

Detecção Bayesiana com Redes NRBF em Canais com Desvanecimento Rayleigh

Wamberto José Lira de Queiroz

Dissertação de Mestrado submetida à Coordenação dos Cursos de Pós-Graduação em Engenharia Elétrica da Universidade Federal da Paraíba - Campus II como parte dos requisitos necessários para obtenção do grau de Mestre no domínio da Engenharia Elétrica .

Área de Concentração: Processamento da Informação

Marcelo Sampaio de Alencar, Ph.D.
Orientador

Campina Grande, Paraíba, Brasil
©Wamberto José Lira de Queiroz, 20-06-00



Q3d

Queiroz, Wamberto Jose Lira de

Detecção bayesiana com redes NRBf em canais com desvanecimento rayleigh / Wamberto Jose Lira de Queiroz. - Campina Grande, 2000.

85 f.

Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal da Paraíba, Centro de Ciências e Tecnologia.

1. Detecção Bayesiana 2. Redes NRBf 3. Efeito Doppler 4. Dissertação I. Alencar, Marcelo Sampaio de II. Universidade Federal da Paraíba - Campina Grande (PB)

CDU 52-67(043)

**DETEÇÃO BAYESIANA COM REDES NRBF EM CANAIS COM
DESVANECIMENTO RAYLEIGH**

WAMBERTO JOSÉ LIRA QUEIROZ

Dissertação Aprovada em 20.06.2000



PROF. MARCELO SAMPAIO DE ALENCAR, Ph.D., UFPB
Orientador



PROF. JOSÉ EWERTON POMBO DE FARIAS, Dr., UFPB
Componente da Banca



PROF. BENEDITO GUIMARÃES AGUIAR NETO, Dr.-Ing., UFPB
Componente da Banca

CAMPINA GRANDE - PB
Junho - 2000

Dedicatória

Dedico este trabalho principalmente a Deus que me deu luz e força para que eu pudesse concluí-lo e aos meus pais, José Velozo de Queiroz e Vanda Lira de Queiroz.

Agradecimentos

- A Deus, por tudo;
- Aos meus pais, Vanda Lira de Queiroz e José Velozo de Queiroz, pelo apoio e incentivo constante;
- Ao meu orientador Marcelo Sampaio de Alencar, pela orientação, disponibilidade e dedicação durante todo o trabalho;
- A Juraci pela ajuda ao acompanhar o trabalho e pelas sugestões;
- Ao pessoal do LABCOM: Waslon e Edmar;
- Ao pessoal do LAPS: Madeiro, Joseana, Luciana, Robson, Rinaldo e Paulo Márcio;
- Em especial às amigas Izabel e Suzete pela sincera amizade e pela força;
- Ao CNPq pelo financiamento do trabalho;
- Aos funcionários da COPELE e do DEE.

Resumo

Neste trabalho foi estudado o problema da estimação de símbolos QPSK em um modelo de canal de comunicações seletivo em frequência e com desvanecimento Rayleigh. O estudo concentrou-se na avaliação do desempenho, por meio de simulação, do método de detecção bayesiano implementado com uma rede neural artificial com funções de base radiais. Duas estruturas de detecção foram avaliadas, usando-se como critério de desempenho a taxa de erro de símbolos: na primeira estrutura desconsiderou-se a informação acerca dos símbolos previamente detectados e na segunda estrutura esses símbolos foram levados em consideração. Essas estruturas são referidas como equalizadores bayesianos sem e com *retroação*.¹ As duas estruturas foram inicialmente avaliadas e comparadas para um modelo de canal conhecido e verificou-se pouca melhora no desempenho da estrutura com realimentação em relação à estrutura sem realimentação. Em seguida as duas estruturas foram avaliadas em um modelo de canal com coeficientes variantes. Os resultados obtidos mostraram que a estrutura com realimentação teve desempenho superior à estrutura sem realimentação, à medida que a frequência Doppler aumentava. Tal característica viabiliza o uso dessa estrutura em sistemas em que o canal apresenta desvanecimento rápido.

¹Embora o termo "equalização" seja mais apropriado para especificar a operação de inversão da resposta ao impulso do canal de transmissão, ele será usado durante todo o texto no sentido de estimação de símbolos enviados pela fonte.

Abstract

This work presents results of the application of Bayesian detection, along with neural networks and radial basis functions (NRBF) to the channel equalization problem. Equalization is used here in the sense of recovering degraded symbols from communication channels. The purpose of this study is therefore, to model the main effects of the channel and propose a solution that relies on NRBF. As a result of this work, two detection structures were implemented in order to evaluate the performance of the detection schemes.

The Bayesian hypothesis testing is a statistical method used in pattern classification and assumes that the decision problem is based on known probabilities. An introduction to the theory is initially presented, using binary symbols for simplicity, to ensure that the concepts can be extended to the equalization problem. The detection structures, with and without feedback, are based on this method.

The adopted modulation scheme was quadrature phase shift keying (QPSK), which is used by several mobile cellular air interfaces, including the code division multiple access (CDMA) system. But, the obtained structures can be used with other modulation schemes.

The equalizer structure is similar to an NRBF. This, certainly, influenced the used of this type of neural network in the implementation of the Bayesian method. Other factors, including implementation facility, fast training and small complexity also were taken into consideration. The detection structures were compared and, for a channel with Doppler effect, the Bayesian feedback detector showed a superior performance.

Lista de Símbolos e Abreviaturas

ISI - Interferência Intersimbólica (*Intersymbolic Interference* - *ISI*)

WSS – US - Estacionário em Sentido Amplo (*Wide Sense Stationary-Uncorrelated Scattering* - *WSS-US*)

NRBF - Redes Neurais com Funções de Base Radiais (*Network Radial Basis Functions* - *NRBF*)

RBF - Funções de Bases Radiais (*Radial Basis Functions* - *RBF*)

LMS - Algoritmo LMS (*Least Mean Square* - *LMS*)

PN - Seqüência Pseudo-Aleatória (*Pseudo Noise* - *PN*)

RNA - Rede Neural Artificial (*Artificial Neural Network* - *RNA*)

ART - Teoria da Ressonância Adaptativa (*Adaptive Resonance Theory* - *ART*)

MLP - Perceptron Multicamadas (*Multilayer Perceptron* - *MLP*)

FIR - Filtro de Resposta ao Impulso Finita (*Finite Impulse Response*- *FIR*)

QPSK - Modulação em Fase e Quadratura (*Quadrature Phase Shift Keying* - *QPSK*)

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

DSP - Densidade Espectral de Potência (*Potency Spectral Density* - *DSP*)

MLSE - Estimção por Seqüências de Máxima Probabilidade - (*Maximum Likelihood Sequence Estimation* - *MLSE*)

DFE - Equalizador com retroação de símbolos (*Decision Feedback Equaliser* - *DFE*)

TE - Equalizador Transversal - (*Transversal Equaliser* - *TE*)

σ_n^2 - Variâncias das amostras de ruído gaussiano aditivo

\hat{r} - Vetor de amostras visto pelo detetor, sem ruído adicionado

$n(i)$ - Amostras de ruído gaussiano aditivo

f_i - Coeficientes do modelo discreto do canal de múltiplos percursos

n_a - Número de coeficientes do modelo do canal
 $p_X(x)$ - Função densidade de probabilidade (fdp)
 $I_0(\cdot)$ - Função de Bessel de ordem zero
 $S(\tau)$ - Perfil de intensidade do desvanecimento de multipercurso
 T_m - Atraso máximo em excesso
 T_s - Tempo de símbolo
 f_o - Banda passante coerente
 θ - Ângulo de incidência de uma onda de rádio recebida por um móvel
 $v(t)$ - Velocidade com que um móvel se afasta ou se aproxima de uma antena transmissora
 β - Constante que depende da permissividade elétrica e da permeabilidade magnética do meio
 $\theta(t)$ - Fase instantânea do sinal transmitido
 $s(t)$ - Sinal contínuo transmitido
 $R_s(\tau)$ - Autocorrelação do sinal $s(t)$
 ϕ - Deslocamento de fase do sinal $s(t)$
 A - Amplitude do sinal transmitido $s(t)$
 $S_s(\omega)$ - Densidade espectral de um sinal proveniente de um canal com efeito Doppler
 ω_m - Frequência Doppler
 ω_c - Frequência da portadora do sinal transmitido
 $h(t, \tau)$ - Processo estocástico usado para modelar o canal
 $S(\tau; \lambda)$ - Função de espalhamento do canal
 $\Phi_C(\Delta f; \Delta t)$ - Função de autocorrelação do processo $h(t, \tau)$
 \mathbf{c} - Vetor de coeficientes do filtro FIR usado na estimação do canal
 $d(n)$ - Amostras desejadas usadas no algoritmo LMS
 J - Função de custo
 $E[\cdot]$ - Operador valor esperado
 ∇ - Operador gradiente
 R_{xx} - Autocorrelação entre dois vetores \mathbf{x} de amostras
 R_{xd} - Correlação entre um vetor \mathbf{x} e uma amostra escalar $d(n)$
 \mathbf{c}_{ot} - Vetor de coeficientes ótimos

μ - Parâmetro de ajuste dos coeficientes do algoritmo LMS
 \mathbf{x}^H - Operação transposta conjugada
 σ_d^2 - Variância da saída desejada $d(n)$ do algoritmo LMS
 $\|\cdot\|$ - Norma euclidiana
 \mathbf{t}_j - j -ésimo vetor centro das funções-núcleo de uma rede RBF genérica
 $\varphi(\cdot)$ - Função núcleo ou função de base radial
 \mathbf{w} - Vetor de coeficientes da camada linear da rede RBF genérica
 ρ - Parâmetro que determina a largura da função núcleo
 Γ - Espaço de observação
 Γ_i - i -ésima partição do espaço de observação Γ
 H_i - i -ésima hipótese associada à i -ésima partição do espaço de observação Γ
 $\eta(x)$ - Variável de decisão bayesiana
 $R_j(\eta)$ - Custo médio associado à regra de decisão $\eta(x)$ quando a hipótese H_j é verdadeira
 $r(\eta)$ - Risco médio de Bayes
 C_{ij} - Custo que se tem ao escolher a hipótese H_i quando a hipótese H_j é verdadeira
 $L(x)$ - Taxa de probabilidade entre as hipóteses H_1 e H_0
 $s(k)$ - Símbolos QPSK em banda-base
 $R_{m,d}$ - Conjunto formado por todos os estados do modelo do canal
 $h^{(i)}$ - Variável associada às probabilidades a priori dos vetores estado $\hat{\mathbf{r}}$
 $\text{sgn}(\cdot)$ - Função sinal complexa
 n_s - Número de estados do modelo do canal
 d - Atraso de decisão do detetor
 n_f - Número de estados de *feedback*
 $\hat{\mathbf{s}}_f$ - Vetor de *feedback*
 m - Ordem direta do detetor
 P_i - Potência normalizada da i -ésima linha de atraso do modelo do canal.

Lista de Figuras

2.1	Modelo geral de um sistema de comunicações	5
2.2	Modelo equivalente discreto no tempo de um canal com ISI	7
2.3	Perfil da Intensidade de Multipercurso	11
2.4	Função de correlação entre dois sinais saídos de um canal dispersivo.	13
4.1	Rede de Funções de Base Radiais	39
5.1	Estrutura genérica do equalizador transversal.	48
5.2	Equalizador bayesiano com feedback	48
5.3	Modelo do sistema de comunicações.	49
5.4	Estrutura do equalizador bayesiano sem <i>feedback</i>	53
5.5	Processamento dos dados no equalizador bayesiano.	57
6.1	Modelo do Sistema Implementado.	60
6.2	Distribuição do módulo das amostras do processo $h(t, \tau)$ entre o seu valor máximo e seu valor mínimo.	61
6.3	Módulo da resposta impulsional do canal.	62
6.4	Autocorrelação do processo $h(t, \tau)$, para τ fixo.	62
6.5	Densidade espectral de potência do processo $h(t, \tau)$ simulado.	63
6.6	Comparação entre o equalizador bayesiano sem <i>feedback</i> e o método de estimação de seqüências de máxima verosimilhança para o modelo de canal conhecido $H(z) = (0.0119 + j0.7589) + (-0.3796 + j0.5059)z^{-1}$	65

6.7	Comparação entre o desempenho do equalizador bayesiano sem <i>feedback</i> , respectivamente com ordem da linha direta $m = 2$, $m = 4$ e $m = 6$ para o modelo de canal conhecido $H(z) = (0.0119 + j0.7589) + (-0.3796 + j0.5059)z^{-1}$	66
6.8	Comparação entre o desempenho do equalizador bayesiano sem <i>feedback</i> e com <i>feedback</i> para o modelo de canal conhecido $H(z) = (0.0119 + j0.7589) + (-0.3796 + j0.5059)z^{-1}$	67
6.9	Comparação entre o desempenho do equalizador bayesiano com <i>feedback</i> para $m = 3$ e $m = 4$ e para o modelo de canal conhecido $H(z) = (0.0119 + j0.7589) + (-0.3796 + j0.5059)z^{-1}$	68
6.10	Diagrama usado para a estimação dos coeficientes do modelo do canal usando o algoritmo LMS.	69
6.11	Curvas de convergência do erro médio quadrático do LMS.	69
6.12	Trajetória dos pesos do processador adaptativo para $\mu = 0.1$	70
6.13	Desempenho do equalizador bayesiano com <i>feedback</i> em um canal variante.	72
6.14	Desempenho do equalizador bayesiano com <i>feedback decision</i> em um modelo de canal variante e com parâmetros $m = 4$, $n_a = 2$ e $d = 1$	73
6.15	Desempenho do equalizador bayesiano sem <i>feedback decision</i> em um modelo de canal variante e com parâmetros $m = 2$, $n_a = 2$ e $d = 1$	73
6.16	Distribuição dos estados do canal nas regiões de decisão sem <i>feedback</i> para o modelo de canal $H(z) = (0.7409 - j0.7406) + (-0.29648 + j0.88884)z^{-1} + (-0.022191 - 0.155562)z^{-2}$	74
6.17	Distribuição dos estados do canal nas regiões de decisão com <i>feedback</i> e considerando-se um vetor de <i>feedback</i> $\hat{s}_f(k) = [1 + j \ 1 + j]^T$ para o modelo de canal conhecido $H(z) = (0.7409 - j0.7406) + (-0.29648 + j0.88884)z^{-1} + (-0.022191 - j0.155562)z^{-2}$	75
6.18	Distribuição dos valores da variável de decisão final em torno dos estados do canal $H(z) = (0.7409 - j0.7406) + (-0.29648 + j0.88884)z^{-1} + (-0.022191 - j0.155562)z^{-2}$	76

Índice

1	Introdução	1
1.1	Motivação	1
1.2	Organização da Dissertação	2
2	Modelamento do Canal de Comunicações	4
2.1	Introdução	4
2.2	Obtenção do Modelo do Canal	5
2.2.1	Caracterização Matemática do Modelo	6
2.3	Categorias de Desvanecimento do Canal de Transmissão	9
2.3.1	Perfil da Intensidade do Desvanecimento de Multipercurso	10
2.3.2	Categorias de Degradação Devido ao Espalhamento no Tempo	11
2.3.3	Espalhamento do Sinal no Tempo Visto no Domínio da Frequência	12
2.4	Obtenção do Espectro de Potência do Sinal Recebido	13
2.5	Modelo do Canal Usado para a Simulação	16
2.6	Conclusão	18
3	Algoritmo dos Passos Descendentes e o Algoritmo LMS	19
3.1	Introdução	19
3.2	O Problema da Filtragem Ótima	20
3.3	Método dos Passos Descendentes	23
3.4	Algoritmo LMS	25
3.5	Conclusão	28

4	Introdução às Redes Neurais Artificiais	29
4.1	Introdução	29
4.2	Treinamento e Teste das RNA's	31
4.3	Definição das Redes Neurais Classificadoras	33
4.3.1	Classificadores Probabilísticos	33
4.3.2	Classificadores de Hiperplano	33
4.3.3	Classificadores de Núcleo - <i>Kernel</i>	34
4.3.4	Classificadores de Exemplos	35
4.4	Redes com Funções de Base Radiais	35
4.5	O Teorema de Cover da Separabilidade de Padrões	36
4.6	Formato da Rede NRBF	38
4.7	Rede NRBF de Valores Complexos	40
4.8	Conclusão	41
5	Detecção Bayesiana	42
5.1	Introdução	42
5.2	Teste de Hipóteses Bayesiano	43
5.3	Aplicação da Detecção Bayesiana à Equalização	47
5.4	Conclusão	58
6	Descrição do Sistema Simulado e Análise dos Resultados	59
6.1	Introdução	59
6.2	Simulação do Canal	60
6.3	Simulação de Equalizador	64
6.4	Conclusão	76
7	Conclusão	78
7.1	Contribuições e Sugestões de Continuidade	80

Capítulo 1

Introdução

1.1 Motivação

A crescente demanda por serviços integrados que reúnem a transmissão de dados escritos, sinais de som e imagem, tem incentivado pesquisadores e empresas prestadoras de serviços de comunicações a buscarem soluções tecnológicas que possibilitem o atendimento das necessidades dos usuários, garantindo qualidade a baixos custos. Essa necessidade de começou a surgir nas grandes redes telefônicas desenvolvidas para a transmissão de sinais de voz. Atualmente, com o avanço tecnológico obtido nos meios de comunicações ópticos e com os modernos *modems* é possível enviar, através de uma linha telefônica, sinais de som, imagem, imagens em movimento e outras formas de informação que os atuais sistemas comunicação permitem difundir.

Em sistemas que fazem a difusão da informação por meio de redes telefônicas fixas é necessário o uso de *modems* que forneçam um sinal modulado e adequado ao meio de transmissão. Já o sinal recebido na estação destino, que pode ser a estação do próprio usuário ou mesmo a prestadora do serviço, é uma forma de onda corrompida por ruído aditivo, devido à possibilidade do surgimento do ruído térmico, e pelo desvanecimento, bem como distorções ocasionadas pelas limitações físicas do canal de transmissão. As alterações impostas ao sinal transmitido são a translação em frequência, a distorção harmônica e a dispersão no tempo. Em uma linha telefônica, por exemplo, a dispersão no tempo ocorre quando a resposta limitada em faixa do canal se desvia da amplitude

constante e da fase linear.

Um dos problemas mais comuns nos meios de transmissão via cabo é a largura de faixa limitada do meio de transmissão. Em determinadas frequências o sinal transmitido é atenuado a ponto de impossibilitar sua detecção no receptor. A limitação da banda passante geralmente resulta em uma distorção que é conhecida como interferência intersimbólica (*Intersymbol Interference - ISI*), na qual os símbolos adjacentes se sobrepõem uns aos outros. Essa distorção é um dos maiores obstáculos a transmissões, com altas taxas de dados, em canais limitados em faixa e com ruído aditivo. Nesses sistemas se faz necessário o uso de estruturas que combatam os efeitos da ISI e permitam a recuperação da informação com menor taxa de erros.

Em canais de rádio e canais submersos, a ISI ocorre devido à propagação do sinal transmitido por meio de múltiplos percursos. Essa propagação pode ser vista como uma transmissão através de um grupo de canais distintos que impõem ao sinal transmitido um atraso e uma variação de amplitude diferentes. Uma exigência especial de um equalizador de canal de rádio é que ele seja capaz de se adaptar às características de desvanecimento variantes no tempo, tipicamente encontradas nesse tipo de canal. Nessas estruturas, a velocidade com que elas conseguem ajustar seus pesos e parâmetros para compensar a variação da característica do canal é um aspecto muito importante. Essa velocidade de ajuste dos pesos é chamada de taxa de convergência e é um dos fatores que determinam a escolha de um determinado algoritmo ou estrutura. Vários outros fatores devem ser levados em consideração ao se preferir uma estrutura a outra. É importante, entretanto, que se tenha conhecimento do problema a ser combatido para que se possa apresentar as possíveis soluções. É por esse motivo que este texto começa com uma análise do canal de comunicações no qual se pretende aplicar o dispositivo equalizador e em seguida são apresentadas algumas possíveis soluções.

1.2 Organização da Dissertação

O texto desta dissertação segue a seguinte linha geral. No Capítulo 2 é feita uma análise do modelo do canal usado na simulação. É apresentado o modelo com desvanecimento por múltiplos percursos e os impactos desse modelo na informação transmitida.

Em seguida é apresentado o modelo matemático que representa o canal e o método de simulação utilizado.

No Capítulo 3 é analisada a equação de Hoff-Wiener, o algoritmo dos passos descendentes e o decorrente algoritmo (*Least Mean Square - LMS*).

No Capítulo 4 é dada uma introdução às redes neurais artificiais e é analisado o uso dessas redes como estruturas de classificação de padrões. São citados os tipos de classificadores e em seguida é dada uma introdução às redes com funções de bases radiais.

No Capítulo 5 é feita uma introdução ao método de detecção bayesiana e a sua aplicação ao problema de equalização usando a rede NRBF. É mostrada a ligação dessa solução com o formato da rede utilizada.

No Capítulo 6 é feita uma análise do procedimento de simulação adotado, dos parâmetros utilizados, dos resultados obtidos e em seguida, no Capítulo 7, são apresentadas conclusões e sugeridas algumas melhorias para o sistema implementado.

Capítulo 2

Modelamento do Canal de Comunicações

2.1 Introdução

A maioria dos sistemas de comunicações apresenta características que os fazem atuar sobre a informação de diferentes formas. A caracterização desses sistemas é um tópico importante e envolve a investigação de diversos fatores.

Em sistemas de comunicações móveis, por exemplo, um dos principais fatores que devem ser levados em consideração para o desenvolvimento de sistemas eficientes e confiáveis é o conhecimento das características do canal. Esse é um tópico que tem recebido muita atenção de pesquisadores interessados na tecnologia das comunicações móveis. A cada dia, aumenta a demanda por serviços que forneçam altas taxas de dados, alto desempenho e baixos custos para os usuários. Essas exigências devem ser satisfeitas independentemente das intempéries às quais a informação se submete em seu percurso da fonte até o receptor.

Na tentativa de melhorar a qualidade do sinal recebido e fazer com que a informação perdida, em decorrência dos erros causados pela natureza do meio e da influência de outros usuários, diminua significativamente, tem-se desenvolvido técnicas que têm aumentado muito o desempenho das estruturas de recepção.

Nesse processo de recuperação da informação são levados em consideração alguns fa-

tores como as relações custo/benefício, desempenho/complexidade, tempo de resposta e vários outros. Tendo em mente a consciência de que esses fatores são imprescindíveis na elaboração de qualquer estudo, são analisadas neste trabalho algumas estruturas que visam minimizar, conjuntamente com os outros componentes da estrutura de recepção, os efeitos destrutivos causados, por exemplo, pelo comportamento dispersivo do canal, pelas variações de amplitude no sinal, pelo efeito Doppler e pela largura de faixa limitada pelo meio físico de comunicação.

As estruturas que, juntas com as outras partes do receptor, podem minimizar os efeitos citados serão chamadas neste trabalho de equalizadores de canais. No diagrama geral de um sistema de comunicações, essas estruturas ocupam posição antes do demodulador e do decodificador. Um diagrama em blocos, como o da Figura 2.1, ilustra melhor o que foi citado.

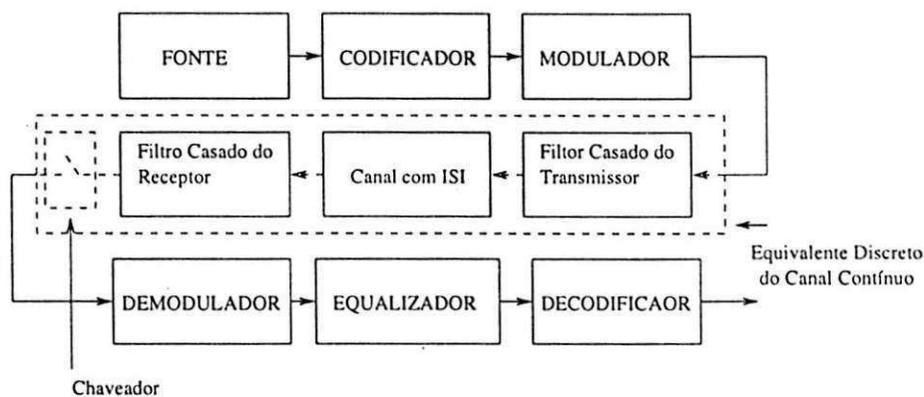


Figura 2.1: Modelo geral de um sistema de comunicações

2.2 Obtenção do Modelo do Canal

De um modo geral a palavra canal pode ser usada para dar significado a tudo que se encontra entre a fonte de sinal e o destino [1]. Desse modo, essa definição geral incluiria também todos os equipamentos de transmissão e recepção bem como o meio físico por meio do qual a informação seria irradiada ou conduzida. Entretanto, neste trabalho o termo canal designará apenas o modelo do meio de transmissão.

Basicamente, um modelo de canal pode ser pensado como uma representação, em forma matemática ou algorítmica, para a função de transferência de cada um dos blocos que compõem o canal. Em relação ao meio externo devem ser levados em consideração todos aqueles parâmetros que exercem uma forte influência na característica do meio. Alguns desses parâmetros incluem a banda passante, as características do ambiente físico e a frequência da portadora [1].

Em ambientes urbanos, por exemplo, o fluxo de automóveis, a topografia do terreno, a quantidade de edificações e a área geográfica são fatores relevantes na elaboração do modelo. Na maioria desses canais o sinal transmitido sofre inversões de fase, devido às reflexões nos obstáculos presentes no meio. Essas variações de fase fazem com que as amostras de sinal recebidas no receptor se somem destrutivamente ou construtivamente.

O sinal que chega ao receptor, composto por todas as componentes atrasadas e distorcidas, é dito desvanecido. Nesse contexto, muitos pesquisadores têm proposto modelos que contemplam uma grande parte dos efeitos indesejáveis dos canais de rádio móvel [1], [2].

2.2.1 Caracterização Matemática do Modelo

Será analisado nesta seção um modelo discreto para o canal com desvanecimento por múltiplos percursos. Como se sabe, os diferentes percursos pelos quais a informação trafega impõem a um mesmo pulso transmitido variações de amplitude e no tempo de chegada desse pulso ao receptor. Essa natureza do canal pode ser modelada por conjunto de linhas de atraso compostas por um coeficiente de valor complexo. Cada linha corresponde a um percurso que afeta de forma diferente a característica original do sinal. Esse conjunto de linhas de atraso é representado por um filtro (*Finite Impulse Response - FIR*) cujos coeficientes representam as variações impostas à amplitude e à fase do sinal transmitido. Pode-se perceber que a saída desse filtro corresponde à soma de todos os componentes de sinais provenientes dos diferentes percursos representados por cada uma das linhas de atraso do filtro. O que essa estrutura faz na verdade é uma convolução da seqüência de amostras de sinais de entrada com os coeficientes do filtro. Esse modelo discreto obtido de forma intuitiva é na verdade um equivalente discreto

para o modelo de canal formado pela ligação em cascata de um filtro casado à resposta do canal no transmissor, da resposta ao impulso no tempo do canal, de um filtro casado no receptor e de um amostrador [3]. A adição de amostras de ruído gaussiano branco de média nula e variância σ_n^2 à saída do equivalente discreto completa o modelo a ser usado e é ilustrado na Figura 2.2.

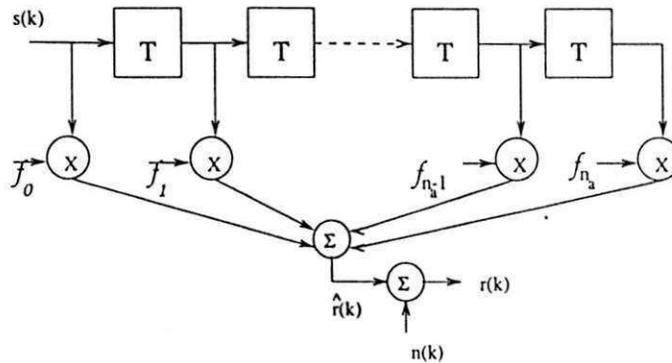


Figura 2.2: Modelo equivalente discreto no tempo de um canal com ISI

De acordo com a Figura 2.2 a saída desse modelo de canal pode ser escrita matematicamente na forma

$$\begin{aligned}\hat{r}(k) &= \sum_{i=0}^{n_a-1} f_i s_{k-i} \\ r(k) &= \hat{r}(k) + n_k,\end{aligned}\tag{2.1}$$

em que $\{n_k\}$ é a seqüência de amostras de ruído gaussiano branco de valor complexo, com a parte real independente da parte imaginária, $\{f_i\}$ é a seqüência de coeficientes do filtro FIR que representa o modelo do canal, n_a é o número de linhas de atraso e T representa o atraso dado ao sinal transmitido e que nesse trabalho é considerado constante. Diz-se que um modelo como esse com n_a linhas de atraso tem memória igual a $n_a - 1$. Os elementos de memória colocados no modelo do canal, inerentes à estrutura do filtro FIR, é que fazem com que um mesmo pulso colocado na entrada da estrutura, chegue na saída com diferentes atrasos.

O modelo de filtro transversal discreto no tempo será usado neste trabalho para modelar os efeitos da ISI que aparecem na transmissão de sinais em canais móveis

sujeitos ao efeito de multipercuso. Esse modelo será usado nas próximas seções para o estudo da técnica compensação da interferência intersimbólica e avaliação da estrutura de detecção utilizada.

Continuando a caracterização do canal, será feita em seguida a análise estatística dos coeficientes usados no filtro FIR. Para essa caracterização existem várias distribuições de probabilidade que podem ser levadas em conta. Nos canais móveis, nos quais as formas de onda transmitidas sofrem um atraso e uma variação na amplitude em cada percurso entre o transmissor e o receptor, um grande número de obstáculos no percurso da informação trafegante faz com que um grande número de versões atrasadas do sinal transmitido chegue ao receptor em um mesmo instante de tempo. Para cada linha de atraso do filtro esse processo resulta, pela aplicação do teorema central do limite, em um processo gaussiano para a resposta ao impulso do canal. Esse processo tem média zero, a envoltória da resposta ao impulso do canal, em qualquer instante de tempo, tem uma distribuição Rayleigh e a fase tem uma distribuição uniforme no intervalo $[0, 2\pi)$ [3], [4]. A distribuição Rayleigh é dada por

$$p_X(x) = \frac{2x}{\sigma_X^2} e^{-x^2/\sigma_X^2}, \quad x \geq 0, \quad (2.2)$$

em que $\sigma_X^2 = E(X^2)$.

Uma outra distribuição que também pode ser usada no modelamento da envoltória da resposta ao impulso do canal é a distribuição de Rice que terá seu uso justificado nas próximas seções e que pode ser expressa pela função densidade de probabilidade (fdp)

$$p_X(x) = \frac{x}{\sigma_X^2} e^{-(x^2+s^2)/2\sigma_X^2} I_0\left(\frac{xs}{\sigma^2}\right), \quad r \geq 0, \quad (2.3)$$

em que $I_0(\cdot)$ é a função de Bessel de ordem zero. Usando os pesos obtidos por uma dessas distribuições citadas, pode-se simular o canal, usando-se a estrutura FIR. Na próxima seção são apresentadas algumas das categorias de desvanecimento encontradas em um canal móvel.

2.3 Categorias de Desvanecimento do Canal de Transmissão

Para iniciar esta seção é conveniente destacar os dois tipos de efeitos de desvanecimento que afetam o canal de comunicações: desvanecimento em pequena escala e desvanecimento em grande escala. O desvanecimento em grande escala representa a atenuação na potência média do sinal transmitido devido ao movimento em grandes áreas. Esse tipo de desvanecimento é afetado por terrenos com contornos acentuados, como morros, florestas, áreas urbanas, etc. As estatísticas do desvanecimento em grande escala fornecem uma maneira de calcular uma estimativa das perdas de percurso como função da distância. Essa estimativa é descrita em termos de uma perda média no percurso e uma variação com distribuição de probabilidade log-normal em torno da média.

O desvanecimento em pequena escala, também conhecido como desvanecimento Rayleigh, refere-se às variações na amplitude e na fase do sinal transmitido, resultantes de pequenas mudanças na separação espacial entre o receptor e o transmissor. Esse tipo de desvanecimento é ocasionado por meio de dois mecanismos que são o espalhamento do sinal no tempo e o comportamento variante no tempo do canal. Em aplicações rádio-móvel, o canal é variante no tempo porque o movimento entre o transmissor e o receptor resulta em mudanças no percurso de propagação. A taxa de mudança dessas condições de propagação contribui para a rapidez do desvanecimento.

Desvanecimento em pequena escala também é conhecido como desvanecimento Rayleigh, porque se o número de percursos reflexivos é grande e se não existe um componente de sinal direto, a envoltória do sinal recebido é estatisticamente descrita por uma fdp de Rayleigh. Por outro lado, quando existe um componente não desvanecido dominante, como em um percurso de propagação direto, o envelope do desvanecimento em pequena escala é descrito por uma fdp de Rice.

Existem três mecanismos básicos que têm impacto na propagação do sinal em um sistema de comunicações móveis. Eles são respectivamente a reflexão, a difração e a dispersão:

- Reflexão ocorre quando uma onda eletromagnética colide com uma superfície lisa, com dimensões muito grandes comparadas ao comprimento de onda λ do sinal de rádio frequência.
- Difração ocorre quando o percurso entre o transmissor e o receptor é obstruído por um corpo denso, com grandes dimensões comparadas a λ , fazendo com que ondas secundárias sejam formadas atrás do corpo denso. Esse é um fenômeno que justifica o transporte de energia do transmissor para o receptor quando não há um percurso direto entre os dois e é também chamado de *sombreamento* porque o campo difratado pode chegar ao receptor mesmo sendo sombreado por uma obstrução impenetrável.
- Dispersão ocorre quando uma onda de rádio colide com uma superfície áspera ou com qualquer superfície cujas dimensões são da ordem de λ ou menores, causando um espalhamento da energia refletida em todas as direções. Em ambientes urbanos, as obstruções típicas que causam a dispersão são os postes de iluminação, as placas de rua e as folhagens das árvores.

A análise da degradação provocada por desvanecimento de larga escala ou pequena escala é feita no domínio do tempo e da frequência e permite classificar o desvanecimento como seletivo ou não seletivo em frequência, com base no espalhamento do sinal no tempo, e como sendo lento ou rápido, com base na natureza variante no tempo do canal.

2.3.1 Perfil da Intensidade do Desvanecimento de Multipercurso

O conhecimento do perfil de intensidade do desvanecimento de multipercurso, $S(\tau)$, serve para que se possa analisar a variação da potência média do sinal recebido em função do atraso de tempo do canal. O termo “atraso de tempo” representa o atraso de propagação que excede o atraso do primeiro sinal que chega no receptor. Para um canal de rádio móvel, o sinal recebido usualmente consiste de vários componentes de multipercurso que podem ser modelados de forma discreta. A medição do perfil de

intensidade do desvanecimento de multipercurso pode ser feita usando-se sinais em banda larga. Para um único impulso transmitido, o tempo T_m , entre o primeiro e o último componente recebido, representa o *atraso máximo em excesso*, durante o qual a potência do sinal transmitido cai para algum nível limite abaixo do nível do componente mais forte. O nível limite poderia ser escolhido em 10 ou 20 dB abaixo do nível de potência do componente mais forte [4]. A Figura 2.3 mostra $S(\tau)$ e o máximo atraso em excesso T_m .

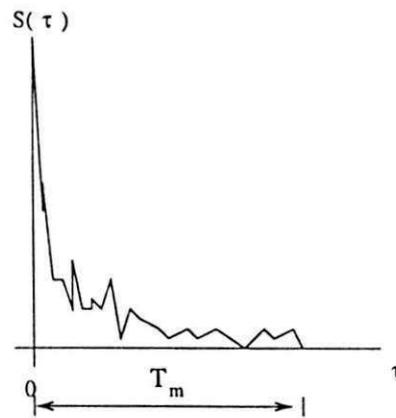


Figura 2.3: Perfil da Intensidade de Multipercurso

2.3.2 Categorias de Degradação Devido ao Espalhamento no Tempo

Em um canal com desvanecimento, a relação entre o atraso máximo em excesso, T_m , e o tempo de símbolo, T_s , pode ser vista em termos de duas categorias de degradação diferentes, que são o desvanecimento seletivo e não seletivo em frequência. Um canal exibe desvanecimento seletivo em frequência se $T_m > T_s$. Essa condição ocorre quando os componentes de multipercurso recebidos se estendem além do tempo de duração do símbolo [4]. Esse tipo de degradação leva ao mesmo tipo de interferência intersimbólica causada por um filtro analógico e é também conhecido como interferência intersimbólica induzida pelo canal.

Um canal exibe um desvanecimento não seletivo em frequência ou plano se $T_m < T_s$. Nesse caso, todos os componentes de multipercurso de um símbolo chegam dentro do tempo de duração do símbolo. Nesse caso não há uma significativa sobreposição entre os símbolos vizinhos transmitidos.

Existe também a degradação no desempenho do sistema, causada pela adição destrutiva ou construtiva dos componentes transmitidos, provocando desse modo uma variação na relação sinal ruído.

2.3.3 Espalhamento do Sinal no Tempo Visto no Domínio da Frequência

Uma caracterização análoga à dispersão no tempo pode ser feita no domínio da frequência [4]. Na Figura 2.4 é vista a função $|R(\Delta f)|$, como uma função de correlação espaçada em frequência, a transformada de Fourier de $S(\tau)$. A função $R(\Delta f)$ representa a correlação entre as respostas do canal a dois sinais, em função da diferença de frequência entre esses dois sinais. O conhecimento de $R(\Delta f)$ serve para que se possa saber qual é a correlação entre sinais recebidos que são espaçados em frequência por um valor $\Delta f = f_1 - f_2$. Os valores assumidos por $R(\Delta f)$ podem ser obtidos transmitindo-se um par de senóides separadas em frequência por um valor Δf , fazendo-se a correlação cruzada entre os dois sinais recebidos e repetindo o processo muitas vezes, aumentando-se igualmente a separação em frequência Δf , entre as senóides. Dessa forma, a medição de $R(\Delta f)$ pode ser feita usando-se uma senóide que é varrida em frequência por toda a faixa de interesse.

A *banda passante coerente* f_0 é uma medida estatística do intervalo de frequências sobre o qual o canal deixa passar todos os componentes espectrais do sinal de entrada, com ganho aproximadamente igual e com fase linear.

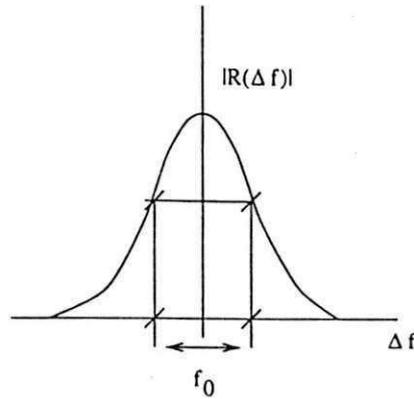


Figura 2.4: Função de correlação entre dois sinais saídos de um canal dispersivo.

2.4 Obtenção do Espectro de Potência do Sinal Recebido

Embora a dedução a seguir seja conhecida ela é colocada aqui porque as equações a serem obtidas serão necessárias na avaliação do efeito Doppler. Seja θ o ângulo de incidência de uma onda de rádio recebida por uma estação móvel e $\omega(\theta)$ a frequência angular instantânea dessa onda. Se a estação móvel se move a uma velocidade constante v , então haverá um deslocamento Doppler igual a $\beta v \cos(\theta)$, em que β depende da permeabilidade magnética e da permissividade elétrica do meio, $\beta = \omega \sqrt{\mu \epsilon}$. Dessa forma, a frequência angular passa a ser dada por

$$\omega(\theta) = \omega_c + \omega_m \cos(\theta), \quad (2.4)$$

em que ω_c é a frequência da portadora e $\omega_m = \beta v$ é o máximo deslocamento Doppler. Como o $\cos(\theta)$ varia entre -1 e 1 na Equação 2.4, tem-se que a banda passante da forma de onda recebida é igual a duas vezes o máximo deslocamento Doppler.

Suponha que o sinal transmitido na saída da antena transmissora, na estação rádio base, seja dado por

$$s(t) = A \cos(\omega_c t + \theta(t) + \phi), \quad (2.5)$$

em que A é a amplitude do sinal transmitido, ω_c é a frequência da portadora, $\theta(t)$ é a fase instantânea e ϕ é o deslocamento de fase.

A autocorrelação de $s(t)$ é dada por

$$\begin{aligned} R_s(\tau) &= E[A^2 \cos(\omega_c t + \theta(t) + \phi) \cos(\omega_c(t + \tau) + \theta(t + \tau) + \phi)] \\ &= \frac{A^2}{2} E[\cos(2\omega_c t + \omega_c \tau + \theta(t) + \theta(t + \tau) + 2\phi) \\ &\quad + \cos(\omega_c \tau + \theta(t + \tau) - \theta(t))]. \end{aligned} \quad (2.6)$$

Considerando que $\theta(t)$ e ϕ sejam variáveis aleatórias independentes, que ϕ seja uniforme no intervalo $[0, 2\pi)$ e que o processo seja estacionário, tem-se que $R_s(\tau)$ pode ser reescrito na forma

$$R_s(\tau) = \frac{A^2}{2} E[\cos(\omega_c \tau + \theta(t + \tau) - \theta(t))]. \quad (2.7)$$

Em termos de exponenciais, $R_s(\tau)$ pode ser reescrita na forma

$$R_s(\tau) = \frac{A^2}{4} E[e^{j\omega_c \tau}] E[e^{j(\theta(t) - \theta(t + \tau))}] + \frac{A^2}{4} E[e^{-j\omega_c \tau}] E[e^{-j(\theta(t) - \theta(t + \tau))}]. \quad (2.8)$$

Considerando que a variação de $\theta(t)$ seja muito lenta e pela minimização do erro médio quadrático, na estimação de uma série temporal, pode-se chegar ao seguinte resultado [5]

$$\begin{aligned} \theta(t + \tau) &\approx \frac{R(\tau)}{R(0)}\theta(t) + \frac{R'(\tau)}{R''(0)}\theta'(t) + \dots \\ &\approx \theta(t) + \tau\theta'(t), \end{aligned} \quad (2.9)$$

em que $\omega(t) = \frac{d\theta(t)}{dt}$.

Desse modo, tem-se que

$$\theta(t + \tau) - \theta(t) \approx \tau\omega(t). \quad (2.10)$$

A autocorrelação $R_s(\tau)$ pode ser reescrita como

$$R_s(\tau) = \frac{A^2}{4} E [e^{j\omega_c \tau}] E [e^{j\omega \tau}] + \frac{A^2}{4} E [e^{-j\omega_c \tau}] E [e^{-j\omega \tau}]. \quad (2.11)$$

Usando o fato de que

$$\begin{aligned} E [e^{j\omega \tau}] &= P_r(\tau) \\ E [e^{-j\omega \tau}] &= P_r(-\tau), \end{aligned} \quad (2.12)$$

que são respectivamente as funções características de $p_\Omega(\omega)$ e de $p_\Omega(-\omega)$, tem-se que

$$R_s(\tau) = \frac{A^2}{4} e^{j\omega_c \tau} P_r(\tau) + \frac{A^2}{4} e^{-j\omega_c \tau} P_r(-\tau). \quad (2.13)$$

Considerando, para tornar a análise mais didática, $P_r(\tau) = P_r(-\tau)$, tem-se que

$$\begin{aligned} R_s(\tau) &= \frac{A^2}{4} P_r(\tau) (e^{j\omega_c \tau} + e^{-j\omega_c \tau}) \\ &= \frac{A^2}{2} P_r(\tau) \cos(\omega_c \tau). \end{aligned} \quad (2.14)$$

Usando as propriedades da transformada de Fourier, tem-se que a densidade espectral de potência é dada por

$$\begin{aligned} S_s(\omega) &= \frac{A^2}{4\pi} p_\Omega(\omega) * \pi[\delta(\omega \pm \omega_c)] \\ &= \frac{A^2}{4} [p_\Omega(\omega + \omega_c) + p_\Omega(\omega - \omega_c)]. \end{aligned} \quad (2.15)$$

A fdp do efeito Doppler, por sua vez, pode ser obtida da seguinte maneira: Sabe-se que, devido ao deslocamento, a frequência $\omega(\theta)$ é dada por

$$\omega(\theta) = \omega_c + \omega_m \cos(\theta), \quad (2.16)$$

em que ϕ é uniforme em $[0, 2\pi]$. Desse modo a fdp $p_\Omega(\omega)$ é dada por

$$p_{\Omega}(\omega) = \frac{p_{\Theta}(\theta)}{|d\omega/d\theta|}. \quad (2.17)$$

Sabendo que

$$\frac{d\omega}{d\theta} = -\omega_m \sin(\theta), \quad (2.18)$$

tem-se que no intervalo $[0, \pi]$ $d\omega/d\theta$ é dada por

$$\frac{d\omega}{d\theta} = -\omega_m \sqrt{1 - \left(\frac{\omega - \omega_c}{\omega_m}\right)^2}; \quad (2.19)$$

de modo que, no intervalo $[0, 2\pi]$, a fdp $p_{\Omega}(\omega)$ é dada por

$$p_{\Omega}(\omega) = \frac{1}{\pi \sqrt{\omega_m^2 - (\omega - \omega_c)^2}}. \quad (2.20)$$

A densidade espectral de potência do efeito Doppler pode ser obtida substituindo a fdp do deslocamento Doppler na Equação 2.15, obtendo desse modo

$$S_s(\omega) = \frac{A^2}{4\pi\omega_m} \left[\frac{1}{\sqrt{1 - \left(\frac{\omega + \omega_c}{\omega_m}\right)^2}} + \frac{1}{\sqrt{1 - \left(\frac{\omega - \omega_c}{\omega_m}\right)^2}} \right]. \quad (2.21)$$

2.5 Modelo do Canal Usado para a Simulação

O modelo de canal usado neste trabalho se baseia no modelo WSS-US (*Wide Sense Stationary-Unrelated Scattering - WSS-US*) para canais variantes no tempo [3]. A simulação do modelo WSS-US na sua forma mais usual (desvanecimento em pequena escala), compreende a geração de pelo menos um processo estocástico gaussiano estacionário, de média nula e com uma forma espectral particular. Para simular o desvanecimento seletivo em frequência é necessária a geração de mais de um processo estocástico descorrelacionado [6].

O canal estudado é visto como um sistema linear variante no tempo com resposta impulsional definida por um conjunto de processos estocásticos no tempo, parametrizados por uma variável τ que representa o atraso de cada componente do sinal transmitido devido aos obstáculos do percurso da fonte ao transmissor. Cada processo estocástico é dado por uma função $h(t, \tau)$, que pode ser entendida como a resposta no instante t a um impulso aplicado ao canal no instante $t - \tau$. Admite-se que para qualquer τ fixo, o processo correspondente $h(t, \tau)$ é estacionário e que para valores τ_1 e τ_2 distintos, os processos $h(t, \tau_1)$ e $h(t, \tau_2)$ são descorrelacionados.

No método de simulação usado neste trabalho (Método de Monte-Carlo), a geração da resposta impulsional do canal é dada pela superposição linear de N percursos elementares, individualmente caracterizados por três variáveis aleatórias (v.a.) que modelam o atraso τ_k , o deslocamento Doppler ν_k e a amplitude complexa a_k . A resposta impulsiva é, desse modo, dada por

$$h(t, \tau) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} a_k \delta(\tau - \tau_k) e^{j2\pi\nu_k t}, \quad (2.22)$$

em que as v.a. τ_k e ν_k possuem função densidade de probabilidade conjunta dada pela função de espalhamento do canal que é dada por [3]

$$S(\tau; \lambda) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Phi_C(\Delta f; \Delta t) e^{-j2\pi\lambda\Delta t} e^{-j2\pi\tau\Delta f} d\Delta t d\Delta f. \quad (2.23)$$

A função de autocorrelação $\Phi_C(\Delta f; \Delta t)$ é dada por

$$\Phi_C(f_1, f_2; \Delta t) = \frac{1}{2} E[H^*(f_1; t) H(f_2; t + \Delta t)], \quad (2.24)$$

em que $H(f; t)$ é dado por

$$H(f; t) = \int_{-\infty}^{\infty} h(t, \tau) e^{-j2\pi f\tau} d\tau, \quad (2.25)$$

que como já foi dito antes, $h(t, \tau)$ é modelado como um processo gaussiano complexo de média nula na variável t .

2.6 Conclusão

A elaboração de um modelo que contemple a maior parte dos efeitos de um canal de comunicações móveis é uma tarefa complicada e dispendiosa. Vários fatores e parâmetros devem ser levados em consideração, dependendo dos efeitos a serem modelados. A escolha de uma determinada técnica deve ser feita de acordo com a precisão que se deseja obter do modelo do canal e de acordo com a aplicação a que esse modelo se destina. Nesse contexto, a resposta impulsional do modelo de canal utilizado foi dada pela superposição linear de N percursos elementares, caracterizados por três variáveis aleatórias que modelam respectivamente o atraso de percurso τ , o deslocamento Doppler e o desvanecimento.

Capítulo 3

Algoritmo dos Passos Descendentes e o Algoritmo LMS

3.1 Introdução

O estudo das estruturas de equalização começou praticamente na década de 60 quando Widrow e Hopf apresentaram uma primeira versão de um algoritmo para filtragem adaptativa baseado no critério do erro médio quadrático. Juntamente com essa teoria de filtros adaptativos começaram a se consolidar também as teorias dos algoritmos de aprendizagem que mais tarde começaram a ser aplicados no processamento adaptativo de sinais, associados a alguma outra forma de diversidade, que faz a recuperação dos sinais recebidos.

Um dos mais populares algoritmos de aprendizagem, desenvolvido por Widrow e Hopf, foi baseado na minimização do erro médio quadrático e ficou conhecido como algoritmo LMS. Devido a algumas de suas propriedades e a sua facilidade de implementação ele passou a ser largamente usado na equalização de canais e no processo de aprendizagem de redes neurais artificiais. Esse algoritmo pode ser visto como sendo a estrutura de uma filtro FIR com seus pesos sendo ajustados recursivamente.

3.2 O Problema da Filtragem Ótima

Considere uma estrutura de um filtro FIR com coeficientes de valores complexos dados por c_0, c_1, c_2, \dots , e seqüência de amostras de entrada dada por $x(0), x(1), x(2), \dots$, também de valores complexos. A saída da estrutura FIR com esses coeficientes e essa seqüência de entrada, em um instante discreto de tempo n , é dada pela soma de convolução dada por

$$y(n) = \sum_{k=0}^{\infty} c_k^* x(n-k), \quad n = 0, 1, 2, \dots, \quad (3.1)$$

em que o asterisco denota uma conjugação complexa. A Equação 3.1 pode ser vista também como o produto interno entre um vetor formado pelos coeficientes do filtro FIR e outro vetor formado pela seqüência de amostras de entrada, dados os atrasos do filtro FIR. Um dos propósitos do filtro FIR adaptativo é produzir em sua saída uma estimativa da resposta desejada $d(n)$. Neste trabalho o filtro FIR juntamente com o algoritmo LMS serão usados para estimar os coeficiente do modelo do canal. Assumindo que a entrada e a saída do filtro são realizações de um processo estocástico conjuntamente estacionário em sentido amplo, ambos com média zero [7], tem-se que o erro de estimação do filtro em relação à resposta desejada é dado por

$$\epsilon(n) = d(n) - y(n). \quad (3.2)$$

Para otimizar o projeto do filtro, escolhe-se então minimizar o valor médio quadrático do erro de estimação. Para isso pode-se definir uma função de custo que conduza a um procedimento linear

$$J = E[\epsilon(n)\epsilon^*(n)] \quad (3.3)$$

e determinar os valores ótimos de c_0, c_1, \dots , para os quais a função de custo J atinge seu valor mínimo. Esses valores podem ser obtidos calculando-se as derivadas parciais de J com relação aos valores c_k e igualando o resultado a zero. Admitindo que os

coeficientes do filtro FIR são dados por

$$c_k = a_k + jb_k \quad (3.4)$$

e aplicando o operador gradiente ∇ ,

$$\nabla_k = \frac{\partial}{\partial a_k} + j \frac{\partial}{\partial b_k} \quad k = 0, 1, 2, \dots, \quad (3.5)$$

à função de custo, tem-se que

$$\nabla_k J = E [\nabla_k (\epsilon(n) \epsilon^*(n))] \quad (3.6)$$

O desenvolvimento e a simplificação de 3.6 resultam em

$$\nabla_k J = -2E[x(n-k)\epsilon^*(n)]. \quad (3.7)$$

O conjunto de pesos ótimos que faz com que o erro de estimação atinja um valor ϵ_{ot} mínimo é obtido igualando-se a Equação 3.7 a zero, ou seja

$$E[x(n-k)\epsilon_{ot}^*(n)] = 0, \quad (3.8)$$

Este resultado obtido na Equação 3.8 constitui o *princípio da ortogonalidade* [7] e estabelece que a condição necessária para que a função de custo J atinja seu valor mínimo é que o valor correspondente do erro de estimação $\epsilon_{ot}(n)$ seja ortogonal a cada amostra que entra no filtro FIR para a estimação da saída desejada no instante n .

A obtenção das equações de Wiener-Hopf segue diretamente do resultado obtido na Equação 3.8 substituindo-se $\epsilon_{ot}(n)$ pela expressão obtida em 3.2, de modo que

$$E \left[x(n-k) \left(d^*(n) - \sum_{j=0}^{\infty} c_{ot,j} x^*(n-j) \right) \right] = 0, \quad k = 0, 1, 2, \dots, \quad (3.9)$$

Reagrupando os termos da Equação 3.9, pode-se obter ainda

$$\sum_{j=0}^{\infty} c_{\text{Ot}j} E[x(n-k)x^*(n-j)] = E[x(n-k)d^*(n)], \quad k = 0, 1, 2, \dots \quad (3.10)$$

Nesta equação:

- O valor esperado no lado esquerdo, $E[x(n-k)x^*(n-j)]$, é igual à função de autocorrelação da entrada do filtro em $j-k$. Esta autocorrelação será denotada por

$$R_{xx}(j-k) = E[x(n-k)x^*(n-j)]. \quad (3.11)$$

- O valor esperado no lado direito, $E[x(n-k)d^*(n)]$ é igual a correlação cruzada entre a entrada do filtro e a saída desejada. Esta correlação cruzada será representada por

$$R_{xd}(-k) = E[x(n-k)d^*(n)]. \quad (3.12)$$

Substituindo essa nova notação, obtém-se as equações de Wiener-Hopf, que podem ser escritas na forma de um conjunto de equações simultâneas, como a seguir

$$\sum_{j=0}^{M-1} c_{\text{Ot}j} R_{xx}(j-k) = R_{xd}(-k), \quad k = 0, 1, \dots, M-1, \quad (3.13)$$

em que $c_{\text{Ot}0}, c_{\text{Ot}1}, \dots, c_{\text{Ot}M-1}$ são os valores ótimos dos pesos do filtro, ou ainda na forma matricial

$$\mathbf{R}_{xx} \mathbf{c}_{\text{Ot}} = \mathbf{R}_{xd}, \quad (3.14)$$

em que

$$\begin{aligned} \mathbf{R}_{xd} &= E[\mathbf{x}(n)d^*(n)] \\ \mathbf{R}_{xd} &= [R_{xd}(0), R_{xd}(-1), \dots, R_{xd}(1-M)]^T \end{aligned} \quad (3.15)$$

e

$$\mathbf{R}_{xx} = E[\mathbf{x}(n)\mathbf{x}^H(n)], \quad (3.16)$$

em que H denota uma transposição hermitiana, que constitui uma operação conjunta de transposição e conjugação complexa. O conjunto de coeficientes ótimos do filtro é então obtido pré-multiplicando a Equação 3.14 pela matriz de autocorrelação inversa, \mathbf{R}^{-1} , obtendo-se desse modo

$$\mathbf{c}_{ot} = \mathbf{R}_{xx}^{-1}\mathbf{R}_{xd}. \quad (3.17)$$

O cálculo do vetor de pesos ótimos requer portanto o conhecimento da matriz de correlação do vetor de entrada $\mathbf{x}(n)$ e do vetor de correlação-cruzada \mathbf{R}_{xd} , entre o vetor $\mathbf{x}(n)$ e a resposta desejada $d(n)$.

3.3 Método dos Passos Descendentes

Da seção anterior pode ser visto que a condição necessária para que um filtro transversal adaptativo encontre sua solução ótima é que a matriz de autocorrelação \mathbf{R}_{xx} seja inversível. Uma maneira de se obter esse conjunto de valores ótimos é resolver, por meio de algum método analítico, a Equação 3.14 para determinar \mathbf{c}_{ot} . Entretanto, esse procedimento poderia resultar em sérias dificuldades computacionais, especialmente quando a estrutura do filtro transversal adaptativo contivesse um número grande de coeficientes. Um procedimento alternativo para esse problema consiste no uso do método dos passos descendentes (*method of steepest descent*) que é um dos mais conhecidos métodos de otimização. A essência do método dos passos descendentes consiste no ajuste dos coeficientes do filtro transversal na direção oposta à direção do gradiente do erro médio quadrático $J(n)$, também chamado de função de custo. Esse ajuste pode ser sumarizado da seguinte maneira:

1. Inicializa-se os pesos da estrutura transversal linear geralmente igual a um vetor nulo, a menos que se tenha uma informação a priori dos valores de \mathbf{c}_{ot} .

2. Usando-se essa suposição inicial dos valores de \mathbf{c}_{ot} calcula-se o vetor gradiente da função de custo. O gradiente $\nabla J(n)$ é obtido aplicando-se o operador

$$\nabla_{\mathbf{c}} = \frac{\partial}{\partial a_k(n)} + j \frac{\partial}{\partial b_k(n)}, \quad k = 1, 2, \dots, M - 1, \quad (3.18)$$

à função de custo $J(n)$.

3. Em seguida faz-se o ajuste dos pesos da próxima iteração fazendo-se uma variação do valor inicial \mathbf{c}_{ot} na direção oposta a do vetor gradiente $\nabla J(n)$.
4. Volta-se então para o passo 2 e repete-se o procedimento até que o erro médio quadrático esteja dentro dos valores requeridos.

É intuitivamente razoável esperar que as correções no vetor de pesos da estrutura transversal linear, na direção oposta a do vetor gradiente, levem a um erro médio quadrado mínimo, que ocorre justamente no ponto em que o vetor $\mathbf{c}(n)$ assume o valor ótimo $\mathbf{c}_{ot}(n)$. De acordo com o método dos passos descendentes os valores do vetor $\mathbf{c}(n)$ podem ser reajustados de acordo com a equação recursiva

$$\mathbf{c}(n + 1) = \mathbf{c}(n) + \mu[-\nabla J(n)], \quad (3.19)$$

em que μ é uma constante positiva de valor real que controla o tamanho do incremento aplicado aos valores \mathbf{c} .

O erro de estimação na saída da estrutura adaptativa linear é dado por

$$\epsilon = d(n) - \mathbf{c}^H(n)\mathbf{x}(n). \quad (3.20)$$

Usando-se a função de custo dada em 3.3, tem-se que

$$\begin{aligned} J(n) &= E[\epsilon(n)\epsilon^*(n)] \\ &= E[d(n)d^*(n)] - E[d(n)\mathbf{x}^H(n)]\mathbf{c}(n) - E[d^*(n)\mathbf{x}(n)] + \\ &+ \mathbf{c}^H(n)E[\mathbf{x}(n)\mathbf{x}^H(n)]\mathbf{c}(n) \\ &= \sigma_d^2 - \mathbf{R}_{xd}^H(n)\mathbf{c}(n) - \mathbf{c}^H(n)\mathbf{R}_{xd}(n) + \mathbf{c}^H(n)\mathbf{R}_{xx}\mathbf{c}(n), \end{aligned} \quad (3.21)$$

em que σ_d^2 é a variância da saída desejada $d(n)$, \mathbf{R}_{xd} é a correlação entre o vetor de amostras de entrada $\mathbf{x}(n)$ e a resposta desejada $d(n)$ e \mathbf{R}_{xx} é a matriz de correlação do vetor de amostras de entrada $\mathbf{x}(n)$. Pode-se obter então o gradiente ∇_c aplicando-se o operador dado na Equação 3.18 ao resultado obtido na Equação 3.21, tendo-se então

$$\nabla_c J(n) = -2\mathbf{R}_{xd} + 2\mathbf{R}_{xx}\mathbf{c}(n). \quad (3.22)$$

A Equação 3.19 pode então ser reescrita na forma

$$\mathbf{c}(n+1) = \mathbf{c}(n) + 2\mu[\mathbf{R}_{xd} - \mathbf{R}_{xx}\mathbf{c}(n)], \quad n = 0, 1, 2, \dots \quad (3.23)$$

Embora a estabilidade desse algoritmo seja muito importante, a análise desse tópico foge dos propósitos deste capítulo. Uma análise detalhada desse assunto pode ser encontrada na referência [7].

3.4 Algoritmo LMS

Na seção anterior foi analisada um pouco da sólida teoria existente por trás do método dos passos descendentes. Como foi visto, esse método requer o conhecimento da matriz de correlação dos dados de entrada e depende do parâmetro de ajuste μ . Nesta seção será analisado um outro algoritmo muito útil no processamento digital de sinais que é o algoritmo LMS. Uma das características básicas desse algoritmo é a sua simplicidade. Uma outra característica interessante é que ele não requer o conhecimento da matriz de correlação dos dados de entrada.

O algoritmo LMS faz parte de uma família de algoritmos conhecidos como algoritmos de gradiente estocástico e são chamados assim para diferenciá-los do algoritmo dos passos descendentes. A estrutura do algoritmo LMS consiste de um filtro FIR com os seus coeficientes ajustados na direção oposta à variação do vetor gradiente do erro médio quadrático. Esse procedimento realizado dessa forma, faz com que os coeficientes da estrutura transversal linear do filtro FIR converjam para um conjunto de valores ótimos que, como já foi citado anteriormente, é conhecido como solução ótima

de Wiener. Esta solução ótima visa a minimização do erro médio quadrático entre a saída estimada pelo algoritmo e a resposta desejada. Por isso o nome do algoritmo é *least mean square*.

Se fosse possível tomar medidas exatas do vetor gradiente $\nabla J(n)$ à cada iteração n , usando-se um parâmetro de ajuste μ apropriadamente escolhido, então o vetor de coeficientes do filtro FIR iria realmente convergir para a solução ótima de Wiener. Na prática, entretanto, não é possível fazer medidas exatas da variação do vetor gradiente, já que isso necessitaria do conhecimento da matriz de correlação \mathbf{R}_{xx} e do vetor de correlação \mathbf{R}_{xd} . Como o LMS não precisa do conhecimento dessas duas entidades, o que se faz na prática é uma estimativa do vetor gradiente. A escolha mais apropriada para os estimadores de \mathbf{R}_{xx} e \mathbf{R}_{xd} é usar estimadores instantâneos que são baseados nas amostras do vetor de entrada e da resposta desejada. Esses estimadores são dados respectivamente por

$$\hat{\mathbf{R}}_{xx}(n) = \mathbf{x}(n)\mathbf{x}^H(n) \quad (3.24)$$

e

$$\hat{\mathbf{R}}_{xd}(n) = \mathbf{x}(n)d^*(n). \quad (3.25)$$

Substituindo estes dois estimadores na Equação 3.22 tem-se que a estimativa do vetor gradiente, $\hat{\nabla}J(n)$, é dada por

$$\hat{\nabla}J(n) = -2\mathbf{x}(n)d^*(n) + 2\mathbf{x}(n)\mathbf{x}^H(n)\hat{\mathbf{c}}(n). \quad (3.26)$$

Note que a estimação do gradiente $\hat{\nabla}J(n)$ também pode ser vista como a aplicação do operador ∇ ao erro quadrático $|\epsilon(n)|^2$. Substituindo o resultado obtido na Equação 3.26 na Equação 3.19, tem-se uma nova relação recursiva dada por

$$\hat{\mathbf{c}}(n+1) = \hat{\mathbf{c}}(n) + \mu\mathbf{x}[d^*(n) - \mathbf{x}^H(n)\hat{\mathbf{c}}(n)]. \quad (3.27)$$

Equivalentemente pode-se ainda escrever o algoritmo LMS em termos de suas três equações básicas que são:

- Saída do filtro FIR

$$y(n) = \hat{\mathbf{c}}^H(n)\mathbf{x}(n) \quad (3.28)$$

- Erro de estimação

$$\epsilon(n) = d(n) - y(n) \quad (3.29)$$

- Equação de ajuste dos coeficientes

$$\hat{\mathbf{c}}(n+1) = \hat{\mathbf{c}}(n) + \mu\mathbf{x}(n)e^*(n) \quad (3.30)$$

As equações do algoritmo LMS podem ainda ser escritas em termos dos componentes reais e imaginários

Usando-se essa notação, a saída da estrutura transversal linear passa a ser

$$\begin{aligned} y.\text{real}(n) &= \sum_{i=1}^m (\hat{c}_i.\text{real} x_i.\text{real} + \hat{c}_i.\text{imag} x_i.\text{imag}) \\ y.\text{imag}(n) &= \sum_{i=1}^m (\hat{c}_i.\text{real} x_i.\text{imag} - \hat{c}_i.\text{imag} x_i.\text{real}), \end{aligned} \quad (3.31)$$

de modo que o erro de estimação pode ser reescrito na forma

$$\begin{aligned} \epsilon.\text{real}(n) &= d.\text{real}(n) - y.\text{real}(n) \\ \epsilon.\text{imag}(n) &= d.\text{imag}(n) - y.\text{imag}(n) \end{aligned} \quad (3.32)$$

e a equação de reajuste dos pesos passa a ser

$$\begin{aligned} \hat{c}.\text{real}(n+1) &= \hat{c}.\text{real}(n) + \mu(\mathbf{x}.\text{real}(n)e.\text{real}(n) + \mathbf{x}.\text{imag}(n)e.\text{imag}(n)) \\ \hat{c}.\text{imag}(n+1) &= \hat{c}.\text{imag}(n) + \mu(\mathbf{x}.\text{imag}(n)e.\text{real}(n) - \mathbf{x}.\text{real}(n)e.\text{imag}(n)) \end{aligned} \quad (3.33)$$

3.5 Conclusão

O algoritmo LMS analisado nesta seção decorre diretamente do algoritmo dos passos descendentes, que requer o conhecimento da matriz de autocorrelação das amostras de entrada. O algoritmo LMS não necessita do conhecimento da matriz de autocorrelação, mas é muito dependente do parâmetro de ajuste de convergência μ . O LMS pode ser usado diretamente como uma estrutura de equalização como pode ser usado na forma de um módulo externo de treinamento de uma rede neural.

Capítulo 4

Introdução às Redes Neurais Artificiais

4.1 Introdução

O funcionamento da mente humana tem sido um grande mistério para o homem. A mente humana é tão poderosa que pode desempenhar uma variedade de atividades relacionadas ao pensar, falar, sentir e aprender. Esse dispositivo tão poderoso tem inspirado muitos cientistas a tentar modelar algumas de suas funções. Na mente humana o neurônio é a unidade celular principal. Cada neurônio recebe e combina sinais de muitos outros neurônios e produz sinais elétricos para o axônio conduzir certas ações. A neurocomputação surgiu desse princípio.

As pesquisas em redes neurais têm sido baseadas na forma como a mente humana processa suas tarefas, de acordo com os conhecimentos disponíveis acerca do funcionamento do cérebro. O trabalho pioneiro de Ramón y Cajál em 1911 introduziu a idéia de *neurônios* como constituintes estruturais da mente, isso colaborou imensamente para o entendimento da mente [7]. Sabe-se que a mente é um sistema de processamento de informações altamente complexo, não-linear que faz cálculos em paralelo e cuja estrutura e capacidade de organizar seus neurônios torna possível ao ser humano realizar tarefas como reconhecimento de padrões, percepção e controle motor, muitas vezes mais rapidamente que o mais rápido computador digital existente hoje.

Os neurônios são conectados por meio de estruturas elementares e unidades funcionais que servem de intermédio para as suas iterações. Essas estruturas, chamadas de *sinapses*, atuam basicamente por reações químicas que operam da seguinte forma: uma substância transmissora que se propaga através das junções sinápticas é liberada durante a comunicação entre os neurônios. Essa comunicação pode ser entendida como uma conversão de um sinal elétrico pré-sináptico em um sinal eletroquímico e depois retorna a um sinal elétrico pós-sináptico. Dispositivos com essa característica, na terminologia elétrica, são chamados dispositivos de duas portas. Na descrição tradicional da organização neuronal, é assumido que a sinapse é uma simples conexão que pode impor uma excitação ou inibição, mas não ambas, ao neurônio receptor.

Na mente humana a plasticidade é seguida por dois mecanismos: um é aquele que se baseia na criação de novas conexões sinápticas entre neurônios e o outro se baseia na modificação das sinapses existentes. Os axônios, que formam a linha de transmissão entre os neurônios e os dendritos, que são as zonas receptoras, constituem dois tipos de filamentos morfológicamente diferentes das células. Um axônio tem uma superfície lisa, poucos ramos e um grande comprimento, enquanto um dendrito tem uma superfície irregular e muitos ramos. A forma e o tamanho dos neurônios pode variar em diferentes partes da mente.

Da mesma forma que a plasticidade parece ser essencial para o funcionamento dos neurônios como unidades de processamento de informações na mente humana, ela é essencial nas redes neuronais artificiais. De acordo com Haykin [7], “em sua forma mais geral, uma rede neural artificial é uma máquina que é projetada para modelar um caminho em que a mente realiza uma determinada tarefa ou função de interesse. Uma RNA geralmente é implementada usando-se componentes eletrônicos ou é simulada por meio de *softwares* em computadores digitais”. Ainda de acordo com Haykin, uma RNA é definida da seguinte forma:

Uma rede neural artificial é um processador distribuído massivamente em paralelo que tem a propensão natural para armazenar conhecimentos experimentais e torná-los disponíveis para o uso.

Uma RNA parece com a mente humana em dois aspectos:

1. O conhecimento é adquirido pela rede no processo de aprendizagem.
2. As conexões entre os neurônios, conhecidas como sinapses, são usadas para armazenar o conhecimento.

O procedimento usado para representar o processo de aprendizagem, comumente chamado de algoritmo de aprendizagem, tem a função de modificar os valores associados às conexões da rede neural buscando alcançar um objetivo inicial projetado.

4.2 Treinamento e Teste das RNA's

Na tarefa de classificação de padrões, realizada por uma RNA, o objetivo é atribuir padrões de entrada para uma classe de saída, dentre um número finito de classes. Os padrões de entrada podem ser dinâmicos ou estáticos. Os padrões de entrada estáticos são os mais comuns e consistem de vetores de entrada \mathbf{x} contendo N elementos, de valores reais ou complexos, denotados por x_1, x_2, \dots, x_N . Segundo Lippmann [8] os padrões de entrada de uma RNA podem ser vistos como sendo pontos de um espaço multidimensional definido pelas medidas das características dos dados de entrada. O propósito de um classificador de padrões é fazer a partição desse espaço multidimensional em regiões de decisão que indiquem a qual das classes de saída a entrada pertence.

A seleção das características dos dados de entrada de uma RNA para um determinado problema é o primeiro passo a ser tomado quando se quer usar uma RNA como um classificador de padrões. A eficiência do classificador está diretamente relacionada à escolha das características que irão representar o problema a ser modelado pela rede. Uma vez que essas características são responsáveis pelas informações que vão permitir a classificação dos padrões de entrada em classes de saída, algumas propriedades são exigidas, como a insensibilidade a variações irrelevantes nos valores de entrada da RNA e o número limitado de valores de entrada, para possibilitar a computação eficiente das funções discriminantes e limitar a quantidade de dados de treinamento.

Uma vez que tenham sido extraídas as características a serem usadas no projeto da rede, seguem as fases de treinamento e testes. Na fase de treinamento uma quantidade limitada de dados de treinamento conhecida a priori, em relação ao domínio

do problema, é usada para ajustar os parâmetros da RNA e aprender a estrutura do classificador. Durante a fase de testes, o classificador projetado na fase de treinamento é avaliado com os dados de teste para que se tenha uma decisão e uma correspondente classificação para cada padrão de entrada. Os parâmetros usados na classificação podem então ser adaptados para tomarem conhecimento de novos dados de treinamento ou para compensar as entradas não-estacionárias, variações em componentes internos da RNA ou falhas internas.

Um caso comum de entradas não-estacionárias ocorre em problemas de recuperação de informação em meios de transmissão com característica variante, como nos canais de rádio móvel. Nesse caso, uma estrutura de recuperação ou um equalizador, implementado com RNA's necessitaria de uma adaptação dos seus parâmetros e pesos a cada vez que a característica do meio de transmissão mudasse. Conseqüentemente seria necessária uma fase de treinamento para que a RNA aprendesse as novas características do meio de transmissão e ajustasse os seus pesos de acordo com essas novas características.

Mais de um procedimento pode ser usado no treinamento de redes neurais para serem usadas em tarefas de classificação. O treinamento de uma RNA pode ser supervisionado, não-supervisionado ou híbrido. Redes neurais treinadas com algoritmos supervisionados requerem dados com informações sobre as classes a que pertencem cada um dos padrões de treinamento. A principal característica do treinamento não-supervisionado ou cego, como também é conhecido, é o desenvolvimento do aprendizado sem que se saiba a qual classe de saída o padrão de entrada pertence. Esse tipo de aprendizagem se aplica principalmente em situações em que a informação a respeito das classes dos padrões de entrada não é disponível durante o treinamento. Em canais de comunicações cujas características são desconhecidas tem-se usado muito estruturas treinadas sem supervisão. Nessas estruturas apenas algumas informações estatísticas a respeito da natureza do canal são passadas para o algoritmo cego. A seqüência de treinamento, que caracteriza o treinamento supervisionado, não é utilizada.

4.3 Definição das Redes Neurais Classificadoras

Para que haja o processo de classificação de padrões por parte da RNA é necessário que haja o mapeamento dos padrões de entrada em um espaço de saída chamado de espaço de classificação. Esse espaço de classificação é definido por regiões de decisão que separam os padrões de entrada, atribuindo a cada padrão uma classe de saída que é estabelecida pelas fronteiras ou regiões de decisão. Nesta seção é apresentada uma análise dos tipos mais usados de classificadores.

4.3.1 Classificadores Probabilísticos

O primeiro grupo de classificadores, os classificadores probabilísticos, se fundamenta numa rigorosa base probabilística, assumindo que a estrutura e a distribuição dos dados de entrada do problema são bem conhecidos, de modo que esses classificadores classificam esses dados de acordo com sua função densidade de probabilidade. Os parâmetros da distribuição utilizada são geralmente estimados usando-se treinamento supervisionado, em que se supõe que todos os dados do treinamento estão disponíveis simultaneamente. A construção dos classificadores se baseia na conversão da probabilidade a priori das classes de padrões de treinamento em medidas de probabilidade condicionadas de um dado padrão pertencer a uma dada classe de treinamento.

4.3.2 Classificadores de Hiperplano

As funções discriminantes usadas nos classificadores de hiperplano são capazes de formar regiões de decisão complexas que usam nós para formar hiperplanos como fronteiras de decisão no espaço de entrada transformado. A transformação do espaço de entrada é feita sobre os nós intermediários formando uma soma ponderada das entradas. Essa soma passa por funções não lineares como a função *sigmóide* ou a *tangente hiperbólica*. Na maioria das vezes esses classificadores têm exigências computacionais e de memória baixas, mas podem precisar de longo tempo de treinamento e/ou algoritmos de treinamento complexos. O perceptron multicamadas treinado com o algoritmo *back-propagation*, os classificadores de árvores binárias, as redes de alta ordem que formam

polinômios de alta ordem a partir das entradas são alguns exemplos de classificadores de hiperplano.

4.3.3 Classificadores de Núcleo - *Kernel*

Assim como nos classificadores de hiperplano, existe nos classificadores de núcleo uma transformação do espaço de entrada para um espaço de saída maior. Os classificadores de núcleo criam regiões de decisão complexas baseadas nos campos receptivos formados pelos nós das funções-núcleo. Os nós do espaço transformado são as saídas proporcionadas pela função núcleo que fornecem um valor maior quando a entrada está mais próxima de um dos centros das funções-núcleo.

Os campos receptivos são centros que fazem com que a saída da j -ésima função-núcleo seja dada por

$$y_j = \varphi_k \left(\frac{\|\mathbf{x} - \mathbf{t}_j\|}{\rho} \right), \quad (4.1)$$

em que φ_k é uma função-núcleo, $\|\cdot\|$ é a norma euclidiana, \mathbf{t}_j é um vetor representando o centro do nó j e ρ é o parâmetro livre que determina a largura da função núcleo. A função mais usada como núcleo é geralmente a função de Gauss, que retorna um valor alto quando a entrada está mais próxima do centro do nó. À medida que a entrada se distancia do centro, os valores retornados pela função de Gauss decrescem monotonicamente. O raio de atuação de cada campo receptivo está intimamente ligado ao parâmetro ρ que determina a região de influência e a suavidade de cada função-núcleo. As decisões de classificação são então baseadas nos nós com maior valor de saída, que são formados pela soma ponderada das funções-núcleo.

O treinamento dos classificadores de núcleo é relativamente rápido. Estes classificadores têm requerimentos computacionais e de memória que são intermediários, quando comparados aos classificadores de hiperplano. Os centros das funções-núcleo podem ser selecionados aleatoriamente dos exemplares de treinamento, utilizando algoritmos de *clustering*, com supervisão ou sem supervisão. Segundo Lippmann [8], classificadores de núcleo incluem classificadores convencionais que estimam funções de distribuição

de probabilidade que usam a abordagem de *Parzen Window* ou distribuições mistas e classificadores que formam funções discriminantes usando funções núcleo [9]. Redes neurais classificadoras de núcleo incluem abordagens baseadas em mapas que usam matrizes de nós que calculam funções-núcleo, classificadores baseados no *Cerebellar Model Articulation Controller* (CMAC) e classificadores que usam o método das *funções potenciais*, às vezes chamadas de funções de bases radiais.

4.3.4 Classificadores de Exemplos

Essa categoria de classificadores se baseia no cálculo da distância entre os padrões de entrada e exemplos treinados previamente. A idéia é atribuir os padrões de entrada aos exemplos mais próximos calculados por alguma métrica. A métrica comumente usada é a euclidiana, que é utilizada para medir a distância entre as entradas e os centróides que representam exemplos previamente selecionados durante o treinamento. O treinamento empregado é geralmente o híbrido ou combinado. As classes dos classificadores associados à maioria dos k nodos exemplos, com as k menores distâncias, formam as fronteiras de decisão do classificador. Essa categoria de classificadores requer grandes quantidades de memória e tempo de computação, porém o aprendizado é rápido.

Existem diferenças na estrutura e na filosofia de cada um dos classificadores que compõem os quatro grupos apresentados, porém, eles proporcionam taxas de erro similares. Para cada tipo de problema existe um classificador mais apropriado, a questão é identificar esse classificador. Na maioria das vezes a opção da escolha está relacionada à quantidade de dados disponíveis para se trabalhar, à origem dos dados, ao tipo de características empregadas e às restrições computacionais.

4.4 Redes com Funções de Base Radiais

Broomhead e Lowe foram os primeiros a explorarem, em 1988, o uso de funções de base radiais no projeto de RNA's [7], com contribuições de Moody e Darken em 1989, Renal em 1990 e Poggio e Girosi em 1990 [10]. Uma rede com funções de base

radiais projetada em sua forma mais básica, consiste apenas de três camadas com papéis totalmente diferentes. A camada de entrada é obtida dos nós fonte, nós de entrada que conectam a rede ao ambiente no qual ela está inserida [11]. A segunda camada, que é a única camada escondida da rede, aplica uma transformação linear do espaço de valores de entrada para o espaço de valores da camada escondida, que na maioria das aplicações tem dimensão bem maior. A camada de saída é linear e fornece a resposta da rede ao modelo de ativação (sinal de entrada) aplicado à entrada da rede. A justificativa matemática para se ter uma transformação não linear seguida por uma transformação linear pode ser obtida de um trabalho de Cover, de 1965. De acordo com esse trabalho, um problema de classificação de modelos, realizado em um espaço de alta dimensão, é mais provável de ser linearmente separável do que em um espaço de baixa dimensão. Por isso a razão de se usar o espaço da camada escondida com dimensão maior que do espaço de entrada. Na próxima seção é apresentado o teorema de Cover da separabilidade de padrões, que apresenta as condições sob as quais dois conjuntos de padrões (dados de entrada) podem ser separados.

4.5 O Teorema de Cover da Separabilidade de Padrões

Quando uma rede NRBF é usada para realizar uma tarefa complexa de classificação de padrões, o problema é resolvido por uma transformação não-linear em um espaço de alta dimensão [7]. A justificativa para essa afirmação é obtida a partir do *teorema de Cover da separabilidade de padrões*, que estabelece que um problema complexo de classificação de padrões, formalizado em um espaço não-linear de alta dimensão, é mais provável de ser linearmente separável do que em um espaço de baixa dimensão

Imagine, por exemplo, um grupo de superfícies em que cada superfície divide um conjunto de valores de entrada, espaço de entrada, em duas regiões. Seja \mathcal{X} um conjunto de N padrões (vetores) $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$, cada um dos quais associado a uma das duas classes \mathcal{X}_1 e \mathcal{X}_2 . Essa partição binária é dita ser separável com respeito à família hipotética de superfícies se existe uma superfície, dentro da sua família, que separe os

pontos da classe \mathcal{X}_1 dos pontos da classe \mathcal{X}_2 . Para cada padrão $\mathbf{x} \in \mathcal{X}$, defina um vetor estabelecido a partir de um conjunto de funções reais $\{\varphi_i(\mathbf{x}) | i = 1, 2, \dots, m_1\}$, de modo que

$$\varphi(\mathbf{x}) = [\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \dots, \varphi_{m_1}(\mathbf{x})]^T. \quad (4.2)$$

Suponha que o padrão \mathbf{x} é um vetor em um espaço de entrada de dimensão m_0 . O vetor $\varphi(\mathbf{x})$ faz então um mapeamento dos pontos do espaço m_0 -dimensional de entrada nos correspondentes pontos do espaço m_1 -dimensional. As funções $\varphi_i(\mathbf{x})$ são chamadas de funções escondidas porque elas desempenham um papel semelhante ao das funções de uma rede neural *feedforward*. Desse modo, o espaço formado pelo conjunto de funções $\{\varphi_i(\mathbf{x})\}_{i=1}^{m_1}$ é referido como espaço escondido.

A partição $\{\mathcal{X}_1, \mathcal{X}_2\}$ de \mathcal{X} é dita ser φ -separável se existe em \mathcal{X} um vetor m_1 -dimensional \mathbf{w} de modo que

$$\begin{aligned} \mathbf{w}^T \varphi(\mathbf{x}) &> 0, & \text{ se } \mathbf{x} \in \mathcal{X}_1 \\ \mathbf{w}^T \varphi(\mathbf{x}) &< 0, & \text{ se } \mathbf{x} \in \mathcal{X}_2. \end{aligned} \quad (4.3)$$

O hiperplano definido pela equação

$$\mathbf{w}^T \varphi(\mathbf{x}) = 0 \quad (4.4)$$

descreve portanto uma superfície de separação no espaço escondido. Analisando-se a separabilidade de padrões do ponto de vista probabilístico, tem-se um evento aleatório que depende da partição binária escolhida e da distribuição dos padrões do espaço de entrada. Suponha que os padrões de ativação $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ são escolhidos independentemente, de acordo com a medida de probabilidade imposta ao espaço de entrada e que todas as possíveis partições binárias de $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^N$ são equiprováveis. A probabilidade $P(N, m_1)$ de que uma particular partição binária do espaço \mathcal{X} , de padrões de entrada, seja φ -separável é dada por

rede em relação à rede com perceptrons multicamadas, treinada com o algoritmo *back-propagation*, está na rapidez do treinamento e na estrutura mais simples da rede. Um diagrama da rede NRBF com três camadas pode ser visto na Figura 4.1.

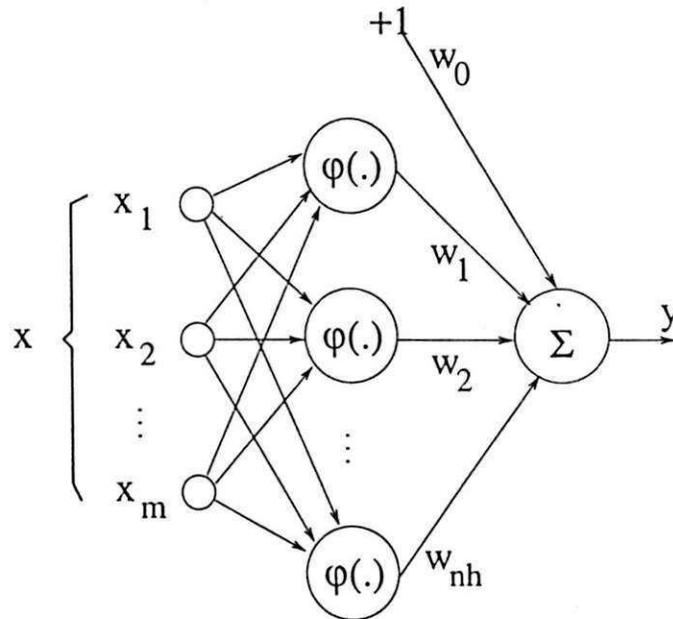


Figura 4.1: Rede de Funções de Base Radiais

Como pode ser visto na Figura 4.1, a rede NRBF consiste de uma camada com m nós de entrada, uma única camada intermediária com n_h unidades de processamento não lineares e uma camada de saída de pesos lineares. Usando as saídas calculadas pela camada intermediária, em resposta ao vetor de entrada, juntamente com uma resposta desejada, apresentada à camada de saída, os pesos da rede são treinados de forma supervisionada.

A rede NRBF difere de uma rede com perceptrons em multicamadas em alguns aspectos estruturais como:

- A rede NRBF tem uma única camada intermediária, enquanto a rede MLP pode ter uma ou mais camadas intermediárias.
- Na rede NRBF, as funções de transferência, conectando a camada de entrada à camada intermediária, são funções não lineares, enquanto que as funções que co-

nectam a camada intermediária à camada de saída são lineares. Na rede MLP, as funções de transferência que conectam cada camada intermediária a sua camada anterior são todas não lineares e as funções de transferência da camada de saída podem ser não lineares ou lineares, dependendo da aplicação de interesse.

- Cada unidade intermediária de uma rede NRBF calcula a distância entre o vetor de entrada e o centro de uma *função de base radial* que caracteriza uma determinada unidade particular da camada intermediária. Por outro lado, cada neurônio de uma rede MLP calcula o produto interno entre o vetor de entrada aplicado ao neurônio e o vetor de pesos associado.

Todas essas definições e essas características serão analisadas considerando-se que a rede processe dados de valores complexos [12], [13]. Isso é feito nesse trabalho devido à utilização da modulação QPSK no sistema de comunicações implementado.

4.7 Rede NRBF de Valores Complexos

Sem perda de generalidade, uma rede NRBF poder ser representada pelo diagrama mostrado na Figura 4.1. De acordo com a terminologia usada nessa figura, o mapeamento entre a entrada e a saída da rede é dado por

$$y(n) = \sum_{k=1}^{n_h} w_k(n) \varphi(\mathbf{x}(n), \mathbf{t}_k(n)) + w_0, \quad (4.7)$$

em que $\varphi(\cdot)$ é a k -ésima função de base radial que recebe o valor da distância ao quadrado entre os vetores de entrada de valores complexos \mathbf{x} e os centros de cada função, \mathbf{t}_k . O sinal de saída produzido pela k -ésima unidade intermediária é uma função não linear da distância entre os vetores \mathbf{x} e \mathbf{t}_k . Os fatores w_k representam os pesos de valores complexos que conectam a k -ésima unidade intermediária ao nó de saída da rede e o termo w_0 representa um termo de polarização.

A função de base radial utilizada é a função Gaussiana, definida pela Equação 4.8, sendo a mais utilizada no problema de equalização por representar a função densidade

de probabilidade do ruído gaussiano comumente empregada nos modelos de sistemas de comunicação [14].

$$\varphi(\mathbf{x}, \mathbf{t}_k) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{t}_k\|^2}{\rho}\right), \quad (4.8)$$

em que ρ é um parâmetro de valor real que controla o raio de alcance da função $\varphi(\cdot)$.

A rede NRBF com valores complexos é apropriada para sistemas de comunicações em que se usa esquemas de modulação como o QPSK. Nesse caso é necessário o uso de uma estrutura complexa para que se possa analisar a sua atuação nas componentes de fase e quadratura do sinal proveniente do modelo de canal utilizado. Como foi visto no Capítulo 2, os coeficientes do modelo do canal dão ao sinal transmitido uma variação em sua amplitude que é conhecida como desvanecimento Rayleigh e à componente de fase do sinal transmitido um deslocamento conhecido como deslocamento Doppler. Para minimizar esses efeitos é analisada no próximo capítulo uma estrutura de equalização que faz uso da rede NRBF para realizar o método de detecção bayesiana.

4.8 Conclusão

Existem várias estruturas de redes neurais que encontram aplicações nas mais variadas áreas da engenharia. Cada estrutura é mais adequada a um determinado problema do que outra e cabe ao projetista que está usando essa ferramenta escolher a estrutura que melhor se ajuste às exigências de seu projeto. Como foi visto neste capítulo, os diferentes modelos de redes neurais possuem diferentes vantagens e desvantagens que devem ser analisadas para que se possa ponderar questões como complexidade, custo, ganhos obtidos e outros fatores. Neste trabalho, por exemplo, foi escolhida a rede NRBF por ser mais parecida com o método de detecção utilizado, por ser mais fácil e rápida de treinar e por apresentar menor complexidade em relação a outras redes.

Capítulo 5

Detecção Bayesiana

5.1 Introdução

A maioria dos problemas de detecção pode ser analisada no campo dos testes das hipóteses, no qual se tem uma observação e se deseja escolher ou classificar essa observação entre M possíveis ocorrências estatísticas. Por exemplo, em um receptor de comunicações M -ário, observa-se uma forma de onda que consiste de uma entre outras M possíveis formas de onda corrompidas por ruído presente no meio de transmissão e deseja-se decidir qual dos M sinais detectados é o que mais se aproxima do sinal que foi transmitido. Sem dúvidas que para qualquer problema de detecção existem várias possíveis estratégias de decisão que poderiam ser aplicadas, entretanto é mais conveniente escolher uma que otimize o critério de desempenho usado no estudo[15].

As estruturas de equalização e detecção de sinais podem ser subdivididas em dois grupos. O primeiro grupo consiste nos equalizadores baseados na estimação de seqüências de máxima probabilidade a posteriori e que são implementados de acordo com o algoritmo de Viterbi. Esses equalizadores são considerados ótimos, mas têm uma complexidade que geralmente os tornam pouco apropriados para aplicações em que o requisito é tempo de processamento baixo e facilidade de implementação. O outro grupo de equalizadores é composto pelas estruturas que tomam decisão a respeito da estimativa que devem fornecer à saída, com base em algum procedimento que minimize o erro de estimação. Uma teoria bem consolidada e que tem sido usada na detecção de sinais é

a teoria da decisão de Bayes. As estruturas baseadas nessa teoria são conhecidas como equalizadores bayesianos e fornecem uma solução ótima para o grupo de equalizadores que fazem estimativa por decisão. Neste capítulo será mostrada a solução bayesiana para o problema da equalização e como essa solução pode ser realizada com a rede NRBF.

Neste trabalho considera-se a formulação de Bayes para o problema da detecção e é analisada uma solução ótima inicialmente para o caso binário, por questões de simplicidade, e em seguida essa teoria é estendida para o problema da equalização. A escolha da teoria Bayesiana se deve ao fato de que ela é ótima no sentido de minimizar a taxa de erro e devido à sua simplicidade em relação, por exemplo, à detecção por estimação de seqüências de máxima probabilidade posteriori.

5.2 Teste de Hipóteses Bayesiano

O problema primário a ser considerado nesta seção é o problema de teste de hipóteses ou de “estados”, em que é assumido, inicialmente para o caso binário, que existem duas hipóteses a serem testadas que são H_0 e H_1 correspondendo respectivamente às probabilidades P_0 e P_1 no espaço de observação Γ . Esse problema pode ser escrito como

$$\begin{aligned} H_0 : X \sim P_0 \\ \text{versus} \\ H_1 : X \sim P_1 \end{aligned}$$

em que a notação $X \sim P$ significa que a variável aleatória observada X tem distribuição de probabilidade P .

A regra de decisão ou teste de hipótese η para H_0 versus H_1 é qualquer partição do conjunto de observação Γ nos conjuntos Γ_1 e $\Gamma_0 = \Gamma_1^c$, de modo que se possa escolher H_j quando $x \in \Gamma_j$ para $j = 0$ ou 1 . No caso binário, o conjunto Γ_1 é chamado de *região de rejeição* e o conjunto Γ_0 é chamado de *região de aceitação*. Pode-se pensar na regra de decisão η como uma função em Γ dada por

$$\eta(x) = \begin{cases} 1 & \text{se } x \in \Gamma_1 \\ 0 & \text{se } x \in \Gamma_1^c \end{cases},$$

de modo que o valor de η para um dado $x \in \Gamma$ é o índice da hipótese aceita por η .

O que se quer nesse caso é escolher Γ_1 de alguma forma ótima e, com isso em mente, pode-se associar custos a uma decisão tomada. Em particular, pode-se utilizar os números positivos C_{ij} , $i = 0, 1$ e $j = 0, 1$ como sendo o custo que se tem ao escolher a hipótese H_i quando H_j é a hipótese verdadeira. Pode-se desse modo definir um *risco condicional* para cada hipótese como sendo o custo médio associado à regra de decisão η quando a hipótese H_j é verdadeira, ou seja

$$R_j(\eta) = C_{1j}P_j(\Gamma_1) + C_{0j}P_j(\Gamma_0), \quad j = 0, 1. \quad (5.1)$$

Note que $R_j(\eta)$ é o custo de se escolher H_1 quando H_j é verdadeira vezes a probabilidade de se fazer essa escolha mais o custo de se escolher H_0 quando H_j é verdadeira vezes a probabilidade de se fazer isso.

Pode-se também associar probabilidades α_0 e $\alpha_1 = (1 - \alpha_0)$ às ocorrências das hipóteses H_0 e H_1 respectivamente, ou seja, α_j é a probabilidade de que a hipótese H_j seja verdadeira independente do valor de X . Essas probabilidades são conhecidas como *probabilidades a priori* das duas hipóteses. Para essas probabilidades a priori, pode-se definir uma média ou *risco médio de Bayes* que é dado por

$$r(\eta) = \alpha_0 R_0(\eta) + \alpha_1 R_1(\eta). \quad (5.2)$$

Pode-se então definir uma regra de decisão ótima para H_0 versus H_1 como sendo uma regra que minimize o risco médio de Bayes.

Combinando as Equações 5.1 e 5.2, tem-se que a função do risco de Bayes passa ser dada por

$$\begin{aligned}
r(\eta) &= \alpha_0(C_{10}P_0(\Gamma_1) + C_{00}P_0(\Gamma_0)) + \alpha_1(C_{11}P_1(\Gamma_1) + C_{01}P_1(\Gamma_0)) \\
&= \alpha_0(C_{10}P_0(\Gamma_1) + C_{00}(1 - P_0(\Gamma_1))) + \alpha_1(C_{11}P_1(\Gamma_1) \\
&\quad + C_{01}(1 - P_1(\Gamma_1)))
\end{aligned} \tag{5.3}$$

ou ainda

$$r(\eta) = \sum_{j=0}^1 \alpha_j C_{0j} + \sum_{j=0}^1 \alpha_j (C_{1j} - C_{0j}) P_j(\Gamma_1), \tag{5.4}$$

em que foi considerado que $P_j(\Gamma_0 = \Gamma_1^c) = 1 - P_j(\Gamma_1)$.

Assumindo que P_j tem função densidade de probabilidade p_j para todo $j = 0$ e 1 e que

$$P_j = \int_A p_j(x) dx, \tag{5.5}$$

em que A é o espaço de observação de X , tem-se que

$$r(\eta) = \sum_{j=0}^1 \alpha_j C_{0j} + \int_{\Gamma_1} \left[\sum_{j=0}^1 \alpha_j (C_{1j} - C_{0j}) p_j(x) \right] dx. \tag{5.6}$$

Sabendo que $\sum_{j=0}^1 \alpha_j C_{0j} > 0$, tem-se que o risco de Bayes é mínimo quando x é escolhido na região

$$\begin{aligned}
\Gamma_1 &= \left\{ x \in \Gamma \mid \sum_{j=0}^1 \alpha_j (C_{1j} - C_{0j}) p_j(x) \leq 0 \right\} \\
&= \left\{ x \in \Gamma \mid \frac{p_1(x)}{p_0(x)} \geq \frac{\alpha_0(C_{10} - C_{00})}{\alpha_1(C_{01} - C_{11})} \right\} \\
&= \{ x \in \Gamma \mid [p_1(x)/p_0(x)] \geq \tau \},
\end{aligned} \tag{5.7}$$

em que $\tau = \frac{\alpha_0(C_{10} - C_{00})}{\alpha_1(C_{01} - C_{11})}$.

A regra de decisão descrita pela região 5.7 é chamada de *teste de taxa de probabilidade* e tem um papel importante na teoria de teste de hipóteses. A quantidade

$$L(x) = \frac{p_1(x)}{p_0(x)}, \quad x \in \Gamma_1 \quad (5.8)$$

é conhecida como *taxa de probabilidade* entre as hipóteses H_0 e H_1 . Desse modo, a regra de decisão correspondente à região Γ_1 calcula a taxa de probabilidade para os valores observados de X e toma sua decisão comparando essa taxa com o limite τ , ou seja

$$\eta(x) = \begin{cases} 1 & \text{se } L(x) \geq \tau \\ 0 & \text{se } L(x) < \tau \end{cases}$$

O custo comumente usado é o custo uniforme, dado por

$$C_{ij} = \begin{cases} 0 & \text{se } i = j \\ 1 & \text{se } i \neq j \end{cases}$$

e o risco de Bayes para uma regra de decisão η com região crítica Γ_1 nesse caso é dado por

$$r(\eta) = \alpha_0 P_0(\Gamma_1) + P_1(\Gamma_0), \quad (5.9)$$

em que $P_i(\Gamma_j)$ é a probabilidade de se escolher H_j quando H_i é verdadeira, ou seja, $P_i(\Gamma_j)$ é a probabilidade condicional de se cometer um erro dado que H_i é verdade. Nesse caso $r(\eta)$ é a probabilidade média de erro associada à regra de decisão η .

A fórmula de Bayes implica que a probabilidade condicional de que a hipótese H_j seja verdadeira dado que a observação aleatória X toma o valor x é dada por

$$P(H_j|X = x) = \frac{p_j(x)\alpha_j}{\alpha_0 p_0(x) + \alpha_1 p_1(x)}, \quad (5.10)$$

em que as probabilidades α_0 e α_1 são chamadas de *probabilidades a posteriori* das duas hipóteses. Usando a Equação 5.10, a região crítica de Bayes passa a ser dada por

$$\begin{aligned} \Gamma_1 &= \{x \in \Gamma | C_{10}\alpha_0(x) + C_{11}\alpha_1(x) \\ &\leq C_{00}\alpha_0(x) + C_{01}\alpha_1(x)\}, \end{aligned} \quad (5.11)$$

em que a quantidade

$$C_{i0}\alpha_0(x) + C_{i1}\alpha_1(x) \quad (5.12)$$

é o custo médio associado à escolha da hipótese H_i dado que $X = x$. Essa quantidade é chamada de *custo a posteriori* e desse modo a regra de Bayes faz a sua decisão escolhendo a hipótese que leva a um mínimo custo a posteriori. A idéia desses conceitos pode ser estendida quase que diretamente para o problema da equalização nas próximas seções.

5.3 Aplicação da Detecção Bayesiana à Equalização

Duas estruturas de equalização bem conhecidas, que fazem estimação símbolo por símbolo, são os equalizadores transversais (*Transversal Equaliser - TE*) e os equalizadores transversais com *feedback* (*Decision Feedback Equaliser - DFE*). A estrutura do TE genérico é mostrada na Figura 5.1. Em essência o processo de equalização consiste na reconstituição do símbolo transmitido com base no vetor de saída do canal, observado pelo equalizador. Para um TE o vetor de saída do canal $\mathbf{r}(k)$ é tal como ilustrado na Figura 5.1.

Como se pode perceber, o vetor de saída do canal, observado pelo equalizador é da forma

$$\mathbf{r}(k) = [r(k) \quad r(k-1) \quad \cdots \quad r(k-m+1)]^T, \quad (5.13)$$

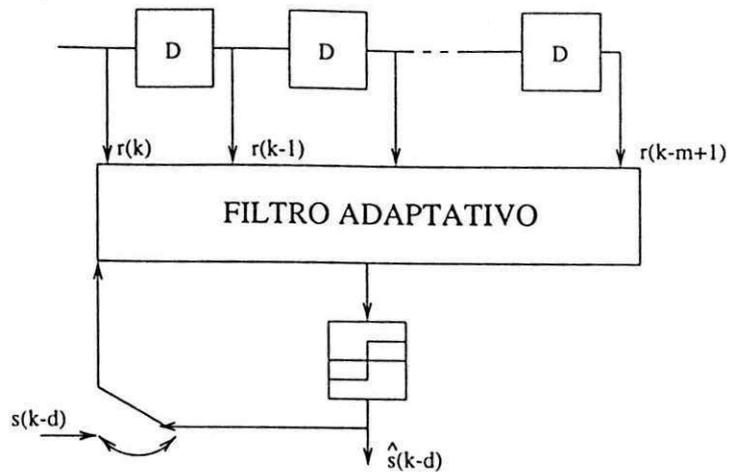


Figura 5.1: Estrutura genérica do equalizador transversal.

em que m é a ordem direta do equalizador. Com base nos estados assumidos pelo vetor \mathbf{r} no instante k o equalizador produz em sua saída uma estimativa $\hat{s}(k-d)$ do símbolo enviado $s(k-d)$, em que d é conhecido como atraso de decisão [16], [17]. Uma técnica usual para melhorar o desempenho de um equalizador é empregar um *link de feedback*. A estrutura genérica de um DFE é mostrada na Figura 5.2

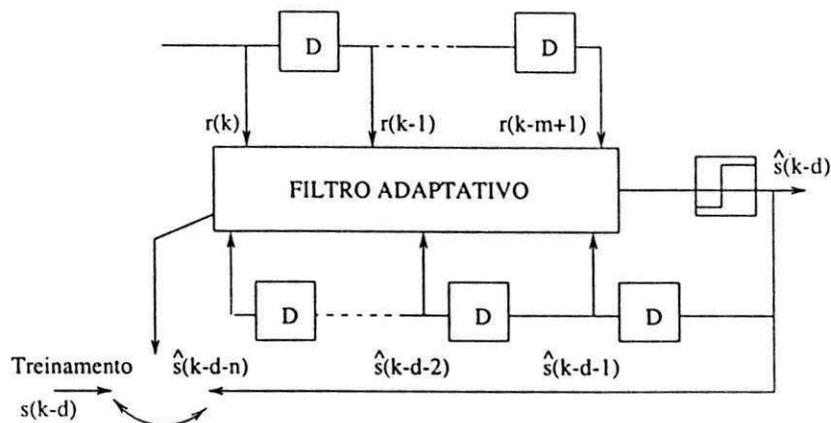


Figura 5.2: Equalizador bayesiano com feedback

em que o inteiro n é a ordem de *feedback*. Essa estrutura é obtida pela expansão das entradas do TE para incluírem os símbolos passados detectados. O TE e o DFE não são as soluções ótimas para essas duas estruturas de detecção símbolo por símbolo. As

soluções ótimas podem ser obtidas usando-se a teoria de decisão de Bayes e os detetores resultantes são referidos como TE bayesiano e DFE bayesiano.

No restante do capítulo serão mostradas duas arquiteturas de equalizadores implementados com a rede NRBF. Nas duas estruturas os dados de entrada são passados para os centros da camada intermediária da rede NRBF através de uma linha transversal com memória, como ilustrado na Figura 5.5. Os vetores de entrada são distribuídos entre os centros da rede de duas maneiras, que serão melhor analisadas juntamente com a organização dos estados do canal em conjuntos. Essa organização dos estados do canal permite que as duas estruturas estudadas sejam referidas respectivamente como equalizadores com *feedback* e sem *feedback*. É interessante lembrar que não será feita nenhuma operação de filtragem, como é feito no DFE convencional. Na Figura 5.3 a seguir é mostrada a posição do equalizador para que se possa analisar a notação utilizada para os vetores de amostras de entrada. Como se pode perceber, o papel do equalizador é estimar os símbolos transmitidos baseado nas observações ruidosas feitas na saída do canal.

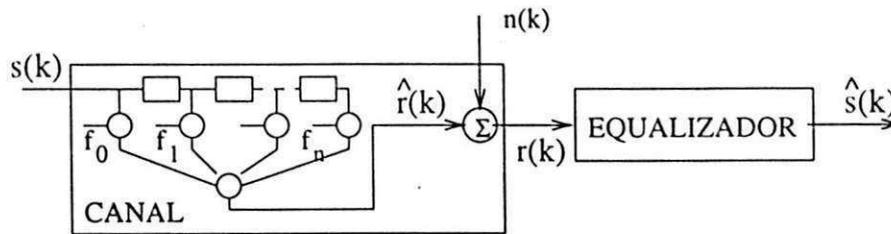


Figura 5.3: Modelo do sistema de comunicações.

em que $s(k)$ são os símbolos transmitidos e que no sistema QPSK podem ser dados por

$$s(k) = \begin{cases} s^{(1)} = 1 + j & \text{ou } (1, -1) \\ s^{(2)} = -1 + j & \text{ou } (-1, 1) \\ s^{(3)} = -1 - j & \text{ou } (-1, -1) \\ s^{(4)} = 1 - j & \text{ou } (1, -1). \end{cases} \quad (5.14)$$

A seguir será introduzido inicialmente o equalizador bayesiano sem *feedback* e em

seguida será mostrada sua ligação com a rede NRBF e a razão da melhora obtida com o *feedback*.

Um processo de equalização pode ser visto como um problema de classificação, ou mais precisamente um problema de classificação de padrões, usando-se a linguagem encontrada nos textos sobre RNA's. Nesse processo o equalizador busca classificar o vetor $\mathbf{r}(k)$ em um dos símbolos $s^{(i)}$, $1 \leq i \leq M$, em que M é igual a 4 no caso de se usar o sistema QPSK.

A partir da ligação em cascata da estrutura transversal linear do filtro FIR que modela o canal e da estrutura transversal do equalizador, pode-se perceber que o primeiro símbolo que influencia a decisão do equalizador no instante k é o símbolo $s(k)$ e o último símbolo a influenciar a decisão é $s(k - m - n_a + 2)$, em que $n_a - 1$ é a memória do canal. Desse modo a seqüência de símbolos que influenciam o processo de decisão do equalizador é dada por

$$\mathbf{s}(k) = [s(k) \ \cdots \ s(k - m - n_a + 2)]^T. \quad (5.15)$$

Desde que cada símbolo $s(k)$ pode assumir um dos quatro valores dados em 5.14, tem-se que a seqüência $\mathbf{s}(k)$ pode ter $n_s = 4^{m+n_a-1}$ combinações possíveis. Considerando-se na Figura 5.3 a ausência de ruído aditivo, tem-se que o vetor de saída do canal, visto pelo equalizador, é denotado por $\hat{\mathbf{r}}(k)$ e é dado por

$$\hat{\mathbf{r}}(k) = [\hat{r}(k) \ \cdots \ \hat{r}(k - m + 1)]^T \quad (5.16)$$

e tem n_s possíveis realizações de saída que são conhecidas como estados do canal. Todos esses vetores estado do canal podem ser agrupados em um conjunto denotado por $R_{m,d}$. O conjunto $R_{m,d}$ formado por todos os possíveis estados do canal pode ainda ser subdividido em M subconjuntos de acordo com os valores do símbolo $s(k - d)$, de modo que

$$R_{m,d} = \bigcup_{1 \leq i \leq M} R_{m,d}^{(i)}, \quad (5.17)$$

em que a notação $R_{m,d}^{(i)}$ indica o subconjunto formado pelos vetores estado $\hat{\mathbf{r}}(k)$ dado que o símbolo $s(k-d) = s_i$ tenha sido enviado, ou seja

$$R_{m,d}^{(i)} = \{\hat{\mathbf{r}}(k) | s(k-d) = s_i\}. \quad (5.18)$$

O número de vetores estado em cada subconjunto $R_{m,d}^{(i)}$ é dado por $n_s(i) = n_s/M = M^{m+n_a-2}$.

Devido ao ruído gaussiano aditivo na saída do canal ter média nula e variância σ_n^2 , o vetor $\mathbf{r}(k)$ é um processo estocástico, de modo que se pode aplicar aqui a teoria da decisão de Bayes, que fornece uma solução ótima para problemas de decisão. De acordo com a teoria de Bayes [16], o primeiro passo para se obter o símbolo correto é calcular as M variáveis bayesianas dadas por

$$\eta_i(k) = \sum_{j=1}^{n_s(i)} \alpha_j^{(i)} p_e(\mathbf{r}(k) - \hat{\mathbf{r}}_j^{(i)}), \quad 1 \leq i \leq M, \quad (5.19)$$

em que cada vetor $\hat{\mathbf{x}}_j^{(i)}$ pertence ao conjunto $R_{m,d}^{(i)}$, as probabilidades $\alpha_j^{(i)}$ representam as probabilidades a priori de $\hat{\mathbf{x}}_j^{(i)}$ e $p_e(\cdot)$ é a função densidade de probabilidade do ruído gaussiano. Cada $\eta_i(k)$ é uma fdp condicional de $\mathbf{r}(k)$ dado $s(k-d) = s_i$. Considerando que todos os estados do canal são equiprováveis (todos os $\alpha_j^{(i)}$ são iguais) e que a fdp do ruído é gaussiana, tem-se que as variáveis de decisão bayesianas podem ser escritas explicitamente na forma

$$\eta_i(k) = \sum_{j=1}^{n_s(i)} h^{(i)} \exp\left(\frac{-\|\mathbf{x}(k) - \mathbf{x}_j^{(i)}(k)\|^2}{2\sigma_n^2}\right), \quad 1 \leq i \leq M, \quad (5.20)$$

em que $h^{(i)}$ é uma variável que está relacionada às probabilidades a priori dos vetores estado $\hat{\mathbf{r}}_j^{(i)}$.

A decisão pelo símbolo correto, que minimiza a probabilidade de erro na detecção, deve ser dada, de acordo com a regra de Bayes, por

$$\hat{s}(k-d) = s^{(i)} \quad \text{se} \quad \eta = \max\{\eta_i(\mathbf{r}(k)), 1 \leq i \leq M\}. \quad (5.21)$$

É interessante, a partir deste ponto, visualizar a relação existente entre a solução bayesiana e a rede NRBF. Como foi visto no capítulo anterior a rede NRBF é uma rede neural de uma única camada escondida em que cada nó escondido tem uma resposta que é radialmente simétrica em torno de um vetor chamado centro e o nó de saída é um combinador linear. Como foi visto, a resposta de uma rede de m entradas, n_h unidades intermediárias e uma única saída é dada por

$$y(\mathbf{x}(k)) = \sum_{i=0}^{n_h-1} w_i \varphi(\|\mathbf{x}(k) - \mathbf{t}_i\|^2 / \rho_i), \quad (5.22)$$

em que \mathbf{t}_i é o i -ésimo vetor centro m -dimensional, ρ são escalares reais e positivos que controlam a suavidade das funções $\varphi(\cdot)$, w_i são os pesos da camada linear que podem ter valores complexos ou não e $\varphi(\cdot)$ são as funções de bases radiais de cada nó escondido. Dado o canal, é possível inicializar todos os parâmetros da rede. O número de nós da camada escondida é igual ao número de estados do canal e esses estados correspondem aos centros da rede NRBF. As funções $\varphi(\cdot)$ são escolhidas como sendo a função gaussiana por motivos já justificados e os parâmetros ρ_i são estabelecidos iguais a $2\sigma_n^2$. Na Equação 5.20 os pesos h_i podem ser estabelecidos iguais a 1 ou -1 correspondendo respectivamente aos centros dos subconjuntos $R_{m,d}^{(1)}$ e $R_{m,d}^{(2)}$ no caso dos símbolos $s^{(i)}$ serem binários ou podem ser estabelecidos iguais a $h^{(1)} = 1 + j1$, $h^{(2)} = -1 + j1$, $h^{(3)} = -1 - j1$ e $h^{(4)} = 1 - j1$, correspondendo respectivamente aos centros dos subconjuntos $R_{m,d}^{(1)}$, $R_{m,d}^{(2)}$, $R_{m,d}^{(3)}$ e $R_{m,d}^{(4)}$ no caso do sistema QPSK. O equalizador bayesiano sem *feedback* pode ser visto como uma estrutura mostrada na Figura 5.4

Pela figura 5.4 pode-se perceber que cada uma das variáveis bayesianas dadas em 5.19 é fornecida por uma rede NRBF de uma única camada escondida. O símbolo estimado na saída corresponderá ao índice da variável bayesiana que possui maior valor e é fornecido pelo bloco MAXNET que é uma função que retorna o valor da

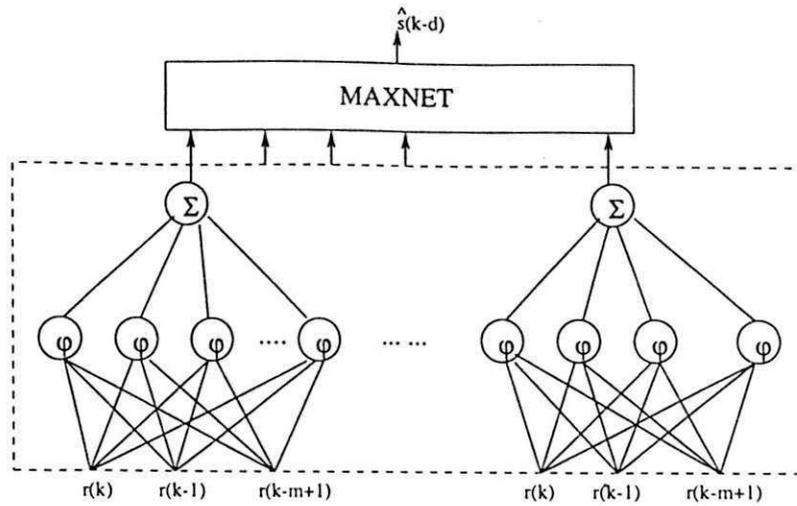


Figura 5.4: Estrutura do equalizador bayesiano sem *feedback*.

maior variável de decisão calculada, de modo que o símbolo estimado $\hat{s}(k)$ é dado na saída dessa função por

$$\hat{s}(k) = \text{sgn} \left(\sum_{l=1}^M \eta_l(\mathbf{r}(k)) \right), \quad (5.23)$$

em que a função $\text{sgn}(\cdot)$ é dada por

$$\text{sgn}(y) = \begin{cases} 1 + j, & \text{Re}[y] \geq 0 \cap \text{Im}[y] \geq 0, \\ 1 - j, & \text{Re}[y] \geq 0 \cap \text{Im}[y] < 0, \\ -1 + j, & \text{Re}[y] < 0 \cap \text{Im}[y] \geq 0, \\ -1 - j, & \text{Re}[y] < 0 \cap \text{Im}[y] < 0. \end{cases} \quad (5.24)$$

Em cada camada escondida o número de centros da rede NRBF corresponde ao número de estados do canal dividido por M , ou seja, $n_s^{(i)} = n_s/M = M^{m+n_a-2}$. Como se pode ver pela Figura 5.4, a solução bayesiana é implementada por uma rede formada por M sub-redes. Na verdade, tem-se apenas uma sub-rede com a camada escondida organizada de modo que seus centros fiquem distribuídos entre os respectivos subconjuntos $R_{m,d}^{(i)}$, $1 \leq i \leq M$.

Uma das maiores desvantagens da estrutura de equalização bayesiana sem *feedback* é a necessidade do uso de todos os estados do canal. Na implementação realizada com a rede NRBF são necessários todos os $n_s = M^{m+n_a-1}$ estados colocados no lugar dos centros da rede. No sistema QPSK, por exemplo, são necessários 4^{m+n_a-1} estados do canal e à medida que m aumenta, aumenta a complexidade computacional.

Uma maneira de sobrepor esse problema é usando o *feedback* com ordem $n = n_a + m - 2 - d$ [16], de modo que o vetor de realimentação $\hat{s}_f(k-d)$ tenha $n_f = M^n$ estados. Os estados do vetor $\hat{s}_f(k-d)$ são formados por todas as possíveis combinações que esse vetor pode assumir com os símbolos $s^{(i)}$, $1 \leq i \leq M$, e são denotados por $s_{f,j}$, $1 \leq j \leq n_f$. Um subconjunto formado pelos estados do canal, $R_{m,d}^{(i)}$, pode ser ainda particionado em n_f subconjuntos, de acordo com os estados do vetor $\hat{s}_f(k-d)$, da seguinte maneira

$$R_{m,d}^{(i)} = \bigcup_{1 \leq j \leq n_f} R_{m,d,j}^{(i)}, \quad (5.25)$$

em que o subconjunto $R_{m,d,j}^{(i)}$ é formado por todos os vetores $\hat{r}(k)$ dado que $s(k-d) = s^{(i)}$ e que $\hat{s}_f(k-d) = s_{f,j}$. Ou seja,

$$R_{m,d,j}^{(i)} = \{\hat{r} | s(k-d) = s^{(i)} \cap \hat{s}_f(k-d) = s_{f,j}\}, \quad 1 \leq j \leq n_f. \quad (5.26)$$

O número de estados do canal $\hat{r}(k)$ em cada subconjunto $R_{m,d,j}^{(i)}$ é igual a $n_{s,j}^{(i)} = n_s^{(i)} / n_f$. Considerando-se o vetor de *feedback* condicionado ao vetor $\hat{s}_f(k-d)$, as M variáveis de decisão bayesianas condicionadas ao vetor $\hat{s}_f(k-d) = s_{f,j}$ são dadas por

$$\eta_i(r(k) | \hat{s}_f(k-d) = s_{f,j}) = \sum_{l=1}^{n_{s,j}^{(i)}} h_l \exp\left(-\frac{\|\mathbf{r}(k) - \hat{\mathbf{r}}_l^{(i)}\|^2}{2\sigma_n^2}\right), \quad 1 \leq i \leq M, \quad (5.27)$$

em que $\hat{\mathbf{r}}_l^{(i)} \in R_{m,d,j}^{(i)}$. A decisão bayesiana condicional é obtida por sua vez substituindo-se a variável bayesiana $\eta_i(k | \hat{s}_f(k-d) = s_{f,j})$ pela variável $\eta_i(k)$ na Equação 5.21, de modo que

$$\hat{s}(k-d) = s^{(i)} \text{ se } \eta = \max\{\eta_i(k) | \hat{s}_f(k-d) = s_{f,j}\}, \quad 1 \leq i \leq M, \quad 1 \leq j \leq n_f\}. \quad (5.28)$$

É interessante notar que o vetor de *feedback* é usado apenas para reduzir o número de estados necessários no cálculo das variáveis de decisão. Sem o *feedback* todos os estados do canal são requeridos para o cálculo das M variáveis de decisão. Com o *feedback* somente uma fração do número total de estados, n_s/n_f , é necessária.

A outra importante característica que torna o equalizador bayesiano com *feedback* superior à estrutura sem *feedback* é a melhora na capacidade de separar padrões. A melhora obtida com o *feedback* se tornará mais clara quando forem examinados os dois exemplos a seguir, para o caso binário.

Exemplo 1 – Suponha a transmissão de símbolos binários por meio de um canal com função de transferência dada por $A(z) = a + bz^{-1}$. Considere que o equalizador tenha ordem direta $m = 3$ e atraso de decisão $d = 1$. Neste caso todas as possíveis combinações da seqüência de símbolos de entrada e os respectivos estados do canal são mostrados na tabela 5.1.

Na parte superior direita da tabela 5.1 pode-se ver todos os vetores estados do canal dado que o símbolo transmitido foi $s(k-1) = -1$ e na parte inferior direita tem-se os vetores estados dado que o símbolo transmitido foi $s(k-1) = 1$. Um diagrama ilustrativo do processo de chegada de símbolo e do deslocamento desses símbolos na linha transversal do equalizador é mostrado na Figura 5.5.

Como se pode perceber na Figura 5.5, os símbolos $s(k-1)$, na saída do canal, entram no equalizador e vão sendo deslocados para a direita à medida que novos símbolos vão chegando. Na Figura 5.5 os vetores formados pelos símbolos deslocados estão colocados uns acima dos outros. Pode-se ver que a partir do $(m+1)$ -ésimo símbolo de entrada, o vetor $\hat{\mathbf{r}}$ visto pelo equalizador é um dos vetores da Tabela 5.1. Desse modo, todas as vezes que um novo símbolo $s(k-1)$ entra no equalizador, um novo vetor estado é apresentado às funções de base radiais que formam a rede NRBF que realiza essa solução bayesiana. Nesse exemplo o ruído gaussiano aditivo do canal foi omitido para

$s(k)$	$s(k-1)$	$s(k-2)$	$s(k-3)$	$\hat{r}(k)$	$\hat{r}(k-1)$	$\hat{r}(k-2)$
-1	-1	-1	-1	$-a-b$	$-a-b$	$-a-b$
-1	-1	-1	1	$-a-b$	$-a-b$	$a-b$
-1	-1	1	-1	$-a-b$	$a-b$	$-a+b$
-1	-1	1	1	$-a-b$	$a-b$	$a+b$
1	-1	-1	-1	$-a+b$	$-a-b$	$-a-b$
1	-1	-1	1	$-a+b$	$-a-b$	$a-b$
1	-1	1	-1	$-a+b$	$a-b$	$-a+b$
1	-1	1	1	$-a+b$	$a-b$	$a+b$
-1	1	-1	-1	$a-b$	$-a+b$	$-a-b$
-1	1	-1	1	$a-b$	$-a+b$	$a-b$
-1	1	1	-1	$a-b$	$a+b$	$-a+b$
-1	1	1	1	$a-b$	$a+b$	$a+b$
1	1	-1	-1	$a+b$	$-a+b$	$-a-b$
1	1	-1	1	$a+b$	$-a+b$	$a-b$
1	1	1	-1	$a+b$	$a+b$	$-a+b$
1	1	1	1	$a+b$	$a+b$	$a+b$

Tabela 5.1: Combinações de entrada e os respectivos estados do canal.

permitir uma melhor análise do exemplo. Quando as amostras de ruído são adicionadas aos estados do modelo do canal, o vetor que passa a ser visto pelo equalizador deixa de ser o vetor $\hat{\mathbf{r}}$ e passa a ser \mathbf{r} . Em cada nó intermediário da rede NRBF é calculada a distância entre o vetor \mathbf{r} e o centro da função NRBF daquele respectivo nó. Se no instante $k = 10$ o vetor visto pelo equalizador, for por exemplo, da forma

$$\mathbf{r} = [a - b + n_1 \quad -a + b + n_2 \quad a - b + n_3]^T, \quad (5.29)$$

em que n_1, n_2, n_3 e n_4 são amostras de ruído gaussiano de variância σ_n^2 e se a relação sinal ruído dada por

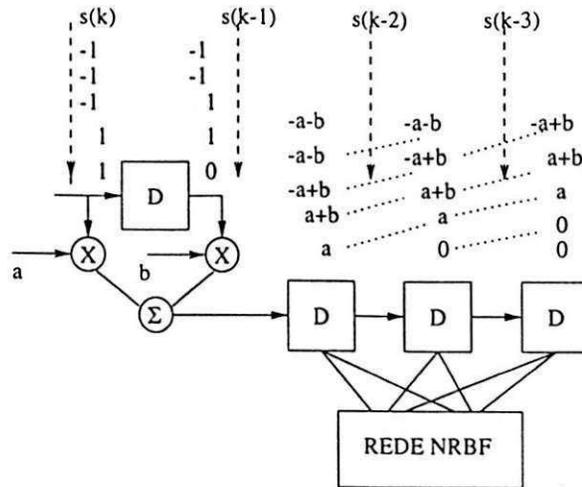


Figura 5.5: Processamento dos dados no equalizador bayesiano.

$$SNR_{dB} = 10 \log \left(\sum_{i=1}^{n_a} f_i^2 \frac{\sigma_s^2}{\sigma_n^2} \right) \quad (5.30)$$

for alta, então o vetor \mathbf{r} , visto pelo equalizador, estará mais próximo do vetor estado

$$\hat{\mathbf{r}}(k) = [a - b \quad -a + b \quad a - b]^T, \quad (5.31)$$

fazendo com que a decisão do equalizador seja pelo símbolo $\hat{s}(k-1) = 1$.

Exemplo 2 – Nas duas colunas da Tabela 5.1 correspondentes aos símbolos $s(k-2)$ e $s(k-3)$, pode-se perceber os estados de *feedback* do equalizador. Está destacado em **negrito** apenas o estado $\hat{\mathbf{s}}_{f,i} = [-1 \quad -1]^T$. Como foi analisado anteriormente, não é necessário o uso de um filtro de *feedback*. O que é feito é apenas uma reorganização dos padrões de classificação, distribuídos no conjunto mostrado na Equação 5.26. Essa nova divisão dos padrões de classificação diminui a complexidade do cálculo das variáveis de decisão bayesianas e permite que se use a ordem m do *link* direto do equalizador maior que o que é possível sem o *feedback*.

5.4 Conclusão

Como foi citado no início deste capítulo, um dos principais motivos que levaram ao uso da teoria bayesiana no problema de detecção de sinais em canais com o comportamento descrito no Capítulo 2 foi o fato dessa teoria minimizar o erro de decisão. Um ganho obtido com o uso do *feedback* é a diminuição do número de saídas da camada intermediária da rede NRBF, que precisa ser somado para que a tomada de decisão seja feita. Como já foi dito, essa diminuição reduz a complexidade computacional do cálculo das variáveis de decisão e reduz o tempo de processamento do equalizador.

Capítulo 6

Descrição do Sistema Simulado e Análise dos Resultados

6.1 Introdução

Nos capítulos anteriores foi mostrada, de acordo com a organização da dissertação, um pouco da teoria relativa ao modelamento do canal de comunicações utilizado neste trabalho e foi dada uma introdução à teoria da detecção bayesiana, às redes neurais artificiais, às redes com funções de base radiais e ao algoritmo LMS, de modo que cada um desses tópicos foi aplicado na implementação de uma parte do sistema completo, desde a fonte de símbolos até o demodulador.

Alguns princípios da detecção bayesiana, por exemplo, foram utilizados na elaboração do detetor ótimo, referido como equalizador ou detetor bayesiano e a rede NRBF foi usada na elaboração do detetor. Já o algoritmo LMS, por sua vez, foi usado na estimação dos coeficientes do modelo de canal utilizado. Na Figura 6.1 é mostrado um diagrama contendo todos os blocos do sistema implementado. Todos esses blocos com exceção do último foram implementados em linguagem C e os programas resultantes foram compilados e executados usando o compilador GCC do sistema UNIX.

A partir da próxima seção serão analisados em detalhes os blocos da Figura 6.1, bem como os parâmetros utilizados, as dificuldades encontradas e os resultados obtidos. Para cada parâmetro utilizado e testado no equalizador foram traçadas curvas de taxa

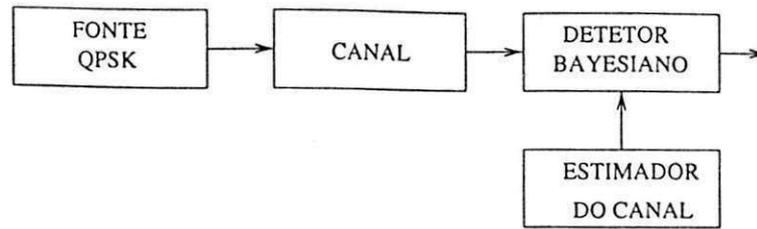


Figura 6.1: Modelo do Sistema Implementado.

de erro de símbolos, já que esse foi o critério de desempenho utilizado na avaliação dos resultados.

6.2 Simulação do Canal

No Capítulo 2 há uma seção que explica com mais detalhes o procedimento utilizado na simulação do canal, de modo que nesta seção será feita apenas uma complementação do que já foi analisado anteriormente, para que se possa em seguida apresentar os resultados obtidos. Como foi visto no capítulo 2, cada linha de atraso do canal seletivo em frequência pode ser modelada por um processo estocástico gaussiano complexo da forma

$$h(t, \tau) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} a_k \delta(t - \tau_k) e^{j2\pi\nu_k t}, \quad (6.1)$$

em que N representa o número de percursos através dos quais a informação é espalhada em seu percurso fonte-receptor.

Na simulação desse método, as variáveis aleatórias a_k têm média nula, variância unitária, são estatisticamente independentes entre si e entre as demais variáveis. Esses parâmetros podem ser modelados por $e^{j\theta_k}$, em que θ_k tem distribuição uniforme no intervalo $[0, 2\pi)$. O valor de N pode ser estabelecido em torno de 20 a 30 que já se obtém uma boa aproximação para $h(t, \tau)$.

Como ilustração, foram geradas 10000 amostras do vetor \mathbf{h} e o resultado dessa geração pode ser visto na Figura 6.2. Como se pode perceber na Figura 6.2, o número de amostras distribuídas nos subintervalos entre o valor mínimo e o valor máximo do

módulo da variável h_k se aproxima da distribuição de Rayleigh. É de se esperar que a medida que o número de amostras de h aumenta, a frequência relativa das amostras geradas no intervalo citado tende a se aproximar da distribuição de Rayleigh.

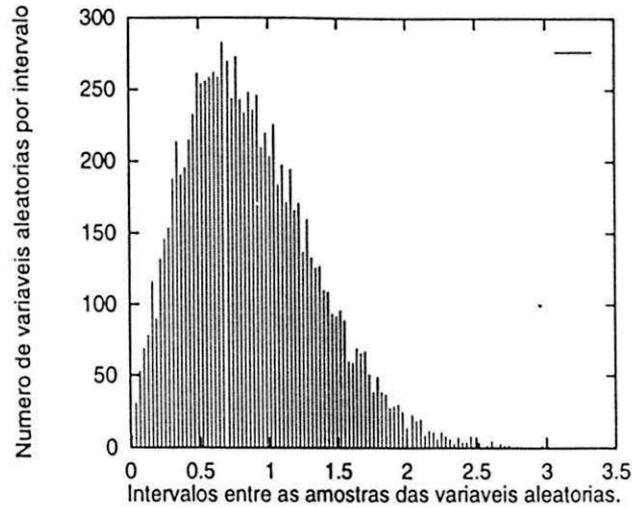


Figura 6.2: Distribuição do módulo das amostras do processo $h(t, \tau)$ entre o seu valor máximo e seu valor mínimo.

Os resultados da simulação dos pesos do canal podem ser avaliados através das Figuras 6.3 e 6.4 dadas a seguir e através da densidade espectral de potência dessas amostras. Na Figura 6.3 é mostrado o módulo da resposta impulsional do canal para uma taxa de transmissão $fs = 24300$ Hz, com frequências Doppler de $fd = 4$ Hz, $fd = 40$ Hz e $fd = 400$ Hz.

Como se pode perceber, o canal varia mais rápido ou mais lento de acordo com a variação da frequência Doppler, em relação à taxa de transmissão. Já na Figura 6.4 pode-se ver a variação da parte real da autocorrelação das amostras de $h(t, \tau)$.

A autocorrelação discreta das amostras de $h(t, \tau)$ é estimada através da média [1]

$$R_{hh} = \frac{1}{NN - k} \sum_{m=1}^{NN-k} h(m)h^*(m+k), \quad k = 0, 1, \dots, NN - 1. \quad (6.2)$$

A FFT dessas amostras dá origem a densidade espectral de potência (DSP) do processo $h(t, \tau)$ que tem a forma do espectro dado pela Equação 2.21. Na Figura 6.5 pode-se ver

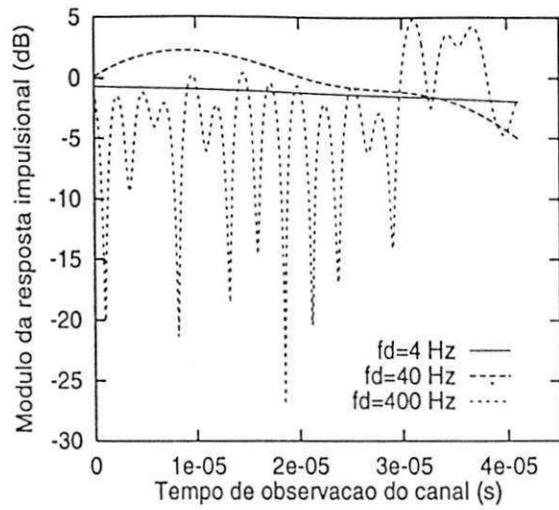


Figura 6.3: M3dulo da resposta impulsional do canal.

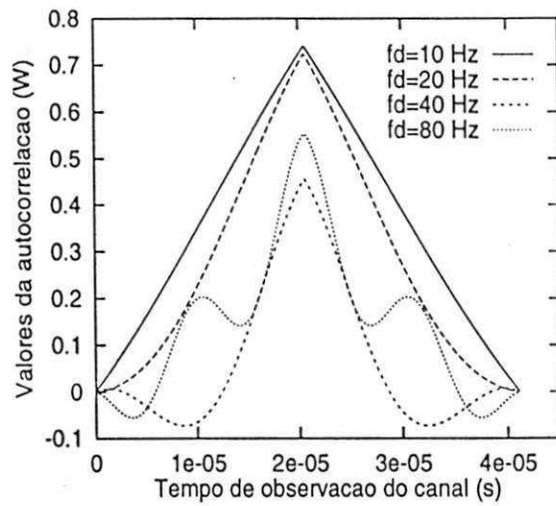
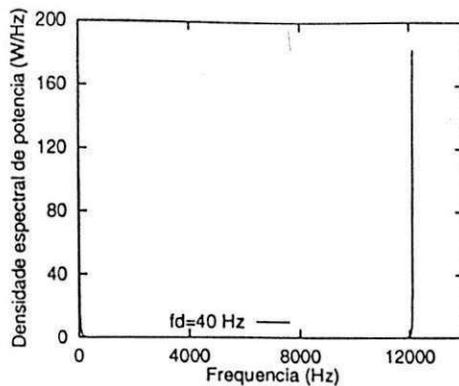
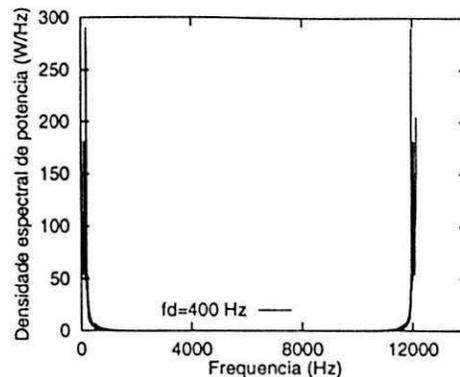


Figura 6.4: Autocorrela3o do processo $h(t, \tau)$, para τ fixo.



(a) DSP para $fd = 40$ Hz.



(b) DSP para $fd = 400$ Hz.

Figura 6.5: Densidade espectral de potência do processo $h(t, \tau)$ simulado.

a densidade espectral de potência do processo $h(t, \tau)$ para 512 amostras, $fs = 24300$ Hz e respectivamente para $fd = 40$ Hz e $fd = 400$ Hz.

Os resultados obtidos nas curvas acima são para um único conjunto de amostras do processo $h(t, \tau)$. Para simular mais de um processo é só chamar a rotina que implementa esse método mais de uma vez. Em um canal com n_a linhas de atraso, são necessários n_a conjuntos de amostras de $h(t, \tau)$ com cada conjunto de amostras multiplicado por um valor de potência P_i , especificado pelo projetista do modelo do canal. Essas potências devem ser tais que

$$\sum_{i=0}^{n_a-1} P_i = 1; \quad (6.3)$$

Antes de passar à próxima seção é interessante citar que o método de somas de senoides também foi analisado e simulado [18], mas não apresentou bons resultados. Esse método é usado para gerar variáveis aleatórias gaussianas e é comumente usado nas simulações em *hardware*.

Feitas estas considerações, pode-se passar à próxima seção em que serão analisados o procedimento de simulação do equalizador e os resultados obtidos.

6.3 Simulação de Equalizador

Os primeiros resultados desta seção foram obtidos usando-se a estrutura da rede NRBF para implementar as variáveis de decisão do método bayesiano de detecção de sinais. Esse método é um procedimento derivado dos detetores que fazem estimação baseados na máxima probabilidade a posteriori. Pode-se dizer que o método bayesiano é um procedimento ótimo de detecção símbolo por símbolo, que apresenta um bom desempenho em relação a outros métodos de detecção, como por exemplo as técnicas baseadas na estimação de seqüências de máxima probabilidade (*Maximum Likelihood Sequence Estimation - MLSE*). Embora a estimação dessas seqüências seja o procedimento que teoricamente fornece o melhor desempenho em termos de taxa de erro, ele apresenta a desvantagem da complexidade de implementação enquanto que a estrutura bayesiana com a rede NRBF é relativamente bem mais simples de se projetar.

Sem dúvida, o método bayesiano também apresenta algumas características indesejáveis. Uma delas é a complexidade computacional, bem diferente da complexidade de implementação, que pode crescer demasiadamente dependendo de fatores como o número de coeficientes do modelo do canal utilizado, a ordem m da linha direta da rede NRBF e o atraso de decisão d . Como foi visto no Capítulo 4, o número de centros necessários para a implementação da camada intermediária da rede NRBF depende intimamente desses parâmetros listados e é dado por

$$n_s = M^{m+n_a-1}, \quad (6.4)$$

em que M é o número de símbolos da constelação M -PSK, m é a ordem da linha direta, que corresponde ao número de entradas da rede NRBF e n_a é o número de coeficientes do modelo do canal. Como se pode perceber, com um número de símbolos maior que 4, a complexidade computacional da rede NRBF já começa a se tornar pouco apropriada. O que permite que esses parâmetros ainda possam aumentar é o uso de um enlace de realimentação (*feedback*). Equalizadores com *feedback* são conhecidos como *feedback decision equalizers* e diminuem bastante o custo computacional, como já foi analisado no Capítulo 5.

Na Figura 6.6, curva superior, pode-se ver, como exemplo, o desempenho de um equalizador bayesiano implementado usando-se um modelo de canal seletivo completamente conhecido com função de transferência dada por

$$H(z) = (0.0119 + j0.7589) + (-0.3796 + j0.5059)z^{-1}. \quad (6.5)$$

Essa estrutura sem *feedback* foi implementada com $m = 2$ e um atraso de decisão $d = 1$. Na curva inferior da mesma figura é mostrada a taxa de erro correspondente à estimação de seqüências que foi reproduzida a partir da referência [13], nas mesmas condições do modelo do canal e usando um equalizador MLSE com atraso de decisão igual a 5.

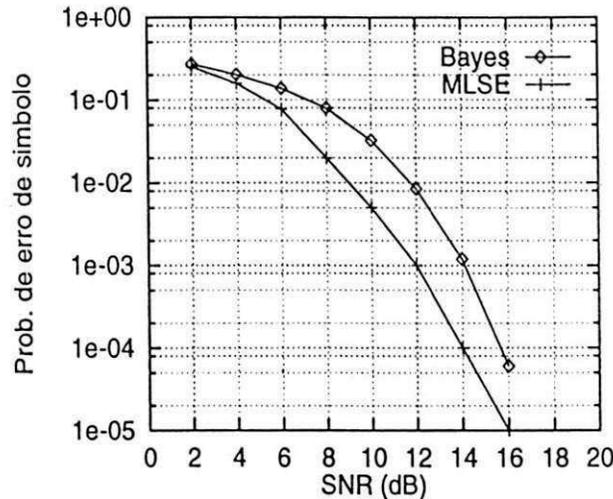


Figura 6.6: Comparação entre o equalizador bayesiano sem *feedback* e o método de estimação de seqüências de máxima verossimilhança para o modelo de canal conhecido $H(z) = (0.0119 + j0.7589) + (-0.3796 + j0.5059)z^{-1}$.

Como se pode perceber na Figura 6.6 o desempenho, em termos da taxa de erro de símbolos, do equalizador MLSE é superior em relação ao equalizador bayesiano sem *feedback*.

Já na Figura 6.7 são mostradas três curvas comparando o desempenho do equalizador bayesiano sem *feedback* e considerando-se que o canal seja completamente conhecido

e tenha a função de transferência dada na Equação 6.5. As curvas foram obtidas com os parâmetros $m = 2$, $m = 4$ e $m = 6$ e com $d = 1$. Pode-se perceber que há uma pequena diminuição na taxa de erro de símbolos quando o parâmetro m , que é a ordem da linha direta do equalizador, aumenta de $m = 2$ para $m = 4$. Com $m = 4$, $n_a = 2$ e $d = 1$, entretanto, já se faz necessário o cálculo de 1024 estados do canal e para $m = 6$ já são necessários 4096. Como já foi dito antes, todos esses estados devem ser passados para os respectivos centros das funções de base radial da rede NRBF.

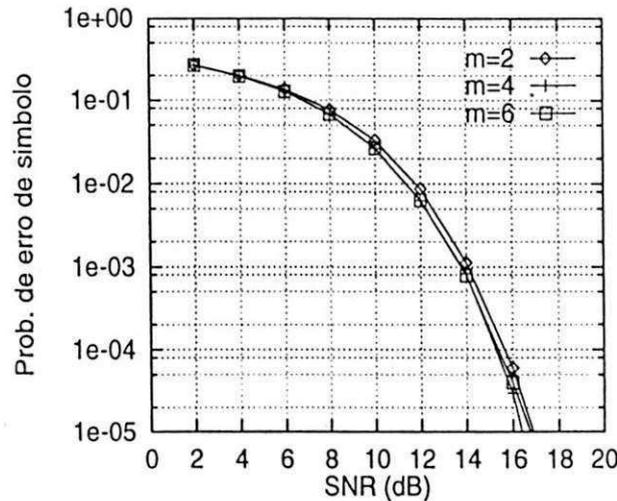


Figura 6.7: Comparação entre o desempenho do equalizador bayesiano sem *feedback*, respectivamente com ordem da linha direta $m = 2$, $m = 4$ e $m = 6$ para o modelo de canal conhecido $H(z) = (0.0119 + j0.7589) + (-0.3796 + j0.5059)z^{-1}$.

Na outra curva mostrada na Figura 6.8 é mostrado o desempenho do equalizador bayesiano com *feedback*, *feedback decision equalizer*, com parâmetros $m = 4$, $d = 1$ e com o canal apresentado na Equação 6.5 comparado com o equalizador bayesiano sem *feedback* e com os mesmos parâmetros.

Embora a melhora na taxa de erro seja, de acordo com a Figura 6.8, praticamente imperceptível, para os parâmetros apresentados, há um considerável ganho no tempo de processamento do sistema e uma considerável diminuição na complexidade computacional, já que o equalizador não tem mais que somar, para cada uma das variáveis de decisão, $n_s/4$ saídas dos n_s nós da camada intermediária da rede NRBF.

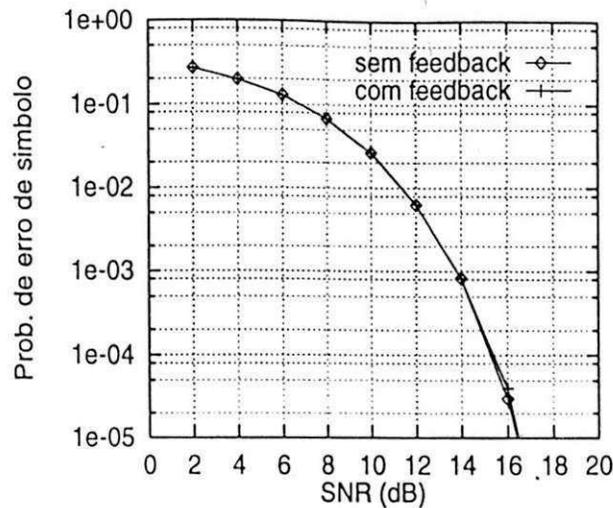


Figura 6.8: Comparação entre o desempenho do equalizador bayesiano sem *feedback* e com *feedback* para o modelo de canal conhecido $H(z) = (0.0119 + j0.7589) + (-0.3796 + j0.5059)z^{-1}$.

Na Figura 6.9 é mostrada uma comparação entre as taxas de erro de símbolos entre duas estruturas de equalizadores usando *feedback*, com $m = 3$, $m = 4$ e $d = 1$. O parâmetro d é dado por $d = n_a - 1$, em que n_a é o número de coeficientes do canal. Como se pode perceber, há uma pequena melhora na taxa de erro de bits quando o aumento na ordem da linha direta do equalizador é apenas de $m = 3$ para $m = 4$. Na referência [13] é mostrado que um detetor bayesiano com $m = d + 1$ tem o mesmo desempenho que um detetor com $m > d + 1$. O ganho obtido na taxa de erro depende dos parâmetros d e m . O parâmetro n é o tamanho do vetor de *feedback*, também depende de m e d .

Até aqui considerou-se o canal conhecido, enquanto que para os resultados obtidos a partir deste ponto, foi considerada a sua característica variante no tempo. O modelo de canal implementado consistiu de um filtro FIR com 2 coeficientes, dados pela Expressão 6.1. Embora haja uma forma mais simples para simular um canal com apenas dois coeficientes, esse procedimento permite uma expansão direta para um canal de ordem n_a superior. Os coeficientes do canal são dados pelo processo estocástico complexo $h(t, \tau)$ dado na Equação 6.1, em que o módulo tem distribuição de probabilidade de Rayleigh e a fase tem distribuição uniforme em $[0, 2\pi)$.

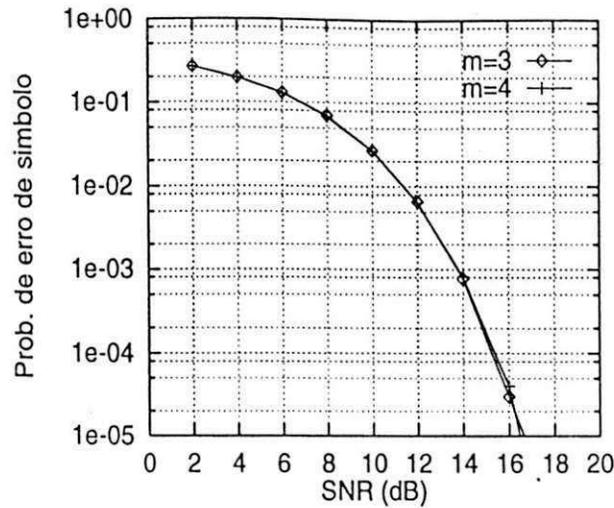


Figura 6.9: Comparação entre o desempenho do equalizador bayesiano com *feedback* para $m = 3$ e $m = 4$ e para o modelo de canal conhecido $H(z) = (0.0119 + j0.7589) + (-0.3796 + j0.5059)z^{-1}$.

À medida que o canal varia, torna-se necessário fazer o treinamento de todos os centros das funções da camada intermediária da rede NRBF. Como já foi dito, esses centros são na verdade os estados do canal que são obtidos passando-se todas as possíveis combinações da sequência de símbolos de entrada pelo modelo discreto do canal. O treinamento desses centros foi feito à parte, externamente à estrutura da rede, usando o algoritmo LMS. O problema de estimação dos coeficientes do modelo do canal pode ser melhor ilustrado pelo diagrama do processador adaptativo na Figura 6.10.

Como se pode perceber na Figura 6.10, as amostras de entrada x_i , passam pelo canal e servem de amostras de treinamento para o processador adaptativo, que é um filtro FIR com os coeficientes ajustados pelo algoritmo LMS. Após um certo número de amostras de treinamento os pesos do processador adaptativo se igualam aos do canal desconhecido. Esse procedimento é bastante comum em problemas de estimação de plantas de sistemas desconhecidos. Na Figura 6.11 pode-se ver, como exemplo, as curvas de convergência do erro médio quadrático do algoritmo LMS para uma relação sinal-ruído SNR=15 dB.

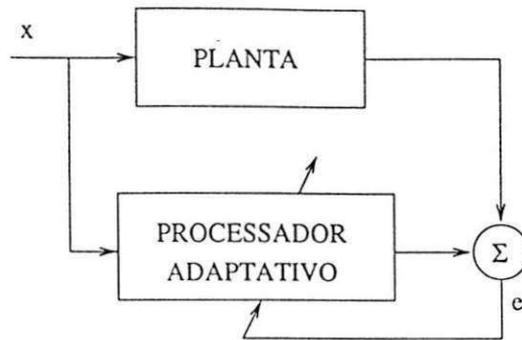


Figura 6.10: Diagrama usado para a estimação dos coeficientes do modelo do canal usando o algoritmo LMS.

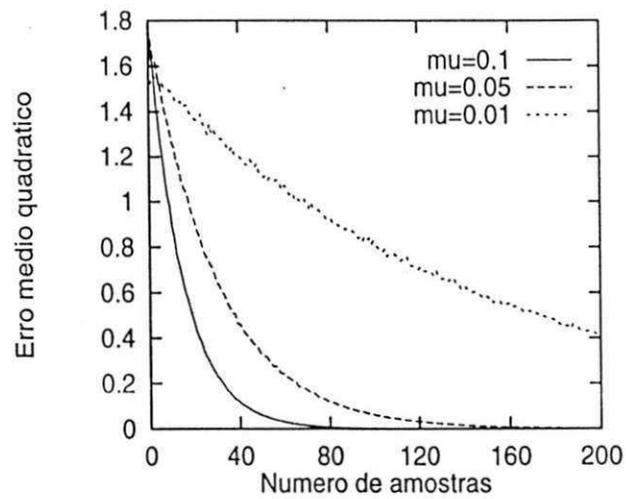


Figura 6.11: Curvas de convergência do erro médio quadrático do LMS.

Como se pode perceber, há uma variação no tempo de convergência do algoritmo LMS de acordo com o parâmetro de ajuste dos passos μ . Em canais cuja característica muda muito rápido, há a necessidade de um ajuste quase contínuo que pode ser feito também por um algoritmo de *clustering*. É interessante também visualizar a trajetória que os pesos do processador adaptativo descrevem até convergirem para os valores que resultam na estimativa do canal. Na Figura 6.12 são mostradas as trajetórias de dois conjuntos com dois pesos cada. Pode-se perceber que os pesos partem de um valor nulo e convergem, no caso deste gráfico após 200 amostras de treinamento, para um conjunto de valores iguais aos pesos do canal.

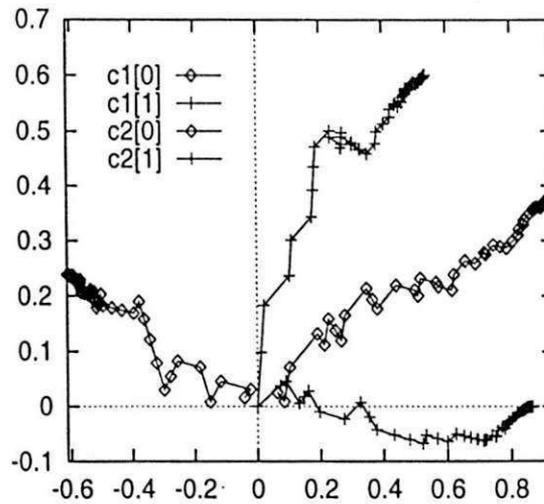


Figura 6.12: Trajetória dos pesos do processador adaptativo para $\mu = 0.1$.

É interessante lembrar que outros algoritmos de estimação poderiam ser usados, entretanto o LMS convencional de valores complexos é um dos mais fáceis de se implementar. Em comparação ao LMS normalizado o LMS convencional apresenta um tempo de convergência menor e em relação ao RLS não há necessidade de avaliar a matriz de correlação dos dados de entrada.

Existem duas formas de usar o LMS para estimar os estados do canal, necessários ao detetor bayesiano [19]. A primeira forma, para canais com efeito Doppler rápido em relação à taxa de transmissão, usa um vetor estimativa $\hat{\mathbf{a}}$ definido como

$$\hat{\mathbf{a}}(k) = [\hat{a}_0(k) \quad \hat{a}_1(k) \quad \cdots \quad \hat{a}_{n_a-1}(k)]^T, \quad (6.6)$$

de modo que na amostra k

$$\begin{cases} \epsilon(k-d) = x(k-d) - \hat{\mathbf{a}}^T(k-d-1)\hat{\mathbf{s}}_a(k-d) \\ \hat{\mathbf{a}}(k-d) = \hat{\mathbf{a}}(k-d-1) + \mu\epsilon(k-d)\hat{\mathbf{s}}_a^*(k-d), \end{cases} \quad (6.7)$$

em que $*$ denota a operação conjugado complexo, μ é o parâmetro de ajuste dos valores de $\hat{\mathbf{a}}$ e

$$\hat{\mathbf{s}}_a(k-d) = [\hat{s}(k-d) \quad \hat{s}(k-d-1) \quad \cdots \quad \hat{s}(k-d-n_a+1)]^T. \quad (6.8)$$

No instante k o detector deve usar a estimativa atrasada $\hat{\mathbf{a}}(k-d)$ como se fosse a mais recente, para fazer a decisão, enquanto os coeficientes do canal podem ter mudado bastante. Por isso é importante usar um atraso de decisão d tipicamente de ordem $d = 1$ ou $d = 2$ para minimizar o erro de rastreamento decorrente desse procedimento.

Uma outra maneira de usar o LMS na estimação do canal é verificando o seu comportamento para determinados valores pequenos de frequência Doppler. Na Figura 6.3, por exemplo, pode-se ver que em relação a uma taxa de transmissão de 24300 Hz, uma frequência Doppler de 4 Hz torna o canal praticamente invariante no tempo. Isso significa que pode-se manter uma estimativa fixa durante um certo conjunto de bits enviados pela fonte. Esse procedimento para canais lentos diminui a necessidade da estimação contínua do canal e o esforço computacional do cálculo dos estados.

Na Figura 6.13 é mostrada uma comparação entre o desempenho de dois equalizadores bayesianos com links de realimentação, em um canal variante de dois raios e com frequência Doppler de 4 Hz. As ordens da linha direta de cada equalizador são respectivamente $m = 2$ e $m = 4$.

Na Figura 6.14 é mostrado o desempenho do equalizador bayesiano com ordem do *link* direto $m = 4$ e ordem do *link* reverso $n = m + d - 2$, para valores de frequência Doppler de $fd = 20$ Hz, $fd = 40$ Hz, $fd = 60$ Hz e $fd = 80$ Hz. O ajuste dos estados

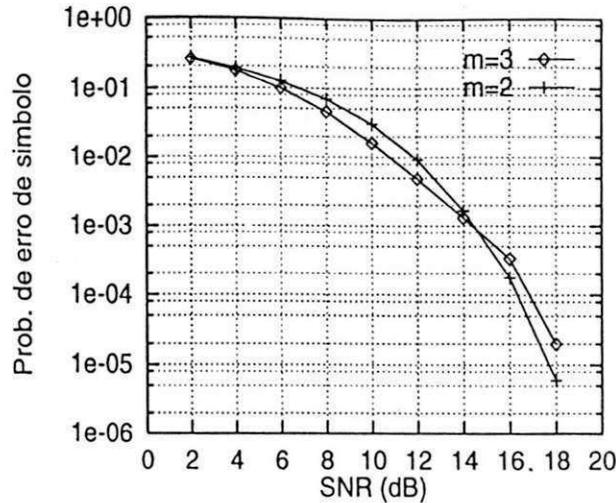


Figura 6.13: Desempenho do equalizador bayesiano com *feedback* em um canal variante.

do canal, neste caso, foi feito usando-se as Equações 6.7. Como se pode perceber, essa estrutura apresenta um desempenho que é pouco influenciado pela variação da frequência Doppler, para esses valores de frequência Doppler.

Sem dúvidas é desejável que uma estrutura de detecção apresente um desempenho que seja pouco afetado pelo deslocamento Doppler. Essa característica faz com que seja possível o uso desse tipo de equalizador em ambientes nos quais a unidade móvel se aproxima com uma certa velocidade da fonte transmissora. Por outro lado, pode-se perceber claramente na Figura 6.15 que o equalizador sem *feedback*, e com ordem direta $m = 2$ apresenta maior sensibilidade à variação da frequência Doppler que o equalizador com ordem $m = 4$ e com *feedback*.

Neste segundo caso uma estimativa do canal foi mantida constante durante uma certa quantidade de símbolos transmitidos pela fonte. Durante, por exemplo, 100 símbolos da fonte a estimativa do canal foi mantida constante para uma Doppler de 12 Hz. Como se pode perceber na Figura 6.14, as curvas de taxa de erro permanecem quase que sobrepostas umas às outras à medida que a SNR varia, enquanto que na Figura 6.15 as curvas de taxa de erro se diferenciam bastante. Uma justificativa para esse resultado pode ser dada com base na distribuição geométrica dos estados do canal, organizados de acordo com os conjuntos 5.18 e 5.26. Essa distribuição dos estados do

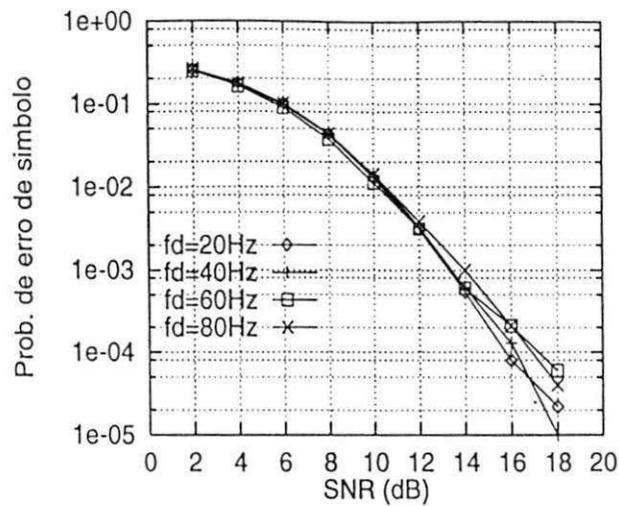


Figura 6.14: Desempenho do equalizador bayesiano com *feedback decision* em um modelo de canal variante e com parâmetros $m = 4$, $n_a = 2$ e $d = 1$.

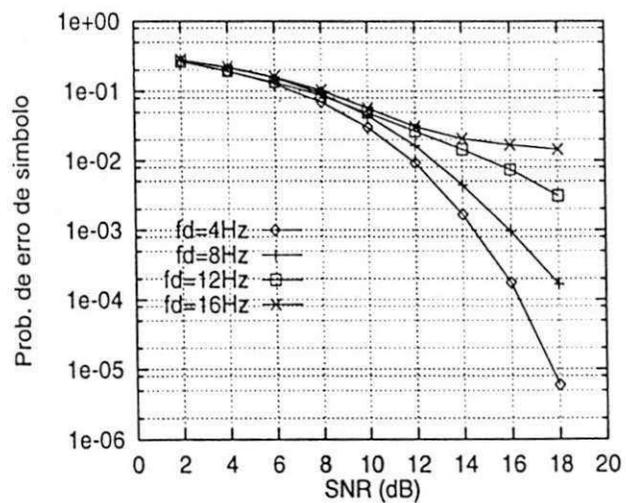


Figura 6.15: Desempenho do equalizador bayesiano sem *feedback decision* em um modelo de canal variante e com parâmetros $m = 2$, $n_a = 2$ e $d = 1$.

canal é que define as regiões de decisão do equalizador que podem ser ilustradas na Figura 6.16, para uma estrutura sem *feedback* e com parâmetros $n_a = 3$, $d = 0$, $m = 1$ e $n = 2$. O parâmetro $m = 1$ foi utilizado nesse exemplo, para permitir a visualização dessas regiões.

$$H(z) = (0.7409 - j0.7406) + (-0.29648 + j0.88884)z^{-1} + (-0.022191 - 0.155562)z^{-2}. \quad (6.9)$$

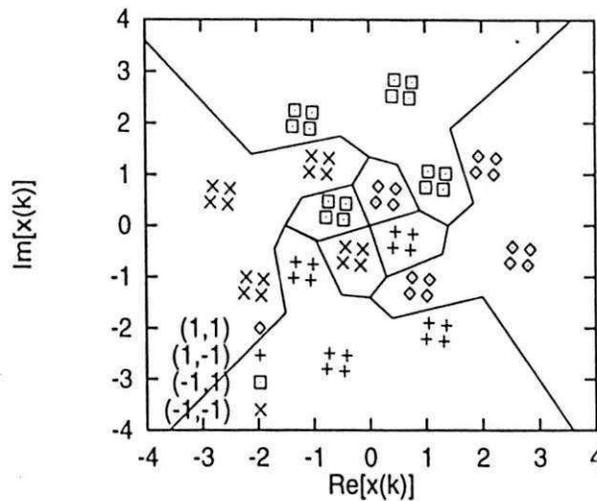


Figura 6.16: Distribuição dos estados do canal nas regiões de decisão sem *feedback* para o modelo de canal $H(z) = (0.7409 - j0.7406) + (-0.29648 + j0.88884)z^{-1} + (-0.022191 - 0.155562)z^{-2}$.

Como pode ser visto na Figura 6.16, as quatro regiões de decisão representadas pelos símbolos \times , \square , \diamond e $+$ estão muito próximas, principalmente na região central da Figura 6.16. Essa proximidade é que contribui para a ocorrência dos erros de decisão do equalizador, principalmente nos modelos de canal com frequência Doppler elevada. Nesses casos os vetores de símbolos de entrada passados para as unidades intermediárias da rede NRBF sofrem um considerável desvio de fase, fazendo com que a variável de decisão bayesiana do equalizador classifique erroneamente o símbolo estimado. Por outro lado, quando o *feedback* é utilizado, as regiões de decisão passam a ser bem mais simples, mais separadas e bem mais claras para o equalizador. Como se pode

perceber na Figura 6.17, as novas regiões de decisão marcadas pelos símbolos \times , \diamond , \square e $+$ são bem mais separadas e desse modo o equalizador toma sua decisão de forma mais precisa.

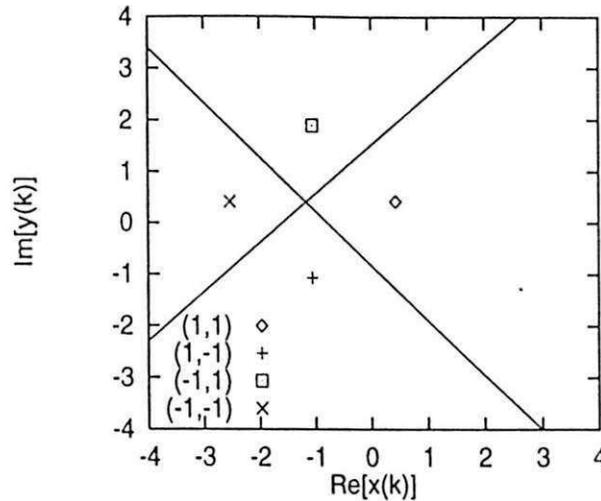
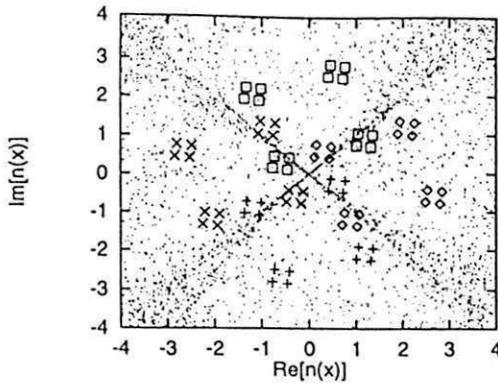


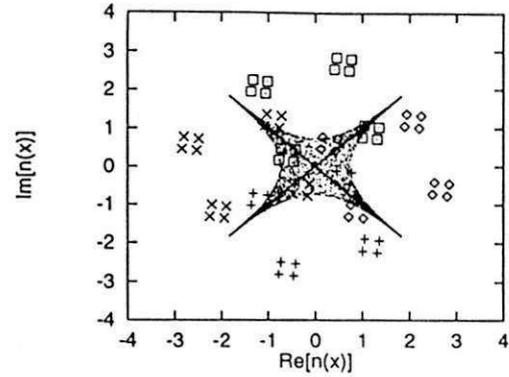
Figura 6.17: Distribuição dos estados do canal nas regiões de decisão com *feedback* e considerando-se um vetor de *feedback* $\hat{s}_f(k) = [1 + j \ 1 + j]^T$ para o modelo de canal conhecido $H(z) = (0.7409 - j0.7406) + (-0.29648 + j0.88884)z^{-1} + (-0.022191 - j0.155562)z^{-2}$.

Um outro resultado interessante, para o equalizador sem *feedback*, é a distribuição dos valores da variável de decisão final, formada pela soma das outras quatro variáveis, em torno dos estados do canal e, conseqüentemente, em torno das quatro regiões de decisão, em função da relação sinal ruído. Na Figura 6.18(b), para uma SNR=2 dB, pode-se perceber que os valores da variável de decisão final se distribuem muito espaçadamente, dificultando bastante a tarefa de classificação do detetor, enquanto que na Figura 6.18(a) percebe-se uma distribuição dos valores de decisão mais concentrados em torno das regiões centrais de decisão.

Estas duas curvas foram obtidas sorteando-se aleatoriamente 10000 estados entre os 64 possíveis estados obtidos para o canal dado pela Equação 6.9 e passando cada estado sorteado pelo detetor. Como se pode perceber, há uma tendência de os valores de decisão final se concentrarem mais em torno das linhas medianas dos dois gráficos e se distribuírem de maneira quase uniforme entre as quatro regiões de decisão.



(a) Distribuição para SNR=2 dB



(b) Distribuição para SNR=15 dB

Figura 6.18: Distribuição dos valores da variável de decisão final em torno dos estados do canal $H(z) = (0.7409 - j0.7406) + (-0.29648 + j0.88884)z^{-1} + (-0.022191 - j0.155562)z^{-2}$.

6.4 Conclusão

Neste capítulo foi analisado o procedimento utilizado na simulação do sistema estudado, desde o canal até o detetor. Na simulação do canal os coeficientes de cada linha de atraso foram gerados independentes uns dos outros de acordo com a Equação 6.1.

Como se pôde perceber nas curvas de taxa de erro obtidas, o desempenho das estruturas implementadas melhorou pouco à medida que a ordem direta do detetor aumentou. Ao mesmo tempo, aumentou também o número de centros-estado do canal, o número de centros da camada intermediária da rede NRBF e conseqüentemente a complexidade computacional final. Como conseqüência desse aumento da complexidade, o detetor bayesiano sem *feedback* passa ser pouco apropriado de ser implementado em sistemas com modulação MPSK em que $M > 4$, pois o número de estados é função do número de coeficientes n_a do canal, do atraso de decisão d e da ordem da linha direta do equalizador m , ou seja

$$n_s = M^{n_a+m-2}. \quad (6.10)$$

Para o caso variante no tempo, o desempenho do detetor foi influenciado pela taxa de variação dos coeficientes do modelo do canal. Como pode ser visto nas Figuras 6.16 e 6.17, o detetor sem *feedback* apresenta uma maior sensibilidade à variação dos coeficientes do modelo do canal em função da frequência Doppler. Por outro lado, o detetor com *feedback* apresenta-se mais insensível às variações de frequência Doppler mais rápidas.

Pôde-se perceber que a estrutura com *feedback* foi mais insensível ao aumento de f_d do que a estrutura sem *feedback*. As curvas de taxa de erro de símbolo apresentaram uma acentuada diferença, no caso do equalizador sem *feedback*, à medida que f_d ia aumentando, enquanto que na estrutura com *feedback* as curvas se mantiveram praticamente sobrepostas.

Como foi mostrado na Equação 5.27, as variáveis de decisão com o *link* de *feedback* realizam a soma dos valores de saída de apenas $n_{s,j}^{(i)}$ unidades intermediárias, ao invés de somar sobre as $n_s^{(i)}$ unidades intermediárias da estrutura sem o *feedback*. No caso de uma estrutura de equalização com atraso de decisão $d = 1$ e com um esquema de modulação QPSK o número $n_{s,j}^{(i)}$ de unidades a serem somadas passa a ser constante e igual a $4^{d+1} = 16$. Há portanto, uma considerável redução de complexidade em relação à estrutura sem *feedback*. Embora não tenha sido implementada nenhuma rotina que contasse o número de multiplicações, o número de adições e de exponenciais é de se esperar que tenha havido uma diminuição no número de operações realizadas, já que com *feedback* o tempo de processamento foi menor.

Um outro fator a que se atribui a melhora dos detetores com *feedback* é a diminuição da complexidade das regiões de decisão da rede NRBF. Essas regiões são formadas pela distribuição geométrica dos estados de valor complexo do canal ao longo dos eixos real e imaginário, de acordo com os conjuntos mostrados nas Equações 5.18 e 5.26. Nas regiões de decisão sem *feedback* é mais provável ocorrerem valores das variáveis de decisão final em cima das linhas que delimitam essas regiões, fazendo com que hajam os erro de classificação.

Quanto ao treinamento da rede utilizada para implementar as variáveis de decisão, foi visto que os únicos parâmetros que precisaram de reajuste foram os centros dos nós da camada intermediária, a cada vez que a característica do canal mudava.

Capítulo 7

Conclusão

Neste capítulo são apresentadas conclusões gerais do trabalho, obtidas a partir daquelas já apresentadas nos outros capítulos anteriores. Ao final serão analisadas as contribuições fornecidas e algumas idéias que podem ser utilizadas para se tentar melhorar o desempenho das estruturas implementadas.

Como as simulações foram feitas a partir do canal, é mais coerente que começar a tecer conclusões a partir do segundo capítulo. Para o modelo de canal utilizado, foram feitas as devidas considerações e simplificações que permitiram que apenas o problema de detecção símbolo-a-símbolo na presença de interferência intersimbólica e ruído gaussiano fosse solucionado. Foi feita uma pesquisa em vários artigos e livros até se chegar a um modelo apropriado. Nas referências [1], [3] e [18], por exemplo, há diferentes formas de simular o canal com diferentes efeitos.

No capítulo 3 foi dada uma introdução às redes neurais artificiais e foram mostradas as principais estruturas de classificadores. Cada tipo de classificador é mais apropriado a uma determinada aplicação do que a outra e nesse ponto é preciso mais uma vez se fazer uma análise de qual estrutura é mais apropriada ao problema estudado. Nesse trabalho foi usada uma rede com classificador de núcleo. Essa escolha foi feita devido à sua similaridade com as variáveis de decisão bayesianas. Seria pouco apropriado resolver um problema de detecção bayesiano em um canal com ruído gaussiano aditivo usando, por exemplo, uma rede neural MLP. Haveria desvantagens em relação ao tempo de treinamento, tamanho da rede e outros fatores que implicariam em uma maior

complexidade do sistema final. Em sistemas que usam redes neurais em tarefas como reconhecimento de caracteres manuscritos, por exemplo, a rede neural é treinada com todas as possíveis formas de escrita de um determinado caracter. A menos que surja uma outra forma muito diferente de se escrever o número "2", por exemplo, a rede ficará treinada com todos os padrões do número "2" que foram passados durante a sua fase de treinamento. Nesse caso, se a fase de treinamento durasse 3 dias não haveria o menor problema. Já em relação a um problema de detecção, seria totalmente impensável se o tempo de estimação dos pesos do canal durasse, por exemplo, um minuto! As redes NRBF utilizadas têm uma estrutura bem mais simples, geralmente são formadas apenas por três camadas e têm seus parâmetros, em aplicações como a apresentada neste trabalho, facilmente ajustados. Sem dúvida que a rede MLP encontra aplicações em muitos problemas de equalização de canais, como pode ser visto nas referências [25], [33] e [34] e dependendo do critério de desempenho utilizado, um equalizador com a rede MLP pode ser mais eficiente que um equalizador com a rede NRBF.

Como pôde ser visto nas curvas de taxa de erro de símbolos, o desempenho dos dois tipos de estruturas implementadas (com e sem *feedback*) depende dos parâmetros m , d , n e n_a . No caso da estrutura implementada sem *feedback*, à medida que a ordem direta m aumenta, a taxa de erro de símbolos pouco diminui e a complexidade também aumenta. Na estrutura com *feedback*, entretanto, o número de estados aumenta com m , mas a complexidade computacional permanece praticamente a mesma. Na estrutura da rede NRBF o maior esforço ocorre no cálculo das variáveis de decisão, mas com o *feedback* esse esforço permanece praticamente constante, no caso de um atraso de decisão $d = 1$. Foi possível também perceber que as estruturas com *feedback* são mais estáveis e são menos afetadas pela frequência Doppler. Isso se deve com certeza ao melhor delineamento das regiões de decisão da estrutura com *feedback* que impede o surgimento de regiões de interseção e desse modo os erros de classificação dos padrões de entrada, que são os símbolos da constelação QPSK, diminuem. Em ambas as estruturas, com e sem *feedback* d é escolhido no máximo igual a 2 para diminuir o erro de rastreamento decorrente do uso das Equações 6.7.

7.1 Contribuições e Sugestões de Continuidade

Como dito antes, o objetivo desse trabalho foi o estudo da aplicação da teoria de detecção bayesiana a modelos de canais com algumas das características que mais influenciam na má qualidade das transmissões. Embora ainda se possa fazer modificações na tentativa de melhorar alguns resultados, pode-se dizer que os objetivos almejados foram alcançados. Como contribuições do trabalho pode-se citar as seguintes:

- Análise do desempenho das estruturas implementadas em função da frequência Doppler.
- Estudo comparativo entre os dois tipos de estruturas de detetor bayesiano: com *feedback* e sem *feedback*
- Estudo comparativo das duas estruturas em função da frequência Doppler.
- Contribuição para a formação, no departamento, de uma nova linha de pesquisa, que é a detecção com rede neural NRBF.

As sugestões para dar continuidade a este trabalho seguem de acordo com alguns resultados que podem ser melhorados. Um estudo a ser feito seria a utilização de rotação dos símbolos transmitidos. Essa rotação propiciaria uma forma de diversidade que poderia ser usada para melhorar as regiões de decisão da rede. Foi feita uma tentativa de rotação da constelação QPSK em um ângulo de 22.5 graus [35], mas a taxa de erro de símbolos não melhorou em relação à taxa de erro obtida sem rotação. Seria necessário testar outros ângulos de rotação para ver se poderia haver algum ganho com esse acréscimo. Este trabalho poderia ainda ser analisado no contexto do CDMA. Nesse caso seria feita uma análise para avaliar a aplicação das estruturas implementadas no sistema CDMA. Enfim, a utilização conjunta dessas estruturas com outras formas de diversidade e outros códigos teria que ser testada para que se pudesse verificar a relevância dessas complementações. Em resumo poder-se-ia

- Analisar o efeito da rotação dos símbolos transmitidos;
- Transmitir os componentes de fase e quadratura alternadamente;

- Utilizar as estruturas estudadas juntamente com códigos de correção de erro;
- Estudar uma outra forma de treinar a rede.

Bibliografia

- [1] C. Jeruchim, K. Sam Shanmugan and Philip Balaban. *Simulation of Communication Systems*. Plenum Press, New York and London, 1992.
- [2] Hirofumi and Suzuki. A Statistical Modell for Urban Radio Propagation. *IEEE Transactions on Communications*, 25(7):673–680, 1997.
- [3] John G. Proakis. *Digital Communications*. Mc Gra-Hill, Inc, 1995.
- [4] Bernard Sklar. Rayleigh Fading Channels in Mobile Digital Communication Systems Part I: Characterization. *IEEE Communications Magazine*, pages 90–100, July 1997.
- [5] Marcelo Sampaio de Alencar. *Comunicações Móveis Celulares*. Departamento de Engenharia Elétrica, UFPB, 1997.
- [6] A. G. Guimarães, C. J. A. da Silva, J. Ferreira Galdino e E. Leite Pinto. Comparação de Desempenho de Simuladores de Canais com Desvanecimento Rápido Parte I- Avaliação Numérica. *Simpósio Brasileiro de Telecomunicações*, pages 426–430, 1997.
- [7] Simon Haykin. *Adaptive Filter Theory*. Prentice Hall, 1996.
- [8] R. P. Lippmann. An Introduction to Computing with Neural Nets. *IEEE ASSP Magazine*, April 1987.
- [9] Richard O. Duda and Peter E. Hart. *Pattern Classification and Scene Analysis*. John Wiley & Sons, 1973.

- [10] Tomaso Poggio and Federico Girosi. Networks for Approximation and Learning. *Proceedings of IEEE*, 78(9):1481–1497, September 1990.
- [11] Mark J. L. Orr. *Introduction to Radial Basis Functions Networks*. Centre for Cognitive Science, University of Edinburgh Buccleuch Place, Edinburgh EH8 9LW, Scotland, 1996.
- [12] I. Cha and S. A. Kassam. Channel equalization using adaptive complex radial basis function networks. *IEEE J. Select. Areas Commun.*, 13:122–131, 1995.
- [13] S. McLaughlin, S. Chen and B. Mulgrew. Complex-valued radial basis function network, Part II: Application to digital communications channel equalization. *IEEE Signal Processing*, 36:175–188, 1994.
- [14] Bernard Mulgrew. Applying Radial Basis Functions. *IEEE Signal Processing Magazine*, pages 50–64, 1996.
- [15] H. Vincent Poor. *An Introduction to Signal Detection and Estimation*. Springer-Verlag, 1988.
- [16] Stephen McLaughlin, Sheng Chen and Bernard Mulgrew. Adaptive Bayesian Equalizer with Decision Feedback. *IEEE Transactions on Signal Processing*, 41:2918–2927, 1993.
- [17] A. Bernardini and D. S. Fina. Optimal Decision Boundaries for M-QAM Signal Formats Using Neural Classifiers. *IEEE Transactions on Neural Networks*, 9:241–246, 1998.
- [18] M. D. Yacoub. *Foundations of Mobile Radio Engineering*. CRC Press, Inc, 1993.
- [19] Sheng Chen, Stephen McLaughlin, Bernard Mulgrew and Peter M. Grant. Adaptive Bayesian Decision Feedback Equalizer for Dispersive Mobile Radio Channels. *IEEE Transactions on Communications*, 43(5):1937–1946, May 1995.
- [20] M. S. Yee and L. Hanzo. Multi-level Radial Basis Function Network Based Equalizers for Rayleigh Channel. *Dept. of Electronics and Computer Science, University of Southampton, UK.*, 1998.

- [21] Leon H. Sibul. *Adaptive Signal Processing*. IEEE PRESS, 1987.
- [22] Sirikiat Ariyavisitakul and Larry J. Greenstein. Reduced-Complexity Equalization Techniques for Broadband Wireless Channels. *IEEE Journal on Selected areas in Communications*, 15(1):5–15, 1997.
- [23] Ali Masoomzadeh-Fard and Subbarayan Pasupathy. Nonlinear Equalization of Multipath Fading Channels with Noncoherent Demodulation. *IEEE Journal on Selected areas in Communications*, 14(3):512–520, 1996.
- [24] Amir Hussain, John J. Soraghan and Tariq S. Durrans. A New Adaptive Functional-Link Neural-Net Based DFE for Overcoming Co-Channel Interference. *IEEE Transactions Communications*, 45(11):1358–1362, 1997.
- [25] A. Hussain, J. J. Sorogahan and T. S. Durrani. A new adaptive functional-link neural-network-based DFE for overcoming co-channel interference. *IEEE Transactions Communications*, 45:1358–1362, 1997.
- [26] M. Abdulrahman, A. U. H. Sheikh and D. D. Falconer. Decision feedback equalization for CDMA in indoor wireless communications. *IEEE J. Selected Areas Communications*, 12:698–706, 1994.
- [27] B. Aazhang, B. P. Paris and G. C. Orsak. Neural networks for multiuser detection in code-division multiple-access channels. *IEEE Transactions Communication*, 40:1212–1222, 1992.
- [28] Kaushik Das and Salvatore D. Morgera. Adaptive interference cancellation for DS-CDMA systems using neural network techniques. *IEEE Journal on Selected Areas in Communications*, 16(9):1774–1784, 1998.
- [29] S. Chen, G. J. Gibson, C. F. N. Cowan and P. M. Grant. Adaptive equalization of finite nonlinear channels using multilayer perceptron. *Signal Processing*, 10:107–119, 1990.
- [30] G. Kechriotis, E. Zervas and E. S. Manolakos. Using recurrent neural networks for adaptive channel equalization. *IEEE Trans. Neural Networks*, 5:267–278, 1994.

- [31] José Homero Feitosa Cavalcanti. Sistemas Inteligentes, Redes Neurais Artificiais e Lógica Fuzzy. *Publicação interna - Departamento de Sistemas e Computação*, Maio 1999.
- [32] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Macmillan publishing company, 1994.
- [33] Khalid A. Al-Mashouq King Suad University and Irving Reed University of Southern California. The Use of Neural Nets to Combine Equalization with Decoding. *ETSI Telecomunicación-UV,47011*, 1996.
- [34] G. J. Gibson, S. Siu and C. F. N. Cowan. The application of nonlinear structures to the reconstruction of binary signals. *IEEE Trans. Signal Processing*, 39:1877–1884, 1991.
- [35] Joseph Boutros and Emanuele Viterbo. Signal Space Diversity: A Power-and-Bandwidth-Efficient Diversity Technique for the Rayleigh Fading Channel. *IEEE Transactions on Information Theory*, 44(4):1453–1467, 1998.