



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA  
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

**REDES NEURAIS E NOVOS MÉTODOS DE REGULARIZAÇÃO EM  
PROBLEMAS INVERSOS DE CONDUÇÃO DE CALOR**

Elcio Hideiti Shiguemori

Dissertação de Mestrado em Computação Aplicada, orientada pelo Dr. Haroldo Fraga de Campos Velho e Dr. José Demísio Simões da Silva, aprovada em Fevereiro de 2002.

INPE  
São José dos Campos  
2002

**MINISTÉRIO DA CIÊNCIA E TECNOLOGIA  
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

**REDES NEURAIS E NOVOS MÉTODOS DE REGULARIZAÇÃO EM  
PROBLEMAS INVERSOS DE CONDUÇÃO DE CALOR**

Elcio Hideiti Shiguemori

Dissertação de Mestrado em Computação Aplicada, orientada pelo Dr. Haroldo Fraga de Campos Velho e Dr. José Demísio Simões da Silva, aprovada em Fevereiro de 2002.

INPE  
São José dos Campos  
2002

xxx.xxx.xx(xxx.x)

SHIGUEMORI, E.H.

Redes Neurais e Novos Métodos de Regularização em Problemas Inversos de Condução de Calor / E. H. Shiguemori. - São José dos Campos: INPE, 2002.

(INPE-xxxx-TDI/xxx).

1. Problemas Inversos. 2. Redes Neurais Artificiais. 3. Condução de Calor. 4. Condição Inicial. 5. Condições de Contorno.

Aprovada pela Banca Examinadora em cumprimento a requisito exigido para a obtenção do Título de Mestre em Computação Aplicada

Dr. Fernando Manoel Ramos

\_\_\_\_\_  
Presidente

Dr. Haroldo Fraga de Campos Velho

\_\_\_\_\_  
Orientador

Dr. José Demisio Simões da Silva

\_\_\_\_\_  
Orientador

Dr. Lamartine Nogueira Frutuoso Guimarães

\_\_\_\_\_  
Membro da Banca

Dr. Antônio de Pádua Braga

\_\_\_\_\_  
Membro da Banca  
- convidado -

Candidato: Elcio Hideiti Shiguemori

São José dos Campos, 27 de Fevereiro de 2002.

*Aos meus pais, oferto o amor e a promessa de tê-los sempre em meu coração; a gratidão, desde já, declaro ser eterna, porque estou ciente de que o meu “tudo” não é suficiente para agradecer. A eles dedico este trabalho como uma singela retribuição pelo amor, carinho e dedicação que, com a ajuda de Deus, me incentivaram e apoiaram nesta difícil empreitada.*

## **AGRADECIMENTOS**

Primeiramente a Deus, que através de sua infinita bondade, nos concede incontáveis oportunidades para o crescimento.

Aos meus orientadores Dr. Haroldo Fraga de Campos Velho e Dr. José Demísio Simões da Silva, a quem tenho profunda admiração, agradeço pela confiança e por terem me orientado de forma segura e amiga.

Aos membros da banca examinadora pela disposição em analisar este trabalho.

À minha querida mamãe Yurico e a minha irmã Erica que com muito amor e carinho sempre me incentivaram muito, me apoiando e acreditando em minha capacidade.

À minha querida namorada Ana Paula por estar sempre ao meu lado me incentivando nas horas que mais precisei.

Em memória ao meu pai Mário e aos meus tios Bruno e Hiro, pois sei que acompanharam em espírito meu esforço.

Agradeço também a meu avô Noboru Gondo que com muito carinho sempre me incentivou.

À minha tia Dica que muito me apoiou, confiou e torceu por mim.

Ao meu tio Massa pela força que sempre me deu.

Ao meu cunhado e amigo Silvío pelo incentivo e por me ajudar nas correções do texto.

Ao meu amigo Dr. Francisco Tavares pelas sugestões e incentivo durante minha formação acadêmica.

Aos amigos do INPE, que acompanharam e apoiaram este trabalho, em especial aos amigos Aritana, Fabrício, Felipe, Isabela e Leonardo.

A todos os professores do INPE-LAC, em especial ao Dr. Fernando Manuel Ramos por ter sugerido o uso de novo operador de máxima entropia e por ter acreditado em minha capacidade.

À FAPESP pelo apoio financeiro e pela compra dos equipamentos.

## RESUMO

Esta dissertação aborda o uso de redes neurais e de um novo operador de regularização, a Máxima Entropia Não-Extensiva, na solução de problemas inversos em condução de calor. As redes neurais artificiais são apresentadas como métodos alternativos na recuperação da condição inicial e da condição de contorno, problemas estes considerados mal-postos requerendo técnicas avançadas, como métodos de regularização, quando resolvidos por métodos clássicos. Para o uso da Máxima Entropia Não-Extensiva é considerado apenas o problema da recuperação da condição inicial. No uso de Redes Neurais Artificiais, estuda-se diferentes redes, como o Perceptron de Camadas Múltiplas com aprendizagem por retropropagação do erro, a rede Funções de Base Radial e uma rede com algoritmo de Correlação em Cascata, treinadas com dois conjuntos de dados diferentes: um contendo funções não-similares e outro funções similares. A análise de desempenho dos modelos de redes é feita apresentando-se dados não usados no treinamento, adicionando-se ou não ruído gaussiano a 5%. Nos experimentos realizados observou-se que na fase de treinamento, a rede Funções de Base Radial convergiu mais rapidamente para a solução, em ambos os problemas inversos estudados. E a estimação da condição inicial exigiu um número muito maior de épocas de treinamento. Com relação à recuperação da condição inicial utilizando a entropia não-extensiva, boas inversões foram recuperadas utilizando o princípio de Hansen e o critério de Morozov na determinação do parâmetro de regularização.



# NEURAL NETWORKS AND NEW REGULARIZATION METHODS IN INVERSE HEAT CONDUCTION PROBLEM

## ABSTRACT

*This dissertation discusses the use of Artificial Neural Networks and Non-extensive Entropy, a New Regularization Operator, to solve two Inverse Heat Conduction Problems. The Artificial Neural Networks are presented as an alternative method in initial condition and boundary condition estimation. These ill-posed problems require advanced techniques to solve them when solved by classic methods, i.e. regularization methods. The New Regularization Operator was used only for the initial condition estimation problem. Using Artificial Neural Network, different kinds of Networks are studied, namely, the Multilayer Perceptron with backpropagation algorithm learning, the Radial Basis Function Network and Cascade Correlation Network. The three networks were trained with two different kinds of data sets: the first containing similar-functions and the second non-similar functions. The performance analysis of the networks models was done by introducing data that had not been used in the training data, adding 5% gaussian noise or not. During the experiments it was observed that the Radial Basis Function converged faster than the others, in both inverse problems under study. The initial condition estimation required a larger number of training epochs. Using Non-extensive Entropy in the initial condition, good estimates were obtained using the Hansen Principle and Morozov criterion in regularization parameter determination.*

## SUMÁRIO

	<u>Pág.</u>
<b>1 INTRODUÇÃO1 .....</b>	<b>15</b>
<b>2 Problemas Inversos .....</b>	<b>19</b>
2.1 Problema Mal-posto .....	20
2.2 Métodos de Solução de Problemas Inversos .....	22
2.2.1 Inversão Direta .....	22
2.2.2 Mínimos Quadrados .....	23
2.2.3 Métodos de Regularização.....	23
2.3 Regularização pela Máxima Entropia Não-extensiva .....	26
<b>3 Redes Neurais Artificiais.....</b>	<b>35</b>
3.1 Neurônios Biológicos .....	36
3.2 Neurônios Artificiais .....	38
3.3 Funções de Ativação.....	39
3.4 Arquitetura das Redes Neurais Artificiais .....	40
3.5 Funcionamento das Redes Neurais.....	41
3.6 Redes Perceptron .....	42
3.7 Rede Funções de Base Radial.....	43
3.8 Rede neural com algoritmo de Correlação em Cascata.....	47
3.9 Rede Perceptrons de Camadas Múltiplas .....	49
<b>4 REDES NEURAIIS EM pROBLEMAS INVERSOS DE CONDUÇÃO DE CALOR.....</b>	<b>52</b>
4.1 Geração dos Conjuntos de Dados .....	53
4.1.1 Problema Direto, Recuperação da Condição Inicial.....	53
4.1.2 Problema Direto, Recuperação da Condição de Contorno .....	54
4.2 Treinamento da Rede Neural .....	55
4.3 Ativação da Rede Neural .....	56
4.4 Critério de Parada .....	57

4.5	Implementação.....	58
<b>5</b>	<b>Resultados. ....</b>	<b>59</b>
5.1	Recuperação da Condição Inicial .....	60
5.1.1	Treinamento da rede Perceptron de Camadas Múltiplas.....	60
5.1.2	Ativação da rede Perceptron de Camadas Múltiplas.....	61
5.1.3	Testes de Generalização da RPCM .....	63
5.1.4	Treinamento da rede de Funções de Base Radial .....	65
5.1.5	Ativação da rede Funções de Base Radial.....	65
5.1.6	Testes de Generalização da RFBR .....	67
5.1.7	Treinamento da Rede com Algoritmo de Correlação em Cascata.....	68
5.1.8	Ativação da rede com algoritmo de correlação em cascata.....	69
5.1.9	Testes de Generalização da RCC .....	70
5.2	Recuperação da Condição de Contorno.....	71
5.2.1	Treinamento da Rede Perceptron de Camadas Múltiplas.....	71
5.2.2	Ativação da Rede Perceptron de Camadas múltiplas .....	72
5.2.3	Testes de Generalização da RPCM .....	73
5.2.4	Treinamento da Rede Funções de Base Radial .....	75
5.2.5	Ativação da Rede Funções de Base Radial .....	75
5.2.6	Testes de Generalização da RFBR .....	77
5.2.7	Treinamento da Rede com Algoritmo de Correlação em Cascata.....	78
5.2.8	Ativação da Rede Treinada com Algoritmo de Correlação em Cascata .....	79
5.2.9	Generalização da Rede Treinada com Algoritmo de Correlação em Cascata	80
<b>6</b>	<b>conclusões. ....</b>	<b>82</b>

## LISTA DE FIGURAS

	<u>Pág.</u>
Figura 1.1 – Problema inverso com RNA versus problema direto.....	17
Figura 2.1 – Problema inverso versus problema direto.....	20
Figura 2.2 – Reconstrução com 5% de ruído, com $\alpha$ determinado pelo princípio de Morozov: (a) $q=0,5$ e (b) $q=1,5$ .....	28
Figura 2.3 – Reconstrução com 5% de ruído, com $\alpha$ determinado pelo princípio de Morozov: (a) $q=2,0$ e (b) $q=2,5$ .....	29
Figura 2.3 – Reconstrução com 5% de ruído, com $\alpha$ determinado pelo princípio de Morozov: (a) $q=0,5$ ; (b) $q=1,5$ ; (c) $q=2,0$ e (d) $q=2,5$ .....	30
Figura 2.4 – <i>Curva-L</i> para a função teste triangular: (a) $q=0,5$ ; (b) $2,0$ .....	31
Figura 2.5 – <i>Curva-L</i> para a função teste semitriangular: (a) $q=0,5$ ; (b) $2,0$ .....	31
Figura 2.6 – Reconstrução com 5% de ruído, $\alpha$ determinado pelo critério de Hansen: (a) $q=0,5$ ; (b) $q=1,5$ ; (c) $q=2,0$ e (d) $q=2,5$ .....	32
Figura 2.7 – Reconstrução com 5% de ruído, $\alpha$ determinado pelo critério de Hansen: (a) $q=0,5$ ; (b) $q=1,5$ ; (c) $q=2,0$ e (d) $q=2,5$ .....	33
Figura 3.1 – Esquema da célula neural.....	37
Figura 3.2 – Representação de um neurônio artificial.....	38
Figura 3.3 – Funções de ativação.....	39
Figura 3.4 – Rede com uma única camada.....	40
Figura 3.5 – Rede de três camadas.....	41
Figura 3.6 – Arquitetura da RCC.....	48
Figura 4.1 – Gráfico esquemático da solução do Problema Inverso.....	52
Figura 4.2 – Variação da temperatura.....	54
Figura 4.3 – Resolução do Problema Direto.....	55
Figura 4.4 – Resolução do Problema Inverso.....	57
Figura 5.1 – (a) Conjunto de funções não similares, (b) Conjunto de funções similares. .....	59
Figura 5.2 – Exemplos de inversões obtidas com a RPCM.....	62

Figura 5.3 – Exemplos dos testes de generalização obtidos com a RPCM.....	63
Figura 5.4 – Exemplos das inversões obtidas com a RFBR.....	66
Figura 5.5 – Exemplos de generalizações obtidas com a RFBR.....	67
Figura 5.6 – Exemplos de inversões obtidas com a RCC.....	69
Figura 5.7 – Exemplos de generalizações obtidos com a RCC.....	70
Figura 5.8 – Exemplos de generalizações obtidas com a RPCM.....	72
Figura 5.9 – Exemplos de inversões obtidas com a RPCM.....	74
Figura 5.10 – Exemplos de ativações obtidas com a RFBR.....	76
Figura 5.11 – Exemplos de generalizações obtidas com a RFBR.....	77
Figura 5.12 – Exemplos de ativações obtidas com a RCC.....	79
Figura 5.13 – Exemplos de generalizações obtidas com a RCC.....	80

## LISTA DE TABELAS

	<u>Pág.</u>
Tabela 2.1 – Função Triangular com $q=0,5$ .....	28
Tabela 2.2 – Função Triangular com $q=2,0$ .....	28
Tabela 2.3 – Função semiTriangular com $q=0,5$ .....	29
Tabela 2.4 – Função semiTriangular com $q=2,0$ .....	29
Tabela 2.5 – Função Triangular .....	32
Tabela 2.6 – Função semi-Triangular .....	34
Tabela 5.1 – Resultados dos treinamentos da RPCM .....	61
Tabela 5.2 – Resultados de ativações da RPCM .....	62
Tabela 5.3 – Resultados de Generalizações da RPCM .....	64
Tabela 5.4 – Resultados dos treinamentos da RFBR .....	65
Tabela 5.5 – Resultados de ativações da RFBR .....	66
Tabela 5.6 – Resultados de Generalizações da RFBR .....	68
Tabela 5.7 – Resultados dos treinamentos da RCC .....	68
Tabela 5.8 – Resultados de ativações da RCC .....	69
Tabela 5.9 – Resultados de generalizações da RCC .....	71
Tabela 5.10 – Resultados dos treinamentos da RPCM .....	72
Tabela 5.11 – Resultados de ativações da RPCM .....	73
Tabela 5.12 – Resultados das generalizações da RPCM .....	74
Tabela 5.13 – Resultados dos treinamentos da RFBR .....	75
Tabela 5.14 – Resultados de Ativações da RFBR .....	76
Tabela 5.15 – Resultados das generalizações da RFBR .....	78
Tabela 5.16 – Resultados dos treinamentos da RCC .....	78
Tabela 5.17 – Resultados de Ativações da RCC .....	80
Tabela 5.18 – Resultados de generalizações da RCC .....	81

## CAPÍTULO 1

### INTRODUÇÃO

As Redes Neurais Artificiais (RNA) têm emergido como uma nova técnica de solução de Problemas Inversos (Hidalgo et al., 1996), (Krejca et al., 1998), (Ogawa et al., 1998), (Woodbury, 1998), (Mikki et al., 1999), (Braga et al., 2000). Neste trabalho estuda-se o uso das RNA na solução de dois Problemas Inversos mal postos em Condução de Calor: a recuperação da condição inicial e a recuperação da condição de contorno. A condição inicial também é recuperada por métodos clássicos de inversão, onde é estudado um novo operador de regularização: a máxima entropia não-extensiva.

Nas soluções de problemas diretos, os efeitos são determinados através de valores dos parâmetros dos modelos, enquanto na solução de Problemas Inversos, causas desconhecidas são determinadas a partir de efeitos desejados ou observados (Engl, 1996). Muitos Problemas Inversos tratam de modelos matemáticos que são “bem-postos”, conceito introduzido por Hadamard (Bertero e Boccacci, 1998) no início do século XX, no qual os problemas devem apresentar soluções que satisfaçam as seguintes propriedades: existência, unicidade e estabilidade. Se uma dessas propriedades não é satisfeita o problema é considerado “mal-posto”.

As técnicas de Inteligência Computacional vêm sendo empregadas na solução de Problemas Inversos com o uso de Algoritmos Genéticos (Raudensky, 1996), (Navarro, 1998) e de Redes Neurais (Woodbury, 1998), (Hidalgo et al., 1996), (Krejca et al., 1998), (Mikki et al., 1999).

Mikki et al. (1999) e Issamoto et al. (1999) utilizaram as RNA para estimar a condição inicial no problema inverso de condução de calor, sendo utilizadas as redes: Perceptron de Camadas múltiplas e Funções de Base Radial. Além da condição inicial também é recuperado o tempo transcorrido entre o resultado inicial e a tomada das observações experimentais. Os autores concluíram que rede Perceptron de Camadas

Múltiplas apresentou resultados melhores na ativação, enquanto a rede Funções de Base Radial apresentou um tempo menor no treinamento. No entanto não foram obtidos resultados da mesma qualidade daqueles obtidos pelos métodos tradicionais como os apresentados em Muniz et al. (1999) e Muniz et al. (2000).

Krejsa et al. (1998) apresentaram a resolução do problema inverso em condução de calor na determinação da condição de contorno, utilizando Redes Neurais Artificiais. Dois experimentos são realizados, no primeiro é utilizado o mapeamento de todo o histórico dos valores de temperatura observados, enquanto no segundo é utilizada uma aproximação sequencial dos valores da superfície de calor. Bons resultados foram obtidos nas inversões, no entanto com uma pequena porcentagem de ruído.

Os problemas inversos em condução de calor buscam estimar: condição inicial, condição de contorno, propriedades térmicas do material e fontes (sumidouros) de calor. O problema direto de condução de calor numa barra é dado matematicamente pelas equações:

$$\frac{\partial^2 T(x,t)}{\partial x^2} = \frac{\partial T(x,t)}{\partial t}, \quad x \in (0,1), \quad t > 0, \quad (1.1)$$

$$\frac{\partial T(x,t)}{\partial x} = 0, \quad x = 1, \quad t > 0 \quad (1.2)$$

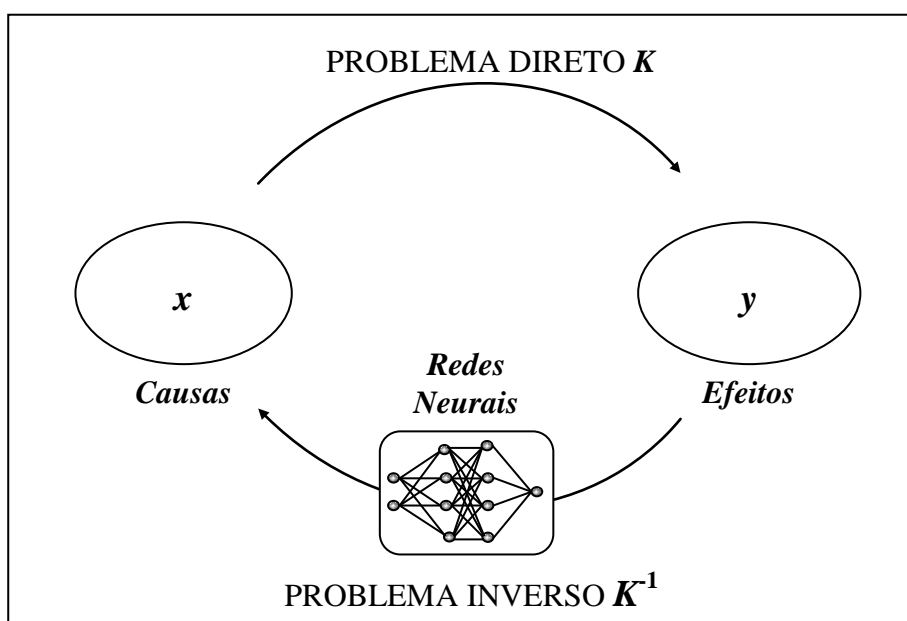
$$\frac{\partial T(x,t)}{\partial x} = q(t), \quad x = 0, \quad t > 0 \quad (1.3)$$

$$T(x,0) = f(x), \quad x \in [0,1], \quad t = 0, \quad (1.4)$$

onde  $T(x,t)$  é a temperatura,  $f(x)$  é a condição inicial,  $q(t)$  o fluxo de calor no contorno,  $x$  (variável espacial) e  $t$  (variável temporal) são quantidades adimensionais.



Nesta dissertação, as Redes Neurais Artificiais são utilizadas para estimar dois problemas inversos em condução de calor: estimação da condição inicial e estimação da condição de contorno. A Figura (1.1) ilustra o uso de redes neurais artificiais na solução de problemas inversos.



**Figura 1.1 – Problema inverso com RNA versus problema direto.**

Na solução do primeiro problema inverso em condução de calor, recupera-se a condição inicial  $f(x)$  dadas as condições de contorno em  $x=0$  e  $x=1$ . Enquanto na solução do segundo problema inverso recupera-se o fluxo de calor  $q(t)$  no contorno  $x=0$ , conhecendo-se a condição inicial e a condição de contorno em  $x=1$ .

Com o objetivo de avaliar o desempenho de diferentes redes neurais na solução dos problemas inversos propostos, são implementadas as RNA: Perceptron de Camadas Múltiplas com aprendizagem por retropropagação do erro (RPCM), Funções de Base Radial (RFBR) e uma terceira com algoritmo de Correlação em Cascata (RCC). São realizadas comparações das fases de treinamento, de ativação e de generalização das redes. Outra análise realizada nesta dissertação é com relação aos conjuntos de treinamento. São gerados dois conjuntos: um contendo 500 funções similares e outro

contendo 500 funções não-similares. Também são realizados experimentos com dados ruidosos, sendo adicionados 5% de ruído gaussiano.

No capítulo 2 são discutidos métodos clássicos de solução de problemas inversos, além da teoria geral para solução de equações mal-postas. É apresentado um novo operador de regularização: a entropia não-extensiva. Também é estimada a condição inicial do problema inverso de condução de calor, sendo variados parâmetros do operador de regularização. Também são implementados dois métodos para a determinação do parâmetro de regularização, a máxima curvatura da *curva-L* e o princípio da discrepância de Morozov.

Os capítulos restantes que compõem este trabalho podem ser assim resumidos: no capítulo 3, são introduzidas as Redes Neurais Artificiais, onde é apresentado um breve histórico, assim como algumas arquiteturas existentes e alguns algoritmos de treinamento das redes. No quarto capítulo é apresentada a metodologia utilizada. No capítulo seguinte são apresentados os resultados do treinamento, ativação e generalização das redes neurais. E para finalizar no capítulo 6 são apresentadas as conclusões e os futuros trabalhos na área.

## CAPÍTULO 2

### PROBLEMAS INVERSOS

A área de Problemas Inversos (PI) tem crescido nos últimos anos, principalmente em aplicações nos campos da indústria e ciência (Dulikravich, 1999). A principal dificuldade no tratamento de PI é a instabilidade de suas soluções com a presença de ruído nos dados medidos ou observados, que é a natureza de problemas mal-postos. Essa dificuldade tem feito com que muitas técnicas de soluções de problemas inversos tenham sido propostas (Bertero e Boccacci, 1998). Algumas destas técnicas são apresentadas no decorrer deste capítulo, bem como alguns exemplos de aplicações das técnicas de problemas inversos.

Os problemas inversos podem ser classificados em três tipos, todos eles baseados em observações na evolução do sistema físico envolvido, sendo eles identificação de parâmetros, determinação do estado inicial do sistema e determinação de condições de contorno (Muniz, 1999).

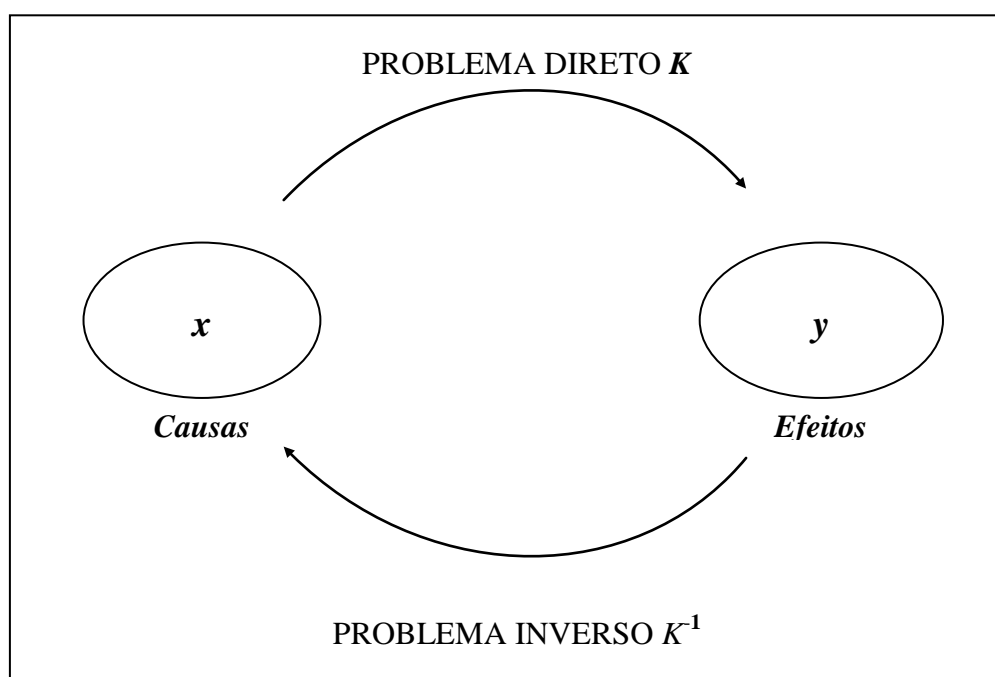
Segundo um dos grandes conhecedores da área de PI, Oleg Alifanov, *a solução de um problema inverso determina causas desconhecidas baseada nos efeitos desejados ou observados*. Existem duas diferentes motivações para o estudo de problemas inversos: quando se deseja conhecer estados passados ou parâmetros de um sistema físico e quando se deseja saber como influenciar um sistema através de seu estado presente ou através dos parâmetros, com o objetivo de dirigi-los para que se obtenha um estado desejado no futuro.

Na solução de um problema direto determina-se o efeito  $y$  de uma dada causa  $x$ , quando um modelo matemático definido  $K$  é colocado:

$$Kx=y \tag{2.1}$$

Para os problemas diretos assume-se que o operador  $K$  esteja bem definido e seja contínuo, assim há um único efeito  $y$  para cada causa  $x$  e pequenas variações em  $x$  resultam em pequenas variações em  $y$ .

A partir do problema direto, apresentado pela equação (2.1), podem ser colocados dois tipos de problemas inversos. O problema inverso de causa (dados  $K$  e  $y$ , determinar  $x$ ) e o de identificação do modelo (dados  $x$  e  $y$ , determinar  $K$ ). A Figura 2.1 mostra a relação entre os problemas direto e inverso.



**Figura 2.1 – Problema inverso versus problema direto.**

## 2.1 Problema Mal-posto

O conceito de problema bem-posto foi proposto pelo matemático francês Hadamard em 1902, num problema de valor de contorno para equações diferenciais parciais e sua interpretação física (Bertero e Boccacci, 1998). Segundo a definição, um problema é considerado bem-posto se satisfaz três propriedades:

- existência,
- unicidade e
- estabilidade, sendo considerado mal-posto, se uma destas propriedades não é satisfeita.

Um modelo físico abstrato pode ser descrito por um mapeamento  $K$  de um conjunto  $X$  de parâmetros teóricos num conjunto  $Y$  de resultados:

$$K : X \rightarrow Y \quad (2.2)$$

$$x \rightarrow K(x) = Kx = y \quad (2.3)$$

Segundo Hadamard (1902), um problema é considerado bem-posto se satisfaz a definição:

*Sejam  $X$  e  $Y$  espaços normados,  $K : X \rightarrow Y$  um operador (linear ou não linear). A equação (ou problema)  $Kx=y$  é dita bem-posta se*

1. *Para qualquer  $y \in Y$  existe (pelo menos um)  $x \in X$  tal que  $Kx=y$ . (Propriedade de Existência).*
2. *Para qualquer  $y \in Y$  existe no máximo um  $x \in X$  tal que  $Kx=y$ , isto é, se  $Kx_1=Kx_2$ ,  $x_j \in X$ , então  $x_1=x_2$ . (Propriedade de Unicidade)*
3. *A solução  $x$  depende continuamente de  $y$ , isto é, para qualquer seqüência  $\{x_n\} \subset X$  tal que  $Kx_n \rightarrow Kx$  ( $n \rightarrow \infty$ ), então  $x_n \rightarrow x$  ( $n \rightarrow \infty$ ). (Propriedade de Estabilidade).*

Existem vários métodos para solução de problemas inversos, como por exemplo: inversão direta, decomposição em valores singulares, mínimos quadrados e variantes,

métodos de regularização, métodos variacionais, entre outros (Tarantola, 1987; Tikhonov e Arsenin, 1977; Bertero e Boccacci, 1998).

## 2.2 Métodos de Solução de Problemas Inversos

Nesta seção são apresentados alguns dos métodos existentes para a solução de problemas inversos:

1. inversão direta,
2. mínimos quadrados e
3. métodos de regularização.

### 2.2.1 Inversão Direta

Métodos explícitos ou inversão direta não são métodos gerais. Normalmente, existe mais um interesse acadêmico do que um esquema metodológico geral a ser seguido. Como exemplo, pode ser citado o problema inverso de condução de calor, onde busca-se a estimação da condição inicial, cuja solução exata do problema direto é dada pela equação

$$T(x, t) = \sum_{m=0}^{+\infty} e^{-\beta_m^2 t} \frac{X(\beta_m, x)}{N(\beta_m)} \int_0^1 X(\beta_m, x') f(x') dx', \quad (2.4)$$

onde  $X(\beta_m, x)$  são as autofunções associadas ao problema de autovalores obtido pelo método de separação de variáveis,  $\beta_m$  são os autovalores e  $N(\beta_m)$  é a norma ou integral de normalização (Ösizik, 1980; Muniz et al, 1999).

A solução exata do problema inverso para um perfil de temperatura  $T(x)$  medida num tempo  $t = \tau$  é (Muniz et al, 1999):

$$f(x) = \sum_{m=0}^{+\infty} e^{\beta_m^2 \tau} \frac{X(\beta_m, x)}{N(\beta_m)} \int_0^1 X(\beta_m, x') T(x', \tau) dx'. \quad (2.5)$$

### 2.2.2 Mínimos Quadrados

Os mínimos quadrados são muito utilizados na solução de problemas inversos devido à facilidade de implementação. No entanto, sua utilização não garante robustez na inversão, por exemplo, são sensíveis a uma pequena quantidade de ruído.

Para encontrar o vetor de parâmetros  $x$ , minimiza-se em relação à  $x$  o funcional:

$$\|Kx - y\|_2^2 \quad (2.6)$$

onde  $Kx$  denota a solução do problema direto,  $y$  representa os dados e  $\|\cdot\|_2$  denota a norma-2. Nesta minimização é necessário que seja calculado o problema direto, dado pela equação 2.4.

### 2.2.3 Métodos de Regularização

O uso de regularização consiste em obter soluções aproximadas de  $Kx=y$  para problemas mal-postos, tal que sejam estáveis para pequenas variações no valor de  $y$ . Tikhonov (1943; 1963) observou que informações a priori poderiam restaurar alguma estabilidade a um problema inverso mal-posto.

A regularização pode ser vista como um procedimento que muda a abordagem por mínimos quadrados (minimização de resíduo) através da adição de termos que pretendem reduzir flutuações na função ou vetor desconhecido  $x$ , flutuações que são

inerentes a problemas mal-postos. Assim, passa-se de uma abordagem natural via mínimos quadrados para uma abordagem com regularização ( $\Omega[x]$ )

$$\min_{x \in X} \|Kx - y\|_Y^2 \Rightarrow \min_{x \in X} \left( \|Kx - y\|_Y^2 + \alpha \Omega[x] \right) \quad (2.7)$$

A partir da base teórica do procedimento de regularização, procedimentos têm sido desenvolvidos baseados nesses conceitos.

### 2.2.3.1 Regularização de Tikhonov

Uma técnica de regularização proposta por Tikhonov e Arsenin (1977) pode ser expressa por:

$$\Omega[f] = \sum_{k=0}^N \alpha_k \|x^{(j)}\|_2^2 \quad (2.8)$$

onde  $x^{(j)}$  denota a  $j$ -ésima derivada em relação a  $x$ , sendo  $x=x(t)$  e os parâmetros  $\alpha_k \geq 0$ .

Se  $\alpha_k = \delta_{kj}$ , delta de Kronecker (Muniz et al. 2000), isto é,

$$\Omega[f] = \|x^{(j)}\|_2^2, \quad (2.9)$$

então, o método é chamado regularização de Tikhonov de ordem  $j$  (Tikhonov- $j$ ). Se  $\alpha_k \rightarrow 0$ , o termo dos mínimos quadrados é sobre-estimado, que pode não apresentar bons resultados com a presença de ruído. Por outro lado, se  $\alpha_k \rightarrow \infty$ , toda consistência com a informação sobre o sistema é perdida.



### 2.2.3.2 Regularização pelo princípio da máxima Entropia

Similarmente à regularização de Tikhonov, o formalismo de máxima entropia busca por regularidade global e gera reconstruções suaves, que são consistentes com os dados avaliáveis (Muniz et al., 1999).

O princípio da máxima entropia foi primeiramente proposto para um processo de inferência global por Jaynes (1957), com base na caracterização axiomática da quantidade de informação de Shannon (1949).

Considerando o vetor de parâmetros  $\bar{f}$  com componentes não negativos, a função de entropia discreta  $S$  de  $\bar{f}$  é definida pelas equações

$$S(\bar{f}) = -\sum_{q=1}^N s_q \ln(s_q), \quad \text{com } f = [f_1, f_2, \dots, f_N]^T, \quad s_q = f_q / \sum_{q=1}^N f_q \quad (2.10)$$

onde  $f_q = f(x_q)$ . A função entropia (não negativa)  $S$  atinge seu máximo global quando todos os  $s_q$  são os mesmos, que corresponde a uma distribuição uniforme com valores de  $S_{\max} = \ln N$ , enquanto o nível de entropia mais baixo,  $S_{\min} = 0$ , é ativado quando todos os elementos  $s_q$  são fixados em zero, exceto num único ponto.

É também possível definir funções de entropia de alta ordem, o que foi proposto por Ramos e Campos Velho (1996), Campos Velho e Ramos (1998) e Ramos et al. (1999). Neste sentido foram definidos dois novos procedimentos de regularização baseados na metodologia de máxima entropia, Muniz et al. (2000).

### 2.3 Regularização pela Máxima Entropia Não-extensiva

Uma forma não-extensiva de entropia foi proposta por Tsallis (1988) e tem sido utilizada em diferentes aplicações, como em certos tipos de processos de difusão anômalos (Tsallis, 1999), bem como o modelo estatístico para dados de fluxo turbulentos (Ramos et al., 2001a, 2001b) e de mercado financeiro (Ramos et al., 2001a). Segundo Plastino e Plastino (1999), a primeira confirmação do formalismo não extensivo é a aproximação de Boghosian Plasma Bidimensional formado somente por elétrons (Boshosian, 1996). A forma da entropia não-extensiva proposta por Tsallis (1988) é dada pela equação:

$$S_q(p) = \frac{k}{q-1} \left( 1 - \sum_{i=1}^{N_p} p_i^q \right) \quad (2.11)$$

onde  $p_i$  é uma probabilidade e  $q$  é um parâmetro livre. Em termodinâmica, o parâmetro  $k$  é conhecido como constante de Boltzmann. Similarmente como na teoria matemática da informação,  $k=1$  é considerado na teoria da regularização. A entropia de Tsallis reduz a uma fórmula usual de Boltzmann-Gibbs-Shanon, quando  $q \rightarrow 1$ .

Nesta dissertação, o princípio da máxima entropia de ordem zero é aplicado como uma função de regularização para estimar a condição inicial em transferência de calor. São estudados dois métodos para determinar o parâmetro de regularização, a máxima curvatura para a *curva-L* e o critério da discrepância de Morozov, alguns resultados são apresentados em (Shiguemori et al., 2002a).

O parâmetro de regularização  $\alpha$  é escolhido por dois métodos: o primeiro numericamente, assumindo que uma medida estatística seja conhecida, esse processo é baseado no princípio da discrepância de Morozov (1966); o segundo graficamente, achando o ponto da máxima curvatura na curva (Hansen, 1992).

Os dados experimentais (medidas da temperatura no tempo  $> 0$ ) são obtidos adicionando uma perturbação de erro aleatória à solução exata do problema direto

$$T^{\text{experimental}} = T^{\text{exato}} + \sigma r, \quad (2.12)$$

onde  $\sigma$  é a divergência padrão do erro e  $r$  é a variável ruidosa. Os dados apresentados foram obtidos utilizando  $t=0.01$  e  $\sigma=0.05$ .

Sendo a resolução espacial  $N=101$ , de acordo com o princípio da discrepância de Morozov, o parâmetro de regularização  $\alpha$  é escolhido de modo que  $R(f) \approx N\sigma^2 = 0.2525$ . O resíduo é definido por:

$$R(f) = \left\| T^{\text{experimental}} - T^{\text{calculado}} \right\|_2^2 \quad (2.13)$$

e o erro entre a solução aproximada (ou calculada) da solução  $f_\alpha$  e a solução exata é definida por:

$$E(f_\alpha) = \left\| f_\alpha - f^{\text{exato}} \right\|_2^2 \quad (2.14)$$

O problema de otimização é iterativamente resolvido pela rotina otimizadora quase-newtoniana da Biblioteca Fortran NAG - função E04UCF (1997), com métrica variável. O algoritmo é projetado para minimizar uma função objetivo arbitrária usando um método de programação seqüencial.

As Tabelas 2.1 e 2.2 apresentam os valores dos resíduos  $R(f_\alpha)$  e erros  $E(f_\alpha)$ , obtidos com diferentes  $\alpha$ . A Tabela 2.1 apresentando os resultados obtidos com função triangular e  $q=0.5$ . Enquanto a Tabela 2.2, apresentando os resultados da função triangular com  $q=2.0$ .

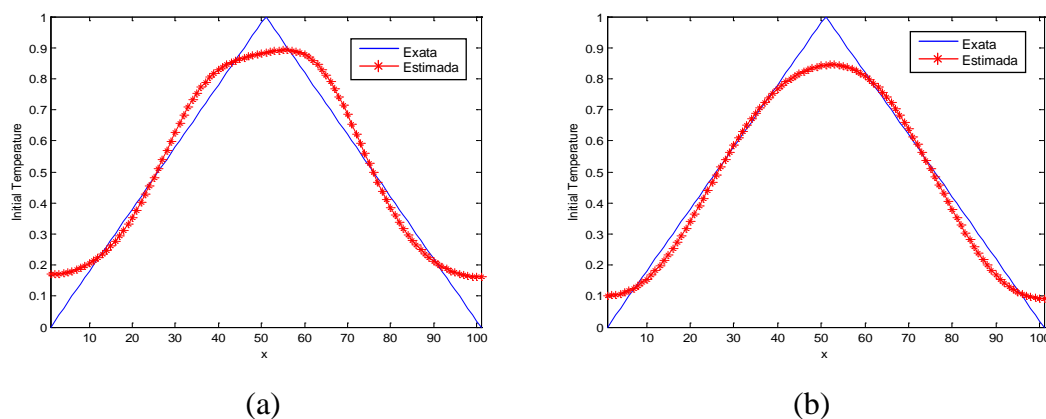
**TABELA 2.1 – FUNÇÃO TRIANGULAR COM  $q=0,5$**

$\alpha$	$R(f_\alpha)$	$E(f_\alpha)$
0.0001	0.1853	2.7298
0.0003	0.1856	0.5388
0.0010	0.1861	0.3443
<b>0.0285</b>	<b>0.2525</b>	<b>0.3994</b>
0.0999	0.7728	0.8684

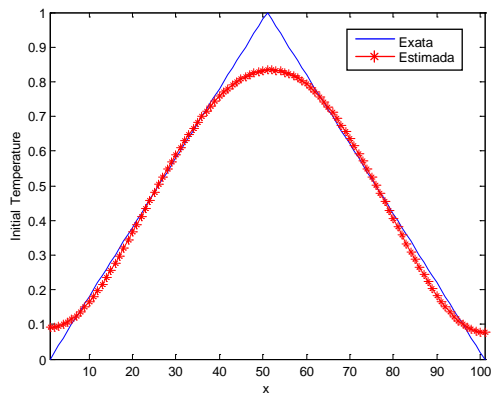
**TABELA 2.2 – FUNÇÃO TRIANGULAR COM  $q=2,0$**

$\alpha$	$R(f_\alpha)$	$E(f_\alpha)$
0.0001	0.1851	4.9359
0.0003	0.1854	0.6740
0.0010	0.1856	0.2113
<b>0.0346</b>	<b>0.2525</b>	<b>0.2400</b>
0.0999	0.6687	0.7207

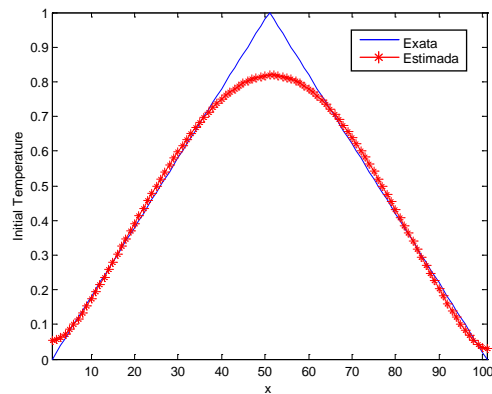
O mesmo processo é realizado para diferentes valores de  $q$ : 1,5 e 2,5. As figuras 2.2 e 2.3 apresentam as estimações das condições iniciais da função triangular, recuperadas utilizando o princípio de Morozov para a determinação do parâmetro de regularização.



**Figura 2.2 – Reconstrução com 5% de ruído, com  $\alpha$  determinado pelo princípio de Morozov: (a)  $q=0,5$  e (b)  $q=1,5$ .**



(a)



(b)

**Figura 2.3 – Reconstrução com 5% de ruído, com  $\alpha$  determinado pelo princípio de Morozov: (a)  $q=2,0$  e (b)  $q=2,5$ .**

Nas tabelas (2.3) e (2.4) são apresentados os resultados obtidos com função semitriangular para  $q=0,5$  e  $q=2,0$ , respectivamente.

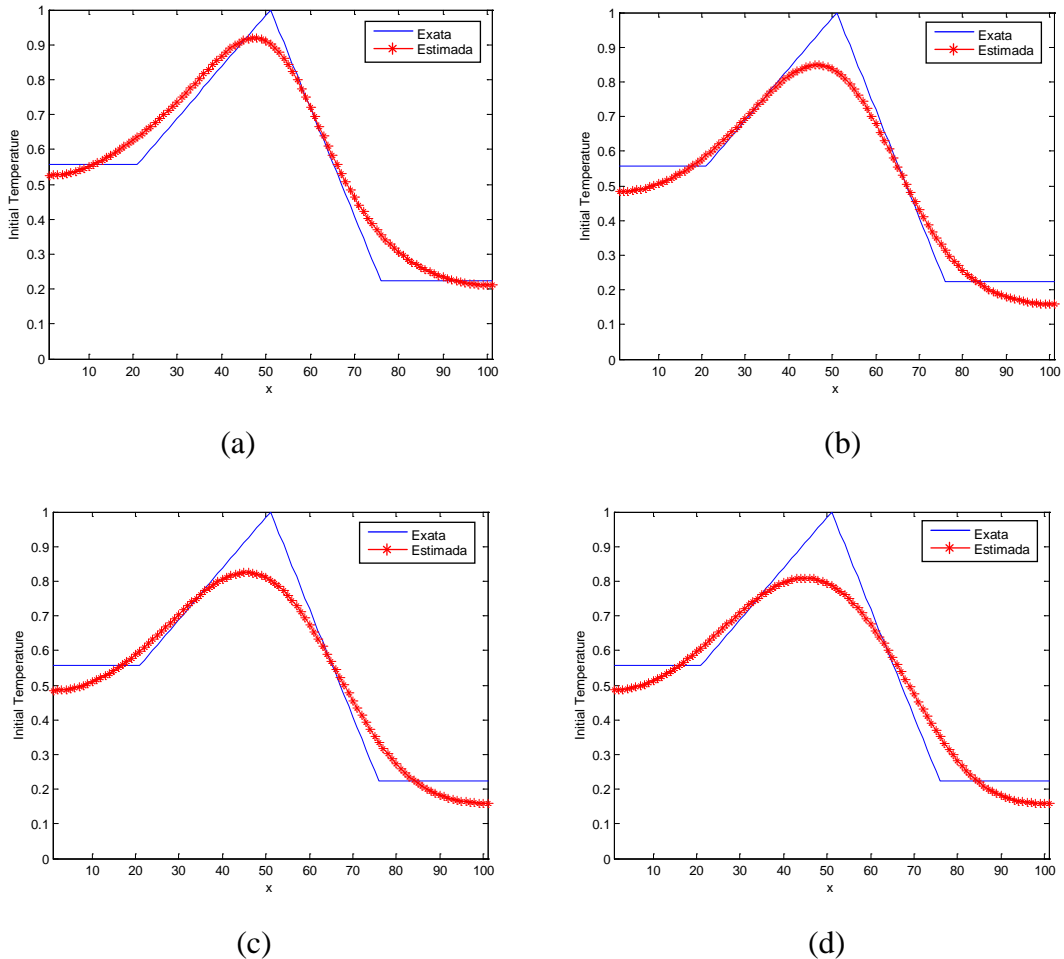
**TABELA 2.3 – FUNÇÃO SEMITRIANGULAR COM  $q=0,5$**

$\alpha$	$R(f_\alpha)$	$E(f_\alpha)$
0.0001	0.1852	2.3979
0.0003	0.1854	0.3578
0.0010	0.1856	0.1383
<b>0.0414</b>	<b>0.2525</b>	<b>0.2370</b>
0.0999	0.5174	0.5977

**TABELA 2.4 – FUNÇÃO SEMITRIANGULAR COM  $q=2,0$**

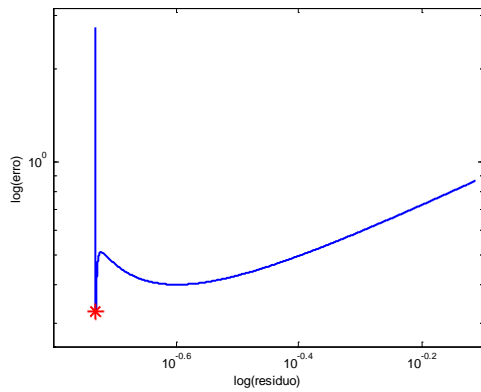
$\alpha$	$R(f_\alpha)$	$E(f_\alpha)$
0.0001	0.1851	4.0920
0.0003	0.1854	0.4455
0.0010	0.1856	0.1807
<b>0.0419</b>	<b>0.2525</b>	<b>0.4379</b>
0.0999	0.5131	0.8651

A figura 2.3 apresenta as estimações das condições iniciais recuperada para a função semitriangular para  $q=0,5$ ;  $q=1,5$ ;  $q=2,0$  e  $q=2,5$ .

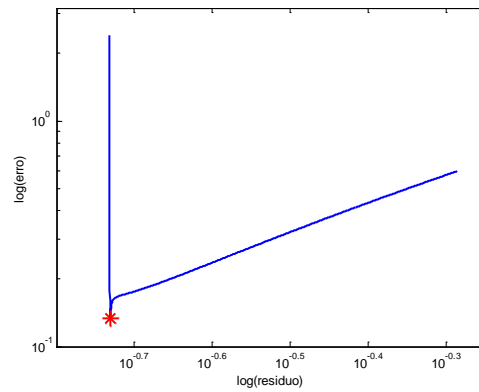


**Figura 2.3 – Reconstrução com 5% de ruído, com  $\alpha$  determinado pelo princípio de Morozov: (a)  $q=0,5$ ; (b)  $q=1,5$ ; (c)  $q=2,0$  e (d)  $q=2,5$ .**

Um outro critério para encontrar o parâmetro de regularização também foi investigado, e é baseado na máxima curvatura na *curva-L* (Hansen, 1992). As Figuras (2.4a) e (2.4b) apresentam as curvas-L para a função teste triangular usando  $q=0,5$  e  $q=2,0$ , respectivamente. As curvas-L para a função teste semitriangular, são apresentadas nas Figuras (2.5a) e (2.5b). O parâmetro de regularização  $\alpha$  é escolhido como o canto da *curva-L*.

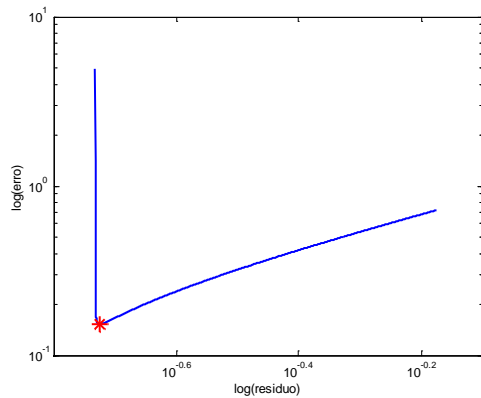


(a)

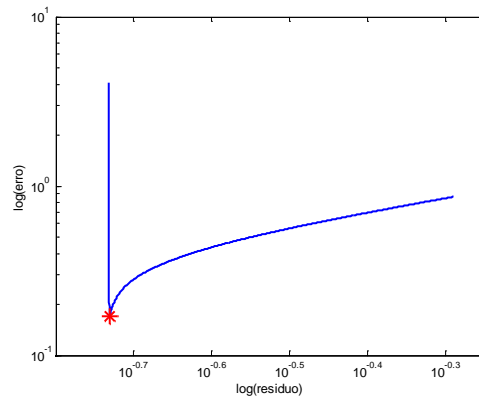


(b)

**Figura 2.4 – Curva-L para a função teste triangular: (a)  $q=0,5$ ; (b) 2,0.**



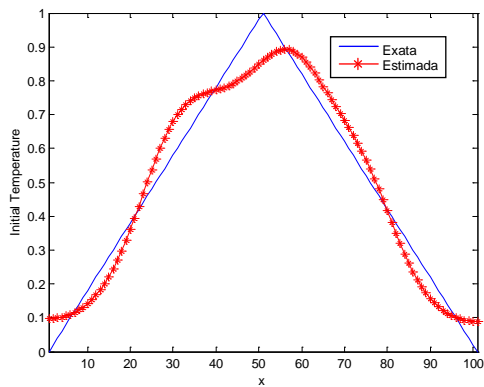
(a)



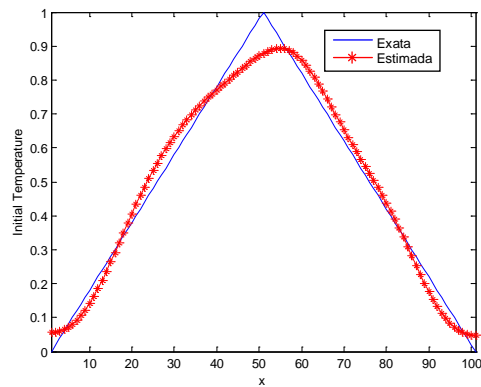
(b)

**Figura 2.5 – Curva-L para a função teste semitriangular: (a)  $q=0,5$ ; (b) 2,0.**

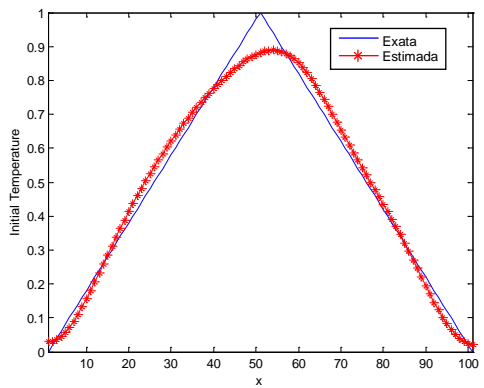
Reconstruções com  $\alpha$  obtidos pelo critério de Hansen são apresentadas nas Figuras (2.6 a-d), para a função teste triangular e as Figuras (2.7 a-d) para a função semitriangular.



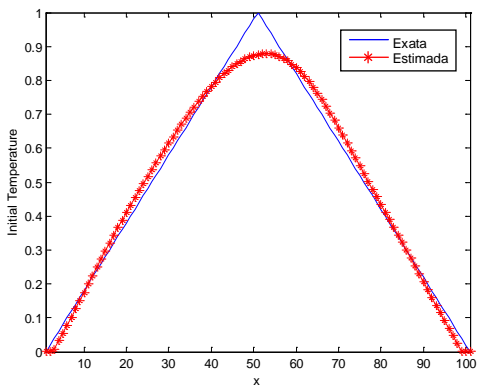
(a)



(b)



(c)



(d)

**Figura 2.6 – Reconstrução com 5% de ruído,  $\alpha$  determinado pelo critério de Hansen: (a)  $q=0,5$ ; (b)  $q=1,5$ ; (c)  $q=2,0$  e (d)  $q=2,5$ .**

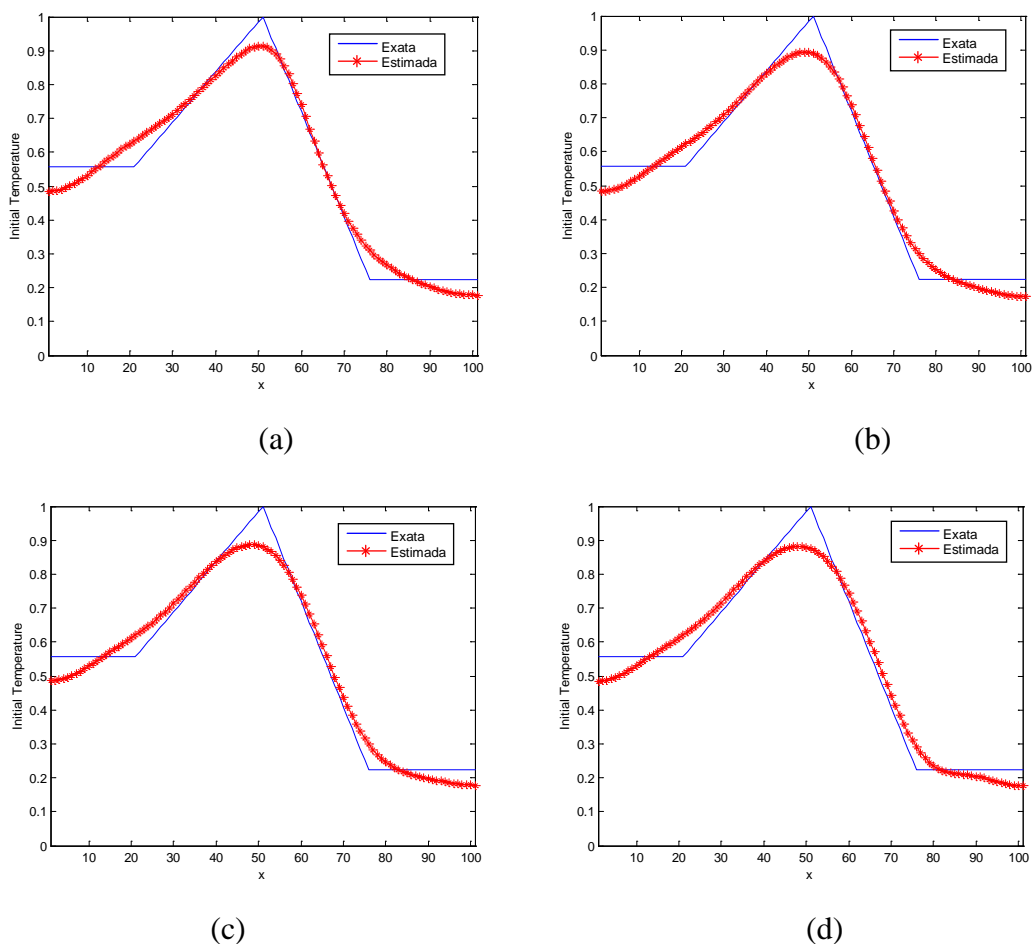
Pode-se observar que foram obtidas boas reconstruções da condição inicial para diferentes valores de  $q$ . Os erros de cada função que são apresentados na tabela 2.5 são calculados pela equação (2.14).

**TABELA 2.5 – FUNÇÃO TRIANGULAR**

$q$	$\alpha$	$E(f_\alpha)$
0,5	0,0006	0,3261
1,5	0,0007	0,1705
2,0	0,0014	0,1339
2,5	0,0055	0,1319



Para mostrar a eficiência do critério baseado na máxima curvatura na *curva-L*, a função semi-triangular também foi utilizada como teste. Os resultados são apresentados na figura 2.7 e os valores dos erros na tabela 2.6.



**Figura 2.7 – Reconstrução com 5% de ruído,  $\alpha$  determinado pelo critério de Hansen: (a)  $q=0,5$ ; (b)  $q=1,5$ ; (c)  $q=2,0$  e (d)  $q=2,5$ .**

Considerando que se trata de um problema mal-posto com 5% de ruído gaussiano, conclui-se que o critério de Hansen também pode ser usado para a estimação do parâmetro de regularização. Os valores dos erros são apresentados na tabela 2.6.

**TABELA 2.6 – FUNÇÃO SEMI-TRIANGULAR**

$q$	$\alpha$	$E(f_\alpha)$
0,5	0,0067	0,1532
1,5	0,0021	0,1682
2,0	0,0023	0,1629
2,5	0,0020	0,1857

Para a função semi-triangular o melhor resultado foi obtido com  $q=0,5$ , no entanto a diferença não prejudica a análise da inversão. Os resultados obtidos pela *curva-L* foram um pouco melhores quando comparados aos obtidos com o uso do critério de Morozov.

Conclui-se que o uso do novo operador da máxima entropia não extensiva apresenta bons resultados na reconstrução da condição inicial da equação de calor. E que tanto o princípio da discrepância de Morozov e o critério de Hansen podem ser usados para a determinação do parâmetro de regularização.

## CAPÍTULO 3

### REDES NEURAIIS ARTIFICIAIS

As Redes Neurais Artificiais também são conhecidas por Sistemas Conexionistas. A pesquisa e o desenvolvimento das várias aplicações destes sistemas deram origem a área de Computação Neural, a qual ficou popularmente conhecida como Neurocomputação.

McCulloch e Pitts (1943) apresentaram um modelo da atividade de um neurônio biológico e usando elementos lógicos simples, mostraram como as redes feitas de muitas dessas unidades interconectadas poderiam realizar operações lógicas; Roseblatt (1958) desenvolveu o “perceptron” e Widrow e Hoff em 1960, concebeu a “ADALINE”. Depois de um período de pouca pesquisa na área, na década de 70, após a publicação do livro *"Perceptrons"* de Minsky e Papert (1969), as pesquisas em redes neurais ressurgiram com a publicação do artigo de Hopfield (1982). Desde então, as atividades na área têm crescido exponencialmente (Braga et al.,1998; Haykin 2001).

As redes neurais artificiais são inspiradas no funcionamento dos neurônios biológicos, têm como principais propriedades e capacidades (Haykin, 2001):

- não-linearidade,
- mapeamento de entrada-saída,
- adaptabilidade,
- aprender com exemplos,
- processamento paralelo e distribuído,
- a capacidade de generalização e
- tolerância à falhas.

A aplicação de uma rede neural é feita em 2 fases: a de aprendizagem e a de ativação.

Uma característica importante das redes neurais artificiais é a aprendizagem através de exemplos. Nos modelos hetero-associativos, a medida que cada exemplo é apresentado a rede automaticamente ajusta-se para relacioná-lo com uma saída. Tipicamente, essa forma constitui um modelo que representa o relacionamento entre as variáveis de entrada e saída.

Nas redes neurais a informação é totalmente distribuída, a informação é propagada entre os pesos, que são ajustados durante a aprendizagem. Esses pesos (conexões) são as unidades de memória das redes e seus valores representam o estado corrente do conhecimento na rede (Tsoukalas e Uhrig, 1995).

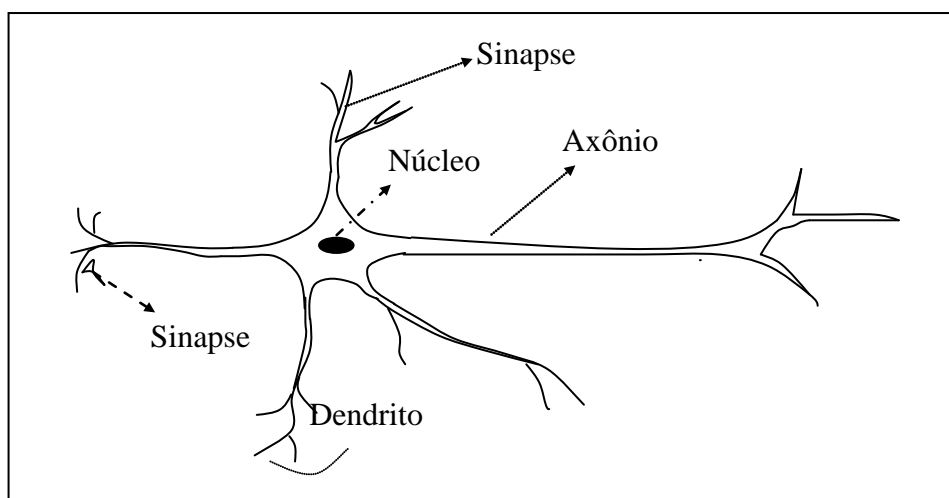
As redes neurais também são tolerantes a falhas, podendo tomar decisões com dados incompletos, ou seja, com o conhecimento totalmente distribuído não serão observadas falhas se houver falha de uma porcentagem das entradas.

A computação das redes neurais consiste em interconectar as unidades que agem instantaneamente nos dados de forma paralela. Embora os computadores digitais tenham que simular esse paralelismo, redes neurais implementadas em hardware realizam essas operações em paralelo, sendo extremamente rápidas (Tsoukalas e Uhrig, 1995).

### **3.1 Neurônios Biológicos**

A célula nervosa, ou neurônio, foi identificada no século XIX pelo neurologista Ramón Cajal. O neurônio possui um corpo celular, ou soma, que é o centro dos processos metabólicos da célula nervosa, a partir dele projetam-se extensões filamentosas, os dendritos e o axônio (Figura 3.1). Os dendritos cobrem um volume muitas vezes maior que o próprio corpo celular e formam uma árvore dendrital, já o axônio conecta uma célula nervosa a outras através de suas ramificações (Kovács, 1996).

O neurônio biológico pode ser visto como sendo o dispositivo computacional elementar básico do sistema nervoso, com muitas entradas e uma saída. As entradas ocorrem através das conexões sinápticas, que conectam a árvore dendrital aos axônios de outras células nervosas. Os sinais que chegam por estes axônios são pulsos elétricos conhecidos como impulsos nervosos e constituem a informação que o neurônio processará, para produzir como saída um impulso nervoso no seu axônio.



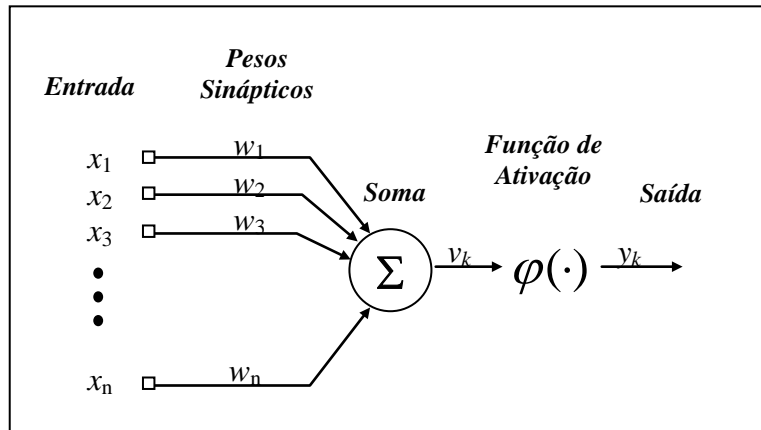
**Figura 3.1 – Esquema da célula neural**

FONTE: Modificado de Braga et al (2000, p. 6).

As sinapses são regiões eletroquímicas ativas, compreendidas entre duas membranas. A pré-sináptica, por onde chega um estímulo proveniente de uma outra célula e a pós-sináptica, que é a do dendrito. Nesta região intersináptica, o estímulo nervoso que chega à sinapse é transferido à membrana dendrital através dos neurotransmissores, que dependendo do tipo, dará origem a uma conexão excitatória ou inibitória. Uma conexão excitatória provoca uma alteração no potencial da membrana que contribui para a formação de um impulso nervoso no axônio de saída, enquanto que uma conexão inibitória age no sentido oposto (Kovács, 1996).

## 3.2 Neurônios Artificiais

Os componentes de um neurônio artificial têm uma analogia direta com os componentes dos neurônios biológicos. A Figura 3.2 mostra a representação esquemática de um neurônio artificial.



**Figura 3.2 – Representação de um neurônio artificial.**

FONTE: Haykin (2000, p. 36).

Assim, como os neurônios biológicos possuem muitas entradas, dadas pelos níveis de estímulos, os neurônios artificiais também têm inúmeras entradas que são apresentadas ao mesmo tempo. Cada uma dessas entradas é amplificada por um peso (ou peso sináptico), representado por  $w_1, w_2, \dots, w_n$ . Os pesos são coeficientes adaptáveis dentro da rede e determinam a intensidade do sinal da entrada. Cada entrada é multiplicada por um peso sináptico, que afeta o seu valor semelhantemente à junção sináptica dos neurônios biológicos. Os produtos das entradas pelos pesos são somados. Matematicamente esse processo pode ser dado pela equação:

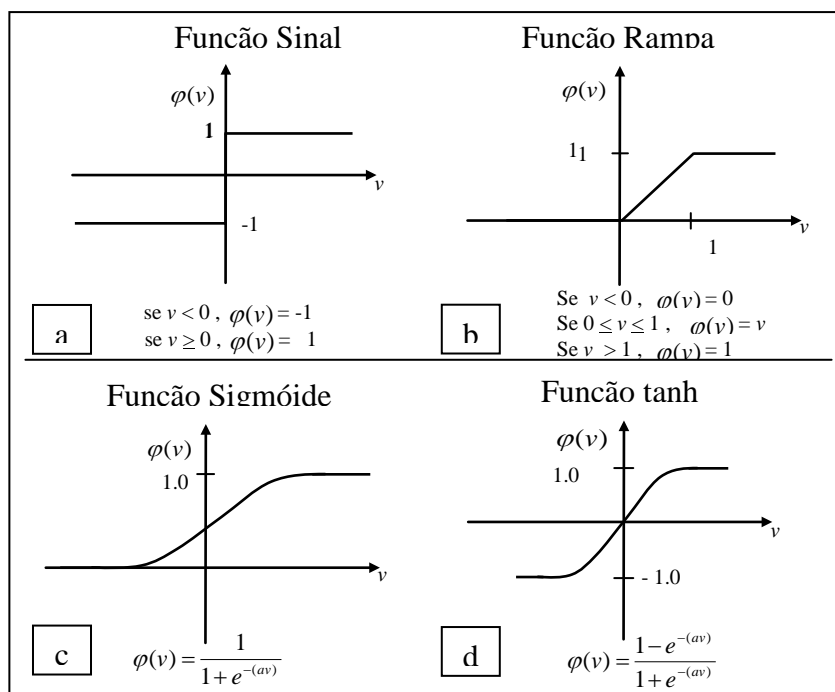
$$v_j = \sum_{k=1}^n w_{jk} x_k \quad (3.1)$$

Na equação (3.1),  $x_k$  para  $k=1, 2, \dots, n$  são os componentes do vetor de entrada e  $w_{jk}$  representa o peso sináptico que conecta a entrada  $j$  com o neurônio  $k$ . O sinal de atividade interna total ( $v$ ) do neurônio  $j$  é o produto interno entre o vetor  $\bar{x}$  e o vetor  $\bar{w}$ .

Semelhantemente aos neurônios biológicos, que emitem um pulso se a atividade elétrica interna supera um limiar – 50mv (Braga et al., 1998), a atividade interna de um neurônio artificial passa por uma função de ativação para produzir um efeito de emissão de um pulso.

### 3.3 Funções de Ativação

A partir do modelo de McCulloch e Pitts foram derivados vários outros modelos que permitem a produção de uma saída qualquer, não necessariamente zero ou um, e com diferentes funções de ativação. A Figura (3.3 a) consiste na função sinal, a (3.3-b) a função rampa, a (3.3-c) a sigmóide e a (3.3-d) na tangente hiperbólica.



**Figura 3.3 – Funções de ativação**

FONTE: modificado de Braga et al (2000, p. 10).

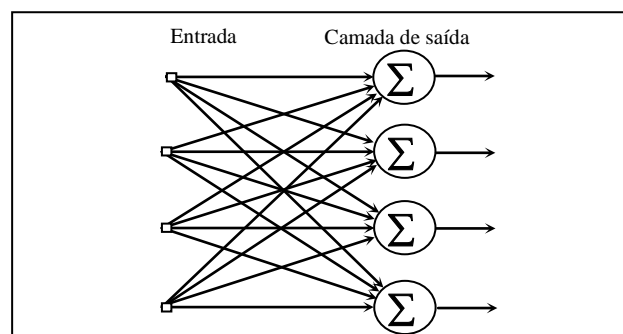
As funções sigmóide e tangente hiperbólica são utilizadas em alguns modelos de redes que requerem a derivada da função no processo de aprendizagem.

### 3.4 Arquitetura das Redes Neurais Artificiais

Através da combinação dos neurônios artificiais pode-se gerar diferentes arquiteturas de redes. Os neurônios em geral são organizados em camadas. Alguns modelos de redes neurais são constituídos somente por uma camada. As redes com duas ou mais camadas são denominadas redes de camadas múltiplas. Quando há realimentação de alguma dessas camadas a rede é chamada de recorrente.

Como mencionado, numa arquitetura mais simples, os neurônios formam uma rede com uma única camada, Figura (3.4). Nela as entradas projetam-se diretamente sobre a camada de saída. Como não há realimentação na rede denomina-se rede “feed-forward” (de alimentação para frente).

A conexão é total quando todas as saídas de uma camada estão conectadas em todos os neurônios da próxima camada. Na falta de alguma dessas conexões, a rede é referida como parcialmente conectada.



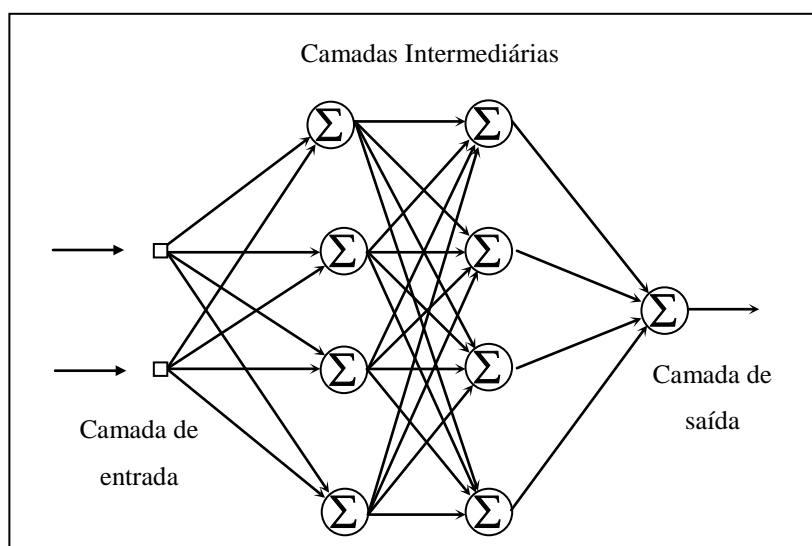
**Figura 3.4 – Rede com uma única camada**

FONTE: Modificada de Haykin (2000, p. 47).



Os neurônios também podem estar dispostos de modo a formar uma arquitetura composta com duas ou mais camadas. Na Figura 3.5 tem-se um exemplo, onde a primeira camada é a entrada, a última camada de saída e as outras, que não têm contato com o meio externo, são denominadas camadas escondidas, compostas pelos chamados neurônios escondidos.

Com a existência de uma ou mais camadas escondidas, a rede torna-se capaz de extrair características de ordem elevada, adquirindo uma perspectiva global, apesar de sua conectividade local (Haykin, 1994).



**Figura 3.5 – Rede de três camadas**

FONTE: Modificada de Haykin (2000, p. 48).

### 3.5 Funcionamento das Redes Neurais

O processo de aprendizagem, também conhecido como treinamento, consiste no processo de adaptação dos pesos sinápticos das conexões e dos níveis de limiar (*bias*) dos neurônios em resposta às entradas. A ativação consiste no processo de receber uma entrada e produzir uma saída de acordo com a arquitetura final da rede.

Existem dois principais paradigmas de aprendizado: o aprendizado supervisionado e o aprendizado não supervisionado, que indicam como os pesos das conexões devem ser ajustados de acordo com a resposta da ativação. No aprendizado supervisionado, apresenta-se uma resposta desejada para uma dada entrada. Já no aprendizado não supervisionado, somente o vetor de características é apresentado à entrada da rede. A rede, então, se auto-organiza para que cada neurônio responda a diferentes conjuntos de entrada (Haykin, 2001; Braga et al., 1998).

Uma rede é treinada até que um número máximo de épocas de treinamento seja atingido ou um erro objetivo seja alcançado. Tem-se uma época de treinamento a cada apresentação de todo o conjunto de treinamento. Deve-se observar que a condição de parada relativa ao erro objetivo, só pode ser usada por redes com aprendizagem supervisionada, pois se pode calcular o erro entre a saída desejada e a saída obtida.

Existem diferentes tipos de redes neurais que se diferenciam principalmente na arquitetura e no algoritmo de aprendizagem. Cada uma delas oferecendo vantagens específicas.

### **3.6 Redes Perceptron**

O Perceptron foi proposto por Rosenblatt em 1958 (Haykin, 2001; Braga et al. 1998), sendo o primeiro modelo de rede neural com aprendizagem supervisionada. É uma rede simples de uma única camada com pesos e limiares ajustáveis. Rosenblatt provou que se os padrões usados para treinar o perceptron são retirados de duas classes linearmente separáveis, então o algoritmo do perceptron converge e posiciona a superfície de decisão na forma de um hiperplano entre as duas classes. A prova de convergência do algoritmo é conhecida com o teorema de convergência do perceptron (Haykin, 2001; Braga et al., 1998). Um único neurônio perceptron limita-se a realizar classificação de padrões com apenas duas classes. Expandindo a camada de saída do perceptron para incluir mais de um neurônio, pode-se classificar mais de duas classes, no entanto as classes devem ser linearmente separáveis.

No processo de adaptação, ou aprendizado, deseja-se obter o valor do incremento  $\Delta \bar{w}$  a ser aplicado ao vetor de pesos  $\bar{w}$  de forma que seu valor atualizado  $\bar{w}(t+1) = \bar{w}(t) + \Delta \bar{w}$  esteja mais próximo da solução desejada que  $\Delta \bar{w}$ .

O erro é calculado através da diferença entre a saída obtida e a saída desejada e é dado pela equação:

$$e_k(n) = d_k(n) - y_k(n) \quad (3.2)$$

onde  $d_k(n)$  e  $y_k(n)$  são os sinais de saída desejado e obtido respectivamente no neurônio  $k$  no tempo  $n$ .

O cálculo dos novos pesos e dos novos limiares é obtido por:

$$w(n+1) = w(n) + \eta x(n)e(n) \quad e \quad (3.3)$$

$$b(n+1) = b(n) + \eta b(n)e(n) \quad (3.4)$$

onde  $w(n)$ ,  $x(n)$ ,  $e(n)$  e  $b(n)$  são respectivamente, os pesos, a entrada, o erro e o limiar no tempo  $n$  e  $\eta$  é a taxa de aprendizagem.

### 3.7 Rede Funções de Base Radial

A Rede de Função Base Radial foi introduzida por Powell (1985), com o intuito de resolver o problema de ajuste de curva num espaço de alta dimensionalidade onde sua generalização é equivalente ao uso da superfície de múltiplas dimensões para interpolar os dados de teste.

A rede é do tipo feed-forward e possui três camadas: a camada de entrada da rede, uma camada escondida e a camada de saída. O algoritmo de aprendizado da rede é supervisionado, no entanto as conexões entre as camadas de entrada e escondida, diferentemente da maioria das redes de camadas múltiplas, representam centros de classes.

O processo de ativação da rede é dado pela equação:

$$F(x) = \sum_{k=1}^m w_k \varphi_k(x) \quad (3.5)$$

onde  $w_k$  são os pesos do neurônio,  $m$  é o número de neurônios e  $\{\varphi_k(x) | k = 1, 2, \dots, m\}$  é conjunto de funções de base, dadas por:

$$\varphi_k(x) = G(\|x - t_k\|), \quad k=1, 2, \dots, m \quad (3.6)$$

onde  $t_k$  para  $k=1, 2, \dots, m$  é o conjunto de centros. Assim, pode-se redefinir  $F(x)$  como sendo

$$F(x) = \sum_{k=1}^m w_k G(x, t_k) \quad (3.7)$$

sendo  $G$ , uma função de Green.

$$F(x) = \sum_{k=1}^m w_k G(\|x - t_k\|) \quad (3.8)$$

A aprendizagem consiste na determinação de um novo conjunto de pesos que minimize o custo  $\varepsilon(F)$  definido por:

$$\varepsilon(F) = \sum_{k=1}^N \left( d_k - \sum_{j=1}^m w_j G \|x_k - t_j\| \right)^2 \quad (3.9)$$

Há algumas estratégias de aprendizado que podem ser implementadas numa Rede Funções Base Radial que dependem de como os centros das Funções Base Radial são especificados (Haykin, 1994). Dois deles são apresentados a seguir:

### ***1. Centros Fixos Selecionados Aleatoriamente.***

Na estratégia mais simples assume-se funções de Base Radial fixas. As localizações dos centros são escolhidas aleatoriamente no conjunto de treinamento. Pode-se empregar funções Gaussianas isotrópicas, cujo desvio padrão é fixado de acordo com espalhamento dos centros. Especificamente, uma função de base radial centrada em  $\bar{t}_k$ , é definida como:

$$G(\|x - t_k\|^2) = \exp\left(-\frac{M}{d^2} \|x - t_k\|^2\right), \quad k=1,2,\dots,m. \quad (3.10)$$

onde  $M$  é o número de centros e  $d$  é a distância máxima entre os centros escolhidos. O desvio padrão ( $\sigma$ ) de todas as funções de base radial são fixadas em:

$$\sigma = \frac{d}{\sqrt{2M}} \quad (3.11)$$

Esta fórmula assegura que as funções de base radial individuais não sejam pontiagudas demais ou planas demais; estas duas condições extremas devem ser evitadas. Com uma alternativa para Equação (3.11), pode-se usar centros escalados

individualmente com larguras maiores em áreas de menos densidade de pontos, o que requer experimentação com os dados do treinamento.

Os únicos parâmetros que devem ser aprendidos nesta abordagem são os pesos lineares na camada de saída da rede.

## 2. Seleção Auto-organizada de Centros

Uma outra estratégia de aprendizado tem um processo híbrido constituído de dois diferentes estágios. O primeiro consiste na aprendizagem auto-organizada, para estimar os centros das funções de base radial da camada escondida. O segundo consiste na aprendizagem supervisionada, na camada de saída.

No estágio da aprendizagem auto-organizada, usa-se um algoritmo de agrupamento para agrupar dados similares, que coloca os centros das funções de base radial nas regiões de espaço de entrada. Um desses algoritmos pode ser o agrupamento de  $k$ -médias (Duda e Hart, 1973), nele os valores dos centros são iniciados aleatoriamente e um vetor de entrada é selecionado do conjunto de treinamento e é feito um casamento de similaridade entre este vetor e o centro da classe, utilizando a distância euclidiana mínima:

$$k(x) = \arg \min \|\bar{x}(n) - \bar{t}_k(n)\|, \quad k = 1, 2, \dots, m \quad (3.12)$$

onde  $\arg \min$  corresponde ao argumento da distância mínima entre a entrada  $x$  e o centro da  $k$ -ésima função de base radial  $t$  na iteração  $n$ ,  $k(x)$  representa o índice do centro com a menor distância.

O ajuste dos centros das funções de base radial é obtido por:

$$t_k(n+1) = \begin{cases} t_k(n) + \eta[x(n) - t_k(n)], & k = k(x) \\ t_k(n), & \text{c.c.} \end{cases} \quad (3.13)$$

onde  $\eta$  é a taxa de aprendizagem.

No processo de ajuste dos pesos e limiares da camada de saída, usa-se o algoritmo de mínimos quadrados.

### 3.8 Rede neural com algoritmo de Correlação em Cascata

Uma das maiores dificuldades no projeto de uma rede neural é obter uma arquitetura adequada para realizar uma determinada tarefa. Diferentes tarefas e aplicações específicas requerem diferentes tamanhos de redes. Pequenas redes podem requerer muitas épocas de treinamento e/ou resultar numa quantidade insuficiente de parâmetros livres para obter todas as características necessárias dos dados de entrada para fazer uma caracterização possível (Silva, 1997).

A Rede neural com algoritmo de Correlação em Cascata (RCC) foi proposta por Falhman (1991), ela tem a capacidade de determinar seu próprio tamanho e topologia. A rede possui duas idéias chave: a arquitetura em cascata, na qual os neurônios são adicionados à rede um de cada vez e não mudam depois de terem sido adicionados, e o algoritmo de aprendizado, que cria e instala os novos neurônios escondidos. A RCC tem como principal característica a determinação de seu próprio tamanho e topologia.

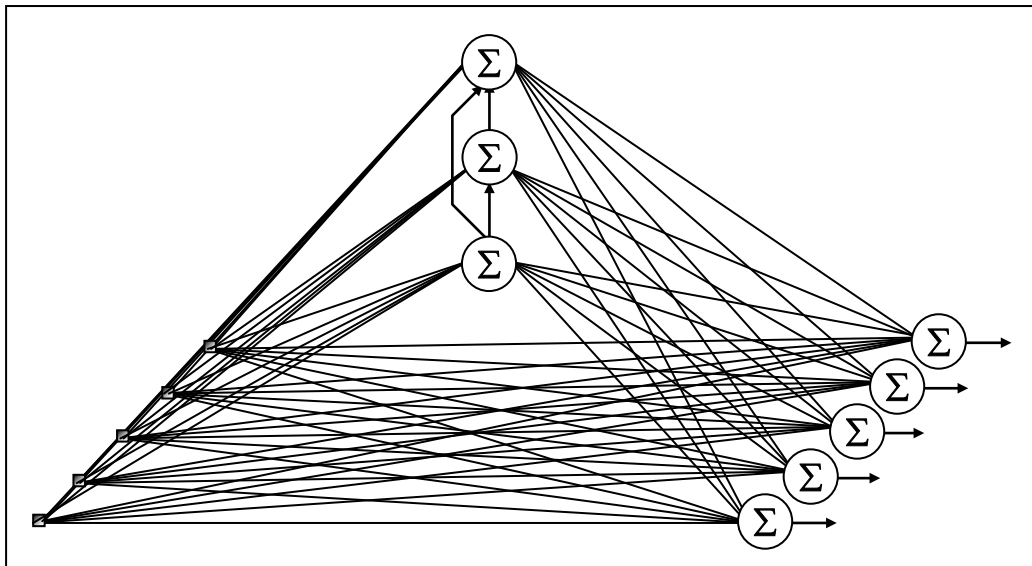
A rede é iniciada sem neurônios na camada escondida e é treinada até que seja atingido o valor do erro residual desejado ou que o erro estabilize num mínimo local. Durante a aprendizagem caso o valor do erro não seja atingido, automaticamente um novo neurônio é treinado e adicionado em forma de cascata, criando uma estrutura de camadas múltiplas. Para cada nova unidade escondida a ser adicionada maximiza-se a magnitude da correlação entre a saída da nova unidade e o sinal do erro residual da rede.

Os pesos então são congelados e somente os pesos e limiares da última camada são treinados. A arquitetura da RCC é apresentada na Figura 3.6.

Um novo neurônio escondido a ser adicionado à rede recebe conexões da entrada externa e de todos os neurônios escondidos já existentes, o limiar e os pesos das conexões são iniciados aleatoriamente, o treinamento desses dados é obtido com o objetivo de maximizar a magnitude da correlação  $S$  entre a saída do novo neurônio e o erro residual que se deseja eliminar:

$$S = \sum_o \left| \sum_i (v_i - \bar{v})(e_{i,o} - \bar{e}_o) \right| \quad (3.14)$$

onde  $v_i$  é o valor de saída da unidade candidata (Equação 3.1),  $\bar{v}$  é a média dos valores de todos os padrões,  $e_{i,o}$  é o erro residual produzido pelo padrão  $i$  na saída  $o$  e  $\bar{e}_o$  é a média do erro residual de todos os padrões na saída  $o$ .



**Figura 3.6 – Arquitetura da RCC**



Para maximizar  $S$ , calcula-se  $\partial S / \partial w_i$ , a derivada parcial de  $S$  com relação a cada peso da unidade candidata,  $w_i$ . De uma maneira similar à derivação ao algoritmo de retropropagação, pode-se expandir e diferenciar a fórmula de  $S$ :

$$\frac{\partial S}{\partial w_i} = \sum_{i,o} \sigma_o (e_{i,o} - \bar{e}_o) \varphi'_i x_{k,i} \quad (3.15)$$

sendo  $\sigma_o$  o sinal de correlação entre o valor da unidade candidata e a saída  $o$ ,  $\varphi'_p$  a derivada da função de ativação correspondente à saída do padrão  $p$  da unidade candidata em relação à soma de suas entradas e  $x_{k,p}$  a entrada que o neurônio recebe da unidade  $k$  para o padrão  $p$ .

O limiar e os pesos das conexões entre os neurônios escondidos e a saída da rede podem ser treinados utilizando-se a regra delta ou de Widrow-Hoff, uma vez que não há a necessidade de realimentar o erro para as unidades escondidas.

### 3.9 Rede Perceptrons de Camadas Múltiplas

As redes de camadas múltiplas alimentadas adiante são uma importante classe de redes neurais. Os Perceptrons de Camadas Múltiplas pertencem a esta classe e têm sido aplicados com sucesso para resolver diversos problemas difíceis, através de seu treinamento de forma supervisionada com um algoritmo bastante popular conhecido como retropropagação do erro (error back-propagation) (Haykin, 2001; Braga et al., 2000).

O algoritmo de retropropagação do erro foi apresentado por Rumelhart et. al (1986), sendo dado um passo importante para o ressurgimento da área de redes neurais.

Entretanto, desenvolvimentos paralelos também foram realizados por Werbos (1989) e Parker em (1987).

A Rede Neural Perceptron de Camadas Múltiplas, é composta por uma camada de entrada, uma ou mais camadas escondidas e uma camada de saída. Seu algoritmo de treinamento, aprendizagem por retropropagação do erro, consiste de dois passos através das diferentes camadas da rede: um passo para frente, a propagação e um passo para trás, a retropropagação.

Na fase de propagação da rede o sinal de entrada se propaga para frente, camada por camada até que um conjunto de saídas seja obtido na última camada. Na primeira camada escondida, multiplica-se cada unidade de entrada por um peso correspondente a cada neurônio. O número de neurônios na camada escondida varia de acordo com a aplicação. Os resultados das multiplicações são somados e a soma (Equação 3.16) passa por uma função de ativação do tipo sigmoidal (Equação 3.17).

$$v_j(n) = \sum_{i=1}^p x_i w_{ij} \quad (3.16)$$

$$y_j(n) = \varphi_j(v_j(n)) \quad (3.17)$$

Nas camadas seguintes o mesmo processo é repetido, sendo que a entrada da próxima camada é a saída da anterior. Nesta rede além de poder variar o número de neurônios na camada escondida, também é possível variar o número de camadas.

Durante o processo de propagação, os pesos sinápticos da rede são todos fixos. Durante o passo para trás, os pesos sinápticos são ajustados de acordo com uma regra de correção do erro. A resposta real da rede é subtraída de uma resposta desejada (alvo) para produzir um sinal de erro. Esse sinal de erro é então propagado para trás através da rede, contra a direção das conexões sinápticas, daí o nome de retropropagação.

Os pesos sinápticos são ajustados para fazer com que a resposta real da rede se mova mais para perto da resposta desejada, num sentido estatístico (Haykin, 2001). O processo de aprendizado termina quando se atinge um erro desejado ou quando é realizado um número máximo de épocas de aprendizado.

Os novos pesos são obtidos pela equação:

$$w_{ji}^L(n+1) = w_{ji}^L(n) + \eta \delta_j^L(n) y_i^{L-1}(n) \quad (3.18)$$

onde o gradiente da última camada é obtido pela equação:

$$\delta_j(n) = -\frac{\partial \varepsilon(n)}{\partial v_j(n)} = -e_j(n) \varphi'_j(v_j(n)). \quad (3.19)$$

O gradiente nas camadas escondidas, é dado pela equação

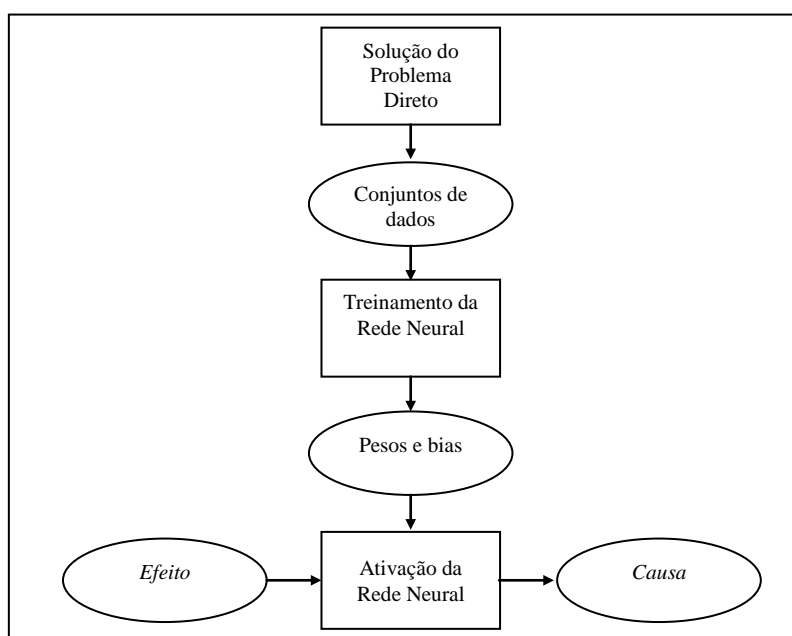
$$\delta_j(n) = \varphi'_j(v_j(n)) \sum_k \delta_k(n) w_{kj}(n) \quad (3.20)$$

## CAPÍTULO 4

### REDES NEURAIAS EM PROBLEMAS INVERSOS DE CONDUÇÃO DE CALOR

Essa dissertação tem como objetivo o uso das Redes Neurais Artificiais na solução de dois Problemas Inversos. Estuda-se a solução do problema inverso de condução de calor na recuperação da condição de contorno e na recuperação da condição inicial. Avalia-se o desempenho de diferentes redes neurais artificiais, treinadas com diferentes conjuntos de dados. Testa-se a capacidade de generalização das redes e o desempenho das inversões quando adicionado ruído.

O uso das RNA na solução de problemas inversos envolve: geração dos conjuntos de dados, treinamento e ativação da rede neural. O gráfico esquemático da solução é apresentado na Figura 4.1.



**Figura 4.1 – Gráfico esquemático da solução do Problema Inverso**

Primeiramente, são gerados os conjuntos de dados a serem utilizados na fase treinamento das redes. Em seguida a rede é treinada, quando os pesos e limiares da rede são ajustados. A ativação consiste em testar as inversões com dados usados e não usados no treinamento.

## 4.1 Geração dos Conjuntos de Dados

Os conjuntos de dados são formados por pares entrada/saída, sendo a entrada gerada pela solução do problema direto e a saída desejada correspondente a cada problema. Na solução do problema direto na recuperação da condição inicial, utiliza-se o método espectral e por superposição linear (Muniz et al, 1999). Enquanto na recuperação da condição de contorno, o problema direto é resolvido numericamente pelo método ATCE – Avançado no Tempo e Centrado no Espaço (Hoffman, 1993).

### 4.1.1 Problema Direto, Recuperação da Condição Inicial

O problema direto da equação de difusão é matematicamente apresentado pelas equações (1.1), (1.2), (1.3) e (1.4). Dada uma distribuição de temperatura inicial  $f(x)$ ,  $0 \leq x \leq 1$ , a solução do problema direto  $T(x,t)$ , para  $0 \leq x \leq 1$  e  $t > 0$ , é formalmente obtida utilizando-se o método espectral e por superposição linear (Muniz et al., 1999).

A solução do problema direto para  $x \in (0,1)$  e  $t > 0$  e para uma dada condição inicial  $f(x)$  é dada por:

$$T(x,t) = \sum_{m=0}^{\infty} e^{-\beta_m^2 t} \frac{1}{N(\beta_m)} X(\beta_m, x) \int_0^1 X(\beta_m, x') f(x') dx' \quad (4.1)$$

onde  $x(\beta_m, x) = \cos(\beta_m, x)$  são as autofunções associadas ao problema,  $\beta_m = m\pi$  são os autovalores e  $N(\beta_m)$  é a normal ou integral de normalização.

Um exemplo da solução do problema direto é apresentado na Figura 4.2. O gráfico apresenta a variação da temperatura, em diferentes instantes de tempo.

#### **Figura 4.2 – Variação da temperatura**

É observado que na medida em que há evolução no tempo, menos informações existem no perfil, dificultando a solução do problema inverso.

#### **4.1.2 Problema Direto, Recuperação da Condição de Contorno**

A equação de difusão dada pelas equações (1.1), (1.2), (1.3) e (1.4), é resolvida numericamente pelo método FTCS (Forward-Time Centered-Space), dado pela equação 4.6 (Hoffman, 1993).

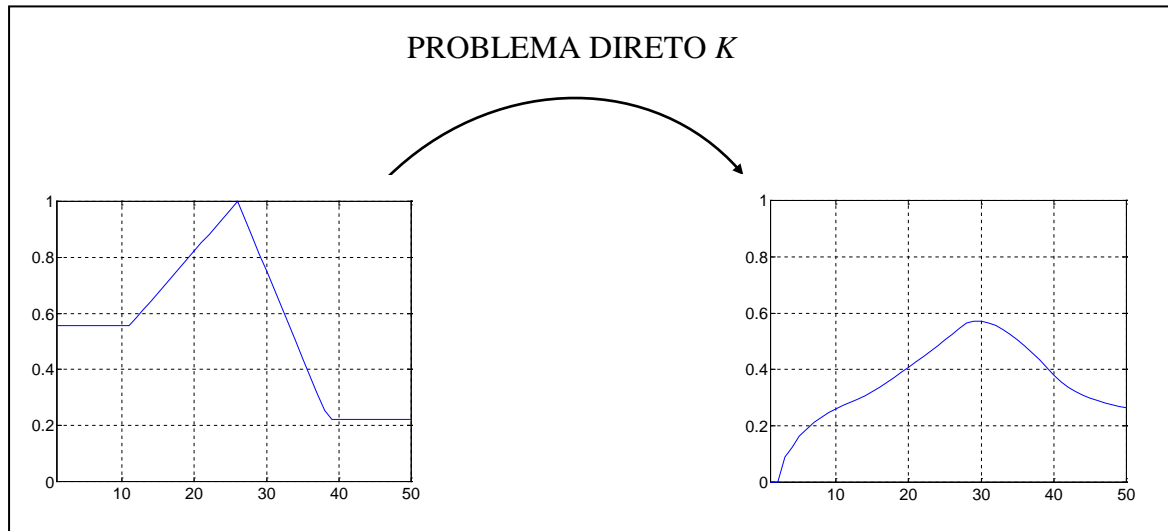
$$f_i^{n+1} = f_i^n + d(f_{i+1}^n - 2f_i^n + f_{i-1}^n) \quad (4.6)$$

onde o número de difusão é dado por:

$$d = \frac{\alpha \Delta t}{\Delta x^2} \quad (4.7)$$

são conhecidas: a condição de contorno em  $x=1$  e a condição inicial  $f(x)$ . O fluxo de calor  $q(t)$  em  $x=0$  é o que se deseja recuperar.

A Figura 4.3 apresenta um exemplo do perfil de temperatura gerado pelo problema direto.



**Figura 4.3 – Resolução do Problema Direto.**

## 4.2 Treinamento da Rede Neural

Além do estudo de diferentes RNA, estuda-se também diferentes estratégias nos algoritmos de aprendizado. Três são as redes utilizadas nas inversões:

- Rede Perceptron de Camadas Múltiplas,
- Rede Funções de Base Radial e

- Rede com algoritmo de Correlação em Cascata.

Cada rede tem características próprias e possui diferentes parâmetros que podem ser variados no treinamento. Por exemplo, na rede Perceptron de Camadas Múltiplas tem-se:

- o número de camadas escondidas,
- o número de neurônios em cada camada,
- taxa de aprendizagem e
- tipos de funções de transferência.

Na rede Funções de Base Radial tem-se:

- o número de neurônios na única camada escondida,
- variação da função de ativação na última camada,
- taxa de aprendizagem e
- diferentes métodos para ajustar as funções de base radial existentes entre a camada de entrada e a camada escondida.

Na rede com algoritmo de correlação em cascata há a adição automática dos neurônios escondidos, portanto pode-se variar:

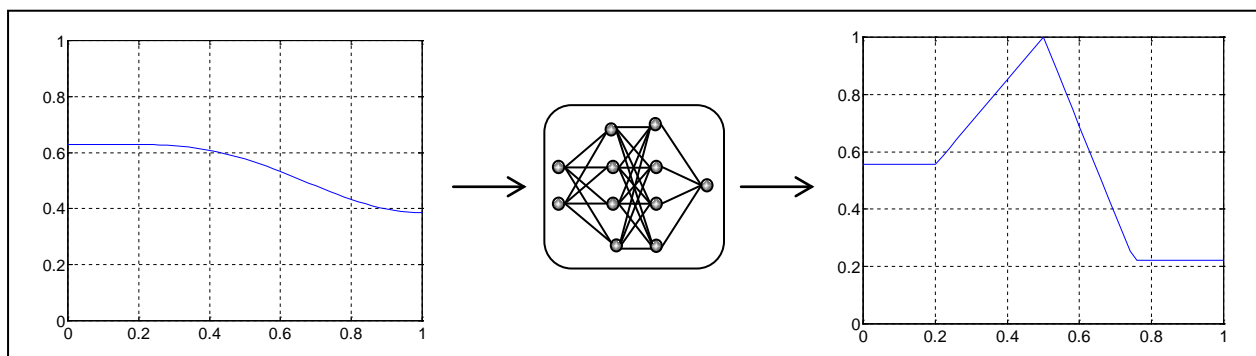
- o número máximo de neurônio a ser adicionado,
- a taxa de aprendizagem da segunda camada e
- as funções de ativação.

### **4.3 Ativação da Rede Neural**

Depois de treinadas, as redes são ativadas para a inversão (Figura 4.4). Cada rede recebe como entrada os dados medidos ou observados, aqui simulados através da



solução do problema direto, com os pesos e valores do limiar obtidos na fase de treinamento geram uma saída. Testa-se a solução do problema inverso utilizando perfis de temperatura usados e também dos não usados no treinamento.



**Figura 4.4 – Resolução do Problema Inverso.**

Estuda-se também a solução dos problemas inversos com dados ruidosos. Adicionando-se 5% de ruído gaussiano nos conjuntos de treinamento e testando-se a generalização das redes.

#### **4.4 Critério de Parada**

Segundo Braga et al. (1998), existem vários métodos para a determinação do momento onde o treinamento deve ser encerrado. Estes métodos são chamados de critérios de parada. Os critérios de parada mais utilizados são:

1. encerrar o treinamento após N ciclos;
2. encerrar o treinamento após o erro quadrático médio ficar abaixo de uma constante  $\alpha$  ;
4. encerrar o treinamento quando a porcentagem de classificações corretas estiver acima de uma constante  $\alpha$  (mais indicado para saídas binárias);
5. combinação dos métodos acima.

Os resultados obtidos no treinamento, apresentados no capítulo 4, foram obtidos utilizando como condição de parada o critério 1 ou o critério 2.

## **4.5 Implementação**

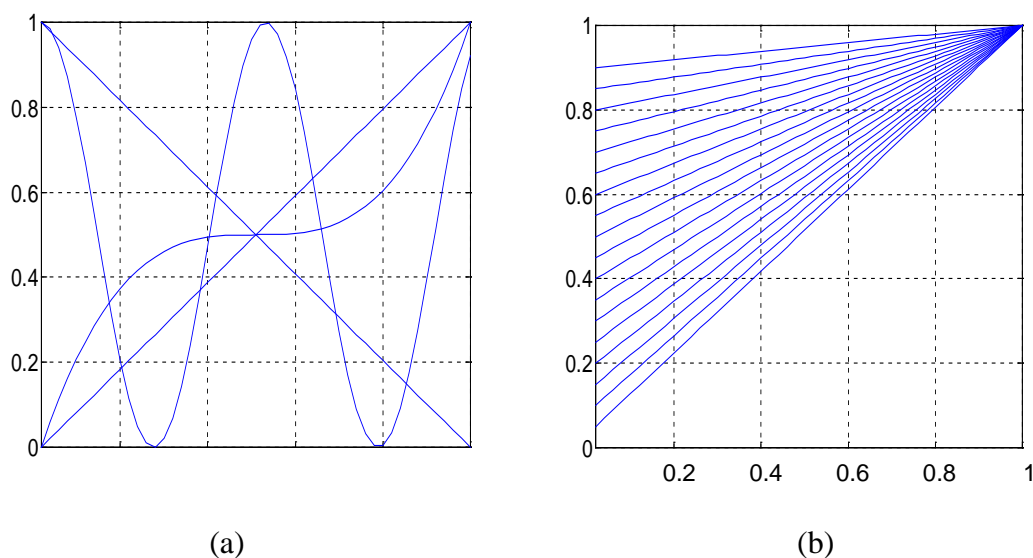
As RNA perceptron de camadas múltiplas, funções de base radial e rede com algoritmo de correlação em cascata foram implementadas em ambiente MatLab© , bem como a geração dos conjuntos de treinamento sem ruído e com 5% de ruído gaussiano.

## CAPÍTULO 5

### RESULTADOS

Neste capítulo são apresentados os resultados obtidos na solução dos dois problemas inversos em condução de calor propostos nesta dissertação resolvidos com o uso das técnicas de redes neurais artificiais. São apresentados os resultados da fase de treinamento, da fase de ativação e e testes de generalização.

Como mencionado na seção 4.1, para cada problema inverso dois conjuntos de funções são utilizados para gerar os conjuntos de dados, cada um contendo 500 funções diferentes. O primeiro é constituído por funções não similares, onde uma função é totalmente independente da outra (Figura 5.1 - a). Enquanto o segundo é constituído por famílias de funções. Por exemplo, uma família pode ser constituída de equações de reta com coeficientes angulares diferentes (Figura 5.1 - b).



**Figura 5.1 – (a) Conjunto de funções não similares, (b) Conjunto de funções similares.**

Este capítulo está dividido em duas partes, a primeira seção (5.1) trata o problema inverso de condução de calor na recuperação da condição inicial, e a segunda (5.2) trata o problema inverso de recuperação da condição de contorno. As partes correspondem a dois problemas inversos diferentes.

## 5.1 Recuperação da Condição Inicial

Como foram discutidos anteriormente, os conjuntos de treinamento são formados por pares entrada/saída, sendo utilizado um tempo  $t$  igual a 0.01 para a geração dos conjuntos.

São realizados testes com dados com 5% de ruído gaussiano e dados com sem ruído, com funções contidas e não contidas no conjunto de treinamento, resultados são apresentados em Shiguemori et al. (2002b).

### 5.1.1 Treinamento da rede Perceptron de Camadas Múltiplas

No treinamento da RPCM utilizou-se uma única camada escondida e funções de ativação sigmóide tanto na camada escondida quanto na camada de saída. Na tabela 5.1 são apresentados os resultados do treinamento da rede RPCM. A primeira coluna consistindo no tipo de conjunto de treinamento, a segunda na porcentagem de ruído adicionada, a terceira no número de épocas de treinamento e a última a média do erro médio quadrático (MEMQ), dada pela equação:

$$\varepsilon_{\text{med}} = \frac{1}{N} \sum_{n=1}^N \varepsilon(n) \quad (5.1)$$

onde  $N$  representa o número total de padrões contidos no conjunto de treinamento,  $n$  é o  $n$ -ésimo exemplo de treinamento e  $\varepsilon(n)$  é a energia instantânea do erro dada pela equação:

$$\varepsilon = \frac{1}{2} \sum_{k \in C} e_k^2(n) \quad (5.2)$$

sendo  $e_k(n)$  o sinal de erro na saída do neurônio  $k$  obtido pela equação 3.2 e  $C$  o conjunto de todos os neurônios da camada de saída da rede.

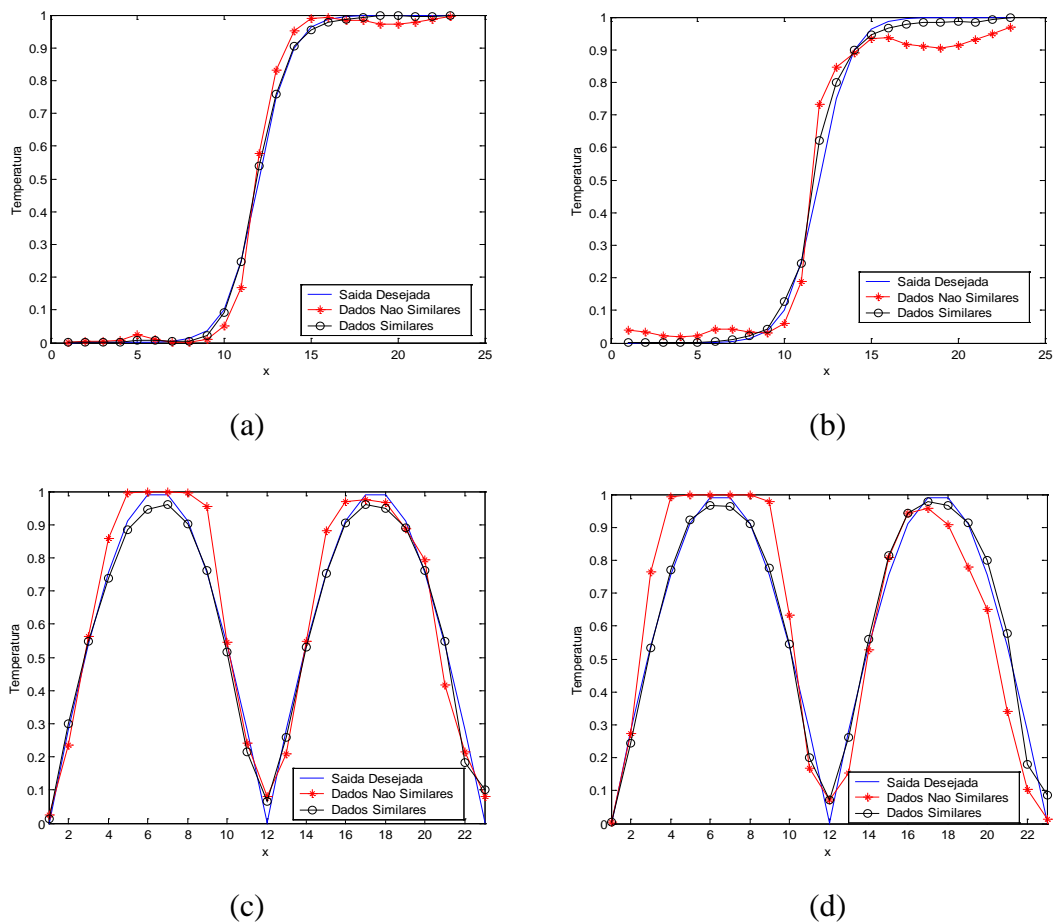
**TABELA 5.1 – RESULTADOS DOS TREINAMENTOS DA RPCM**

Conjunto	Ruído	Número de Neurônios	Épocas de treinamento	MEMQ
Não-similar	0 %	25	150.000	0,0487
Similar	0 %	20	50.000	0,0127
Não-similar	5 %	20	300.000	0,0694
Similar	5 %	20	165.000	0,0144

O treinamento do conjunto de dados não-similares requer um número maior de épocas de treinamento, quando comparado ao treinamento dos dados similares. Também há uma dificuldade maior da convergência do erro quando adicionado ruído nos dados observados.

### 5.1.2 Ativação da rede Perceptron de Camadas Múltiplas

Nesta seção são apresentadas algumas condições iniciais recuperadas, com dados presentes nos conjuntos de treinamento (Figura 5.2). A linha cheia-estrela representa o resultado obtido com a rede treinada com conjunto de funções similares. Enquanto a linha cheia-círculo representa a inversão obtida com o a rede treinada com o conjunto similar. As Figuras 5.2-a e 5.2-c são obtidas com dados sem ruído, enquanto as Figura 5.2-b e 5.2-d com dados ruidosos.



**Figura 5.2 – Exemplos de inversões obtidas com a RPCM.**

**(a) (c) - sem ruído e (b) (d) - com 5% de ruído.**

Como pode ser observado a RPCM recupera bem a condição inicial mesmo sendo adicionado 5% de ruído gaussiano na entrada. A tabela 5.2 apresenta os valores dos erros obtidos pelos exemplos de inversões apresentados na Figura 5.2.

**TABELA 5.2 – RESULTADOS DE ATIVAÇÕES DA RPCM**

<b>Figura</b>	<b>Conjunto</b>	<b>Sem Ruído</b>	<b>EMQ</b>
5.2 (a)	Não-similar	0 %	0.0151
5.2 (a)	Similar	0 %	0.0012
5.2 (b)	Não-similar	5 %	0.0591
5.2 (b)	Similar	5 %	0.0100
5.2 (c)	Não-similar	0 %	0.0651

(continua)

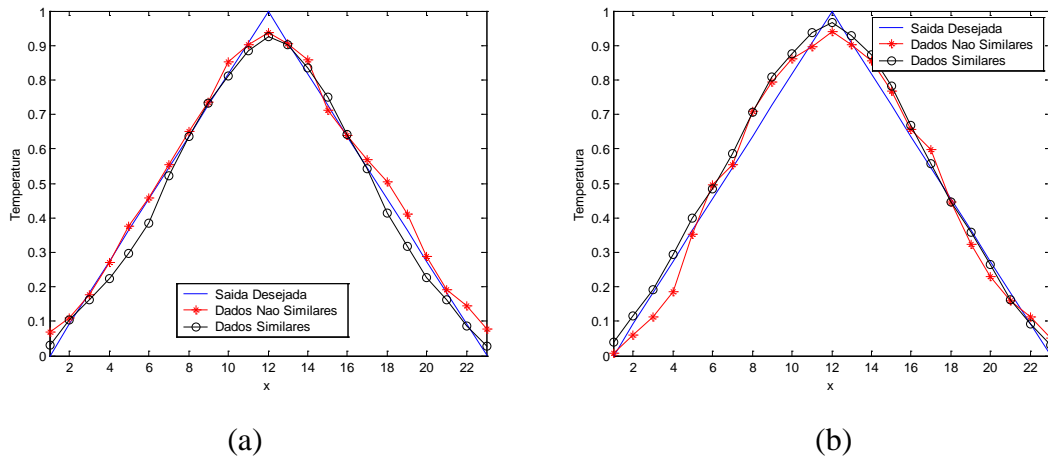
**TABELA 5.2 – Conclusão**

5.2 (c)	Similar	0 %	0.0180
5.2 (d)	Não-similar	5 %	0.1641
5.2 (d)	Similar	5 %	0.0212

As inversões aproximam-se da saída desejada quando a rede é treinada com dados similares. E erros menores são obtidos na ausência de ruído.

### 5.1.3 Testes de Generalização da RPCM

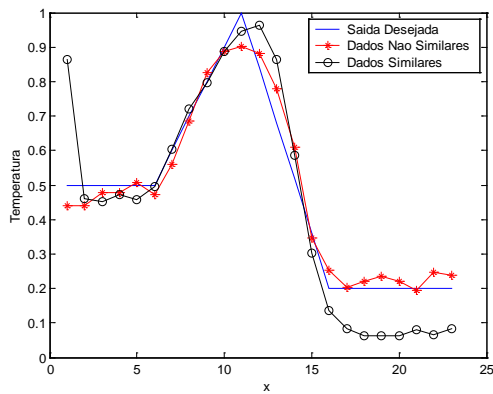
Nos testes de generalização, testes com dados não contidos no conjunto de treinamento, são utilizadas duas funções: triangular e semitriangular. As inversões apresentadas nas Figuras 5.3-a e Figura 5.3-c são obtidas com dados sem ruído, enquanto nas Figuras 5.3-b e 5.3-d as inversões são obtidas com dados com 5% de ruído.



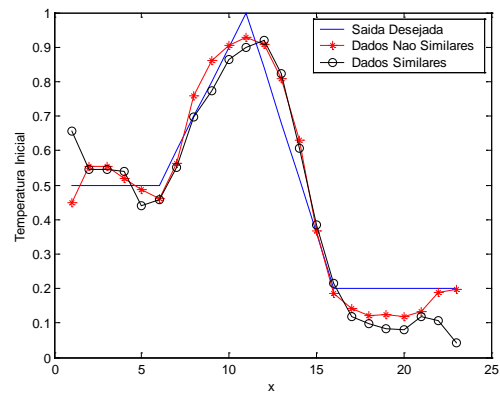
**Figura 5.3 – Exemplos dos testes de generalização obtidos com a RPCM.**

**(a) (c) - sem ruído e (b) (d) - com 5% de ruído.**

(continua)



(c)



(d)

**Figura 5.3 – conclusão**

As condições iniciais recuperadas pela RPCM treinadas com o conjunto não-similar sistematicamente apresentaram erros menores na ativação, embora na fase de treinamento erros menores tenham sido alcançados quando treinadas com dados similares, isso porque a quantidade de informação contida nos pesos sinápticos e nos limiares é maior no conjunto não-similar. Uma análise melhor pode ser realizada através dos valores apresentados na tabela 5.3.

**TABELA 5.3 – RESULTADOS DE GENERALIZAÇÕES DA RPCM**

<b>Figura</b>	<b>Conjunto</b>	<b>Sem Ruído</b>	<b>MEMQ</b>
5.3 (a)	Não-similar	0 %	0.0136
5.3 (a)	Similar	0 %	0.0139
5.3 (b)	Não-similar	5 %	0.0222
5.3 (b)	Similar	5 %	0.0159
5.3 (c)	Não-similar	0 %	0.0246
5.3 (c)	Similar	0 %	0.1599
5.3 (d)	Não-similar	5 %	0.0621
5.3 (d)	Similar	5 %	0.0786



### 5.1.4 Treinamento da rede de Funções de Base Radial

A rede de Funções de Base Radial também foi treinada com os dois conjuntos de treinamento, com dados sem e com 5% de ruído. Os resultados dos treinamentos são apresentados na tabela 5.4.

A rede foi treinada com a função de ativação gaussiana (Equação 5.3).

$$\varphi(v) = \exp\left(-\frac{v^2}{2\sigma^2}\right) \quad (5.3)$$

com  $\sigma > 0$  e  $v \in \mathfrak{R}$ .

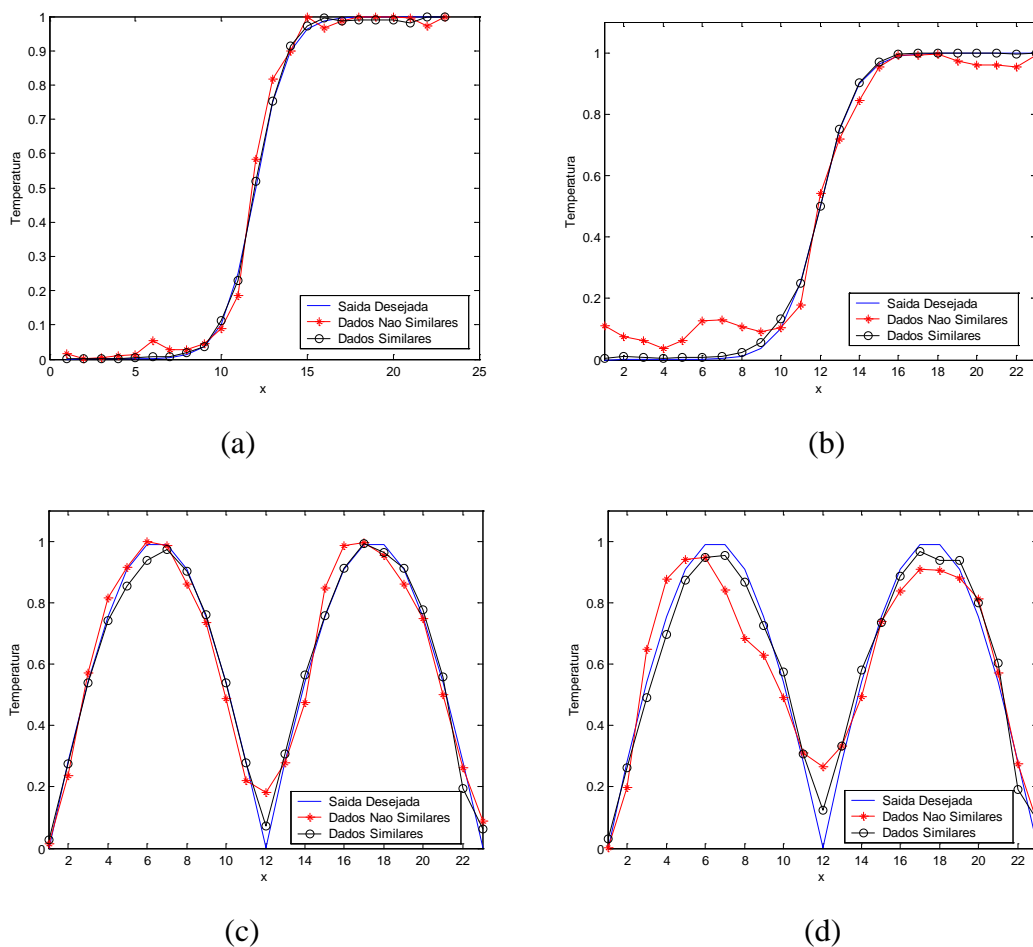
**TABELA 5.4 – RESULTADOS DOS TREINAMENTOS DA RFBR**

Conjunto	Ruído	Número de Neurônios	Épocas de treinamento	MEMQ
Não-similar	0 %	20	50.000	0,0576
Similar	0 %	20	50.000	0,0095
Não-similar	5 %	20	300.000	0,0873
Similar	5 %	20	30.000	0,0123

É observado que também no treinamento da RFBR, há uma dificuldade maior em se treinar o conjunto não similar e os dados com ruído. Embora outros testes tenham sido realizados variando-se o número de neurônios na camada escondida, são apresentados os resultados obtidos com o uso de 20 neurônios.

### 5.1.5 Ativação da rede Funções de Base Radial

As condições iniciais recuperadas com o uso da rede Funções de Base Radial são apresentadas na Figura 5.4. As mesmas funções testadas na RPCM também são utilizadas. As Figuras 5.4-a e 5.4-c com dados sem ruído e as Figuras 5.4-b e 5.4-d, com 5% de ruído.



**Figura 5.4 – Exemplos das inversões obtidas com a RFBR.**  
**(a) (c) - sem ruído e (b) (d) - com 5% de ruído.**

Na Tabela 5.5 são apresentados os valores dos erros obtidos com as inversões.

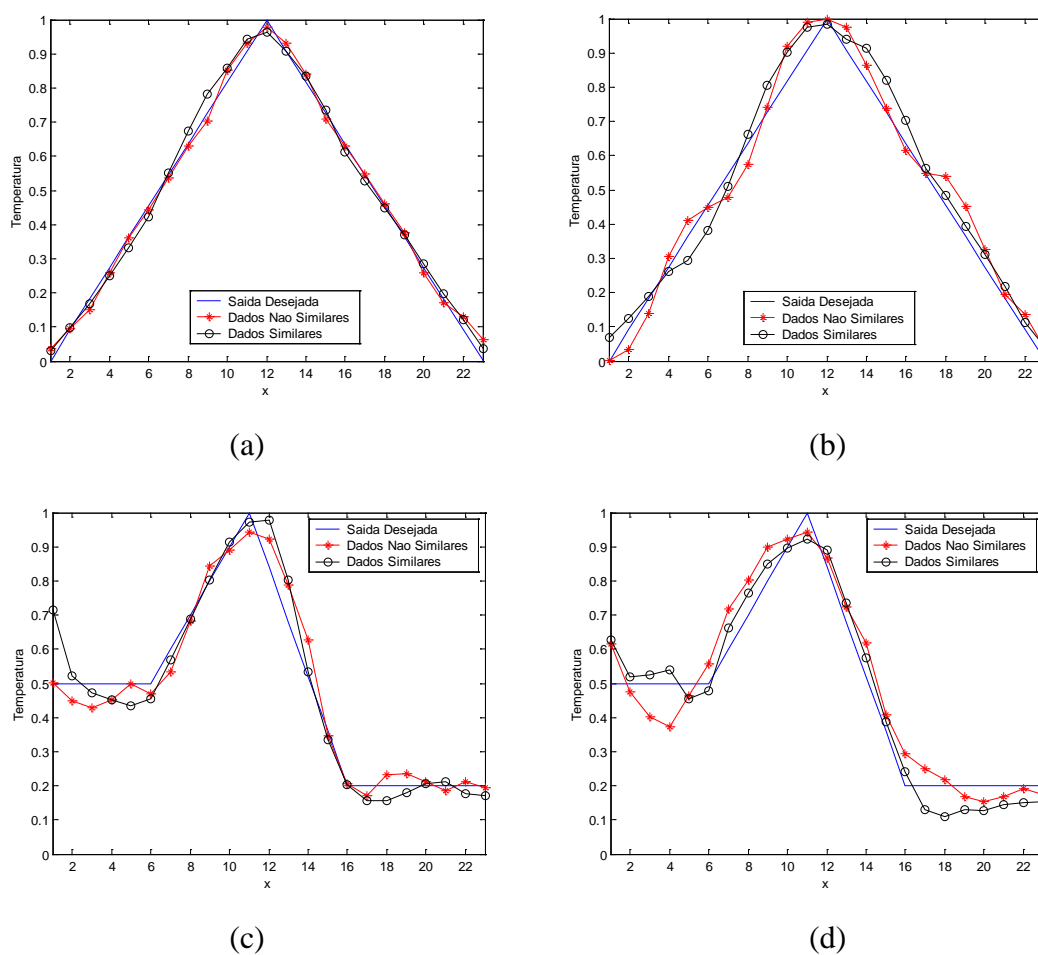
**TABELA 5.5 – RESULTADOS DE ATIVAÇÕES DA RFBR**

<b>Figura</b>	<b>Conjunto</b>	<b>Ruído</b>	<b>MEMQ</b>
5.4 (a)	Não-similar	0 %	0,0116
5.4 (a)	Similar	0 %	0,0011
5.4 (b)	Não-similar	5 %	0,0343
5.4 (b)	Similar	5 %	0,0066
5.4 (c)	Não-similar	0 %	0,0408
5.4 (c)	Similar	0 %	0,0135
5.4 (d)	Não-similar	5 %	0,1161
5.4 (d)	Similar	5 %	0,0311

Na ativação da RFBR valores menores dos erros foram obtidos quando a rede foi treinada com dados similares.

### 5.1.6 Testes de Generalização da RFBR

Os gráficos apresentados na Figura 5.5 apresentam os testes de generalização da RFBR, para estes testes também foram utilizadas as funções triangular e semi-triangular.



**Figura 5.5 – Exemplos de generalizações obtidas com a RFBR.**

(a) (c) - sem ruído e (b) (d) - com 5% de ruído.

**TABELA 5.6 – RESULTADOS DE GENERALIZAÇÕES DA RFBR**

<b>Figura</b>	<b>Conjunto</b>	<b>Ruído</b>	<b>MEMQ</b>
5.5 (a)	Não-similar	0 %	0,0065
5.5 (a)	Similar	0 %	0,0079
5.5 (b)	Não-similar	5 %	0,0308
5.5 (b)	Similar	5 %	0,0331
5.5 (c)	Não-similar	0 %	0,0275
5.5 (c)	Similar	0 %	0,0498
5.5 (d)	Não-similar	5 %	0,0563
5.5 (d)	Similar	5 %	0,0396

Tal qual nos testes de generalização da RPCM, a rede Funções de Base Radial sistematicamente teve um desempenho melhor, para a rede treinada com dados não similares. Com a presença ou ausência de ruído nos dados.

### **5.1.7 Treinamento da Rede com Algoritmo de Correlação em Cascata**

A terceira rede neural utilizada na inversão é a rede com algoritmo de Correlação em Cascata. Os resultados de seu treinamento são apresentados na tabela 5.7. Também foram utilizados dados com e sem ruído.

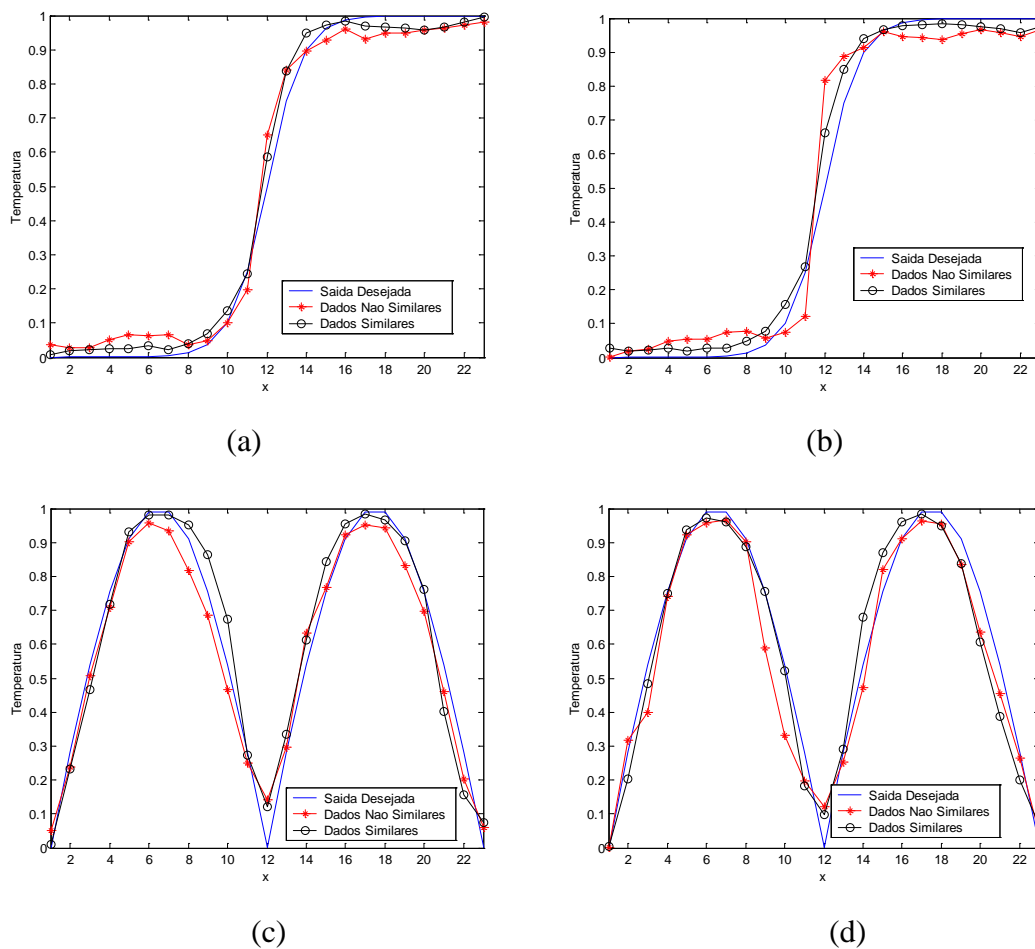
**TABELA 5.7 – RESULTADOS DOS TREINAMENTOS DA RCC**

<b>Conjunto</b>	<b>Ruído</b>	<b>Número de Neurônios</b>	<b>MEMQ</b>
Não-similar	0 %	10	0,0746
Similar	0 %	5	0,0230
Não-similar	5 %	2	0,1389
Similar	5 %	5	0,0318

Tal qual no treinamento das outras duas redes neurais, também foram obtidos valores menores do MEMQ no treinamento do conjunto de dados não-similares. A diferença está no algoritmo de treinamento, pois a RCC tem como característica determinar seu próprio número de neurônios na camada escondida, pode ser observado na tabela 5.7 que a RCC utilizou poucos neurônios na camada escondida.

### 5.1.8 Ativação da rede com algoritmo de correlação em cascata

As condições iniciais apresentadas na Figura 5.6 são obtidas pela RCC. Na tabela 5.8 apresenta-se os valores dos erros.



**Figura 5.6 – Exemplos de inversões obtidas com a RCC.**

**(a) (c) - sem ruído e (b) (d) - com 5% de ruído.**

**TABELA 5.8 – RESULTADOS DE ATIVAÇÕES DA RCC**

<b>Figura</b>	<b>Conjunto</b>	<b>Sem Ruído</b>	<b>MEMQ</b>
5.5 (a)	Não-similar	0 %	0,0336
5.5 (a)	Similar	0 %	0,0152
5.5 (b)	Não-similar	5 %	0,0875

(continua)

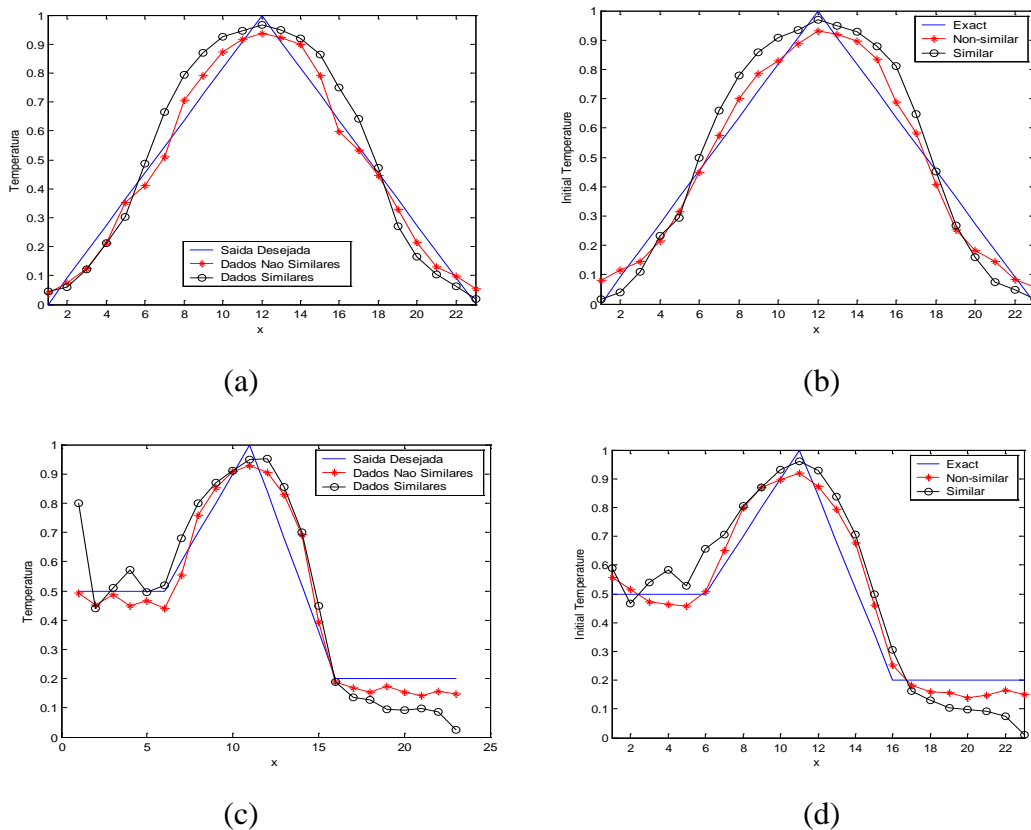
**TABELA 5.8 – Conclusão**

5.5 (b)	Similar	5 %	0,0266
5.5 (c)	Não-similar	0 %	0,0452
5.5 (c)	Similar	0 %	0,0571
5.5 (d)	Não-similar	5 %	0,0783
5.5 (d)	Similar	5 %	0,0652

Na ativação dos dados presentes nos conjuntos de treinamento, na maioria dos casos houve uma melhor inversão quando a rede foi treinada com dados similares.

### 5.1.9 Testes de Generalização da RCC

Os testes de generalização da RCC são apresentados na Figura 5.7. As figura 5.7-a e 5.7-c são obtidas com dados sem ruído. Enquanto nas Figuras 5.7-b e 5.7-d são obtidas com dados ruidosos. Os valores dos erros correspondentes estão na tabela 5.9.



**Figura 5.7 – Exemplos de generalizações obtidos com a RCC.**

**(a) (c) - sem ruído e (b) (d) - com 5% de ruído.**

**TABELA 5.9 – RESULTADOS DE GENERALIZAÇÕES DA RCC**

<b>Figura</b>	<b>Conjunto</b>	<b>Ruído</b>	<b>MEMQ</b>
5.5 (a)	Não-similar	0 %	0,0253
5.5 (a)	Similar	0 %	0,0845
5.5 (b)	Não-similar	5 %	0,0592
5.5 (b)	Similar	5 %	0,1047
5.5 (c)	Não-similar	0 %	0,0471
5.5 (c)	Similar	0 %	0,1462
5.5 (d)	Não-similar	5 %	0,2109
5.5 (d)	Similar	5 %	0,1145

## **5.2 Recuperação da Condição de Contorno**

O uso das redes neurais artificiais na solução do problema inverso na recuperação da condição de contorno envolve o mesmo esquema apresentado na Figura 4.1. Os parâmetros: condições de contorno em  $x=L$  e condição inicial são conhecidos para a solução do problema direto, alguns resultados são apresentados em Shiguemori (2001a) e Shiguemori (2001b).

Nos conjuntos de treinamento, as entradas são geradas utilizando como parâmetros  $\Delta t = 0.5$ ,  $\alpha = 0.0005$  e localização do sensor em  $x = 0.075$ .

Com o objetivo de adequar cada rede ao problema inverso de condução de calor, foram realizados também alguns experimentos variando o número de neurônios na camada escondida e a taxa de aprendizagem. No entanto, são mostrados somente os melhores resultados.

### **5.2.1 Treinamento da Rede Perceptron de Camadas Múltiplas**

Assim como na implementação da RPCM para a recuperação da condição inicial, na recuperação da condição de contorno, a rede foi treinada com uma única camada escondida e funções de ativação sigmóide nas camadas escondidas e na de

saída. Os resultados obtidos nessa fase são apresentados na tabela 5.10. Foram utilizados dado sem ruído e com 5% de ruído gaussiano.

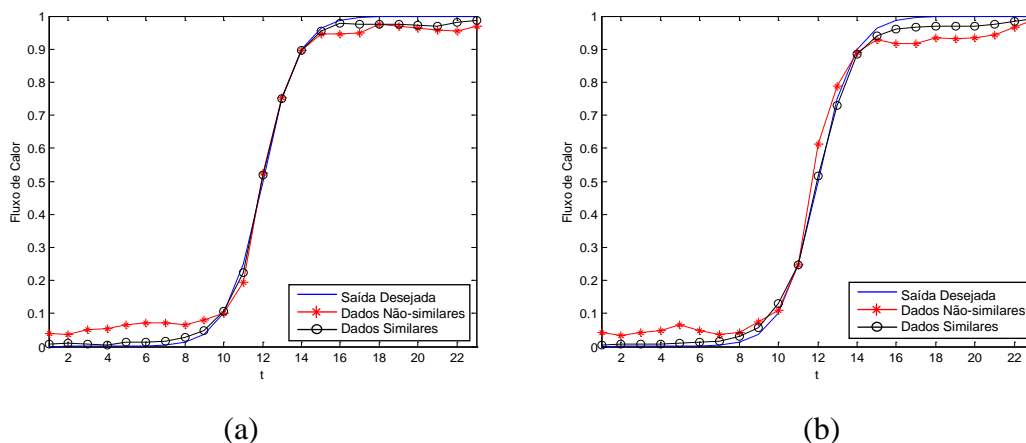
**TABELA 5.10 – RESULTADOS DOS TREINAMENTOS DA RPCM**

Conjunto	Ruído	Número de Neurônios	Épocas de treinamento	MEMQ
Não-similar	0 %	20	10.000	0,0371
Similar	0 %	20	10.000	0,0147
Não-similar	5 %	20	10.000	0,0538
Similar	5 %	20	10.000	0,0205

Tal qual no treinamento da RPCM na solução do primeiro problema proposto nessa dissertação, foram obtidos erros menores no treinamento dos dados similares e sem ruído.

### 5.2.2 Ativação da Rede Perceptron de Camadas múltiplas

Na Figura 5.8 são apresentadas condições de contorno recuperadas pela RPCM com dados presentes no conjunto de treinamento. Na tabela 5.11 são apresentados os valores dos erros correspondentes. São obtidos erros menores quando a rede é treinada com dados similares.

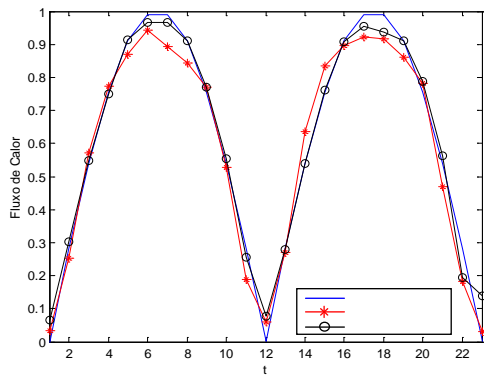


**Figura 5.8 – Exemplos de generalizações obtidas com a RPCM.**

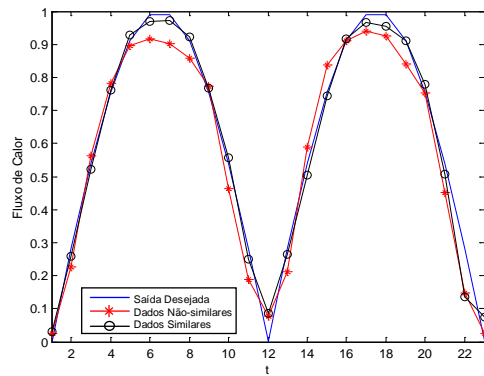
(a) (c) - sem ruído e (b) (d) - com 5% de ruído.

(continua)





(c)



(d)

**Figura 5.8 – Conclusão.****TABELA 5.11 – RESULTADOS DE ATIVAÇÕES DA RPCM**

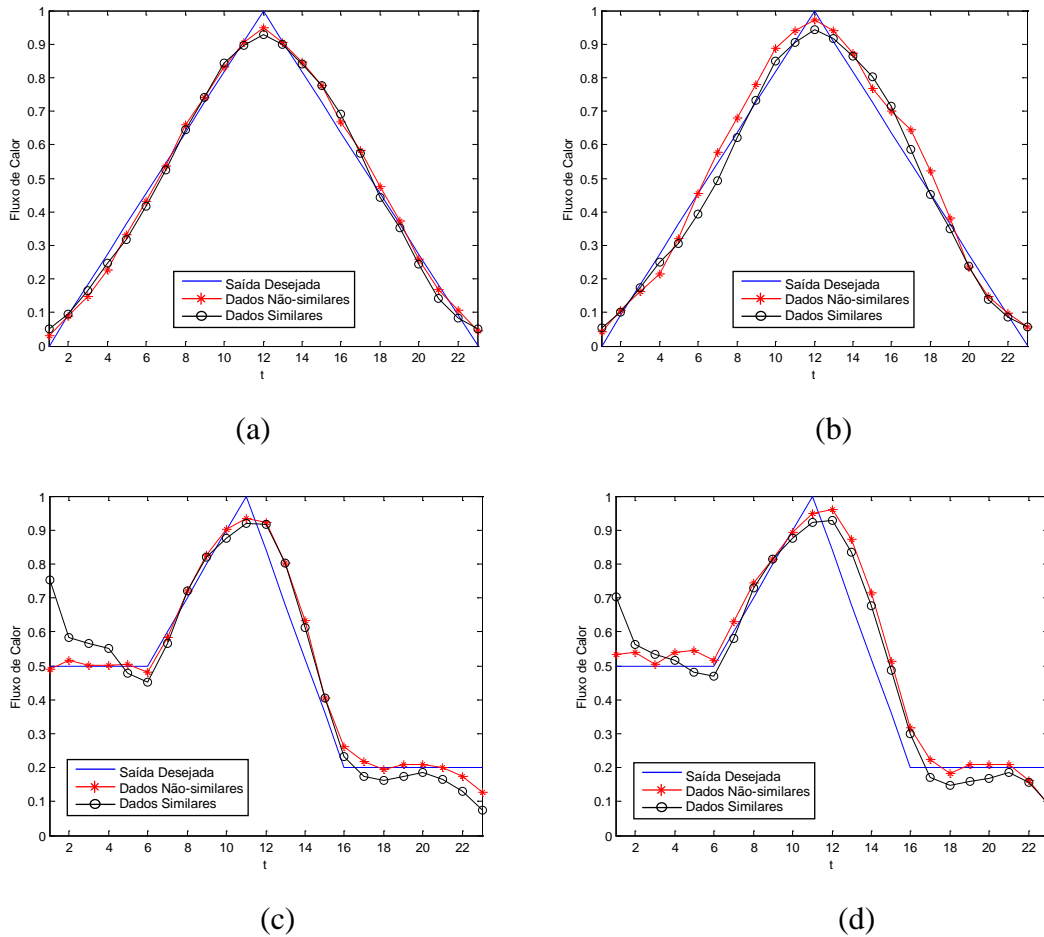
<b>Figura</b>	<b>Conjunto</b>	<b>Ruído</b>	<b>MEMQ</b>
5.5 (a)	Não-similar	0 %	0,0213
5.5 (a)	Similar	0 %	0,0032
5.5 (b)	Não-similar	5 %	0,0307
5.5 (b)	Similar	5 %	0,0044
5.5 (c)	Não-similar	0 %	0,0387
5.5 (c)	Similar	0 %	0,0228
5.5 (d)	Não-similar	5 %	0,0467
5.5 (d)	Similar	5 %	0,0217

Claramente observa-se que a ativação das duas funções presentes no conjunto de treinamento é melhor quando a rede é treinada com dados similares, mesmo com a presença de ruído gaussiano. Isso porque na fase de treinamento foram obtidos valores de erros menores.

### 5.2.3 Testes de Generalização da RPCM

Os testes de generalização da RPCM na recuperação da condição de contorno são apresentados na Figura 5.9. As funções triangular e semi-triangular novamente são

utilizadas. Os valores dos erros obtidos para essas funções são apresentadas na tabela 5.12.



**Figura 5.9 – Exemplos de inversões obtidas com a RPCM.**

(a) (c) - sem ruído e (b) (d) - com 5% de ruído.

**TABELA 5.12 – RESULTADOS DAS GENERALIZAÇÕES DA RPCM**

Figura	Conjunto	Ruído	MEMQ
5.9 (a)	Não-similar	0 %	0,0093
5.9 (a)	Similar	0 %	0,0133
5.9 (b)	Não-similar	5 %	0,0251
5.9 (b)	Similar	5 %	0,0203
5.9 (c)	Não-similar	0 %	0,0267

(continua)

**TABELA 5.12 – Conclusão**

5.9 (c)	Similar	0 %	0,0228
5.9 (d)	Não-similar	5 %	0,0771
5.9 (d)	Similar	5 %	0,0789

#### 5.2.4 Treinamento da Rede Funções de Base Radial

No treinamento da RFBR para a recuperação da condição de contorno também foram utilizadas funções gaussianas com função de ativação dos neurônios escondidos. Variou-se o número de neurônios na camada escondida e também o número de épocas de treinamento, no entanto somente os melhores resultados são apresentados na tabela 5.13, para dados com e sem ruído.

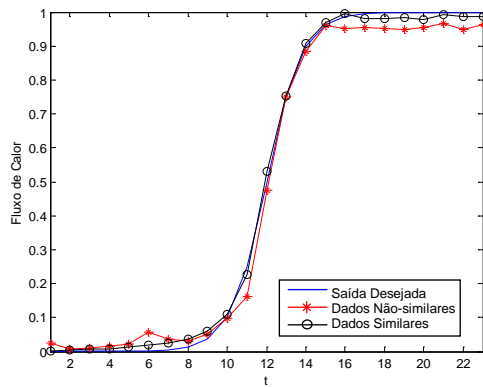
**TABELA 5.13 – RESULTADOS DOS TREINAMENTOS DA RFBR**

Conjunto	Ruído	Número de Neurônios	Épocas de treinamento	MEMQ
Não-similar	0 %	30	10.000	0,0278
Similar	0 %	30	5.348	0,0115
Não-similar	5 %	30	10.000	0,0269
Similar	5 %	30	9.850	0,0115

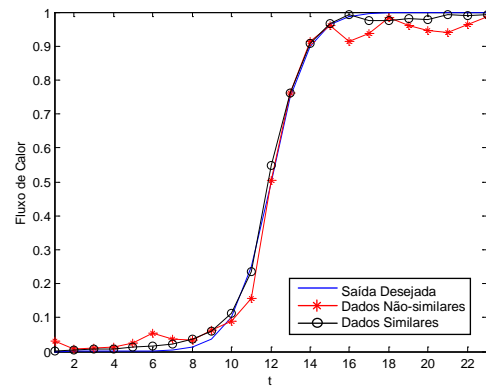
Diferentemente dos treinamentos anteriores, no treinamento da rede RFBR atingiu-se o erro-objetivo (0,0115).

#### 5.2.5 Ativação da Rede Funções de Base Radial

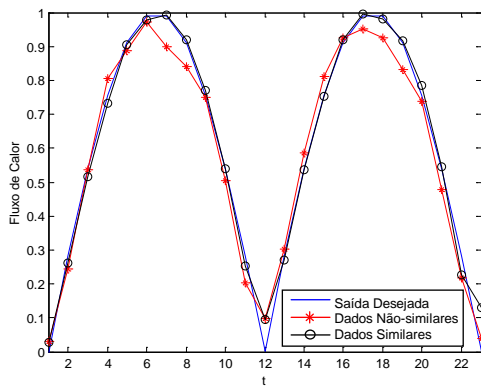
Os resultados da recuperação da condição de contorno obtidos com a RFBR são apresentados (Figura 5.10). E os valores dos erros correspondentes, na tabela 5.14.



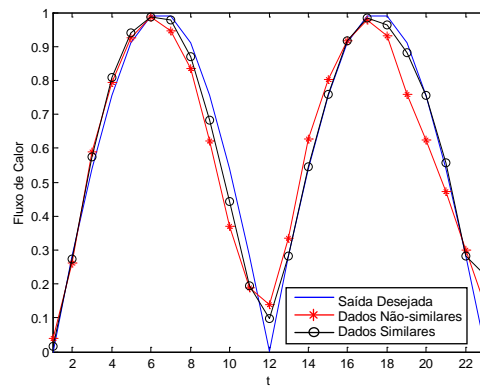
(a)



(b)



(c)



(d)

**Figura 5.10 – Exemplos de ativações obtidas com a RFBR.**

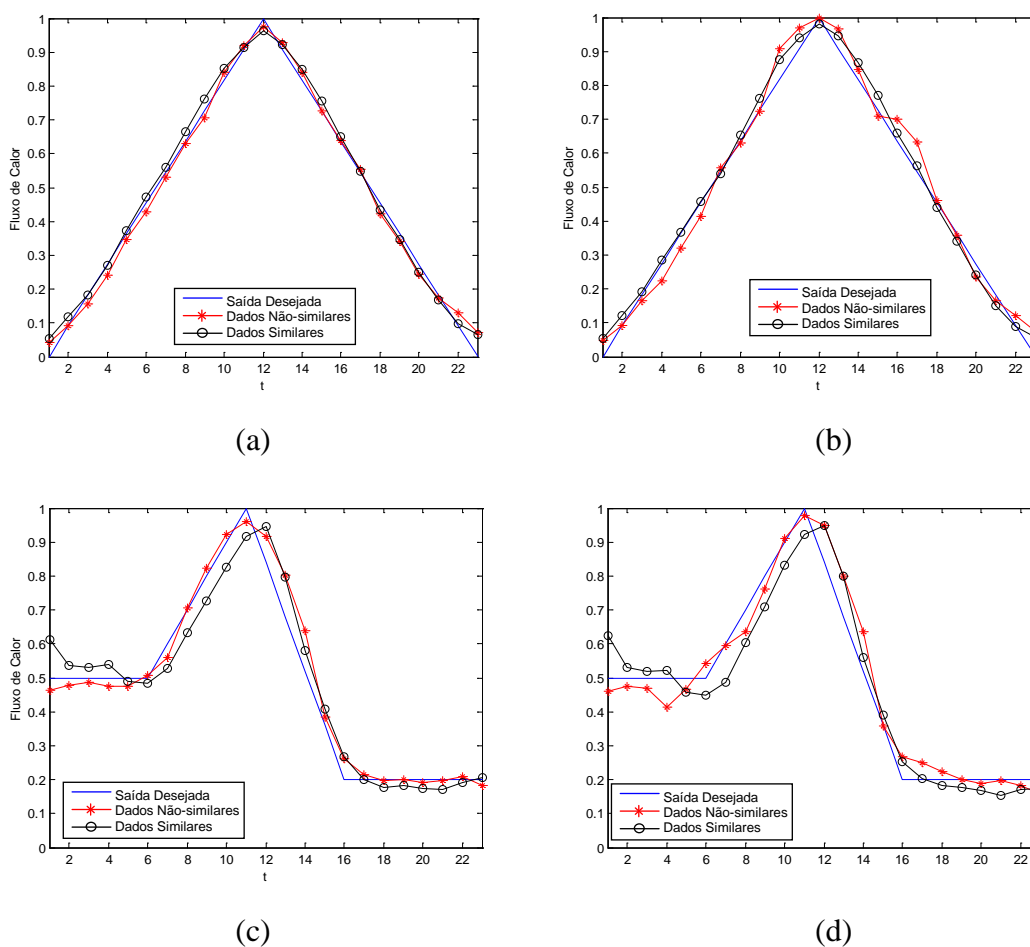
(a) (c) - sem ruído e (b) (d) - com 5% de ruído.

**TABELA 5.14 – RESULTADOS DE ATIVAÇÕES DA RFBR**

<b>Figura</b>	<b>Conjunto</b>	<b>Ruído</b>	<b>MEMQ</b>
5.10 (a)	Não-similar	0 %	0,0314
5.10 (a)	Similar	0 %	0,0168
5.10 (b)	Não-similar	5 %	0,0833
5.10 (b)	Similar	5 %	0,0454
5.10 (c)	Não-similar	0 %	0,0150
5.10 (c)	Similar	0 %	0,0028
5.10 (d)	Não-similar	5 %	0,0171
5.10 (d)	Similar	5 %	0,0035

## 5.2.6 Testes de Generalização da RFBR

São apresentados os testes de generalização da rede RFBR (Figura 5.11) e os valores dos erros (tabela 5.15). Testa-se as funções triangulares e semitriangulares, treinadas com os conjuntos de dados não-similares e similares, além da robustez a dados ruidosos.



**Figura 5.11 – Exemplos de generalizações obtidas com a RFBR.**

**(a) (c) - sem ruído e (b) (d) - com 5% de ruído.**

Pode-se observar que mesmo não estando presente no conjunto de treinamento e mesmo com perturbações de 5% de ruído gaussiano, obteve-se recuperações das condições de contorno bem próximas às desejadas. Existe uma quantidade maior de

informações quando a rede é treinada com funções não-similares, por esse motivo foram obtidos erros menores na generalização para a rede treinada com dados não-similares, como pode ser observado na tabela 5.15.

**TABELA 5.15 – RESULTADOS DAS GENERALIZAÇÕES DA RFBR**

<b>Figura</b>	<b>Conjunto</b>	<b>Ruído</b>	<b>MEMQ</b>
5.11 (a)	Não-similar	0 %	0,0080
5.11 (a)	Similar	0 %	0,0084
5. 11 (b)	Não-similar	5 %	0,0224
5. 11 (b)	Similar	5 %	0,0114
5. 11 (c)	Não-similar	0 %	0,0237
5. 11 (c)	Similar	0 %	0,0412
5. 11 (d)	Não-similar	5 %	0,0343
5. 11 (d)	Similar	5 %	0,0497

Na maioria dos casos obteve-se valores menores dos erros, quando a rede foi treinada com dados não similares.

### **5.2.7 Treinamento da Rede com Algoritmo de Correlação em Cascata**

No treinamento da rede RCC no problema inverso de condução de calor variou-se o número de neurônios máximo a ser adicionado na rede e também o número máximo de épocas de treinamento. São apresentados os valores da MEMQ obtidos pela rede (Tabela 5.16).

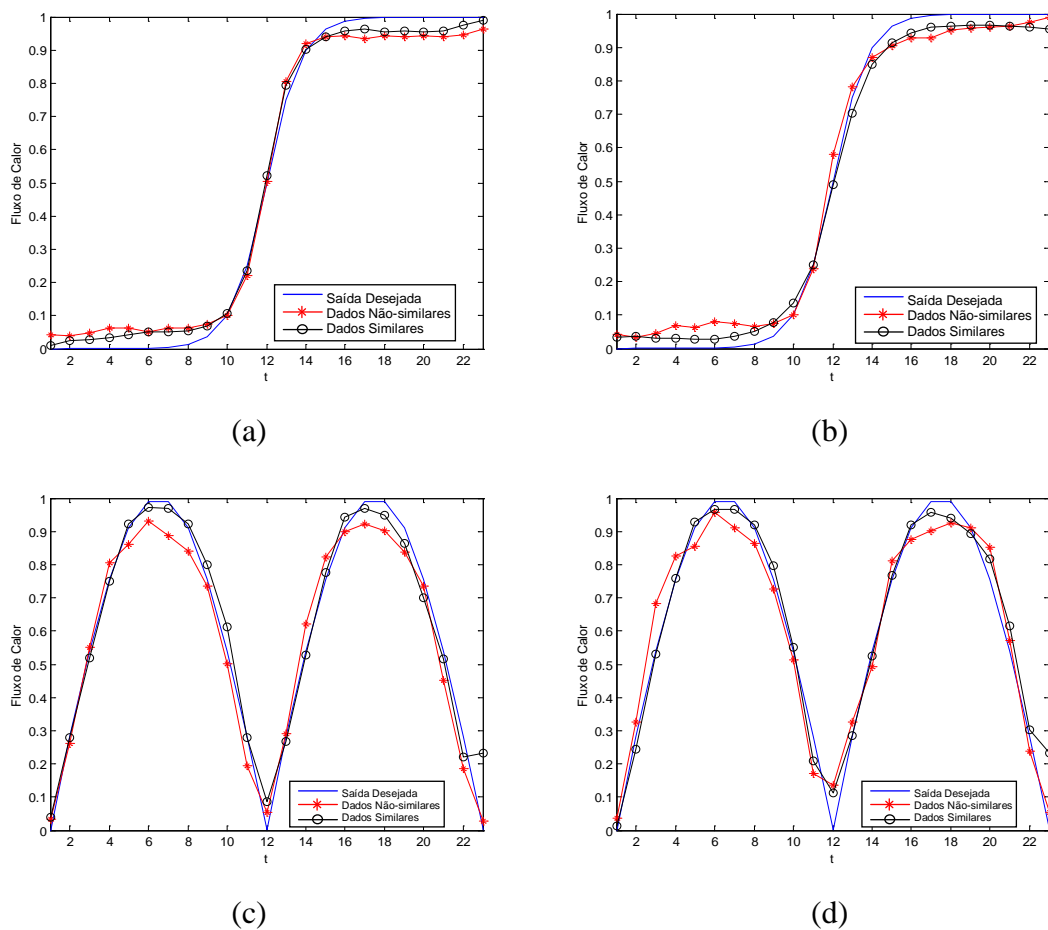
**TABELA 5.16 – RESULTADOS DOS TREINAMENTOS DA RCC**

<b>Conjunto</b>	<b>Ruído</b>	<b>Número de Neurônios</b>	<b>Épocas de treinamento</b>	<b>MEMQ</b>
Não-similar	0 %	14	25.100	0,0313
Similar	0 %	10	10.000	0,0186
Não-similar	5 %	15	19.732	0,0498
Similar	5 %	16	29.859	0,0222

Foram obtidos valores do MEMQ menores no treinamento da rede RCC, para a recuperação da condição de contorno, quando comparada à recuperação da condição inicial.

### 5.2.8 Ativação da Rede Treinada com Algoritmo de Correlação em Cascata

São apresentados exemplos da recuperação da condição de contorno através de dados utilizados no treinamento (Figura 5.12). Além disso, é apresentada uma tabela (5.17) contendo os respectivos erros. São obtidos erros menores quando a rede é treinada com dados similares.



**Figura 5.12 – Exemplos de ativações obtidas com a RCC.**

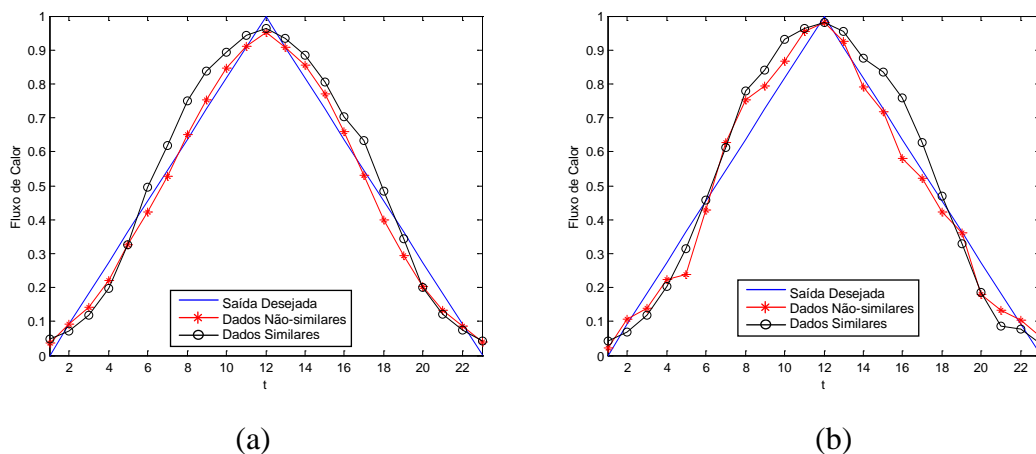
**(a) (c) - sem ruído e (b) (d) - com 5% de ruído.**

**TABELA 5.17 – RESULTADOS DE ATIVAÇÕES DA RCC**

Figura	Conjunto	Ruído	MEMQ
5.12 (a)	Não-similar	0 %	0,0260
5.12 (a)	Similar	0 %	0,0125
5. 12 (b)	Não-similar	5 %	0,0284
5. 12 (b)	Similar	5 %	0,0148
5. 12 (c)	Não-similar	0 %	0,0426
5. 12 (c)	Similar	0 %	0,0427
5. 12 (d)	Não-similar	5 %	0,0547
5. 12 (d)	Similar	5 %	0,0454

### 5.2.9 Generalização da Rede Treinada com Algoritmo de Correlação em Cascata

Nos testes de generalização da RCC também foram utilizadas as funções triangular e semitriangular, com dados sem ruído (Figuras 5.11-a e 5.11-c) e com 5% de ruído gaussiano (Figuras 5.13-a e 5.13-c).

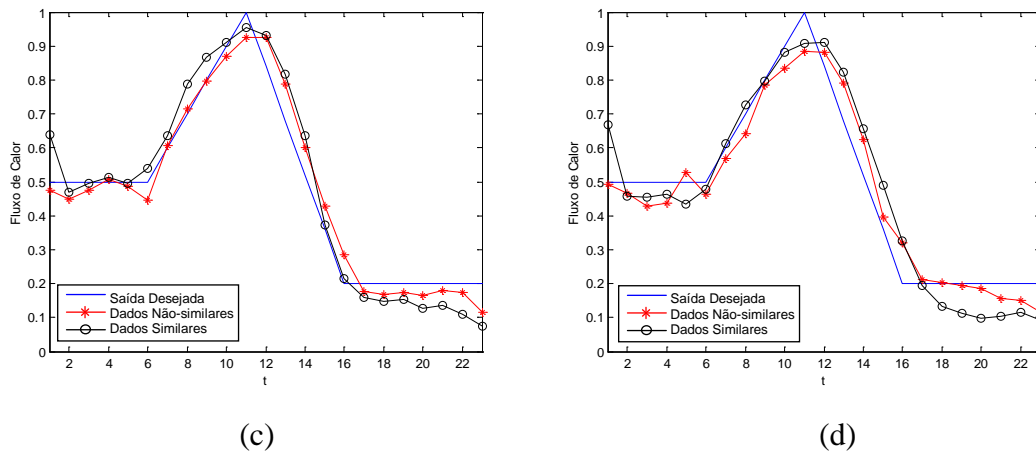


**Figura 5.13 – Exemplos de generalizações obtidas com a RCC.**

**(a) (c) - sem ruído e (b) (d) - com 5% de ruído.**

(continua)





**Figura 5.13 – Conclusão.**

**TABELA 5.18 – RESULTADOS DE GENERALIZAÇÕES DA RCC**

<b>Figura</b>	<b>Conjunto</b>	<b>Sem Ruído</b>	<b>MEMQ</b>
5.12 (a)	Não-similar	0 %	0,0167
5.12 (a)	Similar	0 %	0,0455
5.12 (b)	Não-similar	5 %	0,0356
5.12 (b)	Similar	5 %	0,0626
5.12 (c)	Não-similar	0 %	0,0316
5.12 (c)	Similar	0 %	0,0589
5.12 (d)	Não-similar	5 %	0,0435
5.12 (d)	Similar	5 %	0,0878

Nos testes de generalização, também para a RCC, foram obtidos erros menores na estimação da condição de contorno quando a rede foi treinada com dados não-similares. Verifica-se também que o a rede correlação apresentou erros um pouco maior que os apresentados pela rede Perceptron de Camadas Múltiplas e pela rede Funções de Base Radial.

## CAPÍTULO 6

### CONCLUSÕES

As principais contribuições apresentadas nessa dissertação são:

- estimação da condição inicial no Problema de Condução de Calor, com o uso de três redes neurais artificiais treinadas com dois diferentes conjuntos de treinamento: o primeiro contendo dados similares e o segundo dados não similares;
- estimação da condição de contorno no Problema de Condução de Calor, também com o uso de três redes neurais artificiais treinadas com dois conjuntos de treinamento;
- estudo de um novo operador de regularização: entropia não-extensiva, onde foram investigados:
  - diferentes valores de  $q$ ,  $q < 1$  e  $q > 1$  ( $q = 1$  é a formulação padrão de Boltzmann-Gibbs-Shannon) na estimação da condição inicial no Problema de Condução de calor;
  - métodos para a determinação do parâmetro de regularização: a máxima curvatura da *curva-L* e o princípio da discrepância de Morozov.

Técnicas de Redes Neurais Artificiais foram aplicadas na solução de dois problemas inversos em condução de calor: recuperação da condição inicial e recuperação de contorno. O uso das RNA apresenta uma alternativa para solução de problemas inversos.

Pode-se observar que os conjuntos de treinamento influenciam a qualidade das inversões. Em todos os casos, tanto no treinamento das redes na recuperação da condição inicial, quanto no treinamento das redes na recuperação da condição de contorno, foram obtidos erros menores no treinamento dos dados similares, pois é necessário que seja feito o treinamento de conjuntos de funções. No entanto, nos testes de generalização, na maioria dos casos, erros menores foram obtidos com as redes treinadas com dados não similares, pois existe mais informações num conjunto contendo dados não similares.

Por se tratar de dois problemas inversos diferentes, houve diferenças significativas na fase de treinamento. Foram utilizados os mesmos conjuntos de funções para gerar os conjuntos de treinamento e adicionada a mesma taxa de ruído nos dados. No entanto, o treinamento foi muito mais difícil para o problema de recuperação da condição inicial, sendo necessárias muito mais épocas de treinamento para serem obtidos valores de erros próximos aos obtidos no treinamento no problema de recuperação da condição de contorno, isso porque existem menos informações disponíveis nos dados do problema inverso na recuperação da condição inicial.

As três redes neurais implementadas resolvem os dois problemas inversos de condução de calor. Sendo que a rede treinada com algoritmo de Correlação em Cascata apresentou um desempenho inferior nos testes de generalização. Na recuperação da condição de contorno a rede Funções de Base Radial apresentou resultados melhores na generalização.

A metodologia proposta para abordar os problemas inversos de condução de calor utilizada nesta dissertação mostrou-se eficaz na solução dos problemas que buscam a estimação da condição inicial e da condição de contorno. Assim, conclui-se que as redes neurais podem ser efetivamente usadas como ferramentas para solução de problemas inversos, para os quais dados observados estão disponíveis, ou para os quais uma solução direta pode ser obtida utilizando-se um modelo direto. Os resultados obtidos mostram o potencial das redes neurais na solução de problemas inversos.

Entretanto, dada a existência de um número crescente de modelos de redes neurais, é possível concluir que muita pesquisa ainda precisa ser realizada para explorar todo o potencial dos sistemas de redes neurais artificiais. Assim, como continuidade das pesquisas realizadas até o presente, algumas possibilidades de trabalho que resultaram de observações durante a realização dos experimentos são:

1. investigar o uso de diferentes funções de ativação e estudar a influência do parâmetro de inclinação da função de ativação tangente hiperbólica,
2. estudar a influência da posição dos sensores: nos capítulos anteriores foram somente considerados casos de sondas igualmente espaçadas, coincidindo com o número e posição dos pontos da grade computacional,
3. estudar outros tipos de redes neurais, com por exemplo a rede de Hopfield.

Para finalizar, são apresentadas as principais motivações do uso de redes neurais artificiais em problemas inversos:

- é uma nova metodologia e existem novos horizontes a serem descobertos;
- as RNA são robustas em relação a dados ruidosos;
- as RNA são paralelizáveis;
- o uso das RNA permite implementações em hardware (neurocomputadores);
- e principalmente, esta metodologia é utilizável mesmo quando não se tem um modelo matemático para o problema direto!

Com relação ao estudo do novo operador, a estratégia implícita e as técnicas de regularização adotadas neste trabalho apresentam bons resultados na reconstrução da condição inicial da equação de calor. O princípio da Discrepância de Morozov foi mais eficiente para estimar o parâmetro de regularização dos casos analisados. O critério de Hansen também apresentou boas estimativas para o parâmetro  $\alpha$ . A técnica de regularização usada nesse trabalho recupera bem a condição inicial da equação de calor para todos os parâmetros  $q$  testados.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Bertero M.; Boccacci P. **Introduction to Inverse Problems in Imaging**, Institute of Physics Publishing, Bristol, 1998.
- Boghosian B. M. R. Thermodynamic description of two-dimensional Euler turbulence using Tsallis statistic, **Physical Review**, v. 53, n. 5, p. 4754-4763, May 1996.
- Braga A. P.; Carvalho A. P. L. F.; Ludermir T. B. **Redes Neurais Artificiais** Teorias e aplicações, Rio de Janeiro: LTC, 2000. 262 p.
- Braga J. P.; Almeida M. B.; Braga A. P.; Belchior J. C. Hopfield neural network model for calculating the potential energy function from second virial data. **Chemical Physics**, v. 260, p. 347-352, 2000.
- Campos Velho H.F., Ramos F.M. Numerical Inversion of Two-Dimensional Geoelectric Conductivity Distributions from Eletromagnetic Ground Data, **Revista Brasileira de Geofísica**, v. 15, n. 2, p. 133-143, 1997.
- Duda, R. O.; Hart, P. E. **Pattern Classification and Scene Analysis**. New York: Wiley, 1973. 680 p.
- E04UCF, *NAG Fortran Library Mark 13*, Oxford, UK, 1993.
- Engl, H. W., Hanke, M., Neubauer, A. **Regularization of Inverse Problems: Mathematics and its Applications**, Dordrecht: Kluwer Academic Publishers, July 1996, 332 p.
- Falhman, S. E, Lebiere, C. The Cascade Correlation Learning Architecture, **Advances in Neural Information Processing Systems 2**, p. 524-532, San Mateo, California, 1991.

Hadamard J. **Lectures on Chauchy's Problem in Linear Partial Differential Equations**, Yale University Press, New Haven, 1923.

Hansen, P. C. Analysis of discrete ill-posed problems by means of the L-curve, **SIAM Review**, v. 34, n. 4, p. 561-580, Dec. 1992

Haykin, S. **Neural Networks: A Comprehensive Foundation**, New York: Mcmillan, 1994. 842 p.

Haykin, S. **Redes Neurais: Princípios e práticas**, Porto Alegre: Bookman, 2001. 900 p.

Hidalgo, H.; Gomez-Treviño, E. Application of Constructive Learning Algorithms to the Inverse Problem, **IEEE Transactions on Geoscience and Remote Sensing**, v. 34, n. 4, p. 874-885, Jul. 1996

Hoffman, J. D. **Numerical Methods for Engineers and Scientistis**, McGraw-Hill International Editions, 1993. 823 p.

Hopfield, J. J. Neural Networks and physical systems with emergent collective computational abilities. **Proceedings of the National Academy of Sciences of the United States of America-Biological Science**, n. 79, v. 8, p. 2554-2558, 1982.

Issamoto, E.; Miki, F.T.; Luz J.I.; Silva J.D.; Oliveira P.B.; Velho H. F. C. An Inverse Initial Condition Problem in Heat Conductions: A Neural Network Approach. [CD-ROM]. In Brazilian Congress on Mechanical Engineering (COBEM), Campinas (SP), Brasil, p. 22-26, Nov. 1999.

Jaynes, E. T. Information theory and statistical mechanics, **Physical Review**, v.106, n. 4, p. 620-630, 1957.

Kovács Z. L. **Redes Neurais Artificiais - Fundamentos e Aplicações**, Rio de Janeiro: Ed. Acadêmica 2<sup>a</sup> Ed., 1996. 174 p.

Krejsa, J.; Woodbury K. A.; Ratliff, J. D.; Raudensky, M. Assessment of Strategies and Potential for Neural Networks in Inverse Heat Conduction Problem. **Inverse Problem in Engineering**, v. 7, n. 3, p. 197-213, 1999.

McCulloch, W. S., Pitts W., A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, v. 5, p. 115-133, 1943.

Mikki, F. T., Issamoto, E., Luz, J. I., Oliveira, P. P. B., Velho, H. F. C., Silva, J. D. S. A Neural Network Approach in a Backward Heat Conduction Problem, [CD-ROM]. In: Congresso Brasileiro de Redes Neurais, p. 019-024, São José dos Campos, Brasil, 1999.

Minsky, M. L.; Papert S. **Perceptrons: An Introduction to Computational Geometry**. Cambridge: MIT Press, 1969. 292 p.

Morozov V. A., Solution of functional equations by regularization method. **Doklady Akademii Nauk SSSR**, n. 167, v. 3, p. 510, 1966.

Muniz W. B.; Campos Velho, H. F.; Ramos F. M. A Comparison of Some Inverse Methods for Estimating the Initial Condition of the Heat Equation. **Journal of Computational and Applied Mathematics**, v. 103, n. 1, p. 145-163, 1999.

Muniz W.B.; Ramos F.M.; Campos Velho H.F. Entropy- and Tikhonov-based Regularization Techniques Applied to the Backwards Heat Equation. **Computers & Mathematics with Applications**, v. 40, n. 8-9, p. 1071-1084, 2000.

Navarro P. L. K. G.; Oliveira P. P. B.; Ramos F. M. and Campos Velho H. F. An evolutionary approach in magnetotelluric inversion [CD-ROM]. In: Proc. of the 3rd



- International Conference on Inverse Problems in Engineering, Port Ludlow, USA, 1998.
- Ogawa T.; Kosugi Y; Kanada H. Neural Network Based Solution to Inverse Problems. **Proceedings International Joint Conference on Neural Networks**, p. 2471-2476, 1998.
- Plastino A.; Plastino A. R. Tsallis entropy and Jaynes' information theory formalism. **Brazilian Journal Physics**, v. 29, n.1, p. 50-60, 1999.
- Parker, D. B. Optimal algorithms for adaptive networks: second order back-propagation, second order direct propagation, and second order Hebbian learning. **IEEE 1st International Conference on Neural Network**, v. 2, p. 593-600, 1987.
- Powell, M. J. D. Radial basis functions for multivariable interpolation: A review. **IMA Conference on Algorithms for the Approximation of Functions and Data**, p.143-167, 1985.
- Ramos F. M.; Rodrigues Neto C.; Rosa R. R.; Abreu Sa L. D.; Bolzan M. J. A. Generalized thermostistical description of intermittency and nonextensivity in turbulence and financial markets. **Nonlinear Analysis - Theory**, v. 23, p. 3521-3530, 2001.
- Ramos F. M.; Rodrigues Neto C.; Neto R. R.; Bolzan M. J. A.; Abreu Sa L. D., Campos Velho H. F. Non-extensive statistics and three-dimensional fully developed turbulence, **Physica A**, v. 295, n. 1-2, p. 250-253, 2001.
- Ramos F.M. and Campos Velho H.F. Reconstruction of Geoelectric Conductivity Distributions Using a Minimum First-Order Entropy Technique. **2nd International Conference on Inverse Problems on Engineering**, v. 2, p.199-206, 1996

- Ramos F.M., Campos Velho H.F., Carvalho J.C., Ferreira N.J. Novel Approaches on Entropic Regularization. **Inverse Problems**, v. 15, n. 5, p. 1139-1148, 1999.
- Raudensky, M.; Horsky, J.; Krejsa, J.; Slama, L. Usage of Artificial Intelligence Methods in Inverse problems for Estimation of Material Parâmetros. **International Journal for Numerical Methods in Heat Transfer and Fluid Flow**, v. 6, n. 8, p.19-29, 1996.
- Raudensky, M.; Woodbury, K.A.; Kral, J., Brezina, T. Genetic Algorithm in Solution of Inverse Heat Conduction Problems. **Numerical Heat Transfer, Part B: Fundamentals**, v. 28, n. 3, p. 293-306, 1995.
- Rosenblatt, F. The Perceptron: A probabilistic model for information storage and organization in the brain. **Psychological Review**, v. 65, p. 386-408, 1958.
- Rumelhart D. E.; Hinton G. E.; Williams R. J. Learning Representations by Back-propagation errors. **Nature**, v. 323, n. 6088, p. 533-536, 1986.
- Shannon C. E.; Weaver W. **The Mathematical theory of communication**. Urbana: University of Illinois Press, 1949. 117p.
- Shiguemori E.H., Harter F.P., Campos Velho H.F., Silva J.D.S. Estimation of Boundary Conditions in Heat Transfer by Neural Networks. In XXIV CNMAC - Congresso Nacional de Matemática Aplicada e Computacional, 10-13 Setembro, Belo Horizonte (MG), Brasil, p. 559, 2001.
- Shiguemori E.H., Campos Velho H. F., Silva J.D.S. Redes Neurais Artificiais em Problema Inverso de Condução de Calor: Recuperação da Condição de Contorno. In

IWorcap - I Workshop dos Cursos de Computação Aplicada do INPE, 25 Outubro, São José dos Campos (SP), Brasil, p. 54-56, 2001.

Shiguemori E.H., Campos Velho H.F., Silva J.D.S. Estimation of Initial Condition in Heat Conduction by Neural Network, 4th International Conference on Inverse Problems in Engineering: Theory and Practice, Maio 26-31, Angra dos Reis (RJ), Brasil, 2002.

Shiguemori E.H., Campos Velho H.F., Ramos F.M., Silva J.D.S. A Parametric Study of a New Regularization Operator: Non-extensive Entropy. In 4th International Conference on Inverse Problems in Engineering: Theory and Practice, Maio 26-31, Angra dos Reis (RJ), Brasil, 2002

Shiguemori E. H., Silva J.D.S., Campos Velho H.F. Neural Network for Estimating the Initial Condition in a Heat Conduction Problem. In International Joint Conference on Neural Networks, 12-17 Maio, Honolulu, Hawaii, 2002

Silva, J.D.S. The Cascade-Correlation Neural Network Growing Algorithm using the Matlab Environment [on line] <http://www.lac.inpe.br/~demisio/publica/cascabst.html>, 1997.

Tikhonov A. N. On stability of inverse problems. **Doklady Akademii Nauk SSSR**, v. 39, n. 5, p. 195-198, 1943.

Tikhonov A. N. On solving incorrectly posed problems and methods of regularization. **Doklady Akademii Nauk SSSR**, v. 151, n. 3, p. 501-504, 1963.

Tikhonov A. N.; Arsenin V. Y. **Solutions of Ill-Posed Problems**, New York: Winston & Sons, 1977. 258 p.

Tsallis C. Possible generalization of Boltzmann-Gibbs statistics. **Journal of Statistical Physics**, v. 54, n.1-2, p. 479-487, 1988.

Tsallis C. Nonextensive statistics: theoretical, experimental and computational evidences and connections. **Brazilian Journal of Physics**, p. 1-35, n. 29, v.1, 1999.

Tsoukalas L. H., Uhrig R. E. **Fundamental of Neural Networks**, 1995. 587 p.

Werbos, P. J., Backpropagation and neurocontrol: A review and prospectus. **International Join Conference on Neural Networks**, v. 1, p. 209-216, 1989.

Woodbury, K. A. Neural Networks and Genetic Algorithms in the Solution of Inverse Problems. **Workshop on Novel Approaches for Inverse Problems**, São José dos Campos, Nov. 1998.