



UNIVERSIDADE ESTADUAL PAULISTA

Pós-Graduação em Engenharia Mecânica

***IDENTIFICAÇÃO DE PLANTAS
NÃO LINEARES VIA WAVENETS***

EDISON RIGHETO

Ilha Solteira - SP



**IDENTIFICAÇÃO DE PLANTAS
NÃO LINEARES VIA WAVENETS**

1210001421



Proc. 053/2003-NPD 163

UNESP - "CAMPUS DE ILHA SOLTEIRA"	
SERVIÇO TÉCN. DE BIBLIOTECA E DOCUMENTAÇÃO	
DATA DE CHEGADA	DATA DE TOMBO
02.09.03	30.09.03
REGISTRADO POR	TOMBO
<i>Ailza</i>	<i>Te. 1421</i>
AQUISIÇÃO	CLASSIFICAÇÃO
<i>Ilha Solteira</i> <i>Anta</i> <i>R\$ 10,00</i>	<i>R571L</i>

Co Sys 215 708
Sys 56144

ILHA SOLTEIRA - SP

30500001

BCPLS - FEIS - UNESP



UNIVERSIDADE ESTADUAL PAULISTA

FACULDADE DE ENGENHARIA DE ILHA SOLTEIRA

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA

***IDENTIFICAÇÃO DE PLANTAS
NÃO LINEARES VIA WAVENETS***

EDISON RIGHETO

Dissertação apresentada à Faculdade de Engenharia de Ilha Solteira da Universidade Estadual Paulista “Júlio de Mesquita Filho”, como parte dos requisitos exigidos para a obtenção do título de **Mestre em Engenharia Mecânica**.

Orientador: Prof. Dr. João Antonio Pereira

Co-orientador: Prof. Dr. Vicente Lopes Junior

ILHA SOLTEIRA, FEVEREIRO DE 2003

BCpIS - FEIS - UNESP



FICHA CATALOGRÁFICA

Elaborada pela Seção Técnica de Aquisição e Tratamento da Informação/Serviço Técnico de Biblioteca e Documentação da FEIS/UNESP

R571i Righeto, Edison
Identificação de plantas não lineares via wavenets / Edison Righeto.--
Ilha Solteira : [s.n.], 2003
v, 84 p. : il.

Dissertação (mestrado) – Universidade Estadual Paulista. Faculdade de Engenharia de Ilha Solteira, 2003

Orientador: João Antonio Pereira
Co-Orientador: Vicente Lopes Junior
Bibliografia: p. 80-82

1. Processamento de sinais. ✓

IDENTIFICAÇÃO DE PLANTAS NÃO LINEARES
VIA WAVENETS

Edison Righeto

ESTA DISSERTAÇÃO FOI JULGADA ADEQUADA PARA OBTENÇÃO
DO TÍTULO DE

MESTRE EM ENGENHARIA MECÂNICA

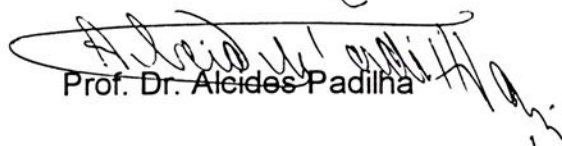
NA ÁREA DE CONCENTRAÇÃO: MECÂNICA DOS SÓLIDOS E
APROVADA EM SUA FORMA FINAL PELO PROGRAMA DE PÓS-
GRADUAÇÃO EM ENGENHARIA MECÂNICA


Prof. Dr. Sérgio Said Mansur
Membro do Conselho
Respondendo pela Coordenação

COMISSÃO EXAMINADORA:


Prof. Dr. João Antonio Pereira - Orientador


Prof. Dr. João Batista Aparecido


Prof. Dr. Alcides Padilha

Ilha Solteira – SP, fevereiro de 2003

AGRADECIMENTOS

Enfaticamente agradeço ao João Antonio Pereira e Vicente Lopes Junior do DEM – UNESP – Ilha Solteira. Ambos exerceram um forte espírito crítico e isto sempre enriquece a vida de um feiticeiro aprendiz. Pelo mesmo motivo também agradeço ao Edvaldo Assunção (DEE) e Marcelo Minhoto de Carvalho Teixeira (DEE). De longa data José Roberto Sanches Mantovani (DEE) tem sido meu parceiro de intermináveis desafios intelectuais. Sinto-me honrado com a amizade e tantas oportunidades de aprender. Agradeço a Ana Julia e a Luzia porque me ensinaram a decifrar a magia nos olhos dos pardais.

Ao Deoclécio Mitsuiti Kosaka agradeço pelo esmero no trabalho datilográfico.

Muito de bom surgiu no decurso desta lida.

As amizades que este trabalho solidificou são os melhores resultados dele.

SUMÁRIO

Página

RESUMO	iii
ABSTRACT	iv
CAPÍTULO 1.....	1
1.1 INTRODUÇÃO	1
1.2 REVISÃO BIBLIOGRÁFICA	3
CAPÍTULO 2.....	5
2.1 FUNDAMENTOS DE REDES NEURAIIS	5
2.2 MODELAMENTO	6
2.3 O PERCEPTRON	7
2.4 PERCEPTRON MULTICAMADAS	8
2.4.1 ALGORITMO DE RETROPROPAGAÇÃO.....	11
CAPÍTULO 3.....	15
3.1 FUNDAMENTOS DA TEORIA DAS WAVELETS	15
3.2 TRANSFORMADA <i>WAVELET</i>	15
3.3 TIPOS DE FILTROS <i>WAVELETS</i>	18
3.3.1 <i>Wavelet</i> Morlet.....	18
3.3.2 <i>Wavelets</i> RASP	19
3.4 TEOREMA DE RESÍDUO	21
3.4.1 <i>WAVELET</i> SLOG	23
3.4.2 <i>WAVELET</i> POLYWOG	26
3.4.3 <i>WAVELET</i> DE SHANNON.....	32

CAPÍTULO 4	35
4.1 <i>WAVENETS</i>	35
4.2 ALGORITMO <i>WAVENET</i>	36
4.3 CONTROLADORES NEURO <i>WAVENET</i> AUTO-AJUSTÁVEIS	41
4.3.1 ESTRUTURA E DEFINIÇÃO	41
CAPÍTULO 5	46
5.1 RESULTADOS	46
5.2 IDENTIFICAÇÃO DE SISTEMAS NÃO LINEARES	47
5.2.1 CASO 1	47
5.2.2 CASO 2	57
5.2.3 CASO 3	64
5.3 CONTROLE ADAPTATIVO	73
CAPÍTULO 6	78
6.1 DISCUSSÃO	78
REFERÊNCIAS	80
APÊNDICE	83
TEOREMA DE SHANNON-WHITTAKER	83



RESUMO

Este trabalho aborda os princípios das redes neurais, os fundamentos das wavelets e como esses conceitos podem ser combinados para formar uma rede neural adaptativa, denominada *wavenet*. As funções de ativação da rede neste caso são definidas a partir de um conjunto de *wavelets* filhas. O escopo central da proposta é a adequação da estrutura deste novo tipo de rede em termos do número de neurônios, da escolha da função de ativação e dos coeficientes de ponderação para que a mesma seja capaz de identificar eficientemente plantas não lineares arbitrárias. O algoritmo *wavenet* consiste no processo de auto construção da rede a partir da minimização do erro quadrático médio (distância na norma L_2 entre a saída da planta e a saída da rede) em relação aos parâmetros escala e deslocamento de cada *wavelet* utilizada na construção da rede. Os parâmetros são atualizados iterativamente através do método do gradiente descendente, algoritmo LMS, e a topologia da rede é preservada. A rede recruta gradualmente janelas *wavelets* as quais devem ser capazes de cobrir eficiente e suficientemente a região tempo-frequência ocupada pela saída da planta. A avaliação da proposta é feita utilizando exemplos simulados de plantas não lineares dadas por equações de diferenças. São discutidas diferentes estruturas de rede para identificação de plantas com não linearidade na dinâmica ou na entrada da planta, sendo que o desempenho do algoritmo é avaliado, quanto à escolha de diferentes famílias *wavelets* no papel da função ativadora bem como em relação ao número de janelas *wavelets* (neurônios) utilizados na rede. E finalmente, é apresentado um exemplo de controle, em que a planta previamente identificada é utilizada em uma proposta de controle auto-ajustado.

Palavras chaves: Redes Neurais, *Wavelets*, *Wavenets*, Identificação.



ABSTRACT

This work approaches the beginnings of the neural network, the foundations of the wavelets and as those concepts can be combined to form a adaptive neural network, named wavenet. The activation functions of the neural network in this case are defined starting from a collection of wavelets daughters. The main core of the proposal is the adaptation of the structure of this new network kind in terms of the number of neurons, of the choice of the activation function and of the consideration coefficients so that the same is capable to identify nonlinear plant efficiently. The algorithm wavenet consists of the process of auto construction of the network starting from the minimization of the mean quadratic error (distance in L2 norm between the plant output and the network output) in relation to the parameters of each wavelet used in the construction of the network. The parameters are updated through the the descending gradient method, algorithm LMS, and the network topology is preserved. The network gradually recruits hidden units to effectively and sufficiently cover the time-frequency region accupied by a given target(plant output). The evaluation of the proposal is made using simulate examples of nonlinear plants given by equations of differences. Different network structures are discussed for identification of nonlinear dynamic or nonlinear input system , and the acting of the algorithm is evaluated, with relationship to the choice of different families wavelets in the role of the function ativadora as well as in relation to the number of windows wavelets (neurons) used in the network. It is finally, a control example is presented, in that the plant previously identified it is used in a proposal of self-tuning control.

Keywords: Neural network, *Wavelets*, *Wavenets*, Identification.



CAPÍTULO 1

1.1 INTRODUÇÃO

O estudo e o desenvolvimento de técnicas de controle adaptativo tem despertado o interesse nas várias áreas das engenharias. A sua aplicação representa um avanço para o tratamento de sistemas mais complexos de controle. Variações nos parâmetros da planta ou nas suas condições de funcionamento podem ter impacto decisivo no desempenho e estabilidade e, portanto no sucesso da ação de controle em andamento. Essas são situações típicas em que o controle adaptativo pode atuar. Por esta razão, é desejável que o algoritmo de controle se adapte automaticamente a essas situações, tão logo sejam notadas as referidas variações. Este fato é da esfera do controle adaptativo. Embora não seja objeto deste trabalho detalhar projetos de controladores, é oportuno lembrar que eles devem considerar vários aspectos, como locação de polos, margem de ganho, resposta ao degrau, e no caso de plantas desconhecidas existe ainda a necessidade de um algoritmo de identificação do modelo da planta. Projetos de controladores devem definir as ações e também definir as avaliações sobre a eficiência das ações. Para controle adaptativo são de interesses os projetos que podem ser acompanhados instantaneamente, sem interrupção de qualquer natureza ou alegação. Estamos na esfera do controle adaptativo auto-ajustável, conceito que será discutido neste trabalho. O método de auto-ajuste é uma importante abordagem na construção de controladores adaptativos para sistemas desconhecidos, variando lentamente. A idéia é estimar os parâmetros de uma planta desconhecida e ao mesmo tempo atualizá-los com base nos sinais medidos, e os parâmetros

atualizados da planta são utilizados na entrada do controle para que ele produza a resposta desejada. Entretanto, técnicas usuais de controle adaptativo auto-ajustável podem apenas tratar de sistemas lineares ou com não-linearidades muito especiais. Habitualmente, essas técnicas assumem que o modelo de controle opera em uma região linear. Os parâmetros do modelo da planta linearizado são estimados recursivamente e usados para atualizar o controlador.

Em situações práticas, freqüentemente não é possível representar através da formulação convencional certas características de um sistema, tais como não-linearidade, atraso temporal, parâmetros que se alteram ligeiramente e outras complexidades. Neste caso, é importante desenvolver uma técnica eficaz na qual a estrutura de um modelo de planta não-linear possa ser identificado por um processo adaptativo. A utilização das redes neurais ativadas por *wavelets* tem se mostrado como alternativa promissora para tratar problemas dessa natureza.

Os próximos capítulos tratarão em separado e com certo nível de detalhes tanto o conceito de redes neurais como o de transformadas *wavelets*, necessários na definição das *wavenets*. No Capítulo 4 será discutido o algoritmo *wavenet*. O Capítulo 5 apresenta resultados de vários exemplos de aproximações de saída da planta pela saída da rede e uma discussão sobre as várias simulações propostas e desenvolvidas com tal fim. Finalmente, no Capítulo 6 delineamos sugestões para trabalhos futuros.



1.2 REVISÃO BIBLIOGRÁFICA

Com o desenvolvimento crescente das técnicas de processamento e análise de sinais, tais como redes neurais, *wavelets* e lógica nebulosa, algoritmos genéticos, projetos de controle adaptativo pôde se expandir significativamente. A literatura atual mostra o conceito de neuro-controladores aplicados em situações variadas, mostrando que eles conseguem tratar não linearidades desconhecidas (Narendra e Parthasarathy, 1990) (Levin e Narendra, 1995). Os autores em (Karakasoglu, Sudharsanan e Sundareshan, 1991) e (Sudharsanan e Sundareshan, 1991) montaram um neuro-controlador utilizando uma rede neural recorrente de três camadas, assumindo, entretanto, que o modelo da planta era acessível. Os autores em (Chen, 1990), (Psaltis e Yamamura, 1988) e (Li e Slotine, 1989) empregaram rede neural *feedforward* multicamadas no projeto de um controlador adaptativo auto-ajustável com razoável sucesso. Segundo os autores tal sucesso não foi maior porque a técnica de treinamento com retropropagação é complexa computacionalmente, exigindo quase sempre computação off-line na minimização do erro.

Em geral, as redes neurais com retropropagação multicamadas ou não, tem demonstrado em muitas aplicações uma surpreendente habilidade para aprender a partir de dados discretos. Um número de provas confiáveis já tem aparecido para explicar esta habilidade incomum das redes neurais com retropropagação no processo de aproximar mapas (Hornik, Stinchcombe e White, 1989). A literatura mostra mais recentemente a importância da combinação da teoria de transformada *wavelet* com a arquitetura de rede neural com retropropagação, surgindo daí uma nova rede mapeadora denominada *wavenet*. O conceito de *wavenet* é baseado na introdução de uma super-*wavelet*, definida a partir de uma combinação linear de *wavelets*-filhas, e tratada como nova *wavelet*. Uma *wavelet*-filha é uma versão dilatada e ou deslocada da *wavelet* original ou mãe.

As características da super-*wavelet* permitem moldá-la para que se torne adequada a um problema particular, processo que vai além da adaptação dos parâmetros de uma *wavelet* de perfil fixo. Nas aplicações em engenharia, *wavenets* tem se mostrado promissoras tanto na representação quanto na classificação de sinais. A literatura mostra contribuições na área de identificação de vozes (Kadambe e Srinivasan, 1994), imagem de radar (Jouny e Kanapathipillai, 1994), equalização de canais não lineares (Chang e Yeh, 1994), detecção de sinais caóticos

(Benton, Ganjoo, Lumetta, Spillman e Ring, 1996), espalhamento acústico (Telfer, Szu e Dobeck, 1994, 1995) e outros.

Diferente do *perceptron* multicamadas, uma rede global, *wavenet* é uma rede local em que a função de saída apresenta características bem localizadas, tanto no domínio do tempo quanto no da frequência (o princípio de Heisemberg define essa “boa localização”). Em uma rede global todos os seus pesos são alterados toda vez que na entrada surgem novos dados e assim eles são atualizados em cada passo do treinamento. Essa globalização na alteração dos pesos tende a obscurecer os detalhes da estrutura local, provocando uma baixa taxa de aprendizado e a possibilidade de tomarmos mínimos locais como globais.

Por outro lado, em uma rede local apenas um reduzido subconjunto de pesos são ativados em cada ponto no espaço de saída. O treinamento da rede em uma parte do espaço de entrada não produz alteração no aprendizado ocorrido em uma região mais distante. Assim, a velocidade de aprendizado da rede local é geralmente maior do que na rede global.

Uma classe de *frame wavelets* foi introduzida mostrando ser simples e adequada para tratar de processos adaptativos quando em combinação com redes neurais (Imhoff, Roeum e Rosiek, 1995), (Marar e Vasconcelos, 1996). As idéias para a construção das novas famílias *wavelets* brotaram a partir da utilização do teorema dos resíduos para funções analíticas, sigmóides, e hermiticidade. Estamos falando das *wavelets* RASP, (funções racionais com pólos de segunda ordem), das *wavelets* SLOG (superposições de funções logísticas) e das *wavelets* POLYWOG (funções polinomiais combinadas com janelas gaussianas).

Essas são as famílias *wavelets* que serão utilizadas no transcurso deste trabalho ativando redes neurais que se alteram e se adaptam procurando identificar uma planta desconhecida.



CAPÍTULO 2

2.1 FUNDAMENTOS DE REDES NEURAIS

Uma rede neural é uma interconexão de elementos de processamento elementar como filtros e escalas. Eles interagem através de caminhos com intensidade de conexão variável que quando convenientemente adaptados podem produzir uma ação global bastante complexa.

O funcionamento do cérebro humano inspirou o projeto das redes neurais artificiais, RNA. O cérebro é uma rede formada por bilhões de processadores elementares, *neurônios*, interligados por uma estrutura ramificada chamada *axônios* e *dentritos*. A meta de uma RNA é imitar o cérebro humano, ligando e interconectando os processadores elementares. A transmissão de informação entre os neurônios ocorre na região da conexão dos axônios e dentritos de um neurônio aos de outros e são denominadas sinapses. As sinapses biológicas são simuladas nas redes neurais pela função *peso*, proporcionando conexões de intensidade variável, neste caso, pretende-se que a RNA, tal como o cérebro humano, seja capaz de ao se deparar com um problema, pelo menos reconhecê-lo e classificá-lo.

As RNAs têm apresentado uma grande evolução nas últimas décadas e sua utilização cobre um vasto campo de aplicações, tais como *classificação de sinais de radar/sonar* (Haykin e Deng, 1991) (Gorman e Sejnowski, 1988), *conversão texto-voz-texto* (Sejnowski e Rosenberg, 1987), *tecnologia VLSI* (Mead, Arreguit e Lazzaro, 1991), *aplicação biomédica* (Donlin e Child,

1992), *controles de processos* (Hammerstrom, 1993), *reconhecimento de caracteres* (Newland, 1995).

Um certo contingente da comunidade científica tem conferido às redes neurais uma forte importância, e esta tem crescido quando se discute a ativação daquelas por certas *wavelets* especiais, fato que se pretende deixar suficientemente claro no desenvolvimento deste trabalho. Passo-a-passo, começa-se com as redes neurais.

2.2 MODELAMENTO

No esforço para modelar certas capacidades do cérebro, Warrem McCulloch e Walter Pitts (McCulloch e Pitts, 1943) conceberam um modelo simplificado de um neurônio biológico, consistindo de múltiplas entradas e uma saída com uma unidade de processamento central. A figura 2.1 mostra esquematicamente o modelo do neurônio descrito por

$$y = f \left(\sum_{i=1}^N w_i x_i - v_t \right) \quad (2.1)$$

em que:

- x_i Sinais entrada, $i = 1, 2, \dots, N$;
- w_i Pesos das sinapses;
- v_t limiar ou *bias*;
- $f(\cdot)$ função de ativação ou elemento processador;
- y sinal de saída do neurônio.

O papel do limiar v_t é proporcionar um disparo da função de ativação f . No modelo de McCulloch e Pitts, os nós ou neurônios não são capazes de se auto-ajustar ou adaptar seus pesos sinápticos no decorrer do aprendizado. Posteriormente, Hebb (Hebb, 1949) sugeriu uma formulação matemática (ampliando e melhorando o modelo original) que podia produzir

mudanças nos pesos, proporcionalmente às atividades entre a pré e pós sinapses do neurônio. A expressão (2.2) mostra esta relação,

$$\Delta w_i = \mu y(n) x_i(n) \quad (2.2)$$

em que μ é uma constante positiva para todo n , denominada *taxa de aprendizado*. Ressalte-se que n está enumerando os dados de aprendizado da rede.

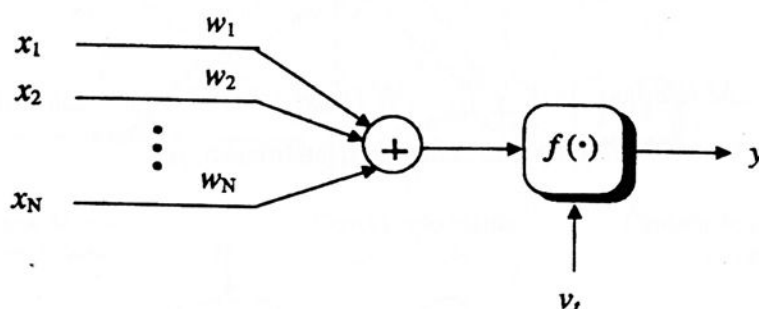


Figura 2.1 - O modelo McCulloch – Pitts de um neurônio

2.3 O PERCEPTRON

O perceptron é uma conexão de um único nível de neurônios McCulloch-Pitts também chamado de rede *feedforward* de camada única. Rosenblatt (Rosenblatt, 1957) demonstrou algumas aplicações práticas utilizando o perceptron. A rede é capaz de separar linearmente os vetores de entrada em classes, por um hiperplano. A memória associativa linear é um exemplo de uma RNA de camada única. Neste caso, a rede associa um padrão de saída (um vetor) a um padrão de entrada (também vetor), e a informação é armazenada na rede através das modificações de seus pesos sinápticos. A equação (2.3) descreve o comportamento do perceptron ilustrado na figura (2.2):

$$y_i = f\left(\sum_{j=1}^N w_{ij} x_j - v_i\right) \quad (2.3)$$

em que $i=1, 2, \dots, M$ (nós de saída), $j=1, 2, \dots, N$ (entrada),

w_{ij} = peso da sinapse j associada ao neurônio i .

Rosenblatt propôs uma regra de aprendizado com base no ajuste dos pesos, proporcionalmente às diferenças (erros) entre as saídas dos neurônios e as saídas desejadas. As adaptações de pesos são calculadas pela expressão (2.4):

$$\Delta w_{ij}(n) = \mu [d_i(n) - y_i(n)] x_j(n) \quad (2.4)$$

em que $i=1, 2, \dots, M$ (saídas), $j=1, 2, \dots, N$ (entradas), $d_i(n)$ é a saída desejada no nó i para a n -ésima apresentação dos dados de treinamento, n .

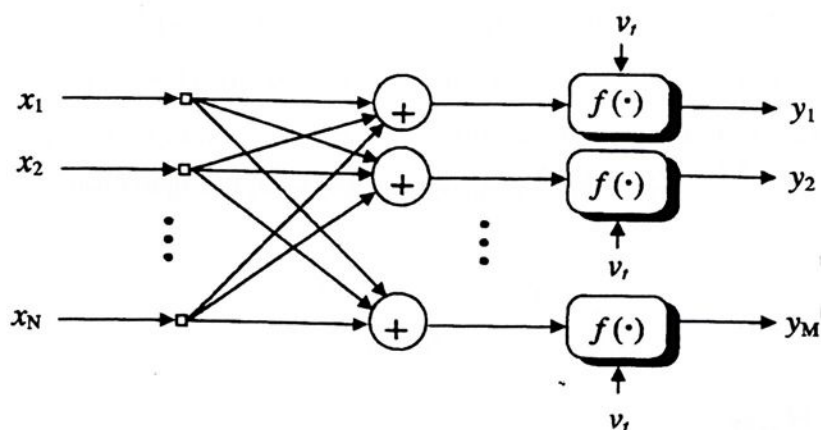


Figura 2.2 - Perceptron de única camada

2.4 PERCEPTRON MULTICAMADAS

Minsky e Papert (Minsky e Papert, 1969) demonstraram que o perceptron de camada única era limitado em sua capacidade computacional. Eles introduziram o perceptron multicamadas com uma ou mais camadas ocultas cujos nós computacionais são chamados neurônios escondidos. Tais neurônios escondidos intervêm entre a entrada e a saída da rede. A figura 2.3 mostra um perceptron *feedforward* de três camadas (entrada, a intermediária e a saída).

Os nós da camada de entrada da rede consiste de N elementos que representam a entrada para os K neurônios na segunda camada (camada escondida). A saída com M neurônios formam a camada final e constituem a resposta da rede ao estímulo oriundo dos nós fontes. Deve-se ter cuidado na escolha apropriada do número de nós da camada oculta para otimizar o desempenho do perceptron.

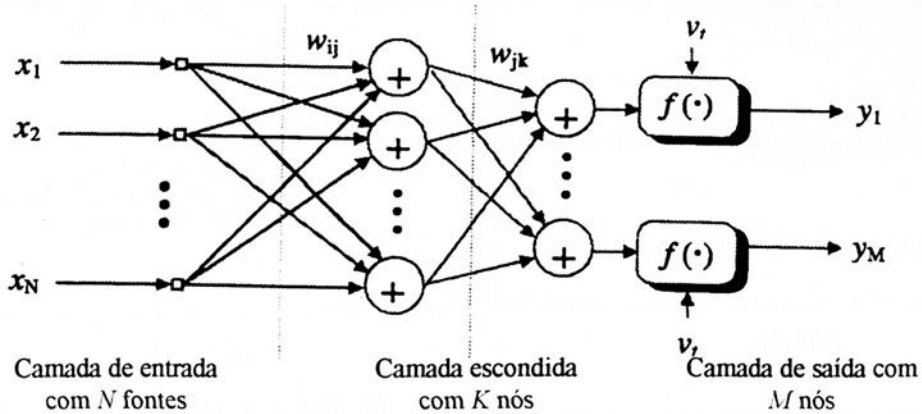


Figura 2.3 - Perceptron de três camadas

A arquitetura das redes neurais descritas até aqui empregam a função de ativação $f(\cdot)$ condicionando a saída de um neurônio ao nível de atividade em sua entrada. A tabela 2.1 resume os tipos básicos de funções de ativações. A sigmóide e a tangente hiperbólica são as funções mais utilizadas porque são diferenciáveis.

Tabela 2.1 - Funções de ativações.

NOMES	DEFINIÇÃO
Linear	$f(x)=kx$
Degrau	$f(x)=\begin{cases} \beta & \text{se } x \geq x_k \\ \delta & \text{se } x < x_k \end{cases}$
Degrau (usualmente: $\beta=1, \delta=0, x_k=0$)	$f(x)=\begin{cases} \beta & \text{se } x \geq x_k \\ \delta & \text{se } x < x_k \end{cases}$
Rampa	$f(x)=\begin{cases} \rho & \text{se } x \geq \rho, \quad \rho > 0 \\ x & \text{se } x < \rho \\ -\rho & \text{se } x \leq -\rho \end{cases}$
Sigmoide	$f(x)=\frac{1}{1+e^{-\alpha x}}, \alpha > 0$
Tangente hiperbólica	$f(x)=\tanh(\gamma x)=\frac{1-e^{-2\gamma x}}{1+e^{-2\gamma x}}, \gamma > 0$
Racional	$f(x)=\begin{cases} \frac{x^2}{1+x^2} & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases}$
Gaussiana	$f(x)=\frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$

2.4.1 ALGORITMO DE RETROPROPAGAÇÃO

Um dos métodos de aprendizado mais utilizado no treinamento de perceptrons multicamadas é o algoritmo de retropropagação desenvolvido por Rumelhart Hinton e Willians (Hinton e Willians, 1986). O aprendizado é feito com base no processo de minimização e emprega-se o método do gradiente descendente para minimizar o erro quadrático médio (norma L_2).

O erro na saída do neurônio j , na n -ésima iteração (isto é, a n -ésima apresentação dos dados de treinamento) é definido por

$$e_j(n) = d_j(n) - y_j(n) \quad (2.5)$$

em que $d_j(n)$ é a saída desejada no nó j após a n -ésima apresentação dos dados para treinamento e $y_j(n)$ é a saída efetiva no mesmo nó e na mesma ocasião.

O valor instantâneo da energia associada ao erro para o neurônio j é expresso como $\frac{1}{2} e_j^2(n)$, o valor instantâneo da energia total associada ao erro, $E(n)$ (distância na norma L_2), é definido como

$$E(n) = \frac{1}{2} \sum_{j \in C} e_j^2(n) \quad (2.6)$$

em que o conjunto C indexa os neurônios na camada de saída da rede, os únicos cujos erros são calculados diretamente.

Seja

$$v_j(n) = v_j^\ell(n) = \sum_{i=1}^N w_{ji}^\ell(n) y_i^{\ell-1}(n) \quad (2.7)$$

o potencial de ativação, isto é, a entrada da função de ativação associada ao neurônio j (na camada ℓ) em que

$y_j(n) = y_j^{\ell-1}(n)$ é o sinal do neurônio j na camada anterior $\ell - 1$ na n -ésima iteração,

$w_{ji}(n) = w_{ji}^{\ell}(n)$ é o peso do neurônio i na camada ℓ , estimulado pelo neurônio j na camada $\ell - 1$.

O sinal $y_j(n)$ surgindo na saída do neurônio j na n -ésima iteração é

$$y_j(n) = f_j(v_j(n)) \quad (2.8)$$

O algoritmo de retro-propagação aplica uma correção $\Delta w_{ji}(n)$ ao peso sináptico $w_{ji}(n)$, proporcional a $\frac{\partial E(n)}{\partial w_{ji}(n)}$. Em consequência da regra da cadeia, pode-se escrever

$$\frac{\partial E(n)}{\partial w_{ji}(n)} = \frac{\partial E(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)} \quad (2.9)$$

Tem-se que $\frac{\partial E(n)}{\partial w_{ji}(n)}$ representa um *fator de sensibilidade*, indicando direção de busca para os pesos sinápticos w_{ji} .

Da equação (2.6) obtém-se

$$\frac{\partial E(n)}{\partial e_j(n)} = e_j(n) \quad (2.10)$$

Da equação (2.5) obtém-se

$$\frac{\partial e_j(n)}{\partial y_j(n)} = -1 \quad (2.11)$$

Da equação (2.8) obtém-se

$$\frac{\partial y_j(n)}{\partial v_j(n)} = f'_j(v_j(n)) \quad (2.12)$$

Da equação (2.7) obtem-se

$$\frac{\partial v_j(n)}{\partial w_{ji}(n)} = y_i(n) \quad (2.13)$$

Retornando para a equação (2.9) obtem-se

$$\frac{\partial E(n)}{\partial w_{ji}(n)} = -e_j(n) f'_j(v_j(n)) y_i(n) \quad (2.14)$$

A correção $\Delta w_{ji}(n)$ aplicada à $w_{ji}(n)$ é definida pela *regra delta*, isto é:

$$\Delta w_{ji}(n) = -\eta \frac{\partial E(n)}{\partial w_{ji}(n)} \quad (2.15)$$

em que $\eta > 0$ é a taxa de aprendizado do algoritmo de retropropagação. O uso do sinal menos (-) na equação (2.15) garante um caminhar na direção do *gradiente descendente*, direção de redução do valor de $E(n)$. Em consequência das equações (2.14) e (2.15) tem-se

$$\Delta w_{ji}(n) = \eta \delta_j(n) y_i(n) \quad (2.16)$$

em que o *gradiente local* $\delta_j(n)$ é definido por

$$\begin{aligned} \delta_j(n) &= -\frac{\partial E(n)}{\partial v_j(n)} \\ &= -\frac{\partial E(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \\ &= e_j(n) f'_j(v_j(n)) \end{aligned} \quad (2.17)$$

O gradiente local aponta na direção da mudança pretendida dos pesos sinápticos. A equação (2.17) mostra que o gradiente local $\delta_j(n)$ para o neurônio de saída j é igual ao produto do seu sinal erro $e_j(n)$ pela derivada $f'_j(v_j(n))$ da função de ativação associada. Tem-se na equação (2.17) a chave para o cálculo do ajuste de peso $\Delta w_{ji}(n)$.

O ajuste nos pesos deve ser reiterado até que o erro atinja um valor abaixo do que for previamente estabelecido como tolerância.

CAPÍTULO 3

3.1 FUNDAMENTOS DA TEORIA DAS WAVELETS

O termo *wavelet* significa ondinha. Essa ondinha para ser admitida como *wavelet* deve apresentar alguma oscilação e cair rapidamente a zero. Esta propriedade, de admissibilidade, é a mesma que se coloca sobre uma função para que ela seja transformável via *wavelet*.

Conjuntos de *wavelets* são empregados para aproximar um sinal e o objetivo é encontrar uma coleção de *wavelets*-filhas (construídas por dilatação ou compressão e translações da *wavelet* original) que melhor se ajuste, que melhor retrate, o sinal analisado. Uma *wavelet* comprimida é adequada para captar mudanças rápidas (altas frequências no sinal sob análise). Uma *wavelet* alongada é adequada para captar mudanças lentas (baixas frequências no sinal sob análise).

3.2 TRANSFORMADA WAVELET

A transformada *wavelet* é uma operação que integra (de menos a mais infinito, em princípio) o produto de uma função por algum núcleo. O núcleo é chamado *wavelet-mãe* e suas modificações são as *filhas*. Para uma função ser uma *wavelet* ela deve ser admissível, isto é,

oscilar decaindo rapidamente a zero. Essa condição vem junto com a exigência que a componente “dc” da *wavelet* deve ser nula. São as *condições* de admissibilidade de Grossmann-Morlet (1984). Mais rigorosamente, uma função $h \in L^2(\mathbb{R})$ (o espaço de todas as funções de quadrado integrável) é admissível se

$$c_h = \int_{-\infty}^{+\infty} \frac{|H(\omega)|^2}{|\omega|} d\omega < \infty \quad (3.1)$$

onde $H(\omega)$ é a transformada de Fourier de $h(t)$. A finitude da constante c_h garante a inversibilidade da transformada *wavelet*. Toda função admissível pode, em princípio, ser uma *wavelet*-mãe. Dada $h(t)$, $c_h < \infty$ ocorre apenas se $H(0)=0$, ou seja, $\int_{-\infty}^{+\infty} h(t)dt = 0$ (valor médio nulo no domínio do tempo), fato que confere às *wavelets* o papel de filtro passa-banda no domínio Fourier.

A *transformada wavelet* de uma função $f \in L^2(\mathbb{R})$ com respeito a uma *wavelet*-mãe $h(t)$ é definida como

$$W_f(a,b) = \int_{-\infty}^{+\infty} f(t)h_{a,b}^*(t)dt \quad (3.2)$$

onde $*$ denota conjugação complexa. Entretanto, quase sempre, *wavelets* assumem valores reais. As *wavelets*-filhas são geradas a partir de $h(t)$ por dilatação e translação:

$$h_{a,b}(t) = \frac{1}{\sqrt{a}} h\left(\frac{t-b}{a}\right) \quad (3.3)$$

onde $\alpha > 0$ é o fator de dilatação e b é o de translação. O termo $\alpha^{-1/2}$ garante a normalização da energia, mantendo a energia de cada uma das filhas igual à da mãe. Ademais, a condição de admissibilidade colocada pela equação (3.1) produz a *resobabilidade* (Daubechies, 1990) cujo significado é dado por:

$$c_h \langle f_1, f_2 \rangle = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \langle f_1, h_{a,b} \rangle \langle h_{a,b}, f_2 \rangle \frac{dad b}{a^2} \quad (3.4)$$

onde $f_1, f_2 \in L^2(\mathfrak{R})$, $h(t)$ admissível, a diferente de zero e $\langle \cdot, \cdot \rangle$ denota o produto interno

$$\langle f_1, f_2 \rangle = \int_{-\infty}^{+\infty} f_1(t) f_2^*(t) dt \quad (3.5)$$

Quando na equação (3.4) faz-se $f_1 = f_2$, a identidade torna-se a *conservação de energia*

$$c_h \int_{-\infty}^{+\infty} |f(t)|^2 dt = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |W_f(a, b)|^2 \frac{dad b}{a^2} \quad (3.6)$$

significando que a energia calculada no domínio do tempo é a mesma daquela calculada no domínio da transformada *wavelet* (como a relação de Parseval no domínio Fourier). A menos de um fator de escala, a transformada *wavelet* é uma *isometria* de $L^2(\mathfrak{R})$ para $L^2(\mathfrak{R})$. A distribuição de energia no domínio da transformada *wavelet*, isto é, o lado direito da equação (3.6) é denominado *escalograma*. O escalograma mostra numa representação visual bidimensional como a energia total de um sinal se distribui no plano tempo-escala (ou tempo-frequência).

A *inversão* da transformada *wavelet* é dada por:

$$f(t) = \frac{1}{c_h} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W_f(a, b) h_{a,b}(t) \frac{dad b}{a^2} \quad (3.7a)$$

Para se constituir em frame, a transformada *wavelet* discreta gerada amostrando-se os parâmetros (tempo/escala) em um reticulado não igualmente espaçado, deve satisfazer a condição de admissibilidade e os pontos do reticulado devem ser suficientemente próximos para

reter as informações básicas. A família de funções $\{h_{a,b}(t)\}$, $a, b \in \mathbb{Z}$ é dita ser um *frame* se existem duas constantes $A > 0$ e $B < \infty$ tais que para todas as funções $h \in L^2(\mathbb{R})$ tem-se

$$A\|f\|^2 \leq \sum_{a,b} |\langle f, h_{a,b} \rangle|^2 \leq B\|f\|^2 \quad (3.7b)$$

3.3 TIPOS DE FILTROS WAVELETS

3.3.1 Wavelet Morlet

A função *wavelet*-mãe de Morlet é o produto da exponencial complexa pela janela gaussiana:

$$h(t) = \exp(j\omega_0 t) \exp(-t^2/2) \quad (3.8)$$

A *wavelet* cosseno-gaussiano, parte real de h , é uma função real par, cuja transformada de Fourier deslocada, respectivamente para ω_0 e $-\omega_0$ é uma função real par e positiva:

$$H(\omega) = \sqrt{2\pi} \left(\exp\left[-(\omega - \omega_0)^2/2\right] + \exp\left[-(\omega + \omega_0)^2/2\right] \right) \quad (3.9)$$

A figura 3.1 mostra os gráficos da *wavelet* Morlet e seu espectro de Fourier, com $\omega_0 = 4$.

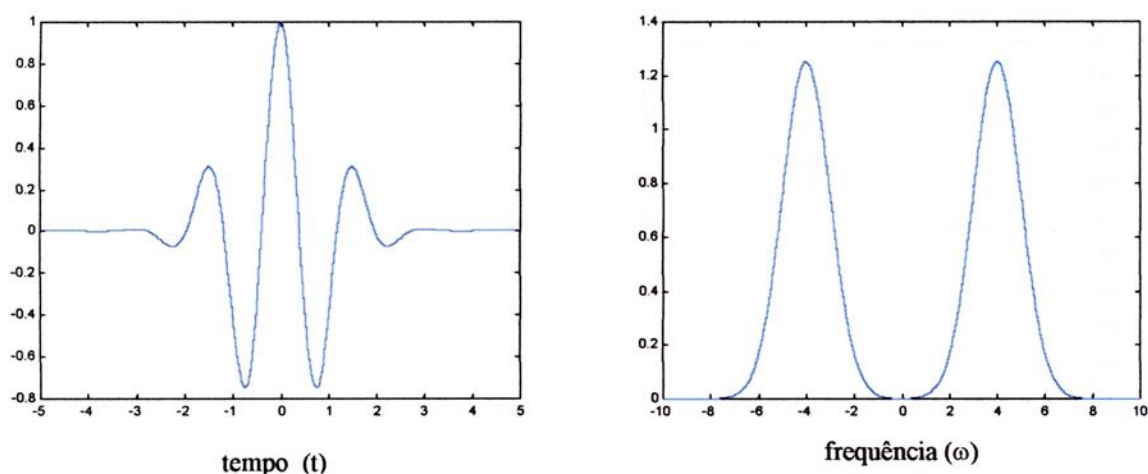


Figura 3.1 - Wavelet cos-gaussiano de Morlet $h(t)$ e sua transformada de Fourier $H(\omega)$

As *wavelets* Morlet não satisfazem, em geral, a condição de admissibilidade porque

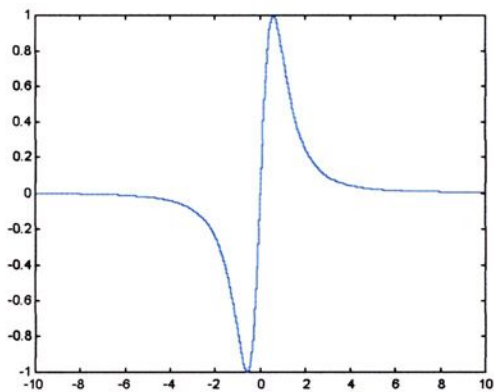
$$H(0) = \sqrt{2\pi} \exp(-\omega_0^2/2) \neq 0 \quad (3.10)$$

e daí $c_h = +\infty$. Entretanto, se ω_0 é bastante grande, digamos $\omega_0 = 4$, $H(0)$ torna-se muito próximo de zero e pode ser tomado como zero, computacionalmente.

3.3.2 Wavelets RASP

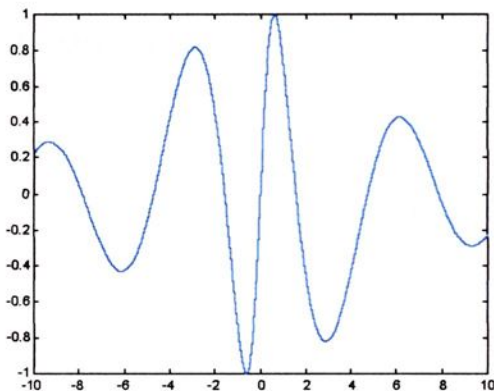
O teorema do resíduo para funções de variáveis complexas pode ser usado para mostrar que algumas funções reais tem média zero, proporcionando a possibilidade de candidaturas a *wavelets*-mães.

É o caso, por exemplo, das funções racionais com polos de segunda ordem (RASP), três delas são apresentadas na figura 3.2.



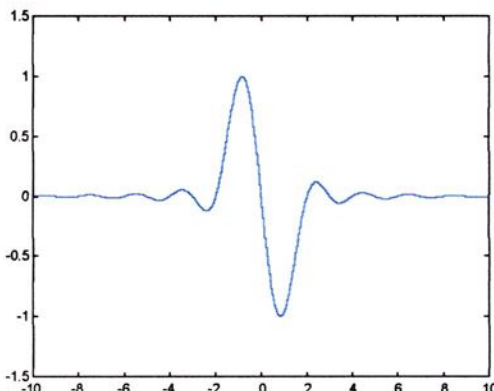
RASP1

$$\frac{k_1 t}{(t^2 + 1)^2} \quad , \quad k_1 = 3,0788$$



RASP2

$$\frac{k_2 t \cos(t)}{(t^2 + 1)} \quad , \quad k_2 = 2,7435$$



RASP3

$$\frac{k_3 \sin(\pi t)}{(t^2 - 1)} \quad , \quad k_3 = 0,6111$$

Figura 3.2 - Funções Racionais com polos de segunda ordem (exemplos de RASP)

3.4 TEOREMA DE RESÍDUO

Seja C um caminho fechado, $C \subset \mathbb{C}$ (corpo dos números complexos), tal que f é analítica sobre C e no interior de C exceto num número finito de pontos singulares $z_1, z_2, \dots, z_n$. Então,

$$\oint_C f(z) dz = 2\pi j \sum_{i=1}^n \text{Res}[f; z_i] \quad (3.11)$$

em que a integral é calculada no sentido anti-horário ao longo de C e se f tem um polo de ordem k em z_i ,

$$\text{Res}[f; z_i] = \lim_{z \rightarrow z_i} \frac{1}{(k-1)!} \frac{d^{k-1}}{dz^{k-1}} \{ (z - z_i)^k f(z) \} \quad (3.12)$$

A RASP $h(t) = \frac{t \cos(t)}{t^2 + 1}$ é aceita como *wavelet*-mãe, pois tem média zero, como se

verifica aplicando o teorema do resíduo, observando-se que outros exemplos são tratados similarmente. Considere-se então a integral no plano complexo

$\oint_C \frac{z e^{jz}}{z^2 + 1} dz$ onde C é a curva fechada simples mostrada na figura 3.3.

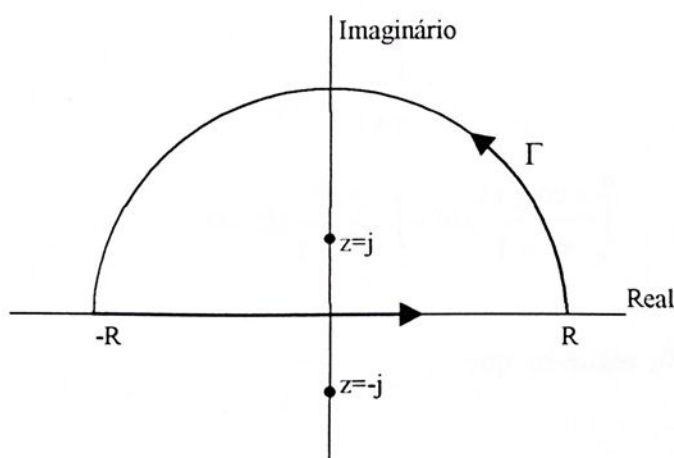


Figura 3.3 - C , uma curva fechada simples

O integrando $\frac{z e^{jz}}{z^2 + 1}$ tem polos simples em $z = \pm j$ e destes apenas $z = j$ ocorre dentro de C .

O residuo de $h(z)$ em $z = j$ é dado por:

$$\text{Res}(h; j) = \lim_{z \rightarrow j} (z - j) \frac{z e^{jz}}{(z + j)(z - j)} = \frac{1}{2e} \quad (3.13)$$

e em virtude de (3.11), tem-se

$$\oint_C \frac{z e^{jz}}{z^2 + 1} dz = 2\pi j \text{Res}[h; j] = \frac{\pi}{e} j \quad (3.14)$$

ou

$$\int_{-R}^R \frac{x e^{jk}}{x^2 + 1} dx + \int_{\Gamma} \frac{z e^{jz}}{z^2 + 1} dz = \frac{\pi}{e} j \quad (3.15)$$

isto é,

$$\int_{-R}^R \frac{x \cos(x)}{x^2 + 1} dx + j \int_{-R}^R \frac{x \sin(x)}{x^2 + 1} dx + \int_{\Gamma} \frac{z e^{jz}}{z^2 + 1} dz = \frac{\pi}{e} j \quad (3.16)$$

e assim a parte real é

$$\int_{-R}^R \frac{x \cos(x)}{x^2 + 1} dx + \int_{\Gamma} \frac{z e^{jz}}{z^2 + 1} dz = 0. \quad (3.17)$$

Fazendo-se $R \rightarrow \infty$, segue-se que

$$\int_{\Gamma} \frac{z e^{jz}}{z^2 + 1} dz = \lim_{R \rightarrow \infty} \int_0^{\pi} \frac{R e^{j\theta} e^{jR e^{j\theta}} R e^{j\theta} j d\theta}{R^2 e^{j2\theta} + 1} \leq \int_0^{\pi} \lim_{R \rightarrow \infty} \frac{R^2 |e^{-R \sin \theta}| d\theta}{R^2 - 1}. \quad (3.18)$$

Note-se que a desigualdade $|A+B| \geq |A|-|B|$ foi usada no denominador e $\left| \int AB d\theta \right| \leq \int |A||B| d\theta$ no numerador da equação (3.18).

Note-se também que $\lim_{R \rightarrow \infty} e^{-R \sin \theta} = 0$ e este fato levado à equação (3.18) assegura que

$$\int_{\Gamma} \frac{z e^{jz}}{z^2 + 1} dz = 0 \text{ e da equação (3.17) segue-se que } \lim_{R \rightarrow \infty} \int_{-R}^R \frac{x \cos(x)}{x^2 + 1} dx = 0 \text{ ou que}$$

$h(t) = \frac{t \cos(t)}{t^2 + 1}$ tem média zero podendo assim ser empregada como *wavelet*-mãe.

3.4.1 WAVELET SLOG

Uma função estritamente crescente, diferenciável (suave) cujo gráfico lembra um S alongado é nomeada de logística, usualmente empregada como ativação no neurônio de saída de uma rede neural [Hay94]. A superposição de um número finito de logísticas (SLOG) devidamente ponderadas e deslocadas constituem uma família de *wavelets*-mães bem adaptadas para problemas de representação e classificação oriundas de rede neural. Uma função logística sigmoidal é por exemplo:

$$f_{\log}(t) = \frac{1}{1 + e^{-t}} \quad (3.19)$$

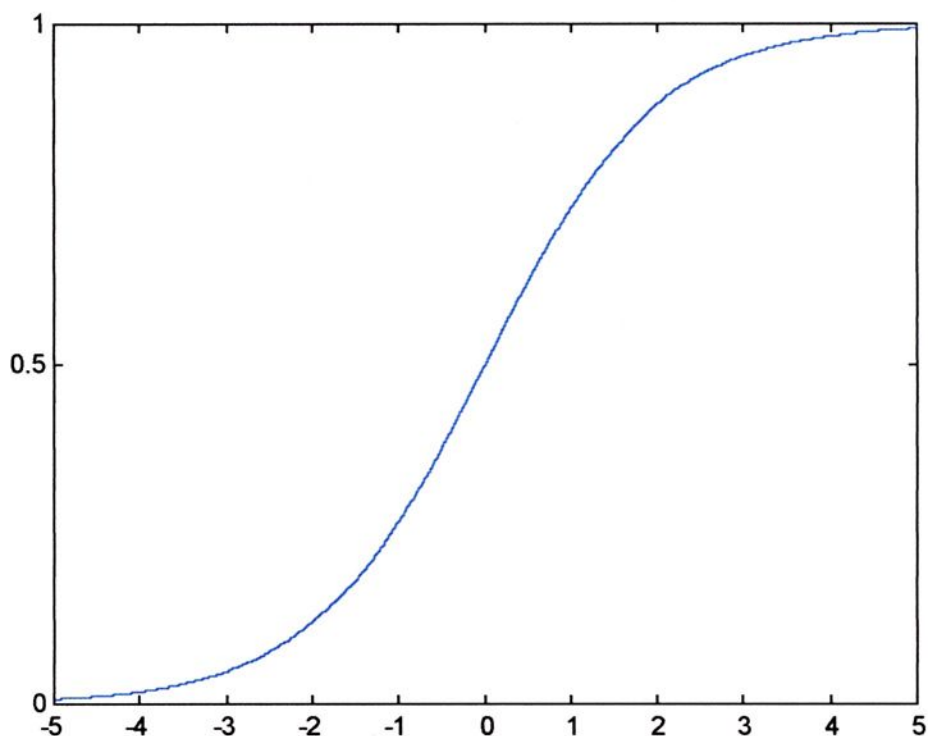


Figura 3.5 - Função Logística Sigmoidal

Essa família de *wavelets*-mães aparecem naturalmente quando redes neurais são utilizadas em questões de representação e classificação. A habilidade com que uma função entrada-saída de rede pode prover uma *wavelet*, sugere que ela tem propriedades de *aproximação* semelhante, isto é, representação de mapas arbitrários de $L^2(\mathbb{R})$. Y. Pati e P. Krishnaprasad (Pati e Krishnaprasad, 1993) provaram construtivamente o teorema de aproximação universal em $L^2(\mathbb{R})$ mostrando que a implementação de uma rede neural feedforward garante a construção de um frame afim para $L^2(\mathbb{R})$. O teorema afirma que *redes neurais feedforward com funções de ativação sigmoidais e uma única camada escondida pode representar qualquer função $f \in L^2(\mathbb{R})$* . A decomposição do sinal em multiresolução sustenta a prova do teorema. Para se constituir um frame, a transformada *wavelet* discreta gerada amostrando os parâmetros (tempo/escala) em um reticulado, não igualmente espaçado, devem satisfazer a condição de admissibilidade e os pontos do reticulado devem estar suficientemente próximos para reter as informações essenciais. Assim, a questão da reconstrução do sinal a partir dos seus valores transformados depende do espaçamento no dito reticulado. Um reticulado mais fino permite reconstrução mais fácil,

embora com possível redundância. Um reticulado mais grosseiro resulta em perda de informação. O cuidado deve estar presente para que se obtenha um conjunto de funções *wavelets* que seja minimal, isto é, não redundante.

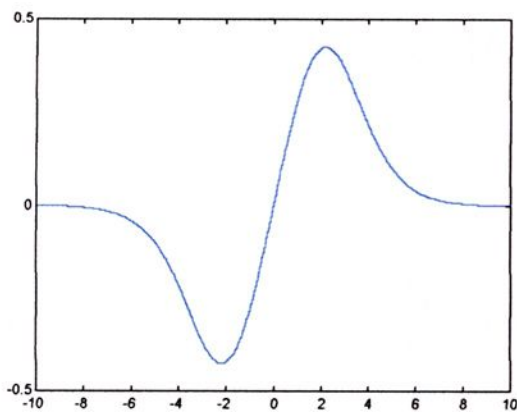
A figura (3.6a) representa a primeira *wavelet*-mãe SLOG exibindo a seguinte superposição de sigmóides logísticas:

$$h_{s\log 1}(t) = \frac{1}{1+e^{-t+1}} - \frac{1}{1+e^{-t+3}} - \frac{1}{1+e^{-t-3}} + \frac{1}{1+e^{-t-1}} \quad (3.20)$$

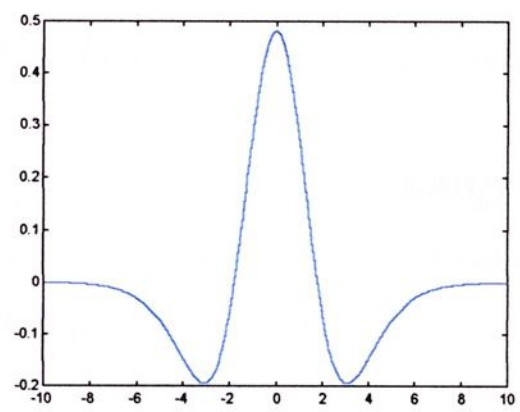
Esta *wavelet*-mãe apresenta uma oscilação mínima. A figura (3.6b) apresenta outra *wavelet* SLOG tendo uma meia oscilação extra. Neste segundo exemplo ainda, uma soma finita de funções logísticas ponderadas e deslocadas é empregada:

$$h_{s\log 2}(t) = \frac{3}{1+e^{-t-1}} - \frac{3}{1+e^{-t+1}} - \frac{1}{1+e^{-t-3}} + \frac{1}{1+e^{-t+3}} \quad (3.21)$$

Da mesma forma, somando e ou subtraindo sigmóides logísticas deslocadas, novas famílias de *wavelets* SLOG podem ser geradas com mais ou menos meia-oscilação.



(a) SLOG1



(b) SLOG2

Figura 3.6 - Exemplos de SLOG

3.4.2 WAVELET POLYWOG

Inúmeras *wavelets*-mães aparecem de funções do tipo gaussiano janeladas por polinômios (POLYWOG).

Por exemplo, todas as derivadas de uma função gaussiana são POLYWOGs sendo pois admissíveis como *wavelets*-mães. A seguir são apresentados quatro exemplos de tais funções:

POLYWOG1:

$$h_{POLYWOG1} = R_1 t \exp\left(-\frac{t^2}{2}\right), \quad R_1 = \sqrt{e} \quad (3.22)$$

POLYWOG2:

$$h_{POLYWOG2} = R_2 (t^3 - 3t) \exp\left(-\frac{t^2}{2}\right), \quad R_2 = 0,7246 \quad (3.23)$$

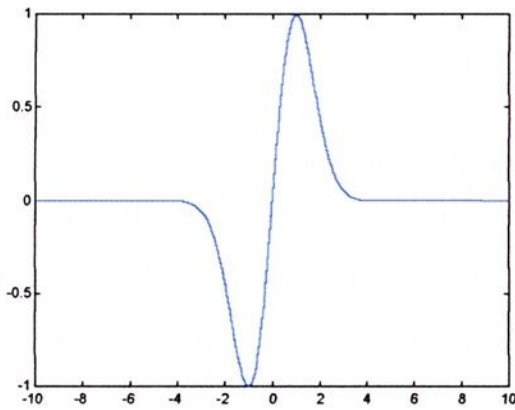
POLYWOG3:

$$h_{POLYWOG3} = R_3 (t^4 - 6t^2 + 2) \exp\left(-\frac{t^2}{2}\right), \quad R_3 = \frac{1}{3} \quad (3.24)$$

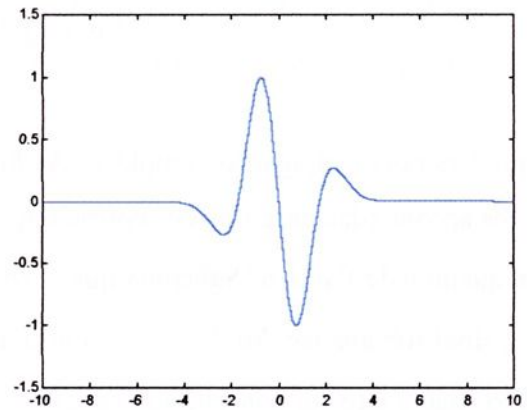
POLYWOG4:

$$h_{POLYWOG4} = (1 - t^2) \exp\left(-\frac{t^2}{2}\right) \quad (3.25)$$

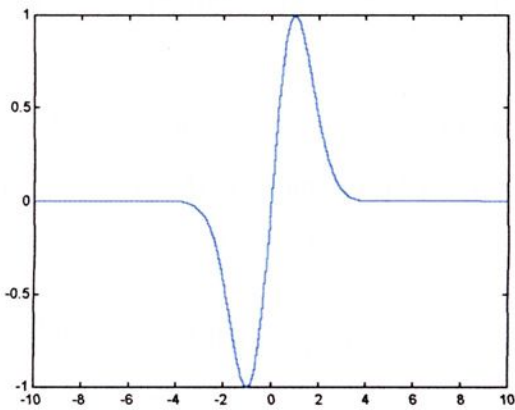
A figura 3.7 mostra exemplos de POLYWOGs.



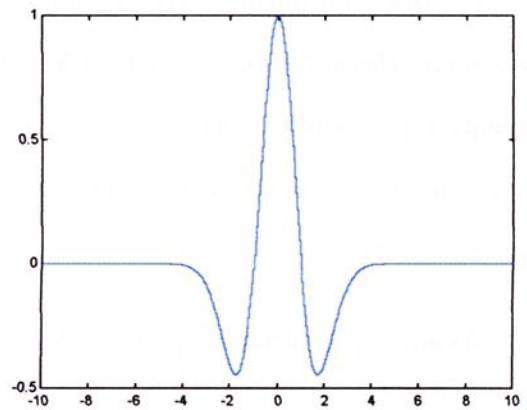
POLYWOG1



POLYWOG2



POLYWOG3



POLYWOG4

Figura 3.7 - Exemplos de wavelets POLYWOG

A *hermiticidade* de um operador pode ser outro caminho para construção de *wavelets* POLYWOG.

É oportuno uma revisão sobre operadores *hermitianos* ou *auto-adjuntos* e como eles marcam presença na representação de quantidades físicas no mesmo sentido que a “*frequência*” surge com a transformada de Fourier. Expanda-se um sinal na forma

$$f(t) = \int_{-\infty}^{\infty} F(a)u(a,t)da \quad (3.26)$$

em que $u(a,t)$ são as funções base e $F(a)$ são os *coeficientes da expansão* ou a “*transformada integral do sinal*”. Será mostrado que $F(a)$ é dado por

$$F(a) = \int_{-\infty}^{+\infty} f(t) u^*(a, t) dt \quad (3.27)$$

em que $*$ denota conjugação complexa. As funções base (base no sentido de Hilbert, conjunto de funções apropriadas para fins de representação), $u(a, t)$ são funções da variável t (tempo) e uma outra quantidade física a . Sabemos que $F(a)$ é uma avaliação da importância do particular a para o sinal sob análise. Se $F(a)$ for significativamente grande apenas em uma região particular, então o sinal é dito ser concentrado naqueles valores de a .

Um operador linear $A: E \rightarrow E$, num espaço vetorial munido de produto interno, chama-se *auto-adjunto (hermitiano)* quando $(Au, v) = (u, Av)$ para quaisquer $u, v \in E$, em que $(\cdot, \cdot)$ representa um produto interno em E . Presentemente, de interesse tem-se no papel de E , o espaço vetorial das funções de quadrado integrável, $E = L^2(\mathfrak{R})$, munido do produto interno usual.

Assim, repetindo, um operador A é *hermitiano* se para todo par de funções $f(t)$ e $g(t)$ tem-se

$$\int_{-\infty}^{+\infty} g^*(t) A(f(t)) dt = \int_{-\infty}^{+\infty} f(t) (A(g(t)))^* dt \quad (3.28)$$

A hermiticidade do operador garante que o conjunto das autofunções é *completo* (no sentido de Riesz-Hilbert) e ortogonal.

Isto significa que as autofunções satisfazem

$$\int_{-\infty}^{+\infty} u^*(a', t) u(a, t) dt = \delta(a - a') \quad (3.29a)$$

$$\int_{-\infty}^{+\infty} u^*(a, t') u(a, t) da = \delta(t - t') \quad (3.29b)$$

Sendo δ a representação clássica da distribuição de Dirac.

São estas propriedades que respaldam as transformadas entre $f(t)$ e $F(a)$ como dadas nas equações (3.26) e (3.27). Em particular, para se obter $F(a)$, multiplica-se (3.26) por $u^*(a', t)$ e integra-se com respeito a t :

$$\int_{-\infty}^{+\infty} f(t) u^*(a', t) dt = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(a) u(a, t) u^*(a', t) da dt \quad (3.30a)$$

$$= \int_{-\infty}^{+\infty} F(a) \delta(a - a') da \quad (3.30b)$$

$$= F(a') \quad (3.30c)$$

provando-se assim, a relação inversa (3.27).

Outro fato notável é que os operadores hermitianos possuem autovalores reais e isto é significativo porque na natureza tudo que é mensurável é real. Entretanto, operadores não-hermitianos, alguns deles, são importantes, embora, não seja possível agregar aos seus autovalores conteúdo físico mensurável. Provaremos a seguir que a diferenciação é um operador hermitiano. Consideremos o operador $\mathbf{A} = \frac{1}{j} \frac{d}{dt}$ definido em $L^2(\mathfrak{R})$ e sejam $f(t)$ e $g(t)$ quaisquer duas de suas funções. Integrando-se por partes,

$$\int_{-\infty}^{+\infty} g^*(t) \frac{1}{j} \frac{d}{dt} f(t) dt = \frac{1}{j} fg^* \Big|_{-\infty}^{+\infty} - \frac{1}{j} \int_{-\infty}^{+\infty} f(t) \frac{d}{dt} g^*(t) dt \quad (3.31)$$

Desde que $f(t), g(t) \in L^2(\mathfrak{R})$, segue que $fg^* \Big|_{-\infty}^{+\infty}$ se anula e a equação (3.31) pode ser

escrita alternativamente como

$$\int_{-\infty}^{+\infty} g^*(t) \frac{1}{j} \frac{d}{dt} f(t) dt = \int_{-\infty}^{+\infty} f(t) \left(\frac{1}{j} \frac{d}{dt} g(t) \right)^* dt$$

ou

$$(\mathbf{A}(f), g) = (f, \mathbf{A}(g)) \quad (3.32)$$

provando assim que $\mathbf{A}$ é hermitiano.

Sejam $f(t)$ e $g(t)$ elementos de $L^2(\mathfrak{R})$ onde $f(t)$ é uma função polinomial janelada e $g(t)$ é uma gaussiana. A partir da equação (3.32) tem-se

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{t^2}{2}\right) \frac{d}{dt} f(t) dt = - \int_{-\infty}^{+\infty} f(t) t \exp\left(-\frac{t^2}{2}\right) dt \quad (3.33)$$

Como exemplo de *wavelets*-mães produzidas por derivação pode-se citar a POLYWOG4 (figura (3.7)), a qual é a derivada segunda da função gaussiana, também conhecida por chapéu mexicano (Gabor):

$$h_{POLYWOG4} = (1-t^2) \exp\left(-\frac{t^2}{2}\right).$$

A transformada de Fourier do chapéu mexicano é

$$H(\omega) = -\omega^2 \exp\left(-\frac{\omega^2}{2}\right), \quad (3.34)$$

uma função real e par, satisfazendo a condição de admissibilidade $H(0)=0$, isto é,

$$\int_{-\infty}^{+\infty} (1-t^2) \exp\left(-\frac{t^2}{2}\right) dt = 0 \quad (3.35)$$

Usando a hermiticidade da diferenciação, combinando as equações (3.33) e (3.35) obtém-se

$$\int_{-\infty}^{+\infty} (1-t^2) \exp\left(-\frac{t^2}{2}\right) dt = \int_{-\infty}^{+\infty} \left(t - \frac{t^3}{3}\right) t \exp\left(-\frac{t^2}{2}\right) dt = 0 \quad (3.36)$$

e daí uma nova *wavelet*-mãe POLYWOG com média zero e boas propriedades de localização.

A figura (3.8) exibe uma nova *wavelet*-mãe, POLYWOG5:

$$h_{POLYWOG5} = R_3 (3t^2 - t^4) \exp\left(-\frac{t^2}{2}\right) \quad (3.37)$$

em que $R_3 = 0,8244$ surge para garantir normalização.

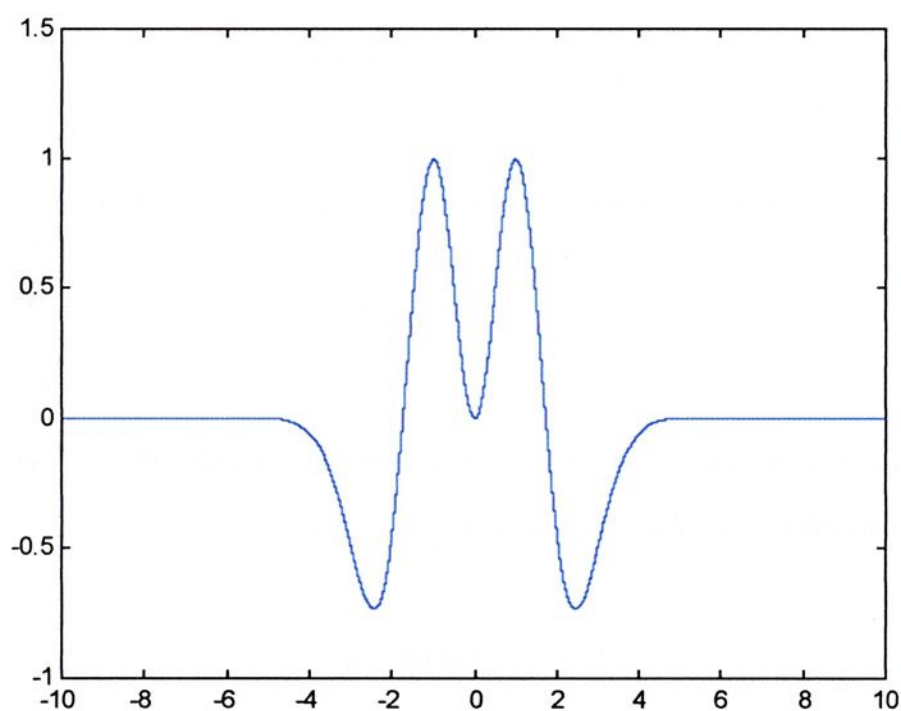


Figura 3.8 - Wavelet POLYWOG5, Hermiticidade da derivada

3.4.3 WAVELET DE SHANNON

Um exemplo de *wavelet*-mãe que gera uma *base ortonormal* para $L^2(\mathfrak{R})$ é a de Shannon, descontínua em frequência e portanto espalhada no tempo. Para que alguma função $\phi(t) \in L_2(\mathfrak{R})$ forme uma base ortogonal, $\{\phi(t-n), n \in \mathbb{N}\}$ deve cumprir a condição de ortonormalidade:

$$(\phi(t-\ell), \phi(t-k)) = \delta(k-\ell) \quad (3.38)$$

ou no domínio de Fourier, basta que

$$\sum_{k=-\infty}^{k=+\infty} |\Phi(\omega + 2k - k\pi)|^2 = 1 \quad (3.39)$$

Por exemplo, o teorema da amostragem de Shannon garante que as funções

$$\phi(t-k) = \frac{\text{sen } \pi(t-k)}{\pi(t-k)} \quad (3.40)$$

constituem uma base ortonormal desde que a transformada de Fourier Φ tem suporte compacto $L^2[-\pi, \pi]$ e daí toda função $f(t)$ pode ser expressa como

$$f(t) = \sum_{k=-\infty}^{\infty} f(k) \phi(t-k) \quad (3.41)$$

A verificação da ortonormalidade é simples, no domínio da frequência. Ora,

$$\phi(t) = \frac{\text{sen}(\pi t)}{\pi t} \leftrightarrow \Phi(\omega) = \begin{cases} 1 & \text{se } |\omega| < \pi \\ 0 & \text{se } |\omega| \geq \pi \end{cases} \quad (3.42)$$

e o produto interno definido em (3.5) garante que

$$\int_{-\infty}^{+\infty} \phi(t-k) \phi^*(t-\ell) dt = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-j\omega k} \Phi(\omega) e^{j\omega \ell} \phi^*(\omega) d\omega \quad (3.43a)$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-j\omega(k-\ell)} d\omega = \delta(k-\ell) \quad (3.43b)$$

Analogamente, se uma função *wavelet* básica $h(t) \in L^2(\mathfrak{R})$ satisfaz a condição de admissibilidade e sua transformada de Fourier $H(\omega)$ satisfaz a condição de ortonormalidade (3.39):

$$\sum_{k=-\infty}^{\infty} |H(\omega + 2k\pi)|^2 = 1, \quad (3.44)$$

as translações inteiras da *wavelet*, $h(t-k)$, também formam uma base ortonormal. A equação (3.44) é satisfeita por, exemplificando,

$$H(\omega) = \begin{cases} 1 & \text{se } \pi < |\omega| < 2\pi \\ 0 & \text{se } |\omega| > 2\pi \text{ ou } |\omega| < \pi \end{cases}$$

produzindo a *wavelet Shannon*

$$h_{Shannon}(t) = \frac{\text{sen}(2\pi t) - \text{sen}(\pi t)}{\pi t}$$

cujo gráfico está mostrado na figura 3.9

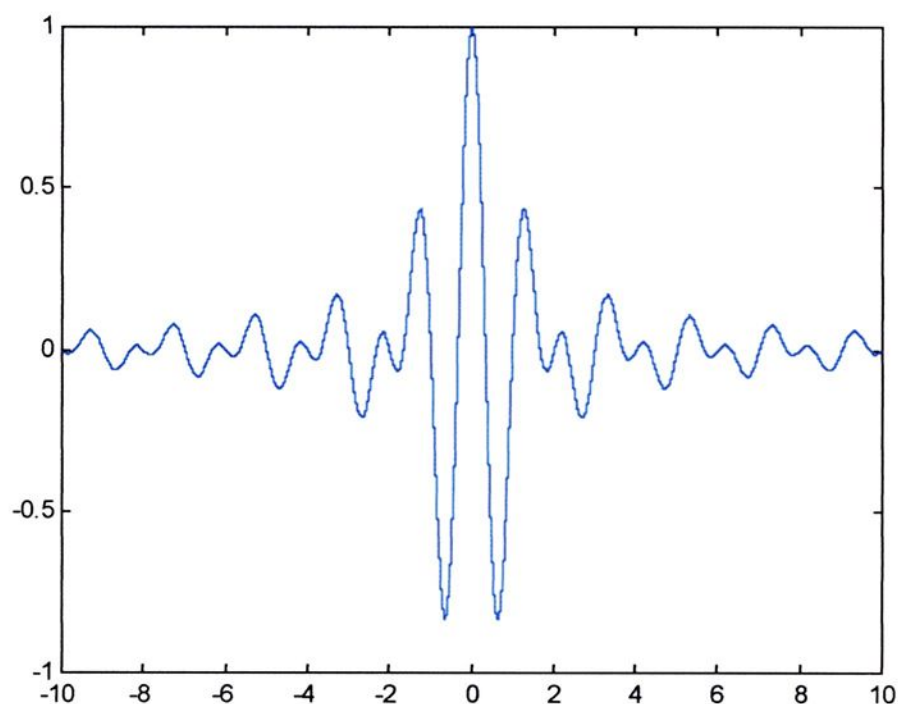


Figura 3.9 - Wavelet-mãe de Shannon

A familiaridade com a teoria da transformada *wavelet* oferece uma alternativa promissora para aproximar funções arbitrárias quando combinada com redes neurais. Uma indicação de como isso ocorre será feita em seguida, no próximo capítulo.

CAPÍTULO 4

4.1 WAVENETS

A transformada *wavelet* combinada com rede neural tem sido utilizada como uma alternativa as redes feedforward no processo de mapeamento de relações entrada/saída, para aproximações de funções arbitrárias. Esta formulação, denominada *wavenet*, pode ser entendida basicamente como uma rede neural ativada por *wavelets* especiais.

O algoritmo *wavenet* consiste de duas etapas, a auto-construção da rede e a minimização do erro. Num primeiro passo é definido a estrutura da rede necessária para representar o sinal e a *wavelet* é utilizada na caracterização dos neurônios. A rede gradualmente recruta neurônios escondidos para cobrir efetiva e suficientemente a região tempo-frequência ocupada pelo sinal e os parâmetros da rede são atualizados simultaneamente. Numa segunda etapa, as aproximações de erros instantâneos são minimizados utilizando-se a técnica baseada em gradiente descendente.

A Figura 4.1 mostra esquematicamente o processo de identificação de uma planta desconhecida, utilizando o conceito *wavenet*. O sinal de entrada da planta $u(t)$ produz a saída $y(t)$ que será comparada com o sinal de saída da rede. Neste caso, a *wavenet* é alimentada com o mesmo sinal de entrada da planta e a saída da rede $\hat{y}(t)$ é comparada com a saída da planta (meta). O erro $e(t)$ é minimizado com relação aos parâmetros da rede, processo repetido iterativamente até que a saída da rede seja aproximadamente igual a saída da planta. A descrição detalhada do algoritmo de identificação (*wavenet*) é apresentada no próximo tópico.

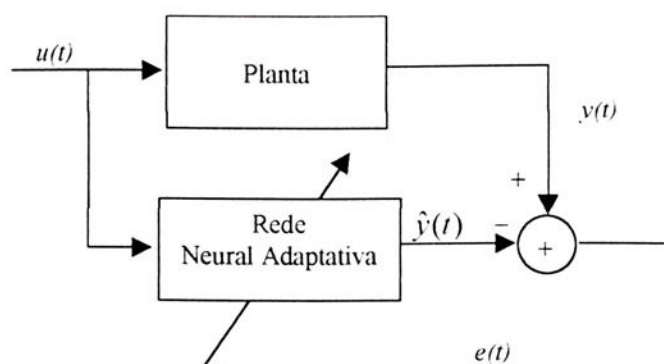


Figura 4.1 - Identificação utilizando rede neural adaptativa

4.2 ALGORITMO WAVENET

A rede neural é arquitetada objetivando produzir um sinal de saída o mais próximo possível da saída da planta. A rede é descrita em termos de componentes *wavelet* e seus respectivos parâmetros a (escala) e b (deslocamento).

A Figura 4.2 mostra a arquitetura do algoritmo. Neste caso, o sinal de entrada $u(t)$ excita cada neurônio (*wavelet*) e a saída da rede aproximante, $\hat{y}(t)$, é dada pela soma ponderada dos vários neurônios. Portanto, o sinal é representado pela combinação linear de um conjunto de *wavelets*-filhas $h_{a,b}(t)$, em que $h_{a,b}(t)$ são geradas por dilatação, a , e translação b , da *wavelet*-mãe $h(t)$

$$h_{a,b}(t) = h\left(\frac{t-b}{a}\right), \quad (4.1)$$

onde $a > 0$ e também desconsiderando aqui, a preocupação com normalização da energia.

O sinal estimado fornecido pela rede, $\hat{y}(t)$ é representado por:

$$\hat{y}(t) = u(t) \sum_{k=1}^K w_k h_{a_k, b_k}(t), \quad (4.2)$$

em que K é um número de janelas *wavelets*, e, w_k são os pesos.

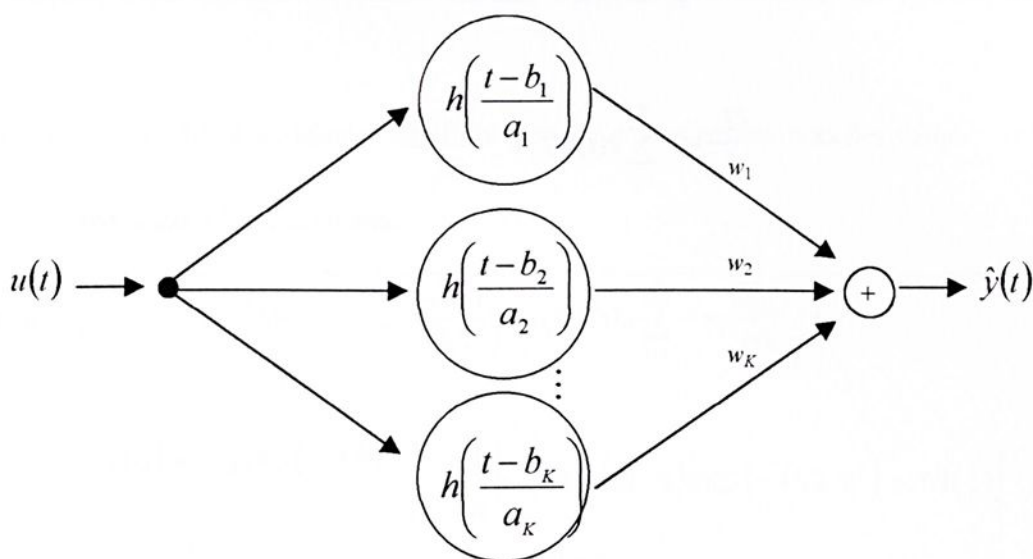


Figura 4.2 - Detalhes da estrutura da rede neural adaptativa

Os parâmetros da rede w_k , a_k e b_k devem ser otimizados no sentido dos mínimos quadráticos, minimizando a função energia, E do sinal. Denotando-se o erro, variável com o tempo, por

$$e(t) = y(t) - \hat{y}(t), \quad (4.3)$$

em que $y(t)$ (Figura 4.1) é a resposta desejada (meta), a função energia (custo) será dada pela expressão (4.4),

$$E = \frac{1}{2} \sum_{t=1}^T e^2(t). \quad (4.4)$$

Os parâmetros ótimos da rede são obtidos através da minimização da função custo. Neste trabalho será utilizado o método do gradiente descendente para minimizar a função custo E , portanto, é necessário o cálculo dos gradientes $\frac{\partial E}{\partial w_k}$, $\frac{\partial E}{\partial a_k}$ e $\frac{\partial E}{\partial b_k}$ no processo de atualização das mudanças incrementais dos parâmetros w_k , a_k , b_k , respectivamente. As equações de (4.5) a (4.7) são utilizadas para o cálculo dos vários gradientes que a tarefa requer.

$$\frac{\partial E}{\partial w_k} = - \sum_{t=1}^T e(t) h(\tau) u(t) \quad (4.5)$$

$$\frac{\partial E}{\partial b_k} = - \sum_{t=1}^T e(t) u(t) w_k \frac{\partial h(\tau)}{\partial b_k} \quad (4.6)$$

$$\frac{\partial E}{\partial a_k} = - \sum_{t=1}^T e(t) u(t) w_k \tau \frac{\partial h(\tau)}{\partial b_k} \quad (4.7)$$

em que

$$\tau = \frac{t - b_k}{a_k} \quad (4.8)$$

As mudanças incrementais dos coeficientes são dados pelas expressões:

$$\Delta w = - \frac{\partial E}{\partial w}, \quad (4.9a)$$

$$\Delta b = - \frac{\partial E}{\partial b}, \quad (4.9b)$$

$$\Delta a = \frac{\partial E}{\partial a} \quad (4.9c)$$

os mesmos serão atualizados de acordo com a regra

$$w(n+1) = w(n) + \mu_w \Delta w \quad (4.10)$$

$$b(n+1) = b(n) + \mu_b \Delta b \quad (4.11)$$

$$a(n+1) = a(n) + \mu_a \Delta a \quad (4.12)$$

em que μ é a taxa de aprendizado, previamente fixada.

O sucesso da identificação da planta, neste caso, está diretamente associada a escolha da família *wavelet* e ao número de janelas usadas, bem como a escolha dos valores iniciais das escalas.

O Quadro 4.1 mostra algumas famílias *wavelet* e suas respectivas derivadas.

Quadro 4.1 - *Wavelets* e suas derivadas

Nome	$h(\tau)$	$\frac{\partial h(\tau)}{\partial b}$
Morlet	$\cos(\omega_0 \tau) \exp(-0,5 \tau^2)$	$\frac{1}{a} [\omega_0 \text{sen}(\omega_0 \tau) \exp(-0,5 \tau^2) + \tau h(\tau)]$
RASP1	$\frac{\tau}{(\tau^2 + 1)^2}$	$\frac{1(3\tau^2 - 1)}{a(\tau^2 + 1)^3}$
RASP2	$\frac{\tau \cos(\tau)}{\tau^2 + 1}$	$\frac{\tau \left(\frac{\tau^2 + 1}{a} \right) \text{sen}(\tau) + \left(\frac{\tau^2 - 1}{a} \right) \cos(\tau)}{(\tau^2 + 1)^2}$
RASP3	$\frac{\text{sen}(\pi \tau)}{\tau^2 - 1}$	$\frac{\left(\frac{2\tau}{a} \right) \text{sen}(\pi \tau) - \pi \left(\frac{\tau^2 - 1}{a} \right) \cos(\pi \tau)}{(\tau^2 - 1)^2}$
SLOG1	$\frac{1}{1 + e^{-\tau+1}} - \frac{1}{1 + e^{-\tau+3}} +$ $-\frac{1}{1 + e^{-\tau-1}} + \frac{1}{1 + e^{-\tau-3}}$	$\frac{1}{a} \left[-\frac{e^{\tau+1}}{(1 + e^{-\tau+1})^2} + \frac{e^{-\tau+3}}{(1 + e^{-\tau+3})^2} + \right.$ $\left. + \frac{e^{-\tau-3}}{(1 + e^{-\tau-3})^2} - \frac{e^{-\tau-1}}{(1 + e^{-\tau-1})^2} \right]$
SLOG2	$\frac{3}{1 + e^{-\tau-1}} - \frac{3}{1 + e^{-\tau+1}} +$	$\frac{1}{a} \left[-\frac{3e^{-\tau-1}}{(1 + e^{-\tau-1})^2} + \frac{3e^{-\tau+1}}{(1 + e^{-\tau+1})^2} + \right.$

	$+ \frac{1}{1+e^{-\tau-3}} + \frac{1}{1+e^{-\tau+3}} + \frac{e^{-\tau-3}}{(1+e^{-\tau-3})^2} - \frac{e^{-\tau+3}}{(1+e^{-\tau+3})^2} \Big]$	
POLYWOG1	$\tau \exp\left(-\frac{\tau^2}{2}\right)$	$\frac{1}{a}(\tau^2 - 1)\exp\left(-\frac{\tau^2}{2}\right)$
POLYWOG2	$(\tau^3 - 3\tau) \exp\left(-\frac{\tau^2}{2}\right)$	$\frac{1}{a}(\tau^4 - 6\tau^2 + 3)\exp\left(-\frac{\tau^2}{2}\right)$
POLYWOG3	$(\tau^4 - 6\tau^2 + 3) \exp\left(-\frac{\tau^2}{2}\right)$	$\frac{1}{a}(\tau^5 - 10\tau^3 + 15\tau)\exp\left(-\frac{\tau^2}{2}\right)$
POLYWOG4	$(1 - \tau^2) \exp\left(-\frac{\tau^2}{2}\right)$	$\frac{1}{a}(3\tau - \tau^3)\exp\left(-\frac{\tau^2}{2}\right)$
POLYWOG5	$(3\tau^2 - \tau^4) \exp\left(-\frac{\tau^2}{2}\right)$	$\frac{1}{a}(-\tau^5 + 7\tau^3 - 6\tau)\exp\left(-\frac{\tau^2}{2}\right)$
Shannon	$\frac{\text{sen } 2\pi\tau - \text{sen } \pi\tau}{\pi\tau}$	$\frac{\pi(-\pi\tau \cos \pi\tau - 2\pi \cos 2\pi\tau + \text{sen } \pi\tau + \text{sen } 2\pi\tau)}{a(\pi\tau)^2}$

Uma outra área de grande interesse de utilização dessa técnica é no processo de controle adaptativo de plantas não lineares. No tratamento e controle de sistemas lineares invariantes no tempo, a questão de estabilidade e robustez é bem conhecida, entretanto, para sistemas não lineares esses aspectos ainda são incipientes. Neste relato, o conceito de redes neurais adaptativas será empregado na identificação e controle de plantas não lineares, conforme será discutido no próximo item.

4.3 CONTROLADORES NEURO WAVENET AUTO-AJUSTÁVEIS

O algoritmo *wavenet*, sua elaboração e aplicação na identificação de plantas não-lineares é tratado neste trabalho com grande deferência. Controle é outra possibilidade de emprego das redes ativadas por *wavelets*. Neste caso, todos os parâmetros envolvidos se alteram e se adaptam segundo preceitos da otimização, trilhando caminho diferente quando a ação de controle faz uso das ferramentas envolvendo as funções de Lyapunov. Nosso enfoque parte da descrição sucinta de uma planta desconhecida através das equações de estado, aqui tratando-se de sistemas de única entrada e única saída.

4.3.1 ESTRUTURA E DEFINIÇÃO

Exercer controle sobre um sistema deve ser entendido como ação corretiva necessária para que a saída da rede $\hat{y}$ se aproxime da saída da planta y e ambas se aproximem do sinal de referência r . Este trabalho tem como preocupação a discussão da estrutura de uma rede neural ativada por *wavelets* especiais, as chamadas *wavenets* utilizadas inicialmente para identificar uma planta eventualmente desconhecida, e em seguida promover uma ação de controle. A Figura 4.3 mostra esquematicamente o neuro controlador.

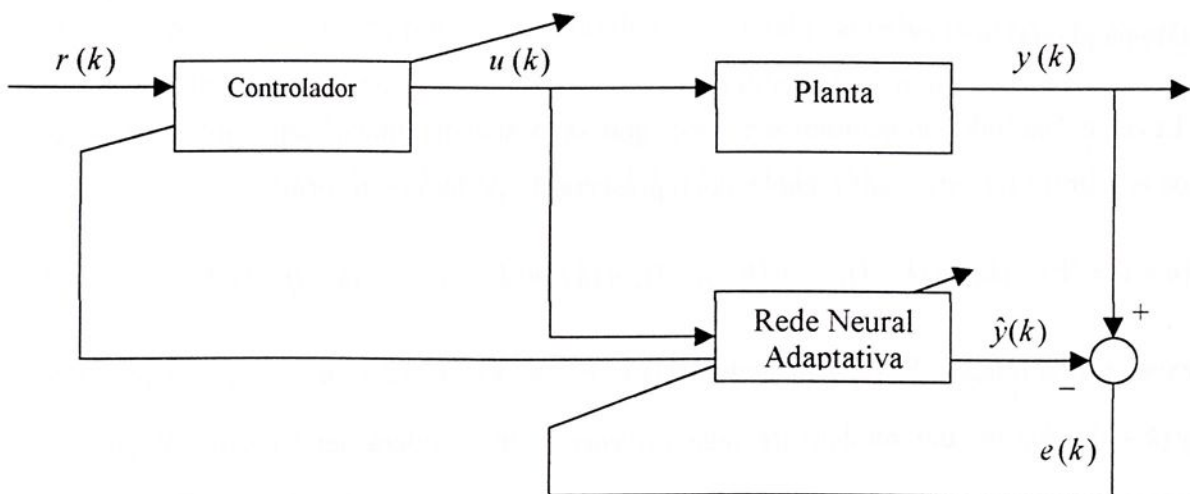


Figura 4.3 - Sistema de neuro controlador auto-ajustável

A criação estruturada da rede e sua preparação adaptativa para torná-la apropriada primeiro para a identificação de uma planta e posterior exercício de controle é basicamente uma questão de otimização, na verdade, uma questão de minimizar o erro médio quadrático que mede as diferenças entre as saídas da rede e da planta. O erro médio quadrático

$$E = \frac{1}{2} \sum_k (y(k) - \hat{y}(k))^2$$

é minimizado (na etapa de controle) através do método do gradiente descendente aplicado para adaptar reiteradamente os coeficientes do filtro (método LMS). Os sistemas dinâmicos que trataremos neste tópico são descritos através de equações de estado.

Consideremos agora um sistema dinâmico de uma única entrada e única saída (SISO) representado pelas equações de estado

$$x(k+1) = f(x(k), u(k), k) \quad (4.13)$$

$$y(k) = g(x(k), k) \quad (4.14)$$

em que $x(k)$, $u(k)$ e $y(k)$ são reais. Admitamos que as funções desconhecidas f e g sejam de classe $C^1(\mathbb{R}^3, \mathbb{R})$. Os únicos dados acessíveis são u (entrada tanto da rede quanto da planta) e y (saída da planta).

Levin e Narendra mostraram em 1995 que se o sistema linearizado em torno de um estado de equilíbrio for *observável*, então sua representação poderá ter a forma

$$y(k+1) = \Psi(y(k), y(k-1), \dots, y(k-n+1), u(k), u(k-1), \dots, u(k-n+1)) \quad (4.15)$$

isto é, existirá uma função $\Psi(\cdot)$ mapeando $y(k)$, e $u(k)$ (e seus $n-1$ valores passados) em $y(k+1)$. Assim, um modelo de rede *wavenet* $\hat{\Psi}$ poderá ser treinada objetivando aproximar (no sentido dos mínimos quadráticos) Ψ numa certa região de interesse. Não importa se o modelo da planta é acessível. Importante é a adaptação do modelo pela atualização dos parâmetros do controle *on-line*. Um enfoque diferente (e adotado aqui) foi proposto por Narendra, Balakrishnan e Ciliz em 1995 para uma planta descrita pela equação

$$y(k+1) = \Phi(y(k), y(k-1), \dots, y(k-n+1), u(k), u(k-1), \dots, u(k-n+1)) + \Gamma(y(k), y(k-1), \dots, y(k-n+1), u(k), u(k-1), \dots, u(k-n+1)) \cdot u(k) \quad (4.16)$$

Em busca de simplicidade e concisão, ela será reescrita na forma

$$y(k+1) = \Phi(y(k)) + \Gamma(y(k)) \cdot u(k) \quad (4.17)$$

em que $y(k)$ e $u(k)$ denotam a saída e entrada (da planta) no k -ésimo instante de tempo.

A descrição expressa na equação (4.17) é uma recomendação técnica feita novamente nos trabalhos de Narendra, Balakrisham e Ciliz.

Se, hipótese improvável, as não linearidades $\Phi(\cdot)$ e $\Gamma(\cdot)$ fossem conhecidas (exatamente), então, o controle $u(k)$ necessário para encontrar a saída desejada, $r(k+1)$, seria calculado em cada instante de tempo por

$$u(k) = \frac{r(k+1) - \Phi(y(k))}{\Gamma(y(k))} \quad (4.18)$$

Entretanto, na prática, $\Phi(\cdot)$ e $\Gamma(\cdot)$ são desconhecidos (ou não importa conhecê-los) e assim que o enfoque é a busca por um modelo de rede neural adaptativa (*wavelets* se adaptando, neste trabalho naturalmente) que se aproxime do sistema dinâmico, isto é,

$$\hat{y}(k+1) = \hat{\Phi}(y(k), \Theta_{\Phi}) + \hat{\Gamma}(y(k), \Theta_{\Gamma}) u(k) \quad (4.19)$$

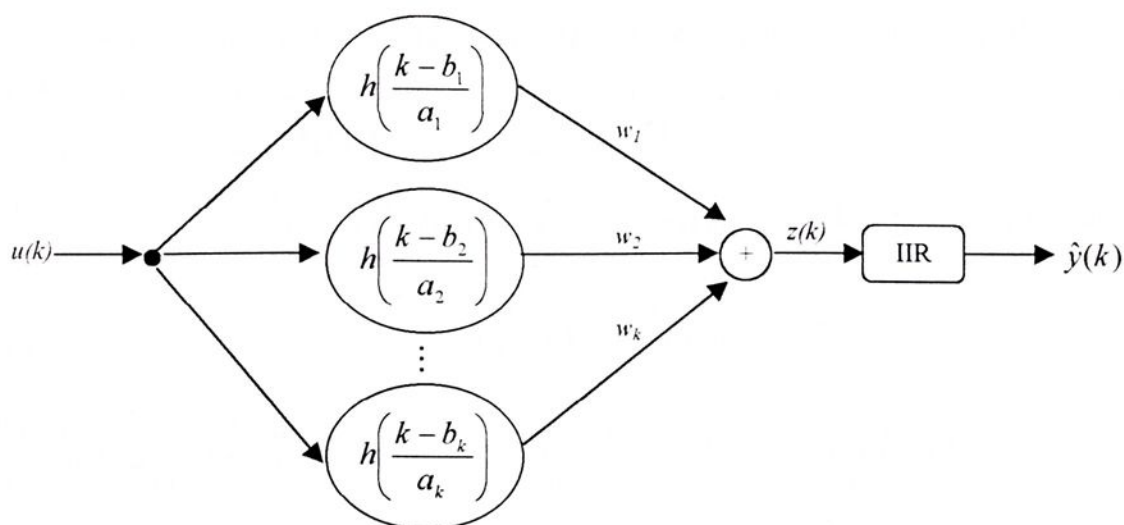


Figura 4.4 - Estrutura da *wavenet* adaptativa

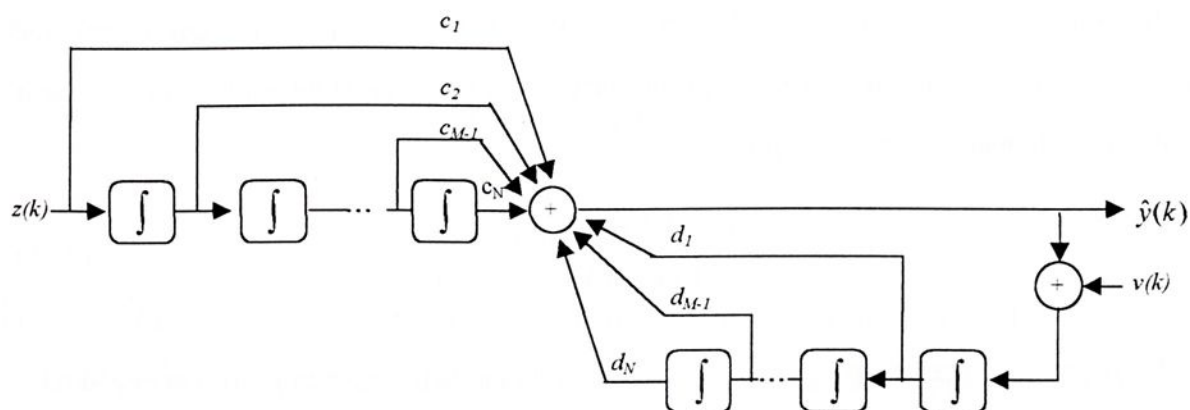


Figura 4.5 - Detalhes do filtro resposta ao impulso infinito (IIR)

Observando que a entrada $u(k)$ excita a rede neural produzindo uma saída $z(k)$, entrada do filtro IIR (Resposta ao Impulso Infinito, detalhado nas figuras 4.4 e 4.5) emergindo $\hat{y}(k)$ modelado por

$$\hat{y}(k) = \sum_{i=0}^M c_i z(k-i) u(k) + \sum_{j=1}^N d_j \hat{y}(k-j) v(k)$$

então, comparando termo a termo, conclui-se que

$$\hat{\Phi}(y(k), \Theta_{\Phi}) = \sum_{j=1}^N d_j \hat{y}(k-j) v(k) \quad (4.20)$$

$$\hat{\Gamma}(y(k), \Theta_{\Gamma}) = \sum_{i=0}^M c_i z(k-i) \quad (4.21)$$

em que

$$z(k) = \sum_{\ell=1}^K w_{\ell} h\left(\frac{k-b_{\ell}}{a_{\ell}}\right) \quad (4.22)$$

e K é o número de janelas *wavelets*. A co-entrada $v(k)$ (figura 4.5) será mantida pequena para garantir estabilidade do processo de aproximação.

Após a substituição das não-linearidades $\Phi(\cdot)$ e $\Gamma(\cdot)$ pelas suas aproximações $\hat{\Phi}(\cdot)$ e $\hat{\Gamma}(\cdot)$ com parâmetros ajustáveis (pesos w_k , escalas a_k , translações b_k , coeficientes feedforward IIR c_k e coeficientes feedback IIR d_k) o controle $u(k)$ necessário para localizar a saída desejada $r(k+1)$ será calculado em todo instante por

$$u(k) = \frac{r(k+1) - \hat{\Phi}(y(k), \Theta_{\Phi})}{\hat{\Gamma}(y(k), \Theta_{\Gamma})} \quad (4.23)$$

em que Θ_{Φ} e Θ_{Γ} sintetizam os parâmetros todos envolvidos no processo de ajuste e adaptação.



CAPÍTULO 5

5.1 RESULTADOS

Os principais resultados e limitações da metodologia são apresentados neste capítulo. Um maior enfoque será dado na discussão e utilização do conceito *wavenet* para a identificação de plantas desconhecidas, visando a sua aplicação posterior no controle de sistemas não lineares. O controle efetivo de uma planta desconhecida, demanda primeiro sua identificação e, a definição posterior de uma lei de controle capaz de produzir a ação desejada. Portanto, a fase de identificação é uma etapa crucial no processo e deve ser perfeitamente entendida. Neste trabalho, a aplicação do conceito *wavenet* será discutida detalhadamente, visando a identificação de sistemas não lineares e posteriormente será apresentado de forma simplificada, uma aplicação em controle.

Várias simulações de identificação foram realizadas utilizando-se sistemas com não linearidade na planta e no sinal de entrada. Os dados mostram uma avaliação do efeito do número de janelas utilizadas na identificação de uma planta, da escolha da família de *wavelets* utilizada como ativadora, bem como o efeito da variação de alguns parâmetros da rede, mais especificamente, as escalas. Finalmente, será discutido a proposta de identificação do algoritmo na identificação e controle de uma planta desconhecida.



5.2 IDENTIFICAÇÃO DE SISTEMAS NÃO LINEARES

Nesta etapa, o algoritmo *wavenet* será utilizado com o propósito de identificação de um sinal desconhecido (uma planta). Três diferentes sinais serão analisados, utilizando *wavelets* das famílias RASP, MORLET e POLYWOG. A influência da escolha do número de janelas *wavelets*, da escala e dos parâmetros da rede serão avaliados buscando definir eventuais critérios para uma boa convergência do processo de minimização.

Os valores iniciais dos coeficientes dos filtros e os de ponderação foram fixos, as janelas foram inicialmente igualmente espaçadas, ou seja, as *wavelets* foram deslocadas ao longo de todo o sinal de modo a cobri-lo inteiramente.

5.2.1 CASO 1

O primeiro sinal a ser tratado, refere-se a uma planta com não linearidade dinâmica, excitada com uma entrada linear. A expressão (5.1) representa tal planta através de uma equação de diferença não linear.

$$y(n) = \frac{y(n-1)}{1 + y^2(n-1)} + u(n) \quad (5.1)$$

em que $u(n) = 0.8 \sin(2\pi n/50)$, $n = 1, 2, \dots, 200$.

Inicialmente, a identificação da planta foi feita utilizando-se 25 janelas *wavelets* POLYWOG5 igualmente espaçadas. Os valores iniciais da escala e dos coeficientes de ponderação e filtros são mostrados abaixo:

- escalas $a(j) = 5$, $j = 1, 2, \dots, 25$
- pesos $w(j) = 1e - 6$, $j = 1, 2, \dots, 25$
- coeficientes de filtro de retroalimentação $d(1) = d(2) = d(3) = 0,01$
- coeficientes de filtro *feedforward* $c(1) = c(2) = c(3) = 0,01$

A Figura 5.1 mostra o sinal original a ser identificado.

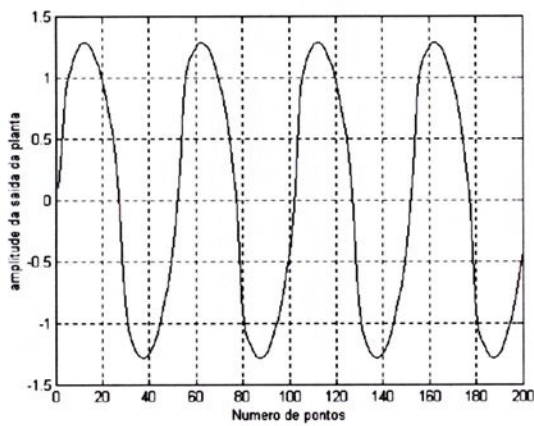
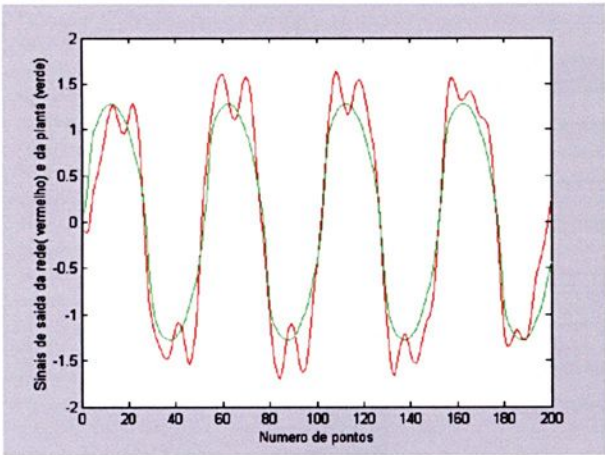


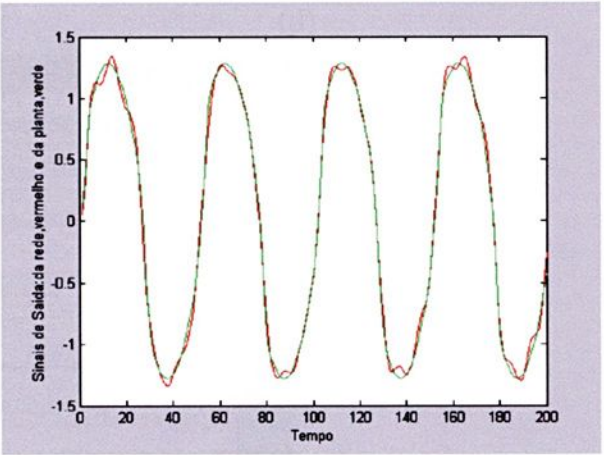
Figura 5.1 – Saída da planta

Os resultados do processo de identificação da planta são discutidas abaixo.

As Figuras 5.2.a e 5.2.b mostram a saída da planta e a saída da rede, após 6 e 38 iterações, respectivamente. Após 38 iterações a rede convergiu e o erro médio quadrático final observado foi de 0,0018. As Figuras 5.3.a e 5.3.b. mostram o erro médio quadrático ao longo do processo.

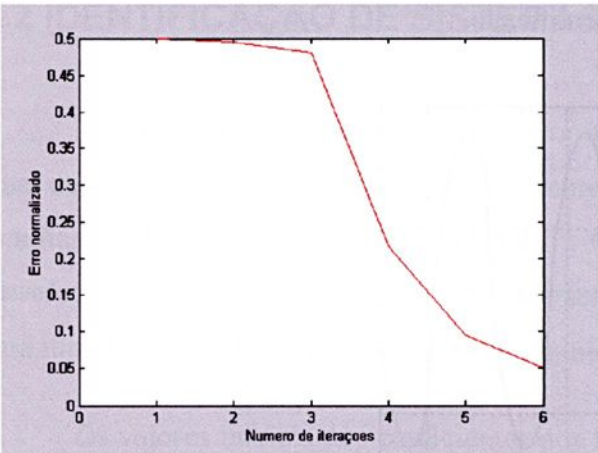


(a)

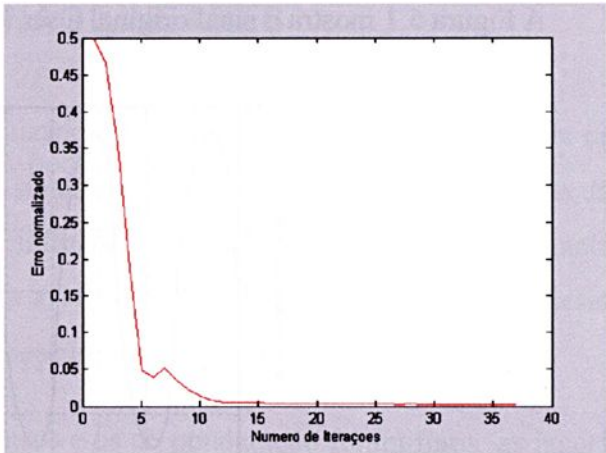


(b)

Figura 5.2 – Saída da rede (sinal aproximante, vermelho) saída da planta (verde)



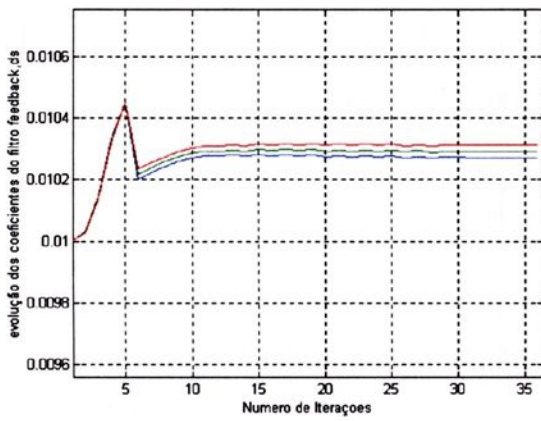
(a)



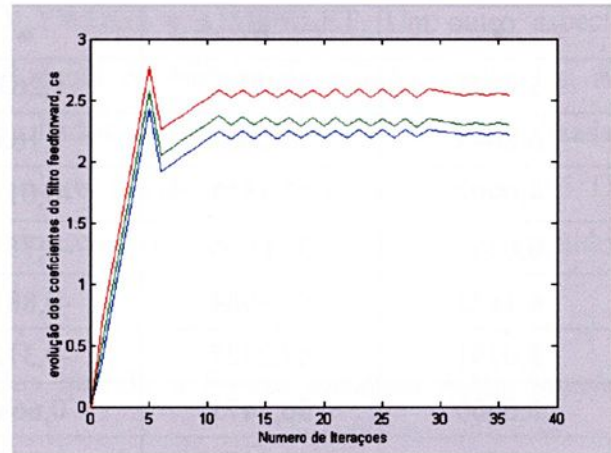
(b)

Figura 5.3 – (a) Erro após 6 iterações (b) Erro após a parada do processo de minimização

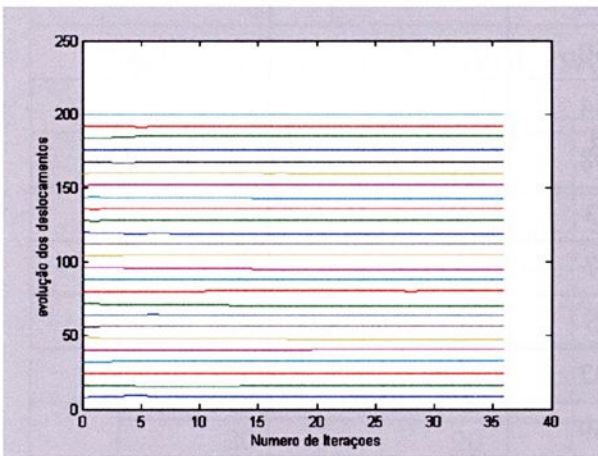
A evolução dos parâmetros da rede ao longo do processo de minimização é apresentada nas figuras seguintes. Os valores finais dos parâmetros da rede (após a convergência) estão na tabela 1.



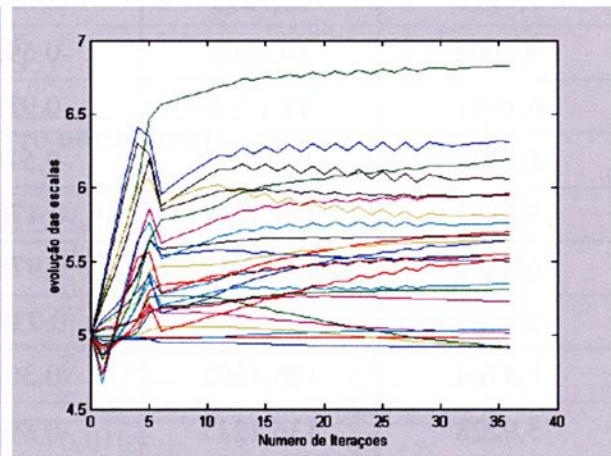
(a)



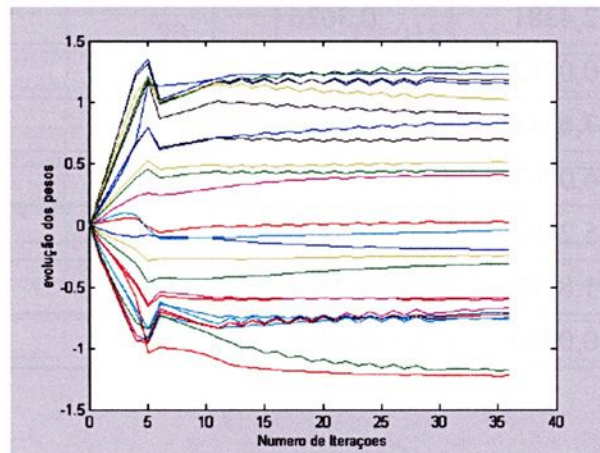
(b)



(c)



(d)



(e)

Figura 5.4 – Evolução dos vários parâmetros no decorrer do processo de minimização.

Tabela 1 – Valores finais dos parâmetros envolvidos.

a	b	w	c	d
5,4928	8,4982	1,2679	2,4286	0,0105
6,8043	15,2467	1,1653	2,3000	0,0105
4,9992	23,8523	-0,0147	2,4974	0,0104
6,6657	32,6066	-0,7513		
6,1875	40,3984	-0,8813		
5,0791	47,2127	-0,3129		
5,6600	56,7471	0,6643		
6,3606	64,2408	0,9453		
5,0576	70,9265	0,4015		
5,7002	80,6206	-0,5950		
6,1941	88,1718	-0,9384		
5,2274	94,9267	-0,5798		
5,6143	104,5818	0,4763		
6,0860	112,1212	0,9757		
5,4687	119,0488	0,7395		
5,3764	128,4502	-0,3637		
5,9228	136,1235	-0,8969		
5,5814	143,2624	-0,7833		
5,2603	152,4381	0,3626		
5,7978	160,0762	0,9417		
5,8613	167,6311	0,8762		
4,9574	176,0961	-0,1319		
6,3078	185,2402	-1,1467		
5,5235	191,8382	-1,1856		
4,9686	200,0880	0,0268		

Para avaliar a dependência da escolha da família *wavelet* nos resultados foram feitas outras simulações, utilizando a RASP1, a POLYWOG5 e a MORLET. Um outro aspecto bastante importante também avaliado foi a influência do número de *wavelets* utilizadas. As tabelas a seguir mostram sinopticamente os resultados destes vários testes. Foram utilizadas redes com 5, 15, 20, 25, 30 e 35 neurônios e os valores iniciais das escalas foram 20, 10 e 5. Os demais parâmetros (pesos, taxa de aprendizado, coeficientes de filtros), foram mantidos inalterados.

A tabela 2 resume os resultados obtidos quando a função ativadora foi a *wavelet* POLYWOG5.

Tabela 2 - Função ativadora (POLYWOG5)

Nº de neurônios (K)	Escalas (a)	Erro normalizado	Nº de iterações
5	20	0,0145	62
5	10	0,4999	4
5	5	0,1881	62
15	20	0,0143	51
15	10	0,0154	72
15	5	0,0207	122
20	20	0,0109	65
20	10	0,0038	65
20	5	0,0050	102
25	20	0,0155	52
25	10	0,0046	54
25	5	0,0018	38
30	20	0,3421	+335
30	10	0,4151	+201
30	5	0,0017	54
35	20	0,0212	38
35	10	0,0059	40
35	5	0,0024	46

A tabela 2 mostra que para esta planta não linear, o número de neurônios mais apropriado sob o ponto de vista de convergência, erro menor, acontece em torno de 25 neurônios. Os erros normalizados são de mesma ordem, uma vez que com o crescimento do número de neurônios o erro sofre uma forte influência dos valores iniciais das escalas. Sob o ponto de vista do tempo de computação (número de iterações necessárias para convergência), os resultados mostram não ser

possível estabelecer uma correlação entre variações para mais ou para menos no número de neurônios.

De modo geral, os resultados para $K=5$ ou $K=15$ janelas mostram que estes valores não são recomendados para análise do sinal da equação (5.1) em virtude da ordem de grandeza dos erros.

A tabela 3 resume os resultados obtidos quando a função ativadora foi a *wavelet* RASP1.

Tabela 3 - Função ativadora (RASP1)

Nº de neurônios (K)	Escala (a)	Erro normalizado	Nº de iterações
5	20	0,4999	4
5	10	0,4999	3
5	5	0,0097	62
15	20	0,0104	65
15	10	0,0027	142
15	5	0,0120	147
20	20	0,0092	56
20	10	0,0031	63
20	5	0,0095	141
25	20	0,0085	94
25	10	0,0052	56
25	5	0,0042	141
30	20	0,4999	3
30	10	0,4308	+350
30	5	0,4999	4
35	20	0,4999	4
35	10	0,4364	+330
35	5	0,4999	4

As tabelas 2 e 3 mostram que redes ativadas por $K=20$ ou $K=25$ janelas (POLYWOG5 ou RASP1) são igualmente promissoras sob o ponto de vista da ordem de grandeza do erro, e no caso da opção pela janela POLYWOG5, as escalas interessantes são 5 e 10; e no caso da opção pela janela RASP1, ganha-se maior flexibilização nas escalas. O processo de identificação oferece bons resultados se a rede neural adaptativa for ativada com $K=20$ ou $K=25$ janelas POLYWOG5, observadas algumas restrições nas escalas. Algumas combinações de número de

neurônios e escalas podem provocar a parada do processo de minimização logo no início ou ao contrário arrastá-lo ao infinito. Quaisquer dessas situações não produzem resultados aceitáveis. Essas situações indesejáveis são observáveis nas duas tabelas acima citadas. Em se tratando da análise do sinal da equação (5.1) a diferença marcante entre as duas redes ativadas é a possibilidade de reconfigurar redes com apenas 5 ou 15 janelas RASP1 e ainda obter bons resultados na aproximação da saída da planta com a saída da rede.

A tabela 4 resume os resultados obtidos quando a função ativadora foi a *wavelet* MORLET.

Tabela 4 - Função ativadora (MORLET)

Nº de neurônios (K)	Escalas (a)	Erro normalizado	Nº de iterações
5	20	0,4805	67
5	10	0,5000	2
5	5	0,5000	2
15	20	0,0103	143
15	10	0,2598	520
15	5	0,4770	36
20	20	0,0129	108
20	10	0,01262	100
20	5	0,4993	11
25	20	0,0132	96
25	10	0,0147	141
25	5	0,3641	237
30	20	0,4586	67
30	10	0,4999	2
30	5	0,4999	2
35	20	0,4557	67
35	10	0,4999	2
35	5	0,4999	2

De modo geral, os resultados obtidos com a utilização da *wavelet* MORLET para a identificação do sinal de equação (5.1), mostraram-se inferiores aos obtidos com as *wavelets* POLYWOG5 e RASP1. Também não foi possível observar diferença entre as várias opções do número de janelas e ou escalas, isto é, quaisquer combinações produzem identificação pobre se comparada com as anteriores, em virtude da ordem de grandeza do erro médio quadrático.

Objetivando discutir a robustez do processo de identificação quanto a contaminação do sinal com ruído, foram realizados vários testes utilizando o sinal original contaminado com ruído. Na Figura 5.5 é mostrado o desempenho do algoritmo na identificação do sinal de equação (5.1) corrompido por ruído (variância 0,05, uniformemente distribuído). A Figura 5.5.a mostra o sinal de saída da planta com ruído na entrada (verde) e a saída da rede (vermelho). A Figura 5.5.b ilustra o decaimento do erro médio quadrático. Após vinte iterações o erro atinge o valor 0,0050.

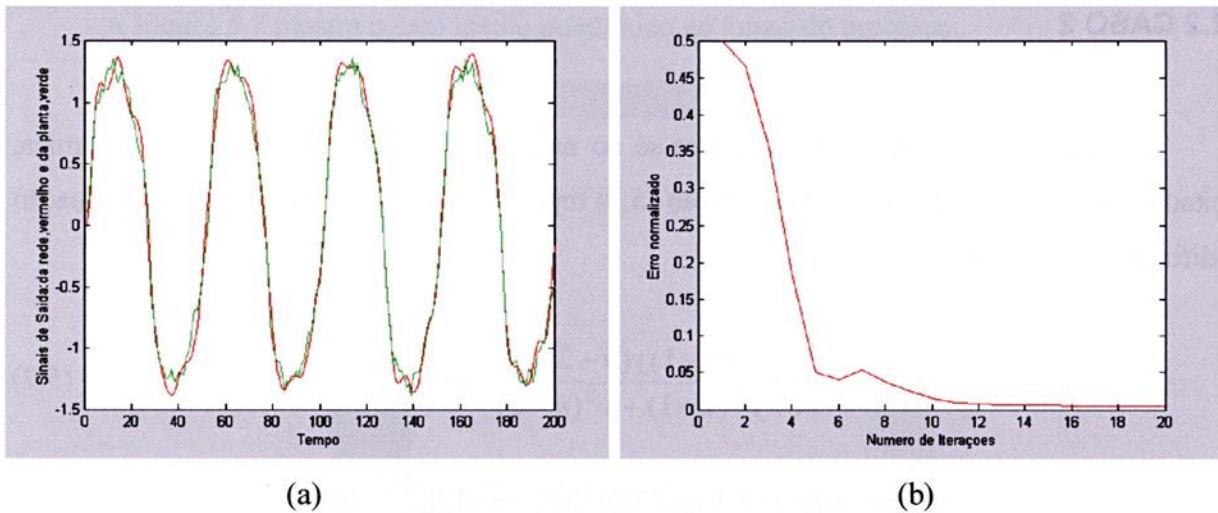


Figura 5.5 – Resultados da simulação feita com sinal ruidoso

A simulação ilustrada na Figura 5.5 foi realizada com $K=25$ janelas POLYWOG5 e escalas inicializadas em $a=5$. No caso da planta não contaminada por ruído e, mantendo todas as outras informações inalteradas, a identificação ocorreu com erro de 0,0018, exigindo 38 iterações, conforme pode ser lido na tabela 2.

5.2.2 CASO 2

O segundo sinal a ser tratado, refere-se a uma planta com não linearidade dinâmica, excitada com uma entrada linear. A expressão (5.2) representa tal planta através de uma equação de diferença não linear.

$$y(n) = \frac{y(n-1)y(n-2)}{1 + y^2(n-1) + y^2(n-2)} + u(n) \quad (5.2)$$

em que $u(n) = 0.8 \sin(2\pi n/50)$, $n = 1, 2, \dots, 200$.

Inicialmente, a identificação da planta foi feita utilizando-se 30 janelas *wavelets* RASP1 igualmente espaçadas. Os valores iniciais da escala e dos coeficientes de ponderação e filtros são mostrados abaixo:

- escalas $a(j) = 5$, $j = 1, 2, \dots, 30$
- pesos $w(j) = 1e-6$, $j = 1, 2, \dots, 30$
- coeficientes de filtro de retroalimentação $d(1) = d(2) = d(3) = 0,01$
- coeficientes de filtro feedforward $c(1) = c(2) = c(3) = 0,01$
- taxa de aprendizado é de 10^{-2} .

A Figura 5.6 mostra a saída da planta e a saída da rede, após 64 iterações, quando a rede convergiu.

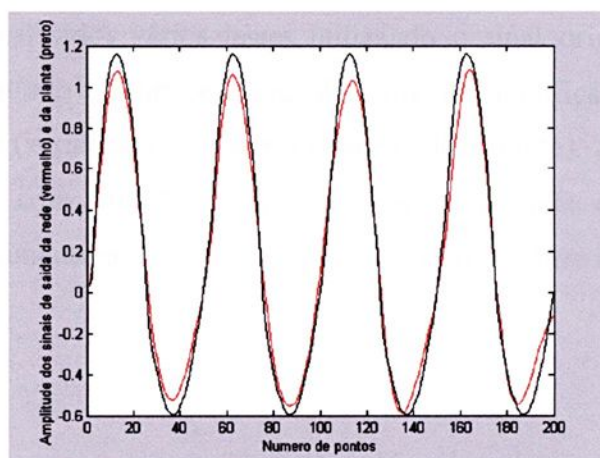


Figura 5.6 – Saída da rede (vermelho) saída da planta (preto)

A Figura 5.7 mostra o erro médio quadrático ao longo do processo.

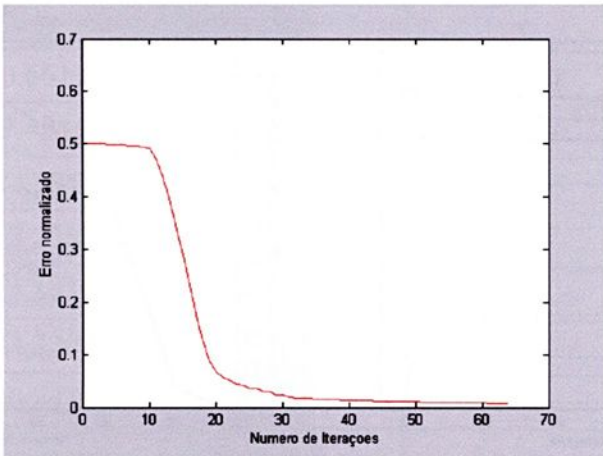


Figura 5.7 – Erro após a parada do processo de minimização

A evolução dos parâmetros da rede ao longo do processo de minimização é apresentada nas figuras seguintes. Os valores finais dos parâmetros da rede (após a convergência) estão na tabela 5.

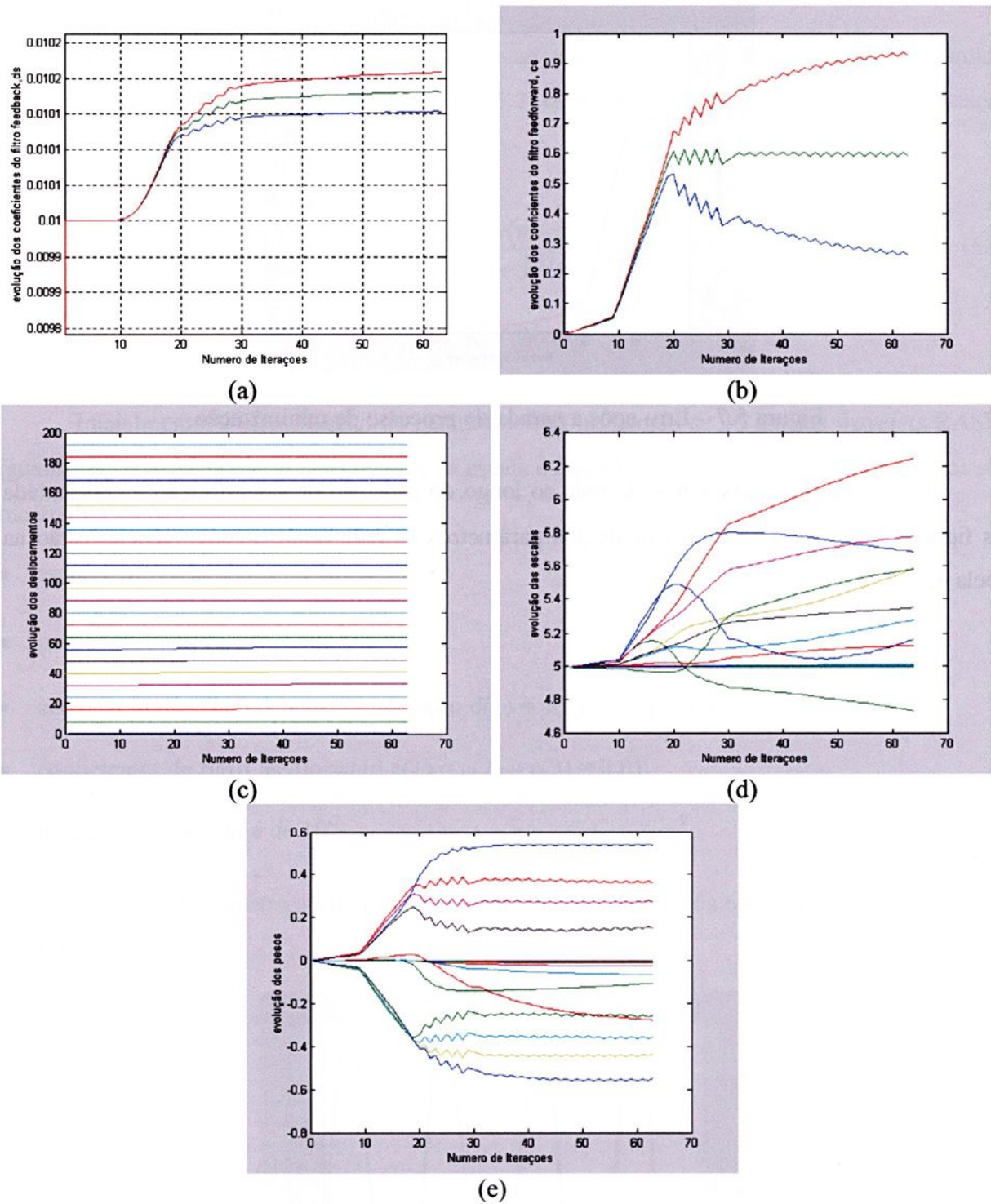


Figura 5.8 – Evolução dos vários parâmetros no decorrer do processo de minimização.

Tabela 5 – Valores finais dos parâmetros envolvidos.

a	b	w	c	d
6,0419	0,6613	0,5496	0,8528	0,0102
5,0993	7,4835	-0,2889	0,5700	0,0102
5,2482	14,2855	0,1863	0,3395	0,0101
4,6682	18,2381	0,3395		
5,6209	24,3457	-0,3335		
6,5506	34,2208	0,3125		
6,1000	40,8455	-0,4556		
5,2628	47,0556	0,0308		
5,3527	51,5017	0,2893		
7,0085	57,5518	-0,5080		
4,7576	66,7663	0,0401		
5,3421	73,1336	-0,4132		
5,0360	79,9916	-0,0919		
5,0044	86,6660	-0,0309		
5,0006	93,3332	-0,0138		
4,9999	100,0000	-0,0073		
4,9998	106,6667	-0,0044		
4,9998	113,3333	-0,0028		
4,9998	120,0000	-0,0020		
4,9998	126,6667	-0,0014		
4,9998	133,3333	-0,0010		
4,9999	140,0000	-0,0008		
4,9999	146,6667	-0,0006		
4,9999	153,3333	-0,0005		
4,9999	160,0000	-0,0004		
4,9999	166,6667	-0,0003		
4,9999	173,3333	-0,0003		
4,9999	180,0000	-0,0002		
4,9999	186,6667	-0,0002		
5,0000	193,3333	-0,0002		

Para avaliar a dependência da escolha da família *wavelet* nos resultados também foram feitas outras simulações, utilizando a RASP1, a POLYWOG5 e a MORLET. Um outro aspecto avaliado foi a influência do número de *wavelets* utilizadas. Foram utilizadas redes com 5, 15, 20, 25, 30 e 35 neurônios e escalas inicializadas em 20, 10 e 5. Os demais parâmetros (pesos, taxa de aprendizado, coeficientes de filtros), foram mantidos inalterados.

A tabela 6 resume os resultados obtidos quando a função ativadora foi a *wavelet* RASP1 e a tabela 7 quando a ativação ocorreu pelo emprego da POLYWOG5

Tabela 6 - Função ativadora (RASP1)

Nº de neurônios (K)	Escalas (a)	Erro normalizado	Nº de iterações
5	20	0,4728	18
5	10	0,3619	352
5	5	0,3619	352
15	20	0,2717	152
15	10	0,1074	69
15	5	0,2449	22
20	20	0,3832	38
20	10	0,0873	166
20	5	0,0311	318
25	20	0,1877	154
25	10	0,0356	27
25	5	0,0093	86
30	20	0,2583	155
30	10	0,0161	126
30	5	0,0008	138
35	20	0,5001	4
35	10	0,0154	36
35	5	0,0076	36

Tabela 7 - Função ativadora (POLYWOG5)

Nº de neurônios (K)	Escalas (a)	Erro normalizado	Nº de iterações
5	20	0,4509	65
5	10	0,2675	112
5	5	0,2152	112
15	20	0,0430	16
15	10	0,0179	314
15	5	0,0146	245
20	20	0,4312	15
20	10	0,0196	207
20	5	0,0013	175
25	20	0,4346	12
25	10	0,0443	75
25	5	0,0171	149
30	20	0,4341	12
30	10	0,0121	52
30	5	0,0690	50
35	20	0,4359	10
35	10	0,0119	52
35	5	0,0043	72

Observamos que a *wavelet* RASP1 oferece bons resultados, em vista da ordem de grandeza dos erros se a rede for configurada com K=25, K=30 ou K=35 janelas e escalas de inicialização a=5 (isto é, cada *wavelet*-filha compondo a estrutura neural inicia com escala 5). Por outro lado, a *wavelet* POLYWOG5 oferece bons resultados com as opções (testadas) K=20 ou K=35.

A tabela 8 resume os resultados obtidos quando a função ativadora foi a *wavelet* MORLET.

Tabela 8 - Função ativadora (MORLET)

Nº de neurônios (K)	Escalas (a)	Erro normalizado	Nº de iterações
5	20	0,0682	+404
5	10	0,0835	+578
5	5	0,1163	+301
15	20	0,0504	+248
15	10	0,0398	+294
15	5	0,0345	+201
20	20	0,0851	558
20	10	0,0510	227
20	5	0,5001	4
25	20	0,0455	501
25	10	0,0693	132
25	5	0,0422	387
30	20	0,0528	487
30	10	0,0451	98
30	5	0,0534	523
35	20	0,3074	64
35	10	0,0385	+204
35	5	0,0364	+217

Uma leitura da tabela 8 mostra, em virtude da ordem de grandeza dos erros normalizados que a utilização da família MORLET como ativadora da rede produz saída da rede muito distante (o erro médio quadrático é de fato uma medida de distância) da saída da planta.

5.2.3 CASO 3

O terceiro sinal, a ser tratado, refere-se a uma planta com não linearidade na entrada. A expressão (5.3) representa tal planta através da equação de diferença

$$y(n) = 0,3y(n-1) - 0,6y(n-2) + f(u(n)) \quad (5.3)$$

em que $u(n) = \sin(2\pi n / 250)$, $n = 1, \dots, 200$. e $f(u) = u^3 + 0,3u^2 - 0,4u$.

A Figura 5.9 mostra o sinal a ser identificado.

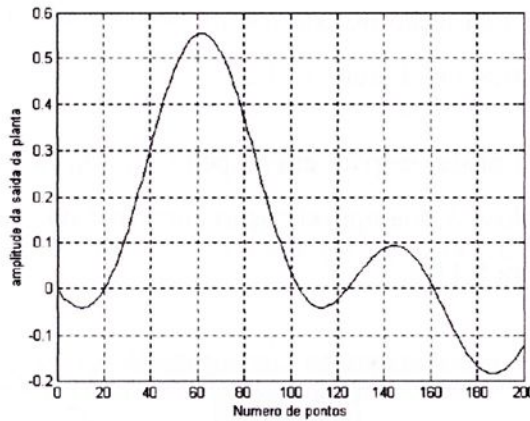


Figura 5.9 – Saída da planta

Inicialmente, a identificação da planta foi feita utilizando-se 32 janelas *wavelets* POLYWOG3 igualmente espaçadas. Os valores iniciais da escala e dos coeficientes de ponderação e filtros são mostrados abaixo:

- escalas $a(j) = 100$, $j = 1, 2, \dots, 22$
- pesos $w(j) = 1e - 6$, $j = 1, 2, \dots, 22$
- coeficientes de filtro de retroalimentação $d(1) = d(2) = d(3) = 0,01$
- coeficientes de filtro *feedforward* $c(1) = c(2) = 0,01$
- taxa de aprendizado é 10^{-2} .

A segunda iteração nesta simulação é ilustrada na Figura 5.10.

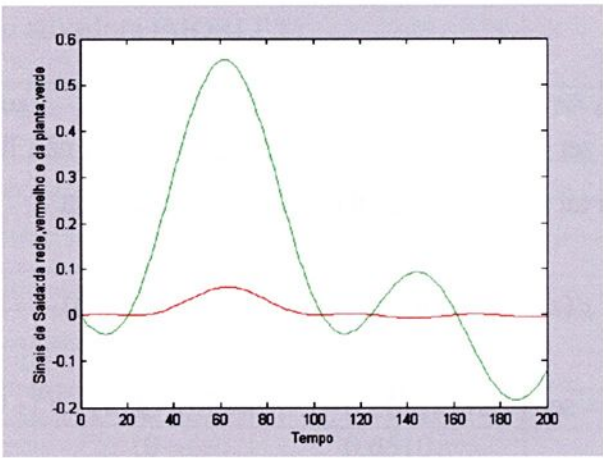


Figura 5.10 – Saída da planta (verde) saída da rede (vermelho)

Na vigésima iteração, a situação de confronto entre a saída da rede e a saída da planta dada pela equação (5.3) é ilustrada na Figura 5.11.a.

Após a 108ª iteração o processo pára em respeito ao critério de nulidade computacional de um dos gradientes envolvidos. A boa aproximação entre as saídas da planta (verde) e da rede (vermelho) é ilustrada na Figura 5.11.b.

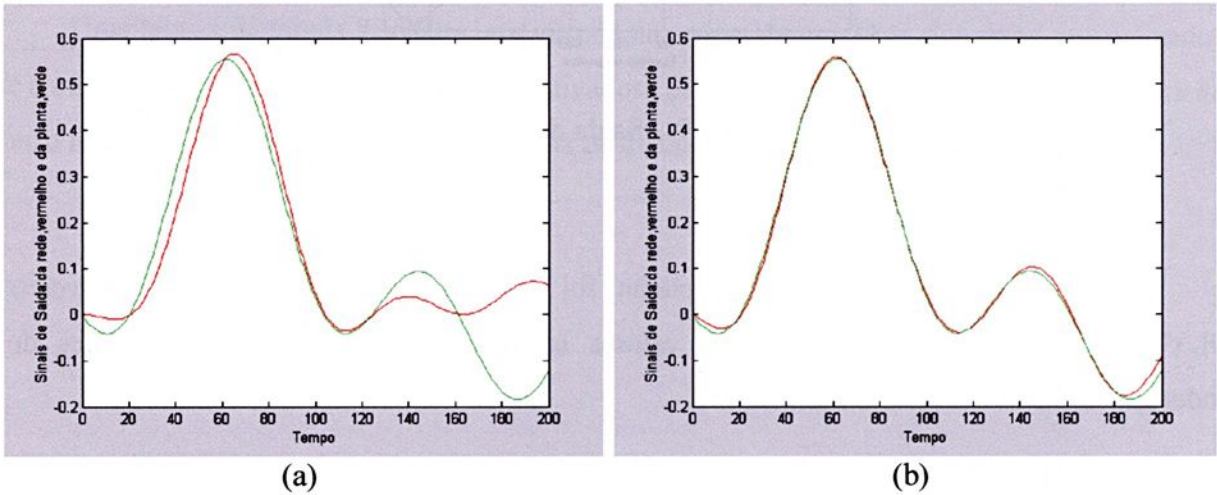


Figura 5.11 – Saída da rede (vermelho) e saída da planta (verde)

A Figura 5.12.a ilustra o comportamento do erro médio quadrático nas primeiras vinte iterações e a Figura 5.12.b ilustra do início até final do mesmo processo de minimização.

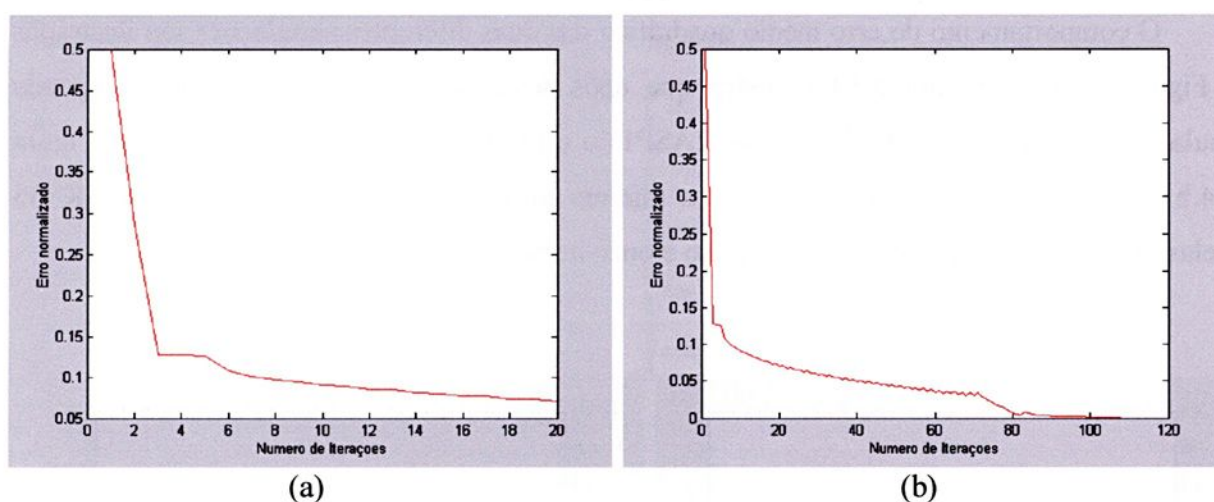


Figura 5.12 – Erros quadrático médio

Após cento e oito interações o processo convergiu e o erro médio quadrático observado foi de 0,0007, sendo que nas primeiras vinte iterações o erro decaiu de 0,5 para 0,07.

O sinal com não linearidade na entrada foi igualmente bem identificado tanto pela POLYWOG5 quanto pela RASP1.

A Figura 5.13.a ilustra o confronto entre a saída da rede e a saída da planta após sessenta iterações de uma outra simulação elaborada com $K=22$ janelas *wavelet* RASP1. A Figura 5.14.b ilustra o mesmo confronto após cento e onze iterações da primeira simulação já mencionada.

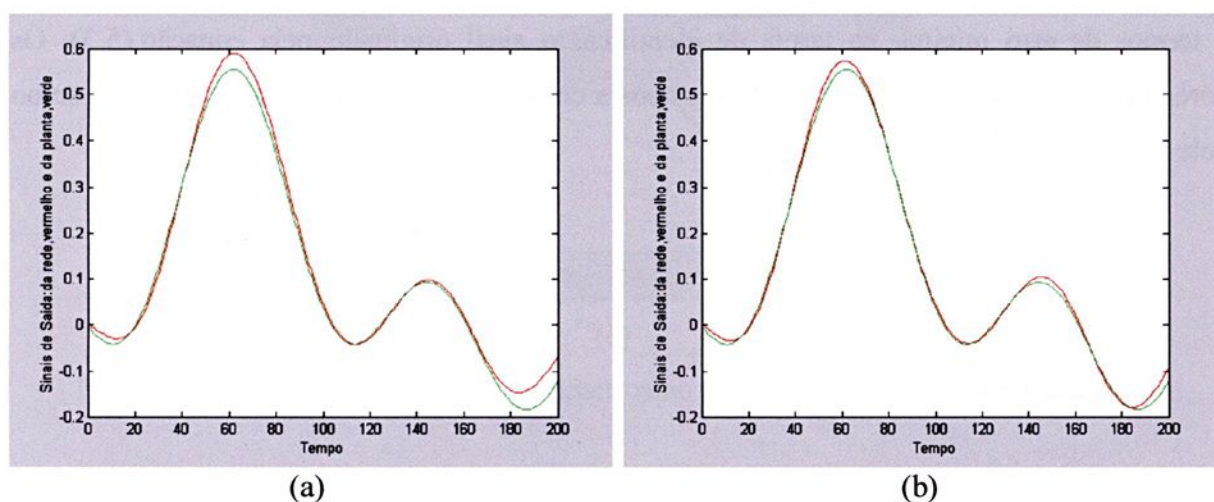


Figura 5.13 – Saída da planta (verde) saída da rede (vermelho)

O comportamento do erro médio quadrático das duas diferentes simulações são ilustradas na Figura 5.14. A Figura 5.14.a ilustra que após sessenta iterações que levou a segunda simulação a convergência (K=22 janelas RASP1) o erro atinge o mínimo de 0,0039. A Figura 5.14.b ilustra o decaimento do erro para o mínimo de 0,0007 na primeira simulação (K=35 janelas POLYWOG5), neste caso após cento e onze iterações.

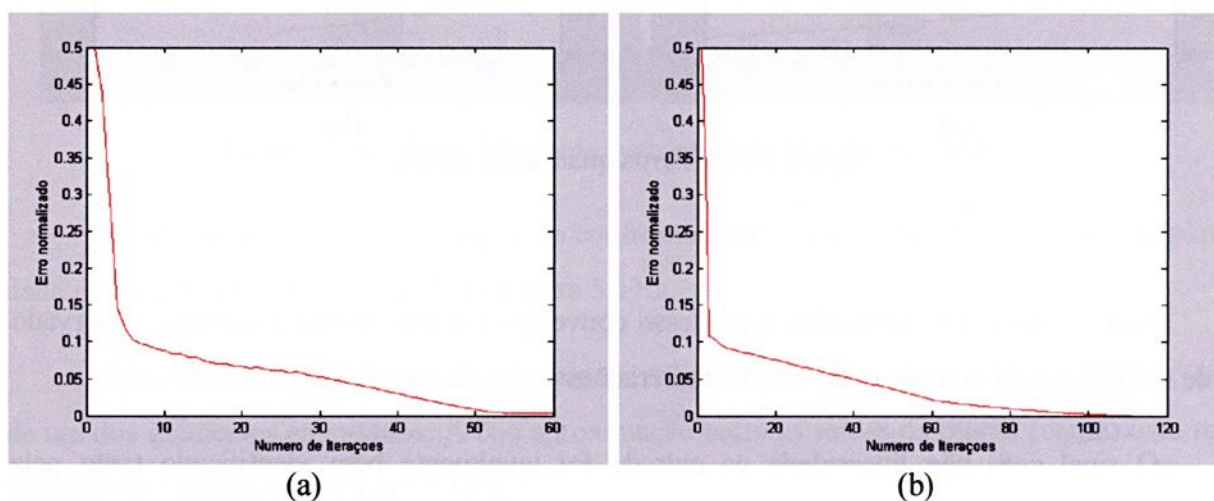
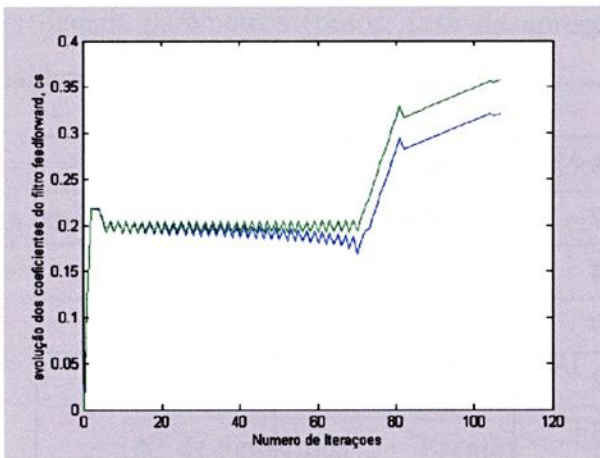
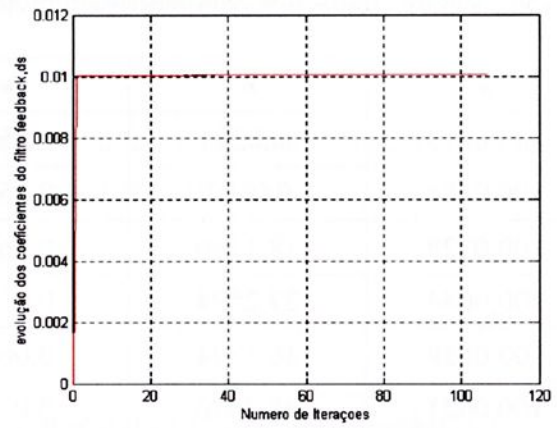


Figura 5.14 – Erros quadráticos médios

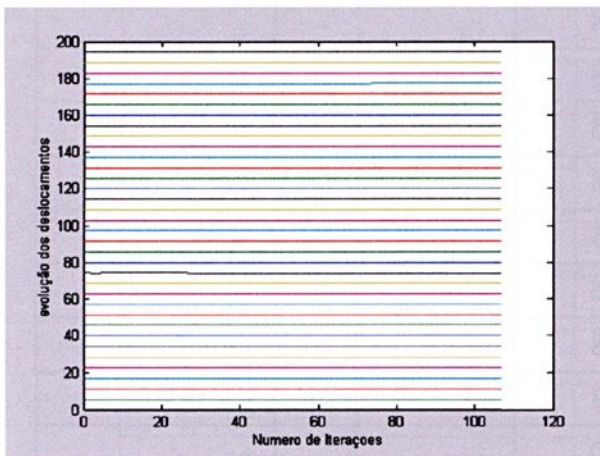
As Figuras 5.15.a, 5.15b, 5.15c, 5.15d, 5.15e ilustram as alterações sofridas pelos vários parâmetros no decorrer do primeiro processo de minimização, utilizando-se K=35 janelas *wavelet* POLYWOG5 simulação que exigiu cento e oito iterações, um dos melhores resultados em termos de erro mínimo na tarefa de identificar o sinal originado pela equação (5.3). Os valores finais dos vários parâmetros obtidos após a convergência da simulação encontram-se no tabela 9.



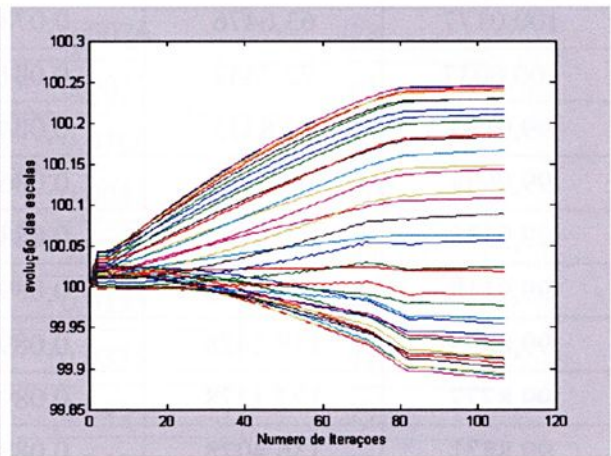
(a)



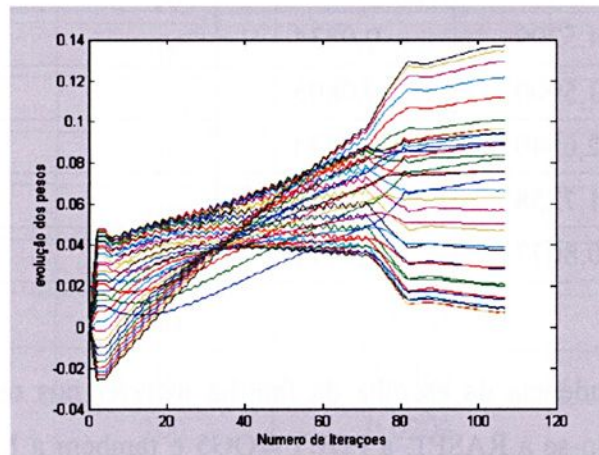
(b)



(c)



(d)



(e)

Figura 5.15 – Evolução do vários parâmetros no decorrer do processo de minimização.

Tabela 9 – Valores finais dos parâmetros envolvidos.

a	b	w	c	d
100,0819	-0,0284	0,0544	0,3805	0,0101
100,0784	9,0623	0,0587	0,3709	0,0101
100,0728	18,1596	0,0624		0,0101
100,0644	27,2594	0,0659		
100,0538	36,3594	0,0692		
100,0421	45,4570	0,0725		
100,0301	54,5526	0,0757		
100,0177	63,6476	0,0788		
100,0037	72,7442	0,0818		
99,9863	81,8435	0,0845		
99,9644	90,9456	0,0868		
99,9385	100,0485	0,0885		
99,9118	109,1488	0,0896		
99,8895	118,2428	0,0898		
99,8777	127,3278	0,0893		
99,8821	136,4028	0,0880		
99,9053	145,4690	0,0860		
99,9461	154,5299	0,0834		
99,9992	163,5900	0,0805		
100,0562	172,6540	0,0774		
100,1078	181,7258	0,0742		
100,1452	190,8077	0,0711		

Para avaliar a dependência da escolha da família *wavelet* nos resultados, foram feitas outras simulações, utilizando-se a RASP1, a POLYWOG5 e também a MORLET. Um aspecto bastante importante avaliado foi a influência do número de janelas utilizadas (número de neurônios) e também como o algoritmo reage com as alterações das escalas.

As tabelas 10 e 11 mostram sinopticamente os resultados dos vários testes. Foram utilizadas redes com 5, 15, 20, 25, 30 e 35 neurônios e escalas de valores iniciais 200, 100 e 50.

Os demais parâmetros (pesos, taxa de aprendizado, coeficientes de filtros), foram mantidos inalterados.

A tabela 10 resume os resultados obtidos quando a função ativadora foi a *wavelet* RASP1 e a tabela 11 quando a ativação ocorreu pelo emprego da POLYWOG5.

Tabela 10 – Função ativadora RASP1

Nº de neurônios (K)	Escalas (a)	Erro normalizado	Nº de iterações
5	50	0,0026	78
5	100	0,0037	61
5	200	0,0160	31
15	50	0,0053	108
15	100	0,0392	33
15	200	0,0153	88
20	50	0,0335	41
20	100	0,0424	29
20	200	0,0199	22
25	50	0,0410	34
25	100	0,0421	28
25	200	0,0196	21
30	50	0,0431	31
30	100	0,0449	25
30	200	0,0198	20
35	50	0,0458	28
35	100	0,0463	24
35	200	0,0203	19

Tabela 11 - Função ativadora POLYWOG5.

Nº de neurônios (K)	Escalas (a)	Erro normalizado	Nº de iterações
5	50	0,0672	404
5	100	0,0770	131
5	200	0,0131	6
15	50	0,0609	19
15	100	0,0638	+189
15	200	0,0094	19
20	50	0,0607	18
20	100	0,0574	+219
20	200	0,0094	18
25	50	0,0610	18
25	100	0,0555	+206
25	200	0,0082	29
30	50	0,0611	18
30	100	0,0479	+203
30	200	0,0073	27
35	50	0,0423	+179
35	100	0,0023	108
35	200	0,0087	27

A *wavelet* POLYWOG5 configurando a rede com K=15, 20, 25 , 30 ou 35 janelas oferecem aproximações entre as saídas da rede e da planta com erro médio quadrático da ordem de 10^{-3} . As boas escalas inicializantes são 100 ou 200. Por outro lado, a *wavelet* RASP1 é mais promissora configurando a rede com K=5 ou 15 janelas. Para K=5, as escalas aceitáveis são 50 ou 100. Podemos saudar resultados bons com relativamente pouco neurônios .

A tabela 12 resume os resultados obtidos quando a função ativadora foi a *wavelet* MORLET.

Tabela 12 - Função ativadora MORLET.

Nº de neurônios (K)	Escala (a)	Erro normalizado	Nº de iterações
5	50	0,07590	319
5	100	0,0622	56
5	200	0,0697	47
15	50	0,0313	139
15	100	0,0602	32
15	200	0,0708	30
20	50	0,0309	116
20	100	0,0602	29
20	200	0,0690	31
25	50	0,0307	107
25	100	0,0602	27
25	200	0,0706	25
30	50	0,0304	107
30	100	0,0606	25
30	200	0,0709	27
35	50	0,0303	108
35	100	0,0608	24
35	200	0,0710	25

De modo geral, os resultados obtidos com a utilização da *wavelet* MORLET para a identificação do sinal (5.3), mostraram-se inferiores aos obtidos com as *wavelets* POLYWOG5 e RASP1. Também não foi possível observar uma nítida diferença entre as várias opções do número de janelas e ou escalas, isto é, quaisquer combinações produzem identificação pobre em virtude da ordem de grandeza do erro médio quadrático observado no tabela 12.

5.3. CONTROLE ADAPTATIVO

Este tópico discute a possibilidade da utilização do conceito *wavenet* na identificação e controle de uma planta desconhecida. A proposta neste caso, é inicialmente identificar a planta em uma fase de reconhecimento que será feita *off line* e posteriormente, exercer uma ação sobre ela.

O exemplo avaliado refere-se a um sistema não linear, em que a não linearidade ocorre na dinâmica e a excitação de entrada, neste caso é linear. O modelo é representado pela equação (5.1)

$$y(n) = \frac{y(n-1)}{1+y^2(n-1)} + u(n).$$

A Figura 5.16 ilustra a fase de identificação, o sinal azul representa a saída da planta e o preto a saída da rede, ou seja, a planta identificada.

A primeira etapa da avaliação do sistema, é a etapa da identificação e a segunda etapa é a do controle da planta.

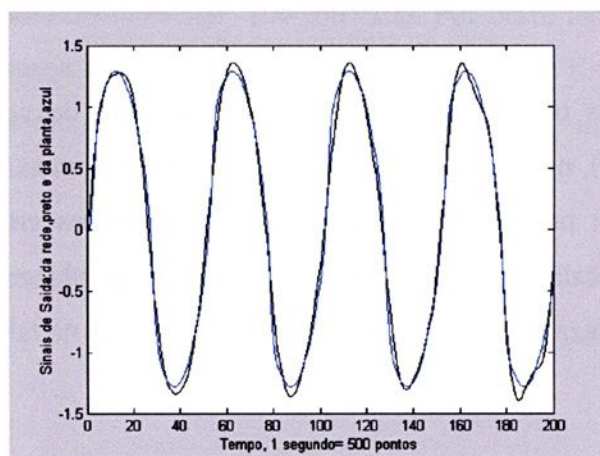


Figura 5.16 – Saída da rede aproximante (preto) saída da planta (azul)

A Figura 5.17 ilustra a evolução dos parâmetros na fase de identificação com taxa de aprendizado 10^{-3} , utilizando-se 15 *wavelets* RASP1.

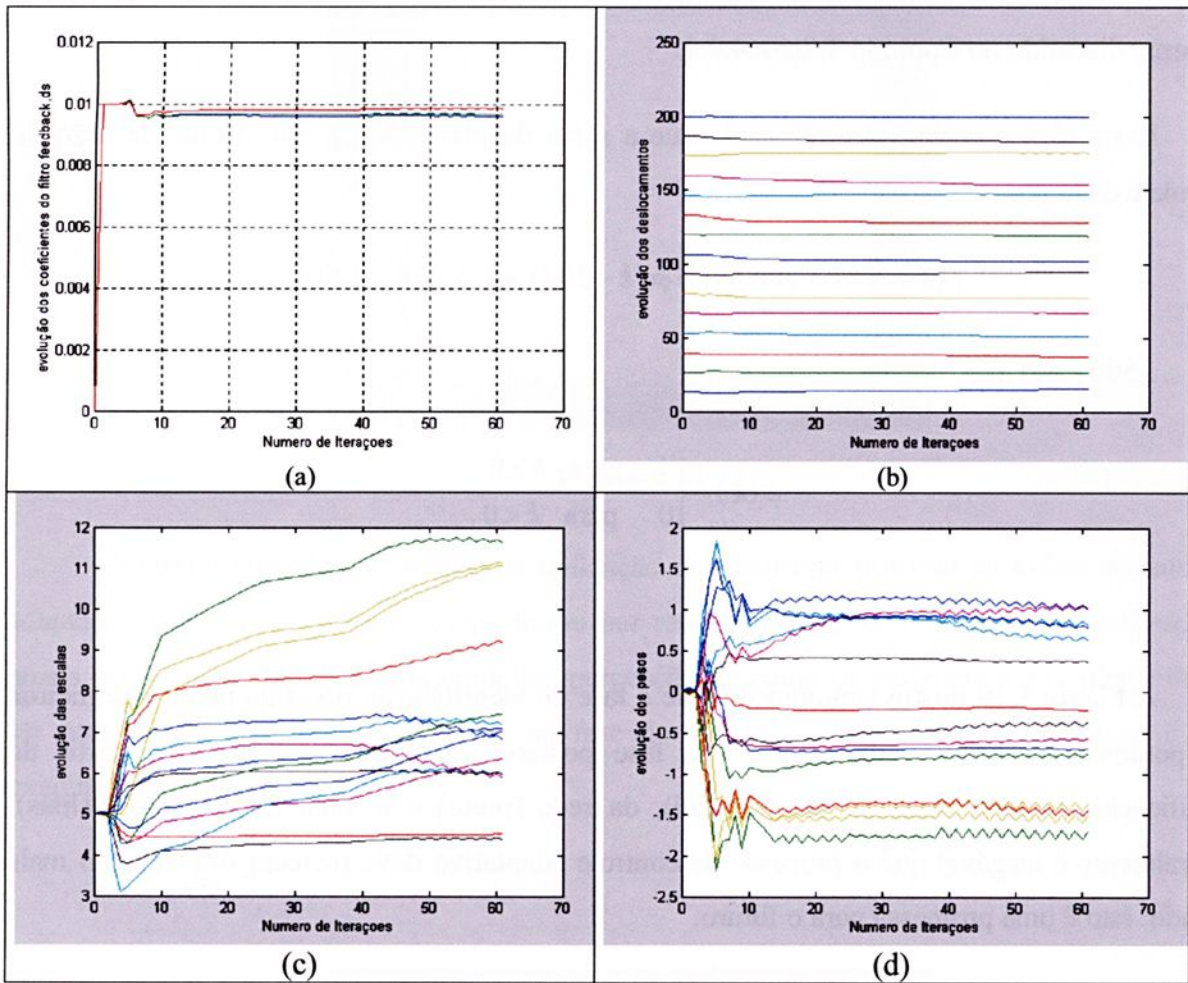


Figura 5.17 – Evolução dos parâmetros

A Figura 5.18 ilustra o decaimento do erro quadrático médio.

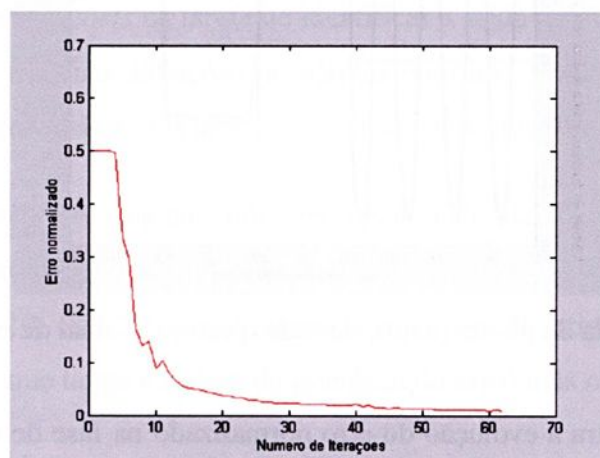


Figura 5.18 – Erro quadrático médio

Após a prévia identificação ocorre a fase de controle adaptativo em que a saída da planta deve trilhar o caminho do reconhecimento do sinal de referência. Assim, na fase de controle a

saída da planta busca pelo sinal de referência e a saída da rede busca pela saída da planta, conforme discutido no Capítulo 4, item 4.2.1.

Neste tópico o sinal de referência que a saída da planta busca em virtude da ação de controle é dado por

$$r(k) = \varphi(k-200) - 1.5\varphi(k-300) + 1.5\varphi(k-400)$$

$k=1, \dots, 500$. em que

$$\varphi(k) = \begin{cases} 1 & \text{para } k \geq 0 \\ 0 & \text{para } k < 0 \end{cases}$$

A Figura 5.19 ilustra simultaneamente a fase de identificação ocorrida para os primeiros 200 pontos sinal oriundo da planta, e a fase posterior de controle ocorrida a partir da identificação, onde a saída da planta (azul), da rede (preto) e o sinal de referência (lilás). Naturalmente é inegável que o processo de controle adaptativo deve merecer um trabalho mais apurado. Isto é uma promessa para o futuro.

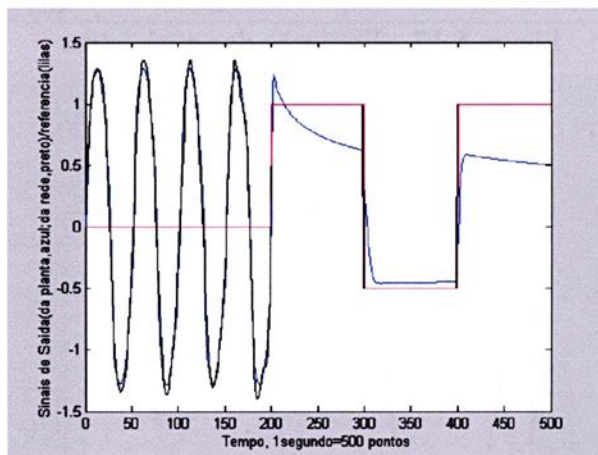


Figura 5.19 - Saída da planta (azul), da rede (preto) e o sinal de referência (lilás)

A Figura 5.20 ilustra a evolução do erro normalizado na fase de controle adaptativo. Após 10 iterações o erro foi de 0,315.

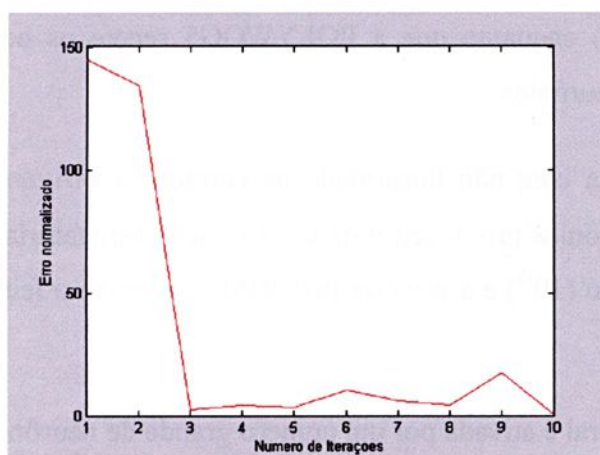


Figura 5.20

A ousadia de apresentar alguma aplicação do algoritmo *wavenet* na esfera do controle adaptativo estabelece um forte compromisso por resultados mais convincentes. As dificuldades atuais serão superadas com uma completa percepção de como os parâmetros dos filtros devem ser atualizados, iteração por iteração na fase de controle. No futuro imediato esta é uma exigência do bom senso.

5.4. DISCUSSÃO

A capacidade das *wavenets* na tarefa de identificar a saída da planta pela saída da rede foi neste trabalho ilustrada em duas situações de plantas com não linearidade na dinâmica e uma com entrada não linear, todas descritas através de equações de diferenças.

No primeiro caso de planta com não linearidade na dinâmica a função ativadora POLYWOG5 proporcionou identificação com erro de mesma ordem de grandeza (10^{-3}) em redes neurais operando com 20, 25, 30 ou 35 neurônios. A função ativadora RASP1 repete um desempenho similar (no que tange a ordem de grandeza do erro) mas operando com 5, 15, 20 ou 25 neurônios.

No segundo caso de planta com não linearidade na dinâmica a *wavelet* RASP1 ativando redes neurais com 25, 30 ou 35 neurônios proporcionou identificação com erros de mesma

ordem de grandeza (10^{-3}) enquanto que a POLYWOG5 repete os bons resultados em redes contando com 20 ou 35 neurônios.

No caso de planta com não linearidade na entrada a *wavelet* RASP1 ativando redes neurais com 5 ou 15 neurônios promoveu uma identificação satisfatória sob o ponto de vista da ordem de grandeza do erro (10^{-3}) e a *wavelet* POLYWOG5 repete o feito com 15, 20, 25, 30 ou 35 neurônios.

Quando a rede neural é ativada por um número grande de neurônios (dentre os testados) a rede pode convergir com menos iterações e isto pode parecer atraente, porém, tem maior interesse encontrar caso a caso, o menor número de neurônios (em cada iteração o número de neurônios influi fortemente na quantidade de operações aritméticas exigidas para a apresentação da saída da rede) capaz de estruturar uma rede cuja saída reconheça a meta (saída da planta) eficazmente e com menor custo computacional.

Além disso, os resultados obtidos para o controle adaptativo ainda não são conclusivos.

CAPÍTULO 6

6.1 CONCLUSÃO

Este trabalho foi elaborado em algumas etapas, a primeira tratando de processos de minimização utilizando o método do gradiente descendente, a segunda abordando as redes neurais e em seguida a construção das *wavenets*, aplicações de filtros retroalimentados e alimentados para frente e finalmente a identificação de plantas não lineares e controle.

As redes neurais ativadas por *wavelets* mostraram-se eficazes no processamento de sinais, particularmente na *identificação* de plantas não lineares e os testes preliminares sugerem a possibilidade subsequente do seu uso em *controle adaptativo*.

As tarefas discutidas neste trabalho, indicam que o sucesso na identificação depende da escolha da família *wavelet* no papel de função ativadora, bem como do número de neurônios necessários para cobrir eficazmente toda a extensão do sinal (planta , meta). Essas escolhas até agora são feitas de forma heurística e a definição de uma arquitetura apropriada da rede neural não é um trabalho trivial.

Assim, as simulações desenvolvidas nos dão a certeza da eficiência das redes neurais ativadas por *wavelets* na tarefa de identificar uma saída da planta pela saída da rede, porém , o número de janelas *wavelets* (neurônios) adequados ao objetivo pretendido ainda não é muito claro, merecendo esforços posteriores de investigação.

O trabalho futuro pode ser resumido nos seguintes tópicos principais: sobre o número mínimo de janelas necessários para tarefas de identificação e a sua relação com a opção da família *wavelet*; isto é, devemos buscar eficiência com menor custo computacional. Também a questão do controle adaptativo deverá receber atenção intensiva, pois proporcionará um emprego mais amplo das ferramentas que estudamos neste trabalho.



REFERÊNCIAS

- BENTON, A. G.; LUMETTA, B.; SPILLMAN, D. ; RING, J. Adaptive *wavelet* transforms of singular and chaotic signals: *wavelet* aplications III, PROCEEDINGS SPIE – INTERNATIONAL SOCIETY FOR OPTICAL ENGINEERING, 1996, Bellingham. **Proceedings SPIE**, v.2762, p.136-143, 1996.
- CHEN, F.C. Back propagation neural for nonlinear self-tuning adaptive control. **IEEE Control Systems Magazine**, New York, Apr., p.44-48, 1990. Special Issue
- COHEN, L. **Time-frequency analysis**. Englewood Cliffs: Prentice Hall, 1995.
- DAUBECHIES, I. **Ten lectures on wavelets**. Philadelphia: SIAM Books, 1992.
- DAUBECHIES, I. The *wavelet* transform, time-frequency localization and signal. **IEEE Transactions on Information Theory**, New York, v36, n.5, p. 961-1005, 1990.
- HAYKIN, S. **Neural networks: a comprehensive foundation**. New York: Macmillan, 1994.
- HEBB, D. O. **The Organization of Behavior: A Neuropsychological Theory**. Wiley, New York, 1949.
- HORNIK, K. ; STINCHCOMBE, M. ; WHITE, H. Multilayer feedforward network and universal approximators. **Neural Networks**, New York, v.2, p. 359-366, 1989.

JOUNY, I. ; KANAPATHIPILLAI, M. Neural network adaptive *wavelet* classification of radar targets. PROCEEDINGS OF THE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 1994, Pasadena. **Proceedings...** Pasadena: IEEE, 1994. v.4, p.1889-1891.

KADAMBE, S. ; SRINIVASAN, P. Applications of adaptive *wavelets* for speech. **Optical Engineering**, Redondo Beach, v.33, n.7, p. 2204-2211, 1994.

KADAMBE, S. ; SRINIVASAN, P. Text independent speaker identification system based on adaptive *wavelets*: *wavelet* applicationd. PROCEEDINGS SPIE THE INTERNATIONAL SOCIETY FOR OPTICAL ENGINEERING, 1994, Bellingham. **Proceedings SPIE**, 1994. v.2242, p. 669-677.

KARAKASOGLU, A. ; SUDHARSANAN, S.I. ; SUNDARESHAN, M. K. Neural network-based identification and adaptive control of nonlinear systems: a novel dynamical network architecture and training policy. PROCEEDINGS OF THE 30TH IEEE CONFERENCE ON DECISION AND CONTROL, 1991, New York. **Proceedings...** New York: IEEE, 1991. v.1, p. 180-181.

McCULLOCH, W.S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, New York, v.5, p.115-133, 1943.

MINSKY, M.L.; PAPERT, S.A. **Perceptrons**. Cambridge: MA, MIT Press, 1969.

NARENDRA, K. S. ; PARTHASARATHY, K. Identification and control of dynamical systems using neural network. **IEEE Transactions on Neural Networks**, New York, v.1, n.1, p.4-27, 1990.

NEWLAND, D. E. **An introduction to random vibrations, spectral & wavelet analysis**. New York: J. Wiley, 1993.

OMATSU, S. Dynamical systems regulation by neuro-controllers," applications and science of artificial neural network. PROCEEDINGS SPIE THE INTERNATIONAL SOCIETY FOR

OPTICAL ENGINEERING, 1995, Bellingham. **Proceedings...** Bellingham: SPIE, 1995. v.2492, p.590-599.

PATY, Y. ; KRISHNAPRASAD, P. Analysis and synthesis of feedforward neural networks using discrete affine wavelet transformation. **IEEE Transactions Neural Networks**, New York, v.4, n.1, p 73-85, 1993.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological Review**, Washington, v.65, p.386-408, 1958.

SUDHARSANAN, S. I. ; SUNDARESHAN, M. K. Training of a three-layer dynamical recurrent neural network for nonlinear input-output Mapping. In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS SINGAPURA, 1991, Washington. **Proceedings...** Washington: IEEE, 1991. v.2, p.111-115, 1991.

SZU, H. H. ; TELFER, B. A. ; KADAMBE, S. Neural network adaptive *wavelets* for signal representation and classification. **Optical Engenering**, Redondo Beach, v.31, n.9, p.1907-1916, 1992.

TELFER, B. A. ; SZU, H. H. ; DOBECK, G. J. Adaptive *wavelet* classification of acoustic backscatter. PROCEEDINGS SPIE – THE INTERNATIONAL SOCIETY FOR OPTICAL ENGINEERING, 1994, Bellingham. **Proceedings...** Bellingham: SPIE, 1994. v.2242, p.661-668.

TELFER, B. A. ; SZU, H. H. ; DOBECK, G. J. Adaptive time-frequency classification of acoustic backscatter. PROCEEDINGS SPIE THE INTERNATIONAL SOCIETY FOR OPTICAL ENGINEERING, 1995, Bellingham. **Proceedings...** Bellingham: SPIE, 1995. v.2491, Pt.1, p.451-460.

ZHANG, Q. ; BENVENISTE, A. *Wavelet* networks. **IEEE Transactions Neural Networks**, New York, v.3, n.6, p. 889-898, 1992.



APÊNDICE

TEOREMA DE SHANNON-WHITTAKER

É possível obter uma boa representação de uma função usando amostragem pontual?

A resposta é sim, se forem colocadas algumas restrições sobre a função f e ao mesmo tempo algumas condições sobre a taxa de amostragem. Inicialmente façamos as seguintes considerações:

- processo de amostragem é uniforme, isto é, os intervalos de amostragem $[t_k, t_{k+1}]$ tem o mesmo comprimento $\Delta t = t_{k+1} - t_k$;
- a transformada de Fourier $\hat{f}$ da função f se anula fora de um intervalo limitado $[-\Omega, \Omega]$ do domínio da frequência, isto é, $\hat{f}$ tem suporte compacto.

Usando as restrições acima podemos enunciar:

TEOREMA DE SHANNON-WHITTAKER

Considere uma função $f : \mathfrak{R} \rightarrow \mathfrak{R}$ com suporte compacto, $\text{supp}(\hat{f}) \subset [-\Omega, \Omega]$, e uma partição uniforme t_i , $i \in \mathbb{Z}$ de números reais tais que $2\Delta t \leq \Omega$. Então f , pode ser reconstruída a partir de sua representação pelos pontos amostrados $(f(t_i))_{i \in \mathbb{Z}}$. A equação de reconstrução é dada por

$$f(t) = \sum_{k=-\infty}^{+\infty} f\left(\frac{k}{2\Omega}\right) \frac{\text{sen } \pi(2\Omega t - k)}{\pi(2\Omega t - k)}.$$

A desigualdade $2\Delta t \leq \Omega$ é chamada de Nyquist. Ela diz que deve-se tomar pelo menos duas amostras para cada ciclo completo de frequência máxima ocorrendo na função.

As hipóteses do teorema são:

- 1- A função f não deve ter frequências muito altas, $\text{supp}(\hat{f}) \subset [-\Omega, \Omega]$;
- 2- As amostras tomadas de f devem ser suficientemente próximas, obedecendo a desigualdade de Nyquist.

A conclusão do teorema é que f pode ser reconstruída exatamente, interpolando suas amostras usando a equação de reconstrução.





UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"

Câmpus de Ilha Solteira
Programa de Pós-Graduação em Engenharia Mecânica
Av. Brasil Centro, 56
15385-000 Ilha Solteira - SP
www.dem.feis.unesp.br

