

Eduardo Pestana de Aguiar

*Sistema de Inferência Fuzzy para Classificação de Distúrbios em
Sinais Elétricos*

Juiz de Fora

Agosto 2011

Eduardo Pestana de Aguiar

*Sistema de Inferência Fuzzy para Classificação de Distúrbios em
Sinais Elétricos*

Dissertação apresentada ao Programa de Pós Graduação em Engenharia Elétrica, área de concentração: Sistemas Eletrônicos, da Universidade Federal de Juiz de Fora, como requisito parcial para obtenção do grau de Mestre.

Orientador: Moisés Vidal Ribeiro

Juiz de Fora

Agosto 2011

Aguiar, Eduardo Pestana de.

Sistema de inferência Fuzzy para classificação de distúrbios em sinais elétricos / Eduardo Pestana de Aguiar. – 2011.

182 f. : il.

Dissertação (Mestrado em Engenharia Elétrica)—Universidade Federal de Juiz de Fora, Juiz de Fora, 2011.

1. Energia elétrica. 2. Algoritmos. I. Título.

CDU 621.3

Termo de Aprovação

Eduardo Pestana de Aguiar

Sistema de Inferência *Fuzzy* para Classificação de Distúrbios em Sinais
Elétricos

Dissertação apresentada ao Programa de Pós Graduação em Engenharia Elétrica, área de concentração: Sistemas Eletrônicos, da Universidade Federal de Juiz de Fora, como requisito parcial para obtenção do grau de Mestre, pela seguinte banca examinadora:

Prof. Dr. Moisés Vidal Ribeiro (Orientador)
Universidade Federal de Juiz de Fora -
UFJF

Prof. Dr. Clodoaldo Aparecido de Moraes
Lima
Universidade de São Paulo - USP

Prof. Dr. Carlos Augusto Duque
Universidade Federal de Juiz de Fora -
UFJF

Prof. Dr. Michel Bortolini Hell
Universidade Federal de Juiz de Fora -
UFJF

Juiz de Fora, 30 de Agosto de 2011

DEDICATÓRIA

Aos meus pais Gilberto e Sônia.

AGRADECIMENTOS

A Deus, por ter me dado saúde, força de vontade para superar os obstáculos e por ter colocado em minha vida as pessoas certas.

Ao Prof. Dr. Moisés Vidal Ribeiro pela orientação, paciência, disponibilidade, incentivo, amizade e crédito ao meu trabalho durante este gratificante processo de aprendizagem.

Ao Prof. Dr. Michel Bortolini Hell pela atenção, colaboração e sobretudo pelas valiosas discussões acerca de sistemas de inferência *fuzzy*.

Ao Prof. Dr. Carlos Augusto Duque pelos conhecimentos transmitidos, pelas discussões sobre qualidade da energia elétrica e pelas prestativas colaborações.

Ao Prof. Dr. Clodoaldo Aparecido de Moraes Lima, pela disponibilidade e contribuição.

Aos meus pais Gilberto e Sônia, exemplos de perseverança e dedicação, a quem devo muito do que sou; creio que mesmo todo agradecimento não seria suficiente para recompensá-los. Obrigado pelos ensinamentos, pelo carinho, estímulo e suporte em todas as etapas de minha vida.

Aos meus familiares que souberam entender minhas ausências devido ao desenvolvimento deste trabalho e mesmo assim sempre me ampararam com palavras de apoio, carinho e compreensão.

À Giselle por todo carinho, incentivo, dedicação e paciência ao longo desta caminhada.

Aos amigos Laryssa e Gléverson, pelas valiosas orientações na formatação deste documento. Agradeço também a amizade, apoio e companhia enquanto trabalhávamos nas nossas respectivas dissertações.

Aos amigos do Laboratório de Processamento de Sinais e Telecomunicações (LAPTEL) pela amizade e convivência, em especial aos amigos Cristiano, Thiago e La-Gatta, pelo apoio concedido em prol do correto desfecho deste trabalho.

Aos amigos da Becton Dickinson, em especial ao gerente de engenharia Francisco Raposo, pois além da disponibilidade concedida para a realização deste trabalho, proporcionaram uma oportunidade singular de continuidade das minhas atividades de pesquisa e de estar envolvido com o ambiente profissional desta empresa.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro concedido durante parte deste trabalho.

Ao Povo Brasileiro que indiretamente ou diretamente financiou este trabalho.

EPÍGRAFE

“Educar a inteligência é dilatar o horizonte dos seus desejos e das suas necessidades.”

James Russell Lowell

SUMÁRIO

Lista de Figuras	
Lista de Tabelas	
Lista de Abreviaturas e Siglas	
Resumo	
Abstract	
1 Introdução	26
1.1 Objetivo do trabalho	28
1.2 Divisão do trabalho	28
1.3 Sumário	29
2 Classificação de distúrbios	30
2.1 Revisão sobre classificação de distúrbios	30
2.2 Classificação de distúrbios baseada em EOS	32
2.2.1 <i>Extração de características</i>	33
2.2.2 <i>Seleção de características</i>	35
2.2.3 <i>Classificadores</i>	35
2.2.3.1 Classificador de Bayes	36
2.2.3.2 Classificador MLP	37
2.3 Sumário	39

3	Formulação do Problema	40
3.1	Sumário	47
4	Sistemas <i>Fuzzy</i>	48
4.1	Conjuntos <i>fuzzy</i>	49
4.2	Normas triangulares	50
4.3	Relações <i>fuzzy</i>	51
4.4	Composição de relações <i>fuzzy</i>	52
4.5	Sistema de inferência <i>fuzzy singleton</i>	53
4.6	Sistema de inferência <i>fuzzy non-singleton</i>	61
4.7	Sumário	66
5	Propostas	67
5.1	Matriz Hessiana para o sistema de inferência <i>fuzzy singleton</i>	68
5.2	Matriz Hessiana para o sistema de inferência <i>fuzzy non-singleton</i>	74
5.3	Método do gradiente conjugado para sistemas de inferência <i>fuzzy</i>	80
5.4	Sumário	84
6	Resultados Experimentais	85
6.1	Convergência	89
6.2	Desempenho	102
6.3	Sumário	118
7	Conclusões	119
	Referências	122
	Apêndice A – Produção Bibliográfica	127
A.1	Artigos em revista	127
A.2	Artigos em congressos	127

Apêndice B – Código em MATLAB[®] utilizado para a geração dos bancos de dados sintéticos	128
B.1 Rotina principal	128
Apêndice C – Matriz Hessiana para o sistema de inferência <i>fuzzy</i>	133
C.1 Dedução da matriz Hessiana para o sistema de inferência <i>fuzzy singleton</i>	133
C.1.1 Dedução de $\mathbf{H}_{m_{F_j^i}(q), \sigma_{F_k^l}(q)}$	133
C.1.2 Dedução de $\mathbf{H}_{m_{F_j^i}(q), \theta_l(q)}$	135
C.1.3 Dedução de $\mathbf{H}_{\sigma_{F_j^i}(q), m_{F_k^l}(q)}$	136
C.1.4 Dedução de $\mathbf{H}_{\sigma_{F_j^i}(q), \sigma_{F_k^l}(q)}$	138
C.1.5 Dedução de $\mathbf{H}_{\sigma_{F_j^i}(q), \theta_l(q)}$	140
C.1.6 Dedução de $\mathbf{H}_{\theta_l(q), m_{F_k^l}(q)}$	142
C.1.7 Dedução de $\mathbf{H}_{\theta_l(q), \sigma_{F_k^l}(q)}$	143
C.1.8 Dedução de $\mathbf{H}_{\theta_l(q), \theta_l(q)}$	145
C.2 Dedução da matriz Hessiana para o sistema de inferência <i>fuzzy non-singleton</i> ..	146
C.2.1 Dedução de $\mathbf{H}_{m_{F_j^i}(q), \sigma_{F_k^l}(q)}$	146
C.2.2 Dedução de $\mathbf{H}_{m_{F_j^i}(q), \theta_l(q)}$	148
C.2.3 Dedução de $\mathbf{H}_{m_{F_j^i}(q), \sigma_X(q)}$	149
C.2.4 Dedução de $\mathbf{H}_{\sigma_{F_j^i}(q), m_{F_k^l}(q)}$	151
C.2.5 Dedução de $\mathbf{H}_{\sigma_{F_j^i}(q), \sigma_{F_k^l}(q)}$	153
C.2.6 Dedução de $\mathbf{H}_{\sigma_{F_j^i}(q), \theta_l(q)}$	156
C.2.7 Dedução de $\mathbf{H}_{\sigma_{F_j^i}(q), \sigma_X(q)}$	157
C.2.8 Dedução de $\mathbf{H}_{\theta_l(q), m_{F_k^l}(q)}$	159
C.2.9 Dedução de $\mathbf{H}_{\theta_l(q), \sigma_{F_k^l}(q)}$	161
C.2.10 Dedução de $\mathbf{H}_{\theta_l(q), \theta_l(q)}$	162
C.2.11 Dedução de $\mathbf{H}_{\theta_l(q), \sigma_X(q)}$	163
C.2.12 Dedução de $\mathbf{H}_{\sigma_X(q), m_{F_k^l}(q)}$	164

C.2.13	<i>Dedução de $\mathbf{H}_{\sigma_X(q), \sigma_{F_k^l}(q)}$</i>	166
C.2.14	<i>Dedução de $\mathbf{H}_{\sigma_X(q), \theta_l(q)}$</i>	168
C.2.15	<i>Dedução de $\mathbf{H}_{\sigma_X(q), \sigma_X(q)}$</i>	170
Apêndice D - Parâmetros utilizados no sistema de inferência <i>fuzzy</i>		172
D.1	Parâmetros do gradiente conjugado para o sistema de inferência <i>fuzzy singleton</i>	172
D.2	Parâmetros do gradiente conjugado para o sistema de inferência <i>fuzzy non-singleton</i>	178

LISTA DE FIGURAS

Figura 1	Técnica padrão para classificação de distúrbios presentes em sinais elétricos	32
Figura 2	Paradigma apresentado em (RIBEIRO; PEREIRA, 2007) para a classificação de distúrbios isolados e múltiplos.	44
Figura 3	Paradigma proposto para a classificação de distúrbios isolados e múltiplos.	44
Figura 4	Comportamento booleano de uma situação.	48
Figura 5	Flexibilização dos limites de uma situação.	49
Figura 6	Sistema de inferência <i>fuzzy singleton</i> utilizada, adaptada de (RIBEIRO, 2005).	54
Figura 7	Exemplo de função de pertinência Gaussiana.	55
Figura 8	Funções de pertinência para a modelar a incerteza no valor da temperatura.	55
Figura 9	Interpretação da máquina de inferência <i>fuzzy</i> como um sistema.	57
Figura 10	Pré-filtragem da entrada para um sistema de inferência <i>fuzzy non-singleton</i>	63

Figura 11	Valores do FDR obtidos para o distúrbio isolado <i>flicker</i>	87
Figura 12	Valores do FDR obtidos para o distúrbio múltiplo <i>sag+flicker</i>	88
Figura 13	Convergência do sistema de inferência <i>fuzzy singleton</i> para o distúrbio <i>sag</i>	89
Figura 14	Convergência do sistema de inferência <i>fuzzy singleton</i> para o distúrbio <i>swell</i>	90
Figura 15	Convergência do sistema de inferência <i>fuzzy singleton</i> para o distúrbio <i>flicker</i>	90
Figura 16	Convergência do sistema de inferência <i>fuzzy singleton</i> para o distúrbio harmônica.	91
Figura 17	Convergência do sistema de inferência <i>fuzzy singleton</i> para o distúrbio <i>sag+flicker</i>	92
Figura 18	Convergência do sistema de inferência <i>fuzzy singleton</i> para o distúrbio <i>sag+harmônica</i>	92
Figura 19	Convergência do sistema de inferência <i>fuzzy singleton</i> para o distúrbio <i>swell+flicker</i>	93
Figura 20	Convergência do sistema de inferência <i>fuzzy singleton</i> para o distúrbio <i>swell+harmônica</i>	93
Figura 21	Convergência do sistema de inferência <i>fuzzy singleton</i> para o distúrbio <i>sag+harmônica+flicker</i>	94

Figura 22	Convergência do sistema de inferência <i>fuzzy singleton</i> para o distúrbio <i>swell+harmônica+flicker</i> .	94
Figura 23	Convergência do sistema de inferência <i>fuzzy non-singleton</i> para o distúrbio <i>sag</i> .	95
Figura 24	Convergência do sistema de inferência <i>fuzzy non-singleton</i> para o distúrbio <i>swell</i> .	96
Figura 25	Convergência do sistema de inferência <i>fuzzy non-singleton</i> para o distúrbio <i>flicker</i> .	96
Figura 26	Convergência do sistema de inferência <i>fuzzy non-singleton</i> para o distúrbio harmônica.	97
Figura 27	Convergência do sistema de inferência <i>fuzzy non-singleton</i> para o distúrbio <i>sag+flicker</i> .	98
Figura 28	Convergência do sistema de inferência <i>fuzzy non-singleton</i> para o distúrbio <i>sag+harmônica</i> .	99
Figura 29	Convergência do sistema de inferência <i>fuzzy non-singleton</i> para o distúrbio <i>swell+flicker</i> .	99
Figura 30	Convergência do sistema de inferência <i>fuzzy non-singleton</i> para o distúrbio <i>swell+harmônica</i> .	100
Figura 31	Convergência do sistema de inferência <i>fuzzy non-singleton</i> para o distúrbio <i>sag+harmônica+flicker</i> .	100
Figura 32	Convergência do sistema de inferência <i>fuzzy non-singleton</i> para o distúrbio	

<i>swell</i> + <i>harmônica</i> + <i>flicker</i>	101
--	-----

LISTA DE TABELAS

Tabela 1	Classes de distúrbios a serem classificados	86
Tabela 2	Desempenho dos classificadores de Bayes e rede neural <i>perceptron</i> multicamadas em (%), para distúrbios isolados e múltiplos.	102
Tabela 3	Desempenho do sistema de inferência <i>fuzzy singleton</i> em (%), para o distúrbio <i>sag</i>	103
Tabela 4	Desempenho do sistema de inferência <i>fuzzy singleton</i> em (%), para o distúrbio <i>swell</i>	103
Tabela 5	Desempenho do sistema de inferência <i>fuzzy singleton</i> em (%), para o distúrbio <i>flicker</i>	104
Tabela 6	Desempenho do sistema de inferência <i>fuzzy singleton</i> em (%), para o distúrbio harmônica.	104
Tabela 7	Desempenho do sistema de inferência <i>fuzzy singleton</i> em (%), para o distúrbio <i>sag+flicker</i>	106
Tabela 8	Desempenho do sistema de inferência <i>fuzzy singleton</i> em (%), para o distúrbio <i>sag+harmônica</i>	106
Tabela 9	Desempenho do sistema de inferência <i>fuzzy singleton</i> em (%), para o distúrbio <i>swell+flicker</i>	107

Tabela 10	Desempenho do sistema de inferência <i>fuzzy singleton</i> em (%), para o distúrbio <i>swell</i> +harmônica.	107
Tabela 11	Desempenho do sistema de inferência <i>fuzzy singleton</i> em (%), para o distúrbio <i>sag</i> +harmônica+ <i>flicker</i>	108
Tabela 12	Desempenho do sistema de inferência <i>fuzzy singleton</i> em (%), para o distúrbio <i>swell</i> +harmônica+ <i>flicker</i>	108
Tabela 13	Desempenho do sistema de inferência <i>fuzzy non-singleton</i> em (%), para o distúrbio <i>sag</i>	110
Tabela 14	Desempenho do sistema de inferência <i>fuzzy non-singleton</i> em (%), para o distúrbio <i>swell</i>	110
Tabela 15	Desempenho do sistema de inferência <i>fuzzy non-singleton</i> em (%), para o distúrbio <i>flicker</i>	111
Tabela 16	Desempenho do sistema de inferência <i>fuzzy non-singleton</i> em (%), para o distúrbio <i>harmônica</i>	111
Tabela 17	Desempenho do sistema de inferência <i>fuzzy non-singleton</i> em (%), para o distúrbio <i>sag</i> + <i>flicker</i>	113
Tabela 18	Desempenho do sistema de inferência <i>fuzzy non-singleton</i> em (%), para o distúrbio <i>sag</i> +harmônica.	113
Tabela 19	Desempenho do sistema de inferência <i>fuzzy non-singleton</i> em (%), para o distúrbio <i>swell</i> + <i>flicker</i>	114
Tabela 20	Desempenho do sistema de inferência <i>fuzzy non-singleton</i> em (%), para o	

	distúrbio <i>swell</i> +harmônica.	114
Tabela 21	Desempenho do sistema de inferência <i>fuzzy non-singleton</i> em (%), para o distúrbio <i>sag</i> +harmônica+ <i>flicker</i>	115
Tabela 22	Desempenho do sistema de inferência <i>fuzzy non-singleton</i> em (%), para o distúrbio <i>swell</i> +harmônica+ <i>flicker</i>	115
Tabela 23	Desempenho dos classificadores de Bayes, rede neural <i>perceptron</i> multicamadas e sistema de inferência <i>fuzzy singleton</i> e <i>non-singleton</i> em (%), para distúrbios isolados e múltiplos.	117
Tabela 24	Parâmetros do método do gradiente conjugado para o sistema de inferência <i>fuzzy singleton</i> aplicado ao distúrbio <i>sag</i>	172
Tabela 25	Parâmetros do método do gradiente conjugado para o sistema de inferência <i>fuzzy singleton</i> aplicado ao distúrbio <i>swell</i>	173
Tabela 26	Parâmetros do método do gradiente conjugado para o sistema de inferência <i>fuzzy singleton</i> aplicado ao distúrbio <i>flicker</i>	173
Tabela 27	Parâmetros do método do gradiente conjugado para o sistema de inferência <i>fuzzy singleton</i> aplicado ao distúrbio harmônica.	174
Tabela 28	Parâmetros do método do gradiente conjugado para o sistema de inferência <i>fuzzy singleton</i> aplicado ao distúrbio <i>sag</i> + <i>flicker</i>	174
Tabela 29	Parâmetros do método do gradiente conjugado para o sistema de inferência <i>fuzzy singleton</i> aplicado ao distúrbio <i>sag</i> +harmônica.	175
Tabela 30	Parâmetros do método do gradiente conjugado para o sistema de inferência <i>fuzzy singleton</i> aplicado ao distúrbio <i>swell</i> + <i>flicker</i>	175

Tabela 31	Parâmetros do método do gradiente conjugado para o sistema de inferência <i>fuzzy singleton</i> aplicado ao distúrbio <i>swell</i> +harmônica.	176
Tabela 32	Parâmetros do método do gradiente conjugado para o sistema de inferência <i>fuzzy singleton</i> aplicado ao distúrbio <i>sag</i> +harmônica+ <i>flicker</i>	176
Tabela 33	Parâmetros do método do gradiente conjugado para o sistema de inferência <i>fuzzy singleton</i> aplicado ao distúrbio <i>swell</i> +harmônica+ <i>flicker</i>	177
Tabela 34	Parâmetro $\alpha_{\max}(q)$ do sistema de inferência <i>fuzzy singleton</i> para o método proposto por Daniel (DANIEL, 1967).	177
Tabela 35	Parâmetros do método do gradiente conjugado para o sistema de inferência <i>fuzzy non-singleton</i> aplicado ao distúrbio <i>sag</i>	178
Tabela 36	Parâmetros do método do gradiente conjugado para o sistema de inferência <i>fuzzy non-singleton</i> aplicado ao distúrbio <i>swell</i>	178
Tabela 37	Parâmetros do método do gradiente conjugado para o sistema de inferência <i>fuzzy non-singleton</i> aplicado ao distúrbio <i>flicker</i>	179
Tabela 38	Parâmetros do método do gradiente conjugado para o sistema de inferência <i>fuzzy non-singleton</i> aplicado ao distúrbio harmônica.	179
Tabela 39	Parâmetros do método do gradiente conjugado para o sistema de inferência <i>fuzzy non-singleton</i> aplicado ao distúrbio <i>sag</i> + <i>flicker</i>	180
Tabela 40	Parâmetros do método do gradiente conjugado para o sistema de inferência <i>fuzzy non-singleton</i> aplicado ao distúrbio <i>sag</i> +harmônica.	180
Tabela 41	Parâmetros do método do gradiente conjugado para o sistema de inferência	

	<i>fuzzy non-singleton</i> aplicado ao distúrbio <i>swell+flicker</i>	181
Tabela 42	Parâmetros do método do gradiente conjugado para o sistema de inferência <i>fuzzy non-singleton</i> aplicado ao distúrbio <i>swell+harmônica</i>	181
Tabela 43	Parâmetros do método do gradiente conjugado para o sistema de inferência <i>fuzzy non-singleton</i> aplicado ao distúrbio <i>sag+harmônica+flicker</i>	182
Tabela 44	Parâmetros do método do gradiente conjugado para o sistema de inferência <i>fuzzy non-singleton</i> aplicado ao distúrbio <i>swell+harmônica+flicker</i>	182
Tabela 45	Parâmetro $\alpha_{\max}(q)$ do sistema de inferência <i>fuzzy non-singleton</i> para o método proposto por Daniel (DANIEL, 1967).	183

LISTA DE ABREVIATURAS E SIGLAS

EOS estatísticas de ordem superior

FBF *fuzzy basis function*

FDR *Fisher's discriminant ratio*

FFNN *feed-forward neural network*

FLS *fuzzy logic system*

GC gradiente conjugado

GCEM método gradiente conjugado, escalonado e modificado

IEC *International Electrotechnical Commission*

IEEE *Institute of Electrical and Electronics Engineers*

IIR resposta ao impulso infinita, do inglês, *infinite impulse response*

i.i.d. independente e identicamente distribuídas

ML critério da máxima verossimilhança - *maximum likelihood*

MLP rede neural artificial perceptron multicamadas - *multilayer perceptron*

MSE erro quadrático médio - *mean-square error*

QEE qualidade da energia elétrica

RNA rede neural artificial

TDNN *time-delay neural network*

SEP sistemas elétricos de potência

SFFS *sequential forward floating search*

SNR relação sinal-ruído - *signal-to-noise ratio*

RESUMO

A presente dissertação tem como objetivo discutir o uso de técnicas de otimização baseadas no gradiente conjugado e de informações de segunda ordem para o treinamento de sistemas de inferência *fuzzy singleton* e *non-singleton*. Além disso, as soluções computacionais derivadas são aplicadas aos problemas de classificação de distúrbios múltiplos e isolados em sinais elétricos. Os resultados computacionais, obtidos a partir de dados sintéticos de distúrbios em sinais de tensão, indicam que os sistemas de inferência *fuzzy singleton* e *non-singleton* treinados pelos algoritmos de otimização considerados apresentam maior velocidade de convergência e melhores taxas de classificação quando comparados com aqueles treinados pelo algoritmo de otimização baseada em informações de primeira ordem e é bastante competitivo em relação à rede neural artificial perceptron multicamadas - *multilayer perceptron* (MLP) e ao classificador de Bayes.

Palavras-chave: Sistema de inferência *fuzzy*; Gradiente conjugado; Classificação de distúrbios; Qualidade da energia elétrica; Sinais elétricos; Algoritmos de otimização.

ABSTRACT

This master dissertation aims to discuss the use of optimization techniques based on the conjugated gradient and on second order information for the training of singleton or non-singleton fuzzy inference systems. In addition, the computational solutions obtained are applied to isolated a multiple disturbances classification problems in electric signals. Computational results obtained from synthetic data from disturbances in electric signals indicate that singleton or non-singleton fuzzy inference systems trained by the considered optimization algorithms present greater convergence speed and better classification rates when compared to those data trained by an optimization algorithm based on first order information and is quite competitive with multilayer perceptron neural network and Bayesian classifier.

Key-words: Fuzzy inference system; Conjugated gradient; Classification of disturbances; Power quality; Electric signals; Optimization algorithms.

1 INTRODUÇÃO

O crescente uso de equipamentos computadorizados e sensíveis às perturbações do sistema elétrico tem exigido requisitos da qualidade da energia elétrica (QEE) cada vez mais rigorosos (HEYDT, 1998), o que, juntamente com a necessidade das concessionárias de energia elétrica de satisfazer os clientes, tem tornado o tema “qualidade da energia” um assunto de extrema importância no mercado atual de energia elétrica.

O aumento do número de cargas de natureza não-linear (sistemas microprocessados, fontes chaveadas, etc) em instalações residenciais, comerciais e industriais provoca o aumento significativo da ocorrência de distúrbios em sistemas de potência assim como o do número de falhas e, conseqüentemente, a diminuição da vida útil dos equipamentos conectados às redes elétricas (RIBEIRO, 2005). Destaca-se ainda a crescente interligação entre processos produtivos no ambiente industrial (ARRILLAGA; BOLLEN; WATSON, 2000).

De maneira geral, os distúrbios de QEE, também denominados eventos, se manifestam como deformidades nas formas de onda de tensão e corrente. Estas deformidades, frequentemente referidas como problemas de QEE, afetam significativamente as indústrias em diversos aspectos (WANG; MAMISHEV, 2004). Dentre eles, destacam-se a interrupção do processo de produção, o que é refletido em significativas perdas financeiras para as indústrias.

O crescente aumento de problemas relacionados à QEE, em sistemas elétricos de potência (SEP), tem levado ao longo dos últimos anos ao desenvolvimento de diversas técnicas de processamento de sinais para o monitoramento e análise de tais problemas. Dentre estas técnicas podemos citar: *i*) detecção e segmentação de distúrbios, *ii*) classificação de distúrbios, *iii*) identificação de fontes de distúrbios, *iv*) localização de fontes de distúrbios, *v*) compressão de sinais, *vi*) estimação de parâmetros do sinal e *vii*) decomposição ou representação de sinais.

O termo QEE não possui uma definição única, fazendo com que existam algumas definições que entrem em conflito umas com as outras. O Instituto de Engenheiros Eletricistas e Eletrônicos, do inglês *Institute of Electrical and Electronics Engineers* (IEEE), define o termo QEE como “O conceito de fornecer e estabelecer a alimentação de um equipamento elétrico sensível de forma adequada ao seu funcionamento”. Já a Comissão Eletrotécnica Internacional, do inglês *International Electrotechnical Commission* (IEC), define o termo como “Características da eletricidade em um dado ponto do sistema elétrico, em relação a um conjunto de parâmetros técnicos de referência”. Note que a definição dada pelo IEEE ressalta a importância da operação do equipamento, não dando muita importância a determinados distúrbios que não afetam a operação do mesmo, já a definição dada pela IEC se preocupa com os parâmetros de referência do sistema, não importando o tipo de equipamento a ele conectado. No entanto, uma definição mais abrangente do termo QEE para processamento de sinais, a qual será adotada nesta dissertação, é encontrado em (BOLLEN; GU, 2006), o qual define: “Qualidade da energia elétrica é a combinação entre a qualidade da tensão e a qualidade da corrente. A qualidade da tensão é um conceito relacionado a uma tensão sem desvios em relação a uma tensão ideal, o mesmo ocorrendo com a qualidade da corrente. Uma tensão ideal é uma tensão senoidal com amplitude e frequência constante, onde ambos apresentam valores nominais”.

A alguns anos, a preocupação com a QEE era relativamente baixa devido ao fato dos equipamentos elétricos não serem muito sensíveis a pequenas variações ocorridas na tensão e na corrente. Este tema tem se tornado bastante importante devido a vários fatores, dentre os quais merecem destaque: *i*) o crescente avanço da eletrônica de potência, fazendo proliferar cargas não-lineares conectadas ao sistema; *ii*) o aparecimento de equipamentos e sistemas de controle utilizando microprocessadores sensíveis a diversos distúrbios relacionados à QEE; *iii*) menor tolerância dos consumidores em relação aos problemas relacionados à QEE, principalmente os consumidores industriais.

Como mencionado, o termo QEE tem sido foco de muitos trabalhos nos últimos anos. Em especial, técnicas de processamento de sinais tem sido desenvolvidas para monitorar e analisar os problemas inerentes a QEE. No que tange a técnicas de classificação, além das mesmas necessitarem apresentar algoritmos de bom desempenho para a classificação de distúrbios em sinais elétricos, também necessitam apresentar complexidade computacional reduzida para que sejam implementadas em hardware de baixo custo, de forma que sua utilização seja viável economicamente. Neste trabalho é proposta ainda a derivação de algoritmos de treinamento baseados no método do Gradiente Conjugado e em informações de segunda ordem (matriz Hessiana), a fim de ajustar os parâmetros dos

classificadores *fuzzy*, com o intuito de tornar o sistema melhor adaptado ao problema.

1.1 Objetivo do trabalho

A presente contribuição discute uma técnica de classificação de distúrbios isolados e múltiplos derivada a partir da proposição introduzida em (RIBEIRO; PEREIRA, 2007). Basicamente, o sinal de tensão é decomposto em duas componentes primitivas, a saber: componente fundamental e distúrbios. A partir dessas componentes, estatísticas de ordens superiores são seletivamente extraídas. A seguir, um classificador baseado em sistemas de inferência *fuzzy singleton* ou *non-singleton* é aplicado. No que tange o classificador *fuzzy*, nesta contribuição, é proposta a derivação de algoritmos de treinamento baseados no método do Gradiente Conjugado e em informações de segunda ordem (matriz Hessiana).

1.2 Divisão do trabalho

Esta dissertação está organizada da seguinte forma: No capítulo 2 é apresentada uma revisão sobre as técnicas de classificação de distúrbios presentes em sinais elétricos. O referido capítulo aborda as metodologias que serão utilizadas para a extração de parâmetros utilizando estatísticas de ordem superior (EOS) e a posterior seleção dos referidos parâmetros utilizando o discriminante de Fisher e as técnicas de classificação de distúrbios baseadas no classificador Bayesiano e em rede neural MLP.

O capítulo 3, apresenta a formulação do problema de classificação de distúrbios da QEE. Primeiramente, discute-se a decomposição do sinal a ser monitorado em diversas componentes e a seguir, apresenta-se o problema de classificação como a decisão entre hipóteses que serão apresentadas.

Já o capítulo 4 aborda a estruturação e projeto de um sistema de inferência *fuzzy*. São apresentados os principais termos e definições para o sistema de inferência *fuzzy singleton* e *non-singleton* juntamente como os respectivos parâmetros do sistema e suas atualizações, que são feitas através do algoritmo de retropropagação.

O Capítulo 5 descreve um sistema de inferência *fuzzy singleton* e *non-singleton* treinado por algoritmos do Gradiente Conjugado e em informações de segunda ordem (matriz Hessiana), visando, sobretudo, mostrar que tal técnica baseada em inteligência computacional pode fornecer uma maior velocidade de convergência e desempenho para aplicações *offline*.

No capítulo 6 são apresentados e discutidos alguns resultados da aplicação das metodologias e técnicas propostas e analisadas nesta dissertação.

Finalmente, o Capítulo 7 apresenta as conclusões e observações finais da presente dissertação.

1.3 Sumário

O presente capítulo apresentou uma breve introdução da dissertação e do conteúdo da mesma, relacionando de forma resumida os principais tópicos envolvidos. A seguir, no Capítulo 2, é apresentada uma revisão das técnicas e metodologias utilizadas para a classificação de distúrbios de QEE.

2 CLASSIFICAÇÃO DE DISTÚRBIOS

Recentemente, uma grande atenção tem sido dada ao uso de técnicas de processamento de sinais e inteligência computacional para o desenvolvimento de ferramentas para caracterizar, analisar e avaliar a QEE e o comportamento de cargas conectadas ao sistema elétrico. Do ponto de vista de processamento de sinais, a análise da QEE pode ser realizada de acordo com os seguintes tópicos (RIBEIRO, 2005): *i*) detecção e segmentação de distúrbios, *ii*) classificação de distúrbios, *iii*) identificação de fontes de distúrbios, *iv*) localização de fontes de distúrbios, *v*) compressão de sinais, *vi*) estimação de parâmetros do sinal e *vii*) decomposição ou representação de sinais.

Neste contexto, o presente capítulo apresenta uma revisão das técnicas e metodologias aplicadas à classificação de distúrbios em sinais elétricos. Primeiramente, a Seção 2.1, apresenta uma breve revisão de técnicas tradicionais de classificação de distúrbios presentes em sinais elétricos. A seguir, a Seção 2.2 apresenta as metodologias que serão utilizadas para a extração de parâmetros utilizando estatísticas de ordem superior (EOS) e a posterior seleção dos referidos parâmetros utilizando o discriminante de Fisher e as técnicas de classificação de distúrbios baseadas no classificador Bayesiano e em rede neural MLP. Por fim, a Seção 2.3 apresenta a síntese do que será discutido no capítulo em questão.

2.1 Revisão sobre classificação de distúrbios

A classificação ou reconhecimento de distúrbios é um importante assunto no que tange ao desenvolvimento da próxima geração de equipamentos de monitoramento da QEE. De fato, tal tarefa requer o uso de ferramentas de reconhecimento de padrões sofisticadas e eficientes para aplicações em tempo real.

Para alcançar esse objetivo, ultimamente técnicas de classificação têm sido ampla-

mente aplicadas para analisar um único distúrbio presente no sinal. Entretanto, sabe-se que durante um comportamento anormal, os sinais de tensão e corrente são corrompidos não apenas por um único distúrbio, mas também por vários. Como a maioria das técnicas desenvolvidas para classificação de distúrbios isolados tem uma aplicação limitada no monitoramento da QEE, recentemente o foco de pesquisa tem sido no desenvolvimento de técnicas capazes de classificar múltiplos distúrbios.

Exemplos atuais de propostas voltadas para a classificação de múltiplos distúrbios são reportados em (RIBEIRO, 2005), (RIBEIRO; PEREIRA, 2007) e (GAING, 2004). No entanto, apesar dos bons resultados apresentados por estas técnicas, nota-se que ainda há grandes desafios para a classificação de múltiplos distúrbios.

A classificação de distúrbios em sinais elétricos, apresentada em IEEE 1159 (QUALITY, 1995) e em EN 50160 (STANDARDIZATION, 1999) motivou o desenvolvimento de diversas técnicas que permitem um reconhecimento automático e confiável dos eventos que se deseja classificar. Tais metodologias são aplicadas como ferramentas para compreender efeitos em cargas e para auxiliar a análise de fontes geradoras de distúrbios (CHUNG et al., 2002a). As diferentes técnicas propostas na literatura para a classificação de distúrbios em sinais elétricos utilizam metodologias que permitem obter características do sinal original que são classificadas segundo a sua magnitude e duração e são apresentadas em (CHOONG; REAZ; MOHD, 2005) e (IBRAHIM; MORCOS, 2002). Para realizar a classificação destes distúrbios, diferentes algoritmos foram definidos, a fim de relacionar as características do sinal com o grupo a que pertencem: Sistema de inferência *fuzzy*, redes neurais artificiais (RNA), classificador de Bayes e outros. Estas metodologias têm um papel importante na classificação de distúrbios, uma vez que o seu desempenho depende dos parâmetros extraídos e da maneira como o classificador a ser utilizado foi projetado. Note que se as características do distúrbio não são capturadas com precisão, o desempenho da classificação conseqüentemente será limitado.

Sabe-se que a rede neural artificial possui como principal vantagem a flexibilidade matemática proporcionada pela sua concepção e é fortemente recomendada para aplicações em tempo real. Todavia, como desvantagem, tem-se que a função de erro a ser minimizada é multimodal e possui muitos mínimos locais, onde o processo de aprendizagem pode ser restrito. Adicionalmente, a velocidade de convergência, robustez e precisão dependem fortemente da arquitetura de rede escolhida, ou seja, se é uma rede com atraso alimentada adiante, do inglês *Feed-Forward Neural Network* (FFNN) ou rede neural de atraso de tempo, do inglês *Time-Delay Neural Network* (TDNN). Exemplos de classi-

ficação de distúrbios presentes em sinais elétricos podem ser verificados em (CHOONG; REAZ; MOHD, 2005; LESKOVICZ et al., 2009; CARPINELLI; CHIODO; LAURIA, 2007; LEE; DASH, 2003; SANTOSO et al., 2000; JANIK; LOBOS, 2006; HOANG; NGUYEN, 2006; REAZ et al., 2007a; GHOSH; LUBKEMAN, 1995)

Já o sistema de inferência *fuzzy* reduz a dificuldade de modelagem e análise de sistemas complexos, uma vez que os graus de pertinência atribuídos às variáveis fornecem suporte no tratamento da incerteza e da ambiguidade. Exemplos de aplicações de sistemas de inferência *fuzzy* voltadas para classificação de distúrbios presentes em sinais elétricos podem ser encontrados em (CHOONG; REAZ; MOHD, 2005; LIAO; LEE, 2004; TEKE; BAYINDIR; TUMAY, 2010; SAMANTARAY, 2010; REAZ et al., 2007b; DUAN et al., 2006; WIJAYAKULASOORIYA J.V., 2002)

Por fim, o classificador de Bayes possui um bom desempenho quando trabalha com funções densidade de probabilidade Gaussiana. Todavia, sabe-se que a função densidade de probabilidade de cada evento devem ser conhecida com antecedência. Não obstante, em muitas situações, o alto custo computacional do classificador de Bayes pode dificultar seu uso prático. Exemplos de aplicações de classificação de distúrbios baseadas na técnica em questão podem ser encontrados em (CHUNG et al., 2002b; CHOONG; REAZ; MOHD, 2005; BOLLEN et al., 2007; RIBEIRO; PEREIRA, 2007; KARIMI-GHARTEMANI; MOKHTARI; IRAVANI, 2000).

2.2 Classificação de distúrbios baseada em EOS

A figura 1 apresenta o paradigma padrão utilizado para a classificação de distúrbios presentes em sinais elétricos. O bloco “Extração de Características” é responsável pela extração de características (parâmetros), para que posteriormente possa haver a seleção de características. Finalmente, uma vez obtido o vetor de características \mathbf{p}_x , o bloco “Classificação” aplica uma das técnicas de classificação a serem apresentadas para decidir o tipo de distúrbio presente no vetor \mathbf{x} , que pode representar amostras de um sinal de tensão ou corrente elétrica.

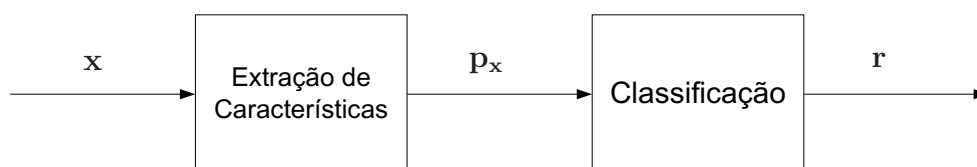


Figura 1: Técnica padrão para classificação de distúrbios presentes em sinais elétricos

Cada bloco e suas respectivas particularidades serão apresentados nas Seções a seguir.

2.2.1 Extração de características

Algumas contribuições como (MENDEL, 1991; NIKIAS; PETROPULU, 1993; FRISCH; MESSER, 1993; COLONNESE; SCARANO, 1999; NIKIAS; MENDEL, 1993; RIBEIRO et al., 2007; FERREIRA et al., 2009), direcionadas a problemas de detecção, classificação e identificação de distúrbios em sistemas elétricos, apresentaram resultados expressivos obtidos através do uso de estatísticas de ordem superior (EOS), do inglês *higher-order statistics*. Isto porque técnicas baseadas em EOS são mais adequadas à processos não gaussianos e a sistemas não lineares, se comparados aquelas que utilizam estatísticas de segunda ordem.

Dada uma sequência $\{z[n]\}$, tal que $E\{z[n]\} = 0$, os cumulantes de segunda, terceira e quarta ordens podem ser calculados a partir de (MENDEL, 1991)

$$c_{2,z}[i] = E\{z[n]z[n+i]\}, \quad (2.1)$$

$$c_{3,z}[i] = E\{z[n]z^2[n+i]\}, \quad (2.2)$$

e

$$c_{4,z}[i] = E\{z[n]z^3[n+i]\} - 3c_{2,z}[i]c_{2,z}[0], \quad (2.3)$$

em que $E\{\cdot\}$ denota o operador valor esperado e i representa o i -ésimo atraso.

Seja agora $\{z[n]\}$ uma sequência de comprimento L finito. Logo, as Equações (2.1) - (2.3) podem ser estocasticamente aproximadas pelas expressões

$$\hat{c}_{2,z}[i] \cong \frac{2}{L} \sum_{n=0}^{L/2-1} z[n]z[n+i], \quad (2.4)$$

$$\hat{c}_{3,z}[i] \cong \frac{2}{L} \sum_{n=0}^{L/2-1} z[n]z^2[n+i], \quad (2.5)$$

e

$$\begin{aligned} \hat{c}_{4,z}[i] &\cong \frac{2}{L} \sum_{n=0}^{L/2-1} z[n]z^3[n+i] - \\ &\quad - \frac{12}{L^2} \sum_{n=0}^{L/2-1} z[n]z[n+i] \sum_{n=0}^{L/2-1} z^2[n], \end{aligned} \quad (2.6)$$

em que $i = 0, 1, \dots, L/2 - 1$.

Uma forma alternativa para o cálculo dos cumulantes pode ser expressa por (RIBEIRO et al., 2007)

$$\tilde{c}_{2,z}[i] \cong \frac{1}{L} \sum_{n=0}^{L-1} z[n]z[\text{mod}(n+i, L)] \quad (2.7)$$

$$\tilde{c}_{3,z}[i] \cong \frac{1}{L} \sum_{n=0}^{L-1} z[n]z^2[\text{mod}(n+i, L)] \quad (2.8)$$

e

$$\begin{aligned} \tilde{c}_{4,z}[i] &\cong \frac{1}{L} \sum_{n=0}^{L-1} z[n]z^3[\text{mod}(n+i, L)] \\ &\quad - \frac{3}{L^2} \sum_{n=0}^{L-1} z[n]z[\text{mod}(n+i, L)] \sum_{n=0}^{L-1} z^2[n], \end{aligned} \quad (2.9)$$

em que $i = 0, 1, \dots, L - 1$ e $\text{mod}(\cdot)$ o operador módulo dado por

$$\text{mod}(n+i, L) = (n+i) - \varrho L, \quad \forall n+i > L-1, \quad (2.10)$$

em que ϱ denota o numero inteiro a que corresponde o resultado da divisão de $n+i$ por L , desprezando-se os valores após a vírgula.

As Equações (2.4) - (2.6) fornecem $L/2$ características, enquanto que aplicando as Equações (2.7) - (2.9), L características são extraídas. Isto porque com o uso do operador $\text{mod}(\cdot)$ é como se a sequência $\{z[n]\}$ fosse periódica, possibilitando assim a extração de mais informações da mesma. Logo, as expressões (2.7) - (2.9) mostram-se mais interessantes, pois resultam em um número maior e mais representativo de características.

Então, para cada \mathbf{x} obtém-se um vetor de características dado por

$$\mathbf{p}_{\mathcal{H}} = [\hat{\mathbf{c}}_{2,z}^T \tilde{\mathbf{c}}_{2,z}^T \hat{\mathbf{c}}_{3,z}^T \tilde{\mathbf{c}}_{3,z}^T \hat{\mathbf{c}}_{4,z}^T \tilde{\mathbf{c}}_{4,z}^T]^T, \quad \mathcal{H} = 0, 1, \quad (2.11)$$

em que $\mathcal{H} = 0$ denota a classe sem distúrbio, enquanto que $\mathcal{H} = 1$ a classe com

distúrbio, $\hat{\mathbf{c}}_{2,z} = [\hat{c}_{2,z}(0), \dots, \hat{c}_{2,z}(L/2 - 1)]^T$, $\tilde{\mathbf{c}}_{2,z} = [\tilde{c}_{2,z}(0), \dots, \tilde{c}_{2,z}(L - 1)]^T$, $\hat{\mathbf{c}}_{3,z} = [\hat{c}_{3,z}(0), \dots, \hat{c}_{3,z}(L/2 - 1)]^T$, $\tilde{\mathbf{c}}_{3,z} = [\tilde{c}_{3,z}(0), \dots, \tilde{c}_{3,z}(L - 1)]^T$, $\hat{\mathbf{c}}_{4,z} = [\hat{c}_{4,z}(0), \dots, \hat{c}_{4,z}(L/2 - 1)]^T$ e $\tilde{\mathbf{c}}_{4,z} = [\tilde{c}_{4,z}(0), \dots, \tilde{c}_{4,z}(L - 1)]^T$.

Dada a quantidade de características extraídas, a Seção 2.2.2 descreve a técnica de seleção de características usada neste trabalho.

2.2.2 Seleção de características

A seleção de características tem por objetivo indicar os K_p parâmetros mais significativos do vetor $\mathbf{p}_{\mathcal{H}}$, de comprimento $\frac{9L}{2}$. Em outras palavras, determina as características que promovem uma melhor separação entre as classes de eventos definidos pelas hipóteses a serem definidas no capítulo 3. Isto faz com que a complexidade do detector seja bastante reduzida, o que torna mais maleável o problema de detecção. Esta etapa é realizada apenas durante o processo de projeto dos detectores.

Neste trabalho foi utilizado o discriminante de Fisher, do inglês *Fisher's discriminant ratio* (FDR), por ser simples e ao mesmo tempo ter proporcionado resultados satisfatórios. Seu cálculo, para um problema envolvendo apenas duas classes distintas, é definido em (THEODORIDIS; KOUTROUMBAS, 1999) por

$$\mathbf{F}_{FDR} = \Lambda_{\mu_0, \mu_1} \Lambda_{\sigma}^{-1}, \quad (2.12)$$

em que $\Lambda_{\sigma} = \text{diag}\{\sigma_{0,0}^2 + \sigma_{1,0}^2, \sigma_{0,1}^2 + \sigma_{1,1}^2, \dots, \sigma_{0, \frac{9L}{2}-1}^2 + \sigma_{1, \frac{9L}{2}-1}^2\}$ é uma matriz diagonal formada pelo vetor de covariância associado a cada classe, enquanto que $\Lambda_{\mu_0, \mu_1} = \text{diag}\{(\mu_{0,0} - \mu_{1,0})^2, (\mu_{0,1} - \mu_{1,1})^2, \dots, (\mu_{0, \frac{9L}{2}-1} - \mu_{1, \frac{9L}{2}-1})^2\}$ é a matriz diagonal formada por seus vetores de média.

Seja $\mathbf{v}_{FDR} \in \mathbb{R}^{\frac{9L}{2} \times 1}$ um vetor constituído pelos elementos da diagonal principal de \mathbf{F}_{FDR} , tal que $v_{FDR}(0) \geq v_{FDR}(1) \geq \dots \geq v_{FDR}(\frac{9L}{2} - 1)$, então as K_p características selecionadas correspondem aos K_p primeiros elementos do vetor \mathbf{v}_{FDR} .

2.2.3 Classificadores

Neste trabalho serão avaliados o desempenho de dois classificadores: Um denominado classificador de Bayes (THEODORIDIS; KOUTROUMBAS, 1999), baseado no critério da máxima verossimilhança - *maximum likelihood* (ML) e outro baseado em uma rede neural MLP (HAYKIN, 1999). Ainda nesta contribuição, no que tange o classificador *fuzzy*, será apresentada a derivação de algoritmos de treinamento baseados no método do Gradiente

Conjugado e em informações de segunda ordem (matriz Hessiana).

A motivação do classificador baseado na regra de Bayes reside no fato de que o mesmo gera resultados que, de certa forma, balizam do ponto de vista comparativo, o projeto de novos classificadores, os quais podem se basear em outras teorias, tais como redes neurais e sistemas de inferência *fuzzy*. De fato, se as suposições consideradas no projeto de um classificador de Bayes não forem suficientes para fornecer bons resultados, há uma indicação da necessidade de novas técnicas de classificação que possam apresentar melhores desempenhos. Neste contexto, a escolha do classificador de Bayes, com as suposições descritas a seguir, visa tão somente explicar as possíveis limitações para esta técnica no que tange a classificação distúrbios presentes em sinais elétricos.

Por outro lado, a escolha da rede neural como classificador deve-se, sobretudo à sua capacidade de descrever superfícies complexas de separação, o suficiente para permitir desempenhos interessantes, mesmo quando os dados são de medição. A maioria dos algoritmos de treinamento de redes neurais se baseia em informações de primeira ordem. Além disso, o uso de algoritmos de treinamento, baseados em informações de segunda ordem, conduzem a resultados bastantes promissores quando a rede neural é submetida a exaustivos períodos de treinamento.

É importante ressaltar que as escolhas dos classificadores, baseados na regra de Bayes e na rede neural, não são as únicas e, possivelmente, não são as melhores de todas as escolhas possíveis. Entretanto, os resultados descritos no Capítulo 6 refletem, de certa forma, o nível de adequação das mesmas para o problema analisado.

2.2.3.1 Classificador de Bayes

Considere um vetor $\mathbf{x}_p \in \mathbb{R}^{K_p \times 1}$ a ser classificado, dentre as duas classes ou hipóteses \mathcal{H}_0 e \mathcal{H}_1 , formado pelos K_p parâmetros de $\mathbf{p}_{\mathcal{H}}$, selecionados a partir da Equação (2.12). Este vetor \mathbf{x}_p tem probabilidade *a priori*, de ser classificado em umas das duas classes, dada por $P(\mathcal{H}_0)$ e $P(\mathcal{H}_1)$. A função densidade de probabilidade condicional é denotada por $p(\mathbf{x}_p|\mathcal{H}_0)$ e $p(\mathbf{x}_p|\mathcal{H}_1)$. Então, de acordo com (THEODORIDIS; KOUTROUMBAS, 1999), a regra de Bayes fornece

$$P(\mathcal{H}_0|\mathbf{x}_p) = \frac{p(\mathbf{x}_p|\mathcal{H}_0)P(\mathcal{H}_0)}{p(\mathbf{x}_p)} \quad (2.13)$$

e

$$P(\mathcal{H}_1|\mathbf{x}_p) = \frac{p(\mathbf{x}_p|\mathcal{H}_1)P(\mathcal{H}_1)}{p(\mathbf{x}_p)}, \quad (2.14)$$

em que $p(\mathbf{x}_p)$ é a função densidade de probabilidade de \mathbf{x}_p , $P(\mathcal{H}_0|\mathbf{x}_p)$ e $P(\mathcal{H}_1|\mathbf{x}_p)$ são as probabilidades à posteriori de inexistência e ocorrência de distúrbios no vetor \mathbf{x}_p .

Portanto, aplicando a regra de Bayes em um vetor \mathbf{x}_p de características, tem-se:

Se $P(\mathcal{H}_0|\mathbf{x}_p) > P(\mathcal{H}_1|\mathbf{x}_p)$, então o vetor \mathbf{x}_p pertence à classe \mathcal{H}_0 ,

Se $P(\mathcal{H}_0|\mathbf{x}_p) \leq P(\mathcal{H}_1|\mathbf{x}_p)$, então o vetor \mathbf{x}_p pertence à classe \mathcal{H}_1 .

Considerando as Equações (2.13) e (2.14), a decisão baseada na regra de Bayes pode ser expressa como

$$p(\mathbf{x}_p|\mathcal{H}_0)P(\mathcal{H}_0) \geq p(\mathbf{x}_p|\mathcal{H}_1)P(\mathcal{H}_1). \quad (2.15)$$

Assuma agora que as probabilidades *a priori* de ocorrência de ambas as classes são iguais, ou seja, $P(\mathcal{H}_0) = P(\mathcal{H}_1) = 1/2$ e, que a função densidade de probabilidade segue uma distribuição Gaussiana. Assim tem-se:

$$p(\mathbf{x}_p|\mathcal{H}_\mathcal{H}) = \frac{1}{(2\pi)^{k_p/2} |\Sigma_\mathcal{H}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}_p - \boldsymbol{\mu}_\mathcal{H})^T \Sigma_\mathcal{H}^{-1} (\mathbf{x}_p - \boldsymbol{\mu}_\mathcal{H})}, \quad \mathcal{H} = 0, 1, \quad (2.16)$$

em que $\boldsymbol{\mu}_\mathcal{H}$ representa o vetor de médias, $|\cdot|$ é o operador determinante e $\Sigma_\mathcal{H}$ é a matriz de covariância referente a cada classe expressa por

$$\Sigma_\mathcal{H} = E\{(\mathbf{x}_p - \boldsymbol{\mu}_\mathcal{H})(\mathbf{x}_p - \boldsymbol{\mu}_\mathcal{H})^T\}, \quad \mathcal{H} = 0, 1. \quad (2.17)$$

Aplicando as considerações anteriores na expressão (2.15), resulta em

$$\frac{|\Sigma_0|^{1/2} e^{-\frac{1}{2}(\mathbf{x}_p - \boldsymbol{\mu}_1)^T \Sigma_1^{-1} (\mathbf{x}_p - \boldsymbol{\mu}_1)}}{|\Sigma_1|^{1/2} e^{-\frac{1}{2}(\mathbf{x}_p - \boldsymbol{\mu}_0)^T \Sigma_0^{-1} (\mathbf{x}_p - \boldsymbol{\mu}_0)}} \geq 1 \quad (2.18)$$

Observa-se em (2.18), que o classificador proposto baseia-se no critério da máxima verossimilhança e que, dado um vetor \mathbf{x}_p , a determinação de qual classe o mesmo corresponde, depende do resultado desta inequação.

2.2.3.2 Classificador MLP

Dados os vetores $\mathbf{x}_{p,i}$, com $i = 1, 2, \dots, N_x$ amostras, constituídos de K_p características extraídas dos vetores $\mathbf{p}_{x,i}$, então as equações de uma rede neural perceptron multicamadas, com K_p entradas, uma camada escondida e camada de saída com uma

saída, são as seguintes

$$\mathbf{s}_i = \mathbf{A}^T \begin{bmatrix} \mathbf{x}_{p,i} \\ 1 \end{bmatrix} \quad (2.19)$$

$$\mathbf{q}_i = \varphi(\mathbf{s}_i), \quad (2.20)$$

e

$$y_{rn,i} = \mathbf{B}^T \begin{bmatrix} \mathbf{q}_i \\ 1 \end{bmatrix}, \quad (2.21)$$

em que $\mathbf{A} \in \mathbb{R}^{(K_p+1) \times N_q}$ e $\mathbf{B} \in \mathbb{R}^{(K_p+1) \times 1}$ são as matrizes de pesos entre as camadas de entrada e escondida/intermediária e entre as camadas escondida/intermediária e de saída, respectivamente; $y_{rn,i}$ é a saída da rede neural MLP associada ao vetor $\mathbf{x}_{p,i}$.

Por sua vez, $\varphi(\cdot)$ é uma função de transferência monotonamente crescente e que neste trabalho será tomada como sendo a tangente hiperbólica. Com isso, a equação (2.20) pode ser reescrita da seguinte maneira:

$$\mathbf{q}_i = \tanh(\mathbf{s}_i). \quad (2.22)$$

Concatenando os vetores colunas das matrizes \mathbf{A} e \mathbf{B} , obtém-se o vetor de pesos dado por

$$\mathbf{w} = [\mathbf{a}^T \ \mathbf{b}^T]^T, \quad (2.23)$$

em que \mathbf{a}^T e \mathbf{b}^T é a concatenação dos vetores colunas das matrizes \mathbf{A} e \mathbf{B} , respectivamente. O vetor \mathbf{w} ótimo (\mathbf{w}_o) pode ser obtido por

$$\mathbf{w}_o = \min_{\mathbf{w}} J(\mathbf{w}), \quad (2.24)$$

em que

$$J(\mathbf{w}) = \frac{1}{2N_x} \sum_{i=1}^{N_x} (y_{rn}(i) - y_d(i))^2 \quad (2.25)$$

é a função custo a ser minimizada e $y_d(i)$ é a i -ésima saída desejada da rede neural MLP. Note que $y_d(i) = 1$ significa que o vetor de parâmetros $\mathbf{x}_{p,i}$ está associado a ocorrência do distúrbio. Por outro lado, $y_d(i) = -1$ informa que o vetor de parâmetros $\mathbf{x}_{p,i}$ está associado à inexistência de distúrbio.

Dentre os vários métodos de otimização disponíveis na literatura para a obtenção de \mathbf{w}_o , neste trabalho optou-se pelo método gradiente conjugado, escalonado e modificado (GCEM) (MEDSKER; JAIN, 2000). Este método foi adotado, posto que ele faz uso de informações de segunda ordem e apresenta baixo custo computacional devido ao uso da

técnica proposta em (PEARLMUTTER, 1994). Assim sendo, tem-se um método de treinamento que apresenta elevada velocidade de convergência e baixo esforço computacional, com relação aos algoritmos baseados em informação de segunda ordem. Uma descrição detalhada do algoritmo do método GCEM para treinamento da rede neural MLP é apresentado em (RIBEIRO, 2005), a qual foi proposto em (SANTOS; ZUBEN, 2000).

2.3 Sumário

O presente capítulo apresentou as técnicas e metodologias utilizadas para classificação de distúrbios presentes em sinais elétricos. Note que as referidas técnicas fazem uso de EOS para a extração de características e FDR para realizar a seleção de características. Tais técnicas foram apresentadas na Seção 2.2.

Por fim, tem-se a apresentação das considerações mais relevantes acerca das abordagens baseados no teorema de Bayes e rede neural MLP para a classificação de distúrbios da qualidade da energia elétrica.

A seguir, o capítulo 3 apresenta a formulação do problema de classificação de distúrbios em sinais elétricos e discute as questões de investigação analisadas na presente contribuição.

3 FORMULAÇÃO DO PROBLEMA

O presente capítulo apresenta a formulação do problema de classificação de distúrbios da QEE. Primeiramente, discute-se a decomposição do sinal a ser monitorado em diversas componentes e a seguir, apresenta-se o problema de classificação como a decisão entre hipóteses que serão apresentadas.

Considerando os distúrbios em tensão e/ou corrente de acordo com os conceitos de QEE, também chamados de eventos ou variações (BOLLEN; GU, 2006), o sinal elétrico (tensão ou corrente monitorada) pode ser decomposto em contribuições aditivas de vários tipos de sinais primitivos, os quais são associados aos fenômenos elétricos da QEE.

Seja $x(t) \in \mathbb{R}$, $-\infty < t < \infty$, um sinal de tensão e/ou corrente de um sistema elétrico de distribuição ou transmissão de energia elétrica, tal que

$$X(\Omega) = \begin{cases} X_R(\Omega) + jX_I(\Omega), & \Omega \leq \Omega_{\max} \\ 0, & \text{caso contrário} \end{cases}, \quad (3.1)$$

em que $X(\Omega)$ é a transformada de *Fourier* de $x(t)$, posto que

$$\int_{-\infty}^{\infty} |x(t)| dt < \infty \quad (3.2)$$

e Ω_{\max} é a frequência angular máxima de $X(\Omega)$. Assim sendo, o sinal em tempo discreto é definido como

$$x[n] := x(t)|_{t=nT}, \quad (3.3)$$

em que

$$\begin{aligned} T &= \frac{\pi}{\Omega_{\max}} \\ &= \frac{1}{2f_{\max}} \\ &= \frac{1}{f_s} \end{aligned} \quad (3.4)$$

é o período de amostragem que atende ao critério de Nyquist e f_s é a frequência de amostragem (amostras por segundo).

O sinal $x[n]$ pode ser expresso por (RIBEIRO, 2005)

$$x[n] := f[n] + h[n] + i[n] + t[n] + v[n], \quad (3.5)$$

em que os sinais $f[n]$, $h[n]$, $i[n]$, $t[n]$ e $v[n]$ representam a componente fundamental, harmônicos, interharmônicos, transitórios e ruídos de fundo, respectivamente. A componente fundamental é representada por

$$f[n] := A_0[n] \cos\left(2\pi \frac{f_0[n]}{f_s} n + \phi_0[n]\right), \quad (3.6)$$

em que $A_0[n]$, $f_0[n]$ e $\phi_0[n]$, referem-se, respectivamente à amplitude, frequência e fase da componente fundamental. Observe que os parâmetros da componente fundamental variam com o tempo. Distúrbios associados a esta componente são: *sags* e *swells* instantâneos, momentâneos e temporários; interrupções momentâneas, temporárias e sustentadas; e sobretensões e subtensões (QUALITY, 1995).

A representação da componente de harmônicos é dada por

$$h[n] := \sum_{m=1}^M h_m[n], \quad (3.7)$$

em que

$$h_m[n] := A_m[n] \cos\left(2\pi m \frac{f_0[n]}{f_s} n + \phi_m[n]\right) (u[n - n_{h_{mi}}] - u[n - n_{h_{mf}}]). \quad (3.8)$$

Note que $h_m[n]$ representa a m -ésima harmônica com amplitude $A_m[n]$ e fase $\phi_m[n]$. O termo $(u[n - n_{h_{mi}}] - u[n - n_{h_{mf}}])$ da equação (3.8) representa a duração do sinal, em que $n_{h_{mi}}$ define a amostra do sinal onde se inicia o componente harmônico e $n_{h_{mf}}$ define a amostra onde tal distúrbio termina e $u[n]$ é a sequência degrau unitário.

A componente de interharmônicos é definida por

$$i[n] := \sum_{j=1}^J i_j[n], \quad (3.9)$$

em que

$$i_j[n] := A_j[n] \cos\left(2\pi \frac{f_j[n]}{f_s} n + \phi_j[n]\right) (u[n - n_{i_{ji}}] - u[n - n_{i_{jf}}]), \quad (3.10)$$

no qual $A_j(n)$, $f_j(n)$ e $\phi_j(n)$ são a magnitude, frequência e a fase da j -ésima interharmônica, respectivamente. Por sua vez, o termo $(u[n - n_{i_{ji}}] - u[n - n_{i_{jf}}])$ da equação (3.10) representa a duração do sinal, em que $n_{i_{ji}}$ define a amostra do sinal onde se inicia

o componente harmônico e $n_{i,jf}$ define a amostra onde tal distúrbio termina.

A componente de transitórios é dada por:

$$t[n] := t_{spi}[n] + t_{not}[n] + t_{dec}[n] + t_{dam}[n], \quad (3.11)$$

em que $t_{spi}[n]$ e $t_{not}[n]$ representam transitórios chamados *spikes* e *notches*, respectivamente e $t_{dec}[n]$ e $t_{dam}[n]$ representam transitórios oscilatórios e transitórios impulsivos respectivamente. Tais fenômenos são expressos, respectivamente, pelas equações:

$$t_{spi}[n] := \sum_{i=1}^{N_{spi}} t_{spi,i}[n] (u[n - n_{spi,i}] - u[n - n_{spi,f}]), \quad (3.12)$$

$$t_{not}[n] := \sum_{i=1}^{N_{not}} t_{not,i}[n] (u[n - n_{not,i}] - u[n - n_{not,f}]), \quad (3.13)$$

$$t_{dec}[n] := \sum_{i=1}^{N_{dec}} A_{dec,i}[n] \cos[\omega_{dec,i}[n]n + \phi_{dec,i}[n]] \exp[-\alpha_{dec,i}[n - n_{dec,i}]] \times [u[n - n_{dec,i}] - u[n - n_{dec,f}]] \quad (3.14)$$

e

$$t_{dam}[n] := \sum_{i=1}^{N_{dam}} A_{dam,i}[n] \exp[-\alpha_{dam,i}[n - n_{dam,i}]] \times [u[n - n_{dam,i}] - u[n - n_{dam,f}]], \quad (3.15)$$

em que o termo $(u[n - n_{spi,i}] - u[n - n_{spi,f}])$ da equação (3.12) representa a duração do sinal, em que $n_{spi,i}$ define a amostra do sinal onde se inicia o distúrbio *spike* e $n_{spi,f}$ define a amostra onde tal distúrbio termina. Por sua vez, o termo $(u[n - n_{not,i}] - u[n - n_{not,f}])$ da equação (3.13) representa a duração do sinal, em que $n_{not,i}$ define a amostra do sinal onde se inicia o distúrbio *notch* e $n_{not,f}$ define a amostra onde o mesmo termina. O termo $[u[n - n_{dec,i}] - u[n - n_{dec,f}]]$ da equação (3.14) representa a duração do sinal, em que $n_{dec,i}$ define a amostra do sinal onde se inicia o transitório oscilatório e $n_{dec,f}$ define a amostra onde tal distúrbio termina. Já o termo $[u[n - n_{dam,i}] - u[n - n_{dam,f}]]$ da equação (3.15) representa a duração do sinal, em que $n_{dam,i}$ define a amostra do sinal onde se inicia a transitório impulsivo e $n_{dam,f}$ define a amostra onde o referido distúrbio termina. Observe que (3.14) se refere tanto a transitórios causados por chaveamentos de capacitor quanto por faltas em sistemas de energia elétrica. A equação (3.15) representa tanto decaimentos exponenciais quanto componentes CC ($\alpha_{dam} = 0$).

O ruído de fundo $\{v[n]\}$ é constituído de variáveis aleatórias independente e identicamente distribuídas (i.i.d.) e independente das componentes $\{f[n]\}$, $\{h[n]\}$, $\{i[n]\}$ e $\{t[n]\}$. Assume-se que $v(n) \sim \mathcal{N}(0, \sigma^2)$, em que $\mathcal{N}(0, \sigma^2)$ define uma distribuição

Gaussiana de média zero e variância σ^2 .

A partir das expressões (3.6)-(3.15), obtêm-se os vetores $\mathbf{x} = [x[n] \dots x[n - N + 1]]^T$, $\mathbf{f} = [f[n] \dots f[n - N + 1]]^T$, $\mathbf{h} = [h[n] \dots h[n - N + 1]]^T$, $\mathbf{i} = [i[n] \dots i[n - N + 1]]^T$, $\mathbf{t} = [t[n] \dots t[n - N + 1]]^T$ e $\mathbf{v} = [v[n] \dots v[n - N + 1]]^T$.

Assim sendo, a classificação de eventos ou distúrbios que possam ocorrer no sinal \mathbf{x} pode ser resolvida de forma efetiva se for possível decompor o sinal \mathbf{x} nas suas componentes primitivas (RIBEIRO; PEREIRA, 2007). Entretanto, há uma grande dificuldade em obter as componentes primitivas do sinal \mathbf{x} , mesmo que seja de forma *offline*, posto que isso pode demandar uma complexidade (espacial e/ou temporal) considerável.

Para contornar este problema, utiliza-se o paradigma baseado no princípio de dividir para conquistar, do inglês *the principle of divide to conquer*, que tem sido amplamente e satisfatoriamente empregado em várias aplicações da engenharia e que pode ser visualizado em (RIBEIRO, 2005; RIBEIRO; PEREIRA, 2007). No referido paradigma, o vetor \mathbf{x} é decomposto em várias componentes primitivas, a partir das quais distúrbios isolados ou múltiplos, chamados de padrões primitivos, podem ser facilmente classificados. A idéia principal é decompor os sinais de tensão e/ou corrente em várias componentes primitivas com técnicas de processamento de sinais e aplicar técnicas de reconhecimento de padrões para classificar os distúrbios, conforme foi inicialmente discutido em (RIBEIRO; PEREIRA, 2007) e (RIBEIRO, 2005) e mostrado na figura 2.

Apesar dos bons resultados obtidos em (RIBEIRO; PEREIRA, 2007) e (RIBEIRO, 2005), pode-se notar que as técnicas apresentadas nessas contribuições demandam um considerável custo computacional para a decomposição do sinal elétrico, posto que o mesmo é decomposto nas componentes \mathbf{f} , \mathbf{h} e \mathbf{u} . Visando reduzir o custo computacional, nesta contribuição propõe-se a decomposição do vetor \mathbf{x} nos vetores \mathbf{f} e \mathbf{u} , conforme mostrado na figura 3. Isso significa que uma simples técnica de filtragem pode ser aplicada para obtenção dos vetores \mathbf{f} e \mathbf{u} .

O esquema da técnica proposta é retratado na figura 3. O bloco decomposição do sinal implementa um filtro *notch* IIR que é capaz de decompor o vetor \mathbf{x} nas componentes \mathbf{f} e \mathbf{u} . O bloco extração de parâmetros é responsável pela extração de características baseado em estatísticas de ordem superior. Finalmente, os blocos de classificação aplicam uma técnica de classificação baseada em sistemas de inferência *fuzzy* para decidir a existência de distúrbios no vetor \mathbf{x} , da mesma forma como foi feito em (RIBEIRO; PEREIRA, 2007) e

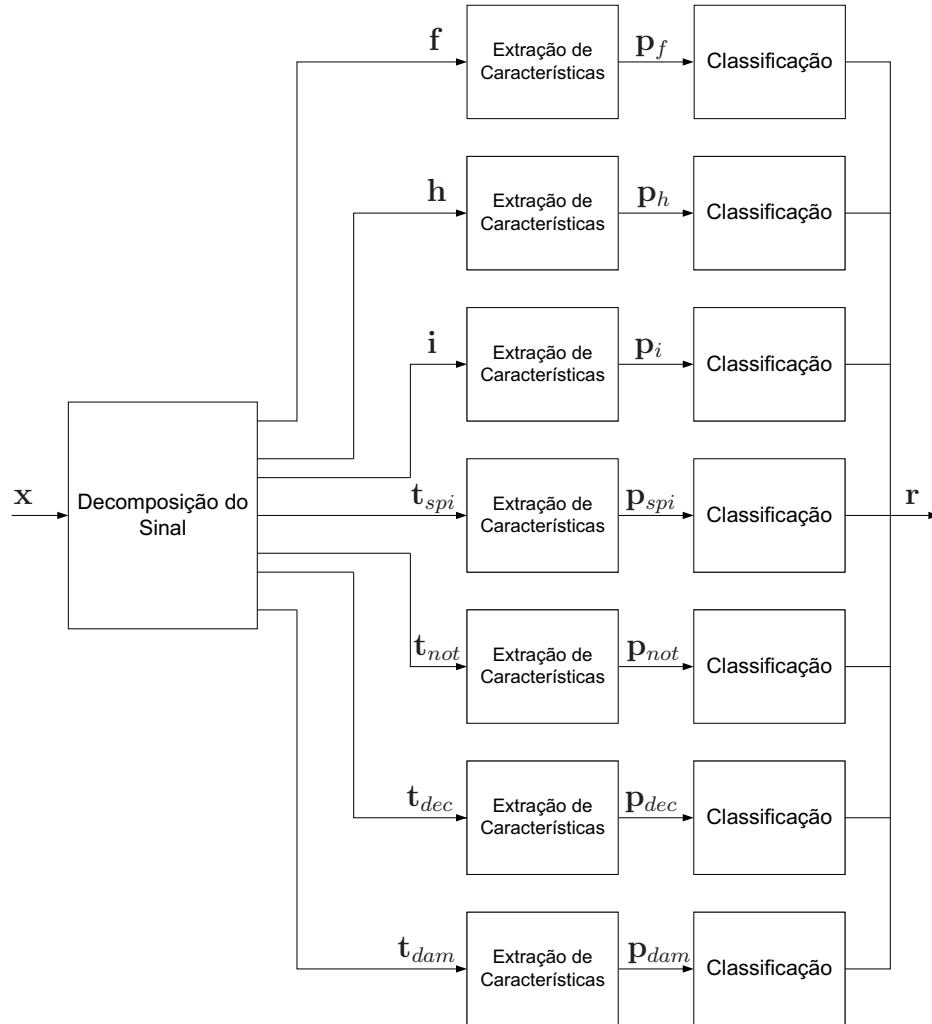


Figura 2: Paradigma apresentado em (RIBEIRO; PEREIRA, 2007) para a classificação de distúrbios isolados e múltiplos.

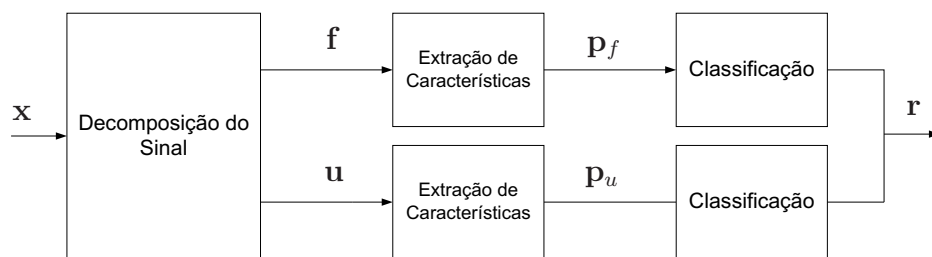


Figura 3: Paradigma proposto para a classificação de distúrbios isolados e múltiplos.

(RIBEIRO, 2005).

Para minimizar eventuais problemas relacionados à classificação de distúrbios, nesta dissertação analisa-se o vetor \mathbf{u} e o vetor \mathbf{f} . Assim sendo, o problema se resume em decompor o vetor \mathbf{x} em duas componentes e realizar a análise conjunta das informações presentes nestas componentes. Utilizando uma técnica simples e eficiente de filtragem, o vetor \mathbf{x} pode ser decomposto nos vetores \mathbf{f} e \mathbf{u} , de forma satisfatória para efeitos de classificação de distúrbios, e satisfazendo a seguinte equação:

$$\mathbf{u} := \mathbf{x} - \mathbf{f}, \quad (3.16)$$

o que pode ser obtido com uma técnica de filtragem simples para a obtenção dos vetores \mathbf{f} e \mathbf{u} .

A partir do vetor \mathbf{f} pode-se facilmente classificar vários conjuntos de distúrbios que estão fortemente vinculados à componente fundamental. Os padrões primitivos associados a essa componente primitiva são chamadas *sag*, *swell*, interrupção e *flicker*. Nesta dissertação, para o sistema de inferência *fuzzy* para classificação de distúrbios de QEE, iremos considerar apenas os distúrbios *sag*, *swell* e *flicker*, uma vez que com os mesmos podemos obter uma satisfatória abrangência do problema. Como resultado, a classificação de distúrbios nesta componente pode ser formulada como a decisão entre hipóteses simples relacionadas à ocorrência de cada um destes distúrbios, conforme mostrada a seguir (TREES, 1968, 1971), (MCDONOUGH; WHALEN, 1995), (RIBEIRO; PEREIRA, 2007):

$$\begin{aligned} \mathcal{H}_{f,0} : \mathbf{f} &= \mathbf{v}_f, \\ \mathcal{H}_{f,1} : \mathbf{f} &= \mathbf{f}_{sag} + \mathbf{v}_f, \\ \mathcal{H}_{f,2} : \mathbf{f} &= \mathbf{f}_{swell} + \mathbf{v}_f, \end{aligned} \quad (3.17)$$

em que \mathbf{v}_f é uma parcela do ruído de fundo presente na componente fundamental. Os vetores \mathbf{f}_{sag} e \mathbf{f}_{swell} denotam a *sag* ou subtensão e *swell* ou sobretensão, respectivamente. Note que as hipóteses expressas em (3.17) podem ser dividida em outras três hipóteses simples, que são expressas por

$$\begin{aligned} \mathcal{H}_{f,i,0} : \mathbf{f} &= \mathbf{v}_f, \\ \mathcal{H}_{f,i,1} : \mathbf{f} &= \mathbf{f}_{dist} + \mathbf{v}_f, \end{aligned} \quad (3.18)$$

em que *dist* denota *sag* e *swell* se $i = 1$ ou 2 , respectivamente.

Já a partir da componente \mathbf{u} , pode-se classificar distúrbios tais como harmônicas, interharmônicas, *spikes*, *notches*, transitórios oscilatórios advindos de chaveamentos

capacitivos e transitórios impulsivos (RIBEIRO, 2005). Nesta dissertação, consideraremos apenas harmônicas e interharmônicas, posto que com tais distúrbios podemos obter uma boa representatividade do problema. Com isso assumindo que seja possível obter o vetor \mathbf{u} , temos que

$$\mathbf{u} = \mathbf{h} + \mathbf{i} + \mathbf{v}, \quad (3.19)$$

em que \mathbf{v}_u é uma parcela do ruído de fundo presente na componente \mathbf{u} . Note que de (3.19), a classificação de distúrbios nesta componente pode ser formulada como a decisão entre hipóteses simples relacionadas à ocorrência de cada um destes distúrbios, conforme mostrada a seguir

$$\begin{aligned} \mathcal{H}_{u,1} : \mathbf{u} &= \mathbf{h} + \mathbf{v}_u, \\ \mathcal{H}_{u,2} : \mathbf{u} &= \mathbf{i} + \mathbf{v}_u. \end{aligned} \quad (3.20)$$

O vetor \mathbf{i} é relacionado à incidência de componentes interharmônicas nos sinais elétricos. Estes componentes estão associados à ocorrência de cintilação (ou *flicker*). Neste caso, o padrão primitivo associado a tal componente recebe o nome de interharmônica. Tal distúrbio pode ser decomposto em outros padrões primitivos se for necessário analisar um grupo específico de componentes interharmônicas. É válido ressaltar que o distúrbio *flicker* é uma classe específica de interharmônicas, conforme discutido em (GROUP, 1992) e será considerado nesta dissertação.

Assim, a classificação de distúrbios pode ser definida como a decisão entre hipóteses simples relacionadas à ocorrência destes distúrbios isolados ou múltiplos da seguinte maneira

$$\begin{aligned} \mathcal{H}_{x,1} : \mathbf{x} &= \mathbf{f}_{sag} + \mathbf{v}, \\ \mathcal{H}_{x,2} : \mathbf{x} &= \mathbf{f}_{swell} + \mathbf{v}, \\ \mathcal{H}_{x,3} : \mathbf{x} &= \mathbf{h} + \mathbf{v}, \\ \mathcal{H}_{x,4} : \mathbf{x} &= \mathbf{i} + \mathbf{v}, \\ \mathcal{H}_{x,5} : \mathbf{x} &= \mathbf{f}_{sag} + \mathbf{i} + \mathbf{v}, \\ \mathcal{H}_{x,6} : \mathbf{x} &= \mathbf{f}_{sag} + \mathbf{h} + \mathbf{v}, \\ \mathcal{H}_{x,7} : \mathbf{x} &= \mathbf{f}_{swell} + \mathbf{i} + \mathbf{v}, \\ \mathcal{H}_{x,8} : \mathbf{x} &= \mathbf{f}_{swell} + \mathbf{h} + \mathbf{v}, \\ \mathcal{H}_{x,9} : \mathbf{x} &= \mathbf{f}_{sag} + \mathbf{h} + \mathbf{i} + \mathbf{v}, \\ \mathcal{H}_{x,10} : \mathbf{x} &= \mathbf{f}_{swell} + \mathbf{h} + \mathbf{i} + \mathbf{v}. \end{aligned} \quad (3.21)$$

em que \mathbf{v} é uma parcela do ruído de fundo presente na componente \mathbf{x} .

Dada a formulação do problema discutida acima, as seguintes questões de inves-

tigaç o podem ser analisadas:

1. Qual   a t cnica de filtragem adequada (desempenho x complexidade) para decompor o sinal?
2. Quais e quantas s o as caracter sticas que devem ser extra das dos vetores \mathbf{f} e \mathbf{u} de forma a reduzir a dimensionalidade do problema de classifica o de dist rbios?
3. Qual   o crit rio de sele o de caracter sticas que garante uma redu o de dimensionalidade  tima?
4. Qual   a t cnica de classifica o adequada (desempenho x complexidade) para o problema em quest o?

Dentre tais quest es de investiga o, neste trabalho   dada especial aten o   quest o 4. Basicamente, no cap tulo 5   descrito e deduzido um sistema de infer ncia *fuzzy* treinado por algoritmos do gradiente conjugado, visando, sobretudo, mostrar que tal tipo de t cnica baseada em intelig ncia computacional pode ser uma candidata interessante para responder a quest o 4.

3.1 Sum rio

Nesta Se o foi apresentada a formula o do problema de classifica o mostrando a composi o do sinal a ser monitorado, assim como a formula o de hip teses proposta baseada na an lise dos vetores \mathbf{u} e \mathbf{f} . A seguir, no Cap tulo 4 discute-se, a partir da formula o do problema apresentado nesta Se o, os sistemas de infer ncia *fuzzy singleton* e *non-singleton* que ser o utilizados para classifica o de dist rbios de QEE.

4 SISTEMAS FUZZY

Criada em 1965 pelo Prof. Lotfi Zadeh, da Universidade da Califórnia, Berkeley, a lógica *fuzzy* (ou lógica nebulosa) desagradou muitos cientistas no início, pois tratava, de certa forma, de dados vagos ou imprecisos. Em outras palavras enquanto a lógica booleana prega os extremos “completamente verdadeiro” ou “completamente falso” para uma premissa, a lógica *fuzzy* permite uma transição gradual de uma proposição dentre os conjuntos em que esta pode pertencer e pela relação de pertinência entre esta em relação aos conjuntos analisados. Isto quer dizer que, em lógica *fuzzy* uma premissa varia em grau de verdade entre 0 e 1, gerando resultados “parcialmente verdadeiros” ou “parcialmente falsos”.

Para exemplificar esta análise de resultados intermediários ao verdadeiro ou falso, tomemos o seguinte: vamos assumir que o conceito de meia idade para o ser humano seja definido pelo período que vai de 35 a 55 anos. Desta forma, como apresentado na figura 4, temos limites rígidos para o conjunto de meia idade e isto não é ideal, pois depende de cada pessoa.

Uma representação para a flexibilização destes limites é apresentada na figura 5, na qual se nota que o grau de certeza que uma pessoa de 25 anos pertença ao grupo de “meia idade” é menor que a certeza de uma pessoa de 45 anos pertencer a este grupo.

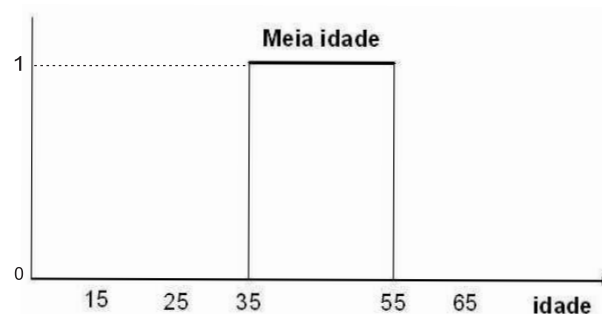


Figura 4: Comportamento booleano de uma situação.

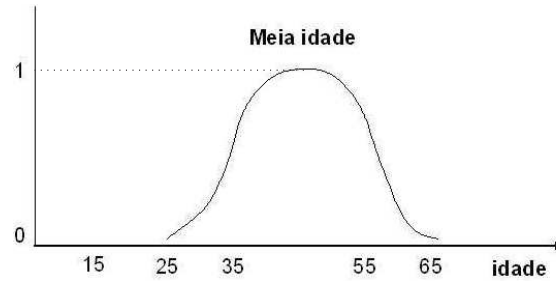


Figura 5: Flexibilização dos limites de uma situação.

O sistema de inferência *fuzzy* imita um comportamento baseado em regras, gerando saídas a partir de um conjunto de entradas não precisas. Para entendermos o raciocínio *fuzzy*, é necessária antes a apresentação de alguns componentes da teoria dos conjuntos *fuzzy*.

4.1 Conjuntos *fuzzy*

Sabemos que um conjunto *fuzzy* é um conjunto em que os limites não são bem definidos, ou seja, a transição entre pertencer e não pertencer ao conjunto é gradual ao invés de abrupta. Um conjunto *fuzzy* F em um universo de discurso X é caracterizado por uma função de pertinência F que associa a cada elemento $x \in X$ um grau $F(x) \in [0, 1]$.

Um conjunto clássico \mathcal{A} em X é um caso particular de um conjunto *fuzzy*, em que os graus de associação (ou graus de pertinência) são definidos conforme a seguir:

$$\mathcal{A}(x) = \begin{cases} 1, & \text{se } x \in \mathcal{A} \\ 0, & \text{se } x \notin \mathcal{A} \end{cases} \quad (4.1)$$

No caso de conjuntos clássicos, $\mathcal{A}(x)$ é a função característica de \mathcal{A} .

Seja X um universo de discurso, ou seja, uma coleção discreta ou contínua de objetos x . Um conjunto *fuzzy* F em X é um conjunto de pares ordenados da forma

$$F = \{(x, F(x)) \mid x \in X\} \quad (4.2)$$

em que $F(x)$ é o grau de pertinência do elemento x no conjunto *fuzzy* F . Note que F mapeia um elemento x de X no intervalo $[0, 1]$, ou seja, $F : X \rightarrow [0, 1]$. Adicionalmente, funções de pertinência para conjuntos *fuzzy* generalizam a idéia de funções características

para conjuntos clássicos. Se o universo $X = \{x_1, x_2, \dots, x_n\}$ é finito, então F pode ser representada por um vetor n -dimensional $F = (f_1, f_2, \dots, f_n)$, com $f_i = F(x_i)$

4.2 Normas triangulares

As normas triangulares formam uma classe geral de operadores de união e interseção para a teoria de conjuntos *fuzzy*. Estas normas fornecem um modelo operacional para os conectivos lógicos E (*and*) e OU (*or*) na lógica proposicional clássica.

Diferente da união e interseção, que trabalham com conjuntos definidos em um mesmo universo, as operações baseadas em normas triangulares podem operar em conjuntos de universos distintos. Na teoria de conjuntos clássicos, a interseção e a união são executadas, em termos de suas funções características, utilizando-se os operadores *min* e *max*, respectivamente.

Por definição, uma norma triangular é uma operação binária t no intervalo $[0, 1]$, que é comutativa, associativa, monotônica e que apresenta 1 como elemento neutro. Com isso, sabemos que que uma t-norma é uma função $t : [0, 1] \times [0, 1] \rightarrow [0, 1]$, tal que para todo o $x, y, z \in [0, 1]$ aplica-se a seguintes propriedades:

- comutatividade: $x t y = y t x$,
- associatividade: $x t (y t z) = (x t y) t z$,
- monotonicidade: $x t y \leq x t z$ quando $y \leq z$,
- elemento neutro: $x t 1 = x$.

Dentre as t-normas mais utilizadas estão:

- interseção: $x t y = \min(x, y)$,
- produto algébrico: $x t y = xy$,
- produto limite: $x t y = \max\{0, x + y - 1\}$,
- produto drástico: $x t y = \begin{cases} x & y = 1 \\ y & x = 1 \\ 0 & x, y < 1 \end{cases}$.

Adicionalmente, uma notação alternativa utilizada para t-normas é dada por $T(x, y) = x \ t \ y$.

Já uma co-norma triangular ou (s-norma) é uma operação binária s no intervalo $[0, 1]$, que é comutativa, associativa, monotônica e que apresenta 0 como elemento neutro. Com isso, sabemos que que uma t-norma é uma função $s : [0, 1] \times [0, 1] \rightarrow [0, 1]$, tal que para todo o $x, y, z \in [0, 1]$:

- comutatividade: $x \ s \ y = y \ s \ x$,
- associatividade: $x \ s \ (y \ t \ z) = (x \ s \ y) \ s \ z$,
- monotonicidade: $x \ s \ y \leq x \ s \ z$ quando $y \leq z$,
- elemento neutro 0: $x \ s \ 0 = x$.

Dentre as s-normas mais utilizadas estão:

- união: $x \ s \ y = \max(x, y)$,
- soma algébrica: $x \ s \ y = x + y - xy$,
- soma limite: $x \ s \ y = \min\{0, x + y - 1\}$,
- soma drástica: $x \ s \ y = \begin{cases} x & y = 0 \\ y & x = 0 \\ 0 & x, y > 0 \end{cases}$.

Por sua vez, uma notação alternativa utilizada para t-normas e s-normas são dadas respectivamente por $T(x, y) = x \ t \ y$ e $S(x, y) = x \ s \ y$ e podem ser encontradas em (PEDRYCZ; GOMIDE, 2007).

4.3 Relações *fuzzy*

Uma relação *fuzzy* R expande a noção de conjuntos *fuzzy* para um universo multi-dimensional e representa a noção de associação parcial entre os elementos dos universos. Por exemplo, um grau de pertinência $R(x, y)$ representa o grau de associação entre o elemento $x \in X$ e $y \in Y$ (PEDRYCZ; GOMIDE, 2007).

Sendo as relações *fuzzy* generalizações das relações tradicionais, elas mapeiam elementos de um universo para o outro através do produto cartesiano dos universos. O grau da relação é definido no intervalo unitário $[0, 1]$.

Com isso, sejam X e Y dois universos. Uma relação *fuzzy* R é um subconjunto *fuzzy* do produto cartesiano $X \times Y$

$$R : X \times Y \rightarrow [0, 1], \quad (4.3)$$

em que $x \in X$, $y \in Y$ e $R(x, y)$ é o grau com que (x, y) são compatíveis com relação a R .

Sabemos que as relações entre variáveis podem ser definidas usando regras *fuzzy* na forma:

$$\text{Se } x_1 \text{ é } F_1 \text{ e } x_2 \text{ é } F_2 \text{ e } \dots x_n \text{ é } F_n, \text{ Então } z \text{ é } Z_l. \quad (4.4)$$

Regras *fuzzy* definem uma relação entre antecedentes (x_1, x_2, \dots, x_n) e consequentes antecedentes (z) em termos dos conjuntos *fuzzy* F_1, F_2, \dots, F_n e Z_l . Uma regra *fuzzy* 4.4 é um exemplo de relação *fuzzy* da forma $R : F_1 \times F_2 \times \dots F_n \times Z_l \rightarrow [0, 1]$, em que $R(x_1, x_2, \dots, z)$ representa o grau de associação entre as variáveis antecedentes e consequentes.

4.4 Composição de relações *fuzzy*

Dentre os mais importantes tipos de composição de relações *fuzzy*, podemos citar as composições denominadas sup-t e inf-s. Supondo que R, S e T são relações *fuzzy* no espaço cartesiano $X \times Y$, $Y \times Z$ e $X \times Z$ respectivamente. Tais relações podem ser combinadas utilizando diferentes operadores de composição. Para definir tais tipos de composição, considere X, Y e Z como universos distintos.

A composição sup-t das relações *fuzzy* $R : X \times Y \rightarrow [0, 1]$ e $S : Y \times Z \rightarrow [0, 1]$ é uma relação *fuzzy* $T : R \times S \rightarrow [0, 1]$ cuja função de pertinência $T(x, z)$ é dada por:

$$T(x, z) = \sup_{y \in Y} \{R(x, y) \ t \ S(y, z)\} \quad \forall x \in X \quad e \quad \forall z \in Z, \quad (4.5)$$

em que t é uma t-norma. Note que a composição sup-t de R e S é denotada por $T = R \circ S$.

Por sua vez, a composição inf-s das relações *fuzzy* $R : X \times Y \rightarrow [0, 1]$ e $S : Y \times Z \rightarrow [0, 1]$ é uma relação *fuzzy* $T : R \times S \rightarrow [0, 1]$ cuja função de pertinência $T(x, z)$

é dada por:

$$T(x, z) = \inf_{y \in Y} \{R(x, y) \text{ s } S(y, z)\} \quad \forall x \in X \text{ e } \forall z \in Z, \quad (4.6)$$

em que s é uma s-norma. A composição inf-s de R e S é denotada por $T = R \bullet S$

O procedimento para se obter a composição inf-s é o mesmo daquele utilizado para a composição sup-t, substituindo-se o supremo pelo ínfimo e a t-norma por uma s-norma. (HELL, 2008)

Com o conhecimento prévio dos elementos dos conjuntos *fuzzy*, pode-se inteirar mais precisamente da forma de raciocínio que um sistema baseado em lógica *fuzzy* utiliza. O fluxograma apresentado na figura 6 ilustra resumidamente os estágios da computação *fuzzy* para um sistema ou processo genérico. Nota-se que como parte fundamental do algoritmo se encontra a base de conhecimento e regras *fuzzy*, que pode ser um banco de dados em memória ou aferições feitas por algum sistema de sensoriamento, e que será utilizada, na máquina de inferência no momento de buscar a melhor resposta para determinada ação.

4.5 Sistema de inferência *fuzzy singleton*

O sistema de inferência *fuzzy singleton* contém quatro blocos ou componentes principais, a saber: regras, *fuzzificador*, inferência e *defuzzificador*. Uma vez definidas as regras, um sistema de inferência *fuzzy* pode ser visualizado como um mapeamento das entradas para as saídas, que pode ser expresso através de modelos matemáticos. Assim, podemos descrever e estabelecer procedimentos para se determinar os parâmetros do sistema de inferência *fuzzy* e com isso otimizar o seu desempenho. Este tipo de sistema de inferência *fuzzy* é largamente utilizado em diversos segmentos da engenharia, tais como em controladores *fuzzy* e processadores de sinais. A figura 6 mostra a estrutura de um sistema de inferência *fuzzy*, em que os termos “Entrada discreta” e “Saída discreta” remetem a um número real relacionado a um conjunto \mathcal{A} que tem a função de pertinência dada por:

$$\mathcal{A} \Rightarrow \mu_{\mathcal{A}}(x) = \begin{cases} 1, \text{ se } x \in \mathcal{A} \\ 0, \text{ se } x \notin \mathcal{A} \end{cases} \quad (4.7)$$

Quando uma entrada discreta é aplicada em tal sistema de inferência *fuzzy*, a mesma é fuzzificada e, em seguida, a máquina de inferência calcula o conjunto *fuzzy* de

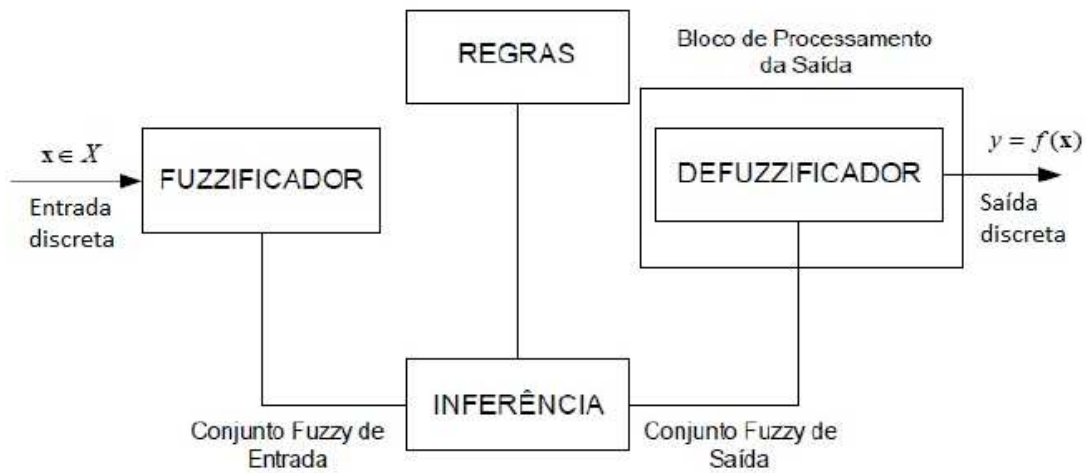


Figura 6: Sistema de inferência *fuzzy singleton* utilizada, adaptada de (RIBEIRO, 2005).

saída correspondente a cada regra. A seguir, o defuzzificador calcula o valor discreto da saída do sistema de inferência *fuzzy*. Esse processo de obtenção da saída do sistema de inferência *fuzzy singleton* é expresso da seguinte maneira:

Seja um conjunto *fuzzy* com p entradas $x_1 \in X_1, \dots, x_p \in X_p$ e uma saída $y^l \in Y$. Supondo que o sistema possui M regras *fuzzy*, em que a l -ésima regra é dada por (MENDEL, 1991, 1995)

$$\text{Se } x_1 \text{ é } F_1^l \text{ e } \dots \text{ e } x_p \text{ é } F_p^l, \text{ Então } y^l \text{ é } G^l, \quad l = 1, \dots, M. \quad (4.8)$$

Esta regra representa uma relação *fuzzy* entre o espaço de entrada $X_1 \times \dots \times X_p$ e o espaço de saída Y do sistema de inferência *fuzzy*. Note que

$$F_k^l = \left\{ \left(x, \mu_{F_k^l}(x) \right) \mid x \in X, \mu_{F_k^l}(x) \in [0, 1] \right\} \quad (4.9)$$

é o k -ésimo conjunto *fuzzy* da l -ésima regra *fuzzy* cuja função de pertinência é dada por $\mu_{F_k^l}(x_k)$. Por sua vez, a expressão linguística “Se x_1 é F_1^l e \dots e x_p é F_p^l ” é o antecedente da l -ésima regra *fuzzy*, a sentença linguística “Então y^l é G^l ” é o conseqüente da l -ésima regra *fuzzy* e G^l um conjunto *fuzzy*. A figura 7 exemplifica uma função Gaussiana de pertinência utilizada no antecedente de uma regra do sistema *fuzzy*, enquanto a figura 8 mostra as funções de pertinência utilizada para modelar a imprecisão no valor de uma variável, por exemplo a temperatura ambiente.

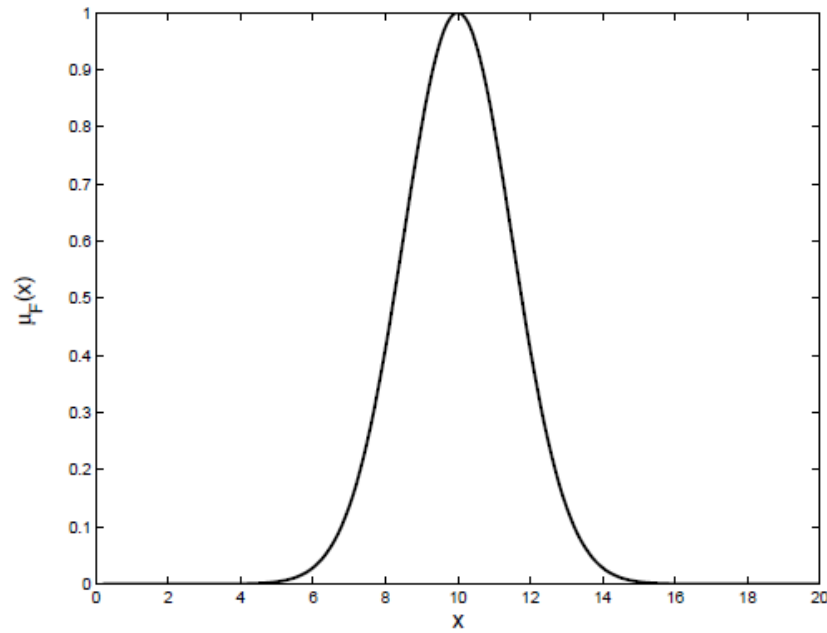


Figura 7: Exemplo de função de pertinência Gaussiana.

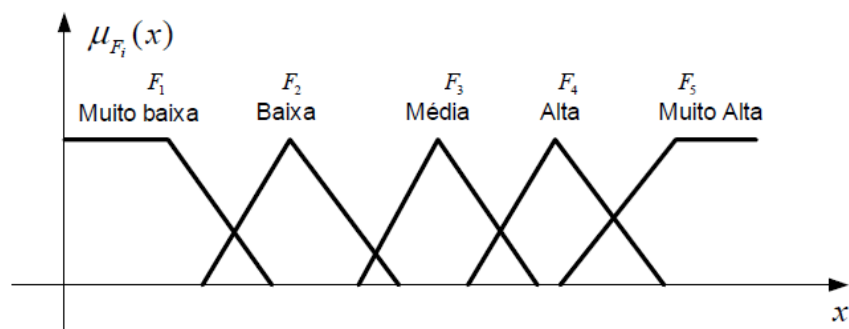


Figura 8: Funções de pertinência para a modelar a incerteza no valor da temperatura.

Na máquina de inferência *fuzzy* (representada pelo bloco “Inferência” na figura 6, utiliza-se princípios de lógica fuzzy para combinar regras *IF-THEN* a partir da base de regras *fuzzy* obtida através de um mapeamento de conjuntos *fuzzy* de entrada em $x_1 \in X_1, \dots, x_p \in X_p$ para os conjuntos *fuzzy* de saída em Y . Cada regra é interpretada como uma implicação *fuzzy*. Assumindo que $F_1^l \times \dots \times F_p^l = A^l$, então a sentença expressa em (4.8) pode ser reescrita da seguinte maneira

$$R^l : F_1^l \times \dots \times F_p^l \rightarrow G^l = A^l \rightarrow G^l, \quad l = 1, \dots, M. \quad (4.10)$$

Note que R^l é descrita pela função de pertinência $\mu_{R^l}(\mathbf{x}, y)$, em que

$$\mu_{R^l}(\mathbf{x}, y) = \mu_{A^l \rightarrow G^l}(\mathbf{x}, y) \quad (4.11)$$

e $\mathbf{x} = [x_1, \dots, x_p]^T$. Com isso, temos que $\mu_{R^l}(\mathbf{x}, y) = \mu_{A^l \rightarrow G^l}(x_1, \dots, x_p, y)$ e assumindo a utilização de implicações Mandami, podemos inferir que

$$\begin{aligned} \mu_{R^l}(\mathbf{x}, y) &= \mu_{A^l \rightarrow G^l}(\mathbf{x}, y) = \mu_{F_1^l \times \dots \times F_p^l \rightarrow G^l}(\mathbf{x}, y) \\ &= \mu_{F_1^l \times \dots \times F_p^l}(\mathbf{x}) \star \mu_{G^l}(y) \\ &= \mu_{F_1^l}(x_1) \star \dots \star \mu_{F_p^l}(x_p) \star \mu_{G^l}(y) \\ &= \left[T_{k=1}^p \mu_{F_k^l}(x_k) \right] \star \mu_{G^l}(y). \end{aligned} \quad (4.12)$$

Note que os múltiplos antecedentes são conectados pelos operadores *and* e por t-normas.

A entrada p -dimensional para R^l é dada por um conjunto $A_{\mathbf{x}}$ cuja a função de pertinência é dada por

$$\begin{aligned} \mu_{A_{\mathbf{x}}}(\mathbf{x}) &= \mu_{X_1}(x_1) \star \mu_{X_2}(x_2) \star \dots \star \mu_{X_p}(x_p) \\ &= T_{k=1}^p \mu_{X_k}(x_k), \end{aligned} \quad (4.13)$$

em que \star e T indicam as t-normas que foram escolhidas. Cada regra R^l determina um conjunto *fuzzy* $B^l = A_{\mathbf{x}} \circ R^l$ em Y tal que

$$\begin{aligned} \mu_{B^l}(y) &= \mu_{A_{\mathbf{x}} \circ R^l}(y) \\ &= \sup_{\mathbf{x} \in X} [\mu_{A_{\mathbf{x}}}(\mathbf{x}) \star \mu_{A^l \rightarrow G^l}(\mathbf{x}, y)], \quad y \in Y. \end{aligned} \quad (4.14)$$

Esta composição, denominada composição *sup-star* é um mapeamento altamente não linear de um vetor \mathbf{x} em um conjunto *fuzzy* escalar $\mu_{B^l}(y)$. Podemos interpretar a máquina de inferência *fuzzy* como um sistema que mapeia conjuntos *fuzzy* através da composição *sup-star* apresentada na equação (4.14). Tal situação é retratada a seguir, na figura 9.

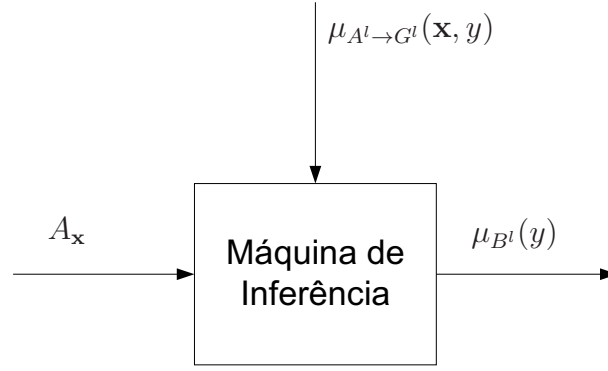


Figura 9: Interpretação da máquina de inferência *fuzzy* como um sistema.

Substituindo as equações (4.12) e (4.13) em (4.14), temos que

$$\begin{aligned}
 \mu_{B^l}(y) &= \sup_{\mathbf{x} \in X} [\mu_{A_X}(\mathbf{x}) \star \mu_{A^l \rightarrow G^l}(\mathbf{x}, y)] \\
 &= \sup_{\mathbf{x} \in X} \left[T_{k=1}^p \mu_{X_k}(x_k) \star \left[T_{k=1}^p \mu_{F_k^l}(x_k) \right] \star \mu_{G^l}(y) \right] \\
 &= \sup_{\mathbf{x} \in X} \left\{ \left[T_{k=1}^p \mu_{X_k}(x_k) \star \mu_{F_k^l}(x_k) \star \mu_{G^l}(y) \right] \right\} \\
 &= \mu_{G^l}(y) \star \left\{ \left[\sup_{x_1 \in X_1} \mu_{X_1}(x_1) \star \mu_{F_1^l}(x_1) \right] \star \right. \\
 &\quad \left. \cdots \star \left[\sup_{x_p \in X_p} \mu_{X_p}(x_p) \star \mu_{F_p^l}(x_p) \right] \right\}, \quad y \in Y.
 \end{aligned} \tag{4.15}$$

Esta é uma sequência muito importante de cálculos e resultados. A última linha decorre da comutatividade de uma *t*-norma e do fato de que $\mu_{X_k}(x_k) \star \mu_{F_k^l}(x_k)$ é apenas uma função de x_k . Note que cada *supremum* na equação (4.15) é uma variável escalar. O conjunto *fuzzy* final B é determinado por todas as M regras e pode ser obtido através da combinação B^l com sua correspondente função de pertinência $\mu_{B^l}(y)$, em que $l = 1, \dots, M$,

$$B = A_{\mathbf{x}} \circ [R^1, \dots, R^M]. \tag{4.16}$$

O conjunto suporte de um conjunto *fuzzy* A é o conjunto de elementos no universo X para os quais $\mu_A(x) > 0$. Um conjunto *fuzzy* cujo suporte é um único ponto x' com $\mu_A(x') = 1$ é chamado de conjunto unitário fuzzy ou singleton. Para a fuzzificação *singleton*, note que a operação *supremum* na composição *sup-star* apresentada na equação (4.15) é de fácil compreensão, uma vez que $\mu_{X_k}(x_k)$ é diferente de zero somente no ponto $x_k = x'_k$. Com isso, temos que

$$\begin{aligned}
\mu_{B^l}(y) &= \mu_{G^l}(y) \star \left\{ \left[\sup_{x_1 \in X_1} \mu_{X_1}(x_1) \star \mu_{F_1^l}(x_1) \right] \star \right. \\
&\quad \left. \cdots \star \left[\sup_{x_p \in X_p} \mu_{X_p}(x_p) \star \mu_{F_p^l}(x_p) \right] \right\} \\
&= \mu_{G^l}(y) \star \left\{ \left[\mu_{X_1}(x'_1) \star \mu_{F_1^l}(x'_1) \right] \star \right. \\
&\quad \left. \cdots \star \left[\mu_{X_p}(x'_p) \star \mu_{F_p^l}(x'_p) \right] \right\} \\
&= \mu_{G^l}(y) \star \left[\mu_{F_1^l}(x'_1) \star \cdots \star \mu_{F_p^l}(x'_p) \right], y \in Y,
\end{aligned} \tag{4.17}$$

Para as implicações produto e mínimo, temos que $\mu_{X_k}(x'_k) \star \mu_{F_k^l}(x'_k) = 1 \star \mu_{F_k^l}(x'_k) = \mu_{F_k^l}(x'_k)$. O termo entre colchetes na última linha da equação (4.17) é denominado nível de disparo (MENDEL, 2001). Note que $\mu_{B^l}(y)$ depende de $\mathbf{x} = \mathbf{x}'$. Logo, se \mathbf{x}' sofre variação então $\mu_{B^l}(y)$ é alterado. A notação $\mu_{B^l}(y|\mathbf{x} = \mathbf{x}')$ é mais adequada, todavia é pouco utilizada na literatura. Por essa razão, iremos adotar a notação $\mu_{B^l}(y)$.

Para cada valor de $y \in Y$, ou seja, y' , sabemos que $\mu_{B^l}(y')$ pode ser visualizado como um valor discreto. Devido ao fato de que $y \in Y$ e $\mu_{B^l}(y)$ estarem em função de y , note que a equação (4.17) pode ser avaliada $\forall y \in Y$ e por isso a mesma é considerada uma função de pertinência.

A defuzzificação pela altura (DRIANKOV; HELLENDORRN; REINFRANK, 1996), também pode ser chamada de defuzzificação pela média dos máximos (WANG, 1997, 1994) substitui cada regra da saída *fuzzy* definida por um *singleton* no ponto que o conjunto possui maior função de pertinência e então calcula-se a centróide do conjunto tipo 1 composta por tais *singletons*.

As t-normas usualmente empregadas em aplicações de engenharia são os operadores produto e mínimo. Considerando que o defuzzificador é baseado na altura, então a saída discreta do sistema de inferência *fuzzy singleton* é dada por

$$y_h(\mathbf{x}) = \frac{\sum_{l=1}^M \theta_l \mu_{B^l}(\theta_l)}{\sum_{l=1}^M \mu_{B^l}(\theta_l)}, \tag{4.18}$$

em que θ_l é o ponto que possui maior valor de função de pertinência no l -ésimo conjunto de saída e μ_{B^l} é o seu grau de pertinência no l -ésimo conjunto de saída. Ou seja, θ_l é o valor no universo correspondente ao centro de gravidade do conjunto *fuzzy* B^l associado ao grau de ativação da regra R^l . Para a fuzzificação *singleton*, sabemos a partir da equação (4.17) que

$$\mu_{B^l}(\theta_l) = \mu_{G^l}(\theta_l) \star \left[\mu_{F_1^l}(x'_1) \star \cdots \star \mu_{F_p^l}(x'_p) \right], y \in Y. \tag{4.19}$$

Este método é simples pois o centro de gravidade das funções de pertinência mais comuns é conhecido *a priori* e para a função de pertinência Gaussiana tal centro de gravidade

é o valor central da função. Tal método de defuzzificação utiliza o centro do suporte \bar{y}^l da função de pertinência do consequente. Por isso, qualquer que seja a largura da função de pertinência, o método fornece o mesmo resultado. Usando-se outros métodos de defuzzificação diferentes saídas podem ser obtidas.

Assumindo a escolha pela fuzzificação *singleton*, composição *max-product*, implicação produto e defuzzificação pela altura e deixando em aberto a escolha pela função de pertinência, é simples mostrar que

$$\begin{aligned} y(\mathbf{x}) &= f_s(\mathbf{x}) \\ &= \frac{\sum_{l=1}^M \theta_l \prod_{k=1}^p \mu_{F_k^l}(x_k)}{\sum_{l=1}^M \prod_{k=1}^p \mu_{F_k^l}(x_k)}. \end{aligned} \quad (4.20)$$

Note que o uso do sobrescrito “s” em $f_s(\mathbf{x})$ remete à utilização do sistema de inferência *fuzzy singleton*. Para obter a equação (4.20), iniciamos com a equação (4.18) e substituímos pelo $\mu_{B^l}(\theta_l)$ da equação (4.21). Adicionalmente, por simplicidade notacional, renomeamos x'_i para x_i , de forma que podemos escrever $f_s(\mathbf{x}')$ como $f_s(\mathbf{x})$. Com isso, assumindo que as funções de pertinência são normalizadas, de modo que $\mu_{G^l}(\theta_l) = 1$, temos que (MENDEL, 2001)

$$\begin{aligned} \mu_{B^l}(\theta_l) &= \left[\prod_{k=1}^p \mu_{F_k^l}(x'_k) \right] \times \mu_{G^l}(\theta_l) \\ &= \prod_{k=1}^p \mu_{F_k^l}(x'_k). \end{aligned} \quad (4.21)$$

Sabendo que as t-normas são operadores produto e cada regra $\mu_{F_k^l}(x_k)$ pode ser visualizada como uma função de pertinência Gaussiana, temos que,

$$\mu_{F_k^l}(x_k) = \exp \left\{ -\frac{1}{2} \left(\frac{x_k - m_{F_k^l}}{\sigma_{F_k^l}} \right)^2 \right\}, \quad (4.22)$$

em que $l = 1, \dots, M$. Note que $m_{F_k^l}$ e $\sigma_{F_k^l}$ são as médias e variâncias, respectivamente.

A equação (4.20) pode ser representado de forma genérica por

$$\begin{aligned} y(\mathbf{x}) &= f_s(\mathbf{x}) \\ &= \sum_{l=1}^M \theta_l \phi_l(\mathbf{x}), \end{aligned} \quad (4.23)$$

em que θ_l é o peso associado à l -ésima regra e $\phi_l(\mathbf{x})$ é chamado de função de base *fuzzy*

(fuzzy basis function - FBF) (WANG; MENDEL, 1992a), dada por

$$\begin{aligned}\phi_l(\mathbf{x}) &= \frac{\prod_{k=1}^p \mu_{F_k^l}(x_k)}{\sum_{l=1}^M \prod_{k=1}^p \mu_{F_k^l}(x_k)} \\ &= \frac{\prod_{k=1}^p \exp\left(-\frac{(x_k^{(q)} - m_{F_k^l})^2}{2\sigma_{F_k^l}^2}\right)}{\sum_{l=1}^M \left[\prod_{k=1}^p \exp\left(-\frac{(x_k^{(q)} - m_{F_k^l})^2}{2\sigma_{F_k^l}^2}\right) \right]},\end{aligned}\quad (4.24)$$

em que $l = 1, \dots, M$ (MENDEL, 2001).

Para o sistema de inferência *fuzzy* projetado, optou-se por fazer uso do algoritmo de treinamento denominado retropropagação, do inglês *back-propagation*. A lógica *fuzzy* abordada visa especificar os parâmetros $m_{F_k^l}$, $\sigma_{F_k^l}$ de tais funções de pertinência, bem como os centros θ_l das consequentes funções de pertinência da equação (4.23).

Dado um conjunto de pares de entradas-saídas ($\mathbf{x}^{(q)} : y^{(q)}$), deseja-se determinar um sistema que remete a um ajuste ótimo para tais conjuntos, com o respaldo da função custo. Existem vários métodos que podem ser utilizados para projetar um sistema de inferência *fuzzy* com diferentes propriedades e características.

$$\begin{aligned}J(\mathbf{w}^{(q)}) &= e^{(q)} \\ &= \frac{1}{2} [f_s(\mathbf{x}^{(q)}) - y^{(q)}]^2,\end{aligned}\quad (4.25)$$

seja minimizada. Note que utilizando o algoritmo de retropropagação para minimizar $e^{(q)}$, podemos facilmente obter as atualizações dos seguintes parâmetros: (MENDEL, 1991)

$$\begin{aligned}m_{F_k^l}(q+1) &= m_{F_k^l}(q) - \alpha_m [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \\ &\quad \times \frac{[x_k^{(q)} - m_{F_k^l}(q)]}{\sigma_{F_k^l}^2(q)} \phi_l(\mathbf{x}^{(q)}),\end{aligned}\quad (4.26)$$

$$\theta_l(q+1) = \theta_l(q) - \alpha_{\theta_l} [f_s(\mathbf{x}^{(q)}) - y^{(q)}] \phi_l(\mathbf{x}^{(q)}), \quad (4.27)$$

e

$$\begin{aligned}\sigma_{F_k^l}(q+1) &= \sigma_{F_k^l}(q) - \alpha_\sigma [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \\ &\quad \times \frac{[x_k^{(q)} - m_{F_k^l}(q)]^2}{\sigma_{F_k^l}^3(q)} \phi_l(\mathbf{x}^{(q)}).\end{aligned}\quad (4.28)$$

Na equação (4.28) optou-se por atualizar $\sigma_{F_k^l}$ ao invés de $\sigma_{F_k^l}^2$, pois $\sigma_{F_k^l}^2$ sempre será

positivo enquanto $\sigma_{F_k^l}$ pode assumir valores positivos ou negativos. É válido ressaltar que os valores atribuídos para $m_{F_k^l}(0)$, $\theta_l(0)$ e $\sigma_{F_k^l}(0)$ devem ser informados para inicializarmos as equações (4.26), (4.27) e (4.28), respectivamente.

Note que é possível obter uma boa inicialização para os parâmetros $m_{F_k^l}$, θ_l e $\sigma_{F_k^l}$. Uma maneira pouco satisfatória de se inicializar tais parâmetros é fazer com que os mesmos sejam inicializados de forma randômica. Tal estratégia poderia fazer com que o algoritmo tenha uma convergência lenta. Inicializar os parâmetros de maneira inteligente, atribuindo valores plausíveis para tais parâmetros pode fazer com que o algoritmo de retropropagação possa convergir mais rapidamente (CHU; MENDEL, 1994).

Os parâmetros de aprendizagem α_m , α_{θ_l} e α_{σ} devem ser escolhidos com parcimônia. Normalmente, atribui-se o valor α para tais parâmetros. A escolha de um valor muito grande para α pode fazer com que o algoritmo não possa convergir. Todavia, atribuir para α um valor pequeno pode fazer com que o algoritmo demore muito tempo para convergir.

Durante o treinamento com o algoritmo de retropropagação, o sistema opera em uma sequência de dois passos. Primeiro, um padrão é apresentado à camada de entrada do sistema de inferência *fuzzy*. A atividade resultante flui através do sistema de inferência *fuzzy*, camada por camada, até que a resposta seja produzida pela camada na saída do sistema. No segundo passo, a saída obtida é comparada à saída desejada para esse padrão particular. Se esta não estiver correta, o erro é calculado e propagado à partir da saída até a entrada. Desta forma, os os parâmetros do sistema de inferência *fuzzy* vão sendo modificados conforme o erro é propagado. As equações (4.26), (4.27) e (4.28) são consideradas, portanto, parte de uma estrutura de um algoritmo de retropropagação devido às suas dependências em relação ao erro ($e^{(i)}$), que se propaga da saída para a entrada.

Os sistemas de inferência *fuzzy singleton* têm sido amplamente utilizados devido à simplicidade e baixo custo computacional destes fuzzificadores. Entretanto, este tipo de sistema de inferência *fuzzy* pode não ser adequado em casos onde há a presença de ruídos nos dados de treinamento ou nos dados que serão futuramente processados pelos sistemas *fuzzy*.

4.6 Sistema de inferência *fuzzy non-singleton*

Para considerar a presença de ruídos nas entradas dos sistemas de inferência *fuzzy*, foram desenvolvidos fuzzificadores *non-singleton* que assumem que os dados na entrada

apresentam incertezas e, conseqüentemente, os mesmos possuem associados à eles funções de pertinência que assumem valores no intervalo $[0, 1]$. Na fuzzificação *non-singleton*, a medição $x_k = x'_k$ é mapeada em um conjunto *fuzzy* (KAUFMAN; GUPTA, 1991) e possui uma função de pertinência associada. Ou seja, de acordo com (MENDEL, 2001), temos que um fuzzificador *non-singleton* é aquele para o qual $\mu_{X_k}(x'_k) = 1$, em que $k = 1, \dots, p$ e $\mu_{X_k}(x_k)$ é decrementado de modo contínuo quando x_k se afasta de x'_k .

Conceitualmente, o fuzzificador *non-singleton* assume que o valor de entrada x'_k é o valor mais provável de ser o correto, se comparado com todos os valores apresentados em sua vizinhança imediata. Contudo, assumindo que a entrada é corrompida por ruído, pontos vizinhos também são suscetíveis de ser o valor correto, todavia em menor grau de afirmação.

Durante o projeto devemos determinar o formato da função de pertinência $\mu_{X_k}(x_k)$ baseado em uma estimativa de tipo e quantidade de ruído presente. São exemplos de funções de pertinência (MENDEL, 2001)

1. Gaussiana: $\mu_{X_k}(x_k) = \exp[-(x_k - x'_k)^2/2\sigma^2]$,
2. Triangular: $\mu_{X_k}(x_k) = \max[0.1 - |(x_k - x'_k)/c|]$,
3. Sino: $\mu_{X_k}(x_k) = 1/(1 + |(x_k - x'_k)/c|^n)$.

Note que x'_k é o valor central dos conjuntos *fuzzy*. Já $\sigma(c)$ define o espaçamento entre tais conjuntos.

Para a fuzzificação *singleton*, simplifica-se a equação (4.15) para a obtenção da forma simplificada demonstrada na equação (4.17). Tal fato é possível pois $\mu_{X_k}(x_k)$ é diferente de zero somente no ponto $x_k = x'_k$. Todavia, é válido ressaltar que tal simplificação não se aplica à fuzzificação *non-singleton*. Note que

$$\mu_{Q_k^l}(x_k) \equiv \mu_{X_k}(x_k) \star \mu_{F_k^l}(x_k), \quad (4.29)$$

em que $k = 1, \dots, p$ e $l = 1, \dots, M$. O valor de X_k no ponto em que ocorre o *supremum* de $\mu_{Q_k^l}(x_k)$ é chamado de $x_{k,\max}^l$. Assim, para a fuzzificação *non-singleton*, a equação (4.15) pode ser reescrita da seguinte maneira

$$\mu_{B^l}(y) = \mu_{G^l}(y) \star \left[T_{k=1}^p \mu_{Q_k^l}(x_{k,\max}^l) \right], \quad (4.30)$$

em que $y \in Y$ e $l = 1, \dots, M$. Esta é a equação fundamental para o sistema de inferência *fuzzy non-singleton*. Assim como apresentado na seção (4.6), o termo $\left[T_{k=1}^p \mu_{Q_k^l}(x_{k,\max}^l) \right]$ é denominado nível de disparo para um sistema de inferência *fuzzy non-singleton*.

Comparando as equações (4.30) e (4.17), temos a presença de um pré-filtro que transforma a entrada \mathbf{x} em \mathbf{x}_{\max}^l , conforme apresentado na figura 10. Tal pré-filtro é o resultado direto da composição *sup-star* e ocorre naturalmente no sistema de inferência *fuzzy non-singleton*.

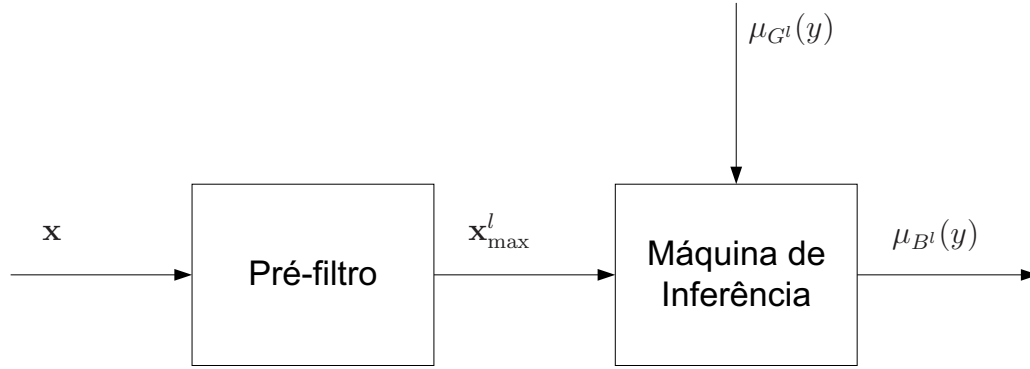


Figura 10: Pré-filtragem da entrada para um sistema de inferência *fuzzy non-singleton*.

Para obter a equação (4.32), iniciamos com a equação (4.18) e combinamos a mesma com o termo $\mu_{B^l}(\theta_l)$ da equação (4.30). Adicionalmente, por simplicidade notacional, renomeamos x'_k para x_k , de forma que escrevemos $f_{ns}(\mathbf{x}')$ como $f_{ns}(\mathbf{x})$. Com isso, assumindo que as funções de pertinência são normalizadas, de modo que $\mu_{G^l}(\theta_l) = 1$, temos que (MENDEL, 2001)

$$\begin{aligned} \mu_{B^l}(\theta_l) &= \left[\prod_{k=1}^p \mu_{Q_k^l}(x_{k,\max}^l) \right] \times \mu_{G^l}(\theta_l) \\ &= \prod_{k=1}^p \mu_{Q_k^l}(x_{k,\max}^l), \end{aligned} \quad (4.31)$$

em que $l = 1, \dots, M$.

Assumindo a escolha pela fuzzificação *non-singleton*, composição *max-product*, implicação produto e defuzzificação pela altura e deixando em aberto a escolha pela função de pertinência, temos que

$$\begin{aligned} y(\mathbf{x}) &= f_{ns}(\mathbf{x}) \\ &= \frac{\sum_{l=1}^M \theta_l \prod_{k=1}^p \mu_{Q_k^l}(x_{k,\max}^l)}{\sum_{l=1}^M \prod_{k=1}^p \mu_{Q_k^l}(x_{k,\max}^l)}. \end{aligned} \quad (4.32)$$

Note que o uso do sobrescrito “*ns*” em $f_{ns}(\mathbf{x})$ remete à utilização do sistema de inferência *fuzzy non-singleton*. O sistema de inferência fuzzy expresso na equação (4.32) pode ser representado de forma genérica por

$$\begin{aligned} y(\mathbf{x}) &= f_{ns}(\mathbf{x}), \\ &= \sum_{l=1}^M \theta_l \phi_l(\mathbf{x}), \end{aligned} \quad (4.33)$$

em que θ_l é o peso associado à l -ésima regra e $\phi_l(\mathbf{x})$ é chamado de função de base *fuzzy* (*fuzzy basis function* - FBF) (WANG; MENDEL, 1992a), dada por

$$\begin{aligned}\phi_l(\mathbf{x}) &= \frac{\prod_{k=1}^p \mu_{Q_k^l}(x_{k,\max}^l)}{\prod_{k=1}^p \mu_{Q_k^l}(x_{k,\max}^l)} \\ &= \frac{\prod_{k=1}^p \exp\left(-\frac{1}{2} \frac{(x_k^{(q)} - m_{F_k^l})^2}{\sigma_{F_k^l}^2 + \sigma_X^2}\right)}{\sum_{l=1}^M \left[\prod_{k=1}^p \exp\left(-\frac{1}{2} \frac{(x_k^{(q)} - m_{F_k^l})^2}{\sigma_{F_k^l}^2 + \sigma_X^2}\right) \right]},\end{aligned}\quad (4.34)$$

em que $l = 1, \dots, M$ (MENDEL, 2001). Com isso, é válido mencionar que a equação (4.34) já se torna uma função de pertinência Gaussiana.

Seja um conjunto de N pares de entradas-saídas de treinamento, $(\mathbf{x}^{(1)} : y^{(1)})$, $(\mathbf{x}^{(2)} : y^{(2)})$, \dots , $(\mathbf{x}^{(N)} : y^{(N)})$, em que \mathbf{x} é o vetor de entrada com a presença ruídos e y é uma saída escalar. A função de pertinência Gaussiana é modelada conforme a seguir:

$$\mu_{F_k^l}(x_k) = \exp\left\{-\frac{1}{2} \left(\frac{x_k - m_{F_k^l}}{\sigma_{F_k^l}}\right)^2\right\}, \quad (4.35)$$

em que $k = 1, \dots, p$ e $l = 1, \dots, M$. Adicionalmente, cada uma das entradas do sistema de inferência *fuzzy* são modeladas como uma incerteza Gaussiana, cuja função de pertinência é dada por

$$\mu_{X_k}(x_k) = \exp\left\{-\frac{1}{2} \left(\frac{x_k - m_{X_k}}{\sigma_{X_k}}\right)^2\right\}, \quad (4.36)$$

em que $k = 1, \dots, p$. Note que $\mu_{X_k}(x_k)$ e $\mu_{F_k^l}(x_k)$ são respectivamente, as funções de pertinência do sinal na entrada do sistema de inferência *fuzzy* e da regra da função de pertinência.

Com isso, após definidas as equações (4.35) e (4.36) e combinando as mesmas com a formulação apresentada em (4.29) e (4.34), temos que $y(\mathbf{x}^{(q)})$ assume a seguinte formatação

$$\begin{aligned}y(\mathbf{x}^{(q)}) &= f_{ns}(\mathbf{x}^{(q)}) \\ &= \frac{\sum_{l=1}^M \theta_l \left[\prod_{k=1}^p \exp\left(-\frac{1}{2} \frac{(x_k^{(q)} - m_{F_k^l})^2}{\sigma_{F_k^l}^2 + \sigma_X^2}\right) \right]}{\sum_{l=1}^M \left[\prod_{k=1}^p \exp\left(-\frac{1}{2} \frac{(x_k^{(q)} - m_{F_k^l})^2}{\sigma_{F_k^l}^2 + \sigma_X^2}\right) \right]}\end{aligned}\quad (4.37)$$

em que $q = 1, \dots, N$, $k = 1, \dots, p$ e $l = 1, \dots, M$.

No algoritmo de retropropagação para o sistema de inferência *fuzzy non-singleton* os parâmetros utilizados não são fixos e sofrem atualizações. Assume-se então, que tais parâmetros devem ser ajustados. As equações do método *non-singleton* diferem-se do método *singleton* devido a presença do ruído σ_X presente em sua concepção. Nota-se que tal método visa estabelecer como especificar os parâmetros $m_{F_k^l}$, $\sigma_{F_k^l}$, σ_X e θ_l , utilizando o conjunto de treinamento $(\mathbf{x}^{(1)} : y^{(1)})$, $(\mathbf{x}^{(2)} : y^{(2)})$, ..., $(\mathbf{x}^{(N)} : y^{(N)})$. Com isso, o objetivo principal é projetar um sistema de inferência *fuzzy* a partir de (4.37), de tal forma que a função de erro instantânea seja minimizada

$$\begin{aligned} J(\mathbf{w}^{(q)}) &= e^{(q)}, \\ &= \frac{1}{2} [f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}]^2. \end{aligned} \quad (4.38)$$

Ao analisar a equação (4.37), é evidente que f_{ns} é caracterizado por θ_l , $m_{F_k^l}$, $\sigma_{F_k^l}$ e σ_X . Utilizando o algoritmo de retropropagação para minimizar $e^{(q)}$, podemos facilmente obter as atualizações dos seguintes parâmetros do sistema de inferência *fuzzy*

$$\begin{aligned} m_{F_k^l}(q+1) &= m_{F_k^l}(q) - \alpha_m [f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_{ns}(\mathbf{x}^{(q)})] \\ &\quad \times \left[\frac{x_k^{(q)} - m_{F_k^l}(q)}{\sigma_X^2(q) + \sigma_{F_k^l}^2(q)} \right] \phi_l(\mathbf{x}^{(q)}), \end{aligned} \quad (4.39)$$

$$\theta_l(q+1) = \theta_l(q) - \alpha_{\theta_l} [f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}] \phi_l(\mathbf{x}^{(q)}), \quad (4.40)$$

$$\begin{aligned} \sigma_{F_k^l}(q+1) &= \sigma_{F_k^l}(q) - \alpha_{\sigma} [f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_{ns}(\mathbf{x}^{(q)})] \\ &\quad \times \sigma_{F_k^l}(q) \left[\frac{x_k^{(q)} - m_{F_k^l}(q)}{\sigma_X^2(q) + \sigma_{F_k^l}^2(q)} \right]^2 \phi_l(\mathbf{x}^{(q)}), \end{aligned} \quad (4.41)$$

e

$$\begin{aligned} \sigma_X(q+1) &= \sigma_X(q) - \alpha_X [f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_{ns}(\mathbf{x}^{(q)})] \\ &\quad \times \sigma_X(q) \left[\frac{x_k^{(q)} - m_{F_k^l}(q)}{\sigma_X^2(q) + \sigma_{F_k^l}^2(q)} \right]^2 \phi_l(\mathbf{x}^{(q)}). \end{aligned} \quad (4.42)$$

Assim, optou-se por atualizar $\sigma_{F_k^l}$ e σ_X ao invés de $\sigma_{F_k^l}^2$ e σ_X^2 , pois $\sigma_{F_k^l}^2$ e σ_X^2 sempre serão valores positivos enquanto $\sigma_{F_k^l}$ e σ_X podem assumir tanto valores positivos quanto negativos. É válido ressaltar que os valores atribuídos para $m_{F_k^l}(0)$, $\theta_l(0)$, $\sigma_{F_k^l}(0)$ e $\sigma_X(0)$ devem ser informados a fim de se inicializar as equações (4.39), (4.40), (4.41) e (4.42),

respectivamente.

4.7 Sumário

Este capítulo abordou a estruturação e projeto de um sistema de inferência *fuzzy*. Os principais termos e definições foram apresentados. Na seção 4.5, foi apresentada o sistema de inferência *fuzzy singleton* bem como os parâmetros $m_{F_k^l}$, θ_l e $\sigma_{F_k^l}$ e suas atualizações, feitas através do algoritmo de retropropagação.

Já a seção 4.6 abordou as principais questões referentes ao sistema de inferência *fuzzy non-singleton*. Da mesma maneira, foram apresentados os parâmetros $m_{F_k^l}$, $\sigma_{F_k^l}$, θ_l e σ_X e suas respectivas atualizações retornadas pelo algoritmo de retropropagação.

5 PROPOSTAS

O presente capítulo descreve um sistema de inferência *fuzzy singleton* e *non-singleton* treinado por algoritmos do gradiente conjugado apresentados na Seção 5.3, visando, sobretudo, mostrar que tal técnica baseada em inteligência computacional pode fornecer uma maior velocidade de convergência para aplicações *offline*.

A otimização com base em algoritmos do gradiente conjugado foi originalmente proposta por Hestenes e Stiefel, em (HESTENES; STIEFEL, 1952). Seu objetivo inicial foi a resolução de problemas quadráticos sem restrições, mas logo o mesmo foi estendido para casos mais gerais. Este método pode ser considerado sob dois pontos de vista: Como um método de descida, com busca linear exata ou como um método de resolução de sistema linear, baseado em um processo de ortogonalização.

O método do gradiente conjugado foi criado visando à resolução de problemas lineares iterativamente (HESTENES; STIEFEL, 1952). Considerando as matrizes de coeficientes simétricas definidas positivamente, o método converge em um número finito de iterações. Contudo, quando se trata de matrizes não-simétricas o método não converge da mesma maneira. Cada nova direção do gradiente conjugado é uma combinação linear de resíduo corrente com a direção anterior.

O método do gradiente conjugado, consiste em um método iterativo de busca do mínimo local da função. Desta forma, geram-se aproximações para a solução e, em cada iteração do método, dois produtos internos são realizados para que se calculem dois escalares definidos de forma que a seqüência obedeça condições de ortogonalidade. Descrições do método podem ser encontradas em (ATKINSON, 1988), (AVRIEL, 2003) e (GOLUB; LOAN, 1996).

Para a utilização do método proposto é preciso projetar matrizes Hessianas durante o período de treinamento. É válido ressaltar que para sistemas de inferência *fuzzy* tal tipo de matriz ainda não foi deduzido, sendo esta uma das principais contribuições deste trabalho.

O capítulo em questão é dividido da seguinte maneira: A Seção 5.1 apresenta as deduções dos termos da matriz Hessiana para o sistema de inferência *fuzzy singleton*. Por sua vez, a Seção 5.2 apresenta as deduções dos termos da matriz Hessiana para o sistema de inferência *fuzzy non-singleton*. Já a Seção 5.3 apresenta o algoritmo do gradiente conjugado para o treinamento dos métodos em questão.

5.1 Matriz Hessiana para o sistema de inferência *fuzzy singleton*

As deduções apresentadas a seguir referem-se ao sistema de inferência *fuzzy singleton* descrito na Seção 4.5. Note que tais deduções são baseadas na equação (4.24), em que assume-se a escolha pela fuzzificação *singleton*, composição *max-product*, implicação produto e defuzzificação pela altura e função de pertinência Gaussiana.

Seja $\mathbf{w}^{(q)}$ e $\nabla J(\mathbf{w}^{(q)})$, respectivamente, o vetor linha e o gradiente dos parâmetros do sistema de inferência *fuzzy*, dados por

$$\mathbf{w}^{(q)} = \begin{bmatrix} m_{F_1^1}(q), \dots, m_{F_p^1}(q), \dots, m_{F_1^M}(q), \dots, m_{F_p^M}(q), \\ \sigma_{F_1^1}(q), \dots, \sigma_{F_p^1}(q), \dots, \sigma_{F_1^M}(q), \dots, \sigma_{F_p^M}(q), \\ \theta_1(q), \dots, \theta_M(q) \end{bmatrix}, \quad (5.1)$$

$$\begin{aligned} \nabla J(\mathbf{w}^{(q)}) = & \\ & \left[\nabla_{m_{F_1^1}(q)} J(\mathbf{w}^{(q)}), \dots, \nabla_{m_{F_p^1}(q)} J(\mathbf{w}^{(q)}), \dots, \nabla_{m_{F_1^M}(q)} J(\mathbf{w}^{(q)}), \dots, \nabla_{m_{F_p^M}(q)} J(\mathbf{w}^{(q)}), \right. \\ & \nabla_{\sigma_{F_1^1}(q)} J(\mathbf{w}^{(q)}), \dots, \nabla_{\sigma_{F_p^1}(q)} J(\mathbf{w}^{(q)}), \dots, \nabla_{\sigma_{F_1^M}(q)} J(\mathbf{w}^{(q)}), \dots, \nabla_{\sigma_{F_p^M}(q)} J(\mathbf{w}^{(q)}), \\ & \left. \nabla_{\theta_1(q)} J(\mathbf{w}^{(q)}), \dots, \nabla_{\theta_M(q)} J(\mathbf{w}^{(q)}) \right], \end{aligned} \quad (5.2)$$

em que q representa a q -ésima iteração.

Recapitulando a função de custo apresentada na equação (4.38), temos que

$$\begin{aligned} J(\mathbf{w}^{(q)}) &= e^{(q)} \\ &= \frac{1}{2} [f_s(\mathbf{x}^{(q)}) - y^{(q)}]^2. \end{aligned} \quad (5.3)$$

Note que utilizando o algoritmo de retropropagação para minimizar $e^{(q)}$, podemos facilmente obter a estrutura do gradiente conjugado de primeira ordem através da derivação de $J(\mathbf{w}^{(q)})$ em função dos parâmetros $m_{F_k^l}(q)$, $\theta_l(q)$ e $\sigma_{F_k^l}(q)$, conforme a seguir

$$\nabla_{m_{F_k^l}(q)} J(\mathbf{w}^{(q)}) = [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q), \quad (5.4)$$

$$\nabla_{\sigma_{F_k^l}(q)} J(\mathbf{w}^{(q)}) = [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \quad (5.5)$$

e

$$\nabla_{\theta_l(q)} J(\mathbf{w}^{(q)}) = [f_s(\mathbf{x}^{(q)}) - y^{(q)}] \phi_l(\mathbf{x}^{(q)}), \quad (5.6)$$

em que $l = 1, \dots, M$ e $k = 1, \dots, p$. Note que

$$a_{F_k^l}(q) = \frac{x_k^{(q)} - m_{F_k^l}(q)}{\sigma_{F_k^l}^2(q)}, \quad (5.7)$$

$$b_{F_k^l}(q) = \frac{\left(x_k^{(q)} - m_{F_k^l}(q)\right)^2}{\sigma_{F_k^l}^3(q)}. \quad (5.8)$$

Para simplificar as deduções que serão apresentadas, adotou-se a seguinte nomenclatura

$$e^{(q)} = f_s(\mathbf{x}^{(q)}) - y^{(q)}, \quad (5.9)$$

$$e_{\theta_l(q)}^{(q)} = \theta_l(q) - f_s(\mathbf{x}^{(q)}). \quad (5.10)$$

A matriz Hessiana é uma matriz quadrada composta pelas derivadas parciais de segunda ordem de uma função. Ou seja, descreve a curvatura local de uma função de várias variáveis. Ser regular e definida positiva são propriedades necessárias à matriz Hessiana para se aplicar algoritmos derivativos de segunda ordem estáveis (WILLE, 1997). É válido ressaltar que uma matriz regular é uma matriz quadrada que possui inversa. Já uma matriz é definida positiva se, e somente se, todos os seus autovalores são positivos.

Sabemos que o algoritmo do gradiente conjugado faz uso da matriz Hessiana $\mathbf{H}(J(\mathbf{w}^{(q)}))$. Para obtenção da mesma, basta derivarmos o gradiente $\nabla J(\mathbf{w}^{(q)})$, apresentado na equação (5.2), em função dos parâmetros do sistema de inferência *fuzzy* $\mathbf{w}^{(q)}$, apresentado na equação (5.1).

Com isso, obtém-se a seguinte estrutura

$$\mathbf{H}(J(\mathbf{w}^{(q)})) = \begin{bmatrix} \mathbf{H}_{m_{F_j^i}(q), m_{F_k^l}(q)} & \mathbf{H}_{m_{F_j^i}(q), \sigma_{F_k^l}(q)} & \mathbf{H}_{m_{F_j^i}(q), \theta_l(q)} \\ \mathbf{H}_{\sigma_{F_j^i}(q), m_{F_k^l}(q)} & \mathbf{H}_{\sigma_{F_j^i}(q), \sigma_{F_k^l}(q)} & \mathbf{H}_{\sigma_{F_j^i}(q), \theta_l(q)} \\ \mathbf{H}_{\theta_i(q), m_{F_k^l}(q)} & \mathbf{H}_{\theta_i(q), \sigma_{F_k^l}(q)} & \mathbf{H}_{\theta_i(q), \theta_l(q)} \end{bmatrix}. \quad (5.11)$$

Note que os termos da matriz Hessiana que representam as derivadas do termo $m_{F_j^i}(q)$ em relação aos gradientes $\nabla_{m_{F_k^l}(q)} J(\mathbf{w}^{(q)})$, $\nabla_{\sigma_{F_k^l}(q)} J(\mathbf{w}^{(q)})$ e $\nabla_{\theta_l(q)} J(\mathbf{w}^{(q)})$ são definidos, respectivamente, nas equações (5.12) - (5.14)

$$\mathbf{H}_{m_{F_j^i}(q), m_{F_k^l}(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}(q))}{\partial m_{F_1^1}^2(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}(q))}{\partial m_{F_1^1}(q) \partial m_{F_p^M}(q)} \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 J(\mathbf{w}(q))}{\partial m_{F_p^M}(q) \partial m_{F_1^1}(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}(q))}{\partial m_{F_p^M}^2(q)} \end{bmatrix}, \quad (5.12)$$

$$\mathbf{H}_{m_{F_j^i}(q), \sigma_{F_k^l}(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}(q))}{\partial m_{F_1^1}(q) \partial \sigma_{F_1^1}(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}(q))}{\partial m_{F_1^1}(q) \partial \sigma_{F_p^M}(q)} \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 J(\mathbf{w}(q))}{\partial m_{F_p^M}(q) \partial \sigma_{F_1^1}(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}(q))}{\partial m_{F_p^M}(q) \partial \sigma_{F_p^M}(q)} \end{bmatrix} \quad (5.13)$$

e

$$\mathbf{H}_{m_{F_j^i}(q), \theta_l(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}(q))}{\partial m_{F_1^1}(q) \partial \theta_1(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}(q))}{\partial m_{F_1^1}(q) \partial \theta_M(q)} \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 J(\mathbf{w}(q))}{\partial m_{F_p^M}(q) \partial \theta_1(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}(q))}{\partial m_{F_p^M}(q) \partial \theta_M(q)} \end{bmatrix}. \quad (5.14)$$

Já os termos da matriz Hessiana que representam as derivadas do termo $\sigma_{F_j^i}(q)$ em relação aos gradientes $\nabla_{m_{F_k^l}(q)} J(\mathbf{w}(q))$, $\nabla_{\sigma_{F_k^l}(q)} J(\mathbf{w}(q))$ e $\nabla_{\theta_l(q)} J(\mathbf{w}(q))$ são definidos, respectivamente, nas equações (5.15) - (5.17)

$$\mathbf{H}_{\sigma_{F_j^i}(q), m_{F_k^l}(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}(q))}{\partial \sigma_{F_1^1}(q) \partial m_{F_1^1}(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}(q))}{\partial \sigma_{F_1^1}(q) \partial m_{F_p^M}(q)} \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 J(\mathbf{w}(q))}{\partial \sigma_{F_p^M}(q) \partial m_{F_1^1}(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}(q))}{\partial \sigma_{F_p^M}(q) \partial m_{F_p^M}(q)} \end{bmatrix}, \quad (5.15)$$

$$\mathbf{H}_{\sigma_{F_j^i}(q), \sigma_{F_k^l}(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}(q))}{\partial \sigma_{F_1^1}^2(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}(q))}{\partial \sigma_{F_1^1}(q) \partial \sigma_{F_p^M}(q)} \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 J(\mathbf{w}(q))}{\partial \sigma_{F_p^M}(q) \partial \sigma_{F_1^1}(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}(q))}{\partial \sigma_{F_p^M}^2(q)} \end{bmatrix} \quad (5.16)$$

e

$$\mathbf{H}_{\sigma_{F_j^i}(q), \theta_l(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}(q))}{\partial \sigma_{F_1^1}(q) \partial \theta_1(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}(q))}{\partial \sigma_{F_1^1}(q) \partial \theta_M(q)} \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 J(\mathbf{w}(q))}{\partial \sigma_{F_p^M}(q) \partial \theta_1(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}(q))}{\partial \sigma_{F_p^M}(q) \partial \theta_M(q)} \end{bmatrix}. \quad (5.17)$$

Por sua vez, os termos da matriz Hessiana que representam as derivadas do termo $\theta_i(q)$ em relação aos gradientes $\nabla_{m_{F_k^l}(q)} J(\mathbf{w}(q))$, $\nabla_{\sigma_{F_k^l}(q)} J(\mathbf{w}(q))$ e $\nabla_{\theta_l(q)} J(\mathbf{w}(q))$ são

definidos, respectivamente, nas equações (5.18) - (5.20)

$$\mathbf{H}_{\theta_i(q), m_{F_k^l}(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_1(q) \partial m_{F_1^l}(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_1(q) \partial m_{F_p^M}(q)} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_M(q) \partial m_{F_1^l}(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_M(q) \partial m_{F_p^M}(q)} \end{bmatrix}, \quad (5.18)$$

$$\mathbf{H}_{\theta_i(q), \sigma_{F_k^l}(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_1(q) \partial \sigma_{F_1^l}(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_1(q) \partial \sigma_{F_p^M}(q)} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_M(q) \partial \sigma_{F_1^l}(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_M(q) \partial \sigma_{F_p^M}(q)} \end{bmatrix} \quad (5.19)$$

e

$$\mathbf{H}_{\theta_i(q), \theta_l(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_1^2(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_1(q) \partial \theta_M(q)} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_M(q) \partial \theta_1(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_M^2(q)} \end{bmatrix} \quad (5.20)$$

Os termos de $\mathbf{H}_{m_{F_j^i}(q), m_{F_k^l}(q)}$, expresso na equação (5.12), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{m_{F_k^l}(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $m_{F_j^i}(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)} = \frac{\partial}{\partial m_{F_j^i}(q)} [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q). \quad (5.21)$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)} = \frac{\partial}{\partial m_{F_j^i}(q)} [XYZ]. \quad (5.22)$$

de tal forma que

$$X = [f_s(\mathbf{x}^{(q)}) - y^{(q)}], \quad (5.23)$$

$$Y = [\theta_l(q) - f_s(\mathbf{x}^{(q)})], \quad (5.24)$$

$$W = \phi_l(\mathbf{x}^{(q)}), \quad (5.25)$$

e

$$Z = a_{F_k^l}(q). \quad (5.26)$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (5.22). Com isso temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)} &= \left(\frac{\partial X}{\partial m_{F_j^i}(q)} \right) YWZ + X \left(\frac{\partial Y}{\partial m_{F_j^i}(q)} \right) WZ \\ &+ XY \left(\frac{\partial W}{\partial m_{F_j^i}(q)} \right) Z + XYW \left(\frac{\partial Z}{\partial m_{F_j^i}(q)} \right). \end{aligned} \quad (5.27)$$

Note que as derivadas parciais retratadas na equação (5.27) podem ser expressas por

$$\frac{\partial X}{\partial m_{F_j^i}(q)} = [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q), \quad \forall i, \quad (5.28)$$

$$\frac{\partial Y}{\partial m_{F_j^i}(q)} = - [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) a_{F_k^l}(q), \quad \forall i, \quad (5.29)$$

$$\frac{\partial W}{\partial m_{F_j^i}(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q), & i \neq l \\ [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] a_{F_k^l}(q), & i = l \end{cases} \quad (5.30)$$

e

$$\frac{\partial Z}{\partial m_{F_j^i}(q)} = \begin{cases} 0, & i \neq l \\ 0, & i = l, \quad j \neq k \\ \frac{-1}{\sigma_{F_k^l}^2(q)}, & i = l, \quad j = k \end{cases}. \quad (5.31)$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)}$, para a situação em que $i \neq l$, temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)} &= [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q) [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &- [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &- [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q) a_{F_k^l}(q), \end{aligned} \quad (5.32)$$

o que remete à seguinte equação

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)} &= [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q) [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &- [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &- [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q) a_{F_k^l}(q). \end{aligned} \quad (5.33)$$

Organizando os termos da equação (5.33), temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)} &= \{ [\theta_i(q) - f_s(\mathbf{x}^{(q)})] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] - [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \\ &- [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q) a_{F_k^l}(q). \end{aligned} \quad (5.34)$$

Substituindo as equações (5.9) e (5.10) em (5.34), temos a seguinte equação simplificada

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)} = \left(e_{\theta_i(q)}^{(q)} e_{\theta_l(q)}^{(q)} - e^{(q)} e_{\theta_i(q)}^{(q)} - e^{(q)} e_{\theta_l(q)}^{(q)} \right) \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q) a_{F_k^l}(q). \quad (5.35)$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)}$ para a situação em que $i = l$ temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)} &= e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_j^l}(q) e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &\quad - e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_j^l}(q) \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] a_{F_j^l}(q) a_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial Z}{\partial m_{F_j^l}(q)}, \end{aligned} \quad (5.36)$$

o que nos remete à seguinte equação

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^l}(q) \partial m_{F_k^l}(q)} &= e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_j^l}(q) e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &\quad - e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_j^l}(q) \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] a_{F_j^l}(q) a_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial Z}{\partial m_{F_j^l}(q)}, \end{aligned} \quad (5.37)$$

ou seja, (5.37) pode ser reescrita da seguinte maneira

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^l}(q) \partial m_{F_k^l}(q)} &= \left[\left(e_{\theta_l(q)}^{(q)} \right)^2 \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} e_{\theta_l(q)}^{(q)} (\phi_l(\mathbf{x}^{(q)}) - 2\phi_l^2(\mathbf{x}^{(q)})) \right] a_{F_j^l}(q) a_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial Z}{\partial m_{F_j^l}(q)}. \end{aligned} \quad (5.38)$$

Com isso, ordenando os termos da equação (5.38), temos a seguinte equação para $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)}$ quando $i = l$

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^l}(q) \partial m_{F_k^l}(q)} &= \left[\left(e_{\theta_l(q)}^{(q)} \right)^2 - 2e^{(q)} e_{\theta_l(q)}^{(q)} \right] \phi_l^2(\mathbf{x}^{(q)}) a_{F_j^l}(q) a_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \left[\frac{\partial Z}{\partial m_{F_j^l}(q)} + a_{F_j^l}(q) a_{F_k^l}(q) \right], \end{aligned} \quad (5.39)$$

em que $\frac{\partial Z}{\partial m_{F_j^l}(q)}$ assume os seguintes valores

$$\frac{\partial Z}{\partial m_{F_j^l}(q)} = \begin{cases} 0, & j \neq k \\ \frac{-1}{\sigma_{F_k^l}^2(q)}, & j = k \end{cases}. \quad (5.40)$$

Note que os demais termos da matriz Hessiana $\mathbf{H}(J(\mathbf{w}^{(q)}))$ expressa na equação

(5.11) são deduzidos e apresentados no apêndice C.

5.2 Matriz Hessiana para o sistema de inferência *fuzzy non-singleton*

As deduções apresentadas a seguir referem-se ao sistema de inferência *fuzzy non-singleton* descrito na Seção 4.6.

Seja $\mathbf{w}^{(q)}$ e $\nabla J(\mathbf{w}^{(q)})$, respectivamente, o vetor linha e o gradiente dos parâmetros do sistema de inferência *fuzzy*, dados por

$$\begin{aligned} \mathbf{w}^{(q)} = & \left[m_{F_1^1}(q), \dots, m_{F_p^1}(q), \dots, m_{F_1^M}(q), \dots, m_{F_p^M}(q), \right. \\ & \left. \sigma_{F_1^1}(q), \dots, \sigma_{F_p^1}(q), \dots, \sigma_{F_1^M}(q), \dots, \sigma_{F_p^M}(q), \right. \\ & \left. \theta_1(q), \dots, \theta_M(q), \sigma_X(q) \right], \end{aligned} \quad (5.41)$$

$$\begin{aligned} \nabla J(\mathbf{w}^{(q)}) = & \left[\nabla_{m_{F_1^1}(q)} J(\mathbf{w}^{(q)}), \dots, \nabla_{m_{F_p^1}(q)} J(\mathbf{w}^{(q)}), \dots, \nabla_{m_{F_1^M}(q)} J(\mathbf{w}^{(q)}), \dots, \nabla_{m_{F_p^M}(q)} J(\mathbf{w}^{(q)}), \right. \\ & \nabla_{\sigma_{F_1^1}(q)} J(\mathbf{w}^{(q)}), \dots, \nabla_{\sigma_{F_p^1}(q)} J(\mathbf{w}^{(q)}), \dots, \nabla_{\sigma_{F_1^M}(q)} J(\mathbf{w}^{(q)}), \dots, \nabla_{\sigma_{F_p^M}(q)} J(\mathbf{w}^{(q)}), \\ & \left. \nabla_{\theta_1(q)} J(\mathbf{w}^{(q)}), \dots, \nabla_{\theta_M(q)} J(\mathbf{w}^{(q)}), \nabla_{\sigma_X(q)} J(\mathbf{w}^{(q)}) \right]. \end{aligned} \quad (5.42)$$

Recapitulando a função de custo apresentada na equação (4.38), temos que

$$\begin{aligned} J(\mathbf{w}^{(q)}) &= e^{(q)} \\ &= \frac{1}{2} [f_s(\mathbf{x}^{(q)}) - y^{(q)}]^2. \end{aligned} \quad (5.43)$$

Note que utilizando o algoritmo de retropropagação para minimizar $e^{(q)}$, podemos facilmente obter a estrutura do gradiente conjugado de primeira ordem através da derivação de $J(\mathbf{w}^{(q)})$ em função dos parâmetros $m_{F_k^l}(q)$, $\theta_l(q)$, $\sigma_{F_k^l}(q)$ e $\sigma_X(q)$ conforme a seguir

$$\nabla_{m_{F_k^l}(q)} J(\mathbf{w}^{(q)}) = [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q), \quad (5.44)$$

$$\nabla_{\sigma_{F_k^l}(q)} J(\mathbf{w}^{(q)}) = [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q), \quad (5.45)$$

$$\nabla_{\theta_l(q)} J(\mathbf{w}^{(q)}) = [f_s(\mathbf{x}^{(q)}) - y^{(q)}] \phi_l(\mathbf{x}^{(q)}) \quad (5.46)$$

e

$$\nabla_{\sigma_X(q)} J(\mathbf{w}^{(q)}) = [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_l(\mathbf{x}^{(q)}) c_{F_k^l}(q), \quad (5.47)$$

em que $l = 1, \dots, M$ e $k = 1, \dots, p$. Note que

$$a_{F_k^l}(q) = \frac{x_k^{(q)} - m_{F_k^l}(q)}{\sigma_{F_k^l}^2(q) + \sigma_{X_k}^2(q)}, \quad (5.48)$$

$$b_{F_k^l}(q) = \frac{\left(x_k^{(q)} - m_{F_k^l}(q)\right)^2}{\left(\sigma_{F_k^l}^2(q) + \sigma_{X_k}^2(q)\right)^2} \sigma_{F_k^l}(q) \quad (5.49)$$

e

$$c_{F_k^l}(q) = \frac{\left(x_k^{(q)} - m_{F_k^l}(q)\right)^2}{\left(\sigma_{F_k^l}^2(q) + \sigma_{X_k}^2(q)\right)^2} \sigma_{X_k}(q), \quad (5.50)$$

Para simplificar as deduções que serão apresentadas, adotou-se a seguinte nomenclatura

$$e^{(q)} = f_s(\mathbf{x}^{(q)}) - y^{(q)}, \quad (5.51)$$

$$e_{\theta_l^{(q)}}^{(q)} = \theta_l(q) - f_s(\mathbf{x}^{(q)}). \quad (5.52)$$

E assim, tal como apresentado na Seção 5.1, obtém-se a seguinte estrutura

$$\mathbf{H}(J(\mathbf{w}^{(q)})) = \begin{bmatrix} \mathbf{H}_{m_{F_j^i}(q), m_{F_k^l}(q)} & \mathbf{H}_{m_{F_j^i}(q), \sigma_{F_k^l}(q)} & \mathbf{H}_{m_{F_j^i}(q), \theta_l(q)} & \mathbf{H}_{m_{F_j^i}(q), \sigma_X(q)} \\ \mathbf{H}_{\sigma_{F_j^i}(q), m_{F_k^l}(q)} & \mathbf{H}_{\sigma_{F_j^i}(q), \sigma_{F_k^l}(q)} & \mathbf{H}_{\sigma_{F_j^i}(q), \theta_l(q)} & \mathbf{H}_{\sigma_{F_j^i}(q), \sigma_X(q)} \\ \mathbf{H}_{\theta_l(q), m_{F_k^l}(q)} & \mathbf{H}_{\theta_l(q), \sigma_{F_k^l}(q)} & \mathbf{H}_{\theta_l(q), \theta_l(q)} & \mathbf{H}_{\theta_l(q), \sigma_X(q)} \\ \mathbf{H}_{\sigma_X(q), m_{F_k^l}(q)} & \mathbf{H}_{\sigma_X(q), \sigma_{F_k^l}(q)} & \mathbf{H}_{\sigma_X(q), \theta_l(q)} & \mathbf{H}_{\sigma_X(q), \sigma_X(q)} \end{bmatrix}. \quad (5.53)$$

Note que os termos da matriz Hessiana que representam as derivadas do termo $m_{F_j^i}(q)$ em relação aos gradientes $\nabla_{m_{F_k^l}(q)} J(\mathbf{w}^{(q)})$, $\nabla_{\sigma_{F_k^l}(q)} J(\mathbf{w}^{(q)})$, $\nabla_{\theta_l(q)} J(\mathbf{w}^{(q)})$ e $\nabla_{\sigma_{X_k}(q)} J(\mathbf{w}^{(q)})$ são definidos, respectivamente, nas equações (5.54) - (5.57)

$$\mathbf{H}_{m_{F_j^i}(q), m_{F_k^l}(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_1^l}^2(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_1^l}(q) \partial m_{F_p^M}(q)} \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_p^M}(q) \partial m_{F_1^l}(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_p^M}^2(q)} \end{bmatrix}, \quad (5.54)$$

$$\mathbf{H}_{m_{F_j^i}(q), \sigma_{F_k^l}(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_1^l}(q) \partial \sigma_{F_1^l}(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_1^l}(q) \partial \sigma_{F_p^M}(q)} \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_p^M}(q) \partial \sigma_{F_1^l}(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_p^M}(q) \partial \sigma_{F_p^M}(q)} \end{bmatrix}, \quad (5.55)$$

$$\mathbf{H}_{m_{F_j^i}(q), \theta_l(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_1^1}(q) \partial \theta_1(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_1^1}(q) \partial \theta_M(q)} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_p^M}(q) \partial \theta_1(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_p^M}(q) \partial \theta_M(q)} \end{bmatrix}, \quad (5.56)$$

e

$$\mathbf{H}_{m_{F_j^i}(q), \sigma_X(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_1^1}(q) \partial \sigma_X(q)} \\ \vdots \\ \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_p^M}(q) \partial \sigma_X(q)} \end{bmatrix}. \quad (5.57)$$

Já os termos da matriz Hessiana que representam as derivadas do termo $\sigma_{F_j^i}(q)$ em relação aos gradientes $\nabla_{m_{F_k^l}(q)} J(\mathbf{w}^{(q)})$, $\nabla_{\sigma_{F_k^l}(q)} J(\mathbf{w}^{(q)})$, $\nabla_{\theta_l(q)} J(\mathbf{w}^{(q)})$ e $\nabla_{\sigma_X(q)} J(\mathbf{w}^{(q)})$ são definidos, respectivamente, nas equações (5.58) - (5.61)

$$\mathbf{H}_{\sigma_{F_j^i}(q), m_{F_k^l}(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_1^1}(q) \partial m_{F_1^1}(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_1^1}(q) \partial m_{F_p^M}(q)} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_p^M}(q) \partial m_{F_1^1}(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_p^M}(q) \partial m_{F_p^M}(q)} \end{bmatrix}, \quad (5.58)$$

$$\mathbf{H}_{\sigma_{F_j^i}(q), \sigma_{F_k^l}(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_1^1}^2(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_1^1}(q) \partial \sigma_{F_p^M}(q)} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_p^M}(q) \partial \sigma_{F_1^1}(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_p^M}^2(q)} \end{bmatrix}, \quad (5.59)$$

$$\mathbf{H}_{\sigma_{F_j^i}(q), \theta_l(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_1^1}(q) \partial \theta_1(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_1^1}(q) \partial \theta_M(q)} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_p^M}(q) \partial \theta_1(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_p^M}(q) \partial \theta_M(q)} \end{bmatrix} \quad (5.60)$$

e

$$\mathbf{H}_{\sigma_{F_j^i}(q), \sigma_X(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_1^1}(q) \partial \sigma_X(q)} \\ \vdots \\ \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_p^M}(q) \partial \sigma_X(q)} \end{bmatrix}. \quad (5.61)$$

Por sua vez, os termos da matriz Hessiana que representam as derivadas do termo $\theta_l(q)$ em relação aos gradientes $\nabla_{m_{F_k^l}(q)} J(\mathbf{w}^{(q)})$, $\nabla_{\sigma_{F_k^l}(q)} J(\mathbf{w}^{(q)})$, $\nabla_{\theta_l(q)} J(\mathbf{w}^{(q)})$ e $\nabla_{\sigma_{X_k}(q)} J(\mathbf{w}^{(q)})$ são definidos, respectivamente, nas equações (5.62) - (5.65)

$$\mathbf{H}_{\theta_i(q), m_{F_k^l}(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_1(q) \partial m_{F_1^l}(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_1(q) \partial m_{F_p^M}(q)} \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_M(q) \partial m_{F_1^l}(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_M(q) \partial m_{F_p^M}(q)} \end{bmatrix}, \quad (5.62)$$

$$\mathbf{H}_{\theta_i(q), \sigma_{F_k^l}(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_1(q) \partial \sigma_{F_1^l}(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_1(q) \partial \sigma_{F_p^M}(q)} \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_M(q) \partial \sigma_{F_1^l}(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_M(q) \partial \sigma_{F_p^M}(q)} \end{bmatrix}, \quad (5.63)$$

$$\mathbf{H}_{\theta_i(q), \theta_l(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_1^2(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_1(q) \partial \theta_M(q)} \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_M(q) \partial \theta_1(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_M^2(q)} \end{bmatrix} \quad (5.64)$$

e

$$\mathbf{H}_{\theta_i(q), \sigma_X(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_1(q) \partial \sigma_X(q)} \\ \vdots \\ \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_M(q) \partial \sigma_X(q)} \end{bmatrix}. \quad (5.65)$$

Já os termos da matriz Hessiana que representam as derivadas do termo $\sigma_X(q)$ em relação aos gradientes $\nabla_{m_{F_k^l}(q)} J(\mathbf{w}^{(q)})$, $\nabla_{\sigma_{F_k^l}(q)} J(\mathbf{w}^{(q)})$, $\nabla_{\theta_l(q)} J(\mathbf{w}^{(q)})$ e $\nabla_{\sigma_X(q)} J(\mathbf{w}^{(q)})$ são definidos, respectivamente, nas equações (5.66) - (5.69)

$$\mathbf{H}_{\sigma_X(q), m_{F_k^l}(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial m_{F_1^l}(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial m_{F_p^M}(q)} \end{bmatrix}, \quad (5.66)$$

$$\mathbf{H}_{\sigma_X(q), \sigma_{F_k^l}(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \sigma_{F_1^l}(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \sigma_{F_p^M}(q)} \end{bmatrix}, \quad (5.67)$$

$$\mathbf{H}_{\sigma_X(q), \theta_l(q)} = \begin{bmatrix} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \theta_1(q)} & \cdots & \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \theta_M(q)} \end{bmatrix}, \quad (5.68)$$

e

$$\mathbf{H}_{\sigma_X(q), \sigma_X(q)} = \left[\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X^2(q)} \right]. \quad (5.69)$$

Os termos de $\mathbf{H}_{m_{F_j^i}(q), m_{F_k^l}(q)}$, expresso na equação (5.54), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{m_{F_k^l}(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro

$m_{F_j^i}(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)} = \frac{\partial}{\partial m_{F_j^i}(q)} [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q). \quad (5.70)$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)} = \frac{\partial}{\partial m_{F_j^i}(q)} [XYZW]. \quad (5.71)$$

de tal forma que

$$X = [f_s(\mathbf{x}^{(q)}) - y^{(q)}], \quad (5.72)$$

$$Y = [\theta_l(q) - f_s(\mathbf{x}^{(q)})], \quad (5.73)$$

$$W = \phi_l(\mathbf{x}^{(q)}), \quad (5.74)$$

e

$$Z = a_{F_k^l}(q). \quad (5.75)$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (5.71). Com isso temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)} &= \left(\frac{\partial X}{\partial m_{F_j^i}(q)} \right) YWZ + X \left(\frac{\partial Y}{\partial m_{F_j^i}(q)} \right) WZ \\ &+ XY \left(\frac{\partial W}{\partial m_{F_j^i}(q)} \right) Z + XYW \left(\frac{\partial Z}{\partial m_{F_j^i}(q)} \right). \end{aligned} \quad (5.76)$$

Note que as derivadas parciais retratadas na equação (5.76) podem ser expressas por

$$\frac{\partial X}{\partial m_{F_j^i}(q)} = [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q), \quad \forall i, \quad (5.77)$$

$$\frac{\partial Y}{\partial m_{F_j^i}(q)} = -[\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) a_{F_k^l}(q), \quad \forall i, \quad (5.78)$$

$$\frac{\partial W}{\partial m_{F_j^i}(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q), & i \neq l \\ [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] a_{F_k^l}(q), & i = l \end{cases} \quad (5.79)$$

e

$$\frac{\partial Z}{\partial m_{F_j^i}(q)} = \begin{cases} 0, & i \neq l \\ 0, & i = l, \quad j \neq k \\ \frac{-1}{\sigma_{F_k^l}^2(q) + \sigma_{X_k}^2(q)}, & i = l, \quad j = k \end{cases}. \quad (5.80)$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)}$, para a situação em que $i \neq l$, temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)} &= [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q) [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &\quad - [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &\quad - [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q) a_{F_k^l}(q), \end{aligned} \quad (5.81)$$

o que remete à seguinte equação

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)} &= [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q) [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &\quad - [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &\quad - [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q) a_{F_k^l}(q). \end{aligned} \quad (5.82)$$

Organizando os termos da equação (5.82), temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)} &= \{ [\theta_i(q) - f_s(\mathbf{x}^{(q)})] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] - [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \\ &\quad - [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q) a_{F_k^l}(q). \end{aligned} \quad (5.83)$$

Substituindo as equações (5.51) e (5.52) em (5.83), temos a seguinte equação simplificada

$$\frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)} = \left(e_{\theta_i(q)}^{(q)} e_{\theta_l(q)}^{(q)} - e^{(q)} e_{\theta_i(q)}^{(q)} - e^{(q)} e_{\theta_l(q)}^{(q)} \right) \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q) a_{F_k^l}(q). \quad (5.84)$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)}$ para a situação em que $i = l$ temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)} &= e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q) e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &\quad - e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] a_{F_j^i}(q) a_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial Z}{\partial m_{F_j^i}(q)}, \end{aligned} \quad (5.85)$$

o que nos remete à seguinte equação

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)} &= e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q) e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &\quad - e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] a_{F_j^i}(q) a_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial Z}{\partial m_{F_j^i}(q)}, \end{aligned} \quad (5.86)$$

ou seja, (5.85) pode ser reescrita da seguinte maneira

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^l}(q) \partial m_{F_k^l}(q)} &= \left[\left(e_{\theta_l(q)}^{(q)} \right)^2 \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} e_{\theta_l(q)}^{(q)} \left(\phi_l(\mathbf{x}^{(q)}) - 2\phi_l^2(\mathbf{x}^{(q)}) \right) \right] a_{F_j^l}(q) a_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial Z}{\partial m_{F_j^l}(q)}. \end{aligned} \quad (5.87)$$

Com isso, ordenando os termos da equação (5.87), temos a seguinte equação para $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial m_{F_k^l}(q)}$ quando $i = l$

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^l}(q) \partial m_{F_k^l}(q)} &= \left[\left(e_{\theta_l(q)}^{(q)} \right)^2 - 2e^{(q)} e_{\theta_l(q)}^{(q)} \right] \phi_l^2(\mathbf{x}^{(q)}) a_{F_j^l}(q) a_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \left[\frac{\partial Z}{\partial m_{F_j^l}(q)} + a_{F_j^l}(q) a_{F_k^l}(q) \right], \end{aligned} \quad (5.88)$$

em que $\frac{\partial Z}{\partial m_{F_j^l}(q)}$ assume os seguintes valores

$$\frac{\partial Z}{\partial m_{F_j^l}(q)} = \begin{cases} 0, & j \neq k \\ \frac{-1}{\sigma_{F_k^l}^2(q) + \sigma_{X_k}^2(q)}, & j = k \end{cases}. \quad (5.89)$$

Note que os demais termos da matriz Hessiana $\mathbf{H}(J(\mathbf{w}^{(q)}))$ expressa na equação (5.53) são deduzidos e apresentados no apêndice C.

5.3 Método do gradiente conjugado para sistemas de inferência *fuzzy*

Métodos de gradiente conjugado (GC) pertencem a uma classe de algoritmos de otimização que são caracterizados por requererem baixa memória e por possuírem propriedades de convergência local e global fortes (HAGER; ZHANG, 2006). Note que tais métodos fazem uso da matriz Hessiana.

As matrizes Hessianas são utilizadas em larga escala em problemas de otimização de métodos de Newton por serem o coeficiente do termo quadrático de uma expansão de Taylor local de uma dada função (ZENG; LI; ZHANG, 2009). As aplicações dessas matrizes abrangem equalização cega e recuperação da fase da portadora em aplicações de telecomunicações, além de otimização de circuitos (BAEK; PARK, 2009) e treinamento de redes neurais não-lineares (YU; MANRY, 2004).

Baseando no roteiro proposto por (LUENBERGER; YE, 2008), o algoritmo que representa a técnica utilizada tem os seus passos descritos abaixo.

1. $q=0$;
2. Inicializar $\mathbf{w}^{(q)}$, $\mathbf{g}(q)$, $\mathbf{d}(q)$ e μ :
 - $\mathbf{g}(0) = -\nabla J(\mathbf{w}^{(0)})$;
 - $\mathbf{d}(0) = -\mathbf{g}(0)$;
3. $q = q + 1$;
4. Calcular $\mathbf{w}^{(q+1)} = \mathbf{w}^{(q)} + \mu\mathbf{d}(q)$;
5. Calcular $\mathbf{g}(q + 1) = -\nabla J(\mathbf{w}^{(q+1)})$;
6. Calcular $\beta(q)$;
7. Verificar a condição:
 - $0 < \beta(q) < \beta_{\max}(q)$

$$\mathbf{d}(q + 1) = -\mathbf{g}(q + 1) + \beta(q)\mathbf{d}(q)$$
 - Senão:

$$\mathbf{d}(q + 1) = -\mathbf{g}(q + 1)$$
8. Voltar ao passo 3.

Note que μ é o passo de adaptação. A variável β é alvo de diversos estudos, dada sua importância para a convergência do algoritmo. Através de pesquisa na literatura, encontram-se 8 maneiras distintas de se calcular esta variável.

Os betas relatados na literatura são (HAGER; ZHANG, 2006)

$$\beta^{HS}(q) = \frac{\mathbf{g}^T(q+1)\zeta(q)}{\mathbf{d}^T(q)\zeta(q)}, \quad (5.90)$$

$$\beta^{FR}(q) = \frac{\|\mathbf{g}^T(q+1)\|^2}{\|\mathbf{g}^T(q)\|^2}, \quad (5.91)$$

$$\beta^{PRP}(q) = \frac{\mathbf{g}^T(q+1)\zeta(q)}{\|\mathbf{g}(q)\|^2}, \quad (5.92)$$

$$\beta^{CD}(q) = \frac{\|\mathbf{g}^T(q+1)\|^2}{-\mathbf{d}^T(q)\mathbf{g}(q)}, \quad (5.93)$$

$$\beta^{LS}(q) = \frac{\mathbf{g}^T(q+1)\zeta(q)}{-\mathbf{d}^T(q)\mathbf{g}(q)}, \quad (5.94)$$

$$\beta^{DY}(q) = \frac{\|\mathbf{g}^T(q+1)\|^2}{\mathbf{d}^T(q)\zeta(q)}, \quad (5.95)$$

$$\beta^N(q) = \left(\zeta(q) - 2\mathbf{d}(q) \frac{\|\zeta(q)\|^2}{\mathbf{d}^T(q)\zeta(q)} \right)^T \frac{\mathbf{g}(q+1)}{\mathbf{d}^T(q)\zeta(q)}, \quad (5.96)$$

$$\beta^D(q) = \frac{\mathbf{g}^T(q+1)\mathbf{H}(J(\mathbf{w}^{(q)}))\mathbf{d}(q)}{\mathbf{d}^T(q)\mathbf{H}(J(\mathbf{w}^{(q)}))\mathbf{d}(q)}, \quad (5.97)$$

em que $\zeta(q) = \mathbf{g}(q+1) - \mathbf{g}(q)$. Note que o índice *HS* remete a Hestenes e Stiefel (HESTENES; STIEFEL, 1952), *FR* a Fletcher e Reeves (FLETCHER; REEVES, 1964), *PRP* a Polak, Ribière e Polyak (POLAK; RIBIÈRE, 1969; POLYAK, 1969), *CD* ao algoritmo de descida conjugada (do inglês *Conjugate Descent*), proposto por Fletcher (FLETCHER, 1987), *LS* a Liu e Storey (LIU; STOREY, 1991), *DY* a Dai e Yuan (DAI Y. H.; YUAN, 1999), *N* a Hager e Zhang (HAGER; ZHANG, 2003) e *D* a Daniel (DANIEL, 1967).

Se a função f a ser otimizada for fortemente quadrática e convexa, então, teoricamente, as 8 escolhas possíveis para β serão equivalentes. Para funções de custo não quadráticas, cada β escolhido leva a diferentes desempenhos. Atualmente, alguns dos algoritmos de GC de melhor desempenho são híbridos, que ajustam a fórmula de atualização de β de acordo com a evolução da iteração (HAGER; ZHANG, 2006).

Para problemas de larga escala, as escolhas de atualização do β que não necessitam do cálculo da matriz Hessiana são normalmente preferidos na prática em relação aos métodos que exigem a Hessiana em cada iteração (HAGER; ZHANG, 2006).

Se $\beta(q) = \beta^D(q)$, o algoritmo possui o seguinte comportamento

1. $q=0$;
2. Inicializar $\mathbf{w}^{(q)}$, $\mathbf{g}(q)$, $\mathbf{d}(q)$ e μ :
 - $\mathbf{g}(0) = -\nabla J(\mathbf{w}^{(0)})$;
 - $\mathbf{d}(0) = -\mathbf{g}(0)$;
3. $q = q + 1$;
4. Calcular $\mathbf{g}(q+1) = -\nabla J(\mathbf{w}^{(q+1)})$;
5. Calcular a matriz Hessiana $\mathbf{H}(J(\mathbf{w}^{(q)}))$;
6. Calcular $\beta^D(q) = \frac{\mathbf{g}^T(q+1)\mathbf{H}(J(\mathbf{w}^{(q)}))\mathbf{d}(q)}{\mathbf{d}^T(q)\mathbf{H}(J(\mathbf{w}^{(q)}))\mathbf{d}(q)}$;
7. Verificar a condição:

- $0 < \beta^D(q) < \beta_{\max}(q)$

$$\mathbf{d}(q+1) = -\mathbf{g}(q+1) + \beta^D(q) \mathbf{d}(q);$$

- Senão:

$$\mathbf{d}(q+1) = -\mathbf{g}(q+1);$$

$$8. \text{ Calcular } \alpha(q) = \frac{-\mathbf{g}^T(q)\mathbf{d}(q)}{\mathbf{d}^T(q)\mathbf{H}(J(\mathbf{w}^{(q)}))\mathbf{d}(q)};$$

9. Verificar a condição:

- Se $0 < \alpha(q) < \alpha_{\max}(q)$:

$$\mathbf{w}^{(q+1)} = \mathbf{w}^{(q)} + \alpha(q)\mathbf{d}(q);$$

- Senão:

$$\mathbf{w}^{(q+1)} = \mathbf{w}^{(q)} + \mu\mathbf{d}(q);$$

10. Voltar ao passo 3.

De acordo com (SANTOS; ZUBEN, 2000), quando a função a ser minimizada é não linear, a computação de α e β requer o cálculo da matriz Hessiana a cada nova iteração do algoritmo. Visando um melhor desempenho do algoritmo de treinamento em questão, optou-se por restringir o valor de $\alpha(q)$ e $\beta(q)$ em uma faixa de valores previamente testada, de forma a evitar que o referido algoritmo de treinamento sofra atualizações inapropriadas e conseqüentemente gere resultados indesejados. Além disso, um problema de preocupação maior é que o caráter positivo definido da matriz Hessiana pode mudar de uma iteração para outra. É importante ressaltar a ocorrência da matriz Hessiana no denominador da expressão de obtenção de α : se, em certa iteração do processo de convergência a matriz Hessiana for negativa definida, então é provável que α será negativo, resultando em um passo na direção de aumentar a função custo, ao invés de diminuí-la, conforme o esperado.

Nos métodos de primeira ordem são usados os valores da função objetivo e de suas derivadas (gradientes) em relação às variáveis de projeto. Entre os métodos mais clássicos existentes temos o método do gradiente, também conhecido como método da descida mais íngreme (do inglês *Steepest Descent*) e o método dos gradientes conjugados. (NOCEDAL; WRIGHT, 1999), (FLETCHER, 1987)

Já nos métodos de segunda ordem são utilizados os valores da função objetivo, de suas derivadas e também da matriz Hessiana. Entre os métodos existentes, destacam-se

os métodos de Newton e Quasi-Newton. A rapidez na convergência do resultado aumenta do primeiro para o segundo método.

O algoritmo baseado em métodos quasi-Newton que foi utilizado nessa dissertação tem os passos descritos abaixo:

1. $q=0$;
2. Inicializar $\mathbf{w}^{(q)}$ e μ ;
3. Calcular o gradiente $\nabla J(\mathbf{w}^{(q)})$;
4. Calcular a matriz Hessiana $\mathbf{H}(J(\mathbf{w}^{(q)}))$;
5. Calcular $\mathbf{w}^{(q+1)} = \mathbf{w}^{(q)} - \mu [\mathbf{H}(J(\mathbf{w}^{(q)}))]^{-1} \nabla J(\mathbf{w}^{(q)})$;
6. $q = q + 1$;
7. Voltar ao passo 3.

Note que $[\mathbf{H}(J(\mathbf{w}^{(q)}))]^{-1}$ é a inversa de $\mathbf{H}(J(\mathbf{w}^{(q)}))$. Em problemas de otimização, métodos Quasi-Newton são algoritmos para encontrar máximos e mínimos locais de funções. Os mesmos são baseados no método de Newton para encontrar o ponto estacionário de uma função, onde o gradiente é zero e assumem que a função pode ser localmente aproximada como uma função quadrática na região em torno do ponto subótimo ideal. Note que tal método utiliza informações de primeira e segunda ordem (gradiente e matriz Hessiana) para encontrar o ponto estacionário (FLETCHER, 1987).

5.4 Sumário

O presente capítulo apresentou a descrição matemática do algoritmo de treinamento do gradiente conjugado para sistemas de inferência *fuzzy singleton* e *non-singleton*.

Para a obtenção do referido algoritmo, utilizamos a matriz Hessiana $\mathbf{H}(J(\mathbf{w}^{(q)}))$, cuja formulação foi demonstrada e analisada de forma detalhada nessa dissertação, podendo ser entendido como evoluções e adaptações de técnicas disponíveis na literatura.

Com o objetivo de analisar qual das técnicas de classificação de distúrbios de QEE possuem melhor desempenho, o próximo capítulo traz diversos resultados numéricos obtidos a partir de simulações computacionais.

6 RESULTADOS EXPERIMENTAIS

Neste capítulo são apresentados e discutidos os resultados de simulações computacionais que foram obtidos para as situações retratadas no Capítulo 3. Inicialmente, a Seção 6.1 abordará a convergência da técnica proposta no Capítulo 5 e fará uma comparação com as técnicas de classificação de distúrbios citadas no Capítulos 2. A seguir, a Seção 6.2 aborda as taxas de classificação obtidas para o sistema de inferência *fuzzy singleton* e *non-singleton* a partir da implementação dos algoritmos propostos no Capítulo 5 e novamente discute uma comparação de tais resultados com as taxas de classificação de distúrbios com os algoritmos propostos no Capítulo 5 e aqueles descritos na Seções 2.2.3.1 e 2.2.3.2 do Capítulo 2.

Foram executadas simulações computacionais a partir de bancos de dados sintéticos gerados conforme citado no Anexo B, conforme recomendado em (QUALITY, 1995). Os bancos de dados gerados, os quais representam as classes de distúrbios listadas na tabela 1, foram gerados com $f_{amostragem} = 256 \times 60$ Hz, $N = 1024$, ou seja, o vetor de dados corresponde a quatro ciclos da componente fundamental. A tabela 1, além de listar os distúrbios isolados e múltiplos analisados, tem como objetivo de resumir diversas combinações de distúrbios que podem ocorrer em sinais elétricos. Apesar desta lista de combinações de eventos não ser completa, note que a mesma é representativa para ilustrar a complexidade envolvida ao problema de classificação de distúrbios.

Para a geração dos vetores \mathbf{f} e \mathbf{u} , a partir dos quais um conjunto reduzido de características são extraídos, optou-se, nesta contribuição pelo uso de um filtro *notch* resposta ao impulso infinita, do inglês, *infinite impulse response* (IIR) digital de 2ª ordem sintonizado na frequência $\omega_0 = 2\pi \times 60/f_s$ rad/s, ou seja, na frequência da componente fundamental, e com um fator *notch* $\rho = 0,98$.

Apesar de tal filtro apresentar inconvenientes, no que tange ao transitório, o mesmo

Distúrbio
<i>sag</i>
<i>swell</i>
<i>flicker</i>
harmônica
<i>sag+flicker</i>
<i>sag+harmônica</i>
<i>swell+flicker</i>
<i>swell+harmônica</i>
<i>sag+harmônica+flicker</i>
<i>swell+harmônica+flicker</i>

Tabela 1: Classes de distúrbios a serem classificados

foi adotado posto que a presente dissertação tem como objetivo avaliar o desempenho dos algoritmos *fuzzy* treinados tal como é descrito no Capítulo 5.

Para a realização das simulações computacionais, considerou-se $v(n) \sim \mathcal{N}(0, \sigma^2)$, tal que a relação sinal-ruído - *signal-to-noise ratio* (SNR) igual a 30 dB é dada por

$$SNR = 10 \log \left\{ \frac{A_0^2/2}{E\{v^2(n)\}} \right\}, \quad (6.1)$$

em que $E\{\}$ é o operador valor esperado, $A_0^2/2$ é a potência da componente fundamental e $v(n)$ é a n -ésima amostra do ruído aditivo, o qual é i.i.d.. Note que a SNR é dada em dB. A quantidade de bancos de dados gerados foram 1000. Esta quantidade foi igualmente dividida para uso nas etapas de projeto e teste (classificador de Bayes) e treinamento e teste (rede MLP e algoritmos *fuzzy*).

O classificador *fuzzy singleton* tem os seguintes parâmetros definidos: $M = 4$ (duas regras para a classe com o distúrbio e duas regras para a classe sem distúrbios); $p = 3$, posto que o FDR foi utilizado para selecionar o vetor de características de menor dimensionalidade possível para o problema de classificação.

Os parâmetros das funções de pertinência foram previamente definidos heurísticamente à partir do cálculo das médias e variâncias dos coeficientes dos vetores de características. O classificador *fuzzy non-singleton* possui as quantidades de regras e entradas iguais ao do classificador *fuzzy singleton*. Além disso, o mesmo possui a variância da função de pertinência de entrada $\sigma_X(q) = 0,46$, uma vez que tal valor gera resultados melhores se comparados aos obtidos através da inicializações com valores diferentes.

O classificador neural baseado na rede neural MLP tem $K_p = 3$ entradas, $N_q = 1$

camada escondida e 1 camada de saída.

A seguinte regra de decisão é adotada para os classificadores *fuzzy* e MLP:

1. Se $y(x) \geq 0$
 - Com ocorrência de distúrbios;
2. Se $y(x) \leq 0$
 - Sem ocorrência de distúrbios;

A regra de decisão do classificador de Bayes é fornecida de acordo com a Seção 2.2.3.1.

Note que as figuras 11 e 12 ilustram exemplos de valores obtidos para o discriminante de Fisher para o distúrbio isolado *flicker* e para o distúrbio *sag+flicker*, respectivamente e as características foram extraídas da componente *f*. É válido ressaltar que resultados semelhantes são obtidos para os demais distúrbios.

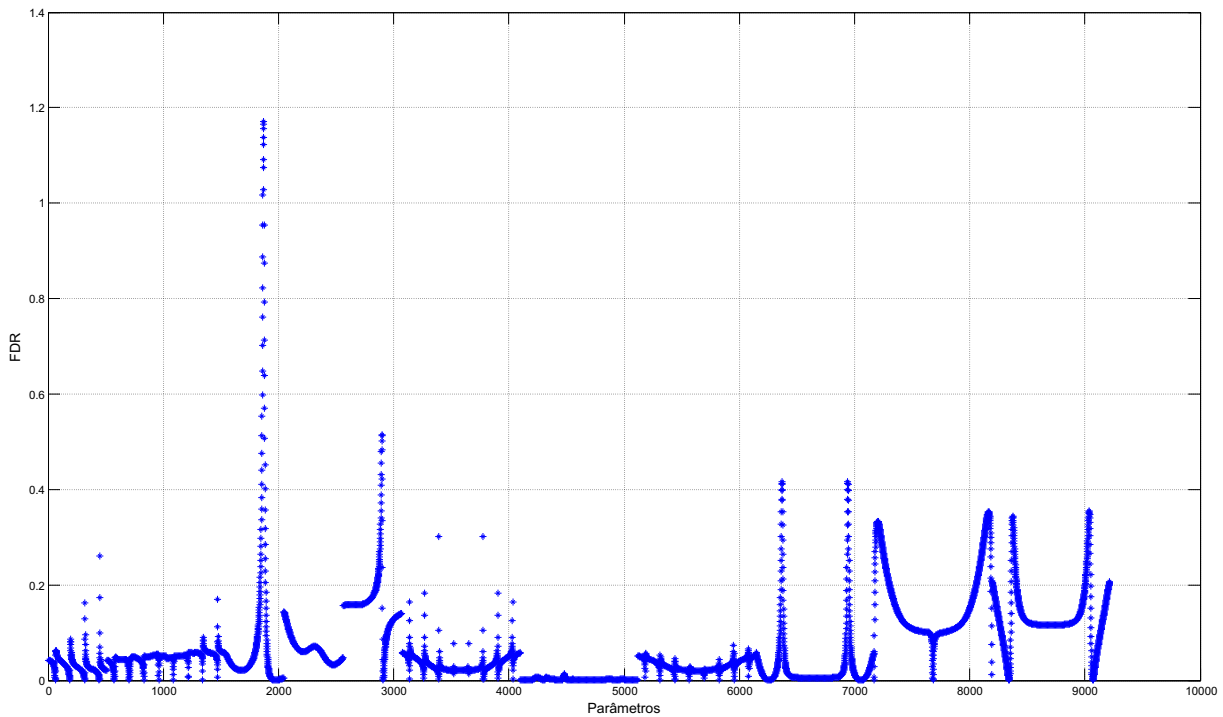


Figura 11: Valores do FDR obtidos para o distúrbio isolado *flicker*.

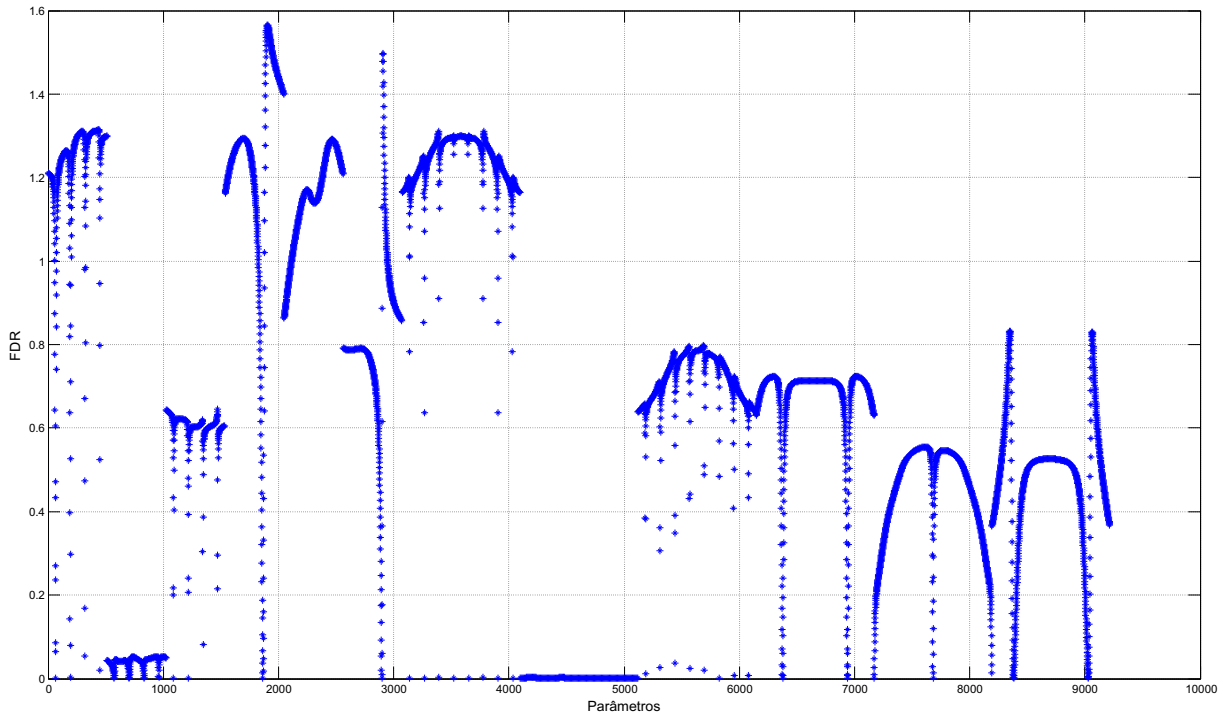


Figura 12: Valores do FDR obtidos para o distúrbio múltiplo *sag+flicker*.

O desempenho da técnica proposta para classificar um ou vários distúrbios em sinais de tensão é avaliada e comparada com a técnica apresentada em (RIBEIRO; PEREIRA, 2007). Assim, foram gerados vários conjuntos de distúrbios isolados (*sag*, *swell*, *flicker* e harmônica) e múltiplos (*sag+flicker*, *sag+harmônica*, *swell+flicker*, *swell+harmônica*, *sag+harmônica+flicker* e *swell+harmônica+flicker*) igualmente distribuídos para teste e treinamento. Adicionalmente, para as simulações do sistema de inferência *fuzzy*, considerou-se 100 épocas para treinamento do algoritmo. Para obtenção dos resultados a serem apresentados, optou-se por dividir os dados igualmente para obtenção do conjunto de treinamento e de teste do algoritmo. Como temos 500 amostras para cada um dos 3 parâmetros extraídos, obtemos assim, 250 amostras para o conjunto de treinamento e 250 amostras para o conjunto de teste.

Além disso, atribuiu-se aos parâmetros $\alpha(q)$ e $\beta(q)$ valores restritivos denominados $\alpha_{\max}(q)$ e $\beta_{\max}(q)$, visando eliminar atualizações indesejadas dos pesos do sistema de inferência *fuzzy* e consequentemente obter um melhor desempenho do algoritmo proposto. Adicionalmente, para o algoritmo do gradiente conjugado, considerou-se um passo de adaptação denominado μ , visando eliminar possíveis atualizações indesejadas dos pesos do sistema de inferência *fuzzy*. Os valores para tais parâmetros foram atribuídos com base

em heurísticas e os mesmos podem ser visualizados no Apêndice D.

6.1 Convergência

Através de simulação computacional, observou-se que o algoritmo de treinamento do gradiente conjugado apresenta uma velocidade de convergência maior se comparado com os algoritmos de retropropagação. Além disso, note que os algoritmos do gradiente conjugado que utilizam a matriz Hessiana $\mathbf{H}(J(\mathbf{w}^{(a)}))$ na composição demonstraram possuir uma menor velocidade de convergência se comparado com os demais que não utilizam a referida matriz. Não obstante, o mesmo ocorre para os algoritmos Quasi-Newton.

As figuras 13 - 16 ilustram a convergência para os distúrbios isolados *sag*, *swell*, *flicker* e harmônica, utilizando um sistema de inferência *fuzzy singleton*.

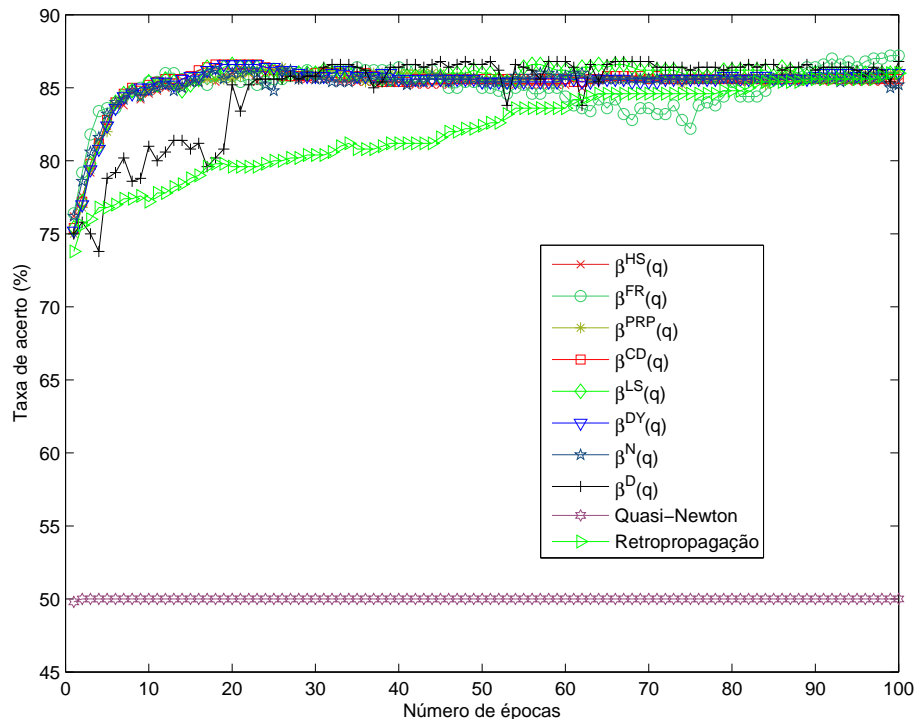


Figura 13: Convergência do sistema de inferência *fuzzy singleton* para o distúrbio *sag*.

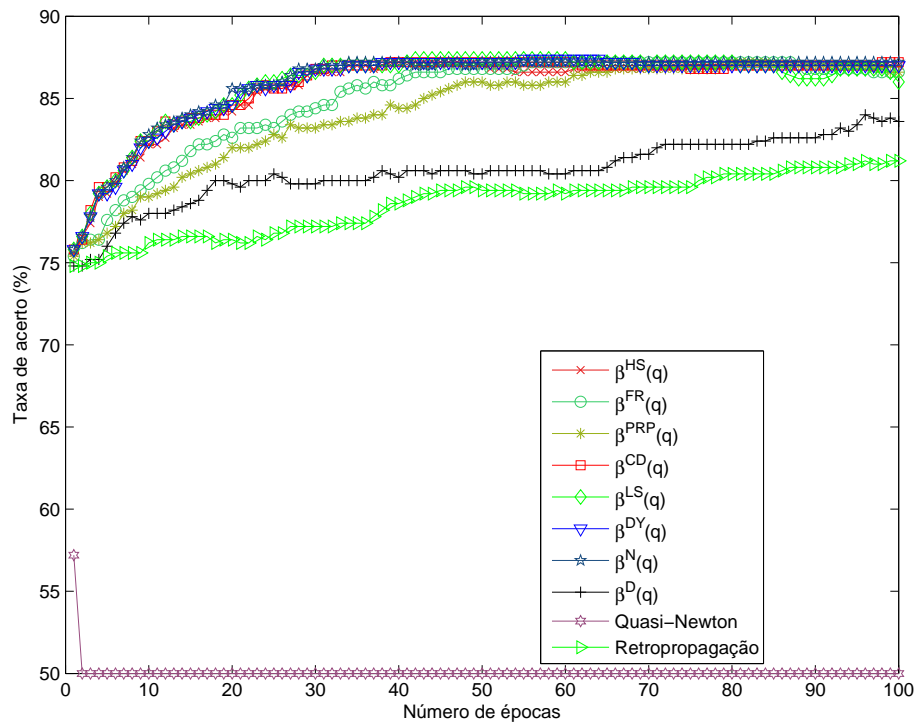


Figura 14: Convergência do sistema de inferência *fuzzy singleton* para o distúrbio *swell*.

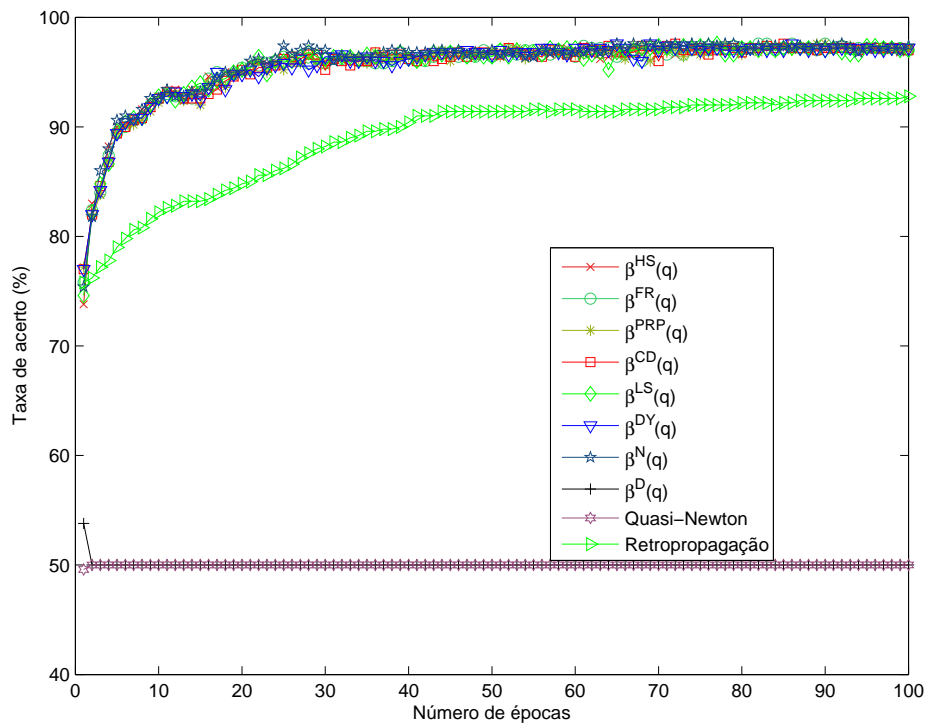


Figura 15: Convergência do sistema de inferência *fuzzy singleton* para o distúrbio *flicker*.

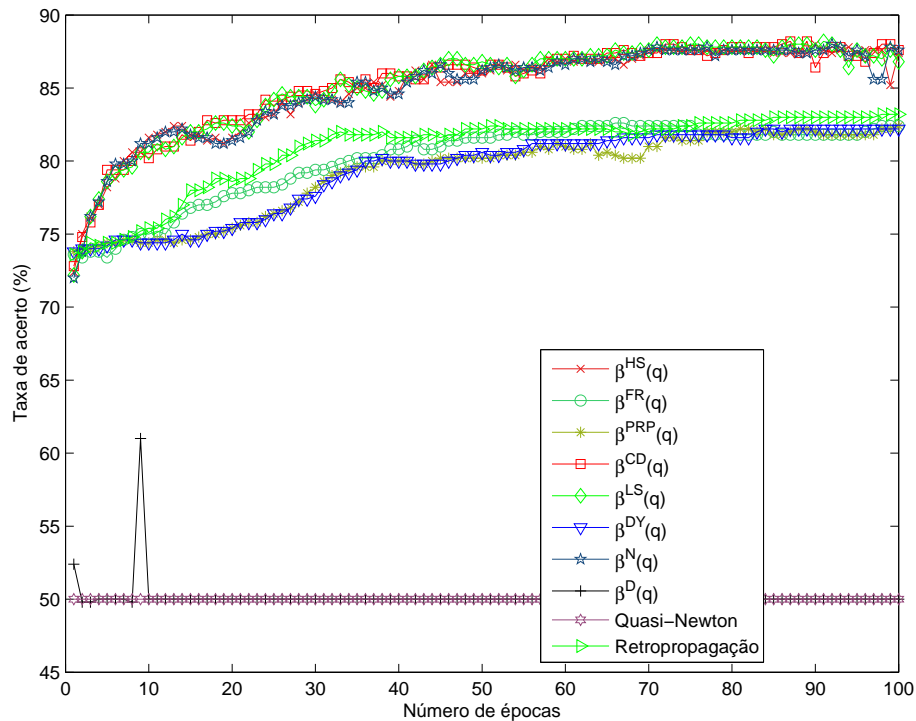


Figura 16: Convergência do sistema de inferência *fuzzy singleton* para o distúrbio harmônica.

Note que para os distúrbios isolados, os algoritmos baseados no gradiente conjugado possuem uma maior velocidade de convergência se comparados com os algoritmos de retropropagação. É válido ressaltar que para todos distúrbios isolados, o algoritmo baseado no método Quasi-Newton não apresentou um comportamento satisfatório, já que o mesmo obteve taxa de acerto de 50% em todos os casos acima apresentados. Tal fato deve-se ao mau condicionamento da matriz Hessiana $\mathbf{H}(J(\mathbf{w}^{(a)}))$, utilizada na referida técnica.

A seguir, as figuras 17 - 22 ilustram a convergência para os distúrbios múltiplos *sag+flicker*, *sag+harmônica*, *swell+flicker*, *swell+harmônica*, *sag+harmônica+flicker* e *swell+harmônica+flicker*, utilizando um sistema de inferência *fuzzy singleton*.

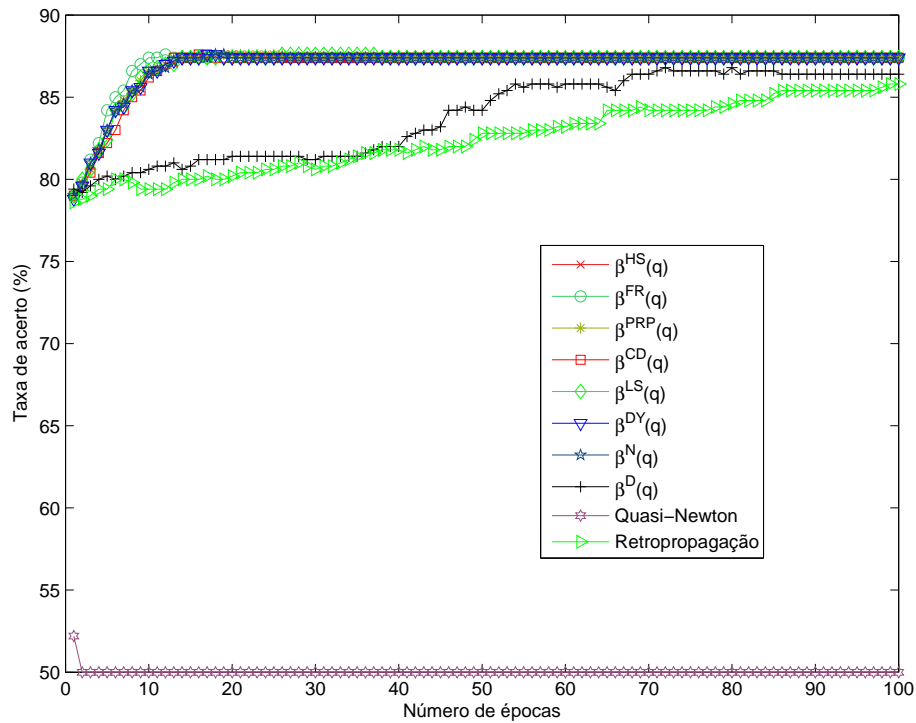


Figura 17: Convergência do sistema de inferência *fuzzy singleton* para o distúrbio *sag+flicker*.

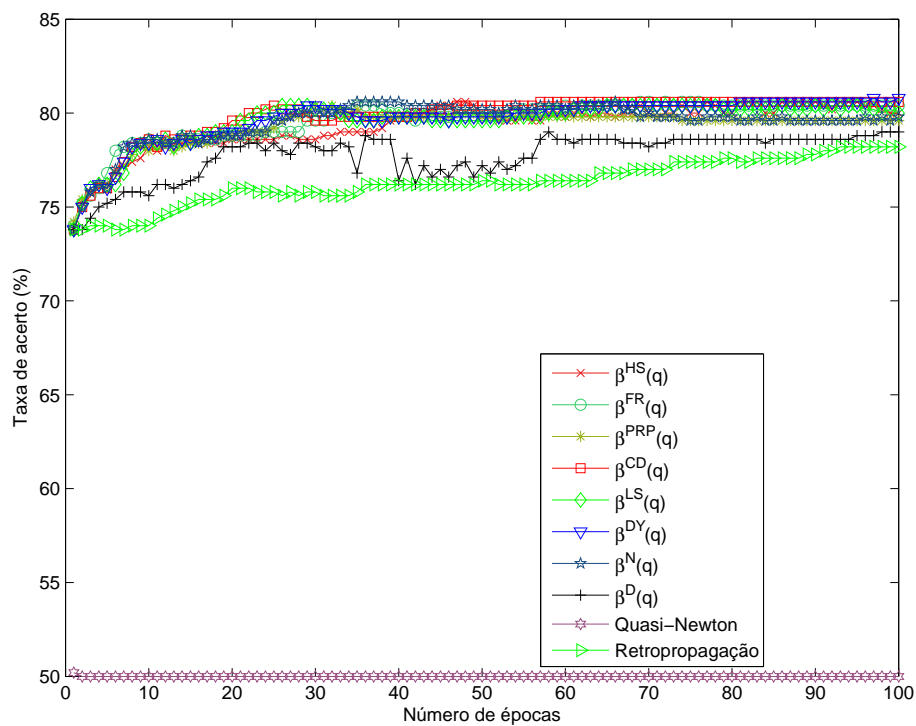


Figura 18: Convergência do sistema de inferência *fuzzy singleton* para o distúrbio *sag+harmônica*.

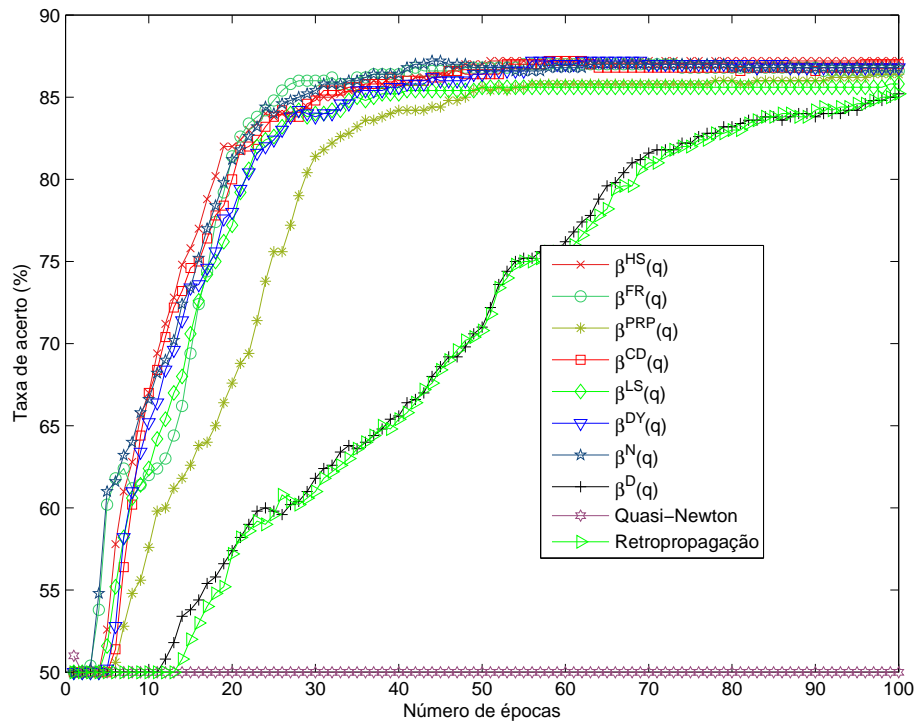


Figura 19: Convergência do sistema de inferência *fuzzy singleton* para o distúrbio *swell+flicker*.

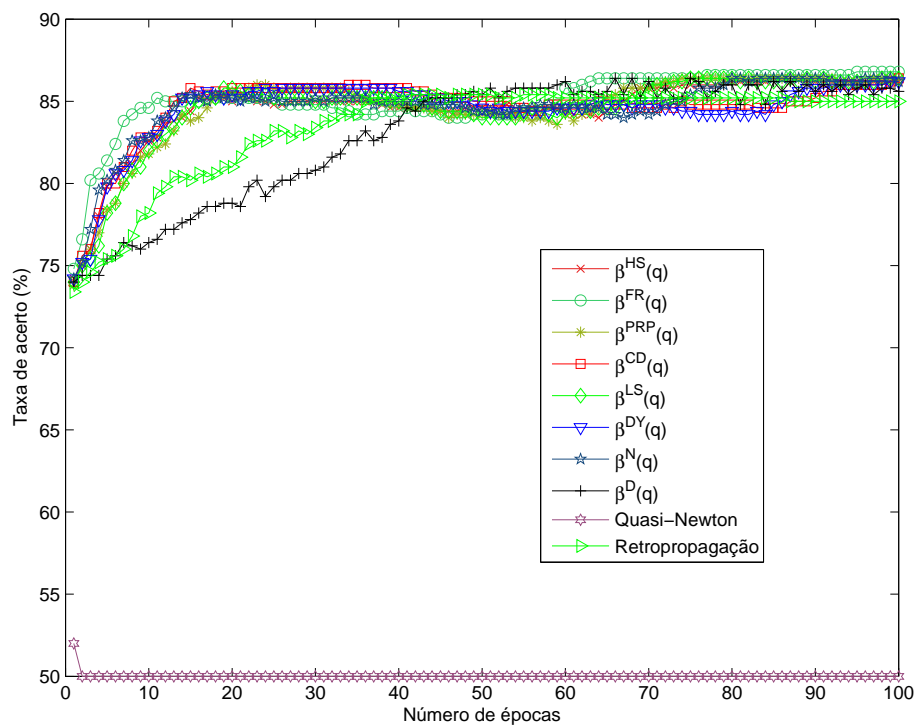


Figura 20: Convergência do sistema de inferência *fuzzy singleton* para o distúrbio *swell+harmônica*.

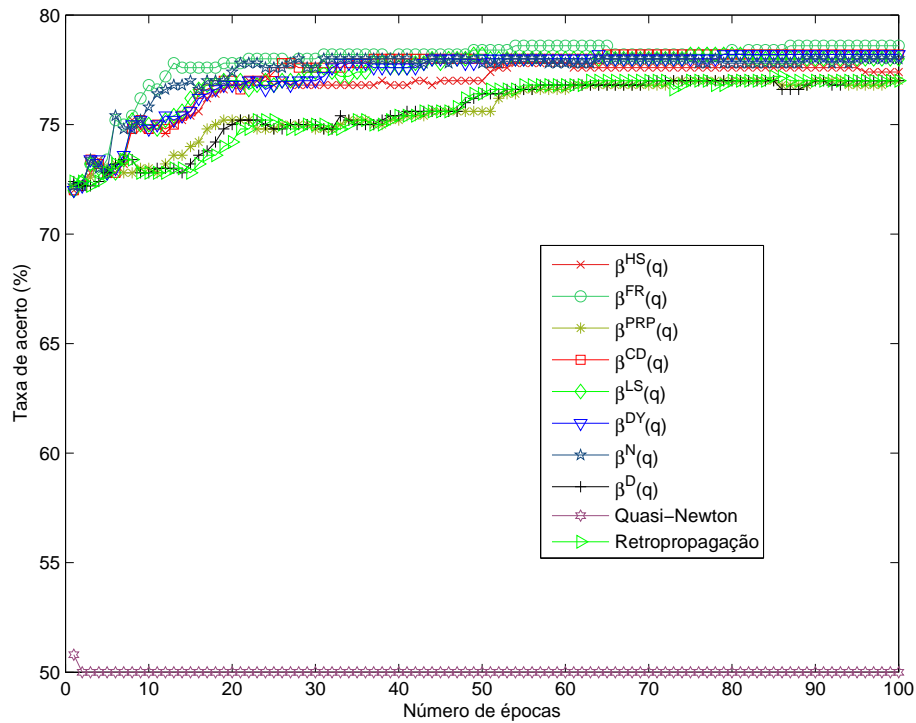


Figura 21: Convergência do sistema de inferência *fuzzy singleton* para o distúrbio *sag+harmônica+flicker*.

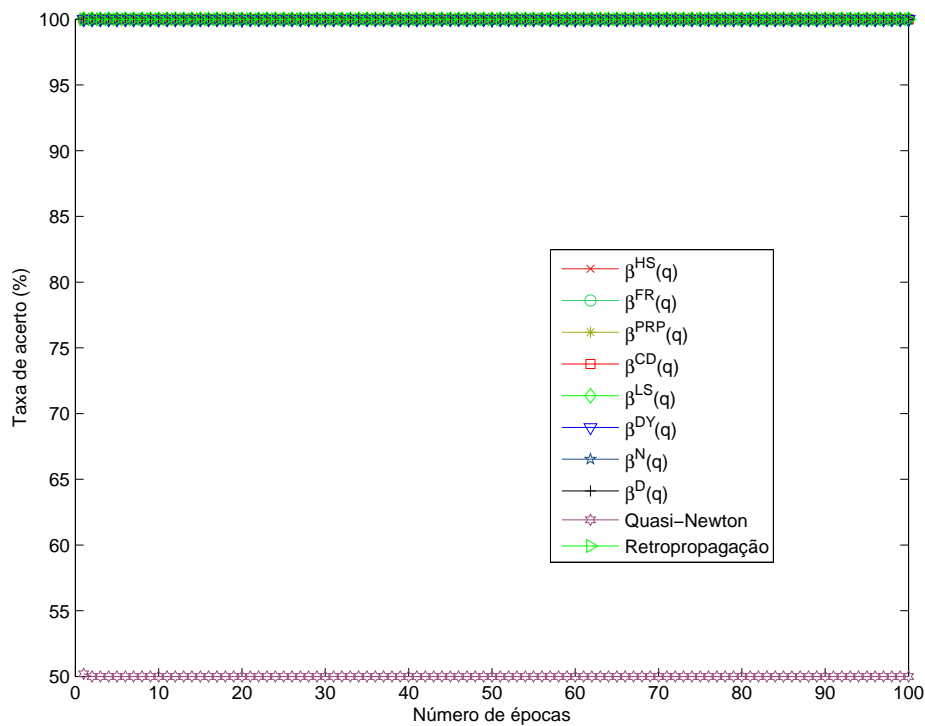


Figura 22: Convergência do sistema de inferência *fuzzy singleton* para o distúrbio *swell+harmônica+flicker*.

Para os distúrbios múltiplos classificados pelo sistema de inferência *fuzzy singleton*, é fácil visualizar que os algoritmos baseado no gradiente conjugado possuem uma melhor convergência se comparados com os demais métodos. Novamente, note que o algoritmo baseado no método Quasi-Newton possui um desempenho pouco satisfatório. Tal fato, deve-se ao mau condicionamento da matriz Hessiana $\mathbf{H}(J(\mathbf{w}^{(a)}))$.

Analisando o distúrbio *swell+harmônica+flicker*, é válido ressaltar o desempenho dos algoritmos propostos, que à exceção do método Quasi-Newton, obtiveram uma alta velocidade de convergência e uma taxa de acerto de 100% ao longo de todas as suas épocas de treinamento devido à grande separabilidade entre as classes de distúrbios que estão sendo classificados.

A seguir, as figuras 23 - 26 ilustram a convergência levando em consideração os distúrbios isolados *sag*, *swell*, *flicker* e harmônica, utilizando um sistema de inferência *fuzzy non-singleton*.

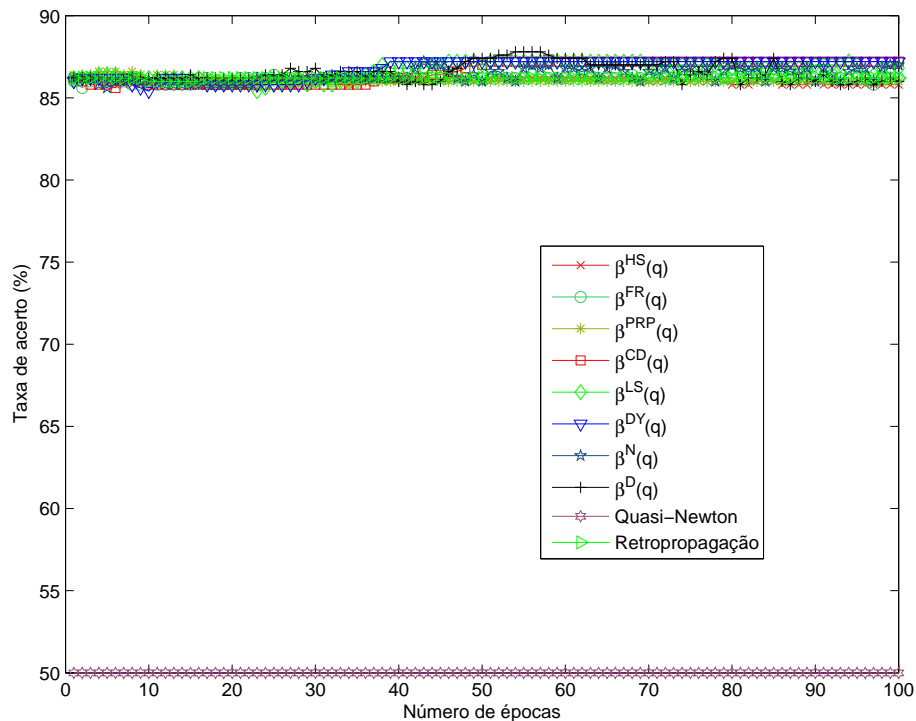


Figura 23: Convergência do sistema de inferência *fuzzy non-singleton* para o distúrbio *sag*.

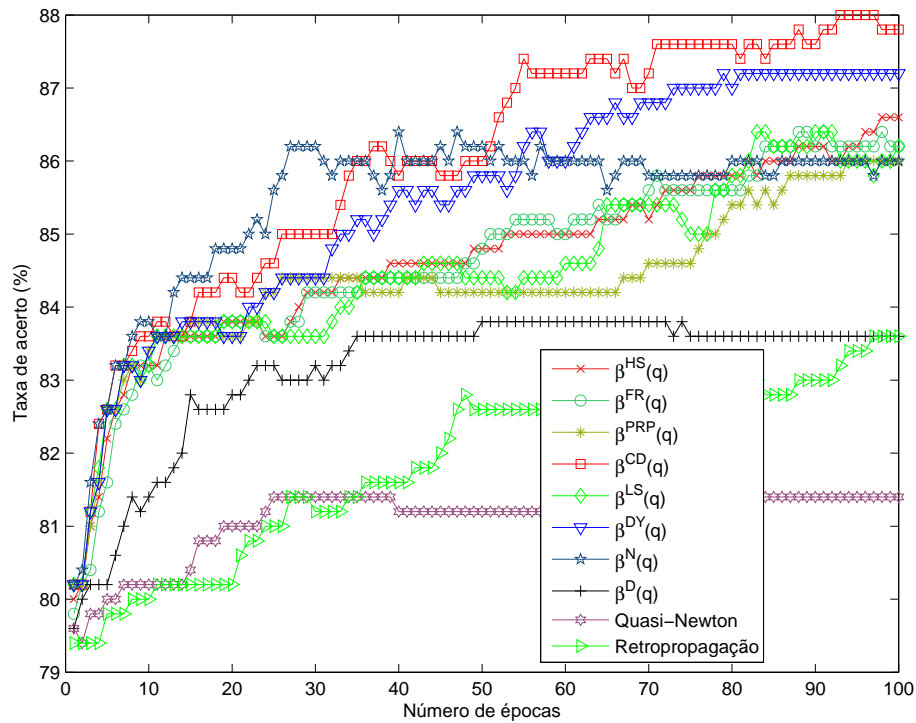


Figura 24: Convergência do sistema de inferência *fuzzy non-singleton* para o distúrbio *swell*.

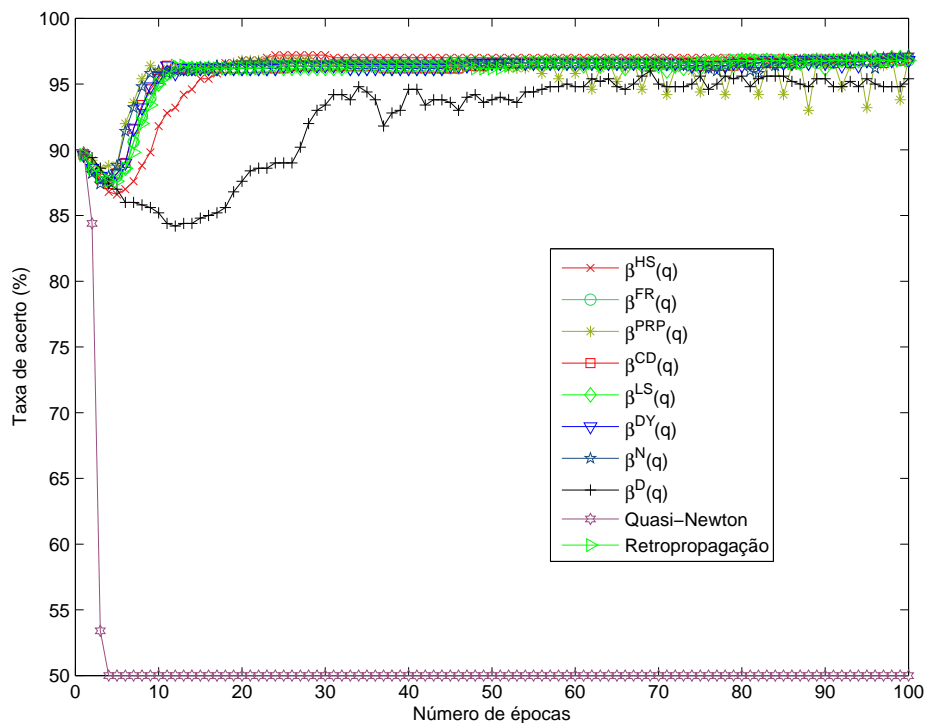


Figura 25: Convergência do sistema de inferência *fuzzy non-singleton* para o distúrbio *flicker*.

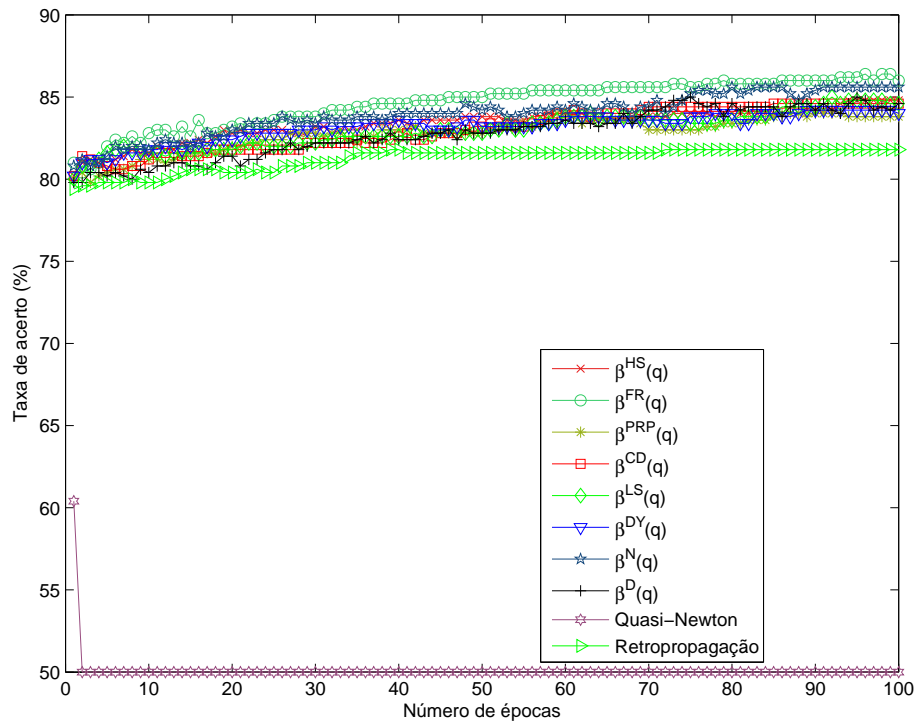


Figura 26: Convergência do sistema de inferência *fuzzy non-singleton* para o distúrbio harmônica.

Assim como foi discutido no sistema de inferência *fuzzy singleton*, note que para os distúrbios isolados, os algoritmos baseados no gradiente conjugado possuem uma maior velocidade de convergência se comparados com os algoritmos de retropropagação. Novamente, é válido ressaltar que para todos distúrbios isolados, o algoritmo baseado no método Quasi-Newton não apresentou um comportamento satisfatório, já que o mesmo obteve um desempenho inferior se comparado com as demais técnicas contempladas nesta dissertação. Tal fato deve-se ao mau condicionamento da matriz Hessiana $\mathbf{H}(J(\mathbf{w}^{(q)}))$, utilizada na referida técnica.

A seguir, as figuras 27 - 32 ilustram a convergência para os distúrbios múltiplos *sag+flicker*, *sag+harmônica*, *swell+flicker*, *swell+harmônica*, *sag+harmônica+flicker* e *swell+harmônica+flicker*, utilizando um sistema de inferência *fuzzy non-singleton*.

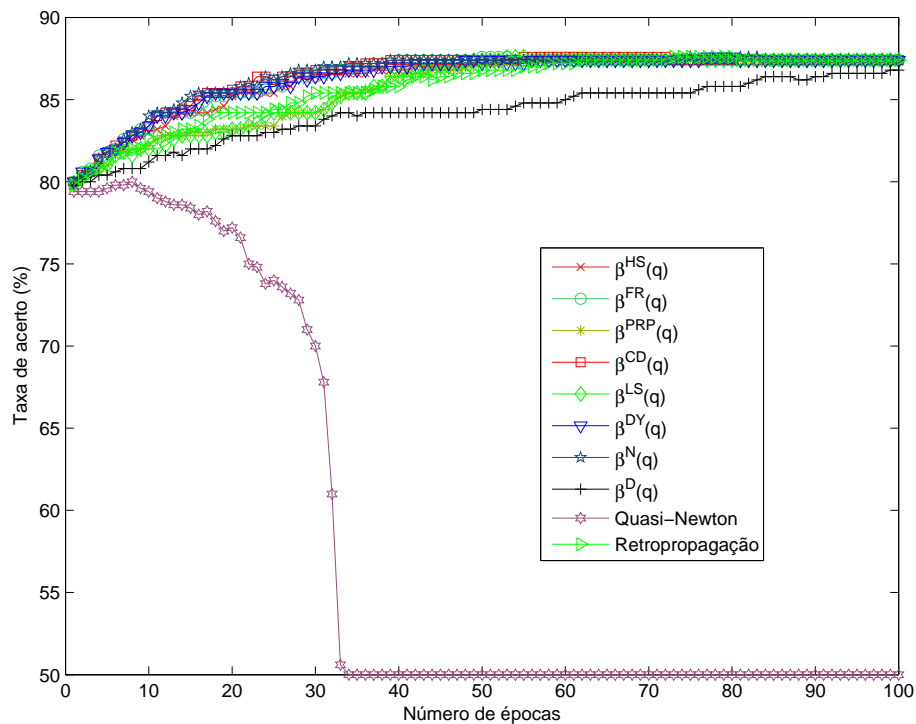


Figura 27: Convergência do sistema de inferência *fuzzy non-singleton* para o distúrbio *sag+flicker*.

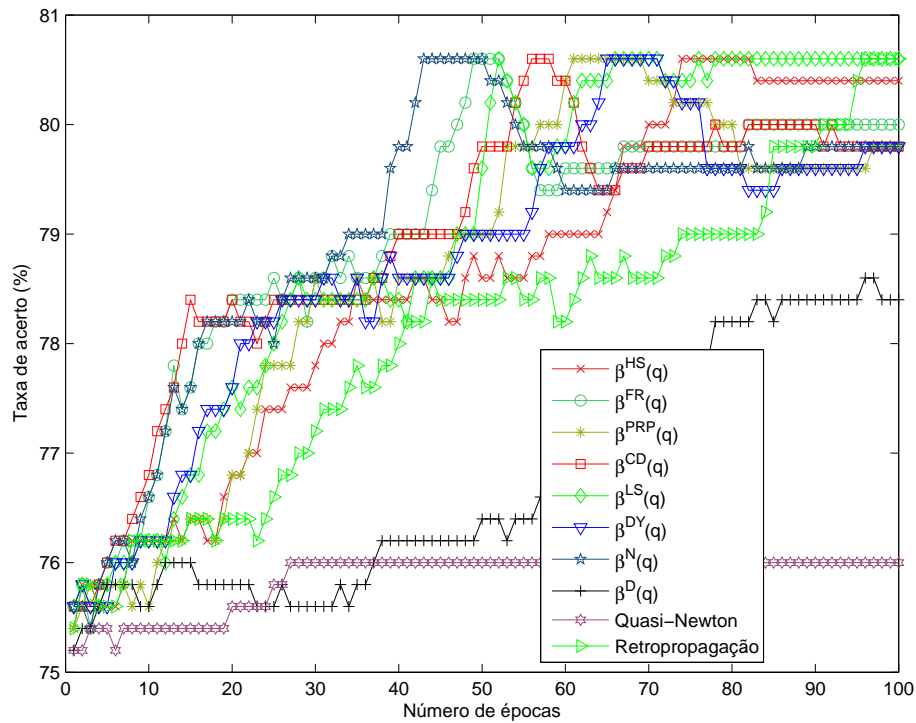


Figura 28: Convergência do sistema de inferência *fuzzy non-singleton* para o distúrbio *sag+harmônica*.

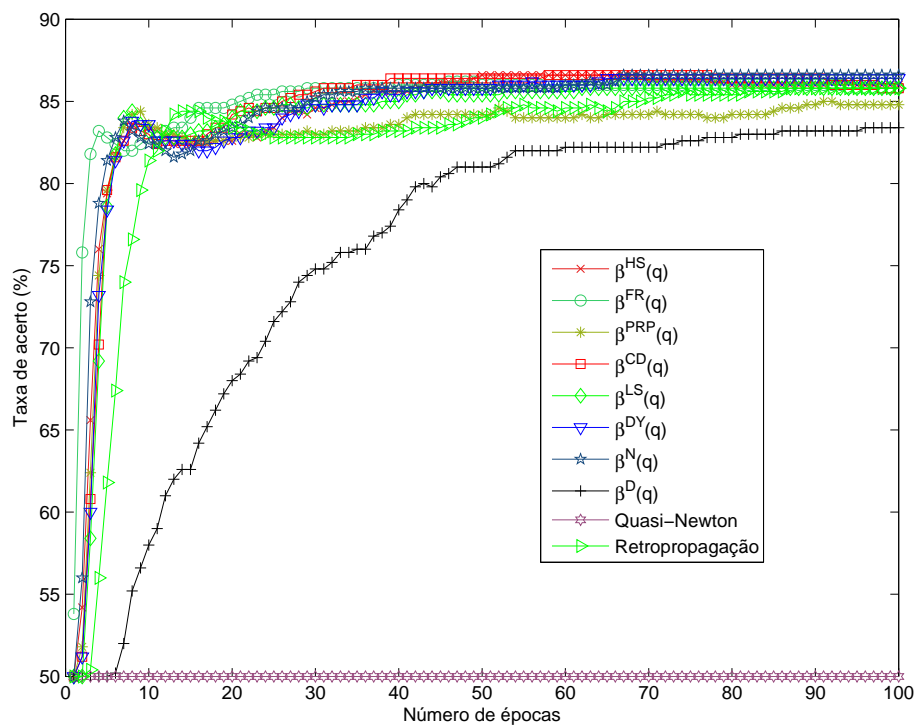


Figura 29: Convergência do sistema de inferência *fuzzy non-singleton* para o distúrbio *swell+flicker*.

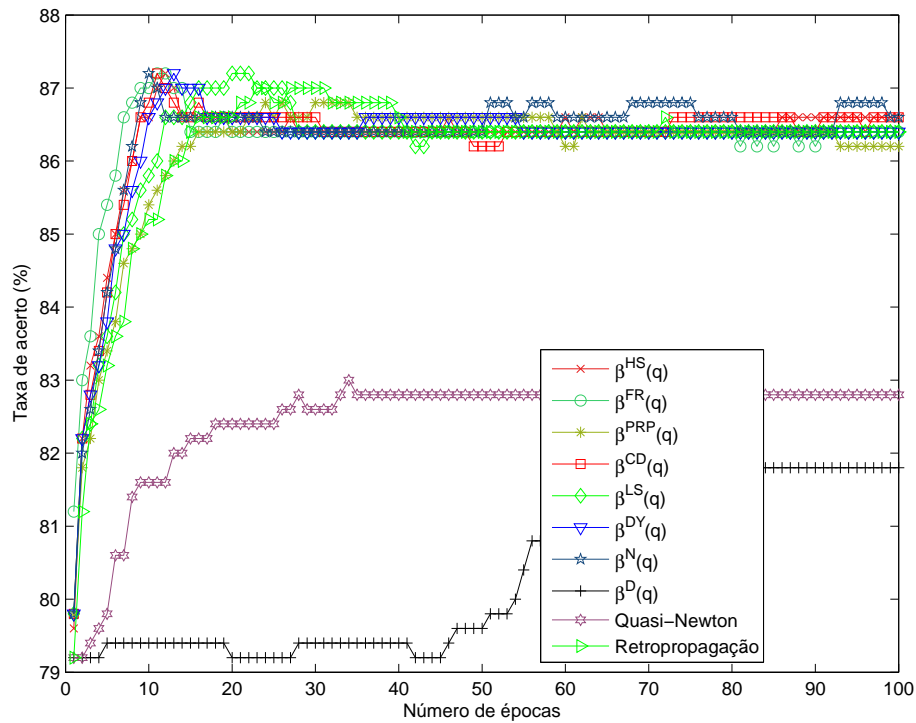


Figura 30: Convergência do sistema de inferência *fuzzy non-singleton* para o distúrbio *swell+harmônica*.

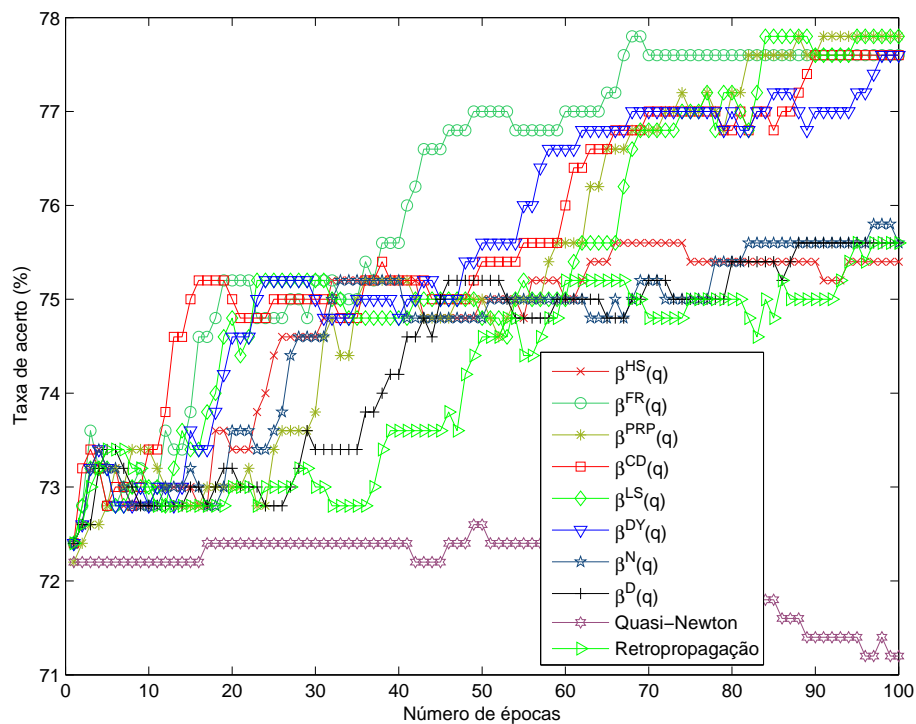


Figura 31: Convergência do sistema de inferência *fuzzy non-singleton* para o distúrbio *sag+harmônica+flicker*.

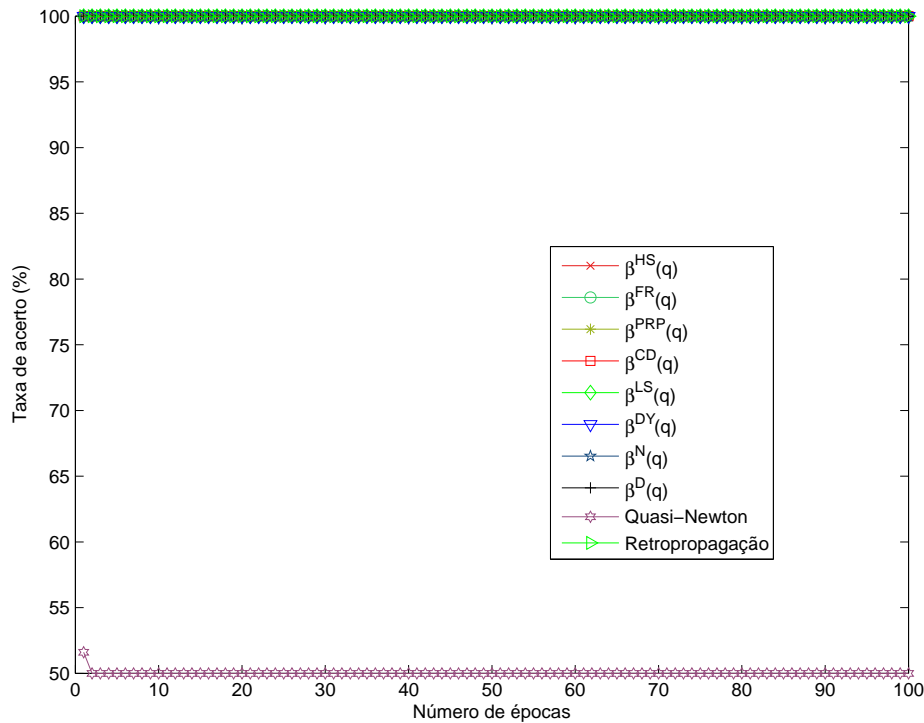


Figura 32: Convergência do sistema de inferência *fuzzy non-singleton* para o distúrbio *swell+harmônica+flicker*.

Para os distúrbios múltiplos classificados pelo sistema de inferência *fuzzy non-singleton*, novamente é fácil visualizar que os algoritmos baseado no gradiente conjugado possuem uma melhor convergência se comparados com os demais métodos. Tal afirmação pode ser facilmente comprovada através de uma simples visualização do comportamento do algoritmo ao longo das 100 épocas de treinamento. Assim como nas situações supracitadas, note que o algoritmo baseado no método Quasi-Newton possui um desempenho pouco satisfatório. Tal fato, deve-se ao mau condicionamento da matriz Hessiana $\mathbf{H}(J(\mathbf{w}^{(q)}))$.

Ao analisar o desempenho dos distúrbios múltiplos *sag+flicker*, temos que a melhor convergência foi obtida quando utilizou-se $\beta^{DY}(q)$. Para o distúrbio múltiplo *sag+harmônica*, note que $\beta^{LS}(q)$ propiciou a melhor velocidade de convergência.

Por fim, assim como ocorrido para o sistema de inferência *fuzzy singleton*, é válido ressaltar o desempenho dos algoritmos propostos para o distúrbio *swell+harmônica+flicker*, que à exceção do método Quasi-Newton, obtiveram uma alta velocidade de convergência e uma taxa de acerto de 100% ao longo de todas as suas épocas de treinamento devido à grande separabilidade entre as classes de distúrbios que estão sendo classificados.

6.2 Desempenho

Nesta seção, serão apresentados os resultados obtidos através de simulações computacionais para os distúrbios isolados e múltiplos que deseja-se classificar.

Note que serão comparados os resultados entre as seguintes classificadores, apresentados e discutidos no capítulo 2: bayes, rede neural *perceptron* multicamadas e sistema de inferência *fuzzy singleton* e *non-singleton*, apresentado e discutido no capítulo 4. Note que tais resultados são expressos na tabela 2

A seguir, para o sistema de inferência *fuzzy*, analisaremos o desempenho para os seguintes distúrbios isolados: *sag*, *swell*, *flicker* e harmônica. Posteriormente, analisaremos o desempenho do classificador proposto para os seguintes distúrbios múltiplos: *sag+flicker*, *sag+harmônica*, *swell+flicker*, *swell+harmônica*, *sag+harmônica+flicker*, *swell+harmônica+flicker*.

Distúrbio	Bayes	Rede MLP
<i>sag</i>	86,2	87,8
<i>swell</i>	82,0	88,8
<i>flicker</i>	99,4	97,8
harmônica	94,2	88,6
<i>sag+flicker</i>	85,8	87,8
<i>sag+harmônica</i>	79,4	82,6
<i>swell+flicker</i>	80,0	87,0
<i>swell+harmônica</i>	86,0	88,0
<i>sag+harmônica+flicker</i>	77,6	79,4
<i>swell+harmônica+flicker</i>	100	100

Tabela 2: Desempenho dos classificadores de Bayes e rede neural *perceptron* multicamadas em (%), para distúrbios isolados e múltiplos.

De uma forma geral, note que a rede neural *perceptron* multicamadas, apresentada no capítulo 2, possui um melhor desempenho para a classificação de distúrbios isolados e múltiplos, se comparado com a técnica de classificação baseada no teorema de Bayes. Essa rede foi escolhida por apresentar elevada velocidade de convergência e baixo esforço computacional com relação aos algoritmos baseados em informação de segunda ordem. A seguir, para o sistema de inferência *fuzzy*, analisaremos o desempenho para os distúrbios isolados: *sag*, *swell*, *flicker* e harmônica.

<i>sag</i>		
Técnica Escolhida	Treinamento	Teste
$\beta^{HS}(q)$ - Hestenes e Stiefel	85,4	86,4
$\beta^{FR}(q)$ - Fletcher e Reeves	87,2	87,8
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	86,2	86,6
$\beta^{CD}(q)$ - Fletcher	85,6	86,8
$\beta^{LS}(q)$ - Liu e Storey	86,0	86,6
$\beta^{DY}(q)$ - Dai e Yuan	86,0	86,6
$\beta^N(q)$ - Hager e Zhang	85,2	86,4
$\beta^H(q)$ - Daniel	86,8	86,6
Quasi-Newton	50,0	50,0
Retropropagação	86,0	86,2

Tabela 3: Desempenho do sistema de inferência *fuzzy singleton* em (%), para o distúrbio *sag*.

<i>swell</i>		
Técnica Escolhida	Treinamento	Teste
$\beta^{HS}(q)$ - Hestenes e Stiefel	87,0	87,0
$\beta^{FR}(q)$ - Fletcher e Reeves	86,6	86,8
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	86,6	86,8
$\beta^{CD}(q)$ - Fletcher	87,2	87,2
$\beta^{LS}(q)$ - Liu e Storey	86,0	86,0
$\beta^{DY}(q)$ - Dai e Yuan	87,0	87,0
$\beta^N(q)$ - Hager e Zhang	87,2	87,2
$\beta^H(q)$ - Daniel	83,6	83,6
Quasi-Newton	50,0	50,0
Retropropagação	81,2	81,2

Tabela 4: Desempenho do sistema de inferência *fuzzy singleton* em (%), para o distúrbio *swell*.

<i>flicker</i>		
Técnica Escolhida	Treinamento	Teste
$\beta^{HS}(q)$ - Hestenes e Stiefel	97,0	97,4
$\beta^{FR}(q)$ - Fletcher e Reeves	97,0	97,6
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	97,2	97,6
$\beta^{CD}(q)$ - Fletcher	97,0	97,6
$\beta^{LS}(q)$ - Liu e Storey	97,0	97,6
$\beta^{DY}(q)$ - Dai e Yuan	97,2	97,6
$\beta^N(q)$ - Hager e Zhang	97,4	97,8
$\beta^H(q)$ - Daniel	50,0	50,0
Quasi-Newton	50,0	50,0
Retropropagação	92,8	93,6

Tabela 5: Desempenho do sistema de inferência *fuzzy singleton* em (%), para o distúrbio *flicker*.

harmônica		
Técnica Escolhida	Treinamento	Teste
$\beta^{HS}(q)$ - Hestenes e Stiefel	87,4	86,2
$\beta^{FR}(q)$ - Fletcher e Reeves	82,4	83,2
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	82,4	82,8
$\beta^{CD}(q)$ - Fletcher	87,6	86,6
$\beta^{LS}(q)$ - Liu e Storey	86,8	87,2
$\beta^{DY}(q)$ - Dai e Yuan	82,2	82,6
$\beta^N(q)$ - Hager e Zhang	87,6	86,8
$\beta^H(q)$ - Daniel	50,0	50,0
Quasi-Newton	50,0	50,0
Retropropagação	83,2	83,8

Tabela 6: Desempenho do sistema de inferência *fuzzy singleton* em (%), para o distúrbio harmônica.

A tabela 3 apresenta os resultados para o distúrbio isolado *sag*. Note que os melhores desempenhos foram obtidos para $\beta^{FR}(q)$, proposto por Fletcher e Reeves, obtendo taxas de classificação de 87,2% para o treinamento e 87,8% para o teste. Todavia, é válido ressaltar que as técnicas que fazem uso de $\beta^{CD}(q)$ conforme Fletcher possui um desempenho próximo da técnica supracitada, podendo ser considerada uma solução satisfatória para a resolução do problema de classificação do distúrbios isolados de QEE denominado *sag*.

Para o distúrbio *swell*, as técnicas que obtiveram melhor desempenho foram as que fizeram uso de $\beta^{CD}(q)$, proposto por Fletcher e também conhecido como algoritmo de descida conjugada (do inglês *Conjugate Descent*), obtendo 87,2% para treinamento e 87,2% para teste e $\beta^N(q)$, proposto por Hager e Zhang, obtendo os mesmos 87,2% para treinamento e 87,2% para teste, conforme apresentado na tabela 4.

Por sua vez, para o distúrbio *flicker*, a técnica proposta por Hager e Zhang, que faz uso de $\beta^N(q)$, obteve o melhor desempenho, com 97,4% de taxa de classificação obtida para o treinamento e 97,8% para o teste, conforme apresentado na tabela 5. É válido ressaltar o desempenho satisfatório oriundo da utilização de $\beta^{DY}(q)$ e $\beta^{PRP}(q)$, propostos respectivamente por Dai e Yuan e Polak, Ribière e Polyak. Note que tais técnicas propiciaram uma taxa de classificação de 97,2% para o treinamento e 97,6% para o teste, fazendo com que também possam ser consideradas abordagens promissoras.

Para o distúrbio harmônica, cujo os resultados foram reportados na tabela 6, temos que o melhor desempenho foi alcançado utilizando a técnica que faz uso de $\beta^{LS}(q)$, com a taxa de classificação para o treinamento de 86,8% e teste de 87,2%.

A seguir, serão apresentados e discutidos os resultados referentes ao desempenho do sistema de inferência *fuzzy singleton* para classificação dos seguintes distúrbios múltiplos: *sag+flicker*, *sag+harmônica*, *swell+flicker*, *swell+harmônica*, *sag+harmônica+flicker* e *swell+harmônica+flicker*.

<i>sag+flicker</i>		
Técnica Escolhida	Treinamento	Teste
$\beta^{HS}(q)$ - Hestenes e Stiefel	87,4	87,4
$\beta^{FR}(q)$ - Fletcher e Reeves	87,4	87,4
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	87,4	87,4
$\beta^{CD}(q)$ - Fletcher	87,4	87,4
$\beta^{LS}(q)$ - Liu e Storey	87,4	87,4
$\beta^{DY}(q)$ - Dai e Yuan	87,4	87,4
$\beta^N(q)$ - Hager e Zhang	87,4	87,4
$\beta^H(q)$ - Daniel	86,4	86,4
Quasi-Newton	50,0	50,0
Retropropagação	85,8	85,8

Tabela 7: Desempenho do sistema de inferência *fuzzy singleton* em (%), para o distúrbio *sag+flicker*.

<i>sag+harmônica</i>		
Técnica Escolhida	Treinamento	Teste
$\beta^{HS}(q)$ - Hestenes e Stiefel	80,0	80,2
$\beta^{FR}(q)$ - Fletcher e Reeves	80,4	80,4
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	79,6	79,6
$\beta^{CD}(q)$ - Fletcher	80,6	80,6
$\beta^{LS}(q)$ - Liu e Storey	80,0	79,8
$\beta^{DY}(q)$ - Dai e Yuan	80,8	80,6
$\beta^N(q)$ - Hager e Zhang	79,8	79,6
$\beta^H(q)$ - Daniel	79,0	79,0
Quasi-Newton	50,0	50,0
Retropropagação	78,2	78,2

Tabela 8: Desempenho do sistema de inferência *fuzzy singleton* em (%), para o distúrbio *sag+harmônica*.

<i>swell+flicker</i>		
Técnica Escolhida	Treinamento	Teste
$\beta^{HS}(q)$ - Hestenes e Stiefel	87,2	87,2
$\beta^{FR}(q)$ - Fletcher e Reeves	86,6	86,8
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	86,4	86,4
$\beta^{CD}(q)$ - Fletcher	87,0	86,8
$\beta^{LS}(q)$ - Liu e Storey	85,8	85,8
$\beta^{DY}(q)$ - Dai e Yuan	86,8	86,8
$\beta^N(q)$ - Hager e Zhang	86,8	86,8
$\beta^H(q)$ - Daniel	85,2	85,2
Quasi-Newton	50,0	50,0
Retropropagação	85,2	85,2

Tabela 9: Desempenho do sistema de inferência *fuzzy singleton* em (%), para o distúrbio *swell+flicker*.

<i>swell+harmônica</i>		
Técnica Escolhida	Treinamento	Teste
$\beta^{HS}(q)$ - Hestenes e Stiefel	86,2	86,4
$\beta^{FR}(q)$ - Fletcher e Reeves	86,8	86,8
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	86,4	86,2
$\beta^{CD}(q)$ - Fletcher	86,4	86,4
$\beta^{LS}(q)$ - Liu e Storey	86,4	86,4
$\beta^{DY}(q)$ - Dai e Yuan	86,2	86,4
$\beta^N(q)$ - Hager e Zhang	86,2	86,2
$\beta^H(q)$ - Daniel	85,6	85,8
Quasi-Newton	50,0	50,0
Retropropagação	85,0	84,8

Tabela 10: Desempenho do sistema de inferência *fuzzy singleton* em (%), para o distúrbio *swell+harmônica*.

<i>sag+harmônica+flicker</i>		
Técnica Escolhida	Treinamento	Teste
$\beta^{HS}(q)$ - Hestenes e Stiefel	77,4	77,4
$\beta^{FR}(q)$ - Fletcher e Reeves	78,6	78,6
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	77,0	77,0
$\beta^{CD}(q)$ - Fletcher	78,2	78,2
$\beta^{LS}(q)$ - Liu e Storey	78,0	78,0
$\beta^{DY}(q)$ - Dai e Yuan	78,2	78,2
$\beta^N(q)$ - Hager e Zhang	78,0	78,0
$\beta^H(q)$ - Daniel	77,0	77,2
Quasi-Newton	50,0	50,0
Retropropagação	77,0	77,0

Tabela 11: Desempenho do sistema de inferência *fuzzy singleton* em (%), para o distúrbio *sag+harmônica+flicker*.

<i>swell+harmônica+flicker</i>		
Técnica Escolhida	Treinamento	Teste
$\beta^{HS}(q)$ - Hestenes e Stiefel	100,0	100,0
$\beta^{FR}(q)$ - Fletcher e Reeves	100,0	100,0
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	100,0	100,0
$\beta^{CD}(q)$ - Fletcher	100,0	100,0
$\beta^{LS}(q)$ - Liu e Storey	100,0	100,0
$\beta^{DY}(q)$ - Dai e Yuan	100,0	100,0
$\beta^N(q)$ - Hager e Zhang	100,0	100,0
$\beta^H(q)$ - Daniel	100,0	100,0
Quasi-Newton	50,0	50,0
Retropropagação	100,0	100,0

Tabela 12: Desempenho do sistema de inferência *fuzzy singleton* em (%), para o distúrbio *swell+harmônica+flicker*.

A tabela 7 apresenta os resultados para o distúrbio múltiplo *sag+flicker*. Note que os melhores desempenhos foram obtidos para $\beta^{HS}(q)$, $\beta^{FR}(q)$, $\beta^{PRP}(q)$, $\beta^{CD}(q)$, $\beta^{LS}(q)$, $\beta^{DY}(q)$, $\beta^N(q)$, propostos respectivamente por Hestenes e Stiefel, Fletcher e Reeves, Polak, Ribière e Polyak, Fletcher, Liu e Storey, Dai e Yuan e Harger e Zhang. Note que as proposições citadas obtiveram taxas de classificação de 87,4% para o treinamento e teste.

Para o distúrbio múltiplo *sag+harmônica*, as técnicas que obtiveram melhor desempenho foram as que fizeram uso de $\beta^{CD}(q)$, proposto por Fletcher, obtendo 80,6% para treinamento e teste e $\beta^{DY}(q)$, proposto por Dai e Yuan, obtendo 80,8 para treinamento e 80,6 para teste, conforme apresentado na tabela 8.

Por sua vez, para o distúrbio múltiplo *swell+flicker*, a técnica proposta por Hestenes e Stiefel, que faz uso de $\beta^{HS}(q)$, obteve o melhor desempenho, com 87,2% de taxa de classificação obtida para o treinamento e teste, conforme apresentado na tabela 9. É válido ressaltar o desempenho satisfatório oriundo da utilização de $\beta^{FR}(q)$, $\beta^{CD}(q)$, $\beta^{DY}(q)$ e $\beta^N(q)$. Note que os mesmos propiciaram uma taxa de acerto de respectivamente 86,6%, 87,0%, 86,8% e 86,8% para o treinamento e 86,8% para o teste, fazendo com que também possa ser considerada uma abordagem satisfatória.

Ao analisar o distúrbio múltiplo *swell+harmônica*, note que o melhor desempenho foi obtido para a técnica que faz uso de $\beta^{FR}(q)$, proposto por Fletcher e Reeves, com 86,8% de taxa de classificação obtida para o treinamento e teste. Note que tais resultados são apresentados na tabela 10.

Para o distúrbio *sag+harmônica+flicker*, cujo os resultados foram reportados na tabela 11, temos que o melhor desempenho foi alcançado utilizando a técnica que faz uso de $\beta^{FR}(q)$, proposto por Fletcher e Reeves, com a taxa de classificação para o treinamento e teste de 78,6%. É válido ressaltar que as técnicas que fazem uso de $\beta^{CD}(q)$ e $\beta^{DY}(q)$, obtiveram um desempenho satisfatório, com taxas de classificação de 78,2% para o treinamento e teste, podendo ser consideradas alternativas promissoras para a resolução do problema de classificação do distúrbio múltiplo em questão.

Por fim, para o distúrbio *swell+harmônica+flicker*, cujo os resultados foram reportados na tabela 12, note que todas as técnicas propostas obtiveram taxa de classificação de 100% para o treinamento e para o teste, à exceção do método Quasi-Newton, que obteve ao longo das 100 épocas de treinamento uma taxa de classificação de 50,0% para o treinamento e teste do algoritmo.

A seguir, para o sistema de inferência fuzzy *non-singleton*, veremos o desempenho para os seguintes distúrbios isolados: *sag*, *swell*, *flicker* e harmônica.

<i>sag</i>		
Técnica Escolhida	Treinamento	Teste
$\beta^{HS}(q)$ - Hestenes e Stiefel	85,8	86,8
$\beta^{FR}(q)$ - Fletcher e Reeves	86,2	86,8
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	86,0	87,0
$\beta^{CD}(q)$ - Fletcher	87,2	87,4
$\beta^{LS}(q)$ - Liu e Storey	86,8	87,2
$\beta^{DY}(q)$ - Dai e Yuan	87,2	87,4
$\beta^N(q)$ - Hager e Zhang	87,0	87,2
$\beta^H(q)$ - Daniel	86,0	86,0
Quasi-Newton	50,0	50,0
Retropropagação	86,2	86,2

Tabela 13: Desempenho do sistema de inferência fuzzy *non-singleton* em (%), para o distúrbio *sag*.

<i>swell</i>		
Técnica Escolhida	Treinamento	Teste
$\beta^{HS}(q)$ - Hestenes e Stiefel	86,6	86,0
$\beta^{FR}(q)$ - Fletcher e Reeves	86,2	86,2
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	86,0	86,2
$\beta^{CD}(q)$ - Fletcher	87,8	87,6
$\beta^{LS}(q)$ - Liu e Storey	86,2	86,2
$\beta^{DY}(q)$ - Dai e Yuan	87,2	87,2
$\beta^N(q)$ - Hager e Zhang	86,0	86,0
$\beta^H(q)$ - Daniel	83,6	83,6
Quasi-Newton	81,4	81,2
Retropropagação	83,6	83,6

Tabela 14: Desempenho do sistema de inferência fuzzy *non-singleton* em (%), para o distúrbio *swell*.

<i>flicker</i>		
Técnica Escolhida	Treinamento	Teste
$\beta^{HS}(q)$ - Hestenes e Stiefel	97,0	97,8
$\beta^{FR}(q)$ - Fletcher e Reeves	96,8	97,2
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	97,0	97,2
$\beta^{CD}(q)$ - Fletcher	97,0	97,2
$\beta^{LS}(q)$ - Liu e Storey	97,0	97,6
$\beta^{DY}(q)$ - Dai e Yuan	96,8	97,4
$\beta^N(q)$ - Hager e Zhang	97,2	97,8
$\beta^H(q)$ - Daniel	95,4	94,8
Quasi-Newton	50,0	50,0
Retropropagação	96,8	97,4

Tabela 15: Desempenho do sistema de inferência *fuzzy non-singleton* em (%), para o distúrbio *flicker*.

<i>harmônica</i>		
Técnica Escolhida	Treinamento	Teste
$\beta^{HS}(q)$ - Hestenes e Stiefel	84,8	84,6
$\beta^{FR}(q)$ - Fletcher e Reeves	86,0	86,4
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	83,8	84,0
$\beta^{CD}(q)$ - Fletcher	84,6	84,8
$\beta^{LS}(q)$ - Liu e Storey	84,6	84,8
$\beta^{DY}(q)$ - Dai e Yuan	84,0	84,2
$\beta^N(q)$ - Hager e Zhang	85,6	85,4
$\beta^H(q)$ - Daniel	84,6	85
Quasi-Newton	50,0	50,0
Retropropagação	81,8	81,8

Tabela 16: Desempenho do sistema de inferência *fuzzy non-singleton* em (%), para o distúrbio *harmônica*.

A tabela 13 apresenta os resultados para o distúrbio isolado *sag*. Os melhores desempenhos foram obtidos para $\beta^{CD}(q)$, proposto por Fletcher e $\beta^{DY}(q)$, proposto por Dai Yuan, obtendo taxas de classificação de 87,2% para o treinamento e 87,4% para o teste.

Para o distúrbio *swell*, a técnica que obteve melhor desempenho utilizou $\beta^{CD}(q)$, que alcançou 87,8% para treinamento e 87,6% para o teste, conforme apresentado na tabela 14.

Por sua vez, para o distúrbio *flicker*, a abordagem proposta por Hager and Zhang, que faz uso de $\beta^N(q)$, obteve o melhor desempenho com 97,2% de taxa de classificação obtida para o treinamento e 97,8% para o teste, conforme mostra a tabela 15. Já a técnica que utiliza $\beta^{HS}(q)$, proposto por Hestenes e Stiefel obteve desempenho de 97,0% de taxa de classificação para o treinamento e também 97,8% para o teste, podendo ser considerada uma alternativa interessante.

Para o distúrbio harmônica, cujo os resultados foram reportados na tabela 16, temos que o melhor desempenho foi alcançado ao utilizar $\beta^{FR}(q)$, proposto por Fletcher and Reeves, com a taxa de classificação para o treinamento de 86,0% e para o teste de 86,4%. Note que novamente, $\beta^N(q)$, proposto por Hager e Zhang teve um desempenho satisfatório, obtendo as taxas de classificação de 85,6% e 85,4%, para o treinamento e teste, respectivamente, podendo ser considerada uma alternativa promissora.

A seguir, serão apresentados e discutidos os resultados referentes ao sistema de inferência *fuzzy non-singleton* para classificação dos seguintes distúrbios múltiplos: *sag+flicker*, *sag+harmônica*, *swell+flicker*, *swell+harmônica*, *sag+harmônica+flicker* e *swell+harmônica+flicker*.

<i>sag+flicker</i>		
Técnica Escolhida	Treinamento	Teste
$\beta^{HS}(q)$ - Hestenes e Stiefel	87,4	87,4
$\beta^{FR}(q)$ - Fletcher e Reeves	87,4	87,4
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	87,4	87,4
$\beta^{CD}(q)$ - Fletcher	87,4	87,4
$\beta^{LS}(q)$ - Liu e Storey	87,4	87,4
$\beta^{DY}(q)$ - Dai e Yuan	87,4	87,4
$\beta^N(q)$ - Hager e Zhang	87,4	87,4
$\beta^H(q)$ - Daniel	86,8	86,8
Quasi-Newton	50,0	50,0
Retropropagação	87,4	87,4

Tabela 17: Desempenho do sistema de inferência *fuzzy non-singleton* em (%), para o distúrbio *sag+flicker*.

<i>sag+harmônica</i>		
Técnica Escolhida	Treinamento	Teste
$\beta^{HS}(q)$ - Hestenes e Stiefel	80,4	80,2
$\beta^{FR}(q)$ - Fletcher e Reeves	80,0	80,0
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	79,8	79,8
$\beta^{CD}(q)$ - Fletcher	79,8	79,8
$\beta^{LS}(q)$ - Liu e Storey	80,6	80,6
$\beta^{DY}(q)$ - Dai e Yuan	79,8	79,8
$\beta^N(q)$ - Hager e Zhang	79,8	79,8
$\beta^H(q)$ - Daniel	78,4	78,4
Quasi-Newton	76,0	76,0
Retropropagação	80,6	80,6

Tabela 18: Desempenho do sistema de inferência *fuzzy non-singleton* em (%), para o distúrbio *sag+harmônica*.

<i>swell+flicker</i>		
Técnica Escolhida	Treinamento	Teste
$\beta^{HS}(q)$ - Hestenes e Stiefel	86,0	85,8
$\beta^{FR}(q)$ - Fletcher e Reeves	85,8	85,8
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	84,8	84,8
$\beta^{CD}(q)$ - Fletcher	85,8	85,8
$\beta^{LS}(q)$ - Liu e Storey	85,8	85,8
$\beta^{DY}(q)$ - Dai e Yuan	86,4	86,4
$\beta^N(q)$ - Hager e Zhang	86,6	86,6
$\beta^H(q)$ - Daniel	83,4	83,4
Quasi-Newton	50,0	50,0
Retropropagação	85,8	85,8

Tabela 19: Desempenho do sistema de inferência *fuzzy non-singleton* em (%), para o distúrbio *swell+flicker*.

<i>swell+harmônica</i>		
Técnica Escolhida	Treinamento	Teste
$\beta^{HS}(q)$ - Hestenes e Stiefel	86,6	86,6
$\beta^{FR}(q)$ - Fletcher e Reeves	86,4	86,4
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	86,2	86,2
$\beta^{CD}(q)$ - Fletcher	86,6	86,6
$\beta^{LS}(q)$ - Liu e Storey	86,4	86,4
$\beta^{DY}(q)$ - Dai e Yuan	86,4	86,4
$\beta^N(q)$ - Hager e Zhang	86,6	86,6
$\beta^H(q)$ - Daniel	81,8	81,8
Quasi-Newton	82,8	82,8
Retropropagação	86,4	86,4

Tabela 20: Desempenho do sistema de inferência *fuzzy non-singleton* em (%), para o distúrbio *swell+harmônica*.

<i>sag+ harmônica+ flicker</i>		
Técnica Escolhida	Treinamento	Teste
$\beta^{HS}(q)$ - Hestenes e Stiefel	75,4	75,4
$\beta^{FR}(q)$ - Fletcher e Reeves	77,6	77,6
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	77,8	77,8
$\beta^{CD}(q)$ - Fletcher	77,6	77,6
$\beta^{LS}(q)$ - Liu e Storey	77,8	78,0
$\beta^{DY}(q)$ - Dai e Yuan	77,6	77,6
$\beta^N(q)$ - Hager e Zhang	75,6	75,6
$\beta^H(q)$ - Daniel	75,6	75,6
Quasi-Newton	71,2	71,2
Retropropagação	75,6	75,6

Tabela 21: Desempenho do sistema de inferência *fuzzy non-singleton* em (%), para o distúrbio *sag+ harmônica+ flicker*.

<i>swell+ harmônica+ flicker</i>		
Técnica Escolhida	Treinamento	Teste
$\beta^{HS}(q)$ - Hestenes e Stiefel	100,0	100,0
$\beta^{FR}(q)$ - Fletcher e Reeves	100,0	100,0
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	100,0	100,0
$\beta^{CD}(q)$ - Fletcher	100,0	100,0
$\beta^{LS}(q)$ - Liu e Storey	100,0	100,0
$\beta^{DY}(q)$ - Dai e Yuan	100,0	100,0
$\beta^N(q)$ - Hager e Zhang	100,0	100,0
$\beta^H(q)$ - Daniel	100,0	100,0
Quasi-Newton	50,0	50,0
Retropropagação	100,0	100,0

Tabela 22: Desempenho do sistema de inferência *fuzzy non-singleton* em (%), para o distúrbio *swell+ harmônica+ flicker*.

A tabela 17 apresenta os resultados do sistema de inferência *fuzzy non-singleton* para o distúrbio múltiplo *sag+flicker*. Note que os melhores desempenhos foram obtidos para $\beta^{HS}(q)$, $\beta^{FR}(q)$, $\beta^{PRP}(q)$, $\beta^{CD}(q)$, $\beta^{LS}(q)$, $\beta^{DY}(q)$, $\beta^N(q)$. Note que as proposições citadas obtiveram taxas de classificação de 87,4% para o treinamento e teste. É válido mencionar o desempenho obtido pelo algoritmo de retropropagação, que também obteve taxa de classificação de 87,4% para treinamento e teste.

Para o distúrbio múltiplo *sag+harmônica*, a técnica que obteve melhor desempenho fez uso de $\beta^{LS}(q)$, proposto por Liu e Storey, obtendo 80,6% para treinamento e teste, conforme apresentado na tabela 18. É válido ressaltar o desempenho oriundo da utilização do algoritmo de retropropagação. Note que o mesmo propiciou uma taxa de classificação de 80,6% para o treinamento e para o teste, podendo também ser considerada uma abordagem satisfatória.

Por sua vez, para o distúrbio múltiplo *swell+flicker*, a técnica que obteve melhor desempenho fez uso de $\beta^N(q)$, proposto por Hager e Zhang, obtendo 86,6% para treinamento e para teste, conforme apresentado na tabela 19.

Para o distúrbio múltiplo *swell+harmônica*, as técnicas que obtiveram melhor desempenho fizeram uso de $\beta^{HS}(q)$, $\beta^{CD}(q)$ e $\beta^N(q)$, proposto respectivamente por Hestenes e Stiefel, Fletcher e Hager e Zhang, obtendo 86,6% para treinamento e teste, conforme apresentado na tabela 20.

Ao analisar o distúrbio *sag+harmônica+flicker*, cujo os resultados foram reportados na tabela 21, temos que o melhor desempenho foi alcançado utilizando a técnica que faz uso de $\beta^{LS}(q)$, proposto por Liu e Storey, com a taxa de classificação para o treinamento de 77,8% e para o teste de 78,0%. É válido ressaltar que a técnica que faz uso de $\beta^{PRP}(q)$ obteve um desempenho satisfatório, com taxa de classificação de 77,8% para o treinamento e teste, sendo considerada uma alternativa interessante para a resolução do problema de classificação do distúrbio múltiplo em questão.

Por fim, para o distúrbio *swell+harmônica+flicker*, cujo os resultados foram reportados na tabela 22, note que todas as técnicas propostas obtiveram taxa de classificação de 100% para o treinamento e para o teste, à exceção do método Quasi-Newton, que obteve ao longo das 100 épocas de treinamento uma taxa de classificação de 50,0% para treinamento e teste do algoritmo.

Com isso, após apresentados os resultados oriundos de simulações computacionais do sistemas de inferência *fuzzy singleton* e *non-singleton*, obtemos a tabela 23, que contém

os melhores desempenhos obtidos para as técnicas de classificação de distúrbios abordadas neste trabalho.

Distúrbio	Bayes	Rede MLP	<i>Fuzzy s</i>	<i>Fuzzy ns</i>
<i>sag</i>	86,2	87,8	87,8	87,4
<i>swell</i>	82,0	88,8	87,2	87,6
<i>flicker</i>	99,4	97,8	97,8	97,8
harmônica	94,2	88,6	87,2	86,4
<i>sag+flicker</i>	85,8	87,8	87,4	87,4
<i>sag+harmônica</i>	79,4	82,6	80,6	80,6
<i>swell+flicker</i>	80,0	87,0	87,2	86,6
<i>swell+harmônica</i>	86,0	88,0	86,8	86,6
<i>sag+harmônica+flicker</i>	77,6	79,4	78,6	78,0
<i>swell+harmônica+flicker</i>	100	100	100	100

Tabela 23: Desempenho dos classificadores de Bayes, rede neural *perceptron* multicamadas e sistema de inferência *fuzzy singleton* e *non-singleton* em (%), para distúrbios isolados e múltiplos.

De uma forma geral, note que a rede neural *perceptron* multicamadas retornou melhores taxas de classificação para o distúrbio isolado *swell* e para os distúrbios múltiplos *sag+flicker*, *sag+harmônica*, *swell+harmônica*, *sag+harmônica+flicker*. O melhor resultado para o distúrbio harmônica foi obtido através da utilização do classificador de Bayes. Já para o distúrbio múltiplo *swell+flicker*, note que o melhor resultado foi retornado pelo sistema de inferência *fuzzy singleton*. Note que resultados iguais foram obtidos para o distúrbio isolado *flicker*, para o qual os classificadores MLP, *fuzzy singleton* e *non-singleton* retornaram o mesmo resultado. Para o distúrbio isolado *sag*, o classificador *fuzzy singleton* e MLP retornaram os melhores resultados dentre os demais classificadores propostos. Por fim, para o distúrbio múltiplo *swell+harmônica+flicker*, resultados idênticos foram obtidos por todos os classificadores abordados na tabela 23.

Por consequência, propõe-se a investigação e aprofundamento do sistema de inferência *fuzzy singleton* e *non-singleton*, através do aumento do número de épocas de treinamento do algoritmo. Como exemplo, simulou-se novamente o sistema de inferência *fuzzy singleton* e *non-singleton* para o distúrbio *sag*. Note que os resultados foram promissores, com retorno de até 88% de taxa de acerto do algoritmo para treinamento e teste. Com isso, acredita-se na viabilidade de melhora do desempenho da técnica proposta.

6.3 Sumário

Neste capítulo foram apresentados detalhes acerca da extração e seleção de características, discutidos respectivamente nas Seções 2.2.1 e 2.2.2 desta dissertação. Não obstante, foram apresentados os resultados obtidos através de simulações computacionais para os classificadores baseados na teoria de Bayes e na rede neural *perceptron* multicamadas, apresentados respectivamente nas Seções 2.2.3.1 e 2.2.3.2. Adicionalmente, foram apresentados os resultados obtidos através de simulações computacionais para os algoritmos baseados no gradiente conjugado, método Quasi-Newton e retropropagação, que foram apresentados no capítulo 5. Note que para tais técnicas, foram analisadas a convergência, bem como o desempenho das mesmas.

No capítulo seguinte são discutidas as conclusões relativas às técnicas de classificação de distúrbios em sinais elétricas apresentadas neste trabalho.

7 CONCLUSÕES

Sistemas de inferência *fuzzy* têm demonstrado grande aplicabilidade nas mais diversas áreas do conhecimento, tais como economia, matemática difusa, teoria de jogos, modelagem de sistemas, entre outras. É válido ressaltar que estes sistemas permitem a construção de modelos robustos e eficientes para a resolução de problemas de modelagem e mostram-se como uma opção atraente para diversos campos. Sabemos que a classificação ou reconhecimento de distúrbios é um tema de grande interesse no que tange, dentre outros, ao desenvolvimento da próxima geração de equipamentos de monitoramento da QEE. Note que, por serem baseados em regras, os sistemas de inferência *fuzzy* reduzem a dificuldade de modelagem e análise de sistemas complexos, pois os graus de pertinência atribuídos às variáveis fornecem suporte no tratamento da incerteza e da ambiguidade.

O presente trabalho discutiu a derivação de uma técnica de classificação de distúrbios isolados e múltiplos a partir da proposta apresentada em (RIBEIRO; PEREIRA, 2007). Os resultados computacionais indicam que a presente contribuição introduz uma certa melhoria de desempenho no processo de classificação de distúrbios. Conforme relatado no capítulo 6, a classificação de múltiplos distúrbios apresenta um resultado aproximado ou em alguns casos superior quando comparado com o obtido para a classificação de distúrbios isolados.

Neste contexto e dado os objetos de pesquisa discutidos nesta contribuição, o presente trabalho também versou sobre técnicas e metodologias aplicadas à extração de parâmetros utilizando estatísticas de ordem superior (EOS) e a posterior seleção dos referidos parâmetros utilizando o discriminante de Fisher. Adicionalmente, esta dissertação discutiu o uso de técnicas de otimização baseadas no gradiente conjugado e de informações de segunda ordem para o treinamento de sistemas de inferência *fuzzy singleton* ou *non-singleton*. Além disso, as soluções computacionais derivadas foram aplicadas aos problemas de classificação de distúrbios múltiplos e isolados em sinais elétricos.

De uma maneira geral, a análise dos métodos de classificação abordados nesta

dissertação foram feitas baseando-se nos seguintes critérios: a convergência do método e o desempenho do mesmo. Portanto, na otimização de um classificador, deve-se buscar maximizar o seu desempenho e minimizar a sua complexidade computacional. Neste trabalho, conclui-se que a otimização de ambos estes critérios não é uma tarefa trivial.

Conforme citado no capítulo 3, dentre tais questões de investigação, neste trabalho foi dada especial atenção à questão 4, que indaga qual é a técnica de classificação adequada (desempenho x complexidade) para o problema em questão. Basicamente, no capítulo 5 foi apresentado um sistema de inferência *fuzzy* treinado por algoritmos do gradiente conjugado e em informações de segunda ordem (matriz Hessiana), visando, sobretudo, mostrar que a referida técnica baseada em inteligência computacional pode ser uma candidata interessante para responder a questão 4.

Os resultados computacionais, obtidos a partir de dados sintéticos de distúrbios em sinais de tensão, indicam que os sistemas de inferência *fuzzy singleton* ou *non-singleton*, treinados pelos algoritmos de otimização considerados, apresentam maior velocidade de convergência e melhores taxas de classificação quando comparados com aqueles treinados pelo algoritmo de otimização baseada em informações de primeira ordem e é bastante competitivo em relação à MLP e ao classificador de Bayes.

Adicionalmente, constatou-se que o desempenho obtido pelo classificador MLP é ligeiramente superior ao obtido pelo classificador de Bayes e pelo sistema de inferência *fuzzy singleton* e *non-singleton*.

Por fim, como trabalhos futuros, pode-se ressaltar diversos desdobramentos a partir da presente contribuição, que, certamente responderão as questões de investigação supracitadas, que, por sua vez, não foram abordadas neste dissertação. Com isso, destacam-se:

1. **Regras do sistema de inferência *fuzzy*:** Visando a obtenção de melhores resultados, propõe-se desdobramentos acerca das regras utilizadas pelo sistema de inferência *fuzzy singleton* e *non-singleton*. Para este trabalho, note que as funções de pertinência são Gaussianas. Como alternativa, o autor sugere, dentre outros, a adoção de outras funções de pertinência, tais como trapezoidal e triangular, por exemplo. Adicionalmente, como foram utilizadas $M = 4$ regras, sendo 2 regras para a classe com o distúrbio e duas regras para a classe sem distúrbios, é proposto o aumento do número de regras, visando a melhor acuracidade do projeto do sistema de inferência *fuzzy* para a classificação de distúrbios em sinais elétricos.

2. **Sistema de inferência *fuzzy* do tipo 2:** É proposta a expansão da técnica abordada nesta dissertação para o sistema de inferência *fuzzy* do tipo 2, apresentada em (MENDEL, 2001). Esta é uma linha de pesquisa bastante promissora, dado o comportamento satisfatório da técnica em questão e sua abordagem frente as incertezas por ela consideradas. Paralelamente, merece destaque o desenvolvimento de técnicas baseadas em sistemas de inferência *fuzzy* para a identificação de fontes geradoras de distúrbios em sinais elétricos, considerada uma questão em aberto e que merece investigação.
3. **Técnicas de decomposição de sinais:** Ainda como proposição para novos estudos, espera-se testar novas estruturas de filtros para a obtenção dos vetores \mathbf{u} e \mathbf{f} , visando um pré-processamento mais robusto, que maximize as fronteiras de separação entre as classes.
4. **Técnicas de seleção e extração de características:** Uma possível solução para projetar um classificador que alcance um bom desempenho com um reduzido custo computacional está no pré-processamento efetuado e na ferramenta de seleção e extração de características a serem utilizadas. Como proposta, deve-se avaliar possíveis alternativas que sejam capazes de simplificar o processo de classificação de distúrbios. Adicionalmente, é proposta a investigação e avaliação do número de parâmetros selecionados, visando a obtenção de melhores resultados para o classificador de distúrbios utilizado.
5. **Aumento do número de épocas de treinamento do sistema de inferência *fuzzy*:** Sabe-se que o aumento do número de épocas de treinamento do algoritmo proposto acarreta em um sistema de inferência *fuzzy* mais apurado e refinado, fazendo consequentemente com que a porcentagem de taxa de acerto do algoritmo venha a aumentar. Com isso, como sugestão, propõe-se refazer as simulações com o aumento da quantidade de épocas que venha a resultar em melhora do desempenho da técnica proposta.
6. **Matriz Hessiana:** Note que o desempenho pouco satisfatório, sobretudo do algoritmo Quasi-Newton, deve-se ao mau condicionamento da matriz Hessiana $\mathbf{H}(J(\mathbf{w}^{(a)}))$. Trabalhos futuros nesta linha de pesquisa serão determinantes para o sucesso das técnicas que fazem uso da matriz Hessiana. Com isso, o entendimento e melhoria na estrutura de composição da mesma são tópicos fundamentais para a obtenção de resultados satisfatórios.

REFERÊNCIAS

- ARRILLAGA, J.; BOLLEN, M.; WATSON, N. Power quality following deregulation. *Proceedings of the IEEE*, v. 88, n. 2, p. 246–261, feb 2000.
- ATKINSON, K. A. *An introduction to numerical analysis*. 2nd. ed. [S.l.]: John Wiley and Sons, 1988. ISBN 0-471-50023-2.
- AVRIEL, M. *Nonlinear Programming: Analysis and Methods*. [S.l.]: Dover Publishing, 2003. ISBN 0-486-43227-0.
- BAEK, S.; PARK, J. Nonlinear controller optimization of a power system based on reduced multivariate polynomial model. In: *International Joint Conference on Neural Networks*. [S.l.: s.n.], 2009. p. 221–228.
- BOLLEN, H. et al. Classification of underlying causes of power quality disturbances: deterministic versus statistical methods. *EURASIP Journal on Advances in Signal Processing*, v. 1, p. 1–17, 2007.
- BOLLEN, M. H. J.; GU, I. Y. H. *Signal processing of power quality disturbances*. [S.l.]: Wiley-Interscience, 2006.
- CARPINELLI, G.; CHIODO, E.; LAURIA, D. Indices for the characterisation of bursts of short-duration waveform distortion. *IET Generation Transmission and Distribution*, v. 1, n. 1, p. 170–175, 2007.
- CHOONG, F.; REAZ, M.; MOHD, F. Advances in signal processing and artificial intelligence technologies in the classification of power quality events: a survey. *Electric Power Components Systems*, v. 33, n. 12, p. 1333–1349, 2005.
- CHU, P.; MENDEL, J. M. First break refraction event picking using fuzzy logic systems. *IEEE Trans. on Fuzzy Systems*, v. 2, p. 255–266, Nov 1994.
- CHUNG, J. et al. Power disturbance classifier using a rule-based method and wavelet packet-based hidden markov model. *IEEE Trans. on Power Delivery*, v. 17, n. 1, p. 233–241, 2002.
- CHUNG, J. et al. Power disturbance classifier using a rule-based method and wavelet packet-based hidden markov model. *IEEE Transactions on Power Delivery*, v. 17, n. 1, p. 233–241, 2002.
- COLONNESE, S.; SCARANO, G. Transient signal detection using higher order moments. *IEEE Transactions on Signal Processing*, v. 47, n. 2, p. 515–520, Feb. 1999.
- DAI Y. H. ; YUAN, Y. A nonlinear conjugate gradient method with a strong global convergence property. *SIAM J. Optim.*, v. 10, p. 177–182, 1999.

- DANIEL, J. W. The conjugate gradient method for linear and nonlinear operator equations. *SIAM J. Numer. Anal.*, v. 4, p. 10–26, 1967.
- DRIANKOV, D.; HELLENDORRN, H.; REINFRANK, M. *A Introduction to Fuzzy Control*. 2nd. ed. [S.l.]: Springer-Verlag, 1996.
- DUAN, X. et al. Synthetic evaluation of power quality based on fuzzy cluster analysis. In: *Power System Technology, 2006. PowerCon 2006. International Conference on*. [S.l.: s.n.], 2006. p. 1–6.
- FERREIRA, D. et al. HOS-based method for classification of power quality disturbances. *Electronics Letters*, v. 45, n. 3, p. 183–185, Jan. 2009.
- FLETCHER, R. *Practical methods of optimization*. Second edition. New York: John Wiley and Sons, 1987. ISBN 978-0-471-91547-8.
- FLETCHER, R.; REEVES, C. Function minimization by conjugate gradients. *Comput. J.*, v. 7, p. 149–154,, 1964.
- FRISCH, M.; MESSER, H. Transient signal detection using prior information in the likelihood ratio test. *IEEE Transactions on Signal Processing*, v. 41, n. 6, p. 2177–2192, June 1993.
- GAING, Z. L. Wavelet-based neural network for power disturbance recognition and classification. *IEEE Trans. on Power Delivery*, v. 19, n. 4, p. 1560–1568, 2004.
- GHOSH, A.; LUBKEMAN, D. The classification of power system disturbance waveforms using a neural network approach. *IEEE Transactions on Power Delivery*, v. 10, n. 1, p. 109–115, 1995.
- GOLUB, G. H.; LOAN, C. F. V. *Matrix computations*. 3rd. ed. [S.l.]: Johns Hopkins University Press, 1996. ISBN 0-8018-5414-8.
- GROUP, I. . W. *IEEE 519-1995 recommended practices and requirements for harmonic control in electrical power systems institute of electrical and electronics engineers*. [S.l.], May 1992.
- HAGER, W. W.; ZHANG, H. A new conjugate gradient method with guaranteed descent and an efficient line search. *to appear in SIAM J. Optim.*, November 17 2003.
- HAGER, W. W.; ZHANG, H. A survey of nonlinear conjugate gradient methods. *Pacific Journal of Optimization*, v. 2, n. 1, p. 35–58, January 2006.
- HAYKIN, S. *Neural networks. A Comprehensive Foundation*. 2. ed. [S.l.]: Prentice Hall, 1999.
- HELL, M. B. *Abordagem Neurofuzzy para Modelagem de Sistemas Dinâmicos Não Lineares*. Dissertação (Doutorado) — Universidade Estadual de Campinas, 2008.
- HESTENES, M. R.; STIEFEL, E. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, v. 49, n. 6, December 1952.

- HEYDT, G. T. Electric power quality: a tutorial introduction. *IEEE Computer Applications in Power*, v. 11, p. 15–19, Janeiro 1998.
- HOANG, T.; NGUYEN, D. Improving training of radial basis function network for classification of power quality disturbances. *IEEE Transactions on Power Delivery*, v. 21, n. 3, p. 1663–1669, 2006.
- IBRAHIM, A.; MORCOS, M. Artificial intelligence and advanced mathematical tools for power quality applications: a survey. *IEEE Transactions on Power Delivery*, v. 17, n. 2, p. 668–673, 2002.
- JANIK, P.; LOBOS, T. Automated classification of power-quality disturbances using svm and rbf networks. *IEEE Transactions on Power Delivery*, v. 21, n. 3, p. 1663–1669, 2006.
- KARIMI-GHARTEMANI, M.; MOKHTARI, H.; IRAVANI, M. Wavelet based on-line disturbance detection for power quality applications. *IEEE Transactions on Power Delivery*, v. 15, n. 4, p. 1212–1220, 2000.
- KAUFMAN, A.; GUPTA, M. M. *Introduction to Fuzzy Arithmetic: Theory and Applications*. NY: Van Nostrand Reinold, 1991.
- LEE, I.; DASH, P. S-transform-based intelligent system for classification of power quality disturbance signals. *IEEE Transactions on Industrial Electronics*, v. 50, n. 4, p. 800–805, 2003.
- LESKOVICZ, M. et al. Power quality analysis applying a hybrid methodology with wavelet transforms and neural networks. *International Journal of Electrical Power and Energy Systems*, v. 31, n. 5, p. 206–212, 2009.
- LIAO, Y.; LEE, J.-B. A fuzzy-expert system for classifying power quality disturbances. *International Journal of Electrical Power and Energy Systems*, v. 26, n. 3, p. 199 – 205, 2004.
- LIU, Y.; STOREY, C. Efficient generalized conjugate gradient algorithms, part 1: Theory. *J. Optim. Theory Appl.*, v. 69, p. 129–137, 1991.
- LUENBERGER, D. G.; YE, Y. *Linear and nonlinear programming*. 3. ed. [S.l.]: Springer, 2008.
- MCDONOUGH, R. N.; WHALEN, A. D. *Detection of Signals in Noise*. London, UK: Academic Press, 1995.
- MEDSKER, L.; JAIN, L. *Recurrent neural networks: design and applications*. [S.l.]: CRC Press, 2000.
- MENDEL, J. M. Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications. *Proc. of the IEEE*, v. 79, n. 3, p. 278–305, March 1991.
- MENDEL, J. M. Fuzzy logic systems for engineering: A tutorial. In: *Proc. of the IEEE*. [S.l.: s.n.], 1995. v. 83, n. 3, p. 345–377.

- MENDEL, J. M. *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. [S.l.]: Prentice Hall PTR, 2001.
- NIKIAS, C.; MENDEL, J. Signal processing with higher-order spectra. *Signal Processing Magazine, IEEE*, v. 10, n. 3, p. 10–37, Jul. 1993.
- NIKIAS, C. L.; PETROPULU, A. P. *Higher-Order Spectra Analysis A Nonlinear Signal Processing Framework*. NJ, USA: Prentice Hall, 1993.
- NOCEDAL, J.; WRIGHT, S. J. *Numerical Optimization*. [S.l.]: Springer-Verlag, 1999. ISBN 0-387-98793-2.
- PEARLMUTTER, B. A. Fast exact multiplication by the hessian. *Neural Computation*, v. 6, p. 147–160, 1994.
- PEDRYCZ, W.; GOMIDE, F. *Fuzzy Systems Engineering: Toward Human-Centric Computing*. [S.l.]: IEEE Press, 2007.
- POLAK, E.; RIBIÈRE, G. Note sur la convergence de directions conjuguées. *Rev. Francaise Informat Recherche Opertionelle*, v. 3e, n. 16, p. 35–43, 1969.
- POLYAK, B. T. The conjugate gradient method in extreme problems. *USSR Comp. Math. Math. Phys.*, v. 9, p. 94–112, 1969.
- QUALITY, I. S. C. C. . on P. *IEEE Recommended Practice For Monitoring Electric Power Quality*. [S.l.], 1995.
- REAZ, M. et al. Expert system for power quality disturbance classifier. *IEEE Transactions on Power Delivery*, v. 22, n. 3, p. 1979–1988, 2007.
- REAZ, M. B. I. et al. Prototyping of wavelet transform, artificial neural network and fuzzy logic for power quality disturbance classifier. *Electric Power Components and Systems*, v. 35, n. 1, p. 1–17, 2007.
- RIBEIRO, M. V. *Técnicas de Processamento de Sinais Aplicadas à Transmissão de Dados via Rede Elétrica e ao Monitoramento da Qualidade de Energia*. Tese (Doutorado) — UNICAMP, Abril 2005.
- RIBEIRO, M. V. et al. Detection of Disturbances in Voltage Signals for Power Quality Analysis Using HOS. *EURASIP Journal on Advances in Signal Processing*, v. 2007, p. 13–25, 2007.
- RIBEIRO, M. V.; PEREIRA, J. L. R. Classification of single and multiple disturbances in electric signals. *EURASIP Journal on Advances in Signal Processing*, v. 2007, n. 56918, p. 18, 2007.
- SAMANTARAY, S. Decision tree-initialised fuzzy rule-based approach for power quality events classification. *Generation, Transmission Distribution, IET*, v. 4, n. 4, p. 530 –537, April 2010. ISSN 1751-8687.
- SANTOS, E. P. d.; ZUBEN, F. J. V. Recurrent neural networks: design and applications. In: _____. [S.l.]: CRC Press, 2000. cap. 3.

- SANTOSO, S. et al. Power quality disturbance waveform recognition using wavelet-based neural classifier. *IEEE Transactions on Power Delivery*, v. 15, n. 1, p. 222–228, 2000.
- STANDARDIZATION, E. C. for Electro technical. *European Standard EN 50160: Voltage characteristics of electricity supplied by public distribution systems*. [S.l.], 1999.
- SWOKOWSKI, E. W. *Cálculo com Geometria Analítica*. São Paulo: Makron Books, 1994.
- TEKE, A.; BAYINDIR, K.; TUMAY, M. Fast sag/swell detection method for fuzzy logic controlled dynamic voltage restorer. *Generation, Transmission Distribution, IET*, v. 4, n. 1, p. 1–12, January 2010. ISSN 1751-8687.
- THEODORIDIS, S.; KOUTROUMBAS, K. *Pattern Recognition*. San Diego: Academic Press, 1999.
- TREES, H. L. van. *Detection, Estimation and Modulation Theory, Part I*. New York, NY, USA: Springer, 1968.
- TREES, H. L. van. *Detection, Estimation and Modulation Theory, Part III*. New York, NY, USA: Springer, 1971.
- WANG, L.-X. *Adaptative Fuzzy Systems and Control: Design and Stability Analysis*. Englewood Cliffs, NJ: PTR Prentice-Hall, 1994.
- WANG, L.-X. *A course in Fuzzy Systems and Control*. Upper Saddle River, NJ: Prentice-Hall, 1997.
- WANG, L.-X.; MENDEL, J. M. Fuzzy basis functions, universal approximation and orthogonal least squares learning. *IEEE Trans. on Neural Networks*, v. 3, p. 807–813, Sept 1992a.
- WANG, M.; MAMISHEV, A. V. Classification of power quality events using optimal time-frequency representations - part 1: Theory. *IEEE Transactions on Power Delivery*, v. 19, n. 3, p. 1488–1945, 2004.
- WIJAYAKULASOORIYA J.V., P. G. M. P. Electric power quality disturbance classification using self-adapting artificial neural networks. *IEE Proceeding Generation, Transmsission and Distribution*, v. 149, n. 1, p. 98–101, 2002.
- WILLE, J. On the structure of the hessian matrix in feedforward networks and second derivative methods. In: *International Conference on Neural Networks*. [S.l.: s.n.], 1997. v. 3.
- YU, C.; MANRY, M. A hessian matrix approach for training nonlinear networks. In: *International Conference on Signal Processing*. [S.l.: s.n.], 2004. v. 2, p. 1514–1517.
- ZENG, W.; LI, X.; ZHANG, X. Adaptive newton algorithms for blind equalization using the generalized constant modulus criterion. In: *IEEE International Conference on Acoustics, Speech and Signal Processing*. [S.l.: s.n.], 2009. p. 2805–2808.

APÊNDICE A - PRODUÇÃO BIBLIOGRÁFICA

A.1 Artigos em revista

de Aguiar, E. P.; Marques, C. A. G.; Duque, C. A.; Ribeiro, M. V. “*Signal Decomposition With Reduced Complexity for Classification of Isolated and Multiple Disturbances in Electric Signals*”. *IEEE Transactions on Power Delivery*, vol.24, no.4, p.2459-2460, Oct. 2009

A.2 Artigos em congressos

Aguiar, E. P.; Marques, Marques, C. A. G.; Cerqueira, A. S.; Duque, C. A.; Ribeiro, M. V. “Classificação de Distúrbios para QEE utilizando EOS e Redes Neurais”. VIII Conferência Brasileira sobre Qualidade da Energia Elétrica, 2009, Blumenau/SC, Brasil, p. 300-305.

APÊNDICE B – CÓDIGO EM MATLAB[®] UTILIZADO PARA A GERAÇÃO DOS BANCOS DE DADOS SINTÉTICOS

B.1 Rotina principal

```

09/08/11 20:40 C:\Documents and Settings\eaquiuar\Desk...\forma banco total.m 1 of 5

clc,clear all,close all

tic

%especificações gerais
Nsinais=500; %número de sinais em cada banco
f0=60; %frequência fundamental do sinal
fs=256*60; %frequência de amostragem
Nciclos=10; %número de ciclos do sinal
pt=6/60; %ponto de início do distúrbio
SNR=30; %SNR presente em cada sinal do banco

%transitório oscilatório
tt_min=.3e-3;
tt_max=50e-3;
ft_min=2000;
ft_max=5000;
at_min=1;
at_max=4;
banco_transitorio_10ciclos = gera_banco_transitorio(Nsinais,f0,fs,Nciclos,pt,tt_min,
tt_max,ft_min,ft_max,at_min,at_max,SNR);

%harmônico
ordem_max=100;
ordens=1;
THD_min=1.5;
THD_max=5;
banco_harmonico_10ciclos = gera_banco_harmonicos(Nsinais,f0,fs,Nciclos,0,ordem_max,
THD_min,THD_max,ordens,SNR);

%swell
a_min=1.1;
a_max=1.9;
banco_swell_10ciclos = gera_banco_variacoes_amplitude(Nsinais,f0,fs,Nciclos,pt,a_min,
a_max,SNR);

%sag
a_min=0.2;
a_max=0.9;
banco_sag_10ciclos = gera_banco_variacoes_amplitude(Nsinais,f0,fs,Nciclos,pt,a_min,
a_max,SNR);

%interrupcao
a_min=0;
a_max=0.1;
banco_interrupcao_10ciclos = gera_banco_variacoes_amplitude(Nsinais,f0,fs,Nciclos,pt,
a_min,a_max,SNR);

%flutuações de tensão
a_min=3/100;
a_max=7/100;
f_min=1;
f_max=25;
banco_flicker_10ciclos = gera_banco_flutuacoes_tensao(Nsinais,f0,fs,Nciclos,pt,a_min,
a_max,f_min,f_max,SNR);

```


09/08/11 20:40 C:\Documents and Settings\eaquiar\Desk...\forma banco total.m 2 of 5

```
%notches e speikes
for n=1:Nsinais
    banco_notch_10ciclos(:,n)=evento_notch;
    banco_spike_10ciclos(:,n)=evento_spike;
end

%sag+flicker
a_minf=3/100;
a_maxf=7/100;
f_min=1;
f_max=25;
banco_flicker_10ciclos_aux1 = gera_banco_flutuacoes_tensao(Nsinais,f0,fs,Nciclos,pt,
a_minf,a_maxf,f_min,f_max,inf);
a_mins=0.2;
a_maxs=0.9;
banco_sag_flicker_10ciclos = aplica_swell_sag(banco_flicker_10ciclos_aux1,SNR,a_mins,
a_maxs,fs,Nciclos,f0,pt,Nsinais);

%sag+harmonico
ordem_max=100;
ordens=1;
THD_min=1.5;
THD_max=5;
banco_harmonico_10ciclos_aux1 = gera_banco_harmonicos(Nsinais,f0,fs,Nciclos,0,
ordem_max,THD_min,THD_max,ordens,inf);
a_mins=0.2;
a_maxs=0.9;
banco_sag_harmonico_10ciclos = aplica_swell_sag(banco_harmonico_10ciclos_aux1,SNR,
a_mins,a_maxs,fs,Nciclos,f0,pt,Nsinais);

%swell+flicker
a_minf=3/100;
a_maxf=7/100;
f_min=1;
f_max=25;
banco_flicker_10ciclos_aux2 = gera_banco_flutuacoes_tensao(Nsinais,f0,fs,Nciclos,pt,
a_minf,a_maxf,f_min,f_max,inf);
a_mins=1.1;
a_maxs=1.9;
banco_swell_flicker_10ciclos = aplica_swell_sag(banco_flicker_10ciclos_aux2,SNR,
a_mins,a_maxs,fs,Nciclos,f0,pt,Nsinais);

%swell+harmonico
ordem_max=100;
ordens=1;
THD_min=1.5;
THD_max=5;
banco_harmonico_10ciclos_aux2 = gera_banco_harmonicos(Nsinais,f0,fs,Nciclos,0,
ordem_max,THD_min,THD_max,ordens,inf);
a_mins=1.1;
a_maxs=1.9;
banco_swell_harmonico_10ciclos = aplica_swell_sag(banco_harmonico_10ciclos_aux2,SNR,
a_mins,a_maxs,fs,Nciclos,f0,pt,Nsinais);
```

09/08/11 20:40 C:\Documents and Settings\eaquiar\Desk...\forma banco total.m 3 of 5

```
%sag+harmonico+flicker
ordem_max=100;
ordens=1;
THD_min=1.5;
THD_max=5;
banco_harmonico_10ciclos_aux3 = gera_banco_harmonicos_aux(Nsinais,f0,fs,Nciclos,0,
ordem_max,THD_min,THD_max,ordens,inf);

a_minf=3/100;
a_maxf=7/100;
f_minf=1;
f_maxf=25;
banco_flicker_10ciclos_aux3 = gera_banco_flutuacoes_tensao(Nsinais,f0,fs,Nciclos,pt,
a_minf,a_maxf,f_minf,f_maxf,inf);

a_mins=0.2;
a_maxs=0.9;
banco_sag_harmonico_flicker_10ciclos = aplica_swell_sag
(banco_harmonico_10ciclos_aux3+banco_flicker_10ciclos_aux3,SNR,a_mins,a_maxs,fs,
Nciclos,f0,pt,Nsinais);

%sag+harmonico+capc
ordem_max=100;
ordens=1;
THD_min=1.5;
THD_max=5;
banco_harmonico_10ciclos_aux4 = gera_banco_harmonicos_aux(Nsinais,f0,fs,Nciclos,0,
ordem_max,THD_min,THD_max,ordens,inf);

tt_min=.3e-3;
tt_max=50e-3;
ft_min=2000;
ft_max=5000;
at_min=1;
at_max=4;
banco_transitorio_10ciclos_aux1 = gera_banco_transitorio(Nsinais,f0,fs,Nciclos,7/60,
tt_min,tt_max,ft_min,ft_max,at_min,at_max,SNR);

a_mins=0.2;
a_maxs=0.9;
banco_sag_harmonico_transitorio_10ciclos = aplica_swell_sag
(banco_harmonico_10ciclos_aux4+banco_transitorio_10ciclos_aux1,SNR,a_mins,a_maxs,fs,
Nciclos,f0,pt,Nsinais);

%sag+harmonico+notch
ordem_max=100;
ordens=1;
THD_min=1.5;
THD_max=5;
banco_harmonico_10ciclos_aux3 = gera_banco_harmonicos_aux(Nsinais,f0,fs,Nciclos,0,
ordem_max,THD_min,THD_max,ordens,inf);

for n=1:Nsinais
    banco_notch_10ciclos_aux(:,n)=evento_notch;
end
```

09/08/11 20:40 C:\Documents and Settings\eaquiar\Desk...\forma banco total.m 4 of 5

```
a_mins=0.2;
a_maxs=0.9;
banco_sag_harmonico_notch_10ciclos = aplica_swell_sag
(banco_harmonico_10ciclos_aux3+banco_notch_10ciclos_aux,inf,a_mins,a_maxs,fs,Nciclos,
f0,pt,Nsinais);

%swell+harmonico+flicker
ordem_max=100;
ordens=1;
THD_min=1.5;
THD_max=5;
banco_harmonico_10ciclos_aux3 = gera_banco_harmonicos_aux(Nsinais,f0,fs,Nciclos,0,
ordem_max,THD_min,THD_max,ordens,inf);

a_minf=3/100;
a_maxf=7/100;
f_minf=1;
f_maxf=25;
banco_flicker_10ciclos_aux3 = gera_banco_flutuacoes_tensao(Nsinais,f0,fs,Nciclos,pt,
a_minf,a_maxf,f_minf,f_maxf,inf);

a_mins=1.1;
a_maxs=1.9;
banco_swell_harmonico_flicker_10ciclos = aplica_swell_sag
(banco_harmonico_10ciclos_aux3+banco_flicker_10ciclos_aux3,SNR,a_mins,a_maxs,fs,
Nciclos,f0,pt,Nsinais);

%swell+harmonico+capc
ordem_max=100;
ordens=1;
THD_min=1.5;
THD_max=5;
banco_harmonico_10ciclos_aux4 = gera_banco_harmonicos_aux(Nsinais,f0,fs,Nciclos,0,
ordem_max,THD_min,THD_max,ordens,inf);

tt_min=.3e-3;
tt_max=50e-3;
ft_min=2000;
ft_max=5000;
at_min=1;
at_max=4;
banco_transitorio_10ciclos_aux1 = gera_banco_transitorio(Nsinais,f0,fs,Nciclos,7/60,
tt_min,tt_max,ft_min,ft_max,at_min,at_max,SNR);

a_mins=1.1;
a_maxs=1.9;
banco_swell_harmonico_transitorio_10ciclos = aplica_swell_sag
(banco_harmonico_10ciclos_aux4+banco_transitorio_10ciclos_aux1,SNR,a_mins,a_maxs,fs,
Nciclos,f0,pt,Nsinais);

%swell+harmonico+notch
ordem_max=100;
ordens=1;
THD_min=1.5;
```

09/08/11 20:40 C:\Documents and Settings\eaquiar\Desk...\forma banco total.m 5 of 5

```
THD_max=5;
banco_harmonico_10ciclos_aux3 = gera_banco_harmonicos_aux(Nsinais,f0,fs,Nciclos,0,
ordem_max,THD_min,THD_max,ordens,inf);

for n=1:Nsinais
    banco_notch_10ciclos_aux(:,n)=evento_notch;
end

a_mins=1.1;
a_maxs=1.9;
banco_swell_harmonico_notch_10ciclos = aplica_swell_sag
(banco_harmonico_10ciclos_aux3+banco_notch_10ciclos_aux,inf,a_mins,a_maxs,fs,Nciclos,
f0,pt,Nsinais);

save banco_transitorio_10ciclos banco_transitorio_10ciclos
save banco_harmonico_10ciclos banco_harmonico_10ciclos
save banco_swell_10ciclos banco_swell_10ciclos
save banco_sag_10ciclos banco_sag_10ciclos
save banco_interrupcao_10ciclos banco_interrupcao_10ciclos
save banco_flicker_10ciclos banco_flicker_10ciclos
save banco_notch_10ciclos banco_notch_10ciclos
save banco_spike_10ciclos banco_spike_10ciclos

save banco_sag_flicker_10ciclos banco_sag_flicker_10ciclos
save banco_sag_harmonico_10ciclos banco_sag_harmonico_10ciclos
save banco_swell_flicker_10ciclos banco_swell_flicker_10ciclos
save banco_swell_harmonico_10ciclos banco_swell_harmonico_10ciclos

save banco_sag_harmonico_flicker_10ciclos banco_sag_harmonico_flicker_10ciclos
save banco_sag_harmonico_transitorio_10ciclos
banco_sag_harmonico_transitorio_10ciclos
save banco_sag_harmonico_notch_10ciclos banco_sag_harmonico_notch_10ciclos
save banco_swell_harmonico_flicker_10ciclos banco_swell_harmonico_flicker_10ciclos
save banco_swell_harmonico_transitorio_10ciclos
banco_swell_harmonico_transitorio_10ciclos
save banco_swell_harmonico_notch_10ciclos banco_swell_harmonico_notch_10ciclos

toc;

tempo=toc/60
```

APÊNDICE C – MATRIZ HESSIANA PARA O SISTEMA DE INFERÊNCIA FUZZY

C.1 Dedução da matriz Hessiana para o sistema de inferência *fuzzy singleton*

C.1.1 Dedução de $\mathbf{H}_{m_{F_j^i}(q), \sigma_{F_k^l}(q)}$

Os termos de $\mathbf{H}_{m_{F_j^i}(q), \sigma_{F_k^l}(q)}$, expresso na equação (5.13), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{\sigma_{F_k^l}(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $m_{F_j^i}(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} = \frac{\partial}{\partial m_{F_j^i}(q)} [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q). \quad (\text{C.1})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} = \frac{\partial}{\partial m_{F_j^i}(q)} [RSTU], \quad (\text{C.2})$$

de tal forma que

$$R = [f_s(\mathbf{x}^{(q)}) - y^{(q)}], \quad (\text{C.3})$$

$$S = [\theta_l(q) - f_s(\mathbf{x}^{(q)})], \quad (\text{C.4})$$

$$T = \phi_l(\mathbf{x}^{(q)}) \quad (\text{C.5})$$

e

$$U = b_{F_k^l}(q). \quad (\text{C.6})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos sim-

plificar a derivada parcial retratada na equação (C.2). Com isso temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} = & \left(\frac{\partial R}{\partial m_{F_j^i}(q)} \right) STU + R \left(\frac{\partial S}{\partial m_{F_j^i}(q)} \right) TU \\ & + RS \left(\frac{\partial T}{\partial m_{F_j^i}(q)} \right) U + RST \left(\frac{\partial U}{\partial m_{F_j^i}(q)} \right), \end{aligned} \quad (\text{C.7})$$

Note que as derivadas parciais retratadas na equação (C.7) podem ser expressas por

$$\frac{\partial R}{\partial m_{F_j^i}(q)} = [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q), \quad \forall i, \quad (\text{C.8})$$

$$\frac{\partial S}{\partial m_{F_j^i}(q)} = -[\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) a_{F_k^l}(q), \quad \forall i, \quad (\text{C.9})$$

$$\frac{\partial T}{\partial m_{F_j^i}(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q), & i \neq l \\ [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] a_{F_k^l}(q), & i = l \end{cases} \quad (\text{C.10})$$

e

$$\frac{\partial U}{\partial m_{F_j^i}(q)} = \begin{cases} 0, & i \neq l \\ 0, & i = l, \quad j \neq k \\ \frac{-2(x_k^{(q)} - m_{F_k^l}(q))}{\sigma_{F_k^l}^3(q)}, & i = l, \quad j = k \end{cases}. \quad (\text{C.11})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \sigma_{F_k^l}(q)}$, para a situação em que $i \neq l$, temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} = & e_{\theta_i}^{(q)} \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q) e_{\theta_l}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \\ & - e^{(q)} e_{\theta_i}^{(q)} \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \\ & - e^{(q)} e_{\theta_l}^{(q)} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q) b_{F_k^l}(q). \end{aligned} \quad (\text{C.12})$$

Organizando os termos da equação (C.12), temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} = \left[e_{\theta_i}^{(q)} e_{\theta_l}^{(q)} - e^{(q)} e_{\theta_i}^{(q)} - e^{(q)} e_{\theta_l}^{(q)} \right] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q) b_{F_k^l}(q). \quad (\text{C.13})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \sigma_{F_k^l}(q)}$ para a situação em que $i = l$ temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} = & e_{\theta_l}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_j^l}(q) e_{\theta_l}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \\ & - e^{(q)} e_{\theta_l}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_j^l}(q) \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \\ & + e^{(q)} e_{\theta_l}^{(q)} [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] a_{F_j^l}(q) \\ & + e^{(q)} e_{\theta_l}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial U}{\partial m_{F_k^l}(q)}, \end{aligned} \quad (\text{C.14})$$

ou seja, (C.14) pode ser reescrita da seguinte maneira

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^l}(q) \partial \sigma_{F_k^l}(q)} = & \left[\left(e_{\theta_l(q)}^{(q)} \right)^2 \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} e_{\theta_l(q)}^{(q)} \left(\phi_l(\mathbf{x}^{(q)}) - 2\phi_l^2(\mathbf{x}^{(q)}) \right) \right] a_{F_j^l}(q) b_{F_k^l}(q) \\ & + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial U}{\partial m_{F_j^l}(q)}. \end{aligned} \quad (\text{C.15})$$

Com isso, ordenando os termos da equação (C.15), temos a seguinte equação para $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \sigma_{F_k^l}(q)}$ quando $i = l$

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^l}(q) \partial \sigma_{F_k^l}(q)} = & \left[\left(e_{\theta_l(q)}^{(q)} \right)^2 - 2e^{(q)} e_{\theta_l(q)}^{(q)} \right] \phi_l^2(\mathbf{x}^{(q)}) a_{F_j^l}(q) b_{F_k^l}(q) \\ & + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \left[\frac{\partial U}{\partial m_{F_j^l}(q)} + a_{F_j^l}(q) b_{F_k^l}(q) \right], \end{aligned} \quad (\text{C.16})$$

em que $\frac{\partial U}{\partial m_{F_j^l}(q)}$ assume os seguintes valores

$$\frac{\partial U}{\partial m_{F_j^l}(q)} = \begin{cases} 0, & j \neq k \\ -2 \left(\frac{x_k^{(q)} - m_{F_k^l}(q)}{\sigma_{F_k^l}^3(q)} \right), & j = k \end{cases}. \quad (\text{C.17})$$

C.1.2 Dedução de $\mathbf{H}_{m_{F_j^i}(q), \theta_l(q)}$

Os termos de $\mathbf{H}_{m_{F_j^i}(q), \theta_l(q)}$, expresso na equação (5.14), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{\theta_l(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $m_{F_j^i}(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \theta_l(q)} = \frac{\partial}{\partial m_{F_j^i}(q)} [f_s(\mathbf{x}^{(q)}) - y^{(q)}] \phi_l(\mathbf{x}^{(q)}). \quad (\text{C.18})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \theta_l(q)} = \frac{\partial}{\partial m_{F_j^i}(q)} [EF]. \quad (\text{C.19})$$

Note que as derivadas parciais retratadas na equação (C.7) podem ser expressas por

$$E = [f_s(\mathbf{x}^{(q)}) - y^{(q)}] \quad (\text{C.20})$$

e

$$F = \phi_l(\mathbf{x}^{(q)}). \quad (\text{C.21})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos sim-

plificar a derivada parcial retratada na equação (C.19). Com isso temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \theta_l(q)} = \left(\frac{\partial E}{\partial m_{F_j^i}(q)} \right) F + E \left(\frac{\partial F}{\partial m_{F_j^i}(q)} \right) \quad (\text{C.22})$$

em que

$$\frac{\partial E}{\partial m_{F_j^i}(q)} = [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q), \quad \forall i, \quad (\text{C.23})$$

e

$$\frac{\partial F}{\partial m_{F_j^i}(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q), & i \neq l \\ [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] a_{F_j^i}(q), & i = l \end{cases}. \quad (\text{C.24})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \theta_l(q)}$, para a situação em que $i \neq l$, temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \theta_l(q)} = e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) - e^{(q)} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q). \quad (\text{C.25})$$

Organizando os termos da equação (C.25), temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \theta_l(q)} = [e_{\theta_i(q)}^{(q)} - e^{(q)}] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q). \quad (\text{C.26})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \theta_l(q)}$ para a situação em que $i = l$ temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \theta_l(q)} = e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_j^l}(q) \phi_l(\mathbf{x}^{(q)}) + e^{(q)} [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] a_{F_j^l}(q), \quad (\text{C.27})$$

o que remete à equação

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \theta_l(q)} = \left[\left(e_{\theta_i(q)}^{(q)} - e^{(q)} \right) \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} \phi_l(\mathbf{x}^{(q)}) \right] a_{F_j^l}(q). \quad (\text{C.28})$$

C.1.3 Dedução de $\mathbf{H}_{\sigma_{F_j^i}(q), m_{F_k^l}(q)}$

Os termos de $\mathbf{H}_{\sigma_{F_j^i}(q), m_{F_k^l}(q)}$, expresso na equação (5.15), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{m_{F_k^l}(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $\sigma_{F_j^i}(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial m_{F_k^l}(q)} = \frac{\partial}{\partial \sigma_{F_j^i}(q)} [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q). \quad (\text{C.29})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial m_{F_k^l}(q)} = \frac{\partial}{\partial \sigma_{F_j^i}(q)} [XYWZ], \quad (\text{C.30})$$

de tal forma que

$$X = [f_s(\mathbf{x}^{(q)}) - y^{(q)}], \quad (\text{C.31})$$

$$Y = [\theta_l(q) - f_s(\mathbf{x}^{(q)})], \quad (\text{C.32})$$

$$W = \phi_l(\mathbf{x}^{(q)}) \quad (\text{C.33})$$

e

$$Z = a_{F_k^l}(q). \quad (\text{C.34})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (C.30). Com isso temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial m_{F_k^l}(q)} &= \left(\frac{\partial X}{\partial \sigma_{F_j^i}(q)} \right) YWZ + X \left(\frac{\partial Y}{\partial \sigma_{F_j^i}(q)} \right) WZ \\ &+ XY \left(\frac{\partial W}{\partial \sigma_{F_j^i}(q)} \right) Z + XYW \left(\frac{\partial Z}{\partial \sigma_{F_j^i}(q)} \right). \end{aligned} \quad (\text{C.35})$$

Note que as derivadas parciais retratadas na equação (C.35) podem ser expressas por

$$\frac{\partial X}{\partial \sigma_{F_j^i}(q)} = [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q), \quad \forall i, \quad (\text{C.36})$$

$$\frac{\partial Y}{\partial \sigma_{F_j^i}(q)} = -[\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) b_{F_k^l}(q), \quad \forall i, \quad (\text{C.37})$$

$$\frac{\partial W}{\partial \sigma_{F_j^i}(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_j^i}(q), & i \neq l \\ [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] b_{F_k^l}(q), & i = l \end{cases} \quad (\text{C.38})$$

e

$$\frac{\partial Z}{\partial \sigma_{F_j^i}(q)} = \begin{cases} 0, & i \neq l \\ 0, & i = l, j \neq k \\ \frac{-2(x_k^{(q)} - m_{F_k^l}(q))}{\sigma_{F_k^l}^3(q)}, & i = l, j = k \end{cases}. \quad (\text{C.39})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial m_{F_k^l}(q)}$ para a situação em que $i \neq l$ temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial m_{F_k^l}(q)} &= e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q) e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &- e^{(q)} e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &- e^{(q)} e_{\theta_l(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_j^i}(q) a_{F_k^l}(q). \end{aligned} \quad (\text{C.40})$$

Organizando os termos da equação (C.40), temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial m_{F_k^l}(q)} = \left[e_{\theta_i(q)}^{(q)} e_{\theta_l(q)}^{(q)} - e^{(q)} e_{\theta_i(q)}^{(q)} - e^{(q)} e_{\theta_l(q)}^{(q)} \right] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_j^i}(q) a_{F_k^l}(q). \quad (\text{C.41})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial m_{F_k^l}(q)}$ para a situação em que $i = l$, temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial m_{F_k^l}(q)} &= e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_j^l}(q) e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &\quad - e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_j^l}(q) \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \left[\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)}) \right] b_{F_j^l}(q) a_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial Z}{\partial \sigma_{F_j^l}(q)}, \end{aligned} \quad (\text{C.42})$$

ou seja, (C.42) pode ser reescrita da seguinte maneira

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^l}(q) \partial m_{F_k^l}(q)} &= \left[\left(e_{\theta_l(q)}^{(q)} \right)^2 \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} e_{\theta_l(q)}^{(q)} \left(\phi_l(\mathbf{x}^{(q)}) - 2\phi_l^2(\mathbf{x}^{(q)}) \right) \right] a_{F_j^l}(q) b_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial Z}{\partial \sigma_{F_j^l}(q)}. \end{aligned} \quad (\text{C.43})$$

Com isso, ordenando os termos da equação (C.43), finalmente obtemos

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^l}(q) \partial m_{F_k^l}(q)} &= \left[\left(e_{\theta_l(q)}^{(q)} \right)^2 - 2e^{(q)} e_{\theta_l(q)}^{(q)} \right] \phi_l^2(\mathbf{x}^{(q)}) a_{F_j^l}(q) b_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \left[\frac{\partial Z}{\partial \sigma_{F_j^l}(q)} + a_{F_j^l}(q) b_{F_k^l}(q) \right], \end{aligned} \quad (\text{C.44})$$

em que $\frac{\partial Z}{\partial \sigma_{F_j^l}(q)}$ assume os seguintes valores

$$\frac{\partial Z}{\partial \sigma_{F_j^l}(q)} = \begin{cases} 0, & j \neq k \\ -2 \frac{(x_k^{(q)} - m_{F_k^l}(q))}{\sigma_{F_k^l}^3(q)}, & j = k \end{cases}. \quad (\text{C.45})$$

C.1.4 Dedução de $\mathbf{H}_{\sigma_{F_j^i}(q), \sigma_{F_k^l}(q)}$

Os termos de $\mathbf{H}_{\sigma_{F_j^i}(q), \sigma_{F_k^l}(q)}$, expresso na equação (5.16), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{\sigma_{F_k^l}(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $\sigma_{F_j^i}(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} = \frac{\partial}{\partial \sigma_{F_j^i}(q)} [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q). \quad (\text{C.46})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} = \frac{\partial}{\partial \sigma_{F_j^i}(q)} [RSTU], \quad (\text{C.47})$$

de tal forma que

$$R = [f_s(\mathbf{x}^{(q)}) - y^{(q)}], \quad (\text{C.48})$$

$$S = [\theta_l(q) - f_s(\mathbf{x}^{(q)})], \quad (\text{C.49})$$

$$T = \phi_l(\mathbf{x}^{(q)}) \quad (\text{C.50})$$

e

$$U = b_{F_k^l}(q). \quad (\text{C.51})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (C.47). Com isso temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} &= \left(\frac{\partial R}{\partial \sigma_{F_j^i}(q)} \right) STU + R \left(\frac{\partial S}{\partial \sigma_{F_j^i}(q)} \right) TU \\ &+ RS \left(\frac{\partial T}{\partial \sigma_{F_j^i}(q)} \right) U + RST \left(\frac{\partial U}{\partial \sigma_{F_j^i}(q)} \right). \end{aligned} \quad (\text{C.52})$$

Note que as derivadas parciais retratadas na equação (C.52) podem ser expressas por

$$\frac{\partial R}{\partial \sigma_{F_j^i}(q)} = [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q), \quad \forall i, \quad (\text{C.53})$$

$$\frac{\partial S}{\partial \sigma_{F_j^i}(q)} = -[\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) b_{F_k^l}(q), \quad \forall i, \quad (\text{C.54})$$

$$\frac{\partial T}{\partial \sigma_{F_j^i}(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_j^i}(q), & i \neq l \\ [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] b_{F_k^l}(q), & i = l \end{cases} \quad (\text{C.55})$$

e

$$\frac{\partial U}{\partial \sigma_{F_j^i}(q)} = \begin{cases} 0, & i \neq l \\ 0, & i = l, \quad j \neq k \\ \frac{-3(x_k^{(q)} - m_{F_k^l}(q))^2}{\sigma_{F_k^l}^4(q)}, & i = l, \quad j = k \end{cases}. \quad (\text{C.56})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \sigma_l(q)}$, para a situação em que $i \neq l$, temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} &= e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q) e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \\ &- e^{(q)} e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \\ &- e^{(q)} e_{\theta_l(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_j^i}(q) b_{F_k^l}(q). \end{aligned} \quad (\text{C.57})$$

Organizando os termos da equação (C.57), temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} = \left[e_{\theta_i^{(q)}}^{(q)} e_{\theta_l^{(q)}}^{(q)} - e^{(q)} e_{\theta_i^{(q)}}^{(q)} - e^{(q)} e_{\theta_l^{(q)}}^{(q)} \right] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_j^i}(q) b_{F_k^l}(q). \quad (\text{C.58})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \sigma_{F_k^l}(q)}$ para a situação em que $i = l$, temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} &= e_{\theta_l^{(q)}}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_j^l}(q) e_{\theta_l^{(q)}}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \\ &\quad - e^{(q)} e_{\theta_l^{(q)}}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_j^l}(q) \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l^{(q)}}^{(q)} [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] b_{F_j^l}(q) b_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l^{(q)}}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial U}{\partial \sigma_{F_j^l}(q)}, \end{aligned} \quad (\text{C.59})$$

ou seja, (C.59) pode ser reescrita da seguinte maneira

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^l}(q) \partial \sigma_{F_k^l}(q)} &= \left[\left(e_{\theta_l^{(q)}}^{(q)} \right)^2 \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} e_{\theta_l^{(q)}}^{(q)} (\phi_l(\mathbf{x}^{(q)}) - 2\phi_l^2(\mathbf{x}^{(q)})) \right] b_{F_j^l}(q) b_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l^{(q)}}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial U}{\partial \sigma_{F_j^l}(q)}. \end{aligned} \quad (\text{C.60})$$

Com isso, ordenando os termos da equação (C.60), temos a seguinte equação para

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^l}(q) \partial \sigma_{F_k^l}(q)} &= \left[\left(e_{\theta_l^{(q)}}^{(q)} \right)^2 - 2e^{(q)} e_{\theta_l^{(q)}}^{(q)} \right] \phi_l^2(\mathbf{x}^{(q)}) b_{F_j^l}(q) b_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l^{(q)}}^{(q)} \phi_l(\mathbf{x}^{(q)}) \left[\frac{\partial U}{\partial \sigma_{F_j^l}(q)} + b_{F_j^l}(q) b_{F_k^l}(q) \right], \end{aligned} \quad (\text{C.61})$$

em que $\frac{\partial U}{\partial \sigma_{F_j^l}(q)}$ assume os seguintes valores

$$\frac{\partial U}{\partial \sigma_{F_j^l}(q)} = \begin{cases} 0, & i \neq l, j \neq k, \\ \frac{-3(x_k^{(q)} - m_{F_k^l}(q))^2}{\sigma_{F_k^l}^4(q)}, & j = k \end{cases}. \quad (\text{C.62})$$

C.1.5 Dedução de $\mathbf{H}_{\sigma_{F_j^i}(q), \theta_l(q)}$

Os termos de $\mathbf{H}_{\sigma_{F_j^i}(q), \theta_l(q)}$, expresso na equação (5.17), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{\theta_l(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $\sigma_{F_j^i}(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \theta_l(q)} = \frac{\partial}{\partial \sigma_{F_j^i}(q)} [f_s(\mathbf{x}^{(q)}) - y^{(q)}] \phi_l(\mathbf{x}^{(q)}). \quad (\text{C.63})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \theta_l(q)} = \frac{\partial}{\partial \sigma_{F_j^i}(q)} [EF]. \quad (\text{C.64})$$

de tal forma que

$$E = [f_s(\mathbf{x}^{(q)}) - y^{(q)}] \quad (\text{C.65})$$

e

$$F = \phi_l(\mathbf{x}^{(q)}). \quad (\text{C.66})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (C.64). Com isso temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \theta_l(q)} = \left(\frac{\partial E}{\partial \sigma_{F_j^i}(q)} \right) F + E \left(\frac{\partial F}{\partial \sigma_{F_j^i}(q)} \right). \quad (\text{C.67})$$

Note que as derivadas parciais retratadas na equação (C.67) podem ser expressas por

$$\frac{\partial E}{\partial \sigma_{F_j^i}(q)} = [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q), \quad \forall i, \quad (\text{C.68})$$

e

$$\frac{\partial F}{\partial \sigma_{F_j^i}(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_j^i}(q), & i \neq l \\ [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] b_{F_j^i}(q), & i = l \end{cases}. \quad (\text{C.69})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \theta_l(q)}$ para a situação em que $i \neq l$ temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \theta_l(q)} = e_{\theta_i^{(q)}}^{(q)} \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) - e^{(q)} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_j^i}(q). \quad (\text{C.70})$$

Organizando os termos da equação (C.70), temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \theta_l(q)} = [e_{\theta_i^{(q)}}^{(q)} - e^{(q)}] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_j^i}(q). \quad (\text{C.71})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \theta_l(q)}$ para a situação em que $i = l$, temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \theta_l(q)} = e_{\theta_l^{(q)}}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) + e^{(q)} [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] b_{F_j^i}(q), \quad (\text{C.72})$$

ou seja, (C.72) pode ser reescrita da seguinte maneira

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^l}(q) \partial \theta_l(q)} = \left[\left(e_{\theta_i(q)}^{(q)} - e^{(q)} \right) \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} \phi_l(\mathbf{x}^{(q)}) \right] b_{F_j^l}(q). \quad (\text{C.73})$$

C.1.6 Dedução de $\mathbf{H}_{\theta_i(q), m_{F_k^l}(q)}$

Os termos de $\mathbf{H}_{\theta_i(q), m_{F_k^l}(q)}$, expresso na equação (5.18), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{m_{F_k^l}(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $\theta_i(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial m_{F_k^l}(q)} = \frac{\partial}{\partial \theta_i(q)} [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q). \quad (\text{C.74})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial m_{F_k^l}(q)} = \frac{\partial}{\partial \theta_i(q)} [XYZ], \quad (\text{C.75})$$

em que

$$X = [f_s(\mathbf{x}^{(q)}) - y^{(q)}], \quad (\text{C.76})$$

$$Y = [\theta_l(q) - f_s(\mathbf{x}^{(q)})], \quad (\text{C.77})$$

$$W = \phi_l(\mathbf{x}^{(q)}) \quad (\text{C.78})$$

e

$$Z = a_{F_k^l}(q). \quad (\text{C.79})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (C.75). Com isso temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial m_{F_k^l}(q)} &= \left(\frac{\partial X}{\partial \theta_i(q)} \right) YWZ + X \left(\frac{\partial Y}{\partial \theta_i(q)} \right) WZ \\ &+ XY \left(\frac{\partial W}{\partial \theta_i(q)} \right) Z + XYW \left(\frac{\partial Z}{\partial \theta_i(q)} \right). \end{aligned} \quad (\text{C.80})$$

Note que as derivadas parciais retratadas na equação (C.80) podem ser expressas por

$$\frac{\partial X}{\partial \theta_i(q)} = \phi_i(\mathbf{x}^{(q)}), \quad \forall i, \quad (\text{C.81})$$

$$\frac{\partial Y}{\partial \theta_i(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)}), & i \neq l \\ 1 - \phi_l(\mathbf{x}^{(q)}), & i = l \end{cases} \quad (\text{C.82})$$

$$\frac{\partial W}{\partial \theta_i(q)} = 0, \quad \forall i \quad (\text{C.83})$$

e

$$\frac{\partial Z}{\partial \theta_i(q)} = 0, \quad \forall i. \quad (\text{C.84})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial m_{F_k^l}(q)}$ para a situação em que $i \neq l$ temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial m_{F_k^l}(q)} = e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) a_{F_k^l}(q) \phi_l(\mathbf{x}^{(q)}) - e^{(q)} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q). \quad (\text{C.85})$$

Organizando os termos da equação (C.85), temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial m_{F_k^l}(q)} = \left[e_{\theta_i(q)}^{(q)} - e^{(q)} \right] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q). \quad (\text{C.86})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_j^{(q)} \partial m_{F_k^l}(q)}$ para a situação em que $i = l$, temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_l(q) \partial m_{F_k^l}(q)} = e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \phi_l(\mathbf{x}^{(q)}) + e^{(q)} [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] a_{F_k^l}(q), \quad (\text{C.87})$$

ou seja, (C.87) pode ser reescrita da seguinte maneira

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_l(q) \partial m_{F_k^l}(q)} = \left[\left(e_{\theta_l(q)}^{(q)} - e^{(q)} \right) \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} \phi_l(\mathbf{x}^{(q)}) \right] a_{F_k^l}(q). \quad (\text{C.88})$$

C.1.7 Dedução de $\mathbf{H}_{\theta_i(q), \sigma_{F_k^l}(q)}$

Os termos de $\mathbf{H}_{\theta_i(q), \sigma_{F_k^l}(q)}$, expresso na equação (5.19), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{\sigma_{F_k^l}(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $\theta_i(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_{F_k^l}(q)} = \frac{\partial}{\partial \theta_i(q)} [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q). \quad (\text{C.89})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_{F_k^l}(q)} = \frac{\partial}{\partial \theta_i(q)} [RSTU], \quad (\text{C.90})$$

em que

$$R = [f_s(\mathbf{x}^{(q)}) - y^{(q)}], \quad (\text{C.91})$$

$$S = [\theta_l(q) - f_s(\mathbf{x}^{(q)})], \quad (\text{C.92})$$

$$T = \phi_l(\mathbf{x}^{(q)}) \quad (\text{C.93})$$

e

$$U = b_{F_k^l}(q). \quad (\text{C.94})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (C.90). Com isso temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_{F_k^l}(q)} &= \left(\frac{\partial R}{\partial \theta_i(q)} \right) STU + R \left(\frac{\partial S}{\partial \theta_i(q)} \right) TU \\ &+ RS \left(\frac{\partial T}{\partial \theta_i(q)} \right) U + RST \left(\frac{\partial U}{\partial \theta_i(q)} \right). \end{aligned} \quad (\text{C.95})$$

Note que as derivadas parciais retratadas na equação (C.95) podem ser expressas por

$$\frac{\partial R}{\partial \theta_i(q)} = \phi_i(\mathbf{x}^{(q)}), \quad \forall i, \quad (\text{C.96})$$

$$\frac{\partial S}{\partial \theta_i(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)}), & i \neq l \\ 1 - \phi_l(\mathbf{x}^{(q)}), & i = l \end{cases} \quad (\text{C.97})$$

$$\frac{\partial T}{\partial \theta_i(q)} = 0, \quad \forall i \quad (\text{C.98})$$

e

$$\frac{\partial U}{\partial \theta_i(q)} = 0, \quad \forall i. \quad (\text{C.99})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_{F_k^l}(q)}$ para a situação em que $i \neq l$ temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_{F_k^l}(q)} = e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) b_{F_k^l}(q) \phi_l(\mathbf{x}^{(q)}) - e^{(q)} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q). \quad (\text{C.100})$$

Organizando os termos da equação (C.100), temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_{F_k^l}(q)} = \left[e_{\theta_i(q)}^{(q)} - e^{(q)} \right] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q). \quad (\text{C.101})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_{F_k^l}(q)}$ para a situação em que $i = l$, temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_l(q) \partial \sigma_{F_k^l}(q)} = e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \phi_l(\mathbf{x}^{(q)}) + e^{(q)} [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] b_{F_k^l}(q), \quad (\text{C.102})$$

ou seja, (C.102) pode ser reescrita da seguinte maneira

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_l(q) \partial \sigma_{F_k^l}(q)} = \left[\left(e_{\theta_l(q)}^{(q)} - e^{(q)} \right) \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} \phi_l(\mathbf{x}^{(q)}) \right] b_{F_k^l}(q). \quad (\text{C.103})$$

C.1.8 Dedução de $\mathbf{H}_{\theta_i(q), \theta_l(q)}$

Os termos de $\mathbf{H}_{\theta_i(q), \theta_l(q)}$, expresso na equação (5.20), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{\theta_l(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $\theta_i(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \theta_l(q)} = \frac{\partial}{\partial \theta_i(q)} [f_s(\mathbf{x}^{(q)}) - y^{(q)}] \phi_l(\mathbf{x}^{(q)}). \quad (\text{C.104})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \theta_l(q)} = \frac{\partial}{\partial \theta_i(q)} [EF], \quad (\text{C.105})$$

em que

$$E = [f_s(\mathbf{x}^{(q)}) - y^{(q)}] \quad (\text{C.106})$$

e

$$F = \phi_l(\mathbf{x}^{(q)}). \quad (\text{C.107})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (C.105). Com isso temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \theta_l(q)} = \left(\frac{\partial E}{\partial \theta_i(q)} \right) F + E \left(\frac{\partial F}{\partial \theta_i(q)} \right). \quad (\text{C.108})$$

Note que as derivadas parciais retratadas na equação (C.108) podem ser expressas por

$$\frac{\partial E}{\partial \theta_i(q)} = \phi_i(\mathbf{x}^{(q)}), \quad \forall i, \quad (\text{C.109})$$

e

$$\frac{\partial F}{\partial \theta_i(q)} = 0, \quad \forall i. \quad (\text{C.110})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \theta_l(q)}$ para a situação em que $i \neq l$ e $i = l$, temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \theta_l(q)} = \begin{cases} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}), & i \neq l \\ \phi_l^2(\mathbf{x}^{(q)}), & i = l \end{cases} \quad (\text{C.111})$$

C.2 Dedução da matriz Hessiana para o sistema de inferência *fuzzy non-singleton*

C.2.1 Dedução de $\mathbf{H}_{m_{F_j^i}(q), \sigma_{F_k^l}(q)}$

Os termos de $\mathbf{H}_{m_{F_j^i}(q), \sigma_{F_k^l}(q)}$, expresso na equação (5.55), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{\sigma_{F_k^l}(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $m_{F_j^i}(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} = \frac{\partial}{\partial m_{F_j^i}(q)} [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \varphi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q). \quad (\text{C.112})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} = \frac{\partial}{\partial m_{F_j^i}(q)} [RSTU], \quad (\text{C.113})$$

de tal forma que

$$R = [f_s(\mathbf{x}^{(q)}) - y^{(q)}], \quad (\text{C.114})$$

$$S = [\theta_l(q) - f_s(\mathbf{x}^{(q)})], \quad (\text{C.115})$$

$$T = \phi_l(\mathbf{x}^{(q)}) \quad (\text{C.116})$$

e

$$U = b_{F_k^l}(q). \quad (\text{C.117})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (C.113). Com isso temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} &= \left(\frac{\partial R}{\partial m_{F_j^i}(q)} \right) STU + R \left(\frac{\partial S}{\partial m_{F_j^i}(q)} \right) TU \\ &+ RS \left(\frac{\partial T}{\partial m_{F_j^i}(q)} \right) U + RST \left(\frac{\partial U}{\partial m_{F_j^i}(q)} \right). \end{aligned} \quad (\text{C.118})$$

Note que as derivadas parciais retratadas na equação (C.118) podem ser expressas por

$$\frac{\partial R}{\partial m_{F_j^i}(q)} = [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q), \quad \forall i, \quad (\text{C.119})$$

$$\frac{\partial S}{\partial m_{F_j^i}(q)} = - [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) a_{F_k^l}(q), \quad \forall i, \quad (\text{C.120})$$

$$\frac{\partial T}{\partial m_{F_j^i}(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)})\phi_l(\mathbf{x}^{(q)})a_{F_j^i}(q), & i \neq l \\ [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] a_{F_k^l}(q), & i = l \end{cases} \quad (\text{C.121})$$

e

$$\frac{\partial U}{\partial m_{F_j^i}(q)} = \begin{cases} 0, & i \neq l \\ 0, & i = l, \quad j \neq k \\ \frac{-2(x_k^{(q)} - m_{F_k^l}(q))}{(\sigma_{F_k^l}^2(q) + \sigma_{X_k}^2(q))} \sigma_{F_k^l}(q), & i = l, \quad j = k \end{cases}. \quad (\text{C.122})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \sigma_{F_k^l}(q)}$, para a situação em que $i \neq l$, temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} &= e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q) e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \\ &\quad - e^{(q)} e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \\ &\quad - e^{(q)} e_{\theta_l(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q) b_{F_k^l}(q). \end{aligned} \quad (\text{C.123})$$

Organizando os termos da equação (C.123), temos que

$$\frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} = \left[e_{\theta_i(q)}^{(q)} e_{\theta_l(q)}^{(q)} - e^{(q)} e_{\theta_i(q)}^{(q)} - e^{(q)} e_{\theta_l(q)}^{(q)} \right] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q) b_{F_k^l}(q). \quad (\text{C.124})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \sigma_{F_k^l}(q)}$ para a situação em que $i = l$ temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^l}(q) \partial \sigma_{F_k^l}(q)} &= e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_j^l}(q) e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \\ &\quad - e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_j^l}(q) \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] a_{F_j^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial U}{\partial m_{F_j^l}(q)}, \end{aligned} \quad (\text{C.125})$$

ou seja, (C.125) pode ser reescrita da seguinte maneira

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^l}(q) \partial \sigma_{F_k^l}(q)} &= \left[\left(e_{\theta_l(q)}^{(q)} \right)^2 \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} e_{\theta_l(q)}^{(q)} (\phi_l(\mathbf{x}^{(q)}) - 2\phi_l^2(\mathbf{x}^{(q)})) \right] a_{F_j^l}(q) b_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial U}{\partial m_{F_j^l}(q)}. \end{aligned} \quad (\text{C.126})$$

Com isso, ordenando os termos da equação (C.126), temos a seguinte equação para

$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \sigma_{F_k^l}(q)}$ quando $i = l$

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^l}(q) \partial \sigma_{F_k^l}(q)} &= \left[\left(e_{\theta_l(q)}^{(q)} \right)^2 - 2e^{(q)} e_{\theta_l(q)}^{(q)} \right] \phi_l^2(\mathbf{x}^{(q)}) a_{F_j^l}(q) b_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \left[\frac{\partial U}{\partial m_{F_j^l}(q)} + a_{F_j^l}(q) b_{F_k^l}(q) \right], \end{aligned} \quad (\text{C.127})$$

em que $\frac{\partial U}{\partial m_{F_j^l}(q)}$ assume os seguintes valores

$$\frac{\partial U}{\partial m_{F_j^l}(q)} = \begin{cases} 0, & j \neq k \\ \frac{-2(x_k^{(q)} - m_{F_k^l}(q))}{(\sigma_{F_k^l}^2(q) + \sigma_{X_k}^2(q))^2} \sigma_{F_k^l}(q), & j = k \end{cases}. \quad (\text{C.128})$$

C.2.2 Dedução de $\mathbf{H}_{m_{F_j^i}(q), \theta_l(q)}$

Os termos de $\mathbf{H}_{m_{F_j^i}(q), \theta_l(q)}$, expresso na equação (5.56), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{\theta_l(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $m_{F_j^i}(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial \theta_l(q)} = \frac{\partial}{\partial m_{F_j^i}(q)} [f_s(\mathbf{x}^{(q)}) - y^{(q)}] \phi_l(\mathbf{x}^{(q)}). \quad (\text{C.129})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial \theta_l(q)} = \frac{\partial}{\partial m_{F_j^i}(q)} [EF]. \quad (\text{C.130})$$

Note que as derivadas parciais retratadas na equação (C.130) podem ser expressas por

$$E = [f_s(\mathbf{x}^{(q)}) - y^{(q)}] \quad (\text{C.131})$$

e

$$F = \phi_l(\mathbf{x}^{(q)}). \quad (\text{C.132})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (C.130). Com isso temos que

$$\frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial \theta_l(q)} = \left(\frac{\partial E}{\partial m_{F_j^i}(q)} \right) F + E \left(\frac{\partial F}{\partial m_{F_j^i}(q)} \right) \quad (\text{C.133})$$

em que

$$\frac{\partial E}{\partial m_{F_j^i}(q)} = [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q), \quad \forall i, \quad (\text{C.134})$$

e

$$\frac{\partial F}{\partial m_{F_j^i}(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q), & i \neq l \\ [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] a_{F_j^i}(q), & i = l \end{cases}. \quad (\text{C.135})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \theta_l(q)}$, para a situação em que $i \neq l$, temos que

$$\frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial \theta_l(q)} = e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) - e^{(q)} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q), \quad (\text{C.136})$$

Organizando os termos da equação (C.136), temos que

$$\frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial \theta_l(q)} = \left[e_{\theta_i(q)}^{(q)} - e^{(q)} \right] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q). \quad (\text{C.137})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \theta_l(q)}$ para a situação em que $i = l$ temos que

$$\frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial \theta_l(q)} = e_{\theta_i(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) + e^{(q)} [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] a_{F_j^i}(q), \quad (\text{C.138})$$

o que remete à equação

$$\frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial \theta_l(q)} = \left[\left(e_{\theta_i(q)}^{(q)} - e^{(q)} \right) \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} \phi_l(\mathbf{x}^{(q)}) \right] a_{F_j^i}(q). \quad (\text{C.139})$$

C.2.3 Dedução de $\mathbf{H}_{m_{F_j^i}(q), \sigma_X(q)}$

Os termos de $\mathbf{H}_{m_{F_j^i}(q), \sigma_X(q)}$, expresso na equação (5.57), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{\sigma_X(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $m_{F_j^i}(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial \sigma_X(q)} = \frac{\partial}{\partial m_{F_j^i}(q)} [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_l(\mathbf{x}^{(q)}) c_{F_k^l}(q). \quad (\text{C.140})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial \sigma_X(q)} = \frac{\partial}{\partial m_{F_j^i}(q)} [RSTU], \quad (\text{C.141})$$

de tal forma que

$$R = [f_s(\mathbf{x}^{(q)}) - y^{(q)}], \quad (\text{C.142})$$

$$S = [\theta_l(q) - f_s(\mathbf{x}^{(q)})], \quad (\text{C.143})$$

$$T = \phi_l(\mathbf{x}^{(q)}) \quad (\text{C.144})$$

e

$$U = c_{F_k^l}(q). \quad (\text{C.145})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (C.141). Com isso temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial \sigma_X(q)} &= \left(\frac{\partial R}{\partial m_{F_j^i}(q)} \right) STU + R \left(\frac{\partial S}{\partial m_{F_j^i}(q)} \right) TU \\ &+ RS \left(\frac{\partial T}{\partial m_{F_j^i}(q)} \right) U + RST \left(\frac{\partial U}{\partial m_{F_j^i}(q)} \right). \end{aligned} \quad (\text{C.146})$$

Note que as derivadas parciais retratadas na equação (C.146) podem ser expressas por

$$\frac{\partial R}{\partial m_{F_j^i}(q)} = [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q), \quad \forall i, \quad (\text{C.147})$$

$$\frac{\partial S}{\partial m_{F_j^i}(q)} = - [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) a_{F_k^l}(q), \quad \forall i, \quad (\text{C.148})$$

$$\frac{\partial T}{\partial m_{F_j^i}(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q), & i \neq l \\ [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] a_{F_k^l}(q), & i = l \end{cases} \quad (\text{C.149})$$

e

$$\frac{\partial U}{\partial m_{F_j^i}(q)} = \begin{cases} 0, & i \neq l \\ 0, & i = l, \quad j \neq k \\ \frac{-2(x_k^{(q)} - m_{F_k^l}(q)) \sigma_X(q)}{(\sigma_{F_k^l}^2(q) + \sigma_X^2(q))^2}, & i = l, \quad j = k \end{cases}. \quad (\text{C.150})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \sigma_X(q)}$, para a situação em que $i \neq l$, temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial \sigma_X(q)} &= e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q) e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) c_{F_k^l}(q) \\ &- e_{\theta_i(q)}^{(q)} e_{\theta_l(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) a_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) c_{F_k^l}(q) \\ &- e_{\theta_l(q)}^{(q)} e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q) c_{F_k^l}(q), \end{aligned} \quad (\text{C.151})$$

o que remete à equação

$$\frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^i}(q) \partial \sigma_X(q)} = \left[e_{\theta_i(q)}^{(q)} e_{\theta_l(q)}^{(q)} - e_{\theta_i(q)}^{(q)} e_{\theta_i(q)}^{(q)} - e_{\theta_l(q)}^{(q)} e_{\theta_l(q)}^{(q)} \right] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_j^i}(q) c_{F_k^l}(q). \quad (\text{C.152})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \sigma_X(q)}$ para a situação em que $i = l$ temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^l}(q) \partial \sigma_X(q)} = & e_{\theta_l^{(q)}}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_j^l}(q) e_{\theta_l^{(q)}}^{(q)} \phi_l(\mathbf{x}^{(q)}) c_{F_k^l}(q) \\ & - e^{(q)} e_{\theta_l^{(q)}}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_j^l}(q) \phi_l(\mathbf{x}^{(q)}) c_{F_k^l}(q) \\ & + e^{(q)} e_{\theta_l^{(q)}}^{(q)} [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] a_{F_j^l}(q) \\ & + e^{(q)} e_{\theta_l^{(q)}}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial U}{\partial m_{F_j^l}(q)}, \end{aligned} \quad (\text{C.153})$$

o que nos remete à seguinte equação

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^l}(q) \partial \sigma_X(q)} = & \left[\left(e_{\theta_l^{(q)}}^{(q)} \right)^2 \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} e_{\theta_l^{(q)}}^{(q)} (\phi_l(\mathbf{x}^{(q)}) - 2\phi_l^2(\mathbf{x}^{(q)})) \right] a_{F_j^l}(q) c_{F_k^l}(q) \\ & + e^{(q)} e_{\theta_l^{(q)}}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial U}{\partial m_{F_j^l}(q)}. \end{aligned} \quad (\text{C.154})$$

Com isso, ordenando os termos da equação (C.154), temos a seguinte equação para

$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial m_{F_j^i}(q) \partial \sigma_X(q)}$ quando $i = l$

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^l}(q) \partial \sigma_X(q)} = & \left[\left(e_{\theta_l^{(q)}}^{(q)} \right)^2 - 2e^{(q)} e_{\theta_l^{(q)}}^{(q)} \right] \phi_l^2(\mathbf{x}^{(q)}) a_{F_j^l}(q) c_{F_k^l}(q) \\ & + e^{(q)} e_{\theta_l^{(q)}}^{(q)} \phi_l(\mathbf{x}^{(q)}) \left[\frac{\partial U}{\partial m_{F_j^l}(q)} + a_{F_j^l}(q) c_{F_k^l}(q) \right], \end{aligned} \quad (\text{C.155})$$

em que $\frac{\partial^2 J(\mathbf{w})}{\partial m_{F_j^l}(q) \partial \sigma_X(q)}$ assume os seguintes valores

$$\frac{\partial U}{\partial m_{F_j^l}(q)} = \begin{cases} 0, & j \neq k \\ \frac{-2(x_k^{(q)} - m_{F_k^l}(q))}{(\sigma_{F_k^l}^2(q) + \sigma_X^2(q))^2} \sigma_X(q), & j = k \end{cases}. \quad (\text{C.156})$$

C.2.4 Dedução de $\mathbf{H}_{\sigma_{F_j^i}(q), m_{F_k^l}(q)}$

Os termos de $\mathbf{H}_{\sigma_{F_j^i}(q), m_{F_k^l}(q)}$, expresso na equação (5.58), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{m_{F_k^l}(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $\sigma_{F_j^i}(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial m_{F_k^l}(q)} = \frac{\partial}{\partial \sigma_{F_j^i}(q)} [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q). \quad (\text{C.157})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial m_{F_k^l}(q)} = \frac{\partial}{\partial \sigma_{F_j^i}(q)} [XYWZ], \quad (\text{C.158})$$

de tal forma que

$$X = [f_s(\mathbf{x}^{(q)}) - y^{(q)}], \quad (\text{C.159})$$

$$Y = [\theta_l(q) - f_s(\mathbf{x}^{(q)})], \quad (\text{C.160})$$

$$W = \phi_l(\mathbf{x}^{(q)}) \quad (\text{C.161})$$

e

$$Z = a_{F_k^l}(q). \quad (\text{C.162})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (C.158). Com isso temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial m_{F_k^l}(q)} &= \left(\frac{\partial X}{\partial \sigma_{F_j^i}(q)} \right) YWZ + X \left(\frac{\partial Y}{\partial \sigma_{F_j^i}(q)} \right) WZ \\ &+ XY \left(\frac{\partial W}{\partial \sigma_{F_j^i}(q)} \right) Z + XYW \left(\frac{\partial Z}{\partial \sigma_{F_j^i}(q)} \right). \end{aligned} \quad (\text{C.163})$$

Note que as derivadas parciais retratadas na equação (C.163) podem ser expressas por

$$\frac{\partial X}{\partial \sigma_{F_j^i}(q)} = [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q), \quad \forall i, \quad (\text{C.164})$$

$$\frac{\partial Y}{\partial \sigma_{F_j^i}(q)} = -[\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) b_{F_k^l}(q), \quad \forall i, \quad (\text{C.165})$$

$$\frac{\partial W}{\partial \sigma_{F_j^i}(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_j^i}(q), & i \neq l \\ [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] b_{F_k^l}(q), & i = l \end{cases} \quad (\text{C.166})$$

e

$$\frac{\partial Z}{\partial \sigma_{F_j^i}(q)} = \begin{cases} 0, & i \neq l \\ 0, & i = l, \quad j \neq k \\ \frac{-2(x_k^{(q)} - m_{F_k^l}(q)) \sigma_{F_k^l}(q)}{(\sigma_{F_k^l}^2(q) + \sigma_X^2(q))}, & i = l, \quad j = k \end{cases}. \quad (\text{C.167})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial m_{F_k^l}(q)}$ para a situação em que $i \neq l$ temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial m_{F_k^l}(q)} &= e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q) e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &- e^{(q)} e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &- e^{(q)} e_{\theta_l(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_j^i}(q) a_{F_k^l}(q). \end{aligned} \quad (\text{C.168})$$

Organizando os termos da equação (C.168), temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial m_{F_k^l}(q)} = \left[e_{\theta_i(q)}^{(q)} e_{\theta_l(q)}^{(q)} - e^{(q)} e_{\theta_i(q)}^{(q)} - e^{(q)} e_{\theta_l(q)}^{(q)} \right] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_j^i}(q) a_{F_k^l}(q). \quad (\text{C.169})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial m_{F_k^l}(q)}$ para a situação em que $i = l$, temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^l}(q) \partial m_{F_k^l}(q)} &= e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_j^l}(q) e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &\quad - e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_j^l}(q) \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \left[\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)}) \right] b_{F_j^l}(q) a_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial Z}{\partial \sigma_{F_j^l}(q)}, \end{aligned} \quad (\text{C.170})$$

ou seja, (C.170) pode ser reescrita da seguinte maneira

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^l}(q) \partial \sigma_{F_k^l}(q)} &= \left[\left(e_{\theta_l(q)}^{(q)} \right)^2 \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} e_{\theta_l(q)}^{(q)} \left(\phi_l(\mathbf{x}^{(q)}) - 2\phi_l^2(\mathbf{x}^{(q)}) \right) \right] a_{F_j^l}(q) b_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial Z}{\partial \sigma_{F_j^l}(q)}. \end{aligned} \quad (\text{C.171})$$

Com isso, ordenando os termos da equação (C.171), temos a seguinte equação para

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^l}(q) \partial m_{F_k^l}(q)} &= \left[\left(e_{\theta_l(q)}^{(q)} \right)^2 - 2e^{(q)} e_{\theta_l(q)}^{(q)} \right] \phi_l^2(\mathbf{x}^{(q)}) a_{F_j^l}(q) b_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \left[\frac{\partial Z}{\partial \sigma_{F_j^l}(q)} + a_{F_j^l}(q) b_{F_k^l}(q) \right], \end{aligned} \quad (\text{C.172})$$

em que $\frac{\partial Z}{\partial \sigma_{F_j^l}(q)}$ assume os seguintes valores

$$\frac{\partial Z}{\partial \sigma_{F_j^l}(q)} = \begin{cases} 0, & j \neq k \\ \frac{-2 \left(x_k^{(q)} - m_{F_k^l}(q) \right) \sigma_{F_k^l}(q)}{\left(\sigma_{F_k^l}^2(q) + \sigma_X^2(q) \right)^2}, & j = k \end{cases}. \quad (\text{C.173})$$

C.2.5 Dedução de $\mathbf{H}_{\sigma_{F_j^i}(q), \sigma_{F_k^l}(q)}$

Os termos de $\mathbf{H}_{\sigma_{F_j^i}(q), \sigma_{F_k^l}(q)}$, expresso na equação (5.59), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{\sigma_{F_k^l}(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $\sigma_{F_j^i}(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w})}{\partial \sigma_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} = \frac{\partial}{\partial \sigma_{F_j^i}(q)} \left[f_s(\mathbf{x}^{(q)}) - y^{(q)} \right] \left[\theta_l(q) - f_s(\mathbf{x}^{(q)}) \right] \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q). \quad (\text{C.174})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w})}{\partial \sigma_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} = \frac{\partial}{\partial \sigma_{F_j^i}(q)} [RSTU], \quad (\text{C.175})$$

de tal forma que

$$R = [f_s(\mathbf{x}^{(q)}) - y^{(q)}], \quad (\text{C.176})$$

$$S = [\theta_l(q) - f_s(\mathbf{x}^{(q)})], \quad (\text{C.177})$$

$$T = \phi_l(\mathbf{x}^{(q)}) \quad (\text{C.178})$$

e

$$U = b_{F_k^l}(q). \quad (\text{C.179})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (C.175). Com isso temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial \sigma_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} &= \left(\frac{\partial R}{\partial \sigma_{F_j^i}(q)} \right) STU + R \left(\frac{\partial S}{\partial \sigma_{F_j^i}(q)} \right) TU \\ &+ RS \left(\frac{\partial T}{\partial \sigma_{F_j^i}(q)} \right) U + RST \left(\frac{\partial U}{\partial \sigma_{F_j^i}(q)} \right). \end{aligned} \quad (\text{C.180})$$

Note que as derivadas parciais retratadas na equação (C.180) podem ser expressas por

$$\frac{\partial R}{\partial \sigma_{F_j^i}(q)} = [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q), \quad \forall i, \quad (\text{C.181})$$

$$\frac{\partial S}{\partial \sigma_{F_j^i}(q)} = - [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) b_{F_k^l}(q), \quad \forall i, \quad (\text{C.182})$$

$$\frac{\partial T}{\partial \sigma_{F_j^i}(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_j^i}(q), & i \neq l \\ [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] b_{F_k^l}(q), & i = l \end{cases} \quad (\text{C.183})$$

e

$$\frac{\partial U}{\partial \sigma_{F_j^i}(q)} = \begin{cases} 0, & i \neq l \\ 0, & i = l, \quad j \neq k \\ \frac{(x_k^{(q)} - m_{F_k^l}(q))^2}{(\sigma_{F_k^l}^2(q) + \sigma_{X_k}^2(q))^2} - \frac{2(x_k^{(q)} - m_{F_k^l}(q))^2 \sigma_{F_k^l}^2}{(\sigma_{F_k^l}^2(q) + \sigma_{X_k}^2(q))^3}, & i = l, \quad j = k \end{cases}. \quad (\text{C.184})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \sigma_{F_k^l}(q)}$, para a situação em que $i \neq l$, temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial \sigma_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} = & e_{\theta_i}^{(q)} \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q) e_{\theta_l}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \\ & - e^{(q)} e_{\theta_i}^{(q)} \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \\ & - e^{(q)} e_{\theta_l}^{(q)} \phi_l(\mathbf{x}^{(q)}) \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q) b_{F_k^l}(q). \end{aligned} \quad (\text{C.185})$$

Organizando os termos da equação (C.185), temos que

$$\frac{\partial^2 J(\mathbf{w})}{\partial \sigma_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} = \left[e_{\theta_i}^{(q)} e_{\theta_l}^{(q)} - e^{(q)} e_{\theta_i}^{(q)} - e^{(q)} e_{\theta_l}^{(q)} \right] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_j^i}(q) b_{F_k^l}(q). \quad (\text{C.186})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \sigma_{F_k^l}(q)}$ para a situação em que $i = l$, temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial \sigma_{F_j^i}(q) \partial \sigma_{F_k^l}(q)} = & e_{\theta_l}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_j^l}(q) e_{\theta_l}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \\ & - e^{(q)} e_{\theta_l}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_j^l}(q) \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \\ & + e^{(q)} e_{\theta_l}^{(q)} \left[\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)}) \right] b_{F_j^l}(q) b_{F_k^l}(q) \\ & + e^{(q)} e_{\theta_l}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial U}{\partial \sigma_{F_j^l}(q)}, \end{aligned} \quad (\text{C.187})$$

ou seja, (C.187) pode ser reescrita da seguinte maneira

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial \sigma_{F_j^l}(q) \partial \sigma_{F_k^l}(q)} = & \left[\left(e_{\theta_l}^{(q)} \right)^2 \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} e_{\theta_l}^{(q)} \left(\phi_l(\mathbf{x}^{(q)}) - 2\phi_l^2(\mathbf{x}^{(q)}) \right) \right] b_{F_j^l}(q) b_{F_k^l}(q) \\ & + e^{(q)} e_{\theta_l}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial U}{\partial \sigma_{F_j^l}(q)}. \end{aligned} \quad (\text{C.188})$$

Com isso, ordenando os termos da equação (C.188), temos a seguinte equação para

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial \sigma_{F_j^l}(q) \partial \sigma_{F_k^l}(q)} = & \left[\left(e_{\theta_l}^{(q)} \right)^2 - 2e^{(q)} e_{\theta_l}^{(q)} \right] \phi_l^2(\mathbf{x}^{(q)}) b_{F_j^l}(q) b_{F_k^l}(q) \\ & + e^{(q)} e_{\theta_l}^{(q)} \phi_l(\mathbf{x}^{(q)}) \left[\frac{\partial U}{\partial \sigma_{F_j^l}(q)} + b_{F_j^l}(q) b_{F_k^l}(q) \right], \end{aligned} \quad (\text{C.189})$$

em que $\frac{\partial U}{\partial \sigma_{F_j^l}(q)}$ assume os seguintes valores

$$\frac{\partial U}{\partial \sigma_{F_j^l}(q)} = \begin{cases} 0, & j \neq k \\ \frac{\left(x_k^{(q)} - m_{F_k^l}(q) \right)^2}{\left(\sigma_{F_k^l}^2(q) + \sigma_{X_k}^2(q) \right)^2} - \frac{2 \left(x_k^{(q)} - m_{F_k^l}(q) \right)^2 \sigma_{F_k^l}^2}{\left(\sigma_{F_k^l}^2(q) + \sigma_{X_k}^2(q) \right)^3}, & j = k \end{cases}. \quad (\text{C.190})$$

C.2.6 Dedução de $\mathbf{H}_{\sigma_{F_j^i}(q), \theta_l(q)}$

Os termos de $\mathbf{H}_{\sigma_{F_j^i}(q), \theta_l(q)}$, expresso na equação (5.60), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{\theta_l(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $\sigma_{F_j^i}(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \theta_l(q)} = \frac{\partial}{\partial \sigma_{F_j^i}(q)} [f_s(\mathbf{x}^{(q)}) - y^{(q)}] \phi_l(\mathbf{x}^{(q)}). \quad (\text{C.191})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \theta_l(q)} = \frac{\partial}{\partial \sigma_{F_j^i}(q)} [EF]. \quad (\text{C.192})$$

de tal forma que

$$E = [f_s(\mathbf{x}^{(q)}) - y^{(q)}] \quad (\text{C.193})$$

e

$$F = \phi_l(\mathbf{x}^{(q)}). \quad (\text{C.194})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (C.192). Com isso temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \theta_l(q)} = \left(\frac{\partial E}{\partial \sigma_{F_j^i}(q)} \right) F + E \left(\frac{\partial F}{\partial \sigma_{F_j^i}(q)} \right). \quad (\text{C.195})$$

Note que as derivadas parciais retratadas na equação (C.195) podem ser expressas por

$$\frac{\partial E}{\partial \sigma_{F_j^i}(q)} = [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q), \quad \forall i, \quad (\text{C.196})$$

e

$$\frac{\partial F}{\partial \sigma_{F_j^l}(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_j^i}(q), & i \neq l \\ [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] b_{F_j^l}(q), & i = l \end{cases}. \quad (\text{C.197})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \theta_l(q)}$ para a situação em que $i \neq l$ temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \theta_l(q)} = e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) - e^{(q)} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_j^i}(q). \quad (\text{C.198})$$

Organizando os termos da equação (C.198), temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \theta_l(q)} = \left[e_{\theta_i}^{(q)} - e^{(q)} \right] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_j^i}(q). \quad (\text{C.199})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \theta_l(q)}$ para a situação em que $i = l$, temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^l}(q) \partial \theta_l(q)} = e_{\theta_l}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_j^l}(q) \phi_l(\mathbf{x}^{(q)}) + e^{(q)} [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] b_{F_j^l}(q), \quad (\text{C.200})$$

ou seja, (C.200) pode ser reescrita da seguinte maneira

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^l}(q) \partial \theta_l(q)} = \left[\left(e_{\theta_l}^{(q)} - e^{(q)} \right) \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} \phi_l(\mathbf{x}^{(q)}) \right] b_{F_j^l}(q). \quad (\text{C.201})$$

C.2.7 Dedução de $\mathbf{H}_{\sigma_{F_j^i}(q), \sigma_X(q)}$

Os termos de $\mathbf{H}_{\sigma_{F_j^i}(q), \sigma_X(q)}$, expresso na equação (5.61), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{\sigma_X(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $\sigma_{F_j^i}(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w})}{\partial \sigma_{F_j^i}(q) \partial \sigma_X(q)} = \frac{\partial}{\partial \sigma_{F_j^i}(q)} [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_l(\mathbf{x}^{(q)}) c_{F_k^l}(q). \quad (\text{C.202})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \sigma_X(q)} = \frac{\partial}{\partial \sigma_{F_j^i}(q)} [RSTU], \quad (\text{C.203})$$

de tal forma que

$$R = [f_s(\mathbf{x}^{(q)}) - y^{(q)}], \quad (\text{C.204})$$

$$S = [\theta_l(q) - f_s(\mathbf{x}^{(q)})], \quad (\text{C.205})$$

$$T = \phi_l(\mathbf{x}^{(q)}) \quad (\text{C.206})$$

e

$$U = c_{F_k^l}(q). \quad (\text{C.207})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (C.203). Com isso temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial \sigma_{F_j^i}(q) \partial \sigma_X(q)} &= \left(\frac{\partial R}{\partial \sigma_{F_j^i}(q)} \right) STU + R \left(\frac{\partial S}{\partial \sigma_{F_j^i}(q)} \right) TU \\ &\quad + RS \left(\frac{\partial T}{\partial \sigma_{F_j^i}(q)} \right) U + RST \left(\frac{\partial U}{\partial \sigma_{F_j^i}(q)} \right). \end{aligned} \quad (\text{C.208})$$

Note que as derivadas parciais retratadas na equação (C.208) podem ser expressas por

$$\frac{\partial R}{\partial \sigma_{F_j^i}(q)} = [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q), \quad \forall i, \quad (\text{C.209})$$

$$\frac{\partial S}{\partial \sigma_{F_j^i}(q)} = - [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q), \quad \forall i, \quad (\text{C.210})$$

$$\frac{\partial T}{\partial \sigma_{F_j^i}(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_j^i}(q), & i \neq l \\ [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] b_{F_j^l}(q), & i = l \end{cases} \quad (\text{C.211})$$

e

$$\frac{\partial U}{\partial \sigma_{F_j^i}(q)} = \begin{cases} 0, & i \neq l \\ 0, & i = l, \quad j \neq k \\ \frac{2(x_k^{(q)} - m_{F_k^l}(q))^2 \sigma_{F_k^l}(q) \sigma_X(q)}{(\sigma_{F_k^l}^2(q) + \sigma_X^2(q))}, & i = l, \quad j = k \end{cases}. \quad (\text{C.212})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \sigma_X(q)}$, para a situação em que $i \neq l$, temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial \sigma_{F_j^i}(q) \partial \sigma_X(q)} &= e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q) e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) c_{F_k^l}(q) \\ &\quad - e^{(q)} e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) c_{F_k^l}(q) \\ &\quad - e^{(q)} e_{\theta_l(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_j^i}(q) c_{F_k^l}(q). \end{aligned} \quad (\text{C.213})$$

Organizando os termos da equação (C.213), temos que

$$\frac{\partial^2 J(\mathbf{w})}{\partial \sigma_{F_j^i}(q) \partial \sigma_X(q)} = \left[e_{\theta_i(q)}^{(q)} e_{\theta_l(q)}^{(q)} - e^{(q)} e_{\theta_i(q)}^{(q)} - e^{(q)} e_{\theta_l(q)}^{(q)} \right] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_j^i}(q) c_{F_k^l}(q). \quad (\text{C.214})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_{F_j^i}(q) \partial \sigma_X(q)}$ para a situação em que $i = l$, temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial \sigma_{F_j^i}(q) \partial \sigma_X(q)} &= e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_j^l}(q) e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) c_{F_k^l}(q) \\ &\quad - e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_j^l}(q) \phi_l(\mathbf{x}^{(q)}) c_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] b_{F_j^l}(q) c_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial U}{\partial \sigma_{F_j^l}(q)}, \end{aligned} \quad (\text{C.215})$$

ou seja, (C.215) pode ser reescrita da seguinte maneira

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial \sigma_{F_j^l}(q) \partial \sigma_X(q)} = & \left[\left(e_{\theta_l(q)}^{(q)} \right)^2 \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} e_{\theta_l(q)}^{(q)} \left(\phi_l(\mathbf{x}^{(q)}) - 2\phi_l^2(\mathbf{x}^{(q)}) \right) \right] b_{F_j^l}(q) c_{F_k^l}(q) \\ & + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial U}{\partial \sigma_{F_j^l}(q)}. \end{aligned} \quad (\text{C.216})$$

Com isso, ordenando os termos da equação (C.216), temos a seguinte equação para

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial \sigma_{F_j^l}(q) \partial \sigma_X(q)} = & \left[\left(e_{\theta_l(q)}^{(q)} \right)^2 - 2e^{(q)} e_{\theta_l(q)}^{(q)} \right] \phi_l^2(\mathbf{x}^{(q)}) b_{F_j^l}(q) c_{F_k^l}(q) \\ & + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \left[\frac{\partial U}{\partial \sigma_{F_j^l}(q)} + b_{F_j^l}(q) c_{F_k^l}(q) \right], \end{aligned} \quad (\text{C.217})$$

em que $\frac{\partial U}{\partial \sigma_{F_j^l}(q)}$ assume os seguintes valores

$$\frac{\partial U}{\partial \sigma_{F_j^l}(q)} = \begin{cases} 0, & j \neq k \\ \frac{2 \left(x_k^{(q)} - m_{F_k^l}(q) \right)^2 \sigma_{F_k^l}(q) + \sigma_X(q)}{\left(\sigma_{F_k^l}^2(q) + \sigma_X^2(q) \right)^3}, & j = k \end{cases}. \quad (\text{C.218})$$

C.2.8 Dedução de $\mathbf{H}_{\theta_i(q), m_{F_k^l}(q)}$

Os termos de $\mathbf{H}_{\theta_i(q), m_{F_k^l}(q)}$, expresso na equação (5.62), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{m_{F_k^l}(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $\theta_i(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial m_{F_k^l}(q)} = \frac{\partial}{\partial \theta_i(q)} \left[f_s(\mathbf{x}^{(q)}) - y^{(q)} \right] \left[\theta_l(q) - f_s(\mathbf{x}^{(q)}) \right] \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q). \quad (\text{C.219})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial m_{F_k^l}(q)} = \frac{\partial}{\partial \theta_i(q)} [XYWZ], \quad (\text{C.220})$$

em que

$$X = [f_s(\mathbf{x}^{(q)}) - y^{(q)}], \quad (\text{C.221})$$

$$Y = [\theta_l(q) - f_s(\mathbf{x}^{(q)})], \quad (\text{C.222})$$

$$W = \phi_l(\mathbf{x}^{(q)}) \quad (\text{C.223})$$

e

$$Z = a_{F_k^l}(q). \quad (\text{C.224})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (C.220). Com isso temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial m_{F_k^l}(q)} &= \left(\frac{\partial X}{\partial \theta_i(q)} \right) YWZ + X \left(\frac{\partial Y}{\partial \theta_i(q)} \right) WZ \\ &+ XY \left(\frac{\partial W}{\partial \theta_i(q)} \right) Z + XYW \left(\frac{\partial Z}{\partial \theta_i(q)} \right). \end{aligned} \quad (\text{C.225})$$

Note que as derivadas parciais retratadas na equação (C.225) podem ser expressas por

$$\frac{\partial X}{\partial \theta_i(q)} = \phi_i(\mathbf{x}^{(q)}), \quad \forall i, \quad (\text{C.226})$$

$$\frac{\partial Y}{\partial \theta_i(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)}), & i \neq l \\ 1 - \phi_l(\mathbf{x}^{(q)}), & i = l \end{cases} \quad (\text{C.227})$$

$$\frac{\partial W}{\partial \theta_i(q)} = 0, \quad \forall i \quad (\text{C.228})$$

e

$$\frac{\partial Z}{\partial \theta_i(q)} = 0, \quad \forall i. \quad (\text{C.229})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial m_{F_k^l}(q)}$ para a situação em que $i \neq l$ temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial m_{F_k^l}(q)} = e_{\theta_l(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) a_{F_k^l}(q) \phi_l(\mathbf{x}^{(q)}) - e^{(q)} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q). \quad (\text{C.230})$$

Organizando os termos da equação (C.230), temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial m_{F_k^l}(q)} = \left[e_{\theta_l(q)}^{(q)} - e^{(q)} \right] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q). \quad (\text{C.231})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_{F_j^i}(q) \partial m_{F_k^l}(q)}$ para a situação em que $i = l$, temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_l(q) \partial m_{F_k^l}(q)} = e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \phi_l(\mathbf{x}^{(q)}) + e^{(q)} [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] a_{F_k^l}(q), \quad (\text{C.232})$$

ou seja, (C.232) pode ser reescrita da seguinte maneira

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_l(q) \partial m_{F_k^l}(q)} = \left[\left(e_{\theta_l(q)}^{(q)} - e^{(q)} \right) \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} \phi_l(\mathbf{x}^{(q)}) \right] a_{F_k^l}(q). \quad (\text{C.233})$$

C.2.9 Dedução de $\mathbf{H}_{\theta_i(q), \sigma_{F_k^l}(q)}$

Os termos de $\mathbf{H}_{\theta_i(q), \sigma_{F_k^l}(q)}$, expresso na equação (5.63), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{\sigma_{F_k^l}(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $\theta_i(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_{F_k^l}(q)} = \frac{\partial}{\partial \theta_i(q)} [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q). \quad (\text{C.234})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_{F_k^l}(q)} = \frac{\partial}{\partial \theta_i(q)} [RSTU], \quad (\text{C.235})$$

em que

$$R = [f_s(\mathbf{x}^{(q)}) - y^{(q)}], \quad (\text{C.236})$$

$$S = [\theta_l(q) - f_s(\mathbf{x}^{(q)})], \quad (\text{C.237})$$

$$T = \phi_l(\mathbf{x}^{(q)}) \quad (\text{C.238})$$

e

$$U = b_{F_k^l}(q). \quad (\text{C.239})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (C.235). Com isso temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_{F_k^l}(q)} &= \left(\frac{\partial R}{\partial \theta_i(q)} \right) STU + R \left(\frac{\partial S}{\partial \theta_i(q)} \right) TU \\ &+ RS \left(\frac{\partial T}{\partial \theta_i(q)} \right) U + RST \left(\frac{\partial U}{\partial \theta_i(q)} \right). \end{aligned} \quad (\text{C.240})$$

Note que as derivadas parciais retratadas na equação (C.240) podem ser expressas por

$$\frac{\partial R}{\partial \theta_i(q)} = \phi_i(\mathbf{x}^{(q)}), \quad \forall i, \quad (\text{C.241})$$

$$\frac{\partial S}{\partial \theta_i(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)}), & i \neq l \\ 1 - \phi_l(\mathbf{x}^{(q)}), & i = l \end{cases} \quad (\text{C.242})$$

$$\frac{\partial T}{\partial \theta_i(q)} = 0, \quad \forall i \quad (\text{C.243})$$

e

$$\frac{\partial U}{\partial \theta_i(q)} = 0, \quad \forall i. \quad (\text{C.244})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_{F_k^l}(q)}$ para a situação em que $i \neq l$ temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_{F_k^l}(q)} = e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) b_{F_k^l}(q) \phi_l(\mathbf{x}^{(q)}) - e^{(q)} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q). \quad (\text{C.245})$$

Organizando os termos da equação (C.245), temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_{F_k^l}(q)} = \left[e_{\theta_i(q)}^{(q)} - e^{(q)} \right] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q). \quad (\text{C.246})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_{F_k^l}(q)}$ para a situação em que $i = l$, temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_l(q) \partial \sigma_{F_k^l}(q)} = e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \phi_l(\mathbf{x}^{(q)}) + e^{(q)} [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] b_{F_k^l}(q), \quad (\text{C.247})$$

ou seja, (C.247) pode ser reescrita da seguinte maneira

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_l(q) \partial \sigma_{F_k^l}(q)} = \left[\left(e_{\theta_l(q)}^{(q)} - e^{(q)} \right) \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} \phi_l(\mathbf{x}^{(q)}) \right] b_{F_k^l}(q). \quad (\text{C.248})$$

C.2.10 Dedução de $\mathbf{H}_{\theta_i(q), \theta_l(q)}$

Os termos de $\mathbf{H}_{\theta_i(q), \theta_l(q)}$, expresso na equação (5.64), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{\theta_l(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $\theta_i(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \theta_l(q)} = \frac{\partial}{\partial \theta_i(q)} [f_s(\mathbf{x}^{(q)}) - y^{(q)}] \phi_l(\mathbf{x}^{(q)}). \quad (\text{C.249})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \theta_l(q)} = \frac{\partial}{\partial \theta_i(q)} [EF], \quad (\text{C.250})$$

em que

$$E = [f_s(\mathbf{x}^{(q)}) - y^{(q)}] \quad (\text{C.251})$$

e

$$F = \phi_l(\mathbf{x}^{(q)}). \quad (\text{C.252})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (C.250). Com isso temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \theta_l(q)} = \left(\frac{\partial E}{\partial \theta_i(q)} \right) F + E \left(\frac{\partial F}{\partial \theta_i(q)} \right). \quad (\text{C.253})$$

Note que as derivadas parciais retratadas na equação (C.253) podem ser expressas por

$$\frac{\partial E}{\partial \theta_i(q)} = \phi_i(\mathbf{x}^{(q)}), \quad \forall i, \quad (\text{C.254})$$

e

$$\frac{\partial F}{\partial \theta_i(q)} = 0, \quad \forall i. \quad (\text{C.255})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \theta_l(q)}$ para a situação em que $i \neq l$ e $i = l$, temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \theta_l(q)} = \begin{cases} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}), & i \neq l \\ \phi_l^2(\mathbf{x}^{(q)}), & i = l \end{cases} \quad (\text{C.256})$$

C.2.11 Dedução de $\mathbf{H}_{\theta_i(q), \sigma_X(q)}$

Os termos de $\mathbf{H}_{\theta_i(q), \sigma_X(q)}$, expresso na equação (5.65), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{\sigma_X(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $\theta_i(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_X(q)} = \frac{\partial}{\partial \theta_i(q)} [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_l(\mathbf{x}^{(q)}) c_{F_k^l}(q). \quad (\text{C.257})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_X(q)} = \frac{\partial}{\partial \theta_i(q)} [RSTU], \quad (\text{C.258})$$

de tal forma que

$$R = [f_s(\mathbf{x}^{(q)}) - y^{(q)}], \quad (\text{C.259})$$

$$S = [\theta_l(q) - f_s(\mathbf{x}^{(q)})], \quad (\text{C.260})$$

$$T = \phi_l(\mathbf{x}^{(q)}) \quad (\text{C.261})$$

e

$$U = c_{F_k^l}(q). \quad (\text{C.262})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (C.258). Com isso temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_X(q)} &= \left(\frac{\partial R}{\partial \theta_i(q)} \right) STU + R \left(\frac{\partial S}{\partial \theta_i(q)} \right) TU \\ &+ RS \left(\frac{\partial T}{\partial \theta_i(q)} \right) U + RST \left(\frac{\partial U}{\partial \theta_i(q)} \right). \end{aligned} \quad (\text{C.263})$$

Note que as derivadas parciais retratadas na equação (C.240) podem ser expressas por

$$\frac{\partial R}{\partial \theta_i(q)} = \phi_i(\mathbf{x}^{(q)}), \quad \forall i, \quad (\text{C.264})$$

$$\frac{\partial S}{\partial \theta_i(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)}), & i \neq l \\ 1 - \phi_l(\mathbf{x}^{(q)}), & i = l \end{cases} \quad (\text{C.265})$$

$$\frac{\partial T}{\partial \theta_i(q)} = 0, \quad \forall i \quad (\text{C.266})$$

e

$$\frac{\partial U}{\partial \theta_i(q)} = 0, \quad \forall i. \quad (\text{C.267})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_X(q)}$ para a situação em que $i \neq l$ temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_X(q)} = e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) c_{F_k^l}(q) \phi_l(\mathbf{x}^{(q)}) - e^{(q)} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) c_{F_k^l}(q). \quad (\text{C.268})$$

Organizando os termos da equação (C.268), temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_X(q)} = \left[e_{\theta_i(q)}^{(q)} - e^{(q)} \right] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) c_{F_k^l}(q). \quad (\text{C.269})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_X(q)}$ para a situação em que $i = l$, temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_X(q)} = e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) c_{F_k^l}(q) \phi_l(\mathbf{x}^{(q)}) + e^{(q)} \left[\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)}) \right] c_{F_k^l}(q), \quad (\text{C.270})$$

ou seja, (C.270) pode ser reescrita da seguinte maneira

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \theta_i(q) \partial \sigma_X(q)} = \left[\left(e_{\theta_l(q)}^{(q)} - e^{(q)} \right) \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} \phi_l(\mathbf{x}^{(q)}) \right] c_{F_k^l}(q). \quad (\text{C.271})$$

C.2.12 Dedução de $\mathbf{H}_{\sigma_X(q), m_{F_k^l}(q)}$

Os termos de $\mathbf{H}_{\sigma_X(q), m_{F_k^l}(q)}$, expresso na equação (5.66), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{m_{F_k^l}(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $\sigma_X(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial m_{F_k^l}(q)} = \frac{\partial}{\partial \sigma_X(q)} \left[f_s(\mathbf{x}^{(q)}) - y^{(q)} \right] \left[\theta_l(q) - f_s(\mathbf{x}^{(q)}) \right] \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q). \quad (\text{C.272})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial m_{F_k^l}(q)} = \frac{\partial}{\partial \sigma_X(q)} [RSTU], \quad (\text{C.273})$$

de tal forma que

$$R = [f_s(\mathbf{x}^{(q)}) - y^{(q)}], \quad (\text{C.274})$$

$$S = [\theta_l(q) - f_s(\mathbf{x}^{(q)})], \quad (\text{C.275})$$

$$T = \phi_l(\mathbf{x}^{(q)}) \quad (\text{C.276})$$

e

$$U = a_{F_k^l}(q). \quad (\text{C.277})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (C.273). Com isso temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial m_{F_k^l}(q)} &= \left(\frac{\partial R}{\partial \sigma_X(q)} \right) STU + R \left(\frac{\partial S}{\partial \sigma_X(q)} \right) TU \\ &+ RS \left(\frac{\partial T}{\partial \sigma_X(q)} \right) U + RST \left(\frac{\partial U}{\partial \sigma_X(q)} \right). \end{aligned} \quad (\text{C.278})$$

Note que as derivadas parciais retratadas na equação (C.278) podem ser expressas por

$$\frac{\partial R}{\partial \sigma_X(q)} = [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) c_{F_j^i}(q), \quad \forall i, \quad (\text{C.279})$$

$$\frac{\partial S}{\partial \sigma_X(q)} = -[\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) c_{F_j^i}(q), \quad \forall i, \quad (\text{C.280})$$

$$\frac{\partial T}{\partial \sigma_X(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) c_{F_j^i}(q), & i \neq l \\ [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] c_{F_j^l}(q), & i = l \end{cases} \quad (\text{C.281})$$

e

$$\frac{\partial U}{\partial \sigma_X(q)} = \begin{cases} 0, & i \neq l \\ 0, & i = l, \quad j \neq k \\ \frac{-2(x_k^{(q)} - m_{F_k^l}(q)) \sigma_X(q)}{(\sigma_{F_k^l}^2(q) + \sigma_X^2(q))^2}, & i = l, \quad j = k \end{cases}. \quad (\text{C.282})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial m_{F_k^l}(q)}$ para a situação em que $i \neq l$ temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial m_{F_k^l}(q)} &= e_{\theta_i}^{(q)} \phi_i(\mathbf{x}^{(q)}) c_{F_j^i}(q) e_{\theta_l}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &- e_{\theta_i}^{(q)} e_{\theta_l}^{(q)} \phi_i(\mathbf{x}^{(q)}) c_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &- e_{\theta_l}^{(q)} e_{\theta_i}^{(q)} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) c_{F_j^i}(q) a_{F_k^l}(q). \end{aligned} \quad (\text{C.283})$$

Organizando os termos da equação (C.283), temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial m_{F_k^l}(q)} = \left[e_{\theta_i(q)}^{(q)} e_{\theta_i(q)}^{(q)} - e^{(q)} e_{\theta_i(q)}^{(q)} - e^{(q)} e_{\theta_i(q)}^{(q)} \right] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) c_{F_j^i}(q) a_{F_k^l}(q) \quad (\text{C.284})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial m_{F_k^l}(q)}$ para a situação em que $i = l$, temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial m_{F_k^l}(q)} &= e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) c_{F_j^l}(q) e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &\quad - e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) c_{F_j^l}(q) \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \left[\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)}) \right] c_{F_j^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial U}{\partial \sigma_X(q)}, \end{aligned} \quad (\text{C.285})$$

ou seja, (C.285) pode ser reescrita da seguinte maneira

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial m_{F_k^l}(q)} &= \left[\left(e_{\theta_l(q)}^{(q)} \right)^2 \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} e_{\theta_l(q)}^{(q)} \left(\phi_l(\mathbf{x}^{(q)}) - 2\phi_l^2(\mathbf{x}^{(q)}) \right) \right] c_{F_j^l}(q) a_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial U}{\partial \sigma_X(q)}. \end{aligned} \quad (\text{C.286})$$

Com isso, ordenando os termos da equação (C.286), temos a seguinte equação para

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial m_{F_k^l}(q)} &= \left[\left(e_{\theta_l(q)}^{(q)} \right)^2 - 2e^{(q)} e_{\theta_l(q)}^{(q)} \right] \phi_l^2(\mathbf{x}^{(q)}) c_{F_j^l}(q) a_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \left[\frac{\partial U}{\partial \sigma_X(q)} + c_{F_j^l}(q) a_{F_k^l}(q) \right], \end{aligned} \quad (\text{C.287})$$

em que $\frac{\partial U}{\partial \sigma_X(q)}$ assume os seguintes valores

$$\frac{\partial U}{\partial \sigma_X(q)} = \begin{cases} 0, & j \neq k \\ \frac{-2 \left(x_k^{(q)} - m_{F_k^l}(q) \right) \sigma_X(q)}{\left(\sigma_{F_k^l}^2(q) + \sigma_X^2(q) \right)^2}, & j = k \end{cases} \quad (\text{C.288})$$

C.2.13 Dedução de $\mathbf{H}_{\sigma_X(q), \sigma_{F_k^l}(q)}$

Os termos de $\mathbf{H}_{\sigma_X(q), \sigma_{F_k^l}(q)}$, expresso na equação (5.67), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{\sigma_{F_k^l}(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $\sigma_X(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \sigma_{F_k^l}(q)} = \frac{\partial}{\partial \sigma_X(q)} \left[f_s(\mathbf{x}^{(q)}) - y^{(q)} \right] \left[\theta_l(q) - f_s(\mathbf{x}^{(q)}) \right] \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q). \quad (\text{C.289})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \sigma_{F_k^l}(q)} = \frac{\partial}{\partial \sigma_X(q)} [RSTU], \quad (\text{C.290})$$

de tal forma que

$$R = [f_s(\mathbf{x}^{(q)}) - y^{(q)}], \quad (\text{C.291})$$

$$S = [\theta_l(q) - f_s(\mathbf{x}^{(q)})], \quad (\text{C.292})$$

$$T = \phi_l(\mathbf{x}^{(q)}) \quad (\text{C.293})$$

e

$$U = b_{F_k^l}(q). \quad (\text{C.294})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (C.290). Com isso temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \sigma_{F_k^l}(q)} &= \left(\frac{\partial R}{\partial \sigma_X(q)} \right) STU + R \left(\frac{\partial S}{\partial \sigma_X(q)} \right) TU \\ &\quad + RS \left(\frac{\partial T}{\partial \sigma_X(q)} \right) U + RST \left(\frac{\partial U}{\partial \sigma_X(q)} \right). \end{aligned} \quad (\text{C.295})$$

Note que as derivadas parciais retratadas na equação (C.295) podem ser expressas por

$$\frac{\partial R}{\partial \sigma_X(q)} = [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) c_{F_j^i}(q), \quad \forall i, \quad (\text{C.296})$$

$$\frac{\partial S}{\partial \sigma_X(q)} = -[\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) c_{F_j^i}(q), \quad \forall i, \quad (\text{C.297})$$

$$\frac{\partial T}{\partial \sigma_X(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) c_{F_j^i}(q), & i \neq l \\ [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] c_{F_j^l}(q), & i = l \end{cases} \quad (\text{C.298})$$

e

$$\frac{\partial U}{\partial \sigma_X(q)} = \begin{cases} 0, & i \neq l \\ 0, & i = l, \quad j \neq k \\ \frac{2 \left(x_k^{(q)} - m_{F_k^l}(q) \right)^2 \sigma_{F_k^l}(q) \sigma_X(q)}{\left(\sigma_{F_k^l}^2(q) + \sigma_X^2(q) \right)^3}, & i = l, \quad j = k \end{cases}. \quad (\text{C.299})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \sigma_{F_k^l}(q)}$, para a situação em que $i \neq l$, temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \sigma_{F_j^i}(q)} &= e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) c_{F_j^i}(q) e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \\ &\quad - e^{(q)} e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) c_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \\ &\quad - e^{(q)} e_{\theta_l(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) c_{F_j^i}(q) b_{F_k^l}(q). \end{aligned} \quad (\text{C.300})$$

Organizando os termos da equação (C.300), temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \sigma_{F_k^l}(q)} = \left[e_{\theta_i(q)}^{(q)} e_{\theta_l(q)}^{(q)} - e^{(q)} e_{\theta_i(q)}^{(q)} - e^{(q)} e_{\theta_l(q)}^{(q)} \right] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) c_{F_j^i}(q) b_{F_k^l}(q). \quad (\text{C.301})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \sigma_{F_k^l}(q)}$ para a situação em que $i = l$, temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \sigma_{F_k^l}(q)} &= e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) c_{F_j^l}(q) e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \\ &\quad - e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) c_{F_j^l}(q) \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \left[\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)}) \right] c_{F_j^l}(q) b_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial U}{\partial \sigma_X(q)}, \end{aligned} \quad (\text{C.302})$$

ou seja, (C.302) pode ser reescrita da seguinte maneira

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \sigma_{F_k^l}(q)} &= \left[\left(e_{\theta_l(q)}^{(q)} \right)^2 \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} e_{\theta_l(q)}^{(q)} \left(\phi_l(\mathbf{x}^{(q)}) - 2\phi_l^2(\mathbf{x}^{(q)}) \right) \right] c_{F_j^l}(q) b_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial U}{\partial \sigma_X(q)}. \end{aligned} \quad (\text{C.303})$$

Com isso, ordenando os termos da equação (C.303), temos a seguinte equação para

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \sigma_{F_k^l}(q)} &= \left[\left(e_{\theta_l(q)}^{(q)} \right)^2 - 2e^{(q)} e_{\theta_l(q)}^{(q)} \right] \phi_l^2(\mathbf{x}^{(q)}) c_{F_j^l}(q) b_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \left[\frac{\partial U}{\partial \sigma_X(q)} + c_{F_j^l}(q) b_{F_k^l}(q) \right], \end{aligned} \quad (\text{C.304})$$

em que $\frac{\partial U}{\partial \sigma_{F_j^l}(q)}$ assume os seguintes valores

$$\frac{\partial U}{\partial \sigma_{F_j^l}(q)} = \begin{cases} 0, & j \neq k \\ \frac{2 \left(x_k^{(q)} - m_{F_k^l}(q) \right)^2 \sigma_{F_k^l}(q) \sigma_X(q)}{\left(\sigma_{F_k^l}^2(q) + \sigma_X^2(q) \right)^3}, & j = k \end{cases}. \quad (\text{C.305})$$

C.2.14 Dedução de $\mathbf{H}_{\sigma_X(q), \theta_l(q)}$

Os termos de $\mathbf{H}_{\sigma_X(q), \theta_l(q)}$, expresso na equação (5.68), são definidos a partir da análise das derivadas parciais do gradiente $\nabla_{\theta_l(q)} J(\mathbf{w}^{(q)})$ em relação ao parâmetro $\sigma_X(q)$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \theta_l(q)} = \frac{\partial}{\partial \sigma_X(q)} \left[f_s(\mathbf{x}^{(q)}) - y^{(q)} \right] \phi_l(\mathbf{x}^{(q)}). \quad (\text{C.306})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \theta_l(q)} = \frac{\partial}{\partial \sigma_X(q)} [EF]. \quad (\text{C.307})$$

de tal forma que

$$E = [f_s(\mathbf{x}^{(q)}) - y^{(q)}] \quad (\text{C.308})$$

e

$$F = \phi_l(\mathbf{x}^{(q)}). \quad (\text{C.309})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (C.307). Com isso temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \theta_l(q)} = \left(\frac{\partial E}{\partial \sigma_X(q)} \right) F + E \left(\frac{\partial F}{\partial \sigma_X(q)} \right). \quad (\text{C.310})$$

Note que as derivadas parciais retratadas na equação (C.310) podem ser expressas por

$$\frac{\partial E}{\partial \sigma_X(q)} = [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) c_{F_j^i}(q), \quad \forall i, \quad (\text{C.311})$$

e

$$\frac{\partial F}{\partial \sigma_X(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) c_{F_j^i}(q), & i \neq l \\ [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] c_{F_j^l}(q), & i = l \end{cases}. \quad (\text{C.312})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \theta_l(q)}$ para a situação em que $i \neq l$ temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \theta_l(q)} = e_{\theta_l(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) c_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) - e^{(q)} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) c_{F_j^i}(q). \quad (\text{C.313})$$

Organizando os termos da equação (C.313), temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \theta_l(q)} = [e_{\theta_l(q)}^{(q)} - e^{(q)}] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) c_{F_j^i}(q). \quad (\text{C.314})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \theta_l(q)}$ para a situação em que $i = l$, temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \theta_l(q)} = e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) c_{F_j^l}(q) \phi_l(\mathbf{x}^{(q)}) + e^{(q)} [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] c_{F_j^l}(q), \quad (\text{C.315})$$

ou seja, (C.315) pode ser reescrita da seguinte maneira

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial \sigma_X(q) \partial \theta_l(q)} = \left[\left(e_{\theta_l(q)}^{(q)} - e^{(q)} \right) \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} \phi_l(\mathbf{x}^{(q)}) \right] c_{F_j^l}(q). \quad (\text{C.316})$$

C.2.15 Dedução de $\mathbf{H}_{\sigma_X(q), \sigma_X(q)}$

Os termos de $\mathbf{H}_{\sigma_X(q), \sigma_X(q)}$, expresso na equação (5.69), são definidos a partir da análise das derivadas parciais do parâmetro $\sigma_X(q)$ em relação ao gradiente $\nabla_{\sigma_X(q)} J(\mathbf{w}^{(q)})$ e são deduzidos conforme a seguir

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial^2 \sigma_X(q)} = \frac{\partial}{\partial \sigma_X(q)} [f_s(\mathbf{x}^{(q)}) - y^{(q)}] [\theta_l(q) - f_s(\mathbf{x}^{(q)})] \phi_l(\mathbf{x}^{(q)}) c_{F_k^l}(q). \quad (\text{C.317})$$

Para simplificar as deduções, fazemos as seguintes atribuições

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial^2 \sigma_X(q)} = \frac{\partial}{\partial \sigma_X(q)} [RSTU], \quad (\text{C.318})$$

de tal forma que

$$R = [f_s(\mathbf{x}^{(q)}) - y^{(q)}], \quad (\text{C.319})$$

$$S = [\theta_l(q) - f_s(\mathbf{x}^{(q)})], \quad (\text{C.320})$$

$$T = \phi_l(\mathbf{x}^{(q)}) \quad (\text{C.321})$$

e

$$U = c_{F_k^l}(q). \quad (\text{C.322})$$

Utilizando o conceito de derivada por partes (SWOKOWSKI, 1994), podemos simplificar a derivada parcial retratada na equação (C.318). Com isso temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial^2 \sigma_X(q)} &= \left(\frac{\partial R}{\partial \sigma_X(q)} \right) STU + R \left(\frac{\partial S}{\partial \sigma_X(q)} \right) TU \\ &\quad + RS \left(\frac{\partial T}{\partial \sigma_X(q)} \right) U + RST \left(\frac{\partial U}{\partial \sigma_X(q)} \right). \end{aligned} \quad (\text{C.323})$$

Note que as derivadas parciais retratadas na equação (C.323) podem ser expressas por

$$\frac{\partial R}{\partial \sigma_X(q)} = [\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q), \quad \forall i, \quad (\text{C.324})$$

$$\frac{\partial S}{\partial \sigma_X(q)} = -[\theta_i(q) - f_s(\mathbf{x}^{(q)})] \phi_i(\mathbf{x}^{(q)}) b_{F_j^i}(q), \quad \forall i, \quad (\text{C.325})$$

$$\frac{\partial T}{\partial \sigma_X(q)} = \begin{cases} -\phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) b_{F_j^i}(q), & i \neq l \\ [\phi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] b_{F_j^l}(q), & i = l \end{cases} \quad (\text{C.326})$$

e

$$\frac{\partial U}{\partial \sigma_X(q)} = \begin{cases} 0, & i \neq l \\ 0, & i = l, \quad j \neq k \\ \frac{(x_k^{(q)} - m_{F_k^l}(q))^2}{(\sigma_{F_k^l}^2(q) + \sigma_X^2(q))^2} - \frac{2(x_k^{(q)} - m_{F_k^l}(q))^2 \sigma_X^2(q)}{(\sigma_{F_k^l}^2(q) + \sigma_X^2(q))^3}, & i = l, \quad j = k \end{cases}. \quad (\text{C.327})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial^2 \sigma_X(q)}$, para a situação em que $i \neq l$, temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial^2 \sigma_X(q)} &= e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) c_{F_j^i}(q) e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) c_{F_k^l}(q) \\ &\quad - e^{(q)} e_{\theta_i(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) c_{F_j^i}(q) \phi_l(\mathbf{x}^{(q)}) c_{F_k^l}(q) \\ &\quad - e^{(q)} e_{\theta_l(q)}^{(q)} \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) c_{F_j^i}(q) c_{F_k^l}(q). \end{aligned} \quad (\text{C.328})$$

Organizando os termos da equação (C.185), temos que

$$\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial^2 \sigma_X(q)} = \left[e_{\theta_i(q)}^{(q)} e_{\theta_l(q)}^{(q)} - e^{(q)} e_{\theta_i(q)}^{(q)} - e^{(q)} e_{\theta_l(q)}^{(q)} \right] \phi_i(\mathbf{x}^{(q)}) \phi_l(\mathbf{x}^{(q)}) c_{F_j^i}(q) c_{F_k^l}(q). \quad (\text{C.329})$$

Analisando $\frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial^2 \sigma_X(q)}$ para a situação em que $i = l$, temos que

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial^2 \sigma_X(q)} &= e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) c_{F_j^l}(q) e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) c_{F_k^l}(q) \\ &\quad - e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) c_{F_j^l}(q) \phi_l(\mathbf{x}^{(q)}) c_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} [\varphi_l(\mathbf{x}^{(q)}) - \phi_l^2(\mathbf{x}^{(q)})] c_{F_j^l}(q) c_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial U}{\partial \sigma_X(q)}, \end{aligned} \quad (\text{C.330})$$

ou seja, (C.330) pode ser reescrita da seguinte maneira

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial^2 \sigma_X(q)} &= \left[\left(e_{\theta_l(q)}^{(q)} \right)^2 \phi_l^2(\mathbf{x}^{(q)}) + e^{(q)} e_{\theta_l(q)}^{(q)} (\phi_l(\mathbf{x}^{(q)}) - 2\phi_l^2(\mathbf{x}^{(q)})) \right] c_{F_j^l}(q) c_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \frac{\partial U}{\partial \sigma_X(q)}. \end{aligned} \quad (\text{C.331})$$

Com isso, ordenando os termos da equação (C.331), temos a seguinte equação para

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w}^{(q)})}{\partial^2 \sigma_X(q)} &= \left[\left(e_{\theta_l(q)}^{(q)} \right)^2 - 2e^{(q)} e_{\theta_l(q)}^{(q)} \right] \phi_l^2(\mathbf{x}^{(q)}) c_{F_j^l}(q) c_{F_k^l}(q) \\ &\quad + e^{(q)} e_{\theta_l(q)}^{(q)} \phi_l(\mathbf{x}^{(q)}) \left[\frac{\partial U}{\partial \sigma_X(q)} + c_{F_j^l}(q) c_{F_k^l}(q) \right], \end{aligned} \quad (\text{C.332})$$

em que $\frac{\partial U}{\partial \sigma_X(q)}$ assume os seguintes valores

$$\frac{\partial U}{\partial \sigma_X(q)} = \begin{cases} 0, & j \neq k \\ \frac{(x_k^{(q)} - m_{F_k^l}(q))^2}{(\sigma_{F_k^l}^2(q) + \sigma_X^2(q))^2} - \frac{2(x_k^{(q)} - m_{F_k^l}(q))^2 \sigma_X^2(q)}{(\sigma_{F_k^l}^2(q) + \sigma_X^2(q))^3}, & j = k \end{cases}. \quad (\text{C.333})$$

**APÊNDICE D – PARÂMETROS UTILIZADOS NO SISTEMA DE
INFERÊNCIA FUZZY**

D.1 Parâmetros do gradiente conjugado para o sistema de inferência *fuzzy singleton*

<i>sag</i>		
Técnica Escolhida	$\mu (\times 10^{-4})$	$\beta_{\max}(q)$
$\beta^{HS}(q)$ - Hestenes e Stiefel	1	5
$\beta^{FR}(q)$ - Fletcher e Reeves	1	5
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	1	2
$\beta^{CD}(q)$ - Fletcher	1	5
$\beta^{LS}(q)$ - Liu e Storey	1	5
$\beta^{DY}(q)$ - Dai e Yuan	1	5
$\beta^N(q)$ - Hager e Zhang	1	5
$\beta^H(q)$ - Daniel	0,1	5
Quasi-Newton	0,1	n.a.
Retropropagação	0,1	n.a.

Tabela 24: Parâmetros do método do gradiente conjugado para o sistema de inferência *fuzzy singleton* aplicado ao distúrbio *sag*.

<i>swell</i>		
Técnica Escolhida	$\mu (\times 10^{-4})$	$\beta_{\max}(q)$
$\beta^{HS}(q)$ - Hestenes e Stiefel	1	5
$\beta^{FR}(q)$ - Fletcher e Reeves	0,5	2
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	0,5	2
$\beta^{CD}(q)$ - Fletcher	1	5
$\beta^{LS}(q)$ - Liu e Storey	1	5
$\beta^{DY}(q)$ - Dai e Yuan	1	5
$\beta^N(q)$ - Hager e Zhang	1	5
$\beta^H(q)$ - Daniel	0,1	5
Quasi-Newton	0,1	n.a.
Retropropagação	0,1	n.a.

Tabela 25: Parâmetros do método do gradiente conjugado para o sistema de inferência fuzzy singleton aplicado ao distúrbio *swell*.

<i>flicker</i>		
Técnica Escolhida	$\mu (\times 10^{-4})$	$\beta_{\max}(q)$
$\beta^{HS}(q)$ - Hestenes e Stiefel	1	2
$\beta^{FR}(q)$ - Fletcher e Reeves	1	0,5
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	1	1
$\beta^{CD}(q)$ - Fletcher	1	1
$\beta^{LS}(q)$ - Liu e Storey	1	2
$\beta^{DY}(q)$ - Dai e Yuan	1	1
$\beta^N(q)$ - Hager e Zhang	1	2
$\beta^H(q)$ - Daniel	0,1	0,5
Quasi-Newton	0,1	n.a.
Retropropagação	0,1	n.a.

Tabela 26: Parâmetros do método do gradiente conjugado para o sistema de inferência fuzzy singleton aplicado ao distúrbio *flicker*.

harmônica		
Técnica Escolhida	$\mu (\times 10^{-4})$	$\beta_{\max}(q)$
$\beta^{HS}(q)$ - Hestenes e Stiefel	0,5	3
$\beta^{FR}(q)$ - Fletcher e Reeves	0,05	3
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	0,05	3
$\beta^{CD}(q)$ - Fletcher	0,05	3
$\beta^{LS}(q)$ - Liu e Storey	0,05	3
$\beta^{DY}(q)$ - Dai e Yuan	0,05	3
$\beta^N(q)$ - Hager e Zhang	0,5	2
$\beta^H(q)$ - Daniel	0,5	2
Quasi-Newton	0,05	n.a.
Retropropagação	0,1	n.a.

Tabela 27: Parâmetros do método do gradiente conjugado para o sistema de inferência *fuzzy singleton* aplicado ao distúrbio harmônica.

sag+flicker		
Técnica Escolhida	$\mu (\times 10^{-4})$	$\beta_{\max}(q)$
$\beta^{HS}(q)$ - Hestenes e Stiefel	1	5
$\beta^{FR}(q)$ - Fletcher e Reeves	1	5
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	1	2
$\beta^{CD}(q)$ - Fletcher	1	5
$\beta^{LS}(q)$ - Liu e Storey	1	5
$\beta^{DY}(q)$ - Dai e Yuan	1	5
$\beta^N(q)$ - Hager e Zhang	1	5
$\beta^H(q)$ - Daniel	0,1	5
Quasi-Newton	0,1	n.a.
Retropropagação	0,1	n.a.

Tabela 28: Parâmetros do método do gradiente conjugado para o sistema de inferência *fuzzy singleton* aplicado ao distúrbio *sag+flicker*.

<i>sag</i>+harmônica		
Técnica Escolhida	$\mu (\times 10^{-4})$	$\beta_{\max}(q)$
$\beta^{HS}(q)$ - Hestenes e Stiefel	1	5
$\beta^{FR}(q)$ - Fletcher e Reeves	1	5
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	1	2
$\beta^{CD}(q)$ - Fletcher	1	5
$\beta^{LS}(q)$ - Liu e Storey	1	5
$\beta^{DY}(q)$ - Dai e Yuan	1	5
$\beta^N(q)$ - Hager e Zhang	1	5
$\beta^H(q)$ - Daniel	0,1	5
Quasi-Newton	0,1	n.a.
Retropropagação	0,1	n.a.

Tabela 29: Parâmetros do método do gradiente conjugado para o sistema de inferência *fuzzy singleton* aplicado ao distúrbio *sag*+harmônica.

<i>swell</i>+<i>flicker</i>		
Técnica Escolhida	$\mu (\times 10^{-4})$	$\beta_{\max}(q)$
$\beta^{HS}(q)$ - Hestenes e Stiefel	1	5
$\beta^{FR}(q)$ - Fletcher e Reeves	1	2
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	1	2
$\beta^{CD}(q)$ - Fletcher	1	5
$\beta^{LS}(q)$ - Liu e Storey	1	5
$\beta^{DY}(q)$ - Dai e Yuan	1	5
$\beta^N(q)$ - Hager e Zhang	1	5
$\beta^H(q)$ - Daniel	0,5	2
Quasi-Newton	0,1	n.a.
Retropropagação	0,5	n.a.

Tabela 30: Parâmetros do método do gradiente conjugado para o sistema de inferência *fuzzy singleton* aplicado ao distúrbio *swell*+*flicker*.

<i>swell</i>+harmônica		
Técnica Escolhida	$\mu (\times 10^{-4})$	$\beta_{\max}(q)$
$\beta^{HS}(q)$ - Hestenes e Stiefel	1	5
$\beta^{FR}(q)$ - Fletcher e Reeves	1	5
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	1	2
$\beta^{CD}(q)$ - Fletcher	1	5
$\beta^{LS}(q)$ - Liu e Storey	1	5
$\beta^{DY}(q)$ - Dai e Yuan	1	5
$\beta^N(q)$ - Hager e Zhang	1	5
$\beta^H(q)$ - Daniel	0,1	5
Quasi-Newton	0,1	n.a.
Retropropagação	0,5	n.a.

Tabela 31: Parâmetros do método do gradiente conjugado para o sistema de inferência *fuzzy singleton* aplicado ao distúrbio *swell*+harmônica.

<i>sag</i>+harmônica+<i>flicker</i>		
Técnica Escolhida	$\mu (\times 10^{-4})$	$\beta_{\max}(q)$
$\beta^{HS}(q)$ - Hestenes e Stiefel	1	5
$\beta^{FR}(q)$ - Fletcher e Reeves	1	5
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	0,5	2
$\beta^{CD}(q)$ - Fletcher	1	5
$\beta^{LS}(q)$ - Liu e Storey	1	5
$\beta^{DY}(q)$ - Dai e Yuan	1	5
$\beta^N(q)$ - Hager e Zhang	1	5
$\beta^H(q)$ - Daniel	0,5	2
Quasi-Newton	0,1	n.a.
Retropropagação	0,5	n.a.

Tabela 32: Parâmetros do método do gradiente conjugado para o sistema de inferência *fuzzy singleton* aplicado ao distúrbio *sag*+harmônica+*flicker*.

<i>swell+harmônica+flicker</i>		
Técnica Escolhida	$\mu (\times 10^{-4})$	$\beta_{\max}(q)$
$\beta^{HS}(q)$ - Hestenes e Stiefel	1	5
$\beta^{FR}(q)$ - Fletcher e Reeves	1	5
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	1	5
$\beta^{CD}(q)$ - Fletcher	1	5
$\beta^{LS}(q)$ - Liu e Storey	1	5
$\beta^{DY}(q)$ - Dai e Yuan	1	5
$\beta^N(q)$ - Hager e Zhang	1	5
$\beta^H(q)$ - Daniel	0,1	5
Quasi-Newton	0,1	n.a.
Retropropagação	0,5	n.a.

Tabela 33: Parâmetros do método do gradiente conjugado para o sistema de inferência fuzzy singleton aplicado ao distúrbio *swell+harmônica+flicker*.

Parâmetro $\alpha_{\max}(q)$ - método Daniel	
Distúrbio	$\alpha_{\max}(q) (\times 10^{-3})$
<i>sag</i>	1
<i>swell</i>	1
<i>flicker</i>	1
harmônica	1
<i>sag+flicker</i>	1
<i>sag+harmônica</i>	1
<i>swell+flicker</i>	0,5
<i>swell+harmônica</i>	1
<i>sag+harmônica+flicker</i>	0,5
<i>swell+harmônica+flicker</i>	1

Tabela 34: Parâmetro $\alpha_{\max}(q)$ do sistema de inferência fuzzy singleton para o método proposto por Daniel (DANIEL, 1967).

D.2 Parâmetros do gradiente conjugado para o sistema de inferência *fuzzy non-singleton*

<i>sag</i>		
Técnica Escolhida	$\mu (\times 10^{-4})$	$\beta_{\max}(q)$
$\beta^{HS}(q)$ - Hestenes e Stiefel	1	5
$\beta^{FR}(q)$ - Fletcher e Reeves	1	5
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	1	2
$\beta^{CD}(q)$ - Fletcher	1	5
$\beta^{LS}(q)$ - Liu e Storey	1	5
$\beta^{DY}(q)$ - Dai e Yuan	1	5
$\beta^N(q)$ - Hager e Zhang	1	5
$\beta^H(q)$ - Daniel	0,1	5
Quasi-Newton	0,5	n.a.
Retropropagação	0,1	n.a.

Tabela 35: Parâmetros do método do gradiente conjugado para o sistema de inferência *fuzzy non-singleton* aplicado ao distúrbio *sag*.

<i>swell</i>		
Técnica Escolhida	$\mu (\times 10^{-4})$	$\beta_{\max}(q)$
$\beta^{HS}(q)$ - Hestenes e Stiefel	1	2
$\beta^{FR}(q)$ - Fletcher e Reeves	0,5	2
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	1	2
$\beta^{CD}(q)$ - Fletcher	1	5
$\beta^{LS}(q)$ - Liu e Storey	1	5
$\beta^{DY}(q)$ - Dai e Yuan	1	5
$\beta^N(q)$ - Hager e Zhang	1	5
$\beta^H(q)$ - Daniel	0,5	2
Quasi-Newton	0,5	n.a.
Retropropagação	0,1	n.a.

Tabela 36: Parâmetros do método do gradiente conjugado para o sistema de inferência *fuzzy non-singleton* aplicado ao distúrbio *swell*.

flicker		
Técnica Escolhida	$\mu (\times 10^{-4})$	$\beta_{\max}(q)$
$\beta^{HS}(q)$ - Hestenes e Stiefel	0,5	5
$\beta^{FR}(q)$ - Fletcher e Reeves	1	0,3
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	1	2
$\beta^{CD}(q)$ - Fletcher	1	0,5
$\beta^{LS}(q)$ - Liu e Storey	1	1
$\beta^{DY}(q)$ - Dai e Yuan	1	1
$\beta^N(q)$ - Hager e Zhang	1	2,5
$\beta^H(q)$ - Daniel	0,1	2,5
Quasi-Newton	0,1	n.a.
Retropropagação	0,1	n.a.

Tabela 37: Parâmetros do método do gradiente conjugado para o sistema de inferência fuzzy non-singleton aplicado ao distúrbio flicker.

harmônica		
Técnica Escolhida	$\mu (\times 10^{-4})$	$\beta_{\max}(q)$
$\beta^{HS}(q)$ - Hestenes e Stiefel	1	5
$\beta^{FR}(q)$ - Fletcher e Reeves	1	5
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	1	3
$\beta^{CD}(q)$ - Fletcher	1	5
$\beta^{LS}(q)$ - Liu e Storey	1	5
$\beta^{DY}(q)$ - Dai e Yuan	1	5
$\beta^N(q)$ - Hager e Zhang	1	5
$\beta^H(q)$ - Daniel	0,5	2
Quasi-Newton	0,1	n.a.
Retropropagação	0,1	n.a.

Tabela 38: Parâmetros do método do gradiente conjugado para o sistema de inferência fuzzy non-singleton aplicado ao distúrbio harmônica.

<i>sag+flicker</i>		
Técnica Escolhida	$\mu (\times 10^{-4})$	$\beta_{\max}(q)$
$\beta^{HS}(q)$ - Hestenes e Stiefel	1	5
$\beta^{FR}(q)$ - Fletcher e Reeves	1	2
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	1	2
$\beta^{CD}(q)$ - Fletcher	1	5
$\beta^{LS}(q)$ - Liu e Storey	1	5
$\beta^{DY}(q)$ - Dai e Yuan	1	5
$\beta^N(q)$ - Hager e Zhang	1	5
$\beta^H(q)$ - Daniel	0,5	1
Quasi-Newton	0,5	n.a.
Retropropagação	1	n.a.

Tabela 39: Parâmetros do método do gradiente conjugado para o sistema de inferência fuzzy non-singleton aplicado ao distúrbio *sag+flicker*.

<i>sag+harmônica</i>		
Técnica Escolhida	$\mu (\times 10^{-4})$	$\beta_{\max}(q)$
$\beta^{HS}(q)$ - Hestenes e Stiefel	1	2
$\beta^{FR}(q)$ - Fletcher e Reeves	1	2
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	1	2
$\beta^{CD}(q)$ - Fletcher	1	5
$\beta^{LS}(q)$ - Liu e Storey	1	5
$\beta^{DY}(q)$ - Dai e Yuan	1	2
$\beta^N(q)$ - Hager e Zhang	1	5
$\beta^H(q)$ - Daniel	0,1	5
Quasi-Newton	0,3	n.a.
Retropropagação	1	n.a.

Tabela 40: Parâmetros do método do gradiente conjugado para o sistema de inferência fuzzy non-singleton aplicado ao distúrbio *sag+harmônica*.

<i>swell+flicker</i>		
Técnica Escolhida	$\mu (\times 10^{-4})$	$\beta_{\max}(q)$
$\beta^{HS}(q)$ - Hestenes e Stiefel	1	5
$\beta^{FR}(q)$ - Fletcher e Reeves	1	2
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	1	2
$\beta^{CD}(q)$ - Fletcher	1	5
$\beta^{LS}(q)$ - Liu e Storey	1	5
$\beta^{DY}(q)$ - Dai e Yuan	1	5
$\beta^N(q)$ - Hager e Zhang	1	5
$\beta^H(q)$ - Daniel	0,1	5
Quasi-Newton	0,1	n.a.
Retropropagação	1	n.a.

Tabela 41: Parâmetros do método do gradiente conjugado para o sistema de inferência fuzzy non-singleton aplicado ao distúrbio *swell+flicker*.

<i>swell+harmônica</i>		
Técnica Escolhida	$\mu (\times 10^{-4})$	$\beta_{\max}(q)$
$\beta^{HS}(q)$ - Hestenes e Stiefel	1	5
$\beta^{FR}(q)$ - Fletcher e Reeves	1	2
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	1	2
$\beta^{CD}(q)$ - Fletcher	1	5
$\beta^{LS}(q)$ - Liu e Storey	1	5
$\beta^{DY}(q)$ - Dai e Yuan	1	5
$\beta^N(q)$ - Hager e Zhang	1	5
$\beta^H(q)$ - Daniel	0,1	5
Quasi-Newton	0,5	n.a.
Retropropagação	1	n.a.

Tabela 42: Parâmetros do método do gradiente conjugado para o sistema de inferência fuzzy non-singleton aplicado ao distúrbio *swell+harmônica*.

<i>sag+harmônica+flicker</i>		
Técnica Escolhida	$\mu (\times 10^{-4})$	$\beta_{\max}(q)$
$\beta^{HS}(q)$ - Hestenes e Stiefel	1	5
$\beta^{FR}(q)$ - Fletcher e Reeves	1	2
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	1	2
$\beta^{CD}(q)$ - Fletcher	1	5
$\beta^{LS}(q)$ - Liu e Storey	1	5
$\beta^{DY}(q)$ - Dai e Yuan	1	5
$\beta^N(q)$ - Hager e Zhang	1	2
$\beta^H(q)$ - Daniel	0,5	5
Quasi-Newton	0,1	n.a.
Retropropagação	1	n.a.

Tabela 43: Parâmetros do método do gradiente conjugado para o sistema de inferência fuzzy non-singleton aplicado ao distúrbio *sag+harmônica+flicker*.

<i>swell+harmônica+flicker</i>		
Técnica Escolhida	$\mu (\times 10^{-4})$	$\beta_{\max}(q)$
$\beta^{HS}(q)$ - Hestenes e Stiefel	1	5
$\beta^{FR}(q)$ - Fletcher e Reeves	1	0,5
$\beta^{PRP}(q)$ - Polak, Ribière e Polyak	1	5
$\beta^{CD}(q)$ - Fletcher	1	5
$\beta^{LS}(q)$ - Liu e Storey	1	5
$\beta^{DY}(q)$ - Dai e Yuan	1	5
$\beta^N(q)$ - Hager e Zhang	1	5
$\beta^H(q)$ - Daniel	0,1	5
Quasi-Newton	0,1	n.a.
Retropropagação	1	n.a.

Tabela 44: Parâmetros do método do gradiente conjugado para o sistema de inferência fuzzy non-singleton aplicado ao distúrbio *swell+harmônica+flicker*.

Parâmetro $\alpha_{\max}(q)$ - método Daniel	
Distúrbio	$\alpha_{\max}(q) (\times 10^{-3})$
<i>sag</i>	1
<i>swell</i>	1
<i>flicker</i>	1
harmônica	1
<i>sag+flicker</i>	0,5
<i>sag+harmônica</i>	1
<i>swell+flicker</i>	1
<i>swell+harmônica</i>	1
<i>sag+harmônica+flicker</i>	1
<i>swell+harmônica+flicker</i>	1

Tabela 45: Parâmetro $\alpha_{\max}(q)$ do sistema de inferência fuzzy non-singleton para o método proposto por Daniel (DANIEL, 1967).