $\hat{x}(n)$ possui quase toda a informação em $x(n)$, exceto a informação veiculada por $a(n)$. Então, reduzir a redundância entre $x(n)$ e suas amostras passadas (minimização do erro de estimação) corresponde a extrair $a(n)$ de $x(n)$. Assim, reescrevendo a Equação (6.6) agrupando os termos com as mesmas variáveis dependentes tem-se

$$\begin{aligned} e_p(n) &= a(n)f_0 + a(n-1)[f_1 - f_0p_1] + a(n-2)[f_2 - f_1p_1 - f_0p_2] \\ &- a(n-3)[f_2p_1 + f_1p_2] - a(n-4)f_2p_2 + b(n) + b(n-1)p_1 + b(n-2)p_2. \end{aligned} \quad (6.7)$$

Obviamente, para recuperar o símbolo $a(n)$, que corresponde à *inovação* de $x(n)$ com relação às amostras passadas, as constantes que multiplicam os outros símbolos devem ser *zero*. Por exemplo, isso corresponde à escolher \mathbf{p} , tal que

$$\begin{aligned} p_1 &= \frac{f_1}{f_0} \quad \text{e} \\ p_2 &= \frac{f_2}{f_0} - \left(\frac{f_1}{f_0}\right)^2. \end{aligned}$$

A atenuação provocada em $a(n)$ por f_0 é compensada pelo CAG.

Entretanto, não é possível anular todos os termos simultaneamente, então pode-se notar que na Equação (6.7), alguns termos permanecem não nulos, um **resíduo**, $(a(n-3)[\cdot] - a(n-4)[\cdot] + b(n) + b(n-1)p_1 + b(n-2)p_2)$. Para garantir que o erro de predição seja uma sequência descorrelacionada, este resíduo deve ser de valor desprezível e é necessário que o coeficiente f_0 seja o mais importante na RI do canal, caracterizando

assim um canal de fase mínima. Para conseguir tal efeito, ou seja, o resíduo ser mínimo, deve-se aumentar a ordem do filtro tentando colocar a dependência temporal para os instantes mais distantes.

6.1.2 Influência do ruído e atraso

Uma questão importante é que a melhor predição fornece uma solução para a Equação (6.7) do tipo

$$e_p(n) = a(n)f_0 + b(n). \quad (6.8)$$

Logo, se o ruído for bastante forte, a estimação não é realizada adequadamente.

Uma outra questão é que, caso o canal seja de FNM, de um modo geral a componente mais importante de sua RI não é a primeira ou última, assim, o uso de estruturas simples (apenas preditiva ou retropreditiva), utilizando estatísticas de ordem 2 não é suficiente.

Como já mencionado no capítulo 2, é possível separar um canal de FNM em suas partes FMín e FMáx, e com isso utilizar-se estruturas combinadas (retropreditores + preditores) para a equalização.

A combinação das partes de FMín e FMáx corresponde também a mudança no *atraso de decisão*, pois no caso do preditor o atraso é sempre zero ($\delta = 0$) e no retropreditor é sempre igual à M ($\delta = M$). O uso de estruturas combinadas seguramente modifica o instante de decisão para favorecer o coeficiente mais importante da RI do canal. No entanto, a adoção de um critério de equalização cega baseado nas estatísticas de 2ª ordem não permite ter sucesso na equalização de canais FNM [Mota, 1992].

6.2 Preditores não lineares

Outra estratégia para a equalização cega de canais de FNM através de técnicas preditivas é baseada na utilização de estruturas não lineares ou ainda em critérios que usem MOS.

Várias estratégias de equalização podem ser listadas nesta seção, mas serão descritas apenas duas:

- a equalização de magnitude e fase proposta em [Rocha et al., 1995], que serviu de base para os equalizadores propostos em [Labat et al., 1998; Labat & Laot, 1999].
- a equalização que utiliza o critério PCM³, proposto em [Cavalcanti, 1996].

Ambas as estratégias apresentam a semelhança de fazer uso de estruturas combinadas para equalização.

³Do inglês “Predictive Constant Modulus”.

6.2.1 Equalizador de magnitude e fase

O equalizador de magnitude e fase apresenta a estrutura de filtragem em cascata entre um retropreditor e um preditor recursivo, como exemplificado na Figura 6.2, na qual A e B são filtros transversais e g é um CAG.

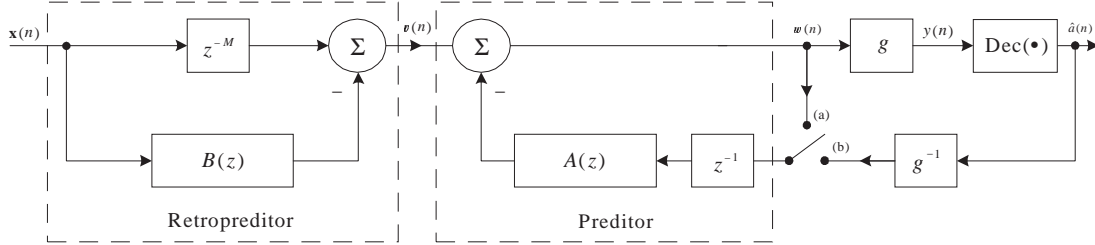


Figura 6.2: Equalizador de magnitude e fase.

A estrutura apresenta, além do preditor e do retropreditor, um CAG e uma chave de posição para um período inicial, posição (a) e em seguida para a posição (b) que corresponde a uma estrutura realimentada por decisão, o DFE⁴.

Para equalização, é necessário que as seqüências $\mathbf{w}(n)$ e $\mathbf{y}(n)$ sejam super-brancas, tal como $\mathbf{a}(n)$ [Mota, 1992]. De um modo geral, não há garantia que a seqüência $\mathbf{v}(n)$ seja branca e $\mathbb{E}\{|v(n)|^2\}$ não é necessariamente mínima, nem mesmo a quantidade

$$\mathcal{M}(A, B) = \mathbb{E}\{|w(n)|^2\}. \quad (6.9)$$

Sabe-se que a minimização da Equação (6.9) leva à equalização para os casos em que o canal é somente de FMín ou FMáx. Entretanto, para os canais de FNM a minimização da potência, Equação (6.9) e não realiza a equalização, é necessário então a adoção de um critério, tal como mostrado a seguir, que permita uma minimização global (qualquer tipo de canal). Neste caso, um funcional multi-objetivo considerando o erro de decisão (ε_{DD}), é então dado por

$$J_\rho(A, B) = \rho g^2 \mathcal{M}(A, B) + (1 - \rho) \varepsilon_{DD}(A, B), \quad (6.10)$$

em que $\rho \in [0, 1]$ e o fator g^2 assegura a homogeneidade [Rocha, 1996].

Dentro do processo de minimização algorítmica e durante a convergência, o fator ρ é controlado para decrescer de 1, contemplando o critério $\mathcal{M}(A, B)$ à 0, quando o critério ε_{DD} é predominante, através da regra

$$\rho = (\tanh |\varepsilon_{DD}|). \quad (6.11)$$

O momento de chaveamento da posição (a) para a posição (b) é determinado pelo instante em que o *olho está aberto*, situação que garante uma convergência para a estrutura de realimentação de decisão.

⁴Do inglês “Decision Feedback Equalizer”.

6.2.2 Critério “Predictive Constant Modulus” (PCM)

A estratégia proposta em [Cavalcanti, 1996] utiliza uma estrutura de filtragem ARMA⁵ composta de uma cascata de um filtro de erro de retropredição (BPEF⁶) e um filtro de erro de predição (FPEF⁷) e do critério do módulo constante (CM⁸). A Figura 6.3 ilustra a estrutura de filtragem utilizada.

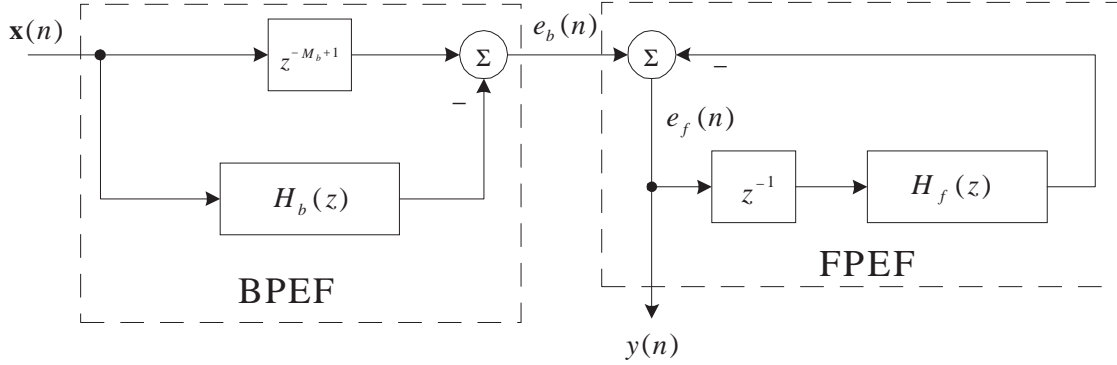


Figura 6.3: Estrutura ARMA composta de BPEF e FPEF.

Ao invés da minimização da potência do erro de predição é utilizado o critério CM, em que o funcional é $J(\mathbf{h}_f, \mathbf{h}_b) = E\{|y(n)|^2 - R_2\}^2$, resultando no algoritmo de adaptação do tipo LMS

$$\mathbf{h}_b(n+1) = \mathbf{h}_b(n) + \mu_b y(n) (R_2 - |y(n)|^2) \mathbf{x}^*(n) \quad (6.12a)$$

$$\mathbf{h}_f(n+1) = \mathbf{h}_f(n) + \mu_f y(n) (R_2 - |y(n)|^2) \mathbf{e}_f^*(n-1), \quad (6.12b)$$

em que \mathbf{h}_b e \mathbf{h}_f são os vetores dos coeficientes do BPEF e do FPEF, respectivamente, $\mathbf{e}_f(n-1) = [e_f(n-1) \cdots e_f(n-M_f)]^T$, M_b é o comprimento da RI do BPEF, M_f é comprimento da RI do FPEF e μ_b e μ_f são os passos de convergência para o BPEF e FPEF, respectivamente.

Esta estratégia marca uma grande diferença entre as demais estratégias preditivas por não realizar um processo de branqueamento e sim uma penalização das saídas fora do raio de equalização R_2 .

A diferença principal entre estas duas estratégias preditivas reside no tipo de não linearidade incorporada aos funcionais. O desempenho de ambas é equiparável em simulações em que predomina a parte FMín num canal FNM.

⁵Do inglês “Auto Regressive Moving Average”.

⁶Do inglês “Backward Prediction Error Filter”.

⁷Do inglês “Forward Prediction Error Filter”.

⁸Do inglês “Constant Modulus”

6.3 DFE como preditor não linear

Sob um enfoque preditivo, pode-se entender o DFE como sendo um preditor não linear. Por utilizar uma estrutura de realimentação, o DFE guarda as características de um preditor, como mostrado na Figura 6.4. É válido observar que o DFE corresponde ao equalizador de magnitude e fase quando este encontra-se com a chave na posição (a) e o filtro $H(z)$, mostrado na Figura 6.4, é um FPEF.

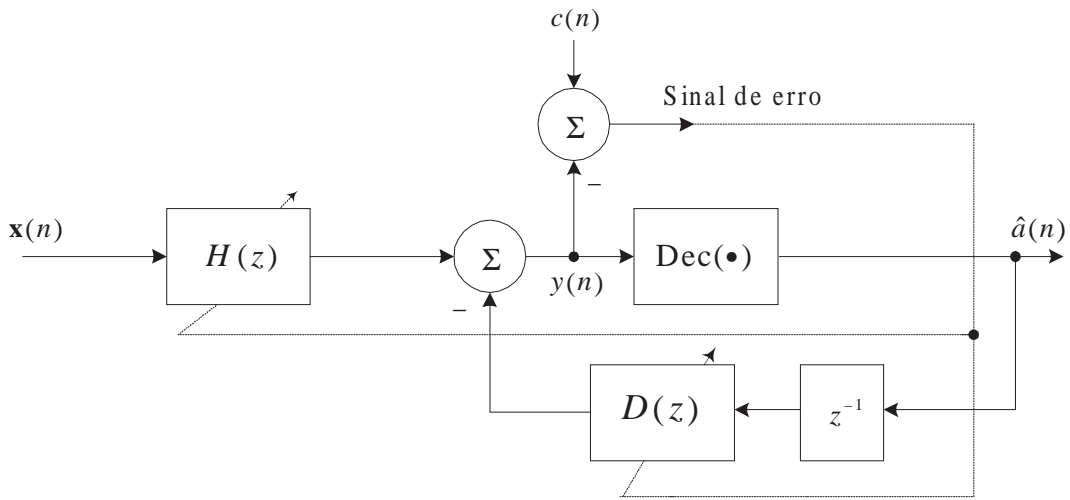


Figura 6.4: Diagrama do DFE.

Na realidade, o DFE não é exatamente uma estrutura de predição no sentido de *inovação* como o preditor da Figura 6.1, mas ele é uma estrutura de predição da IES residual no sinal de saída do filtro do ramo direto como mostrado a seguir.

Considere um canal cuja FT é dada por $F(z) = f_0 + f_1 z^{-1} + f_2 z^{-2}$, tomando-se o DFE como possuindo apenas duas entradas no ramo reverso e uma no ramo direto, tem-se na saída do equalizador, desconsiderando o ruído

$$y(n) = a(n)f_0h_0 + a(n-1)f_1h_0 + a(n-2)f_2h_0 - \hat{a}(n-1)d_1 - \hat{a}(n-2)d_2, \quad (6.13)$$

em que $\mathbf{h} = [h_0 \ h_1 \ \dots \ h_{M_d-1}]^T$ é a RI do filtro do ramo direto, $\mathbf{d} = [d_1 \ d_2 \ \dots \ d_{M_r}]^T$ é a RI do filtro do ramo reverso na sua forma vetorial, M_d é o comprimento da RI do filtro no ramo direto e M_r é o comprimento da RI do filtro no ramo reverso. Então, se as decisões forem corretas, ou seja, $\hat{a}(n-1) = a(n-1)$ e $\hat{a}(n-2) = a(n-2)$ tem-se:

$$y(n) = a(n)f_0h_0 + a(n-1)[f_1h_0 - d_1] + a(n-2)[f_2h_0 - d_2]. \quad (6.14)$$

Logo, para recuperar-se o sinal $a(n)$ deve-se ter:

$$h_0 = \frac{1}{f_0}, \quad d_1 = \frac{f_1}{f_0} \quad \text{e} \quad d_2 = \frac{f_2}{f_0}. \quad (6.15)$$

O que significa que o ramo reverso executa o papel do preditor e o ramo direto funciona como o CAG.

6.4 Predição × Interpolação

Ao analisar-se a equalização do ponto de vista da classificação, vê-se que a predição se enquadra no contexto como uma técnica de *interpolação*.

Isto porque um filtro de erro de predição busca implementar uma superfície (função) de separação entre os símbolos de classes diferentes, cuja diferença entre o símbolo considerado para estimação ($x(n)$ ou $x(n - M)$) e a superfície de interpolação seja igual ao símbolo transmitido possivelmente atenuado por uma constante. A Figura 6.5 ilustra os casos para uma predição utilizando uma estrutura linear, Figura 6.5(a), e utilizando uma estrutura não linear, Figura 6.5(b). Por simplicidade de visualização, utiliza-se a predição no espaço de classificação de duas dimensões, ou seja, tenta-se estimar $x(n)$ a partir de $x(n - 1)$. Também será desconsiderado o ruído.

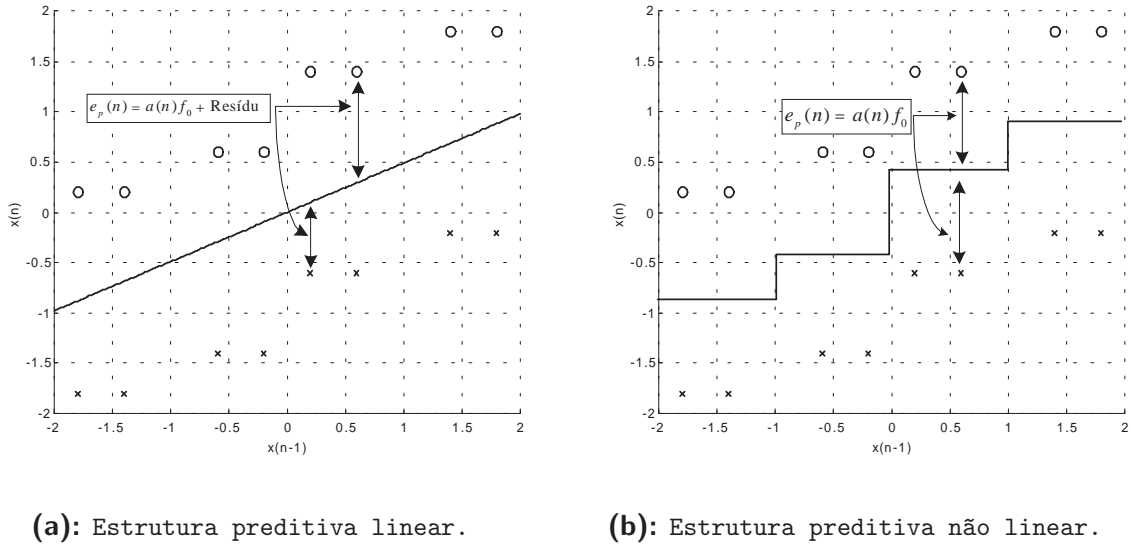


Figura 6.5: Predição vista como um problema de interpolação - exemplo de um caso bidimensional.

É possível verificar através da Figura 6.5 que o *mapeamento* realizado pela função não linear é melhor, pois este consegue uma anulação do resíduo existente no mapeamento linear. Além disso, existe ainda a limitação da necessidade de aumento na ordem do espaço de classificação, para possibilitar a equalização de canais com forte IES e o fato de que uma estrutura linear não soluciona os casos de canais de FNM, como já mencionado anteriormente, para critérios baseados na potência.

Por sua vez estruturas não lineares permitem a construção de uma função de

mapeamento sem aumentar a ordem do espaço de classificação, ou seja, a ordem do equalizador, conforme mostrado a seguir [Cavalcante et al., 2000].

Lema: *Uma função (superfície) de interpolação não linear para equalização baseada em estruturas preditivas é possível de ser implementada utilizando-se apenas uma entrada.*

Prova:

Seja o erro de predição dado por:

$$e_p(n) = x(n) - \hat{x}(n). \quad (6.16)$$

No caso de uma estrutura não linear, a estimativa do sinal $x(n)$ será dada por uma função de mapeamento:

$$\hat{x}(n) = \Psi(\mathbf{x}(n-1)). \quad (6.17)$$

Expandindo-se a Equação (6.16) utilizando-se as Equações (6.17) e (6.3), tem-se

$$\begin{aligned} e_p(n) = & \underbrace{a(n)f_0 + a(n-1)f_1 + \cdots + a(n-N+1)f_{N-1} + b(n)}_{x(n)} \\ & - \underbrace{\Psi(x(n-1), x(n-2), x(n-3), \cdots)}_{\hat{x}(n)}. \end{aligned} \quad (6.18)$$

Entretanto, ao observar-se a Equação (6.3) verifica-se que *toda* a informação sobre a seqüência de dados transmitida contida em $x(n)$ *exceto* a inovação, é disponível em $x(n-1)$. Desta maneira a função da estimativa $\hat{x}(n)$ pode ser escrita em função apenas da entrada do instante anterior $x(n-1)$. Formalizando, para retirar a inovação de $x(n)$, a função de mapeamento deve fornecer

$$\begin{aligned} & \Psi(x(n-1), x(n-2), \cdots) \\ & = a(n-1)f_1 + a(n-2)f_2 + \cdots + a(n-N+1)f_{N-1} + b(n), \end{aligned} \quad (6.19)$$

em que se considera $E\{a(n-i)b(n)\} = 0$, $i = 1, \cdots, N-1$, e como todas as constantes f_i e variáveis necessárias estão presentes em $x(n-1)$ com exceção de $b(n)$, é suficiente uma *dependência somente* em $x(n-1)$, ou seja

$$\begin{aligned} & \Psi\left(\underbrace{a(n-1)f_0 + a(n-2)f_1 + \cdots + a(n-N+2)f_{N-1} + b(n-1)}_{x(n-1)}\right) \\ & = a(n-1)f_1 + a(n-2)f_2 + \cdots + a(n-N+1)f_{N-1} + b(n), \end{aligned} \quad (6.20)$$

e então o erro de predição será dado por

$$e_p(n) = x(n) - \Psi(x(n-1)). \quad (6.21)$$



Obviamente, uma outra hipótese está sendo considerada que é o caso de não ocorrência da superposição de estados, pois neste caso seria necessário a inserção de redundância para extração da informação, ou seja, o símbolo $a(n)$.

É importante notar que não se consegue suprimir o termo $b(n)$, visto que o mesmo é originário de um processo aleatório independente das observações passadas. Desta maneira a melhor resposta da predição para o caso não linear é a mesma que para o caso linear (dada pela Equação (6.8)), ou seja:

$$e_p(n) = a(n)f_0 + b(n), \quad (6.22)$$

que novamente remete à influência do ruído no processamento preditivo do sinal.

A técnica proposta na seção seguinte baseia-se sobre esta possibilidade do tratamento preditivo do sinal utilizando-se uma memória igual a 1.

6.5 Preditor neural

Dado o fato que a predição pode ser realizada através da utilização de uma única entrada (memória 1), é possível realizar a equalização de canais de FNM, como mostrado na Figura 6.6, com atraso $\delta = 0$.

Então uma função capaz de realizar a separação das classes pode ser realizada por uma rede neural e escrita como:

$$\Psi_{RN}(x, \boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_i \beta_i \cdot \text{sgn}(x - \theta_i), \quad (6.23)$$

em que $\boldsymbol{\theta} = [\theta_1 \ \theta_2 \ \cdots \ \theta_M]^T$ é o vetor de bias dos neurônios, $\boldsymbol{\beta} = [\beta_1 \ \beta_2 \ \cdots \ \beta_M]^T$ é o vetor de coeficientes de saída de cada neurônio e o subscrito RN indica *rede neural*.

A estrutura que realiza a função descrita na Equação (6.23), será denominada por “**Neural Predictive Structure**” (NPS) e encontra-se esquematizada na Figura 6.7.

A NPS é um caso particular de uma rede MLP, onde a 1ª camada apresenta apenas uma entrada e a saída é um somador.

O modelo do neurônio que será utilizado é o perceptron de Rosenblatt, onde a função de ativação é dada por $\varphi_{NPS}(\cdot) = \text{sgn}(\cdot)$. Nesta estrutura, trabalha-se com os neurônios

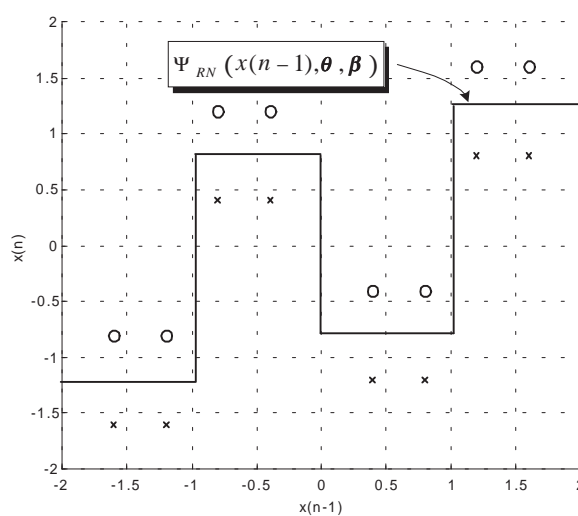


Figura 6.6: Superfície de interpolação não linear para equalização de um canal de FNM.

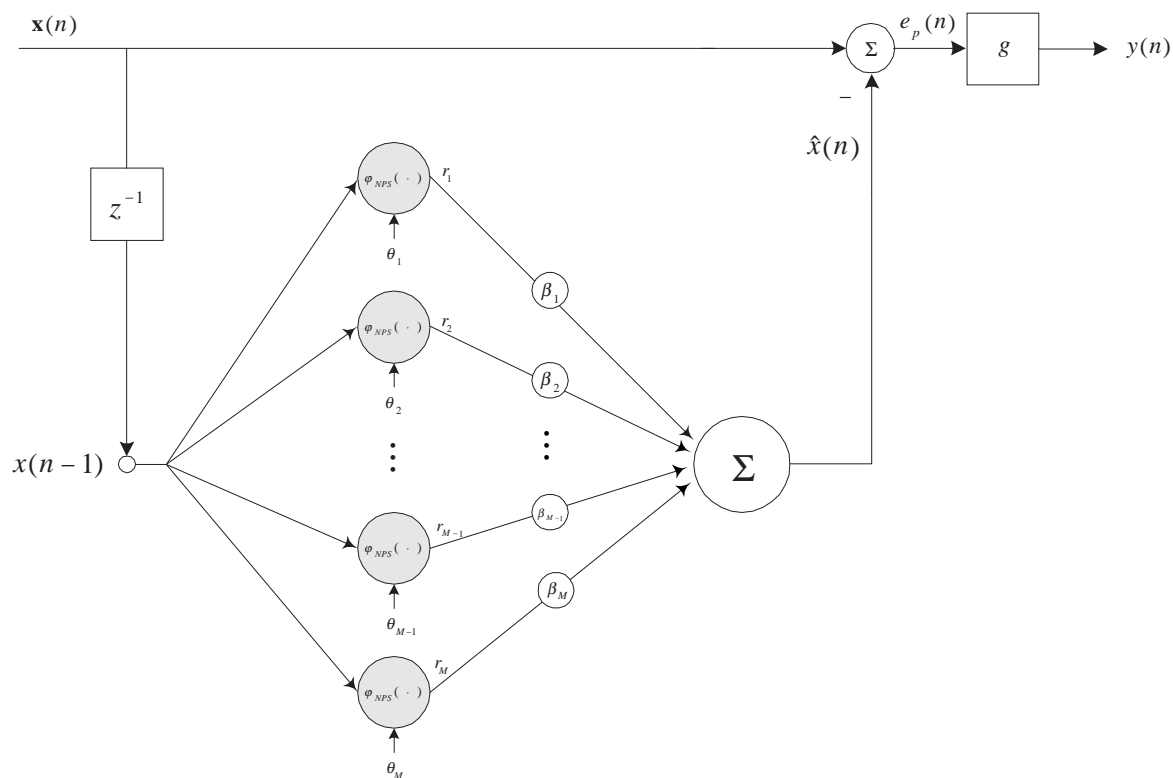


Figura 6.7: Esquema representando a arquitetura da “Neural Predictor Structure” (NPS).

ditos saturados, de tal maneira que a camada de entrada não apresenta pesos sinápticos ajustáveis, ou seja os pesos da camada de entrada são iguais a 1.

Uma estrutura neuronal convencional, ou seja, uma rede MLP com três camadas em que a primeira camada com várias entradas e um certo número de neurônios na camada escondida e sendo a saída um somador das saídas da camada dos neurônios da camada escondida ponderadas pelos pesos sinápticos da camada de saída foi utilizada em [Cavalcante et al., 1999], mas naquele caso apenas canais com IES fraca (e que respeitavam a condição de olho aberto) foram considerados, utilizando o BPA como algoritmo de adaptação dos parâmetros da RNA.

Uma verificação das características do problema levou à proposição de um novo algoritmo de adaptação, descrito na seção seguinte.

6.5.1 Aprendizado da NPS

O primeiro passo para elaboração de um novo algoritmo para adaptação foi a determinação da fdp do sinal recebido (saída do canal).

Teorema: *Dado um canal linear de comprimento finito N , e considerando-se o ruído distribuído no receptor como sendo gaussiano de média nula e variância σ_b^2 , então o sinal no receptor apresenta uma fdp formada por uma composição de gaussianas centradas nos estados do canal e cada uma com variância σ_b^2 .*

Prova:

Sendo um canal linear cuja RI é dado por \mathbf{f} de comprimento finito N , e sendo uma sequência de transmissão dada por $\mathbf{a}(n)$ também de comprimento N , a saída do canal é modelada por

$$x(n) = \bar{x}(n) + b(n), \quad (6.24)$$

em que $\bar{x}(n) = \mathbf{a}^T(n)\mathbf{f}$ é um estado do canal e $b(n)$ é uma variável aleatória cuja fdp é dada por:

$$p(b(n)) = \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp\left(-\frac{|b(n)|^2}{2\sigma_b^2}\right). \quad (6.25)$$

Então, dado que $\bar{x}(n)$ é conhecido, a fdp do sinal recebido é dada por

$$p(x(n)) = \frac{1}{\sqrt{2\pi\sigma_b^2}} \sum_{i=0}^{C-1} \exp\left(-\frac{|x(n) - \mathbf{a}_i^T \mathbf{f}|^2}{2\sigma_b^2}\right) \Pr(\mathbf{a}(n) = \mathbf{a}_i), \quad (6.26)$$

em que \mathbf{a}_i é uma sequência determinística de símbolos, $C = S^N$ é o número de sequências determinísticas \mathbf{a}_i e S é o número de símbolos no alfabeto de transmissão \mathcal{A} .



Somente para ilustração, a Figura 6.8 esboça a fdp do sinal recebido para um canal genérico de 3 coeficientes e modulação BPSK⁹, gerando os estados do canal $\pm\bar{x}_0, \pm\bar{x}_1, \pm\bar{x}_2$ e $\pm\bar{x}_3$.

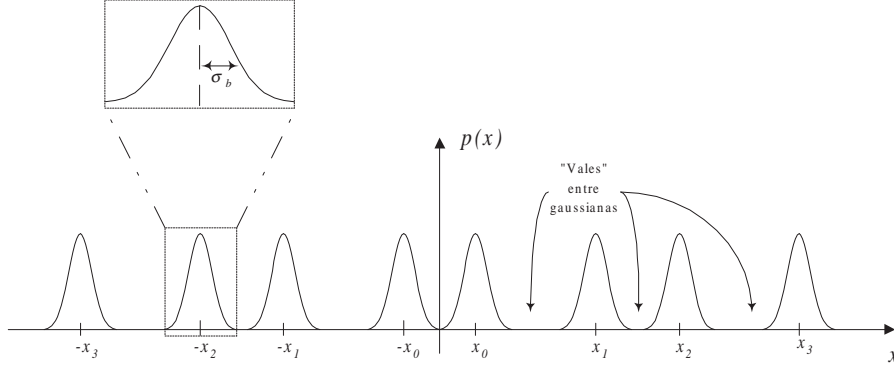


Figura 6.8: Função densidade de probabilidade do sinal na saída de um canal linear para uma modulação BPSK.

Ao observar-se as representações dos estados do canal e suas respectivas funções de interpolação (Figuras 6.5(b) e 6.6), é fácil notar que a função de interpolação possui *transições* nos pontos **entre** os estados do canal e que essas transições apresentam variações rápidas. Desta maneira, a função $\Psi(x, \theta, \beta)$ deve, para solucionar adequadamente o problema (equalização), localizar o *bias* de cada neurônio (θ_i) na região dos vales entre as gaussianas que compõem a fdp do sinal $x(n)$.

A estratégia de aprendizado da NPS adotada nesta dissertação é dividida em duas etapas:

- **1ª** : uma auto-organizada, na qual apenas os θ_i são adaptados de maneira a serem localizados *entre* as gaussianas de $p(x(n))$;
- **2ª** : após a convergência da 1ª etapa, um aprendizado supervisionado é feito para adaptação dos β_i .

Esta estratégia foi a que apresentou os melhores resultados de simulação dentre várias outras testadas. A seguir descreve-se cada etapa da estratégia.

6.5.1.1 Aprendizado auto-organizado

O aprendizado auto-organizado para a NPS tem como meta, encontrar os *vales* na fdp do sinal recebido. É então necessário encontrar uma função que seja capaz de realizar tal tarefa.

⁹ “Binary Phase Shifting Keying”.

Levando-se em consideração a medida de similaridade de funções, a maneira mais simples de fazê-lo pode ser [Small & McLeish, 1994]:

$$\int_{-\infty}^{\infty} f(x) \cdot g(x) dx, \quad (6.27)$$

em que $f(x)$ e $g(x)$ são duas funções normalizadas, ou seja,

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \int_{-\infty}^{\infty} |g(x)|^2 dx = 1.$$

No caso em questão, $f(x)$ é a fdp $p(x)$ do sinal recebido e $g(x)$ será uma função que será utilizada para comparar com $p(x)$ e minimizar a Equação (6.27).

No problema específico, essa medida deve ser capaz de associar um mínimo local para cada *vale* entre as gaussianas. Isto permitirá encontrar os parâmetros θ_i .

Dois casos foram analisados:

- **Caso (1):**

A função mais simples capaz de adequar-se aos *vales* é a função módulo $N(x, \theta_i) = |x - \theta_i|$, mostrada na Figura (a).

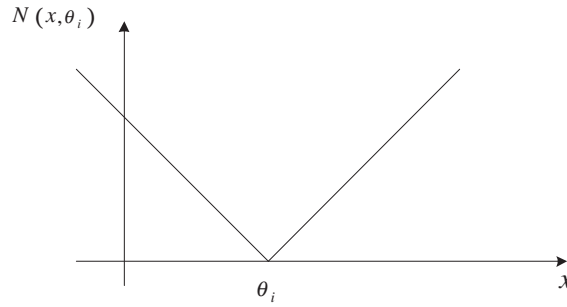


Figura 6.9: Função módulo.

A escolha dessa função é também justificada pela escolha do modelo de neurônio artificial escolhido para a NPS. Isto porque dado que os neurônios estão na região de saturação e que os pesos sinápticos da camada de entrada são iguais a 1, a saída de um neurônio k será dada por $r_k = \text{sgn}(x - d)$ que é a derivada da função $N(x, d)$.

Para essa função tem-se que a medida de similaridade dada pelo funcional:

$$J_1(d) = \int_{-\infty}^{\infty} p(x) \cdot N(x, d) dx \quad (6.28)$$

$$J_1(d) = \mathbb{E} \{N(x, d)\},$$

em que d é o parâmetro de posição que deve ser encontrado e que corresponde aos parâmetros θ_i .

A expressão do gradiente para a Equação (6.28) é então dada por:

$$\nabla J_1(d) = \mathbb{E} \{ \text{sgn}(x - d) \}, \quad (6.29)$$

e a sua versão estocástica para a busca de seu mínimo relativo aos parâmetros θ_i , é dada por:

$$\theta_i(n+1) = \theta_i(n) + \lambda \text{sgn}(x - \theta_i(n)), \quad (6.30)$$

em que λ é o passo de convergência (ou taxa de aprendizado) do algoritmo.

O efeito da busca pelas posições corretas para minimização da Equação (6.28) é ilustrado na Figura 6.10. Esse tipo de processamento em classificação é conhecido como *contraste* [Touzni, 1998].

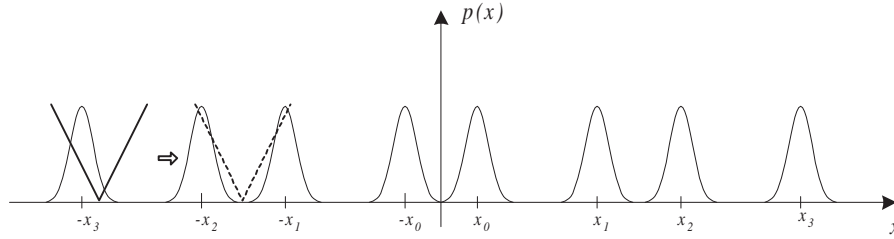


Figura 6.10: Esquema do processamento para encontrar os *vales* na fdp do sinal recebido.

Entretanto, resultados de simulação (caso ilustrado) mostraram que a minimização da Equação (6.28) não necessariamente fornece mínimos locais para **cada** *vale* da fdp do sinal recebido, como exemplificado na Figura 6.11.

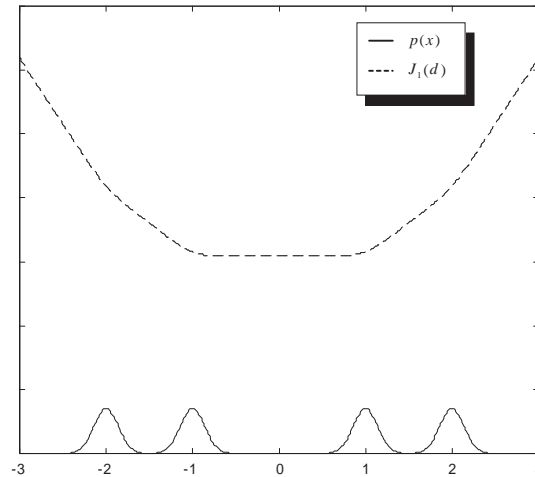


Figura 6.11: Funcional J_1 e fdp do sinal recebido $p(x)$.

- **Caso (2):**

Como a função $N(x, d)$ não se mostrou capaz de encontrar todos os vales, isto parece indicar a necessidade de utilizar uma função que seja mais adequada. Observando-se os resultados da Figura 6.11, leva-se a crer que o problema tem a relação com a *estreiteza* da função.

Uma das maneiras mais simples de tornar a função mais *estreita* é através da aplicação do logaritmo natural, conforme mostrado na Figura 6.12.

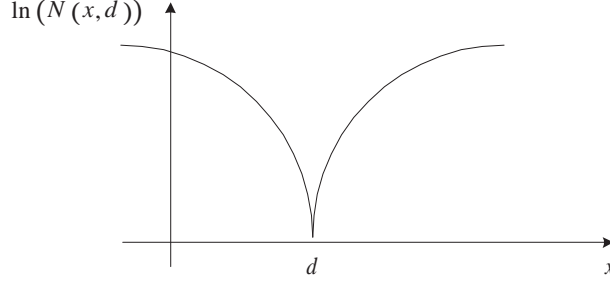


Figura 6.12: Função $\ln(N(x, d))$.

Substituindo-se na função custo tem-se:

$$J_2(d) = \int_{-\infty}^{\infty} p(x) \ln(N(x, d)) dx \quad (6.31)$$

$$J_2(d) = \mathbb{E} \{ \ln(N(x, d)) \}.$$

Deve-se notar que a Equação (6.31) é semelhante à medida de divergência de Kullback-Leibler (KLD) entre as funções $p(x)$ e $N(x, d)$. Mas para garantir que ambas as funções sejam estritamente positivas deve-se substituir a função $N(x, d)$ por uma função $N(x, d, \kappa) = |x - d| + \kappa$. Desta maneira, pod-se escrever:

$$D_{p(x)||N(x,d,\kappa)} = \int_{-\infty}^{\infty} p(x) \ln \left(\frac{p(x)}{N(x, d, \kappa)} \right) dx$$

$$D_{p(x)||N(x,d,\kappa)} = \int_{-\infty}^{\infty} p(x) \ln(p(x)) dx - \int_{-\infty}^{\infty} p(x) \ln(N(x, d, \kappa)) dx. \quad (6.32)$$

Dessa forma, o funcional será então descrito por uma equação baseada na KLD. Considerando na Equação (6.32) apenas o termo que depende de $N(x, d, \kappa)$, dada pelo funcional:

$$J_2(d) = - \int_{-\infty}^{\infty} p(x) \ln(N(x, d, \kappa)) dx \quad (6.33)$$

$$J_2(d) = -\mathbb{E} \{ \ln(N(x, d, \kappa)) \}.$$

E, calculando-se o gradiente obtém-se:

$$\nabla J_2(d) = -\mathbb{E} \left\{ \frac{\text{sgn}(x - d)}{|x - d| + \kappa} \right\}, \quad (6.34)$$

e os valores mínimos para d (ou θ_i) podem ser obtidos através da versão estocástica do gradiente:

$$\theta_i(n+1) = \theta_i(n) - \lambda \frac{\text{sgn}(x - \theta_i(n))}{|x - \theta_i(n) + \kappa|}, \quad (6.35)$$

em que λ é o passo de convergência (ou taxa de aprendizado) do algoritmo.

Pode-se então observar que o algoritmo proposto é um tipo de lei anti-Hebbiana [Haykin, 1998].

Outro aspecto de importante conceituação é o parâmetro κ que além de evitar problemas numéricos para $x = \theta_i$ e tornar a função $N(x, d)$ estritamente positiva também atua como um fator de controle da *estreiteza* da função [Schraudolph, 1995].

Os resultados de simulação mostraram que o funcional $J_2(d)$ baseado na KLD mostrou-se capaz de associar mínimos locais para **cada** *vale* de $p(x)$, como mostrado na Figura 6.13.

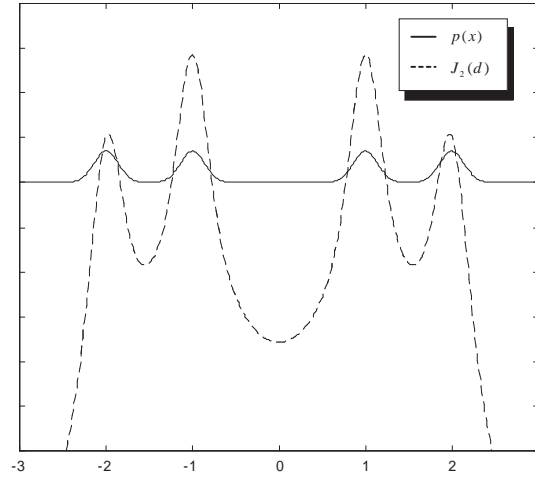


Figura 6.13: Funcional J_2 e fdp do sinal recebido $p(x)$.

Este algoritmo será aqui denominado “**Self-Organized for Finding Valleys Algorithm**” (SOFVA) e será utilizado nas simulações que utilizam um aprendizado auto-organizado.

A Tabela 6.1 resume a dinâmica do algoritmo SOFVA.

Pode-se então esquematizar o aprendizado da NPS através da Figura 6.14 em que observa-se o chaveamento do algoritmo auto-organizado (para ajuste dos parâmetros θ_i) para o supervisionado (ajuste dos parâmetros lineares β_i).

Algoritmo SOFVA
(1) Apresentação do sinal $x(n-1)$
(2) Cálculo do gradiente: $\Delta\theta_i(n) = \lambda \frac{\text{sgn}(x(n-1)-\theta_i(n))}{ x(n-1)-\theta_i(n) +\kappa}$
(3) Atualização: $\theta_i(n+1) = \theta_i(n) - \Delta\theta_i(n)$

Tabela 6.1: Resumo do algoritmo auto-organizado SOFVA.

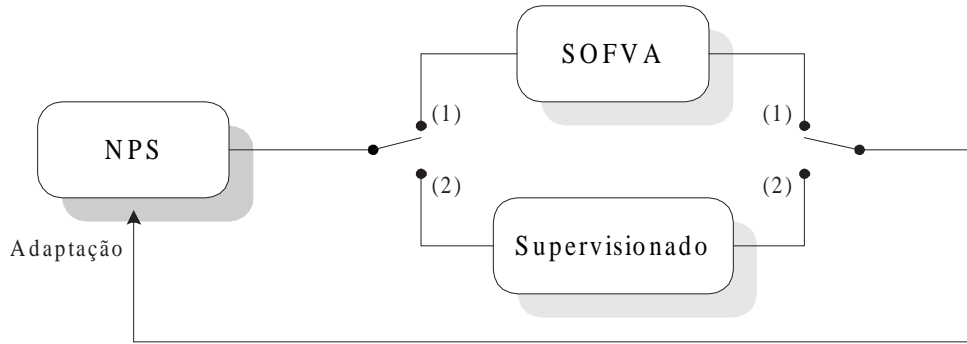


Figura 6.14: Esquema do aprendizado para a NPS - auto-organizado e supervisionado.

6.5.1.2 Aprendizado supervisionado

Uma vez que, na estratégia adotada nessa dissertação, a NPS tenha *aprendido* as posições dos parâmetros θ_i é suficiente a adaptação dos parâmetros β_i , (lineares) para determinar a função de interpolação $\Psi(x, \theta, \beta)$.

Nesse ponto, como o aprendizado supervisionado corresponde apenas a uma adaptação linear de parâmetros baseados nos dados pode-se escolher dois algoritmos, descritos a seguir, para encontrar os valores de β .

Vale salientar que, o aprendizado é supervisionado para a NPS (rede neural) pois existe uma referência para a adaptação já que há um cálculo direto do erro. Entretanto, para a equalização é um processamento *cego* já que não se baseia no conhecimento dos dados transmitidos.

- **Minimização do erro de predição (MEP)**

Essa estratégia usa o mesmo critério da filtragem linear para o branqueamento do

signal de saída do preditor. Dessa maneira o funcional a ser minimizado é $\mathbb{E} \{|e_p(n)|^2\}$ e a equação de adaptação pode ser dada por:

$$\boldsymbol{\beta}(n+1) = \boldsymbol{\beta}(n) + \mu_\beta \cdot \mathbf{r}(n) \cdot e_p(n), \quad (6.36)$$

em que $\mathbf{r} = [r_1 \ r_2 \ \cdots \ r_{M-1} \ r_M]^T$ são as saídas dos neurônios.

A estrutura quando utilizar esse critério será denominada “NPS-Minimized for Prediction Error” (NPS-MPE).

• Critério CM (CMC)

O critério CM, como já descrito anteriormente, foi utilizado para predição em [Cavalcanti, 1996]. Nesse caso minimiza-se o funcional $\mathbb{E} \{(|y(n)|^2 - R_2)^2\}$ em que $y(n) = g \cdot (x(n) - \hat{x}(n))$, g é o CAG e a equação de adaptação é dada por:

$$\boldsymbol{\beta}(n+1) = \boldsymbol{\beta}(n) + \mu_\beta \cdot \mathbf{r}(n) \cdot y(n) \cdot (R_2 - |y(n)|^2), \quad (6.37)$$

Quando da utilização desse critério a estrutura será denominada “NPS-Constant Modulus Criterion” (NPS-CMC).

A Tabela 6.2 resume a dinâmica dos algoritmos supervisionados utilizados para adaptação dos parâmetros lineares.

Algoritmo MEP	Algoritmo CMC
<p>(1) Apresentação do sinal $x(n-1)$</p> <p>(2) $e_p(n) = x(n) - \Psi(x(n-1), \boldsymbol{\theta}, \boldsymbol{\beta}(n-1))$</p> <p>(3) Cálculo do gradiente: $\Delta\boldsymbol{\beta}(n) = \mu_\beta \cdot \mathbf{r}(n) \cdot e_p(n)$</p> <p>(4) Atualização: $\boldsymbol{\beta}(n+1) = \boldsymbol{\beta}(n) + \Delta\boldsymbol{\beta}(n)$</p>	<p>(1) Apresentação do sinal $x(n-1)$</p> <p>(2) $e_p(n) = x(n) - \Psi(x(n-1), \boldsymbol{\theta}, \boldsymbol{\beta}(n-1))$</p> <p>(3) Cálculo do gradiente: $\Delta\boldsymbol{\beta}(n) = \mu_\beta \cdot \mathbf{r}(n) \cdot y(n) \cdot (R_2 - y(n) ^2)$</p> <p>(4) Atualização: $\boldsymbol{\beta}(n+1) = \boldsymbol{\beta}(n) + \Delta\boldsymbol{\beta}(n)$</p>

(a):Algoritmo MPE.

(b):Algoritmo CMC.

Tabela 6.2: Resumo dos algoritmos supervisionados para a NPS.

6.5.2 Limites e complexidade

A complexidade da NPS está relacionada à RI do canal. Isto porque, como cada neurônio deve estar localizado *entre* duas gaussianas, então o número mínimo de neurônios necessários para a correta identificação dos estados do canal e conseqüente construção da função (superfície) de interpolação é igual à:

$$M = S^N - 1, \quad (6.38)$$

em que S é o número de símbolos no alfabeto de transmissão \mathcal{A} e N é a ordem do canal.

Obviamente, para que o SOFVA seja capaz de colocar corretamente os parâmetros θ_i *entre* as gaussianas, o número de neurônios deve ser suficiente para permitir uma identificação de todos os *vales* de $p(x)$, entretanto não se pode inicializar a NPS com apenas o número mínimo de neurônios, uma vez que se desconhece o canal.

Uma questão também igualmente importante é a forte dependência do desempenho do SOFVA com a *inicialização* dos parâmetros θ_i , visto que, o funcional associa mínimos locais para cada um dos *vales*, a inicialização pode determinar uma convergência para pontos em que existam neurônios *colapsados*, isto é, sendo localizados no mesmo mínimo local [Haykin, 1998].

Essa preocupação de evitar o *colapso* de neurônios é natural do processamento com RNA. E uma das soluções factíveis é a inicialização aleatória do *bias* em distribuição uniforme dentro de uma determinada faixa de valores para que os neurônios sejam colocados em regiões diferentes.

No problema da predição utilizando a NPS é necessário que o número de neurônios que sejam inicializados seja muito maior do que o mínimo, para garantir que toda a faixa de valores dos estados do canal, que são os centros das gaussianas, seja mapeada.

Assim, provavelmente, alguns (ou mesmo vários) neurônios estarão colapsados, mas isso não representa prejuízo na performance. Entretanto há um aumento na complexidade, pois o número de parâmetros β_i a serem adaptados pelo critério supervisionado será igual ao número de neurônios (ver Figura 6.7).

Uma *possível* solução para esse problema é a *regularização* (ou “prunning”) da rede que implementa a NPS. A regularização corresponde a descobrir qual, ou quais, neurônios estão colapsados e retirá-los da rede, ou melhor, substituí-los por *apenas um* que represente o aprendizado do conjunto, com isso diminuindo a complexidade da rede através da redução do número de parâmetros livres. Entretanto, a construção de um funcional é uma tarefa complexa num ambiente autodidata, pois não se detém nenhum conhecimento sobre o canal, podendo levar a uma extração de neurônios não colapsados. Esse procedimento, torna ainda possível estimar a ordem do *canal* pois, uma vez encontrado o número mínimo de neurônios pode-se estimar a ordem do canal através da Equação (6.38).

Ainda na questão de limites, o bom desempenho está vinculado a uma boa estimação dos *vales* de $p(x)$. Mas a estimação é relacionada com a distância entre as gaussianas

podendo apresentar dois problemas:

- (1) Ruído: quando o ruído é forte, a variância de cada uma das gaussianas aumenta, podendo fazer com que duas gaussianas adjacentes praticamente não apresentem um *vale* entre elas. Dessa maneira, o SOFVA não consegue associar um mínimo local àquela posição, não possibilitando a estimação da correta função de interpolação.
- (2) Canal: quando a resposta do canal fornece estados adjacentes muito próximos, mesmo uma pequena contribuição do ruído pode ser suficiente para tornar as gaussianas indistinguíveis para o SOFVA. Outro aspecto é que, como está sendo considerada apenas uma entrada no ramo preditivo, e se o canal possuir zeros superpostos, também não é possível construir uma função de interpolação para separar os sinais pertencentes à diferentes classes. Isso se deve ao fato de que a projeção em uma dimensão apresenta duas (ou mais) gaussianas superpostas [Montalvão et al., 1999; Montalvão, 2000].

Em algumas situações, como para os canais de FMín, esses problemas são minorizados, pois, como a função de interpolação apresenta uma forma mais simples (próxima da linear), uma má estimação de um parâmetro *pode* não comprometer o desempenho da NPS. Entretanto, para canais de FNM, a função de interpolação a ser implementada é bastante complexa e uma boa estimação dos parâmetros θ_i é essencial para o bom desempenho da estrutura.

Quanto ao processo de aprendizado, esse é bem menos restritivo visto que corresponde a um processamento bem mais simples. Dessa maneira, não representa nenhuma questão sobre a limitação da estratégia.

A seção seguinte ilustra os resultados de simulação obtidos para a estratégia.

6.6 Simulações computacionais

As simulações computacionais realizadas têm como meta a comparação do desempenho da proposição frente a outras estratégias de equalização.

As simulações foram calculadas sobre 100 simulações de Monte Carlo. No caso das curvas de evolução do erro de decisão quadrático (DSE¹⁰) para suavização das curvas foi também utilizada uma filtragem passa-baixas através de um filtro de ordem 1 e frequência de corte $\omega_c = 10^{-2}$.

O CAG, quando utilizado, é atualizado pelo algoritmo proposto em [Macchi, 1995, 1998]:

$$\begin{aligned} G(n+1) &= G(n) + \mu_{\text{CAG}} \cdot (\sigma_a^2 - |y(n)|^2) \\ g(n) &= \sqrt{|G(n)|}. \end{aligned} \tag{6.39}$$

¹⁰Do inglês “Decision Squared Error”.

A descrição da metodologia e parâmetros de simulação encontram-se descritas após cada caso.

6.6.1 Canais de fase mínima

Canal 1 - Canal de fase mínima para $MD(\mathbf{f}) < 1$:

O canal apresenta a seguinte RI:

$$\mathbf{f}_1 = \begin{bmatrix} 1 & 0.6 & 0.2 \end{bmatrix}^T. \quad (6.40)$$

E como $MD(\mathbf{f}_1) = \frac{|1.8| - |1|}{|1|} = 0.8$, o canal apresenta a condição de *olho aberto*.

As Figuras 6.15 e 6.16 mostram a evolução do DSE para as estruturas NPS, CMA, DFE e preditor linear.

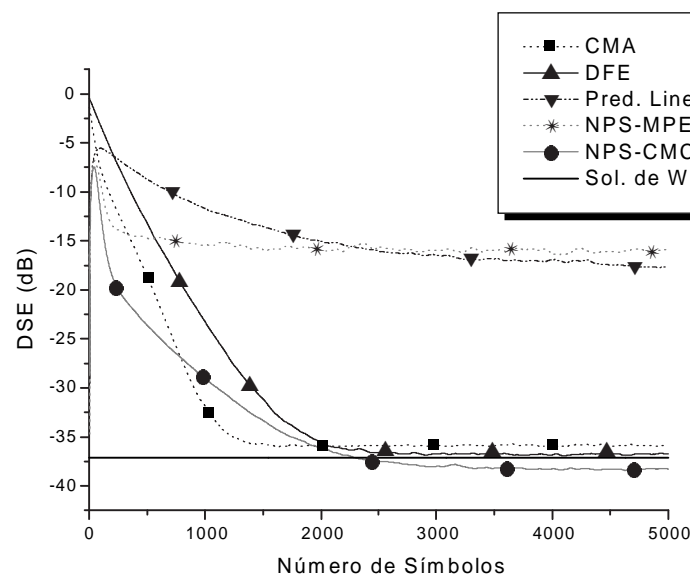


Figura 6.15: Evolução do DSE para o canal \mathbf{f}_1 para uma $SNR = 40$ dB com modulação BPSK.

Observa-se um melhor desempenho da NPS-CMC que a NPS-MPE. Além disso, o NPS-CMC fornece um desempenho em termos de DSE da ordem do DSE fornecido pelo DFE. Isso pode ser justificado pelos seguintes fatores:

- a NPS-MPE utiliza como critério de erro para a rede, o erro de predição, e esse não apresenta uma convergência para zero, o que dificulta bastante a tarefa de aprendizado;

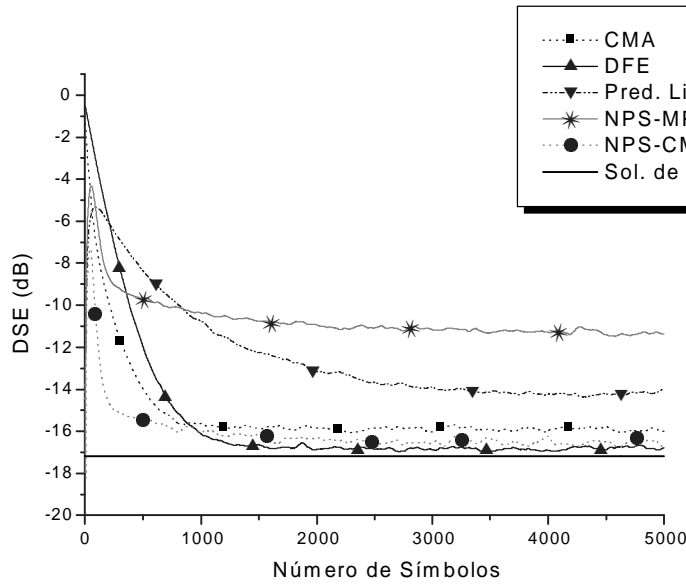


Figura 6.16: Evolução do DSE para o canal f_1 para uma $SNR = 20$ dB com modulação BPSK.

- por sua vez, a NPS-CMC utiliza um critério de erro assintoticamente convergente para zero e, por utilizar apenas uma entrada, a amplificação de ruído ocorre de maneira menos intensa que em uma estrutura com várias entradas;
- uma superfície de interpolação não linear é mais adequada que um hiperplano para separação de classes, resultando num melhor desempenho na equalização.

Essas justificativas são pertinentes devido ao fato da fase ser mínima e o atraso $\delta = 0$ ser favorável para a NPS e para o DFE.

Entretanto, foi observado que os parâmetros θ_i inicializados *linearmente espaçados* (LS¹¹) ao invés de uma inicialização aleatória, a estrutura NPS-MPE apresentou um melhor desempenho em termos de DSE, o que não ocorreu com a NPS-CMC, a qual não modificou seu desempenho com uma mudança na estratégia de inicialização. A Figura 6.17 compara o desempenho dessa estratégia, que será nomeada NPS-MPE-“Linearly Spaced” (NPS-MPE-LS), com o do preditor linear e o da NPS-MPE.

Essa diferença se deve, provavelmente, ao fato de que ao usar uma inicialização LS, pode-se utilizar um número menor de neurônios na estrutura (foram utilizados 20 para as simulações com o canal f_1). Isso aliado ao fato de que o erro de predição não converge para zero, pode ocasionar um melhor desempenho pela adaptação de um número menor de parâmetros.

A Figura 6.18 ilustra as superfícies de interpolação obtidas pelas estruturas NPS-MPE, NPS-MPE-LS e NPS-CMC no final da convergência para uma $SNR = 40$ dB.

¹¹Do inglês “Linearly Spaced”.

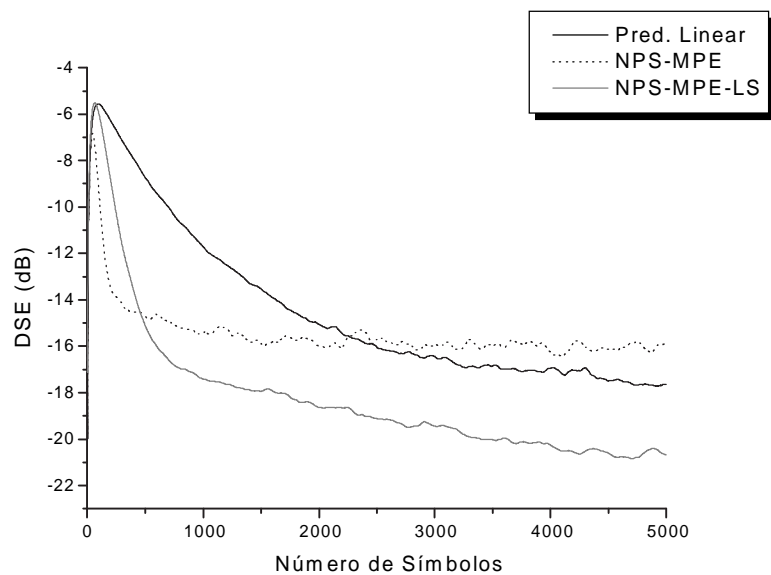
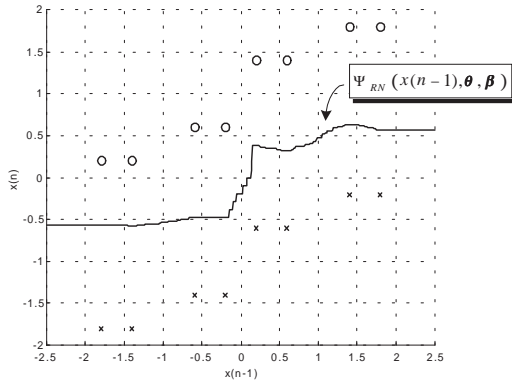


Figura 6.17: Evolução do DSE para o canal f_1 para uma $SNR = 40$ dB com modulação BPSK - comparação entre preditor linear, NPS-MPE e NPS-MPE-LS.

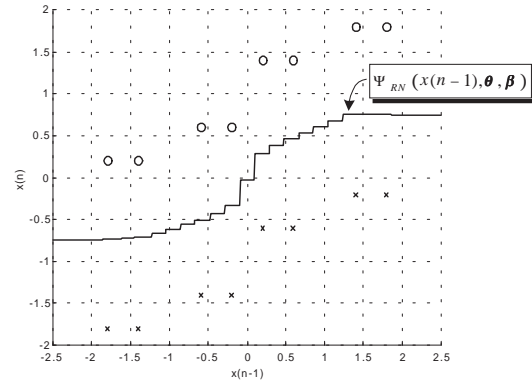
Deve-se notar que a NPS-CMC fornece uma melhor superfície de interpolação, já que a mesma se localiza simétrica em relação aos estados e com um valor constante nessa faixa. Pode-se ainda verificar que o mesmo não ocorre com a NPS-MPE e NPS-MPE-LS que apresentam falta de simetria em relação aos estados e também variação na faixa entre os estados.

Simulações também foram realizadas para avaliar o comportamento da NPS utilizando-se um treinamento em paralelo dos parâmetros θ_i e β_i , ou seja, o algoritmo SOFVA e uma das estratégias supervisionadas são executados simultaneamente. Verificou-se que o desempenho em termos de evolução do DSE é o mesmo mas o algoritmo SOFVA leva os parâmetros θ_i para os valores dos *bias* ótimos (valores médios entre as médias das gaussianas adjacentes), favorecendo assim uma superfície de interpolação mais compacta. Esse resultado leva a uma consideração da aplicabilidade da estratégia para canais variantes no tempo. A Figura 6.19 mostra uma evolução temporal dos *bias* em relação aos valores ótimos, que são aqueles referentes aos valores médios entre cada dois centros de gaussianas adjacentes.

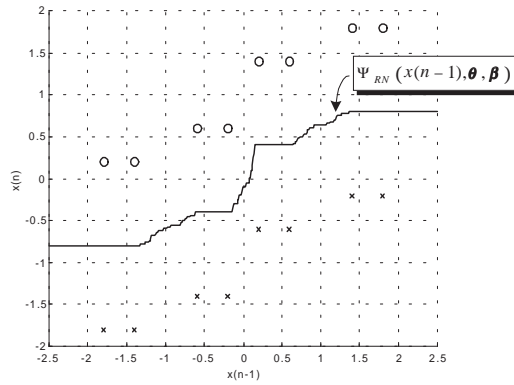
A Tabela 6.3 resume os parâmetros de simulação utilizados.



(a): Superfície de interpolação da NPS-MPE.



(b): Superfície de interpolação da NPS-MPE-LS.



(c): Superfície de interpolação da NPS-CMC.

Figura 6.18: Superfícies de interpolação para as estruturas NPS - canal f_1 , SNR = 40 dB e modulação BPSK.

Canal 2 - Canal de fase mínima para $MD(f) > 1$:

O canal apresenta a seguinte RI:

$$\mathbf{f}_2 = \begin{bmatrix} 1 & 0.8 & 0.4 \end{bmatrix}^T. \quad (6.41)$$

E como $MD(\mathbf{f}_2) = \frac{|2.2| - |1|}{|1|} = 1.2$, o canal apresenta a condição de *olho fechado*.

A evolução do DSE para as estruturas NPS, CMA, DFE e preditor linear é ilustrada nas Figuras 6.20 e 6.21.

A exemplo do caso 1, observou-se um melhor desempenho da NPS-MPE com uma inicialização LS dos parâmetros θ_i , como mostrado na Figura 6.22, relativo ao erro

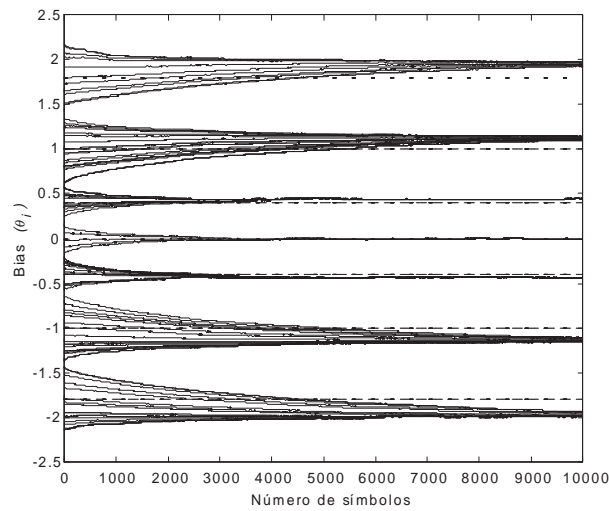


Figura 6.19: Evolução temporal dos parâmetros θ_i - (--) valores ótimos e (-) valores obtidos.

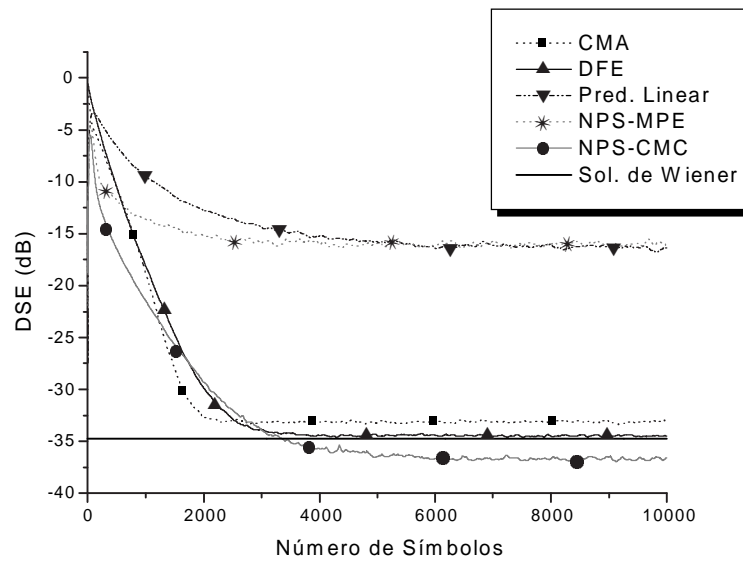


Figura 6.20: Evolução do DSE para o canal f_2 para uma $SNR = 40$ dB com modulação BPSK.

de decisão mínimo. Porém, quanto à velocidade de convergência, a NPS-MPE com inicialização aleatória teve um desempenho melhor.

As superfícies de interpolação obtidas ao final da convergência para a NPS-MPE, NPS-MPE-LS e NPS-CMC são ilustradas na Figura 6.23.

Novamente, a superfície gerada pela NPS-CMC mostrou uma melhor disposição de simetria em relação aos estados do canal bem como uma menor variação na faixa dos estados.

Parâmetros de simulação para o canal f_1	
<i>CMA</i>	$\mu = 5 \cdot 10^{-3}$ Comprimento do filtro: 25 taps Inicialização: $\mathbf{h} = [0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0]^T$
<i>DFE</i>	$\mu_h = 5 \cdot 10^{-3}, \mu_d = 10^{-4}$ Comprimento do filtro: 15 (direto) e 5 (realim.) Inicialização: $\begin{cases} \mathbf{h} = \mathbf{0} \\ \mathbf{d} = \mathbf{0} \end{cases}$
<i>Pred. Linear</i>	$\mu_h = 10^{-3}, \mu_{\text{CAG}} = 10^{-3}$ Comprimento do filtro: 25 Inicialização: $\begin{cases} \mathbf{h} = \mathbf{0} \\ g = G = 1 \end{cases}$
<i>NPS-MPE</i>	$\lambda = 10^{-4}, \mu_\beta = 5 \cdot 10^{-4}, \mu_{\text{CAG}} = 10^{-3}, \kappa = 10^{-7}$ Número de neurônios: 100 Num. símbolos para SOFVA: 50 Inicialização: $\begin{cases} \theta_i = U[-2, 2] \\ \beta = \mathbf{0} \\ g = G = 1 \end{cases}$
<i>NPS-MPE-LS</i>	$\lambda = 10^{-4}, \mu_\beta = 5 \cdot 10^{-4}, \mu_{\text{CAG}} = 10^{-3}, \kappa = 10^{-7}$ Número de neurônios: 20 Num. símbolos para SOFVA: 50 Inicialização: $\begin{cases} \theta_i = \text{LS}[-2, 2] \\ \beta = \mathbf{0} \\ g = G = 1 \end{cases}$
<i>NPS-CMC</i>	$\lambda = 10^{-4}, \mu_\beta = 5 \cdot 10^{-4}, \mu_{\text{CAG}} = 10^{-3}, \kappa = 10^{-7}$ Número de neurônios: 100 Num. símbolos para SOFVA: 50 Inicialização: $\begin{cases} \theta_i = U[-2, 2] \\ \beta = \mathbf{0} \\ g = G = 1 \end{cases}$

Tabela 6.3: Sumário da simulação para o canal f_1 .

A Tabela 6.4 resume os parâmetros de simulação utilizados.

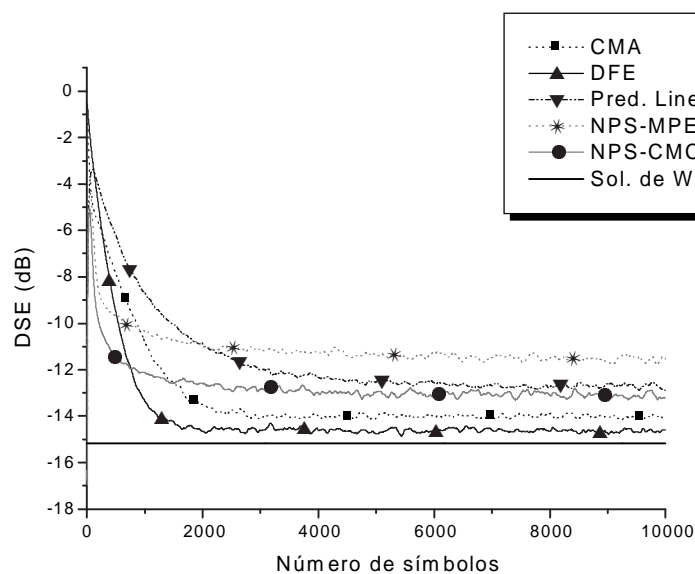


Figura 6.21: Evolução do DSE para o canal f_2 para uma $SNR = 20$ dB com modulação BPSK.

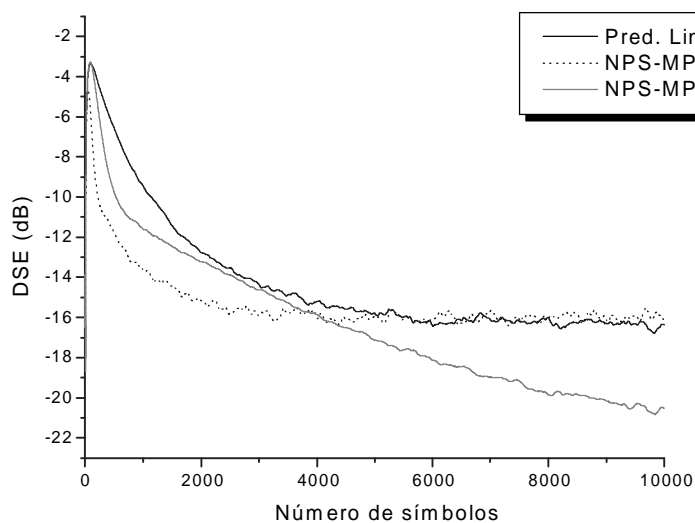


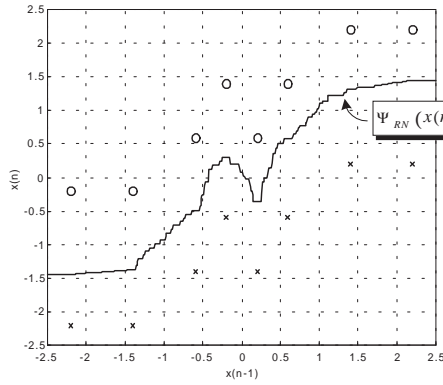
Figura 6.22: Evolução do DSE para o canal f_2 para uma $SNR = 40$ dB com modulação BPSK - comparação entre preditor linear, NPS-MPE e NPS-MPE-LS.

6.6.2 Canais de fase não mínima

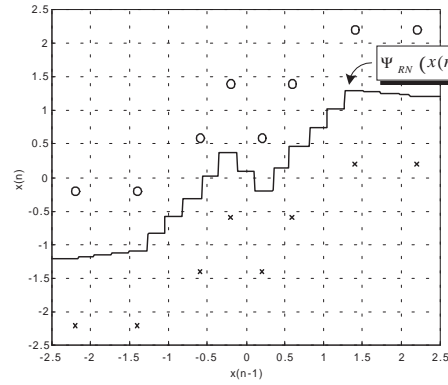
Canal 3 - Canal de fase mínima para $MD(f) > 1$:

O canal apresenta a seguinte RI:

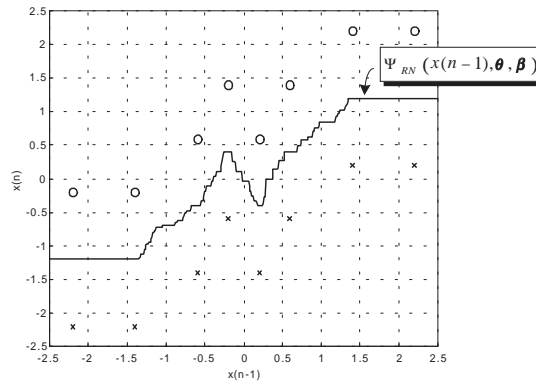
$$\mathbf{f}_3 = \begin{bmatrix} 0.5 & 1 & -0.6 \end{bmatrix}^T. \quad (6.42)$$



(a): Superfície de interpolação da NPS-MPE.



(b): Superfície de interpolação da NPS-MPE-LS.



(c): Superfície de interpolação da NPS-CMC.

Figura 6.23: Superfícies de interpolação para as estruturas NPS - canal f_2 , SNR = 40 dB e modulação BPSK.

E como $MD(\mathbf{f}_3) = \frac{|2.1| - |1|}{|1|} = 1.1$, o canal apresenta a condição de *olho fechado*.

A Figura 6.24 mostra a evolução do DSE para as estruturas NPS, CMA e DFE. Como é conhecido que um preditor linear utilizando estatísticas de ordem 2 não é capaz de equalizar um canal de FNM, este não é apresentado nas curvas.

Neste caso é possível observar as limitações do método pois, embora o algoritmo seja capaz de realizar a equalização utilizando apenas **uma** entrada, as características do canal e a SNR não possibilitam uma melhora substancial no desempenho, mesmo utilizando um maior número de símbolos para a adaptação dos θ_i (nesse caso foram utilizados 500 símbolos). Além disso apresenta um comportamento não satisfatório para SNR inferiores a 35 dB, como exemplifica a Figura 6.25.

Nesse caso não foi observado nenhuma melhora do desempenho ao utilizar-se a

Parâmetros de simulação para o canal f_2	
<i>CMA</i>	$\mu = 3 \cdot 10^{-3}$ Comprimento do filtro: 25 taps Inicialização: $\mathbf{h} = [0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0]^T$
<i>DFE</i>	$\mu_h = 5 \cdot 10^{-3}, \mu_d = 10^{-4}$ Comprimento do filtro: 15 (direto) e 5 (realim.) Inicialização: $\begin{cases} \mathbf{h} = \mathbf{0} \\ \mathbf{d} = \mathbf{0} \end{cases}$
<i>Pred. Linear</i>	$\mu_h = 10^{-3}, \mu_{CAG} = 10^{-3}$ Comprimento do filtro: 25 Inicialização: $\begin{cases} \mathbf{h} = \mathbf{0} \\ g = G = 1 \end{cases}$
<i>NPS-MPE</i>	$\lambda = 10^{-4}, \mu_\beta = 5 \cdot 10^{-4}, \mu_{CAG} = 10^{-3}, \kappa = 10^{-7}$ Número de neurônios: 100 Num. símbolos para SOFVA: 50 Inicialização: $\begin{cases} \theta_i = U[-2, 2] \\ \beta = \mathbf{0} \\ g = G = 1 \end{cases}$
<i>NPS-MPE-LS</i>	$\lambda = 10^{-4}, \mu_\beta = 5 \cdot 10^{-4}, \mu_{CAG} = 10^{-3}, \kappa = 10^{-7}$ Número de neurônios: 20 Num. símbolos para SOFVA: 50 Inicialização: $\begin{cases} \theta_i = LS[-2, 2] \\ \beta = \mathbf{0} \\ g = G = 1 \end{cases}$
<i>NPS-CMC</i>	$\lambda = 10^{-4}, \mu_\beta = 5 \cdot 10^{-4}, \mu_{CAG} = 10^{-3}, \kappa = 10^{-7}$ Número de neurônios: 100 Num. símbolos para SOFVA: 50 Inicialização: $\begin{cases} \theta_i = U[-2, 2] \\ \beta = \mathbf{0} \\ g = G = 1 \end{cases}$

Tabela 6.4: Sumário da simulação para o canal f_2 .

NPS-MPE-LS. E visto que seu desempenho foi o mesmo da NPS-MPE, não foi incluído nas curvas.

A Figura 6.26 ilustra as funções de interpolação, atingidas ao final da convergência para as SNR anteriormente citadas, e para as estratégias NPS-MPE e NPS-CMC.

Vale notar que as superfícies são muito similares para ambos valores de SNR, sendo que o mal desempenho refletido na Figura 6.25 é assumido da própria estrutura da função de interpolação. As transições rápidas são também responsáveis pela mudança abrupta de direção e referência. No caso de um ruído mais forte, o sinal pode então não estar sendo

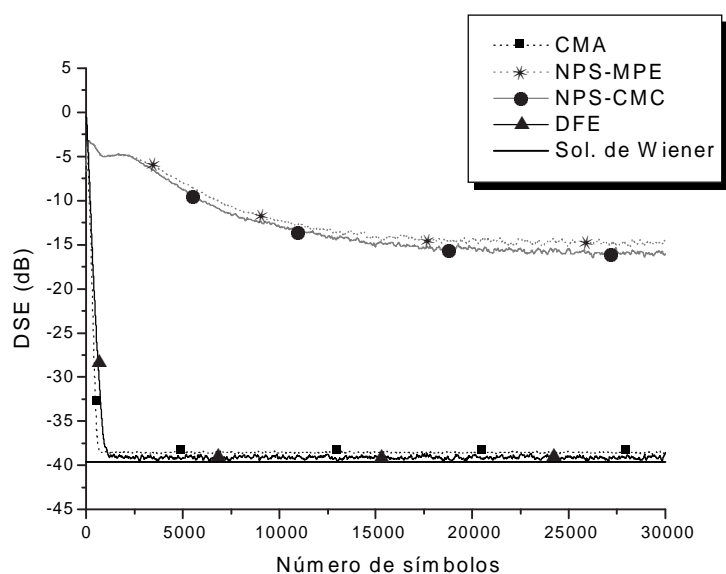


Figura 6.24: Evolução do DSE para o canal f_3 para uma $SNR = 40$ dB com modulação BPSK.

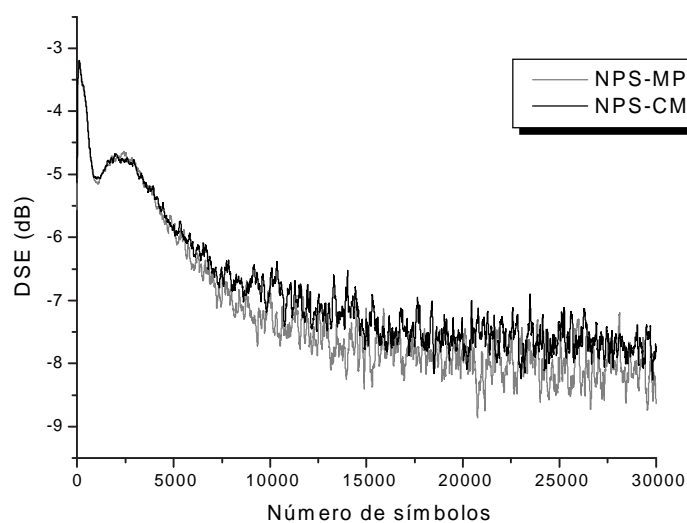
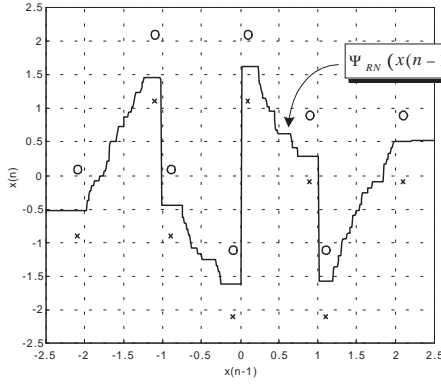


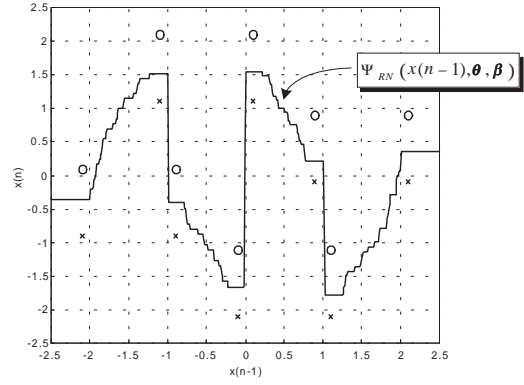
Figura 6.25: Evolução do DSE para o canal f_3 para uma $SNR = 30$ dB com modulação BPSK.

comparado com o valor de $\Psi(x(n-1), \theta, \beta)$ associado ao estado que gera o sinal, e sim ao seu estado vizinho na representação dos estados. Com isso o erro passa a ter uma medida muito maior (ou mesmo menor) do que o seu valor ideal.

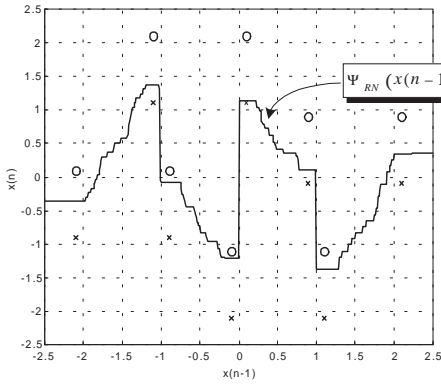
A Tabela 6.5 resume os parâmetros de simulação utilizados.



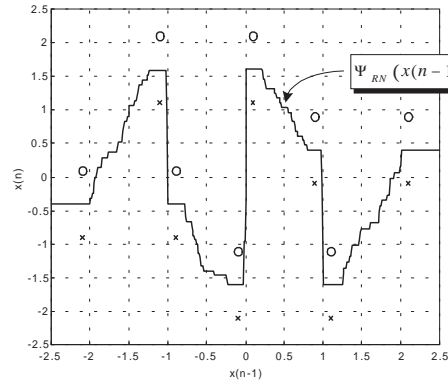
(a): NPS-MPE (SNR = 30 dB)



(b): NPS-MPE (SNR = 40 dB)



(c): NPS-CMC (SNR = 30 dB)



(d): NPS-CMC (SNR = 40 dB)

Figura 6.26: Superfícies de interpolação para as estruturas NPS - canal f_3 e modulação BPSK.

Canal 4 - Canal de fase mínima para $MD(f) < 1$:

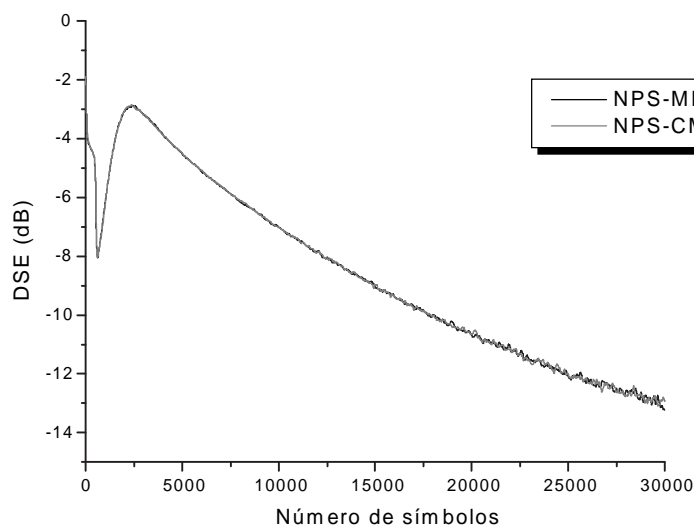
O canal apresenta a seguinte RI:

$$\mathbf{f}_4 = [0.2 \quad 1 \quad 0.6]^T. \quad (6.43)$$

E como $MD(\mathbf{f}_1) = \frac{|1.8| - |1|}{|1|} = 0.8$, o canal apresenta a condição de *olho aberto*.

Nesse caso, embora o canal respeite a condição de *olho aberto*, o valor da componente f_0 é muito baixo em comparação com as outras o que traz problemas para a construção da superfície e conseqüentemente uma má estimação do sinal na saída do equalizador. A Figura 6.27 ilustra as curvas de evolução do DSE para ambas estratégias, NPS-MPE e NPS-CMC. Aqui fica evidenciado um limite do método, associado ao uso de uma única entrada.

Parâmetros de simulação para o canal f_3	
<i>CMA</i>	$\mu = 5 \cdot 10^{-3}$ Comprimento do filtro: 25 taps Inicialização: $\mathbf{h} = [0 \ \dots \ 0 \ 1 \ 0 \ \dots \ 0]^T$
<i>DFE</i>	$\mu_h = 5 \cdot 10^{-3}, \mu_d = 10^{-4}$ Comprimento do filtro: 25 (direto) e 5 (realim.) Inicialização: $\begin{cases} \mathbf{h} = \mathbf{0} \\ \mathbf{d} = \mathbf{0} \end{cases}$
<i>NPS-MPE</i>	$\lambda = 10^{-4}, \mu_\beta = 5 \cdot 10^{-4}, \mu_{CAG} = 10^{-3}, \kappa = 10^{-7}$ Número de neurônios: 100 Num. símbolos para SOFVA: 500 Inicialização: $\begin{cases} \theta_i = U[-2.5, 2.5] \\ \beta = \mathbf{0} \\ g = G = 1 \end{cases}$
<i>NPS-CMC</i>	$\lambda = 10^{-4}, \mu_\beta = 5 \cdot 10^{-4}, \mu_{CAG} = 10^{-3}, \kappa = 10^{-7}$ Número de neurônios: 100 Num. símbolos para SOFVA: 500 Inicialização: $\begin{cases} \theta_i = U[-2.5, 2.5] \\ \beta = \mathbf{0} \end{cases}$

Tabela 6.5: Sumário da simulação para o canal f_3 .Figura 6.27: Evolução do DSE para o canal f_4 para uma SNR = 40 dB com modulação BPSK.

6.7 Síntese e conclusões

Esse capítulo apresentou uma das propostas dessa dissertação, versando sobre a utilização de uma estratégia não linear baseada em técnicas preditivas. Ao utilizar-se uma RNA como

dispositivo não linear para a predição, procurou-se tomar proveito das suas características de não linearidade e capacidade de aprendizado.

Na questão do aprendizado, a inserção de uma abordagem inédita para a construção de uma função de interpolação capaz de implementar uma separação entre classes, foi baseada na divisão do aprendizado em duas etapas: auto-organizado e supervisionado. Foi então proposto um algoritmo de aprendizado auto-organizado, com base na minimização da divergência de Kullback-Leibler, para possibilitar uma boa estimação da função de interpolação.

Na etapa supervisionada dois critérios de minimização foram avaliados para a adaptação da parte linear (camada de saída) da estrutura de filtragem: um baseado na minimização do erro de predição e outro na penalização dos sinais exteriores a um raio de equalização, o critério CM.

As limitações da estratégia são discutidas e algumas possibilidades de solução para problemas inerentes à técnica de processamento não linear através de RNA também são apresentadas visando enriquecer a discussão.

A análise dos resultados de simulação mostraram um desempenho bastante razoável para as estruturas neurais propostas para os casos de canais de fase mínima, tanto para a condição de *olho aberto* quanto para a condição de *olho fechado*. A inserção do critério CM na adaptação da parte linear mostrou-se, naqueles casos, uma solução bastante atrativa, cujo desempenho foi melhor que a própria estratégia supervisionada com o DFE e superando o desempenho do equalizador de Wiener. Esse resultado pode ser melhor explicado a partir da teoria da classificação, fundamentando-se na comparação entre os critérios de Wiener (minimização do erro quadrático) e de Bayes (minimização da probabilidade de erro) [Montalvão, 2000].

Já para os resultados obtidos para os casos de canais de fase não mínima, a possibilidade de realizar a equalização através do uso de estatísticas de ordem 2 é mostrada como plausível, embora o desempenho seja muito inferior àqueles das estratégias clássicas de equalização (filtragem linear), em que o baixo desempenho está associado à questão do valor do coeficiente f_0 do canal (atraso de decisão nulo) e às características de ruído.

Pode-se imaginar algumas possíveis melhoras no próprio algoritmo SOFVA, para tentar solucionar o problema de encontrar os *vales* na fdp do sinal recebido, principalmente atuando no parâmetro κ , entretanto outros aspectos da própria construção da função de interpolação, tal como as transições rápidas que geram uma estreita faixa de ruído para que a NPS seja capaz de realizar a *correta* separação, limitam a técnica. Uma outra possível solução seria a utilização de uma estratégia *semi-cega* para treinamento da RNA, mas essa alternativa não foi testada.

Finalmente, a proposição esclarece alguns pontos importantes sobre a aplicação de uma estrutura neural para equalização autodidata, buscando fornecer uma série de elementos como contribuição para a análise e utilização de estruturas neurais como equalizadores.

A seguir, a equalização será estruturada baseando-se em um novo critério cujo funcional

reflete o desejo de ser resgatada a função de densidade de probabilidade dos dados transmitidos.

CRITÉRIO ESTATÍSTICO BASEADO NA ESTIMAÇÃO DA FUNÇÃO DE DENSIDADE DE PROBABILIDADE

CRITÉRIO estatístico para a otimização de parâmetros é uma ferramenta bastante desenvolvida na aplicação de equalização adaptativa. Um grande exemplo são as técnicas de Bussgang em que é assumida uma determinada fdp para o sinal de saída do equalizador.

Tais técnicas utilizam o modelo da fdp da saída do equalizador para a construção do estimador capaz de gerar um símbolo que exerça o papel de referência no cálculo do erro, necessário para a adaptação dos parâmetros.

A proposta deste capítulo consiste em um novo critério baseado na estimativa da fdp do sinal na saída do equalizador através de uma fdp assumida para os dados transmitidos. A minimização da divergência de Kullback-Leibler é então utilizada na construção do funcional a ser otimizado.

A organização deste capítulo é feita da seguinte forma: seção 7.1 aborda alguns dos critérios que embasam a compreensão da proposição; a seção 7.2 descreve a proposição com as justificativas matemáticas para o modelo; seção 7.3 apresenta as simulações exemplificando o desempenho da proposta frente a outras técnicas e finalmente a seção 7.4 apresenta as conclusões.

7.1 Características do sinal na saída do equalizador

A realização da equalização de um canal de comunicação através de uma estratégia adaptativa requer a minimização (ou maximização) de algum funcional que conduza à recuperação dos dados transmitidos.

Nas estratégias autodidatas, as técnicas de Bussgang são bastante conhecidas. Elas empregam o conhecimento da fdp dos dados transmitidos e um modelo para os sinais na saída do equalizador realizando uma estimativa de máxima verossimilhança entre as duas fdps através de uma função não linear sem memória. Esta estimativa é de fundamental importância para a eficiência do critério.

Em [Sala-Alvarez & Vázquez-Grau, 1997], um critério baseado na estimação da fdp do sinal na saída do equalizador é proposto. Naquele trabalho, a confecção de um funcional de referência estatística é investigado e comparado com alguns dos algoritmos pertencentes à família de Bussgang. É mostrado que o algoritmo decorrente produz uma melhor velocidade de convergência e um desempenho em regime um pouco superior.

O algoritmo SOFVA, proposto no capítulo 6, utiliza o conhecimento do modelo paramétrico da fdp do sinal recebido (saída do canal) para descobrir suas características (*vales* entre as gaussianas), adaptando os parâmetros θ_i de maneira a localizá-los nos vales. Mas, devido ao número de gaussianas e suas localizações serem função da RI do canal, que é desconhecida, não é possível extrair o sinal transmitido apenas com esta estratégia.

Entretanto, utilizando o mesmo conceito, pode-se recuperar a informação se for conhecida a fdp do sinal na saída do equalizador, implementado por uma função linear (filtro transversal). A seguir é mostrada a dedução da fdp do sinal na saída do equalizador.

7.1.1 fdp do sinal na saída do equalizador

Sendo um sistema de comunicação digital representado em seu MBBE na Figura 7.1.

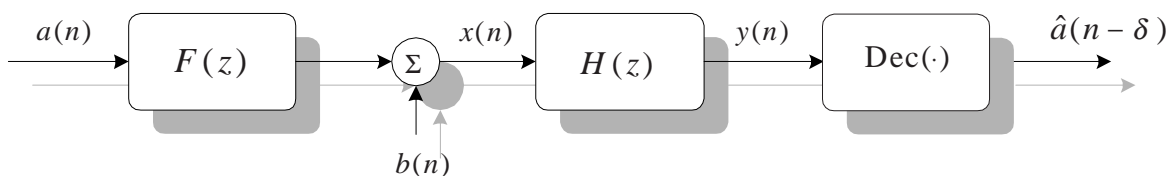


Figura 7.1: Modelo em banda básica equivalente do sistema digital.

Considerando que o equalizador funcione idealmente (sem ruído), tem-se $F(z) \cdot H(z) = z^{-\delta}$, em que $F(z)$ é a FT do canal, $H(z)$ é a RI do equalizador e δ é o atraso de decisão. O equalizador que realiza tal aproximação é o equalizador ideal (sem ruído). Desta maneira será utilizada a notação $\mathbf{h}_{\text{ideal}}$ para o vetor de seus coeficientes.

Sendo o sinal na saída do equalizador dado por:

$$y(n) = \mathbf{x}^T(n) \mathbf{h}_{\text{ideal}}, \quad (7.1)$$

em que:

$$\mathbf{x}(n) = \mathbf{F}^T \mathbf{a}(n) + \mathbf{b}(n), \quad (7.2)$$

e ainda:

$$\mathbf{a}(n) = \begin{bmatrix} a(n) \\ a(n-1) \\ \vdots \\ a(n-N+M+2) \end{bmatrix}, \quad \mathbf{b}(n) = \begin{bmatrix} b(n) \\ b(n-1) \\ \vdots \\ b(n-N+M+2) \end{bmatrix},$$

e \mathbf{F} é a matriz de convolução definida como:

$$\mathbf{F} = \begin{bmatrix} f_0 & & \mathbf{0} \\ \vdots & \ddots & \\ f_{N-1} & \ddots & f_0 \\ & \ddots & \vdots \\ \mathbf{0} & & f_{N-1} \end{bmatrix}_{(N+M-1) \times (M)}. \quad (7.3)$$

Assim ao reescrever-se a Equação (7.1) utilizando a Equação (7.2):

$$\begin{aligned} y(n) &= (\mathbf{F}^T \mathbf{a}(n) + \mathbf{b}(n))^T \mathbf{h}_{\text{ideal}} \\ &= \mathbf{a}^T(n) \mathbf{F} \mathbf{h}_{\text{ideal}} + \mathbf{b}^T(n) \mathbf{h}_{\text{ideal}} \\ &= \mathbf{a}^T(n) \underbrace{\mathbf{F} \mathbf{h}_{\text{ideal}}}_{\mathbf{g}_{\text{ideal}}} + \mathbf{b}^T(n) \mathbf{h}_{\text{ideal}} \\ &= \mathbf{a}^T(n) \mathbf{g}_{\text{ideal}} + \vartheta(n) \\ &= a(n-\delta) + \vartheta(n), \end{aligned} \quad (7.4)$$

em que é considerado $\vartheta(n)$ uma v.a. gaussiana de média nula. Esta consideração tem sustentação no fato de que uma combinação linear de variável gaussiana é também uma variável gaussiana.

Uma outra consideração que pode ser feita para a obtenção do mesmo resultado da Equação (7.4) é através da substituição do equalizador ideal por um equalizador não ideal em que a interferência residual apresente uma distribuição gaussiana. Essa aproximação é a mesma feita nas técnicas de Bussgang.

Então, ao calcular-se a fdp de y e considerando que $a(n-\delta)$ pertence ao alfabeto de símbolos transmitidos \mathcal{A} , tem-se:

$$p_{Y, \text{ideal}}(y) = \frac{1}{\sqrt{2\pi\sigma_\vartheta^2}} \sum_{i=1}^S \exp\left(-\frac{|y(n) - a_i|^2}{2\sigma_\vartheta^2}\right) \Pr(\mathbf{a}(n) = \mathbf{a}_i), \quad (7.5)$$

em que a variância de cada uma das gaussianas é dada por:

$$\sigma_{\vartheta}^2 = \sigma_b^2 \cdot \sum_i |h_i|^2, \quad (7.6)$$

em que os termos h_i são os coeficientes da RI do filtro equalizador. É interessante observar da Equação (7.6), que σ_b^2 pode ser estimada a partir da estimação de σ_{ϑ}^2 , impondo a normalização dos coeficientes do equalizador, ou seja, $\sigma_{\vartheta}^2 = \sigma_b^2$ se $\sum |h_i|^2 = 1$.

7.2 Critério estatístico para equalização

Uma vez conhecida a fdp do sinal equalizado, ao analisar-se de um ponto de vista de *classificação* e, novamente, tomando como rótulos das classes os símbolos do alfabeto de transmissão \mathcal{A} , a tarefa de equalização consiste na separação destas classes.

Uma maneira natural de tentar realizar a tarefa de separação é *minimizando a dispersão intraclasse*. A dispersão intraclasse pode ser entendida como sendo a variação dos sinais da mesma classe em torno do baricentro [Montalvão, 2000]. Esta dispersão é responsável por uma possível má classificação das classes de observações. Uma minimização da dispersão intraclasse pode ser obtida através da estimação das características da fdp do sinal de saída do equalizador.

Neste caso, a exemplo do algoritmo SOFVA apresentado no capítulo anterior, a separação das classes é conseguida se for possível identificar os *vales* entre as gaussianas de $p_{Y,\text{ideal}}(y)$.

Utilizando-se novamente o critério de medida de similaridade entre funções da Equação (6.27):

$$\int_{-\infty}^{\infty} f(x) \cdot g(x) dx,$$

deseja-se encontrar as características de $p_{Y,\text{ideal}}(y)$.

Ao idealizar-se funções que sejam capazes de executar a tarefa de identificação dos vales através do *contraste*, pode-se escolher funções como aquelas representadas na Figura 7.2.

É interessante ressaltar que a função da Figura 7.2(a) é aquela utilizada para a elaboração do algoritmo SOFVA.

Entretanto, no estudo de simulações, tais funções não se mostraram capazes de realizar a tarefa desejada de forma aceitável. Acredita-se que o motivo desta dificuldade deva ser a não penalização de termos que estejam fora da área de interesse, uma vez que tais funções existem no intervalo $[-\infty, \infty]$.

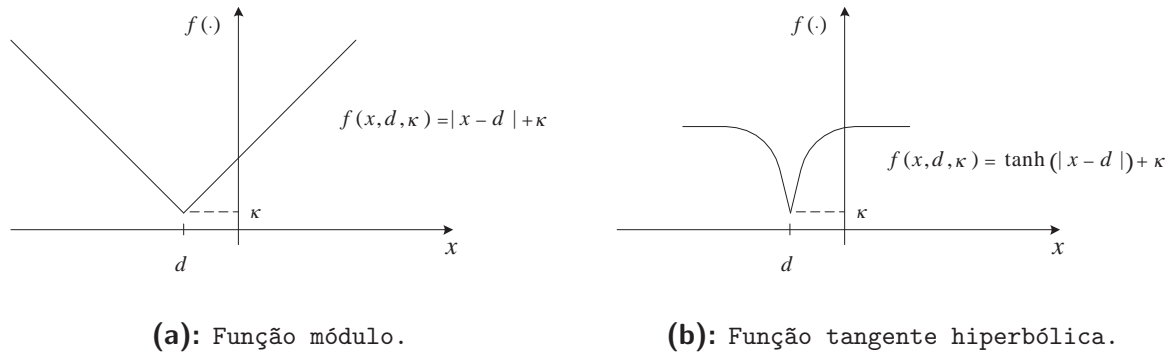


Figura 7.2: Funções de contraste.

Desta maneira, a função de interesse deve ser capaz de, além de encontrar as características (*vales*) da fdp do sinal de saída do equalizador, deve também penalizar as componentes fora da região de interesse. Assim, é razoável admitir que a função seja constituída de um somatório de gaussianas, tal como $p_{Y,\text{ideal}}(y)$. Esta função deverá então procurar o ponto de mínima distorção.

A mínima distorção é então dada quando o funcional:

$$J_{\text{ideal}}(\mathbf{h}_{\text{ideal}}, \sigma_{\vartheta}^2, \mathbf{h}, \sigma_r^2) = - \int p_{Y,\text{ideal}}(y, \mathbf{h}_{\text{ideal}}, \sigma_{\vartheta}^2) \cdot f(y, \mathbf{h}, \sigma_r^2) dy, \quad (7.7)$$

for mínimo, o que significa $p_{Y,\text{ideal}}(y, \mathbf{h}_{\text{ideal}}, \sigma_{\vartheta}^2) = f(y, \mathbf{h}, \sigma_r^2)$.

Assim, igualando-se as funções pode-se garantir que a equalização é obtida [Benveniste et al., 1980].

Novamente, busca-se uma medida de similaridades entre duas funções, ambas fdps. Então uma maneira imediata de comparação de fdps é através da medida de divergência de Kullback-Leibler (KLD).

Tomando como *função de referência* uma determinada função $\Phi(y)$ que é a fdp do sinal de saída do equalizador, tem-se:

$$\begin{aligned} D_{p_{Y,\text{ideal}}(y)||\Phi(y)} &= \int_{-\infty}^{\infty} p_{Y,\text{ideal}}(y) \cdot \ln \left(\frac{p(y)}{\Phi(y)} \right) dy \\ &= \int_{-\infty}^{\infty} p_{Y,\text{ideal}}(y) \cdot \ln(p(y)) dy - \int_{-\infty}^{\infty} p(y) \cdot \ln(\Phi(y)) dy, \end{aligned} \quad (7.8)$$

em que $y = \mathbf{h}^T \mathbf{x}$ e

$$\Phi(y) = \underbrace{\frac{1}{\sqrt{2\pi\sigma_r^2}}}_A \cdot \sum_{i=1}^S \exp \left(-\frac{|y - a_i|^2}{2\sigma_r^2} \right). \quad (7.9)$$

Desta maneira, minimizando o segundo termo corresponde à encontrar a entropia de y se $\Phi(y)$ for a verdadeira fdp [Bishop, 1995, p. 59].

O funcional é então dado por:

$$\begin{aligned} J(\mathbf{h}) &= \int_{-\infty}^{\infty} p_{Y,\text{ideal}}(y) \cdot \ln \left(\frac{1}{\Phi(y)} \right) dy \\ &= -E \{ \ln [\Phi(y)] \} \\ &= -E \left\{ \ln \left[A \cdot \sum_{i=1}^S \exp \left(-\frac{|y - a_i|^2}{2\sigma_r^2} \right) \right] \right\}. \end{aligned} \quad (7.10)$$

Este critério será denominado **“Fitting pdf Criterion” (FPC)** e será representado por J_{FPC} .

Para a obtenção das expressões de adaptação é necessário o gradiente estocástico, que é então dado por:

$$\nabla J_{\text{FPC}}(\mathbf{h}) = \frac{\sum_{i=1}^S \exp \left(-\frac{|y - a_i|^2}{2\sigma_r^2} \right) (y - a_i)}{\sigma_r^2 \cdot \sum_{i=1}^S \exp \left(-\frac{|y - a_i|^2}{2\sigma_r^2} \right)} \mathbf{x}^*, \quad (7.11)$$

de modo que a equação de adaptação é então escrita como:

$$\mathbf{h}(n+1) = \mathbf{h}(n) - \mu_h \nabla J_{\text{FPC}}(\mathbf{h}). \quad (7.12)$$

Este algoritmo de adaptação será denominado por **“Fitting pdf Algorithm” (FPA)**.

Observe que o parâmetro σ_r^2 funciona como o parâmetro de *temperatura* no “annealing” [Rose, 1998; Montalvão, 2000]. Isto porque σ_r^2 funciona como um parâmetro de *esfriamento* do sistema controlando a velocidade de convergência.

A Figura 7.3 ilustra o funcional $J_{\text{FPC}}(\mathbf{h})$ para um filtro de ordem 2, modulação BPSK e um canal cuja RI é dada por $\mathbf{f} = [1 \ 0.6]^T$. A Figura 7.4 mostra as curvas de nível para o mesmo caso.

O FPC consiste na estimação paramétrica do sinal na saída do equalizador. Neste caso os parâmetros ajustáveis são a média e a variância das gaussianas. Uma pequena diferença do processo de estimação paramétrica da fdp em relação à máxima verossimilhança é que o critério utilizado para encontrar os valores dos parâmetros é a minimização da entropia utilizando a KLD.

Uma abordagem a partir da teoria da classificação pode ser feita para comparar o critério proposto com outros existentes para equalização autodidata. A próxima seção aborda este aspecto.

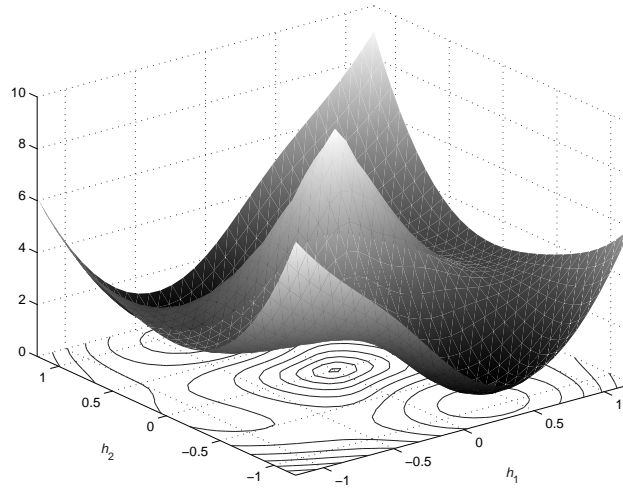


Figura 7.3: Funcional $J_{\text{FPC}}(\mathbf{h})$ em função dos parâmetros h_1 e h_2 do equalizador.

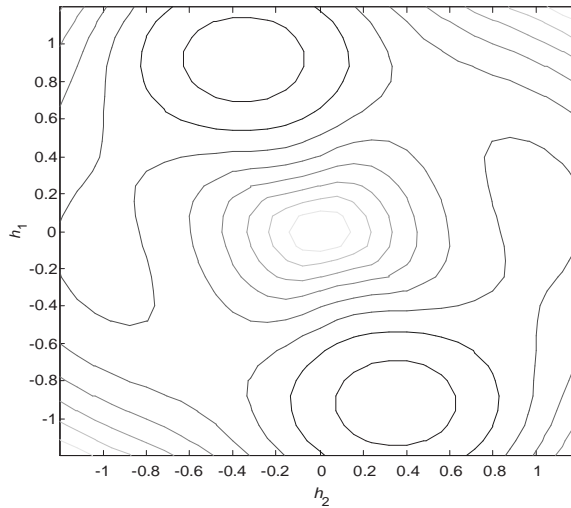


Figura 7.4: Curvas de nível para o funcional $J_{\text{FPC}}(\mathbf{h})$ em função dos parâmetros h_1 e h_2 do equalizador.

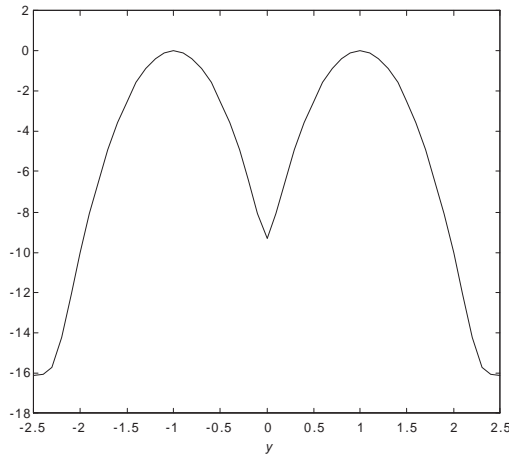
7.2.1 Critérios cegos: abordagem de classificação

Dos algoritmos existentes em equalização cega, aqueles baseados no critério CM são, provavelmente, os mais populares. Desta família, o CMA, ou CM^{2-2} [Haykin, 2000], é membro mais conhecido. Na literatura, vários trabalhos, como por exemplo [Sala-Alvarez & Vázquez-Grau, 1997; Touzni, 1998; Regalia, 1999] dentre outros, mostraram a ligação de vários critérios com o CM. Além disto, várias publicações discutindo seus limites e garantias de convergência tornam interessante também mostrar uma ligação do FPC com o critério CM.

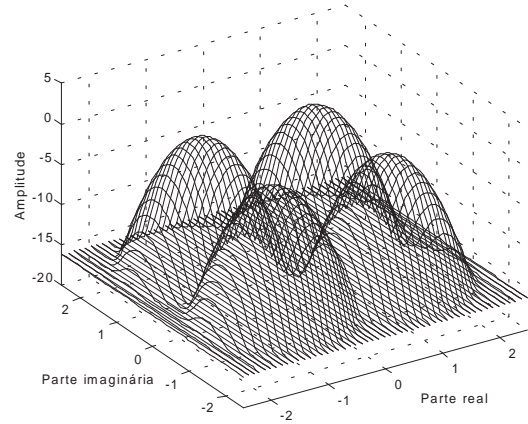
Do ponto de vista da classificação, estimar parametricamente a fdp do sinal na saída

do equalizador através do FPC, consiste em ajustar os parâmetros do filtro para que os dados na saída do equalizador sejam *adequados* sob a função $\ln(\Phi(y))$.

A Figura 7.5 ilustra a função $\ln(\Phi(y))$ para alguns casos de modulação. Foi utilizado $\sigma_r^2 = 0.05$.



(a): Modulação BPSK.



(b): Modulação 4QAM.

Figura 7.5: Função $\ln(\Phi(y))$ para alguns tipos de modulação.

Então, pode-se, pela mesma abordagem observar o funcional do critério CM.

Como o funcional CM é dado por $J_{\text{CM}}(\mathbf{h}) = \mathbb{E} \{ (|y|^p - R_p)^2 \}$, deve-se então procurar uma função de referência, ou *função alvo*, de tal maneira que o critério CM seja equivalente ao FPC.

Então, escrevendo:

$$J_{\text{CM}}(\mathbf{h}) = -\mathbb{E} \{ \ln(f_{\text{CM}}(y, \mathbf{h}, p)) \}, \quad (7.13)$$

deve-se obter:

$$f_{\text{CM}}(y, \mathbf{h}, p) = \frac{\exp [(|y|^p - R_p)^2]}{\int_{-\infty}^{\infty} \exp (|\xi|^p - 1) d\xi}, \quad (7.14)$$

para que a Equação (7.13) seja satisfeita. O termo no denominador é utilizado para garantir que a função tenha módulo unitário.

Então pode-se mostrar uma equivalência entre o FPC e os critérios CM: $f_{\text{CM}}(y, \mathbf{h}, 1)$ (critério de Sato) e $f_{\text{CM}}(y, \mathbf{h}, 2)$ (critério de Godard), sendo que a melhor aproximação ocorre para uma escolha apropriada do valor σ_r^2 .

A Figura 7.6 mostra as funções alvo para o FPC, o critério de Sato e o critério de Godard para a modulação BPSK. Observa-se que a função alvo do FPC varia para os

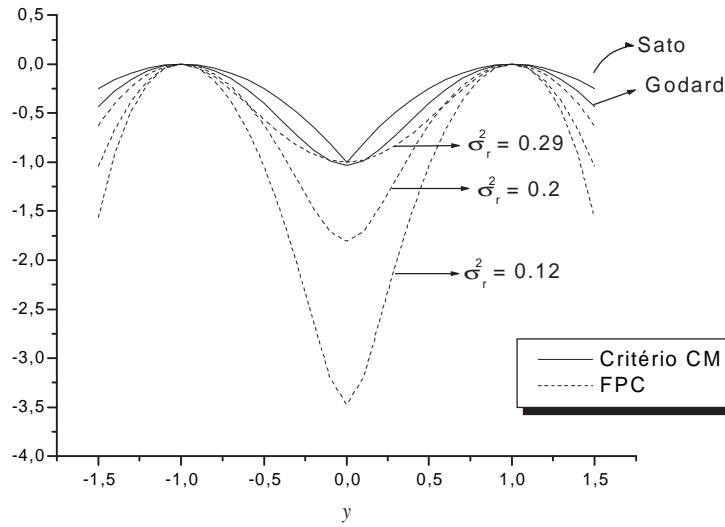


Figura 7.6: Comparação entre funções alvo para os critérios de Sato, Godard e FPC.

valores de σ_r^2 e que a equivalência entre os critérios FPC, Sato e Godard ocorre para um valor $\sigma_r^2 \cong 0.3$.

Uma importante diferença entre os critérios CM e o FPC é a não necessidade de utilização de um dispositivo de correção de fase por parte do FPC, e necessário no critério CM. Isso se deve ao fato de que o cálculo da adaptação do FPC, Equação (7.11), considera o desvio de fase, enquanto que no critério CM apenas o módulo é considerado.

!

7.3 Simulações computacionais

Simulações computacionais foram realizadas para avaliar o desempenho do FPA, comparado com o desempenho do CMA.

As simulações foram mediadas sobre 100 simulações de Monte Carlo para a evolução do DSE. Também para efeito de suavização das curvas foi, após as mediações, aplicado um filtro passa-baixas de ordem 1 e $\omega_c = 10^{-2}$.

Nos casos em que se fez necessário a utilização de um dispositivo de recuperação de fase para o CMA, o critério de minimização do erro de decisão quadrático $J(\psi)$ foi adotado, em que ψ é a correção de fase mostrada na Figura 7.7.

Então, tem-se:

$$J(\psi) = E \{ |\hat{a}(n) - y(n)|^2 \}, \quad (7.15)$$

resultando-se numa equação de adaptação dada pela Equação (7.16) [Cavalcanti, 1996]:

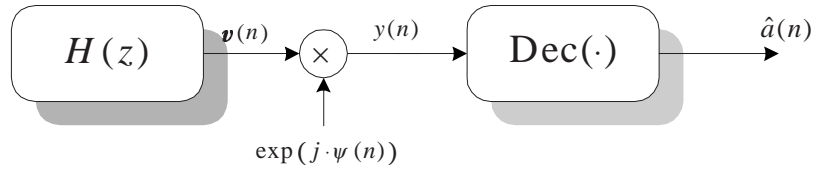


Figura 7.7: Esquema de recuperação de fase em um sistema com equalizador.

$$\psi(n+1) = \psi(n) + \mu_\psi \cdot \text{Im} \{ [\hat{a}(n) - y(n)] \cdot y^*(n) \}. \quad (7.16)$$

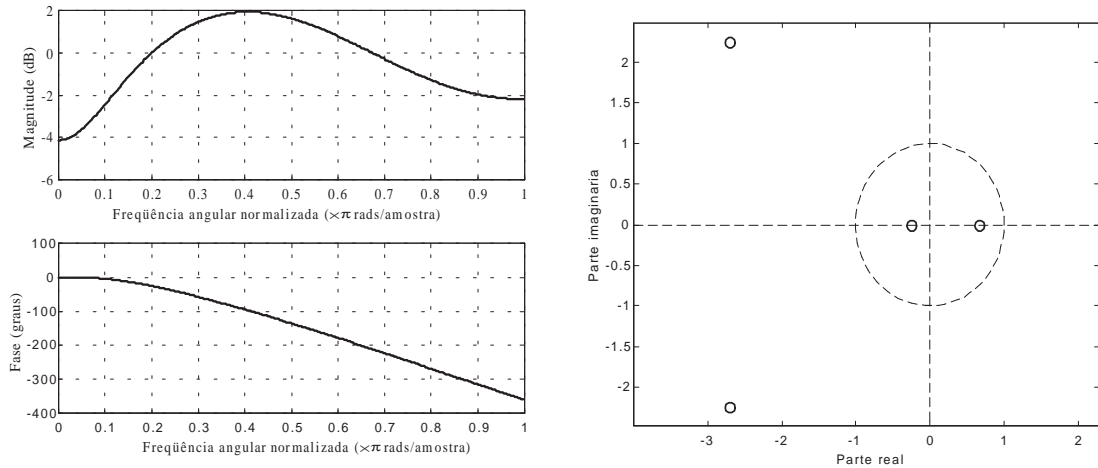
Os parâmetros de simulação de cada um dos experimentos encontram-se descritos nas tabelas após os resultados.

Experimento 1 - Canal de fase não mínima com $\text{MD}(\mathbf{f}) > 1$:

O canal apresenta a seguinte RI:

$$\mathbf{f}_5 = [0.1 \quad 0.5 \quad 1 \quad -0.6 \quad -0.2]^T. \quad (7.17)$$

A Figura 7.8 ilustra as resposta em frequência (magnitude e fase) e o diagrama de zeros do canal.



(a): Resposta em frequência.

(b): Diagrama de zeros.

Figura 7.8: Características do canal \mathbf{f}_5 - resposta em frequência e diagrama de zeros.

Facilmente pode-se ver que o canal não apresenta uma forte distorção de fase, concentrando-se na distorção de amplitude (magnitude).

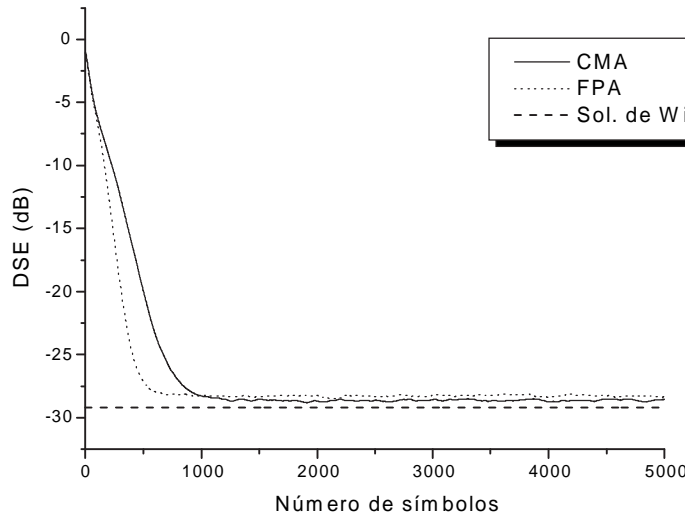


Figura 7.9: Evolução do DSE para o canal f_5 - modulação BPSK e SNR = 30 dB.

A Figura 7.9 ilustra o desempenho em termos da evolução do DSE.

Nota-se aqui um desempenho similar entre o CMA e o FPA embora a velocidade de convergência seja maior por parte do FPA. Isso se deve ao parâmetro livre σ_r^2 que permite um melhor ajuste para aumentar a velocidade de convergência. Em [Montalvão et al., 2000] foi analisado o caso da escolha do valor de σ_r^2 para que o desempenho final do CMA e do FPA fosse equivalente sem preocupação com a velocidade de convergência.

Uma outra verificação importante é a maior velocidade de convergência do FPA. Isto deve-se a, no algoritmo, uma medida direta entre o sinal de saída do equalizador e os símbolos do alfabeto \mathcal{A} (Equação (7.11)) e não uma estimativa do símbolo real para tal cálculo.

A Tabela 7.1 resume os parâmetros de simulação para o experimento 1.

Parâmetros de simulação para o canal f_5	
<i>CMA</i>	$\mu_h = \mu_\psi = 10^{-3}$ Comprimento do filtro: 30 taps Inicialização: $\begin{cases} \mathbf{h} = [0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0]^T \\ \psi = 0 \end{cases}$
<i>FPA</i>	$\mu_h = 25 \cdot 10^{-4}$ $\sigma_r^2 = 0.15$ Comprimento do filtro: 30 taps Inicialização: $\mathbf{h} = [0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0]^T$

Tabela 7.1: Sumário da simulação para o canal f_5 .

Experimento 2 - Canal de fase não mínima com $MD(\mathbf{f}) > 1$:

O canal apresenta a seguinte RI:

$$\mathbf{f}_6 = \begin{bmatrix} 0.5679 & -0.1136 & 0.5849 & 0.1124 & 0.5556 \end{bmatrix}^T. \quad (7.18)$$

O canal proposto em questão apresenta as seguintes características, ilustradas na Figura 7.10.

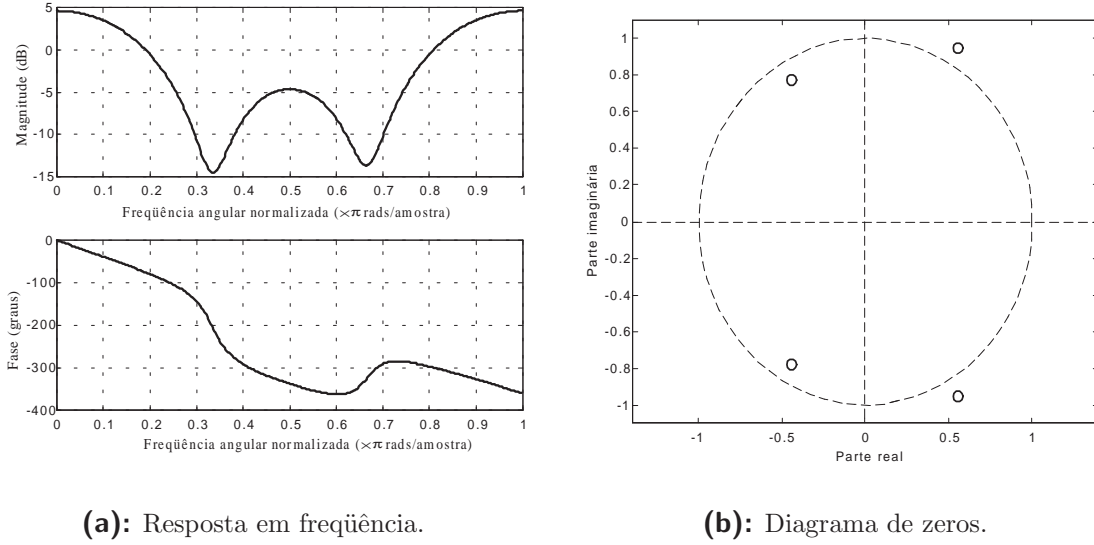


Figura 7.10: Características do canal \mathbf{f}_6 - resposta em frequência e diagrama de zeros.

Como visto, o canal apresenta uma forte distorção em amplitude e em fase. A Figura 7.11 ilustra a evolução do DSE.

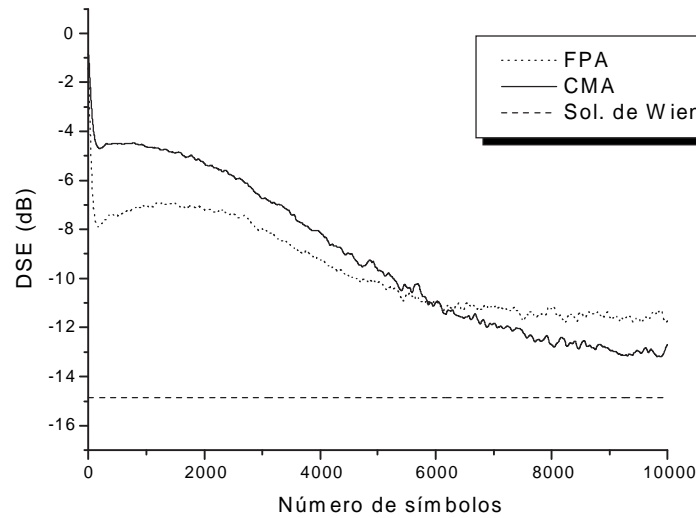


Figura 7.11: Evolução do DSE para o canal \mathbf{f}_6 - modulação BPSK e SNR = 30 dB.

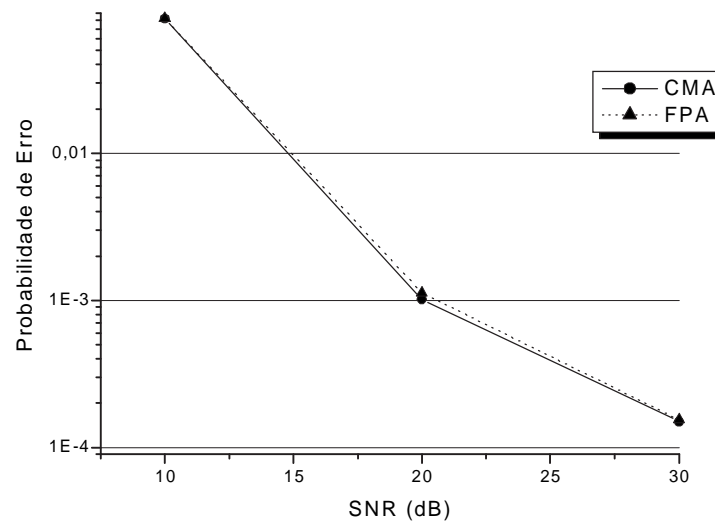


Figura 7.12: Curvas de taxa de símbolos errados para os algoritmos CMA e FPA - canal f_6 .

Como mencionado anteriormente, no caso de uma forte distorção de fase, o desempenho do FPA é ligeiramente pior que o do CMA, embora o FPA apresente uma maior velocidade de convergência.

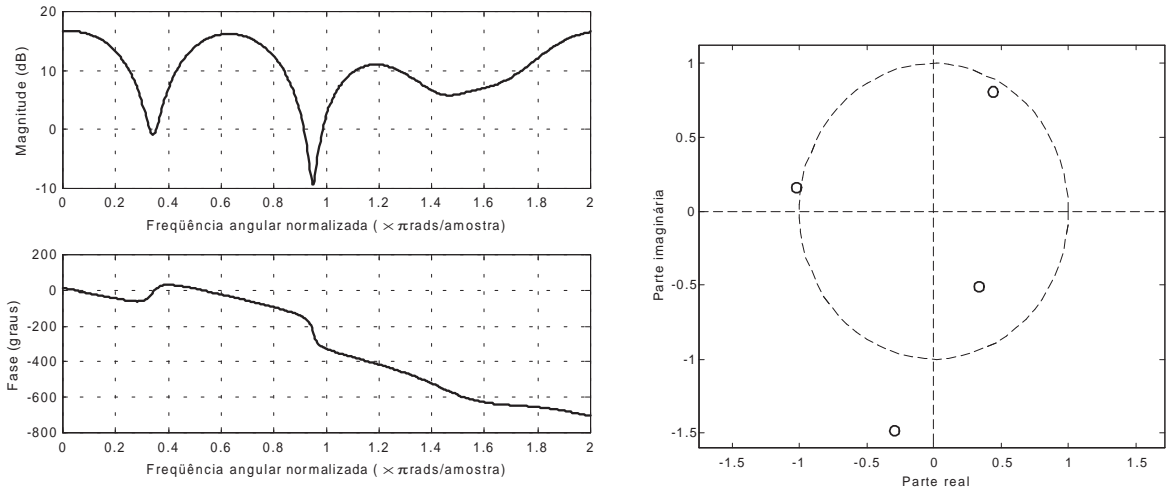
A Figura 7.12 ilustra o cálculo da taxa de símbolos errados (TSE) para os dois algoritmos. Pode-se verificar que o desempenho é praticamente o mesmo.

A Tabela 7.2 resume os parâmetros de simulação para o experimento 2.

Parâmetros de simulação para o canal f_6	
<i>CMA</i>	$\mu_h = \mu_\psi = 10^{-3}$ Comprimento do filtro: 30 taps Inicialização: $\begin{cases} \mathbf{h} = [0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0]^T \\ \psi = 0 \end{cases}$
<i>FPA</i>	$\mu_h = 25 \cdot 10^{-4}$ $\sigma_r^2 = 0.29$ Comprimento do filtro: 30 taps Inicialização: $\mathbf{h} = [0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0]^T$

Tabela 7.2: Sumário da simulação para o canal f_6 .

Experimento 3 - Canal de fase não mínima com $MD(\mathbf{f}) > 1$:



(a): Resposta em frequência.

(b): Diagrama de zeros.

Figura 7.13: Características do canal f_7 - resposta em frequência e diagrama de zeros.

O canal apresenta a seguinte RI:

$$\mathbf{f}_7 = \begin{bmatrix} 0.4854 - j \cdot 0.0971 \\ 0.3640 + j \cdot 0.4368 \\ 0.2427 \\ 0.2912 - j \cdot 0.3155 \\ 0.1941 + j \cdot 0.3883 \end{bmatrix}. \quad (7.19)$$

Neste experimento, o canal novamente apresenta uma forte distorção de fase e de magnitude como pode ser visto na Figura 7.13.

A Figura 7.14 ilustra a evolução do DSE para o canal f_7 .

O desempenho um pouco pior (em torno de 1 dB) do FPA pode ser novamente causado devido a forte distorção de fase imposta pelo canal. Neste caso o CMA também apresentou uma convergência mais rápida.

A Tabela 7.3 resume os parâmetros de simulação para o experimento 3.

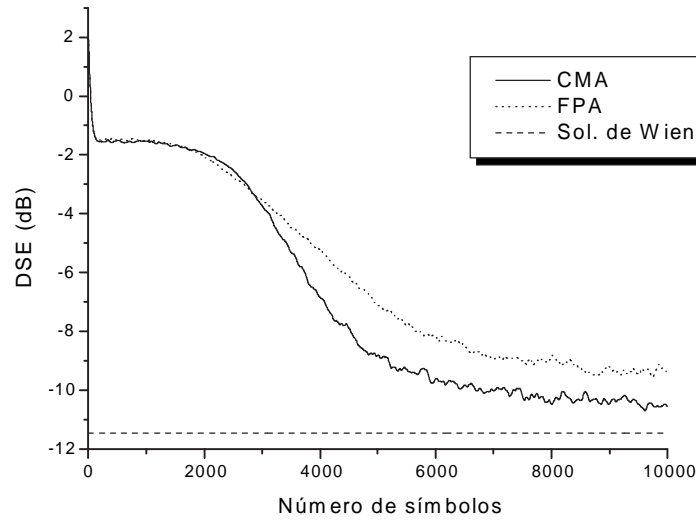


Figura 7.14: Evolução do DSE para o canal f_7 - modulação 4QAM e SNR = 30 dB.

Parâmetros de simulação para o canal f_7	
<i>CMA</i>	$\mu_h = \mu_\psi = 10^{-3}$ Comprimento do filtro: 30 taps Inicialização: $\begin{cases} \mathbf{h} = [0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0]^T \\ \psi = 0 \end{cases}$
<i>FPA</i>	$\mu_h = 25 \cdot 10^{-4}$ $\sigma_r^2 = 0.45$ Comprimento do filtro: 30 taps Inicialização: $\mathbf{h} = [0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0]^T$

Tabela 7.3: Sumário da simulação para o canal f_7 .

7.4 Síntese e conclusões

Neste capítulo foi apresentado um novo critério de equalização ideal baseado na referência estatística da fdp do sinal na saída do equalizador linear transversal.

Inicialmente são feitas algumas considerações sobre os critérios conhecidos e a dedução da fdp do sinal na saída de um equalizador suposto funcionando perfeitamente. Na sequência, é descrito a dedução do critério proposto (FPC).

Foi também discutida uma abordagem do ponto de vista da teoria da classificação tanto da proposição quanto de alguns dos critérios CM (Sato e Godard). Neste ponto, observou-se uma equivalência entre os critérios a partir de uma escolha adequada do parâmetro σ_r^2 , e deduziu-se um melhor desempenho do FPC visto ser um funcional capaz de melhor realizar a minimização da variância intraclasse da sequência de sinais na saída do equalizador.

Nas simulações computacionais a comparação fica evidente. Para o caso do canal com distorção de fase praticamente linear o desempenho do FPA é praticamente o mesmo do CMA com uma pequena variação, mas apresentando uma velocidade de convergência melhor.

Nos casos de uma forte distorção de fase, há uma perda de desempenho em relação ao CMA, mas é importante notar que em todos os casos a modulação apresenta o módulo constante, o que facilita a tarefa do critério CM. Acredita-se que para modulações onde a condição de módulo constante não é satisfeita pode haver um melhor desempenho por parte do FPA.

CONCLUSÕES E PERSPECTIVAS

POR se tratar de um campo extremamente vasto, a equalização autodidata experimenta uma também extensa lista de contribuições para a solução do problema, sem mostrar ainda uma resposta definitiva.

Embora a grande maioria das implementações utilizem estruturas de filtragem lineares e algoritmos de baixo custo computacional, várias das contribuições que têm surgido utilizam também estruturas de filtragem não lineares e critérios mais elaborados na busca de uma solução capaz de fornecer um melhor desempenho.

A proposta dessa dissertação se enquadra no contexto de investigação sobre uso de uma estrutura de filtragem não linear e também na proposição de um funcional mais elaborado.

Na investigação sobre a aplicabilidade de uma estrutura de filtragem neural, foi analisado sob a óptica da classificação uma estrutura preditiva com apenas uma entrada que fosse capaz de, utilizando estatísticas de ordem 2, equalizar canais de fase não mínima. Nessa tarefa, a análise das características do problema forneceram elementos para a concepção de um novo algoritmo auto organizado para a rede neural baseado em uma técnica, até onde se tem conhecimento, inédita.

Pode-se então listar as seguintes contribuições:

- ✓ abordagem do problema de equalização preditiva através da teoria de classificação;
- ✓ possibilidade da utilização de um modelo não linear com apenas uma entrada;
- ✓ divisão do aprendizado em duas etapas distintas: auto organizado e supervisionado;

- ✓ proposição de um novo algoritmo auto organizado para a rede tomando como base a minimização da divergência de Kullback-Leibler.

Essas contribuições embora forneçam alguma informação adicional, e um ganho no processamento de alguns tipos de canais como os canais de fase mínima apresentados, não são suficientes para fornecer uma pesquisa exaustiva sobre a estratégia proposta e algumas perspectivas de continuidade desse trabalho podem também ser listadas:

- estudo de estruturas de filtragem recursivas para viabilizar uma realimentação da informação tentando obter mais dados para realização da extração da informação desejada;
- estudo da aplicação de estratégias semi-cegas para equalização utilizando uma estrutura de filtragem neural;
- análise de uma combinação de retropreditor e preditor para buscar uma solução mais adequada para a equalização de canais de fase não mínima;
- estudo de algoritmos capazes de melhorar o aprendizado na etapa auto-organizada;
- aplicações em canais não lineares, os quais podem se beneficiar do uso de uma estrutura não linear.

Na parte do critério de referência estatística, a lista de contribuições é bem menor que a de perspectivas dado que se trata apenas de um trabalho inicial onde, acredita-se, pode-se ainda analisar vários aspectos.

Nas contribuições pode-se listar:

- ✓ a proposição de um novo critério de minimização para a utilização de estruturas de filtragem lineares que tem como base a estimativa da fdp real dos símbolos equalizados e também utiliza o conceito de minimização da entropia para a construção do funcional;
- ✓ abordagem do ponto de vista da classificação dos critérios CM e sua ligação com o critério proposto buscando, a exemplo de outros trabalhos, um suporte sobre aspectos bastante sedimentados no campo dos algoritmos cegos.

E nas perspectivas, que buscam refletir algumas das preocupações iniciais que surgiram com o estudo iniciado nessa dissertação, pode-se listar:

- análise de convergência do algoritmo para mínimos locais;
- estudo comparativo sobre a utilização de esquemas de modulação mais completos;

- utilização de um critério de otimização para o parâmetro σ_r^2 , pois concluiu-se que ele executa um papel fundamental no desempenho do algoritmo;
- utilização em cascata de uma realimentação das decisões e estudo comparativo com o DFE clássico;
- aplicação em detecção multiusuário;
- aplicação em canais não lineares.

Finalmente, pode-se também concluir que o campo de classificação pode promover algum avanço sobre as estratégias clássicas de equalização e, a partir de resultados recentes, já fornece uma abordagem bastante diferente e também elucidativa sobre algumas características do problema de equalização.

PARTE **III**

APÊNDICES

DERIVAÇÃO DE FUNÇÕES ESCALARES EM RELAÇÃO A UM VETOR

O estudo da filtragem adaptativa tem por necessidade a minimização de uma função custo em relação a um vetor de parâmetros (coeficientes do filtro). Faz-se necessário então, o conhecimento de algumas regras de derivação em relação a um vetor, real ou complexo.

Na seção A.1 são apresentadas os conceitos básicos sobre derivação em relação a vetores. Na seção A.2 faz-se uso dos conceitos previamente descritos e mostra-se alguns resultados particulares. Finalmente na seção A.3 é mostrada a relação entre a derivada em relação a um vetor e o vetor gradiente.

A.1 Definições básicas

Seja \mathbf{h} um vetor, onde pode-se escrever em termos de sua k -ésima componente:

$$h_k = h_{k,R} + j \cdot h_{k,I}$$

em que os índices subscritos R e I indicam, respectivamente, as componentes real e imaginária.

Então as partes real ($h_{k,R}$) e imaginária ($h_{k,I}$) podem ser escritas como:

$$\begin{aligned} h_{k,R} &= \frac{1}{2} \cdot (h_k + h_k^*) \\ h_{k,I} &= \frac{1}{2} \cdot (h_k - h_k^*) \end{aligned}$$

logo, pode-se representar a derivada em relação à k -ésima componente do vetor \mathbf{h} como:

$$\begin{aligned}\frac{\partial}{\partial h_k} &= \frac{1}{2} \cdot \left(\frac{\partial}{\partial h_{k,R}} - j \cdot \frac{\partial}{\partial h_{k,I}} \right) \\ \frac{\partial}{\partial h_k^*} &= \frac{1}{2} \cdot \left(\frac{\partial}{\partial h_{k,R}} + j \cdot \frac{\partial}{\partial h_{k,I}} \right)\end{aligned}\tag{A.1}$$

e sendo y um escalar complexo, a sua derivada em relação ao escalar complexo h é dada pela seguinte equação:

$$\begin{aligned}\frac{\partial y}{\partial h} &= \frac{\partial y_R}{\partial h_R} - \frac{\partial y_I}{\partial h_I} + j \cdot \left(\frac{\partial y_R}{\partial h_I} + \frac{\partial y_I}{\partial h_R} \right) \\ \frac{\partial y^*}{\partial h} &= \frac{\partial y_R}{\partial h_R} + \frac{\partial y_I}{\partial h_I} + j \cdot \left(\frac{\partial y_R}{\partial h_I} - \frac{\partial y_I}{\partial h_R} \right)\end{aligned}\tag{A.2}$$

As derivadas definidas na Equação (A.1) satisfazem as seguintes condições:

$$\frac{\partial h_k}{\partial h_k} = 1 \quad \text{e} \quad \frac{\partial h_k}{\partial h_k^*} = \frac{\partial h_k^*}{\partial h_k} = 0$$

(Uma função analítica satisfaz $\frac{\partial f}{\partial z^*} = 0 \quad \forall z$ [Oppenheim & Schafer, 1989].)

Então considerando que o vetor \mathbf{h} tenha M componentes ($M \times 1$) tem-se:

$$\frac{\partial}{\partial \mathbf{h}} = \frac{1}{2} \begin{bmatrix} \frac{\partial}{\partial h_{0,R}} - j \cdot \frac{\partial}{\partial h_{0,I}} \\ \frac{\partial}{\partial h_{1,R}} - j \cdot \frac{\partial}{\partial h_{1,I}} \\ \vdots \\ \frac{\partial}{\partial h_{M-1,R}} - j \cdot \frac{\partial}{\partial h_{M-1,I}} \end{bmatrix} \quad \text{e} \quad \frac{\partial}{\partial \mathbf{h}^*} = \frac{1}{2} \begin{bmatrix} \frac{\partial}{\partial h_{0,R}} + j \cdot \frac{\partial}{\partial h_{0,I}} \\ \frac{\partial}{\partial h_{1,R}} + j \cdot \frac{\partial}{\partial h_{1,I}} \\ \vdots \\ \frac{\partial}{\partial h_{M-1,R}} + j \cdot \frac{\partial}{\partial h_{M-1,I}} \end{bmatrix}\tag{A.3}$$

A.2 Alguns casos particulares

(a) $\frac{\partial(\mathbf{h}^T \mathbf{x})}{\partial \mathbf{h}}$

Pod-se reescrever:

$$\mathbf{h}^T \mathbf{x} = \sum_i h_i x_i$$

Devido a linearidade dos operadores pod-se inverter a ordem:

$$\frac{\partial(\mathbf{h}^T \mathbf{x})}{\partial \mathbf{h}} = \frac{\partial}{\partial \mathbf{h}} \sum_i h_i x_i = \sum_i \frac{\partial(h_i x_i)}{\partial h_i}$$

e como:

$$h_i x_i = (h_{i,R} x_{i,R} - h_{i,I} x_{i,I}) + j \cdot (h_{i,R} x_{i,I} + h_{i,I} x_{i,R}) \quad (\text{A.4})$$

e aplicando (A.3) em (A.4) tem-se:

$$\frac{\partial (h_i x_i)}{\partial h_i} = 2x_i$$

Logo,

$$\frac{\partial (\mathbf{h}^T \mathbf{x})}{\partial \mathbf{h}} = 2\mathbf{x} \quad (\text{A.5})$$

(b) $\frac{\partial (\mathbf{x}^H \mathbf{h}^*)}{\partial \mathbf{h}}$

Reescrevendo o argumento:

$$\mathbf{x}^H \mathbf{h}^* = \sum_i x_i^* h_i^*$$

Usando o fato da linearidade dos operadores tem-se:

$$\frac{\partial (\mathbf{x}^H \mathbf{h}^*)}{\partial \mathbf{h}} = \frac{\partial}{\partial \mathbf{h}} \sum_i x_i^* h_i^* = \sum_i \frac{\partial (x_i^* h_i^*)}{\partial h_i}$$

Sabe-se que,

$$x_i^* h_i^* = (x_{i,R} h_{i,R} - x_{i,I} h_{i,I}) - j \cdot (x_{i,R} h_{i,I} + x_{i,I} h_{i,R}) \quad (\text{A.6})$$

Pode-se também observar,

$$x_i^* h_i^* = (x_i h_i)^*$$

e então aplicando (A.3) em (A.6) tem-se:

$$\frac{\partial (x_i h_i)^*}{\partial h_i} = 0$$

Logo,

$$\frac{\partial (\mathbf{x}^H \mathbf{h}^*)}{\partial \mathbf{h}} = \mathbf{0} \quad (\text{A.7})$$

(c) $\frac{\partial (\mathbf{x}^H \mathbf{h})}{\partial \mathbf{h}}$

Reescrevendo o argumento:

$$\mathbf{x}^H \mathbf{h} = \sum_i x_i^* h_i$$

Utilizando a linearidade dos operadores:

$$\frac{\partial (\mathbf{x}^H \mathbf{h})}{\partial \mathbf{h}} = \frac{\partial}{\partial \mathbf{h}} \sum_i x_i^* h_i = \sum_i \frac{\partial (x_i^* h_i)}{\partial h_i}$$

E como,

$$x_i^* h_i = (x_{i,R} h_{i,R} + x_{i,I} h_{i,I}) + j \cdot (x_{i,R} h_{i,I} - x_{i,I} h_{i,R}) \quad (\text{A.8})$$

Utilizando-se então, (A.3) em (A.8) obtém-se:

$$\frac{\partial (x_i^* h_i)}{\partial h_i} = 2x_i^*$$

Logo,

$$\frac{\partial (\mathbf{x}^H \mathbf{h})}{\partial \mathbf{h}} = 2\mathbf{x}^* \quad (\text{A.9})$$

(d) $\frac{\partial (\mathbf{h}^H \mathbf{x})}{\partial \mathbf{h}}$

Reescrevendo o argumento:

$$\mathbf{h}^H \mathbf{x} = \sum_i h_i^* x_i$$

Fazendo uso da linearidade dos operadores tem-se:

$$\frac{\partial (\mathbf{h}^H \mathbf{x})}{\partial \mathbf{h}} = \frac{\partial}{\partial \mathbf{h}} \sum_i h_i^* x_i = \sum_i \frac{\partial (h_i^* x_i)}{\partial h_i}$$

Sabendo que,

$$h_i^* x_i = (h_{i,R} x_{i,R} + h_{i,I} x_{i,I}) + j \cdot (h_{i,R} x_{i,I} - h_{i,I} x_{i,R}) \quad (\text{A.10})$$

e utilizando (A.3) em (A.10) tem-se:

$$\frac{\partial (h_i^* x_i)}{\partial h_i} = 0$$

Logo,

$$\frac{\partial (\mathbf{h}^H \mathbf{x})}{\partial \mathbf{h}} = \mathbf{0} \quad (\text{A.11})$$

(e) $\frac{\partial(\mathbf{h}^H \mathbf{R} \mathbf{h})}{\partial \mathbf{h}}$ onde \mathbf{R} é uma matriz quadrada de ordem $(M \times M)$

Considerando:

$$g = \mathbf{h}^H \mathbf{R} \mathbf{h} \quad , \quad \mathbf{g}_1 = \mathbf{R}^H \mathbf{h} \quad \text{e} \quad \mathbf{g}_2 = \mathbf{R} \mathbf{h}$$

obtem-se:

(i) de (c):

$$\frac{\partial(\mathbf{g}_1^H \mathbf{h})}{\partial \mathbf{h}} = 2\mathbf{g}_1^* \quad (\text{A.12})$$

(ii) de (d):

$$\frac{\partial(\mathbf{h}^H \mathbf{g}_2)}{\partial \mathbf{h}} = \mathbf{0} \quad (\text{A.13})$$

Então, das Equações (A.12) e (A.13) tem-se:

$$\frac{\partial(\mathbf{h}^H \mathbf{R} \mathbf{h})}{\partial \mathbf{h}} = \frac{\partial(\mathbf{h}^H \mathbf{g}_2)}{\partial \mathbf{h}} + \frac{\partial(\mathbf{g}_1^H \mathbf{h})}{\partial \mathbf{h}} = 2\mathbf{R} \mathbf{h}^*. \quad (\text{A.14})$$

A.3 Relação: vetor gradiente - derivada em relação a um vetor

Sendo $J(\mathbf{h})$ uma função custo que define uma superfície de erro em relação ao vetor de parâmetros \mathbf{h} . Define-se o vetor gradiente da superfície de erro como:

$$\nabla J = \begin{bmatrix} \frac{\partial J}{\partial h_{0,R}} + j \cdot \frac{\partial J}{\partial h_{0,I}} \\ \frac{\partial J}{\partial h_{1,R}} + j \cdot \frac{\partial J}{\partial h_{1,I}} \\ \vdots \\ \frac{\partial J}{\partial h_{M-1,R}} + j \cdot \frac{\partial J}{\partial h_{M-1,I}} \end{bmatrix} \quad (\text{A.15})$$

O gradiente é normal (perpendicular) à superfície de erro. Comparando as Equações (A.3) e (A.15), verifica-se que:

$$\nabla J = 2 \frac{\partial J}{\partial \mathbf{h}^*} \quad (\text{A.16})$$

SUMÁRIO DE ALGORITMOS PROPOSTOS NA DISSERTAÇÃO

NESSE apêndice são descritos os algoritmos propostos na dissertação visando uma melhor referência para futuras consultas e implementações.

B.1 “Self Organized for Finding Valleys Algorithm” (SOFVA)

- Variáveis:

- θ_i : bias do neurônio i no instante n
- $\boldsymbol{\theta}$: vetor de parâmetros θ_i ($\boldsymbol{\theta} = [\theta_1 \ \theta_2 \ \cdots \ \theta_M]^T$)
- $x(n-1)$: saída do canal no instante $n-1$
- κ : controle de finura (constante para evitar problemas numéricos)
- λ : taxa de aprendizado

- Inicialização:

- $\theta_i(0)$: v.a. de $U[a, b]$
- κ, λ : > 0

- Dinâmica:

$$\theta_i(n+1) = \theta_i(n) - \lambda \frac{\text{sgn}[x(n-1) - \theta_i(n)]}{|x(n-1) - \theta_i(n)| + \kappa} \quad (\text{B.1})$$

Uma variação do algoritmo é através da inicialização dos parâmetros θ_i . Pode-se realizá-la através de uma distribuição linearmente espaçada entre os valores **a** e **b**. Nesse caso, $\boldsymbol{\theta} = \text{LS}[\mathbf{a}, \mathbf{b}]$.

B.2 Adaptação NPS-MPE

A adaptação NPS-MPE corresponde ao critério MPE como sendo o critério adotado pelo algoritmo supervisionado da estrutura NPS.

- Variáveis:

- $\beta_i(n)$: coeficiente linear da saída do neurônio i no instante n
- $\boldsymbol{\beta}(n)$: vetor dos parâmetros β_i no instante n ($\boldsymbol{\beta} = [\beta_1 \ \beta_2 \ \cdots \ \beta_M]^T$)
- $r_i(n)$: saída do neurônio i
- $\mathbf{r}(n)$: vetor de parâmetros r_i no instante n ($\mathbf{r} = [r_1 \ r_2 \ \cdots \ r_M]^T$)
- $e_p(n)$: erro de predição ($e_p(n) = x(n) - \hat{x}(n)$)
- $\hat{x}(n)$: saída da rede neural $\hat{x}(n) = \Psi(x(n-1))$
- μ_β : passo de convergência

- Inicialização:

- $\boldsymbol{\beta}(n) : \mathbf{0}$
- $\mu_\beta : > 0$

- Dinâmica:

$$x(n) = \Psi(x(n-1)) \quad (\text{B.2})$$

$$e_p(n) = x(n) - \hat{x}(n) \quad (\text{B.3})$$

$$\boldsymbol{\beta}(n+1) = \boldsymbol{\beta}(n) + \mu_\beta \cdot \mathbf{r}(n) \cdot e_p(n) \quad (\text{B.4})$$

B.3 Adaptação NPS-CMC

A adaptação NPS-CMC (“Neural Predictive Structure with Constant Modulus Criterion”) corresponde ao critério CM como sendo o critério adotado pelo algoritmo supervisionado da estrutura NPS.

- Variáveis:

- $\beta_i(n)$: coeficiente linear da saída do neurônio i no instante n
- $\boldsymbol{\beta}(n)$: vetor dos parâmetros β_i no instante n ($\boldsymbol{\beta} = [\beta_1 \ \beta_2 \ \cdots \ \beta_M]^T$)
- $r_i(n)$: saída do neurônio i
- $\mathbf{r}(n)$: vetor de parâmetros r_i no instante n ($\mathbf{r} = [r_1 \ r_2 \ \cdots \ r_M]^T$)
- $y(n)$: saída do NPS ($y(n) = e_p(n) \cdot g$)
- g : controle automático de ganho (CAG)
- R_2 : raio de equalização $\left(R_2 = \frac{\mathbb{E}\{a^4(n)\}}{\mathbb{E}\{a^2(n)\}} \right)$
- μ_β : passo de convergência

- Inicialização:

- $\boldsymbol{\beta}(n) : \mathbf{0}$
- $\mu_\beta : > 0$

- Dinâmica:

$$y(n) = e_p(n) \cdot g \quad (\text{B.5})$$

$$\boldsymbol{\beta}(n+1) = \boldsymbol{\beta}(n) + \mu \cdot y(n) \cdot (R_2 - |y(n)|^2) \cdot \mathbf{r}(n) \quad (\text{B.6})$$

B.4 “Fitting pdf Algorithm” (FPA)

O algoritmo FPA corresponde à regra de adaptação do “Fitting pdf Criterion” (FPC).

- Variáveis:

- $\mathbf{h}(n)$: vetor dos coeficientes do equalizador no instante n
- $\mathbf{x}(n)$: seqüência de entrada do equalizador no instante n
- $y(n)$: saída do equalizador ($y(n) = \mathbf{h}^T \mathbf{x}(n)$)
- σ_r^2 : parâmetro ajustável (variância de $p(y)$)
- a_i : símbolo do alfabeto de transmissão \mathcal{A}
- μ_h : passo de convergência

- Inicialização:

- $\mathbf{h}(0) : \mathbf{h} = [0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0]^T$
- $\mu_h : > 0$
- $\sigma_r^2 : > 0$

- Dinâmica:

$$\nabla J_{\text{FPC}}(\mathbf{h}) = \frac{\sum_{i=1}^S \exp\left(-\frac{(y-a_i)^2}{2\sigma_r^2}\right) (y - a_i)}{\sigma_r^2 \cdot \sum_{i=1}^S \exp\left(-\frac{(y-a_i)^2}{2\sigma_r^2}\right)} \mathbf{x}^* \quad (\text{B.7})$$

$$\mathbf{h}(n+1) = \mathbf{h}(n) - \mu_h \cdot \nabla J_{\text{FPC}}(\mathbf{h}) \quad (\text{B.8})$$

ARTIGOS PUBLICADOS

AQUI são listadas as publicações geradas durante a concepção desse trabalho e as cópias de alguns dos artigos.

- (1) ***“Preliminary Results in Blind Equalization with Neural Network-Based Prediction”***, Cavalcante,C.C., Mota,J.C.M., Dorizzi,B. e Montalvão,J.R., Anais do XVII Simpósio Brasileiro de Telecomunicações, pp. 508-513, Vila Velha-ES, Brasil, 1999.
- (2) ***“A Neural Predictor for Blind Equalization in Digital Communication”***, Cavalcante,C.C., Montalvão,J.R., Dorizzi,B. e Mota,J.C.M., Proceedings of Adaptive Systems for Signal Processing, Communication and Control (AS-SPCC), 1-3 October, Lake Louise, Canada, 2000.
- (3) ***“A Neural Predictor for Blind Equalization in Digital Communication: Is It Plausible”***, Cavalcante,C.C., Montalvão,J.R., Dorizzi,B. e Mota,J.C.M., Proceedings of IEEE Neural Networks for Signal Processing (NNSP2000), 11-13 December, Sydney, Australia, 2000.
- (4) ***“On The Nonlinear Prediction for Blind Equalization: Is That a Good Choice?”***, Cavalcante,C.C., Mota,J.C.M., Dorizzi,B. e Montalvão,J.R., Proceedings of XVIII Simpósio Brasileiro de Telecomunicações (SBT2000), 3-6 Setembro, Gramado, Brasil, 2000.
- (5) ***“A Simple PDF Fitting Approach for Blind Equalization”***, Montalvão,J.R., Cavalcante,C.C., Dorizzi,B. e Mota,J.C.M., Proceedings of XVIII Simpósio Brasileiro de Telecomunicações (SBT2000), 3-6 Setembro, Gramado, Brasil, 2000.

A Neural Predictor for Blind Equalization of Digital Communication Systems

Charles C. Cavalcante*
Univ. Fed. do Ceará
Fortaleza CE, Brazil
charlesc@dee.ufc.br

Jugurta R. Montalvão F.*
Univ. Tiradentes
Aracaju SE, Brazil
Jugurta.Montalvao@int-evry.fr

Bernadette Dorizzi
*Inst. Nat. des Télécom.
Evry, France
Bernadette.Dorizzi@int-evry.fr

João Cesar M. Mota
Univ. Fed. do Ceará
Fortaleza CE, Brazil
mota@dee.ufc.br

Abstract

In digital channel equalization, self-learning techniques are used in the cases where a training period is not available. Considering the transmitted sequence as composed of independent random variables, the equalization task can be done by means of prediction. In this work we propose to use Artificial Neural Networks (ANN), instead of a linear prediction device, in order to obtain a better performance. Prediction concepts are revisited and a new self-organized algorithm is proposed to update the first layer in the nonlinear predictor whose aim is to avoid local minimum points in the applied cost function. The second layer is updated by using a classical supervised algorithm. Simulation results are presented which illustrate the performance of this technique.

1. Introduction

Equalization of digital communication channels is usually done by using a transmitted sequence also known to the receiver during a preamble period. Figure 1 depicts a simplified digital communication system.

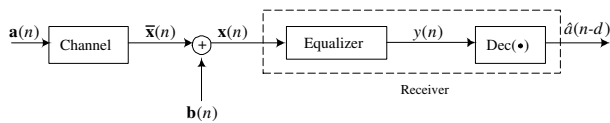


Figure 1: Digital Communication System

where $\mathbf{a}(n)$ is the transmitted sequence, $\mathbf{b}(n)$ is the noise sequence and $\hat{\mathbf{a}}(n-d)$ is the estimated symbol after a delay d .

Self-learning (blind) equalizers are used in order to provide the correct identification of transmitted symbols when one does not have a training period or when it is not practical to use such a strategy, as in digital TV broadcasting and multipoint networks where training has to be redone whenever one single receiver is inserted in the system. Another example is mobile communication systems, where due to

multipath fading, the received signal may be so low that the receiver does not synchronize adequately.

Some classical strategies for blind equalization are the following related algorithms: Direct Decision (DD), Sato, Godard [1], Benveniste-Goursat [2] and Shalvi-Weinstein [3].

Considering the transmitted symbols to be uncorrelated, it is possible to deal with the blind equalization problem by means of prediction [4]. In this context, the pioneer work is that of Macchi and Hachicha, in 1986, who used a linear filter as a predictor device. The symbol with the desired information is recovered, in this case, by elimination of the existing redundancy in the time sequence formed by the channel outputs.

In classical implementations for minimum phase channels, the prediction filter is linear and has a finite impulse response which is adapted to minimize the prediction squared error. This, indeed, is equivalent to a whitening process over the received time sequence and the white sequence obtained in the predictor error is the same as that of the transmitted symbols apart for a scale factor. The prediction error sequence will be i.i.d. if the transmitted sequence $\{\mathbf{a}(n)\}$ is also i.i.d and the noise negligible.

Nevertheless, in spite of the well established theory behind linear prediction, a crucial point should be mentioned: if the channel is nonminimum phase, even an infinite length predictor cannot provide a super-whitened error sequence. In other words, the original transmitted sequence cannot be recovered as result of the intrinsic linear mapping of past samples on the current estimated one. Nonetheless, it is quite easy to show that, in most cases, the ideal mapping is nonlinear (see example in Section 2).

Therefore, in this work, we propose a nonlinear structure based on Artificial Neural Networks as a prediction device. Moreover, in order to improve the adaptive solution, we divided the learning task in to two steps. First a new self-organized learning algorithm is proposed to adapt the first layer then the second layer connections are updated by

In Section 2, we explore the prediction concepts. Section 3 is dedicated to the new proposed self-organized algorithm. In Section 4, some simulation results are presented to illustrate the performance of this new strategy and, in the last section, conclusions are presented.

2. Prediction Concepts

In digital communication systems, the implicit goal of applying prediction is to remove temporal redundancies in the received signal, which can be used in blind equalization. The representation of a prediction-based equalizer is shown in Figure 2, where $\mathbf{x}(n)$ is the noisy channel output sequence, $\hat{\mathbf{x}}(n)$ is the predicted signal, $e(n)$ is the prediction error, \mathbf{P} is a prediction filter and g is an Automatic Gain Control (AGC).

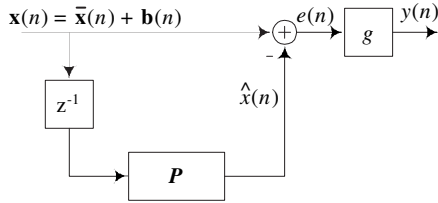


Figure 2: Prediction-Based Equalizer

The channel is modeled as a linear filter with discrete impulse response represented by

$$F(z) = \sum_{i=0}^{N-1} f_i z^{-i} \quad (1)$$

where f_i are the channel coefficients and N is the channel length. We can also represent the channel model in a vectorial form: $\mathbf{f} = [f_0 \ f_1 \ f_2 \ \cdots \ f_{N-1}]^T$.

Therefore, the noiseless channel outputs, which we call *channel states*, can be written as:

$$\begin{aligned} \bar{x}(n) &= a(n)f_0 + \cdots + a(n-N+1)f_{N-1} \\ \bar{x}(n-1) &= a(n-1)f_0 + \cdots + a(n-N+2)f_{N-1} \\ \bar{x}(n-2) &= a(n-2)f_0 + \cdots + a(n-N+3)f_{N-1} \\ &\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \end{aligned} \quad (2)$$

And then, the prediction error corresponds to:

$$e(n) = x(n) - \mathbf{P}(\mathbf{x}(n-1)) \quad (3)$$

where $\mathbf{x}(n-1) = [x(n-1) \ x(n-2) \ \dots]^T$, $x(n) = \tilde{x}(n) + b(n)$ and \mathbf{P} is a function which provides a prediction of $x(n)$.

The simplest structure is obtained by choosing \mathbf{P} to be linear. Although it is the most frequently used function, the application of the corresponding linear predictor is limited to minimum and maximum phase channels [1, 4]. Some works (for instance [4] and references therein) have proposed a combined structure to treat nonminimum phase channels.

To remove temporal redundancies, the prediction error equation is rewritten in the form:

$$e(n) = a(n)f_0 + a(n-1)f_1 + \cdots + b(n) - \mathbf{P}(\bar{x}(n-1) + b(n-1) + \bar{x}(n-2) + b(n-2) + \cdots) \quad (4)$$

Using a linear filter with discrete finite impulse response $\mathbf{p} = [p_1 \ p_2 \ p_3 \ \cdots \ p_k]$, as a predictor device, we have:

$$e(n) = \underbrace{a(n)f_0 + a(n-1)f_1 + \cdots + b(n)}_{x(n)} - \underbrace{[x(n-1)p_1 + x(n-2)p_2 + \cdots + x(n-k)p_k]}_{\hat{x}(n)} \quad (5)$$

where p_i is the i -th prediction filter coefficient.

Expanding $\hat{x}(n)$ leads to:

$$\begin{aligned} \hat{x}(n) = & (a(n-1)f_0 + a(n-2)f_1 + \cdots + b(n-1))p_1 \\ & + (a(n-2)f_0 + a(n-3)f_1 + \cdots + b(n-2))p_2 + \cdots \\ & + (a(n-k)f_0 + a(n-k+1)f_1 + \cdots + b(n-k))p_k \end{aligned} \quad (6)$$

Combining Equations (5) and (6) leads to:

$$\begin{aligned} e(n) = & a(n)f_0 + b(n) + a(n-1)[f_1 - f_0p_1] \\ & - b(n-1)p_1 + a(n-2)[f_2 - f_1p_1 - f_0p_2] - b(n-2)p_2 \\ & + \cdots - a(n-N+1)f_{N-1}[p_k] - b(n-N+1)p_k \end{aligned} \quad (7)$$

The goal here is to recover $a(n)f_0$. For this purpose, we must remove the undesired symbols by adapting the prediction filter in order to force them to zero. Unfortunately, not all coefficients can be canceled at once.

It becomes evident that, there is a *residue* in the prediction error expression and this residue cannot be cancelled by a finite linear filter. For equalization to be achieved, the samples of the prediction error sequence must to be uncorrelated and this residue must also be negligible with respect

to $a(n)f_0$. A possible solution to this problem is to increase the predictor order which decreases the contribution of the residue. The scale factor multiplying $a(n)$ is recovered by the AGC that matches the power of sequence $e(n)$ and transmitted sequence $\mathbf{a}(n)$.

However it is known that in the nonminimum phase channel case, it does not work and in any case the noise itself cannot be removed.

Since the linear mapping of a linear predictor may be not enough for equalization, we have tried to find a structure able to perform a nonlinear mapping in a satisfactory way. We chose the function implemented by an ANN that has the following form:

$$\mathbf{F} = \sum_i \beta_i \cdot \text{sign}(x - \theta_i) \quad (8)$$

where $\text{sign}(\cdot)$ is the signum function.

In the nonlinear case, Equation (3) is rewritten by replacing function \mathbf{P} by a nonlinear function \mathbf{F}_{NN} where the subscript stands for a neural network.

$$e(n) = x(n) - \mathbf{F}_{NN}(\mathbf{x}(n-1)) \quad (9)$$

According to Equation (8), \mathbf{F}_{NN} is a sum of weighted and shifted copies of $\text{sign}(\cdot)$, whose parameters β and θ would be found by means of an a priori knowledge of the channel coefficients. However, since we do not have such an a priori knowledge, all parameters of \mathbf{F}_{NN} are stochastically adjusted by means of the new algorithm described in Section 3.

Expanding Equation (9), it follows that:

$$e(n) = a(n)f_0 + a(n-1)f_1 + \dots + b(n) - \mathbf{F}_{NN}(x(n-1), x(n-2), \dots) \quad (10)$$

It is possible to find a function \mathbf{F}_{NN} such that we can exactly cancel the term: $a(n-1)f_1 + \dots + a(n-N+1)f_{N-1}$. Moreover this function can only explicitly depend on $x(n-1)$ since it has all information about past symbols that we need. So, rewriting Equation (10) it follows that:

$$e(n) = a(n)f_0 + a(n-1)f_1 + \dots + b(n) - \underbrace{\mathbf{F}_{NN}(a(n-1)f_0 + a(n-2)f_1 + \dots + b(n-1))}_{a(n-1)f_1 + a(n-2)f_2 + \dots + a(n-N+1)f_{N-1}} \quad (11)$$

In this case we can obtain no residue. It is worth noting that the noise $b(n)$ is assumed to be an white Gaussian random variable and, consequently, it is not predictable, therefore

the best the ANN can do is to cancel redundancies in the time sequence $e(n)$.

Figure 3 shows, a two-dimensional illustration of a nonlinear mapping done by the ANN.

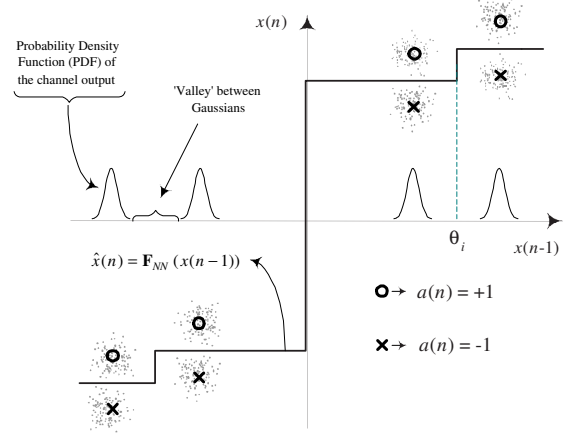


Figure 3: Nonlinear Mapping Function

Clearly, the parameters θ_i , in Equation (8), have a crucial role on the construction of ψ_{NN} . So, the problem of finding good parameters for the ANN is addressed in Section 3.

3. New Self-Organized Learning Algorithm

In classical techniques, such as backpropagation [5], the updating of parameters θ_i and of the synaptic weights is usually done through the same procedure. However, in this equalization problem such a procedure may not be able to quickly realize the fast transitions shown in Figure 3¹.

In order to solve this problem in a satisfactory way, we propose a new self-organized learning algorithm that is based on the minimization of a cost function in order to correctly find the θ_i of the neurons and to simplify the task of finding the interpolation surface.

It is easy to show that the probability density function (PDF) of the received signal is a mixture of Gaussians. Furthermore, the variance of each Gaussian is the noise variance and their means are channel-characteristic-related.

It can be seen from Figure 3 that the function referred to in Equation (8) can achieve correct interpolation if parameters θ_i , associated to the step transitions, are well placed between the 'valleys'. Then, we minimize a cost function that permits us to find those parameters by looking for a function that can fit in those 'valleys'. Since the valleys have the

¹In a previous work we have used the backpropagation algorithm for the equalization of channels with low intersymbol interference (ISI) [6].

shape of a “V”, perhaps the simplest function, similar to a “V” we can use is $N(x, \theta_i) = |x - \theta_i| + \kappa$ where $\kappa > 0$ (see Figure 4). We use this function in order to simplify the resulting algorithm.

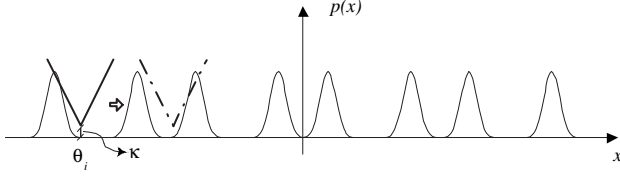


Figure 4: Looking for the ‘valleys’

The constant κ is inserted to avoid instability problems when $|x - \theta_i|$ is very small and to guarantee a strictly positive function $N(x, \theta_i)$.

In order to measure function similarities, we have chosen the Kullback-Leibler Divergence (KLD), which is indeed a distance measure in the Riemann space [5] given by:

$$D_{h(x)||g(x)} = \int_{-\infty}^{\infty} h(x) \ln \left(\frac{h(x)}{g(x)} \right) dx \quad (12)$$

where $h(x)$ and $g(x)$ are two strictly positive functions. Since the PDF of $x(n-1)$ and $N(x, \theta_i)$ are strictly positives, we apply KLD in order to measure similarities between them.

Eliminating the term which does not depend on θ_i we obtain a cost function given by:

$$J(\theta_i) = \int_{-\infty}^{\infty} p(x) \cdot \ln \left(\frac{1}{N(x, \theta_i)} \right) dx \quad (13)$$

$$J(\theta_i) = -\mathbb{E}\{\ln(N(x, \theta_i))\}$$

In order to find the minima of this cost function we have to set $\frac{\partial J(\theta_i)}{\partial \theta_i} = 0$, where

$$\frac{\partial J(\theta_i)}{\partial \theta_i} = -\mathbb{E} \left\{ \frac{\text{sign}(x - \theta_i)}{|x - \theta_i| + \kappa} \right\} \quad (14)$$

The minimum value of θ_i can thus be obtained by using a simple stochastic version of the gradient algorithm:

$$\theta_i(n+1) = \theta_i(n) - \lambda \frac{\text{sign}(x(n-1) - \theta_i)}{|x(n-1) - \theta_i| + \kappa} \quad (15)$$

which is a local adaptation rule of the Anti-Hebbian kind [5].

Finally, this adaptation rule is applied on the first layer while the second one is updated by a stochastic LMS algorithm.

4. Simulation Results

Performance measures of the proposed nonlinear predictor are compared with those of the linear predictor using two kinds of channels, firstly the minimum phase, where the linear predictor works, and then the nonminimum phase, where it gives bad results.

Using BPSK modulation and an SNR defined as $\text{SNR} = 10 \log_{10} \left(\frac{\sigma_a^2 \sum_{i=0}^{N-1} f_i^2 + \sigma_b^2}{\sigma_b^2} \right)$ where σ_a^2 and σ_b^2 are the symbol and noise variance respectively, we simulated a 100 Monte-Carlo trials the Decision Squared Error (DSE) $(\varepsilon(n) = y - \text{Dec}(y))^2$ performance and the measure of Symbol Error Ratio (SER) for both cases.

Minimum Phase Channel

The minimum phase channel used in the computational simulations has the following impulse response:

$$f(z) = 1 + 0.8z^{-1} + 0.4z^{-2}$$

In the linear predictor we used a filter with 25 coefficients, a step factor equal to 10^{-3} and the initialization is done by setting the vector of filter coefficients at zero.

The nonlinear predictor has one input, 15 neurons in the hidden layer and one output. For this structure we used the following parameters: supervised learning rate equals 10^{-3} , λ equals $5 \cdot 10^{-4}$. The number of symbols for finding the θ_i was set to 50 and $\kappa = 10^{-7}$. The algorithm for the AGC [4] has a step size equal to 10^{-3} and the weights in the output layer were initialized at zero and the θ_i were randomly initialized from an uniformly distributed interval: $[-1.5, 1.5]$.

We can see in Figures 5 and 6 the performance achieved when the SNR is equal to 40 dB and 30 dB, respectively. Figure 7 shows the Symbol Error Rate (SER) for some SNRs computed for both predictors.

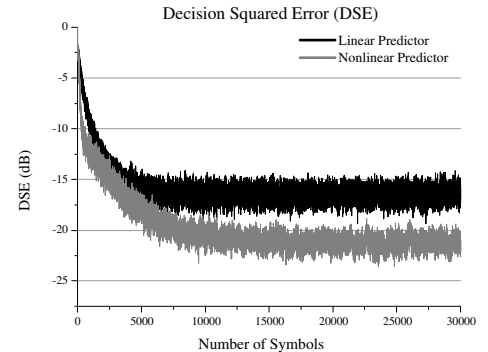


Figure 5: Decision-Squared Error (SNR = 40 dB)
Minimum Phase Channel

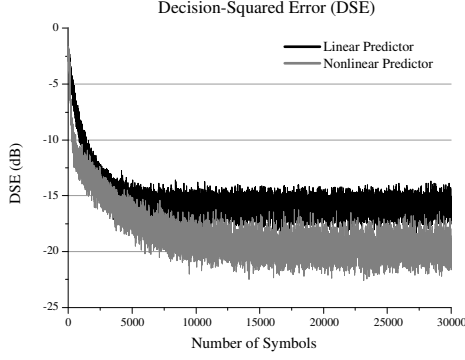


Figure 6: Decision-Squared Error (SNR = 30 dB)
Minimum Phase Channel

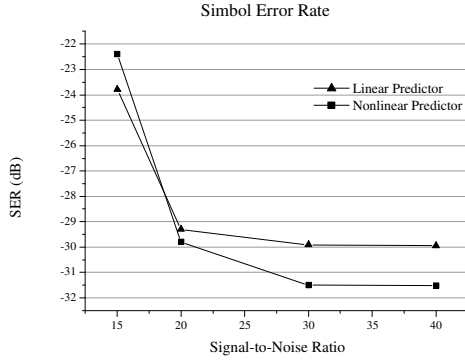


Figure 7: Symbol Error Rate
Minimum Phase Channel

One can easily see that the performance of the nonlinear predictor is better than that of the linear one when $\text{SNR} > 20$ dB.

Nonminimum Phase Channel

The channel considered has the following impulse response:

$$f(z) = 0.5 + 1z^{-1} - 0.6z^{-2}$$

The linear predictor has the same characteristics as in the previous case.

The nonlinear predictor has one input, 10 neurons in the hidden layer and one output. The parameters are: supervised learning rate equal to $5 \cdot 10^{-3}$ and λ equals 10^{-4} . The number of symbols for finding the θ_i was set to 500 symbols and $\kappa = 10^{-7}$. The step size of the AGC equals 10^{-3} and weights in the output layer were initialized at zero and the θ_i were randomly initialized from an uniformly distributed interval: $[-1.5, 1.5]$.

Table 1 shows the SER of both predictors, for an SNR of 35 dB. As expected, the linear predictor has a poor performance

while the nonlinear one performs pretty well.

	Linear Predictor	Nonlinear Predictor
SER	-3.82 dB	-12.1 dB

Table 1: SER for an SNR of 35 dB
Nonminimum Phase Channel

5. Conclusions

The strategy presented in this paper proposes a nonlinear prediction device based on Artificial Neural Networks. The use of prediction is extended to some cases of nonminimum phase channels and the nonlinear predictor outperforms the linear one even in the cases where it realizes channel equalization.

The division on the learning task in to two steps, a self-organized and a supervised one, avoids reaching points of local minimum in the cost function.

However, this strategy is limited in the situations where the ‘valleys’ between the Gaussians of the PDF of $x(n-1)$ are deep enough, this deepness is the function of the noise power and channel characteristics. In cases where this condition does not hold, we must consider an adaptation of the previous algorithms, particularly acting on the parameter κ . This improvement is actually under development.

Acknowledgments

The authors would like to thank FUNCAP and CNPq for the partial financial support. They would also like to thank INT-Evry France for the logistical support.

6. References

- [1] Haykin, S., *Adaptive Filter Theory*. 3rd edition. Prentice-Hall International, 1996.
- [2] Benveniste, A. and M. Goursat, “Blind Equalizers”, IEEE Trans. Communication, Vol. COM-32, No. 8, August 1984.
- [3] Shalvi O. and E. Weinstein, “New criteria for blind deconvolution of non minimum phase systems”, IEEE Trans. on Inform. Theory, Vol. 36, pp. 312-321. 1990.
- [4] Macchi, O. “L’égaliisation numérique en communications”, Annales des Télécomm. 53, No 1-2, pp. 39-58, 1998.
- [5] Haykin, S., *Neural Networks: A Comprehensive Foundation*. 2nd edition. Prentice Hall International, 1998.
- [6] Cavalcante C.C., J.C.M. Mota, B. Dorizzi and J.R. Montalvão, “Preliminary Results in Blind Equalization with Neural Network-Based Prediction”, In Proc. XVII Brazilian Telecommunication Symposium, pp.508-513, Vila Velha, Brazil, 1999.

A SIMPLE PDF FITTING APPROACH FOR BLIND EQUALIZATION

JUGURTA R. MONTALVÃO F.⁽¹⁾, CHARLES C. CAVALCANTE⁽²⁾, B. DORIZZI⁽³⁾, JOÃO C. M. MOTA⁽⁴⁾

(1) Universidade Tiradentes, Aracaju-SE, Brazil, Jugurta.Montalvao@int-evry.fr

(2) Universidade Federal do Ceará, Fortaleza-CE, Brazil, charlesc@dee.ufc.br

(3) Institut National des Télécomm., Evry, France, Bernadette.Dorizzi@int-evry.fr

(4) Universidade Federal do Ceará, Fortaleza-CE, Brazil, mota@dee.ufc.br

ABSTRACT

A new blind equalization algorithm for digital communication systems is presented. This algorithm is based on the adjusting of a linear equalizer in such a way that the probability density function (PDF) of its output matches a parametric target function. A link between the proposed cost function and that used by the Constant Modulus Algorithm is also pointed out. Some simulation results are presented and compared to that provided by the Godard's equalizer.

1. INTRODUCTION

This work deals with blind equalization of linear channels in digital communication systems. A new blind equalizer is presented, based on the adaptation of the equalizer taps until that of the output equalizer PDF matches a target multimodal function. In this moment we can guarantee that equalization is performed [1]. In Section 2, the system model is presented along with the new algorithm. A link between this new algorithm and the CMA is presented in Section 3. Simulations and conclusions are presented in the last two sections.

2. A NEW BLIND ALGORITHM

Fig. 1 shows the single-input-single-output system model we are concerned with. The equalizer input is obtained after demodulation and sampling of the received signal.

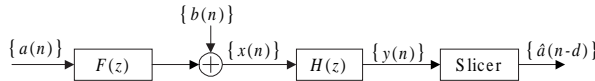


Figure 1: Baseband System Model.

The stream $a(n)$ carries the information and each element in this stream is modeled as a random variable which takes values from an S -sized symbolic alphabet. We further assume that this alphabet is complex-valued and symmetric w. r. t. the origin. The noise sequence $b(n)$ is i.i.d., zero-mean Gaussian and statistically independent of $a(n)$. Then, and assuming that the channel transfer function, $F(z)$, has no spectral nulls, a possible equalization strategy is to make the global response $G(z) = F(z) \cdot H(z)$ as

close as possible to the ideal one: $G(z) = z^{-d}$, where d is a suitable decision delay. This is, in fact, the so-called zero forcing (ZF) equalization [2].

Assuming that perfect ZF is obtained, it is easy to show that the PDF of $y(n)$ is:

$$p_{Y,ideal}(y; \mathbf{h}_{ZF}, \sigma_b^2) = \frac{1}{\sqrt{2\pi\sigma_b^2}} \sum_{s=1}^S \exp\left(\frac{-|y(n) - a_s|^2}{2\sigma_b^2}\right) \quad (1)$$

where $\mathbf{h}_{ZF} = [h_0^{ZF} \ h_1^{ZF} \ \dots \ h_{M-1}^{ZF}]^T$ is a normalized vector of equalizer coefficients and σ_b^2 is the noise variance. In fact, we assume $M \gg 1$ to perform equalization.

Then, in order to match this desired PDF of a perfectly equalized system, we construct a parametric target function given by $f(y, \mathbf{h}, \hat{\sigma}_b^2) = p_{Y,ideal}(y; \mathbf{h}, \hat{\sigma}_b^2)$, and we compare it to the actual PDF, $p_Y(y)$, by means of the following measurement of the extent to which the model density and the true density agree [3, p. 59]:

$$J = - \int p_Y(y; \mathbf{h}_{ZF}, \sigma_b^2) \ln(f(y, \mathbf{h}, \hat{\sigma}_b^2)) dy \quad (2)$$

where the minimum of J , obtained when the two functions are equal, is the entropy of Y .

Clearly, the estimated noise variance $\hat{\sigma}_b^2$ depends on the equalizer coefficients. However, in order to be simple, we can replace it by a constant parameter Υ which, in fact, can play a crucial role on the final algorithm (see Section 3). Then, expanding Equation (2) and eliminating the constant terms, our effective cost function becomes

$$J = -E \left\{ \ln \left[\sum_{s=1}^S \exp\left(\frac{-|y(n) - a_s|^2}{2\Upsilon}\right) \right] \right\} \quad (3)$$

Now, taking the stochastic gradient of Equation (3) w.r.t. the equalizer parameters, our blind equalization algorithm can be summarized as follows:

$$\mathbf{h}(n+1) = \mathbf{h}(n) - \alpha_h \nabla J$$

where

$$\nabla J = \frac{\sum_{s=1}^S \exp\left(\frac{-|y(n) - a_s|^2}{2\Upsilon}\right) (y(n) - a_s)}{\Upsilon \sum_{s=1}^S \exp\left(\frac{-|y(n) - a_s|^2}{2\Upsilon}\right)} \mathbf{x}^* \neq 0$$

3. A LINK TO THE CM CRITERION

After the milestone paper of Sato [4], in 1975, blind equalization to compensate for intersymbol interference (ISI) in digital communication has been intensively studied. Among the most known

blind equalization criteria, those based on Constant Modulus are probably the most popular. For instance, we can point out the so-called Godard equalizer [5]. Moreover, since a great number of publications on CMA performance limits and convergence issues is available, it is useful to find similarities between new approaches and CMA. Indeed, some relevant works like [6],[7] and [8] point out links between other approaches and the CM criterion.

Likewise, a first step toward the performance characterization of our proposal has been to find a link between our approach and CMA. This link is clearer when a BPSK (± 1) modulation scheme is applied. Then, a parametric target function can be defined as follows:

$$f_{CM}(y, \mathbf{h}, p) = \frac{\exp(|y|^p - 1|^2)}{\int_{-\infty}^{+\infty} \exp(|\xi|^p - 1|^2) d\xi}$$

where p is a strictly positive integer parameter.

It is straightforward to show that, applying it in Equation (2), the resulting cost function, apart from some constants, is indeed the well-known constant modulus criterion:

$$J_{CM} = -E\{\ln[f_{CM}(y, \mathbf{h}, p)]\} = E\{|y|^p - 1|^2\}$$

As a result, from this point of view we can regard the CM criterion as being a measure of divergence between the actual equalizer output PDF and the target function $f_{CM}(\cdot)$. Moreover, we can show that our proposed target function $f(y, \mathbf{h}, \Upsilon)$ can be set close to either $f_{CM}(y, \mathbf{h}, 1)$ (i.e., the Sato's criterion) or $f_{CM}(y, \mathbf{h}, 2)$ (i.e., the Godard's criterion), by choosing a suitable value for Υ .

Fig. 2 shows some such target functions. Similarities between the target functions, along with some simulation results, have shown that an appropriate choice of Υ can afford performance characteristics close to those of either the Godard or Sato equalizer, when the modulation scheme is the BPSK.

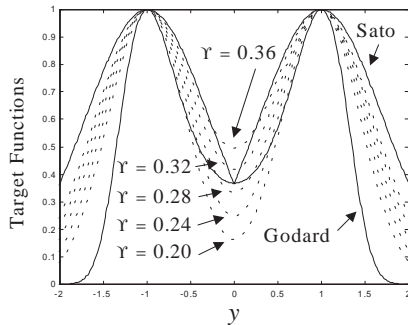


Figure 2: Comparison between Target Functions with some Υ values.

Moreover, for more complex modulation schemes, the new family of target functions can better fits the idealized equalizer output PDF with S Gaussian kernels. Thus, since Equation (1) can take into account even complex symbol parts and multilevel alphabets, we can predict an improved performance w.r.t. the CMA.

4. SIMULATION RESULTS

All simulations presented in this section were done in order to verify the relationship between our proposal and CMA for $p = 2$ (i.e., the Godard's equalizer). Moreover, only two modulation schemes were considered, namely, BPSK and 4-QAM.

The signal-to-noise ratio is defined as $\text{SNR} = 10 \log_{10} \left(\frac{\sigma_a^2 \sum_{i=0}^{N-1} f_i^2 + \sigma_b^2}{\sigma_b^2} \right)$ where σ_a^2 and σ_b^2 are the symbol and noise variance respectively. Performance measures (i.e., Decision Squared Error (DSE) $(\varepsilon(n) = y - \text{Dec}(y))^2$ and Symbol Error Rate (SER)) are averaged from 200 Monte-Carlo trials.

For clarity's safe, our approach is called NBE (New Blind Equalizer) in the figures.

Simulations with BPSK modulation were done with two channels. The discrete impulse response of such channels can be represented in vectorial form as follows:

$$\mathbf{f}_1 = [0.1 \quad 0.5 \quad 1 \quad -0.6 \quad -0.2]$$

and

$$\mathbf{f}_2 = [1 \quad -0.2 \quad 0.71 \quad 0.282 \quad 0.8658]$$

Channel \mathbf{f}_1 has no spectral near-nulls as we can see in Fig. 3 (The representation of channel zeros is also presented in Fig. 4). Therefore this channel can be "easily equalized" by means of a linear transversal equalizer (LTE).

Fig. 5 shows the performance of both NBE and CM equalizers measured in terms of DSE. It can be easily seen that our approach is equivalent to the CM criterion after convergence. However, the convergence of the NBE is faster than that of the CMA because its adaptation step size is the greater. We highlight that a step size greater than 10^{-3} provokes divergence of the CM Algorithm. Table 1 specifies the simulation parameters. That fact indicates higher robustness to the noise by the NBE.

	CMA	NBE
TAPs	30	30
Step size	10^{-3}	$25 \cdot 10^{-3}$
Υ	—	0.3
Initial TAPs	$[0 \cdots 0 \ 1 \ 0 \cdots 0]$	$[0 \cdots 0 \ 1 \ 0 \cdots 0]$

Table 1: Simulation Parameters for \mathbf{f}_1

Channel \mathbf{f}_2 has two spectral near-nulls, as we can see in Fig. 6 that shows the frequency response of the channel and Fig. 7 shows the channel zeros. It is worth noting that those spectral near-nulls rend the equalization with a LTE more difficult.

Fig. 8 shows the performance of both the NBE and CMA in an environment with $\text{SNR} = 30$ dB. Table 2 specifies simulation parameters.

	CMA	NBE
TAPs	30	30
Step size	10^{-3}	$25 \cdot 10^{-3}$
Υ	—	0.29
Initial TAPs	$[0 \cdots 0 \ 1 \ 0 \cdots 0]$	$[0 \cdots 0 \ 1 \ 0 \cdots 0]$

Table 2: Simulation Parameters for \mathbf{f}_2

Simulations with 4-QAM modulation scheme were done with the channel proposed in [9] that has the discrete impulse response given by:

$$\mathbf{f}_3 = [2 - 0.4j \quad 1.5 + 1.8j \quad 1 \quad 1.2 - 1.3j \quad 0.8 + 1.6j]$$

This channel has been used in some recent works to illustrate the robustness of algorithms facing a very distorsive channel. The frequency response of this channel is depicted in Fig. 9 whereas Figure 10 shows the channel zeros.

The performances of both equalizers are depicted in Fig. 11. As we can see, they have an almost identical performance. However, we emphasize that in this case the CMA requires a phase recovering device to work whereas the NBE compensates for phase distortion by itself. Table 3 specifies simulation parameters.

	CMA	NBE
TAPs	30	30
Step size	10^{-3}	2.10^{-3}
Υ	—	0.35
Step Size Phase	10^{-3}	—
Initial Phase	0	—
Initial TAPs	$[0 \cdots 0 \ 1 \ 0 \cdots 0]$	$[0 \cdots 0 \ 1 \ 0 \cdots 0]$

Table 3: Simulation Parameters for \mathbf{f}_3

5. CONCLUSIONS AND PERSPECTIVES

Simulation results had confirmed that when $\Upsilon \cong 0.3$ our proposal has a final performance practically equivalent to that of the Godard's equalizer for both BPSK and 4-QAM modulation schemes. On the other hand, we also conclude that, thanks to its capacity of convergence with step-sizes greater than Godard's, the NBE can provide faster convergence in some cases.

Furthermore, as illustrated in the last simulation result, the NBE has no need for a phase recovering device.

Actually, this work has just started and the continuity of it includes tests involving multi-amplitude complex modulation, like 16-QAM schemes, and the study of an adaptation strategy for the parameter Υ . The use of a feedback structure cascaded with the filter is also under study.

ACKNOWLEDGEMENTS

The authors would like to thank CNPq and FUNCAP from Brazil by their partial financial support.

6. REFERENCES

- [1] Benveniste A., M. Goursat and G. Ruget, "Robust Identification of a Nonminimum Phase System: Blind Adjustment of a Linear Equalizer in Data Communications", IEEE Trans. on Automatic Control, Vol. AC-25, No. 3, Pp. 385-399, June, 1980.
- [2] Haykin, S., *Adaptive Filter Theory*. 3rd Edition. Prentice-Hall International, 1996.
- [3] Bishop C. M., *Neural Networks for Pattern Recognition*, Oxford Uk University Press, 1995.
- [4] Sato, Y., "A Method of Self-Recovering Equalization for Multi-level Amplitude Modulation", IEEE Trans. on Comm., Vol. 23, Pp. 679-682, 1975.
- [5] Godard, D. N., "Self-Recovering Equalization and Carrier Tracking in a Two-Dimensional Data Communication System (Channels)", IEEE Trans. on Communications, Vol. COM-28, Pp. 1867-1875, 1980.
- [6] Sala-Alvarez, J. and G. Vázquez-Grau "Statistical Reference Criteria for Adaptive Signal Processing in Digital Communications", IEEE Transactions on Signal Processing, Vol. 45, No. 1, Pp.14-30. 1997.
- [7] Touzni, A. *Performance et Robustesse en Égalisation Aveugle Spatio-Temporelle*, Ph.D. dissertation (in french), Université de Cergy-Pontoise, France, November, 1998.
- [8] Regalia, P.A. "On the Equivalence Between the Godard and Shalvi-Weinstein Schemes of Blind Equalization", Signal Processing, Vol. 73, No. 1-2, Pp.185-190. 1999.
- [9] Porat B., B. Friendlander, "Blind Equalization of Digital Communication Channels using High-order Moments", IEEE Trans. on Signal Processing, Vol. 39, No. 2, Pp.522-526, February, 1991.

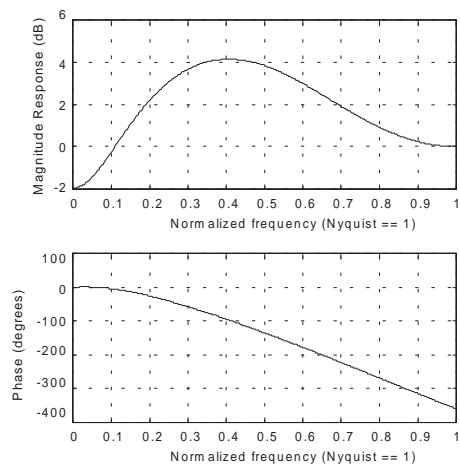


Figure 3: Frequency Response of Channel f_1 .

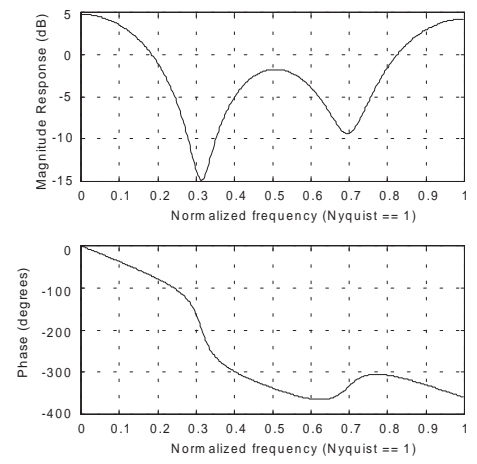


Figure 6: Frequency Response of Channel f_2 .

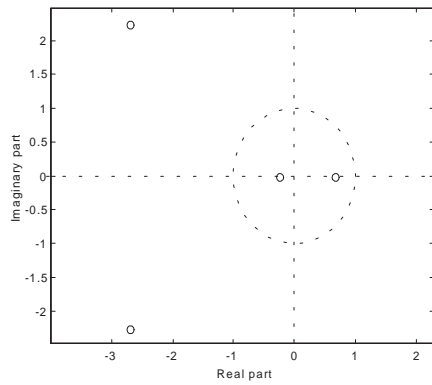


Figure 4: Root Locus of Channel f_1 .

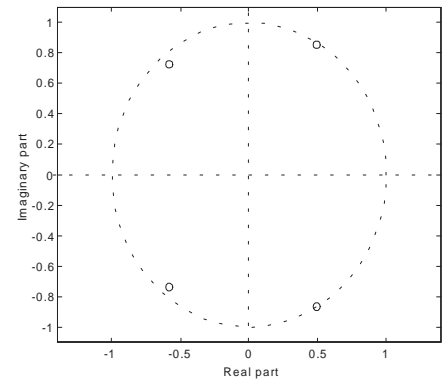


Figure 7: Root Locus of Channel f_2 .

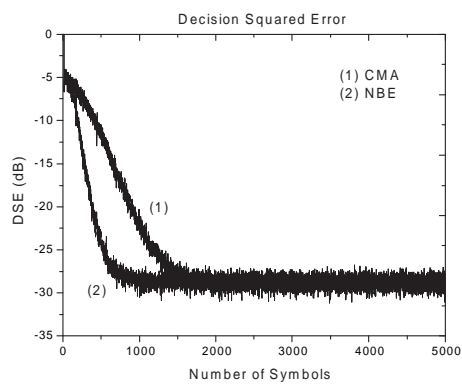


Figure 5: DSE for Channel f_1 (SNR = 30 dB).

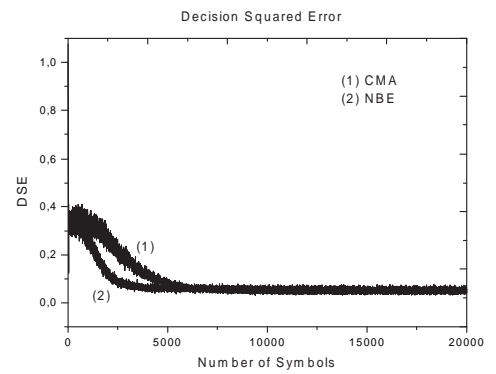


Figure 8: DSE for Channel f_2 (SNR=30 dB).

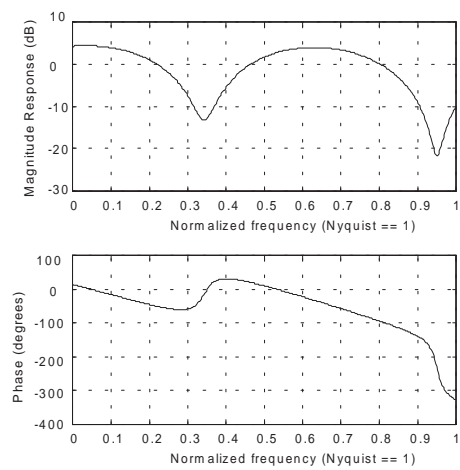


Figure 9: Frequency Response of Channel f_3 .

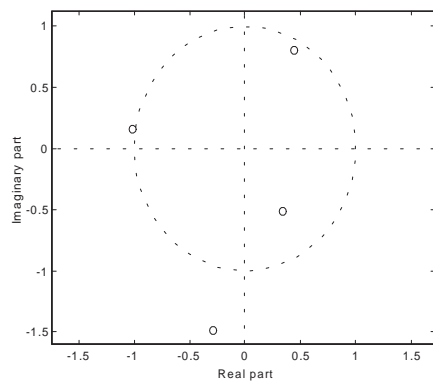


Figure 10: Root Locus of Channel f_3 .

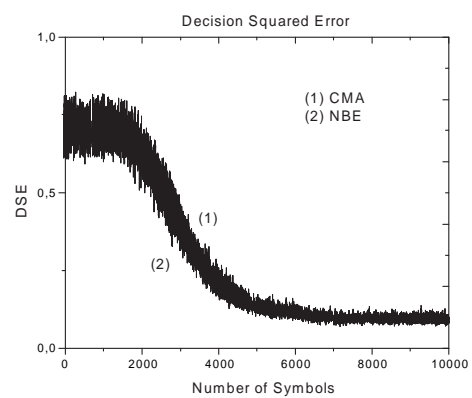


Figure 11: DSE for Channel f_3 (SNR = 30 dB).

REFERÊNCIAS

- BALAY, P. (1995). *Application des Techniques Neuronales à la Demodulation Numérique*. Tese de doutorado, L'Université de Rennes I, France. Citado na(s) pág(s): 8, 61
- BARRETO, G. A. (1998). *Redes Neurais Não-Supervisionadas para Processamento de Sequências Temporais*. Dissertao de mestrado, Escola de Engenharia de São Carlos, Universidade de São Paulo. Citado na(s) pág(s): 8
- BENEDETTO, S., BIGLIERI, E., & CASTELLANI, V. (1987). *Digital Trasmission Theory*. Prentice Hall International. Citado na(s) pág(s): 14, 24
- BENVENISTE, A., & GOURSAT, M. (1984). Blind Equalizers. *IEEE Transaction on Communications*, Vol. COM-32(No. 8), Pp. 871-883. Citado na(s) pág(s): 71
- BENVENISTE, A., GOURSAT, M., & RUGET, G. (1980). Robust Identification of a Nonminimum Phase System: Blind Adjustment of a Linear Equalizer in Data Communications. *IEEE Trans. on Automatic Control*, Vol. AC-25(No. 3), Pp. 385-399. Citado na(s) pág(s): 121
- BISHOP, C. M. (1995). *Neural Networks for Pattern Recognition*. UK: Oxford University Press. Citado na(s) pág(s): 32, 39, 41, 42, 43, 44, 122
- BONÉ, R. (2000). *Reseaux de Neurones Recurrents pour la Prévission de Series Temporelles*. Tese de doutorado, Université François Rabelais Tours, France. Citado na(s) pág(s): 8
- BROSSIER, J.-M. (1997). *Signal et Communication Numérique: égalisation et synchronisation*. HERMES. Citado na(s) pág(s): 22, 66, 68

- CAVALCANTE, C. C., MONTALVÃO, J. R., FILHO, DORIZZI, B., & MOTA, J. C. M. (2000). A Neural Predictor in Blind Equalization for Digital Communication Systems. In *Adaptive System in Signal Processing, Communications and Control (AS-SPCC)* (p. 493-497). Lake Louise, Canada. Citado na(s) pág(s): 90
- CAVALCANTE, C. C., MOTA, J. C. M., DORIZZI, B., & MONTALVÃO, J. R., FILHO. (1999). Preliminary Results in Blind Equalization with Neural Networks Based-Prediction. In *Proc. XVII Simpósio Brasileiro de Telecomunicações - SBT 99* (p. 508-513). Vila Velha, ES - Brasil. Citado na(s) pág(s): 93
- CAVALCANTI, F. R. P. (1996). *Equalização Adaptativa e Autodidata em Sistemas de Comunicação Digitais com Ênfase em Técnicas Preditivas*. Dissertao de mestrado, UFC, Brasil. Citado na(s) pág(s): 8, 68, 85, 87, 100, 125
- CZERNICHOW, T. (1996). *Apport des Réseaux Recurrents à la Prévision des Series Temporelles, Application à la Prévision de Consommation d'Électricité*. Tese de doutorado, Université Paris VI, France. Citado na(s) pág(s): 8
- DECO, G., & OBRADOVIC, D. (1996). *An Information-Theoretic Approach to Neural Computing*. Springer-Verlag. Citado na(s) pág(s): 37, 44
- DUDA, R. O., & HART, P. E. (1973). *Pattern Classification and Scene Analysis*. New York, USA: Wiley. Citado na(s) pág(s): 39, 44
- FORNEY, G. D., JR. (1973). The Viterbi Algorithm. *Proceedings of the IEEE, Vol. 61*(No. 3), PP. 268-278. Citado na(s) pág(s): 67
- GIBSON, J. D. (1993). *Principles of Digital and Analog Communications* (2nd ed.). New Jersey, USA: Prentice-Hall. Citado na(s) pág(s): 31
- GLENTIS, G.-O., BERBERIDIS, K., & THEODORIDIS, S. (1999). A Efficient Least Squares Adaptive Algorithms for FIR Transversal Filtering. *IEEE Signal Processing Magazine, Vol. 16*(No. 4), Pp. 13-41. Citado na(s) pág(s): 71
- GODARD, D. N. (1980). Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems. *IEEE Trans. on Communications, Vol. COM-28*(No. 11), Pp. 1867-1875. Citado na(s) pág(s): 70
- HARVEY, R. L. (1994). *Neural Network Principles*. Prentice Hall International. Citado na(s) pág(s): 47, 49
- HASSOUN, M. H. (1995). *Fundamentals of Artificial Neural Networks*. Cambridge, Massachusetts: MIT Press. Citado na(s) pág(s): 47, 51
- HAYKIN, S. (Ed.). (1994a). *Blind Deconvolution*. NJ, USA: Prentice-Hall, Englewood Cliffs. Citado na(s) pág(s): 4, 69
- HAYKIN, S. (1994b). *Communication Systems* (3rd ed.). John Wiley and Sons. Citado

- na(s) pág(s): 14, 29
- HAYKIN, S. (1996). *Adaptive Filter Theory* (3rd ed.). Prentice-Hall. Citado na(s) pág(s): 7, 56, 60, 67, 69, 70, 83
- HAYKIN, S. (1998). *Neural Networks: A Comprehensive Foundation* (2nd ed.). Prentice Hall. Citado na(s) pág(s): 9, 31, 32, 34, 35, 37, 44, 47, 55, 56, 60, 61, 62, 98, 101
- HAYKIN, S. (Ed.). (2000). *Unsupervised adaptive filtering* (Vol. 2). John Wiley. Citado na(s) pág(s): 123
- HILAL, K. (1993). *Algorithmes Accélérés d'Egalisation Adaptative Autodidacte*. Tese de doutorado, ENST/Paris, France. Citado na(s) pág(s): 69
- JOHNSON, C. R., JR., SCHNITER, P., ENDRES, T. J., BEHM, J. D., BROWN, D. R., & CASAS, R. A. (1998). Blind Equalization using the Constant Modulus Criterion: A Review. *Proceedings of the IEEE*, 86(10), 1927-1950. Citado na(s) pág(s): 70
- KOVÁCS, Z. L. (1996). *Redes Neurais Artificiais: Fundamentos e Aplicações* (2^a edição ed.). São Paulo: Collegium Cognitio. Citado na(s) pág(s): 46, 47, 49, 50, 51, 52, 56, 58
- KOVÁCS, Z. L. (1997). *O Cérebro e a sua Mente: Uma Introdução à Neurociência Computacional*. São Paulo: Edição Acadêmica São Paulo. Citado na(s) pág(s): 47
- LABAT, J., & LAOT, C. (1999). Blind Spatio-Temporal Decision Feedback Equalization: A Self-Adaptive Approach. In *Proc. of Global Telecommunications Conference - GLOBECOM'99* (p. Pp.2301-2305). Rio de Janeiro - Brazil. Citado na(s) pág(s): 85
- LABAT, J., MACCHI, O., & LAOT, C. (1998). Adaptive Decision Feedback Equalization: Can You Skip the Training Period? *IEEE Transactions on Communications*, Vol. 46(No. 7), pp. 921-930. Citado na(s) pág(s): 8, 85
- MACCHI, O. (1995). *Adaptive Processing: The Least Mean Squares with Applications in Transmission*. John Wiley & Sons. Citado na(s) pág(s): 83, 102
- MACCHI, O. (1998). L'égalisation numérique en communications. *Annales de Télécommunications*, Vol. 53(No. 1-2), pp. 39-58. Citado na(s) pág(s): 8, 102
- MACCHI, O., & GU, Y. (1987). Self-Adaptive Equalization with a Mixed Backward and Forward Predictors. In *Proceedings Int. Symp. on Electronics Devices, Circuits and Systems* (p. Pp. 437-440). Kharagpur. Citado na(s) pág(s): 7
- MACCHI, O., & HACHICHA, A. (1986). Self-Adaptive Equalization Based on a Prediction Principle. In *Proceedings GLOBECOM 86*. Houston, USA. Citado na(s) pág(s): 7
- MCCULLOCH, W., & PITTS, W. (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, 5(115), 33. Citado na(s) pág(s): 50

- MONTALVÃO, J. R., FILHO. (2000). *Égalisation et Identification de Canaux de Communication numérique: Une Approche par Reconnaitance des Formes et Mélange de Gaussiennes*. Tese de doutorado, Université Paris XI, France. (Em francês) Citado na(s) pág(s): 17, 23, 63, 65, 66, 67, 68, 76, 78, 102, 115, 120, 122
- MONTALVÃO, J. R., FILHO, CAVALCANTE, C. C., DORIZZI, B., & MOTA, J. C. M. (2000). A Simple PDF Fitting Approach for Blind Equalization. In *Simpósio Brasileiro de Telecomunicações (SBT'2000)*. Gramado, Brasil. Citado na(s) pág(s): 127
- MONTALVÃO, J. R., FILHO, DORIZZI, B., & MOTA, J. C. M. (1999). Some Theoretical Limits for Linear and Nonlinear Equalizers. *Revista da Sociedade Brasileira de Telecomunicações*, Vol. 3(No. 2), pp. 22-31. Citado na(s) pág(s): 4, 102
- MONTALVÃO, J. R., FILHO, MOTA, J. C. M., DORIZZI, B., & CAVALCANTE, C. C. (1998). Reducing Bayes Equalizer Complexity: A New Approach for Clusters Determination. In *Proceedings ITS'98/IEEE*. São Paulo, Brasil. Citado na(s) pág(s): 39, 63, 67, 72, 76
- MOTA, J. C. M. (1992). *Equalização Adaptativa em Sistemas Rádio-Digitais: das Técnicas em Frequência à Teoria da Desconvolução Autodidata*. Tese de doutorado, DECOM/FEEC/UNICAMP, Brasil. Citado na(s) pág(s): 7, 8, 20, 21, 67, 68, 85, 86
- MULGREW, B. (1996). Applying Radial Basis Functions. *IEEE Signal Processing Magazine*, Vol. 13(No. 2), Pp. 50-65. Citado na(s) pág(s): 8, 63, 72
- MULGREW, B. (1998). Nonlinear Signal Processing for Adaptive Equalization and Multi-user Detection. In *Proc. IX European Signal Processing Conference, EUSIPCO 98* (p. Pp. 312.8.1-312.5.5). Rhodes, Greece. Citado na(s) pág(s): 63
- OPPENHEIM, A. V., & SCHAFER, R. W. (1989). *Discrete-Time Signal Processing* (1st ed.). New Jersey: Prentice Hall. Citado na(s) pág(s): 15, 18, 140
- PAPADIAS, C. (1995). *Methods for Blind Equalization and Identification of Linear Channels*. Tese de doutorado, Ecole National Supérieure des Télécommunications (ENST), France. Citado na(s) pág(s): 14, 70
- PAPOULIS, A. (1991). *Probability, Random Variables and Stochastic Processes* (3rd ed.). McGraw-Hill International. Citado na(s) pág(s): 17, 41, 82
- PAULRAJ, A., ROYCHOWDHURY, V., & SCHAPER, C. (Eds.). (1997). *Communications, Computation, Control and Signal Processing: A Tribute to Thomas Kailath*. USA: Kluwer Academics. Citado na(s) pág(s): 7
- PICCHI, G., & PRATI, G. (1987). Blind Equalization and Carrier Recovery Using a 'Stop-and-Go' Decision-Direct Algorithm. *IEEE Trans. on Communications*, Vol. COM-35(No. 9), Pp. 877-887. Citado na(s) pág(s): 71

- PORAT, B., & FRIEDLANDER, B. (1991). Blind Equalization of Digital Communication Channels using High-order Moments. *IEEE Trans. on Signal Processing*, Vol. 39(No. 2), Pp. 522-526. Citado na(s) pág(s): 69
- PROAKIS, J. G. (1995). *Digital Communications* (3rd ed.). McGraw-Hill. Citado na(s) pág(s): 4, 14, 15, 18, 22, 24, 67
- QURESHI, S. U. H. (1985). Adaptive Equalization. *Proceedings of the IEEE*, Vol. 73(No. 9), pp. 1349-1387. Citado na(s) pág(s): 4
- REGALIA, P. A. (1999). On the Equivalence Between the Godard and Shalvi-Weinstein Schemes of Blind Equalization. *Signal Processing*, 73(1-2), 185-190. Citado na(s) pág(s): 123
- RENYI, A. (1987). *A Diary on Information Theory*. Wiley. Citado na(s) pág(s): 33
- ROCHA, C. A. F. DA. (1996). *Técnicas Preditivas para Equalização Autodidata*. Tese de doutorado, DECOM/FEEC/UNICAMP, Brasil. Citado na(s) pág(s): 6, 16, 69, 82, 86
- ROCHA, C. A. F. DA, & MACCHI, O. (1994). A Novel Self-Learning Adaptive Recursive Equalizer with Unique Optimum for QAM. In *Proceedings ICASSP-94* (p. Pp. III.481-III.484). Adelaide, Australia. Citado na(s) pág(s): 8, 83
- ROCHA, C. A. F. DA, ROMANO, J. M. T., & MACCHI, O. (1995). Égalisation Adaptative Autodidacte en Cascadant Égaliseurs d'Amplitude et de Phase. In *Proc. XVI GRETSI Colloquium*. Juan les Pins, France. Citado na(s) pág(s): 85
- ROSE, K. (1998). Deterministic Annealing for Clustering Compression, Classification, Regression and Related Optimization Problems. *Proceedings of the IEEE*, 86(11), 2210-2239. Citado na(s) pág(s): 122
- SALA-ALVAREZ, J., & VÁZQUEZ-GRAU, G. (1997). Statistical Reference Criteria for Adaptive Signal Processing in Digital Communications. *IEEE Trans. on Signal Processing*, Vol. 45(No. 1), Pp. 14-31. Citado na(s) pág(s): 9, 118, 123
- SATO, Y. (1975). A Method of Self-recovering Equalization for Multi-level Amplitude Modulation. *IEEE Trans. on Communications*, Vol. 23, Pp. 679-682. Citado na(s) pág(s): 70
- SCHARF, L. L. (1991). *Statistical Signal Processing: Detection, Estimation and Time Series Analysis*. USA: Addison-Wesley. Citado na(s) pág(s): 83
- SCHRAUDOLPH, N. N. (1995). *Optimization of Entropy with Neural Networks*. Tese de doutorado, University of California, San Diego, USA. Citado na(s) pág(s): 98
- SHALVI, O., & WEINSTEIN, E. (1990). New Criteria for Blind Deconvolution of Nonminimum Phase Systems (Channels). *IEEE Trans. on Information Theory*, Vol. 36(No. 2), Pp. 312-321. Citado na(s) pág(s): 7

- SHANNON, C. E. (1948). A Mathematical Theory of Communication. *Bell Systems Technical Journal*, Vol. 27, Pp. 379-423, 623-656. Citado na(s) pág(s): 29, 33
- SILVA, M. T. M. DA, & GERKEN, M. (2000). Um Algoritmo Acelerador para Treinamento de Redes Neurais MLP. In *Simpósio Brasileiro de Telecomunicações (SBT'2000)*. Gramado, Brasil. Citado na(s) pág(s): 60
- SMALL, C. G., & MCLEISH, D. L. (1994). *Hilbert Space Methods in Probability & Statistical Inference*. USA: John Willey & Sons. Citado na(s) pág(s): 95
- SVARER, C. (1991). *Neural Networks for Signal Processing*. Tese de doutorado, Technical University of Denmark. Citado na(s) pág(s): 8
- THERRIEN, C. W. (1992). *Discrete Random Signals and Statistical Signal Processing*. Prentice-Hall International. Citado na(s) pág(s): 41
- TOUZNI, A. (1998). *Performance et Robustesse en Egalisation Aveugle Spatio-Temporelle*. Tese de doutorado, Université de Cergy-Pontoise, France. Citado na(s) pág(s): 96, 123
- TREICHLER, J. R., & AGEE, B. G. (1983). A New Approach to Multipath Correction of Constant Modulus Signals. *IEEE Trans. on Adaptive Speech and Signal Processing*, Vol. ASSP-31 (No. 4), Pp. 349-472. Citado na(s) pág(s): 70

ÍNDICE REMISSIVO

A

Algoritmo

- back-propagation, 58
- Benveniste-Goursat, 71
- de Hopfield, 60
- Decisão Dirigida, 70
- Godard, 70
- LMS, 7
- Sato, 70
- SOFVA, 98, 99, 145
- Stop-and-Go, 71

Aprendizado, 56

- auto-organizado, 94
- competitivo, 61
- hebbiano, 61
- não supervisionado, 60
- supervisionado, 57, 99

Arvore dendrítica, 46

Atraso de decisão, 85

Atraso de decisão, 22

Axônio, 46

B

Bit, 30

Bomba de sódio, 47

Bussgang, 7, 69, 117

C

Canal

- AWGN, 18
- capacidade de, 36
- estados do, 19
- fase mínima, 103
- fase mínima, 20
- fase máxima, 20
- fase não mínima, 109
- fase não mínima, 20
- ordem do, 19
- saídas do, 19
- tipos de, 19

Canal de comunicação, 3

Classificação, 65, 72

- influência da ordem, 74

- influência do atraso, 77

Critério

- ótimo, 67

- CM, 70, 100

- máxima distorção, 68

- MD, *ver* máxima distorção

- PCM, 87

Critério sub-ótimo, 67

Critério

- cego, 123

- FPC, 120

D

Desconvolução, 4
DFE, 88
Diagrama de olho, 22
Distância
 de Mahalanobis, 39
Divergência
 assimétrica, 37
 de Kullback-Leibler, 37
 propriedades, 37

E

Entropia, 31
 condicional, 35
 cruzada, 37
 da fonte, 32
 de Renyi, 33
 diferencial, 37
 limites, 33
Equalização, 65
 adaptativa, 68
 autodidata (cega), 69
 não-supervisionada, *ver* autodidata
 (cega)
 supervisionada, 68
 clássica, 66
Equalizador
 de magnitude e fase, 86
Erro
 de predição, 82
Estimação
 de fdp, 38
 não paramétrica, 43
 paramétrica, 38
Estimador
 Busgang, 72
 não linear, 69

F

Filtragem
 de Wiener, 67
 inversa, 67
Filtro
 transversal, 66
Fitting pdf Algorithm, 122, 147

Fitting pdf Criterion, 122
FPA, *ver* Fitting pdf Algorithm, *ver*
 Fitting pdf Algorithm
FPC, *ver* Fitting pdf Criterion
Função
 de ativação, 47, 51
 gaussiana (RBF), 55
 interpolação não linear, 91
 multiquadrática, 55
 multiquadrática inversa, 55
 Siebert, 52, 53
 sigmóide, 52, 53
 tangente hiperbólica, 52, 53
Função
 alvo, 124
 de referência, 37
 de transferência, 18
 densidade de probabilidade (fdp)
 gaussiana, 38
 densidade gaussiana
 propriedades, 39

I

Identificação inversa, 67
Identificação inversa, 4
Incerteza, *ver* Informação
Inferência bayesiana, 41
Informação, 30
 fonte, 16
 mútua, 34
 propriedades, 35
 quantidade de, 30
 propriedades, 31
 taxa de, 34
 teoria da, 29
Interferência entre símbolos, 21
Interpolação, 89

L

Lei
 anti-hebbiana, 62
 hebbiana, 62
 não hebbiana, 62
Lema
 predição com memória unitária, 90

Limiar de disparo, 47

M

Máquina de aprendizado, 57

Mapas de Kohonen, 61

Matriz de convolução, 119

Medidas de desempenho, 23

erro quadrático médio, 24

métodos de Monte Carlo, 25

relação interferência entrada-saída, 24

relação Sinal-Ruído, 23

taxa de símbolos errados, 24

Modelo

canal de comunicação, 17

em banda básica, 15

ruído gaussiano, 17

Máxima verossimilhança, 41

N

Neurônio

artificial, 48

de Macculloch-Pitts, 49

de Rosenblatt, 50

biológico, 46

Neural Predictor Structure, 91

NPS, *ver* Neural Predictive Structure

NPS-CMC, 100

adaptação, 146

NPS-MPE, 100

adaptação, 146

P

Período

de treinamento, 68

Perceptron, 50

Pesos sinápticos, 50

Potencial de ação, 47

Predição

linear, 82

não linear, 81

Preditor

não linear, 85

Processo de branqueamento, 81, 82

R

Recuperação de fase, 125

Rede multiponto, 5

Redes neurais artificiais, 45, 52

de Hoppfield, 56

multicamadas, 52

RBF, 54

recorrentes, 56

Relação

sinal-interferência, 24

sinal-ruído, 23

Resposta

global, 68

RNA, *ver* Redes Neurais Artificiais

S

Saída do equalizador

fdp, 118

Sinapse, 47

hebbiana, 62

Sistema

de comunicação digital, 13

T

Teorema

fdp sinal recebido, 93

W

Wiener, 7

ÍNDICE DE AUTORES

A

Agee, B. G., 70, 166

B

Balay, P., 8, 61, 161
Barreto, G. A., 8, 161
Behm, J. D., 163
Benedetto, S., 14, 24, 161
Benveniste, A., 71, 121, 161
Berberidis, K., 162
Biglieri, E., 161
Bishop, C. M., 32, 39, 41–44, 122, 161
Boné, R., 8, 161
Brossier, J.-M., 22, 66, 68, 161
Brown, D. R., 163

C

Casas, R. A., 163
Castellani, V., 161
Cavalcante, C. C., 90, 93, 162, 164
Cavalcanti, F. R. P., 8, 68, 85, 87, 100, 125, 162
Czernichow, T., 8, 162

D

Deco, G., 37, 44, 162
Dorizzi, B., 162, 164

Duda, R. O., 39, 44, 162

E

Endres, T. J., 163

F

Forney, G. D., Jr., 67, 162
Friedlander, B., 69, 165

G

Gerken, M., 60, 166
Gibson, J. D., 30, 162
Glentis, G.-O., 71, 162
Godard, D. N., 70, 162
Goursat, M., 71, 161
Gu, Y., 7, 163

H

Hachicha, A., 7, 163
Hart, P. E., 39, 44, 162
Harvey, R. L., 47, 49, 162
Hassoun, M. H., 47, 51, 162
Haykin, S., 4, 7, 9, 14, 29, 30, 32, 34, 35, 37, 44, 47, 55, 56, 60–62, 67, 69, 70, 83, 98, 101, 123, 162, 163
Hilal, K., 69, 163

J

Johnson, C. R., Jr., 70, 163

K

Kovács, Z. L., 46, 47, 49–52, 56, 58, 163

L

Labat, J., 8, 85, 163

Laot, C., 85, 163

M

Macchi, O., 7, 8, 83, 102, 163, 165

McCulloch, W., 50, 163

McLeish, D. L., 95, 166

Montalvão, J. R., Filho, 4, 17, 23, 39, 63,
65–68, 72, 76, 78, 102, 115, 120,
122, 127, 162, 164

Mota, J. C. M., 7, 8, 20, 21, 67, 68, 85,
86, 162, 164

Mulgrew, B., 8, 63, 72, 164

O

Obradovic, D., 37, 44, 162

Oppenheim, A. V., 15, 18, 140, 164

P

Papadias, C., 14, 70, 164

Papoulis, A., 17, 41, 82, 164

Paulraj, A., 7, 164

Picchi, G., 71, 164

Pitts, W., 50, 163

Porat, B., 69, 165

Prati, G., 71, 164

Proakis, J. G., 4, 14, 15, 18, 22, 24, 67,
165

Q

Qureshi, S. U. H., 4, 165

R

Regalia, P. A., 123, 165

Rényi, A., 33, 165

Rocha, C. A. F. da, 6, 8, 16, 69, 82, 83,
85, 86, 165

Romano, J. M. T., 165

Rose, K., 122, 165

Roychowdhury, V., 164

Ruget, G., 161

S

Sala-Alvarez, J., 9, 118, 123, 165

Sato, Y., 70, 165

Schafer, R. W., 15, 18, 140, 164

Schaper, C., 164

Scharf, L. L., 83, 165

Schniter, P., 163

Schraudolph, N. N., 98, 165

Shalvi, O., 7, 165

Shannon, C. E., 29, 33, 166

Silva, M. T. M. da, 60, 166

Small, C. G., 95, 166

Svarer, C., 8, 166

T

Theodoridis, S., 162

Therrien, C. W., 41, 166

Touzni, A., 96, 123, 166

Treichler, J. R., 70, 166

V

Vázquez-Grau, G., 9, 118, 123, 165

W

Weinstein, E., 7, 165