

Universidade Estadual de Campinas
Faculdade de Engenharia Elétrica e de Computação
Depto. de Engenharia de Computação e Automação Industrial

Modelagem de Sistemas Não-lineares por Base de Funções Ortonormais Generalizadas com Funções Internas

Autor: M.Sc. Jeremias Barbosa Machado

Orientador: Prof. Dr. Wagner Caradori do Amaral

**Co-orientador: Prof. Dr. Ricardo José Grabrielli Barreto
Campello**

Tese de Doutorado apresentada à Faculdade de Engenharia Elétrica e de Computação como parte dos requisitos para obtenção do título de Doutor em Engenharia Elétrica. Área de concentração: **Automação**.

Banca Examinadora

Prof. Dr. Wagner Caradori do Amaral (Presidente) .. DCA/FEEC/UNICAMP
Prof. Dr. Gustavo Henrique da Costa Oliveira UFPR
Prof. Dr. Roberto Kawakami Harrop Galvão ITA
Prof. Dr. Paulo Augusto Valente Ferreira DT/FEEC/UNICAMP
Prof. Dr. Rafael Santos Mendes DCA/FEEC/UNICAMP

Campinas, SP
Fevereiro/2011

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA E ARQUITETURA - BAE - UNICAMP

M119m Machado, Jeremias Barbosa
Modelagem de sistemas não-lineares por base de
funções ortonormais generalizadas com funções internas.
/ Jeremias Barbosa Machado. – Campinas, SP: [s.n.], 2011.

Orientadores: Wagner Caradori do Amaral; Ricardo José
Gabrielli Barreto Campello.

Tese de Doutorado - Universidade Estadual de
Campinas, Faculdade de Engenharia Elétrica e de
Computação.

1. Sistemas dinâmicos. 2. Sistemas não-lineares.
3. Identificação de sistemas. 4. Volterra, Equações de.
5. Lógica Difusa. I. Amaral, Wagner Caradori do. II. Campello,
Ricardo José Gabrielli Barreto. III Universidade Estadual de
Campinas. Faculdade de Engenharia Elétrica e de
Computação. IV. Título.

Título em Inglês : Nonlinear systems modeling based on ladder-structured generalized
orthonormal basis functions

Palavras-chave em Inglês: Dynamical systems, Nonlinear systems, System
identification, Volterra equations, Fuzzy logic

Área de concentração: Automação

Titulação: Doutor em Engenharia Elétrica

Banca examinadora: Gustavo Henrique da Costa Oliveira, Roberto Kawakami Harrop
Galvão, Paulo Augusto Valente Ferreira, Rafael Santos Mendes

Data da defesa: 24/02/2011

Programa de Pós-Graduação: Engenharia Elétrica

COMISSÃO JULGADORA - TESE DE DOUTORADO

Candidato: Jeremias Barbosa Machado

Data da Defesa: 24 de fevereiro de 2011

Título da Tese: "Modelagem de Sistemas Não-lineares por Base de Funções Ortonormais Generalizadas com Funções Internas "

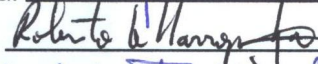
Prof. Dr. Wagner Caradori do Amaral (Presidente):



Prof. Dr. Gustavo Henrique da Costa Oliveira:



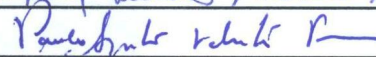
Prof. Dr. Roberto Kawakami Harrop Galvão:



Prof. Dr. Rafael Santos Mendes:



Prof. Dr. Paulo Augusto Valente Ferreira:



Resumo

Este trabalho enfoca a modelagem e identificação de sistemas dinâmicos não-lineares estáveis através de modelos *fuzzy* Takagi-Sugeno (TS) e/ou Volterra, ambos com estruturas formadas por bases de funções ortonormais (BFO), principalmente as bases de funções ortonormais generalizadas (GOBF - *Generalized Orthonormal Basis Functions*) com funções internas. As GOBF's com funções internas modelam sistemas dinâmicos com múltiplos modos através de uma parametrização que utiliza somente valores reais, sejam os polos do sistema reais e/ou complexos. Uma das principais contribuições desta tese concentra-se na proposta da otimização e ajuste fino dos parâmetros destes modelos não-lineares. Realiza-se a identificação dos modelos *fuzzy* TS-BFO utilizando-se de medidas dos sinais de entrada e saída do sistema a ser modelado. Os modelos *fuzzy* TS-BFO são inicialmente determinados utilizando-se uma técnica de agrupamento *fuzzy* (*fuzzy clustering*) e simplificados por algoritmos que eliminam eventuais redundâncias. Em sequência desenvolve-se o cálculo analítico dos gradientes da saída do modelo TS-BFO em relação aos parâmetros do modelo (polos da BFO, coeficientes da expansão da BFO e parâmetros das funções de pertinência). Utilizando-se técnicas de otimização não-linear e o valor dos gradientes, realiza-se a sintonia fina dos parâmetros dos modelos inicialmente obtidos. Para os modelos de Volterra-GOBF desenvolve-se uma nova abordagem utilizando-se GOBF com funções internas nos *kernels* dos modelos. São calculados os gradientes analíticos da saída do modelo de Volterra-GOBF, seja com *kernels* simétricos ou não simétricos, com relação aos parâmetros a serem determinados. Estes valores são utilizados em algoritmos de otimização que possibilitam a obtenção de modelos mais precisos do sistema sem nenhum conhecimento a priori de suas características. Além da identificação de sistemas não-lineares por modelos BFO, abordou-se também, nesta tese, uma nova metodologia para a otimização de modelos lineares BFO no domínio da frequência. Neste contexto, destaca-se como principal contribuição o desenvolvimento, no domínio da frequência, do cálculo analítico dos gradientes da resposta em frequência das funções de Kautz e Laguerre, com relação aos seus parâmetros de projeto. Os valores dos gradientes fornecem a direção de busca dos parâmetros dos modelos em processos de otimização não-linear. Também foram otimizados os modelos GOBF com funções internas, com o cálculo numérico dos seus gradientes, pois, ainda não foi possível estabelecer uma fórmula genérica para o cálculo analítico dos gradientes dos modelos GOBF, de qualquer ordem, em relação aos parâmetros a serem determinados. Exemplos ilustram a aplicação e eficiência dos métodos de identificação e otimização propostos na modelagem de sistemas lineares (domínio do tempo e da frequência) e não-lineares utilizando BFO's.

Palavras-chave: Modelagem e Identificação de Sistemas Não-lineares, Base de Funções Ortonormais, Modelos *Fuzzy* Takagi-Sugeno, Modelos de Volterra, Modelagem no domínio da Frequência.

Abstract

This work is concerned with the modeling and identification of stable nonlinear dynamic systems using Takagi-Sugeno fuzzy and Volterra models within the framework of orthonormal basis functions (OBF), mainly ladder-structured generalized orthonormal basis functions (GOBF). The ladder-structured GOBFs allows to model dynamic systems with multiple modes, real and/or complex poles, through a parameterization, which uses only real values. The main contribution of this thesis is the optimization and fine tuning of the parameters of OBF nonlinear models. The GOBF models identification are performed using only input and output measurements. The initial GOBF-TS fuzzy model is obtained using a fuzzy clustering technique and simplified by algorithms that eliminate any redundancies. Next, the analytical calculation of the gradients of GOBF-TS model concerning model parameters (GOBF poles, OBF expansion coefficients and the parameters of membership functions) is developed. A fine tuning of the model parameters is obtained by using a nonlinear optimization technique and the calculated gradients. For Volterra-GOBF models a new approach using kernels with ladder-structured GOBF is also proposed. Furthermore, Volterra-GOBF model optimization, with symmetrical or asymmetrical kernels, using an analytical gradients calculation of the output model regarding their parameters is presented. Following, a new approach for linear OBF models optimization, in frequency domain, is also addressed. In this context, the analytical calculation of the gradients of the Laguerre and Kautz frequency response concerning its parameters is presented. The ladder-structured GOBF models optimization, in the frequency domain, is performed using only numerical calculation of its gradients, as it has not yet been possible to derive a generic analytical gradients. Examples illustrate the performance and effectiveness of identification methods proposed here in the modeling and optimization of linear (time domain and frequency) and non-linear systems.

Keywords: Modeling and Identification of Nonlinear Systems, Orthonormal Basis Functions, Takagi-Sugeno Fuzzy Models, Volterra Models, Modeling on Frequency Domain.

*Dedico esta tese aos meus pais Mário e
Terezinha e à minha noiva Andrezza*

Agradecimentos

A Deus pela vida e pela iluminação na execução deste trabalho.

Aos meus orientadores Wagner Caradori do Amaral e Ricardo J. G. B. Campello pela orientação, amizade e dedicação nestes seis anos de mestrado e doutorado. Meus especiais agradecimentos ao professor Ricardo pelas valiosas ideias que tornaram possível este trabalho. Ao professor Wagner minha eterna gratidão pelo companheirismo e confiança demonstrados, além do auxílio prestado no momento que mais precisei. Que Deus lhes retribua em dobro tudo o que fizeram por mim!

Aos meus pais Mário e Terezinha, e meu irmão Juliano pelo apoio, amor e incentivo durante esta jornada. Amo vocês!

À minha noiva Andrezza pelo amor, carinho, compreensão, paciência e constante incentivo durante todos esses anos. Te amo muito!

Aos professores Paulo Valente e Amauri da FEEC/UNICAMP pelo auxílio e importantes contribuições e sugestões.

Aos amigos Pasquini, Luciano, Daniel e tantos outros que fiz no LCA e na FEEC neste período, espero que nossa amizade possa continuar apesar da distância e que possamos repetir nossas aventuras culinárias de vez em quando.

A todos os professores, funcionários da UNICAMP ou aqueles que de maneira direta ou indireta me ajudaram, ensinaram e participaram da minha vida na caminhada até aqui.

Ao CNPq pelo apoio financeiro com o qual foi possível a dedicação integral na elaboração deste trabalho.

A Funcamp pelo apoio financeiro na minha viagem à Grécia.

“Os anos de busca ansiosa no escuro, com um intenso almejar, as alternâncias entre a confiança e a exaustão e emergir, afinal, à luz - apenas os que já passaram por isso são capazes de entendê-lo.”

Albert Einstein

Sumário

Lista de Figuras	xvii
Lista de Tabelas	xix
Lista de Siglas e Abreviações	xxi
1 Introdução	1
1.1 Modelagem por Base de Funções Ortonormais	1
1.1.1 Modelagem de Sistemas Não-lineares	3
1.1.2 Modelagem de Sistemas Lineares no Domínio da Frequência	6
1.2 Organização do Trabalho e Contribuições	7
2 Modelagem de Sistemas Lineares por Base de Funções Ortonormais	11
2.1 Introdução	11
2.2 Base de Funções Ortonormais	12
2.3 Base de Funções Ortonormais Generalizadas	19
2.3.1 Funções Internas	20
2.4 Ajuste dos parâmetros das BFO's	27
2.4.1 Retropropagação através do tempo	28
2.4.2 Métodos de otimização não-linear	29
2.4.3 Estimação e Otimização dos Parâmetros dos modelos por BFO's	35
2.5 Algoritmo	50
2.6 Resultados	50
2.7 Conclusão	57
3 Modelagem de Sistemas Não-lineares por Base de Funções Ortonormais	59
3.1 Introdução	59
3.2 Modelos <i>Fuzzy</i> TS-BFO	62
3.3 Projeto de modelos <i>Fuzzy</i> TS-BFO	66
3.3.1 Premissa das Regras	66
3.3.2 Algoritmo de Gustafson-Kessel	67
3.3.3 Número de Regras	69
3.4 Simplificação da base de Regras dos Modelos <i>Fuzzy</i>	75
3.5 Ajuste global dos parâmetros do modelo final	83

3.6	Algoritmo de Otimização dos modelos <i>fuzzy</i> TS-BFO	92
3.7	Modelos de Volterra	94
3.7.1	Modelo Volterra-GOBF	95
3.7.2	Otimização dos modelos de Volterra-GOBF	97
3.8	Algoritmo de Otimização dos modelos Volterra GOBF	102
3.9	Resultados	103
3.9.1	Sistemas de Volterra Simulados	104
3.9.2	Modelagem de nível líquido em um tanque	110
3.9.3	Planta para produção de Etanol	113
3.9.4	Modelo Racional NARX	117
3.10	Conclusão	120
4	Identificação de modelos BFO no domínio da frequência	123
4.1	Introdução	123
4.2	Identificação na frequência	124
4.3	Algoritmo	130
4.4	Resultados	131
4.5	Conclusão	141
5	Comentários Finais, Contribuições e Perspectivas	143
5.1	Contribuições	145
5.2	Perspectivas e Atividades Futuras	146
	Referências bibliográficas	148
A	Modelo TS-BFO com modelos de Kautz	159
A.1	Ajuste global dos parâmetros do modelo final	159
B	Modelo Não-lineares com GOBF's	163
B.1	Tabelas com os modelos identificados	163
B.1.1	Modelagem de nível líquido em um tanque	163
B.1.2	Planta para produção de Etanol	165
B.1.3	Modelo Racional NARX	166
C	Trabalhos publicados e em desenvolvimento	167

Lista de Figuras

2.1	Representação em Blocos das Funções de Laguerre.	15
2.2	Representação em Blocos das Funções de Kautz.	17
2.3	Representação em blocos de um modelo GOBF.	20
2.4	Representação em blocos da estrutura interna parametrizada por γ	23
2.5	Representação em blocos das funções da base de uma GOBF.	24
2.6	Representação de um modelo com recorrência.	28
2.7	Espaço de busca dos parâmetros b e c	43
2.8	Sinais utilizados na identificação do Modelo de GOBF com 3 polos	51
2.9	Valor da função de custo	53
2.10	Respostas do modelo com 1 polo com multiplicidade 3	53
2.11	Sinais utilizados na identificação do Modelo de GOBF com 4 polos	54
2.12	Respostas do modelos com 2 polos e dois blocos G_b	55
2.13	Respostas do modelos com 3 polos distintos e 3 γ 's	56
3.1	Intersecção de dois conjuntos <i>fuzzy</i> A_1 e A_2 (em escuro).	63
3.2	Exemplo de agrupamento GK e projeção.	67
3.3	Dados de identificação contaminados com ruído.	74
3.4	Resultado dado pela combinação dos critérios de análise de agrupamento.	74
3.5	Resultado do agrupamento dos objetos por GK.	75
3.6	Funções de pertinência dos antecedentes das regras.	75
3.7	Comparação entre os valores simulados e os estimados.	76
3.8	Casos de redundâncias do modelo <i>fuzzy</i>	76
3.9	Grupos de dados definidos por hiperelipsoides.	77
3.10	Fusão de grupos similares	79
3.11	Resultado de fusão entre conjuntos similares	82
3.12	Modelo TS-BFO	85
3.13	Função de pertinência suavizada	87
3.14	Função de pertinência trapezoidal	89
3.15	Sinais utilizados na identificação do modelo de Volterra de segunda ordem.	105
3.16	Valor da função de custo	106
3.17	Respostas do modelo de Volterra com <i>Kernel</i> simétrico	107
3.18	Sinais de identificação - Modelo de Volterra de segunda ordem, <i>kernel</i> não simétrico	108
3.19	Respostas do modelo de Volterra com <i>kernel</i> não simétrico	110
3.20	Estrutura esquemática do tanque de líquido.	111

3.21	Dados utilizados na estimação do modelo.	111
3.22	Saídas do modelo <i>Fuzzy</i> TS-GOBF - pólos iniciais e otimizados.	113
3.23	Figura esquemática da planta industrial para produção de etanol.	114
3.24	Dados de identificação na estimação do processo de Etanol.	115
3.25	Saídas do modelo <i>Fuzzy</i> TS-GOBF - pólos iniciais e otimizados - Sistema de Etanol.	117
3.26	Dados utilizados na estimação do Sistema NARX.	118
3.27	Saídas do modelo <i>Fuzzy</i> TS-GOBF - pólos iniciais e otimizados - Sistema NARX.	120
4.1	Sinais utilizados na identificação do Modelo de Terceira Ordem.	132
4.2	Relação dos módulos da transformada de Fourier discreta dos sinais de saída e entrada.	132
4.3	Relação dos módulos da transformada de Fourier discreta dos sinais de saída e entrada do modelo com 1 polo com multiplicidade 3.	134
4.4	Sinais utilizados na identificação do sistema de quarta ordem.	135
4.5	Relação dos módulos da transformada de Fourier discreta dos sinais de saída e entrada do sistema de quarta ordem.	135
4.6	Relação dos módulos da transformada de Fourier discreta dos sinais de saída e entrada dos modelos com 4 funções de Kautz.	137
4.7	Relação dos módulos da transformada de Fourier discreta dos sinais de saída e entrada do sistema de terceira ordem: Polos complexos + polo real.	137
4.8	Relação dos módulos da transformada de Fourier discreta dos sinais de saída e entrada do modelo GOBF com 3 polos.	138
4.9	Relação dos módulos da transformada de Fourier discreta dos sinais de saída e entrada do modelo GOBF com 5 polos.	139
4.10	Relação dos módulos da transformada de Fourier discreta dos sinais de saída e entrada do modelo GOBF com 3 polos; Frequência específica 0,15 <i>rad</i> a 0,50 <i>rad</i>	140

Lista de Tabelas

2.1	Resultados dos modelos de Laguerre para o primeiro exemplo.	51
2.2	Resultados dos modelos GOBF para o primeiro exemplo.	52
2.3	Resultados dos modelos GOBF para o segundo exemplo.	55
2.4	Resultados dos modelos GOBF para o terceiro exemplo.	56
3.1	Resultados dos modelos Volterra-GOBF	105
3.2	Resultados dos modelos Volterra-GOBF não simétrico	109
3.3	Valores iniciais e finais do EQM e dos polos dos modelos empregados - Sistema de Tanques.	112
3.4	Informação dos modelos do Sistema de Tanques.	112
3.5	Valores iniciais e finais do EQM e dos polos dos modelos empregados - Sistema de Etanol.	116
3.6	Informação dos modelos do Sistema de Etanol.	116
3.7	Valores iniciais e finais dos Parâmetros dos modelos GOBF - Sistema de Etanol.	116
3.8	Valores iniciais e finais dos EQM's e polos dos modelos empregados - Sistema NARX.	118
3.9	Valores iniciais e finais dos Parâmetros dos modelos GOBF - Sistema NARX.	119
3.10	Informação dos modelos do Sistema NARX.	119
4.1	Resultados dos modelos de Laguerre para Ident. na frequência.	133
4.2	Resultados dos modelos GOBF para Ident. na frequência - polos reais.	133
4.3	Resultados dos modelos Kautz para o segundo exemplo.	136
4.4	Resultados dos modelos GOBF para Ident. na frequência - polos complexos.	136
4.5	Resultados dos modelos GOBF para Ident. na frequência - múltiplos modos.	138
4.6	Resultados do modelo GOBF para Ident. na frequência em uma área do espectro de freq.	140
B.1	Modelos para Sistema de nível de líquidos em um tanque.	164
B.2	Modelos para Sistema de Etanol.	165
B.3	Modelos para Sistema Racional NARX.	166

Lista de Siglas e Abreviações

APD	- <i>Average Partition Density</i>
AWCD	- <i>Average Within-Cluster Distance</i>
BFGS	- B royden- F letcher- G oldfarb- S hanno
BPTT	- <i>BackPropagation-Through-Time</i>
BFO	- B ase de F unções O rtormais
CCM	- <i>Compatible Cluster Merging</i>
CSTR	- <i>Continuous Stirred Tank Reactor</i>
EQM	- E rro Q uadrático M édio
FHV	- <i>Fuzzy Hypervolume</i>
FIR	- <i>Finite Impulse Response</i>
FS	- <i>Fuzzy Silhouette</i>
FFT	- <i>Fast Fourier Transform</i>
GK	- G ustafson- K essel
GOBF	- <i>Generalized Orthonormal Basis Functions</i>
IIR	- <i>Infinite Impulse Response</i>
KT	- K uhn- T ucker
NARX	- <i>Nonlinear AutoRegressive with eXogenous inputs</i>
NARMAX	- <i>Nonlinear AutoRegressive Moving Average with eXogenous inputs</i>
NF	- N úmero de F unções
NP	- N úmero de P olos
NF	- N úmero de R egras no Modelo <i>fuzzy</i>
PRBS	- <i>Pseudo Random Binary Sequence</i>
SISO	- <i>Single Interface Single Output</i>
SLTI	- <i>Stable Linear Time Invariant</i>
SQP	- <i>Sequential Quadratic Programming</i>
TS	- T akagi- S ugeno

Lista de Símbolos

$u(k)$	- Entrada de excitação do sistema
$\phi(k)$	- Resposta impulsiva da função ortonormal
$\psi(k)$	- Resposta de uma função ortonormal a uma entrada $u(k)$
K	- Número de regras <i>fuzzy</i> - modelos locais
$\Phi(k)$	- Vetor com as saídas das funções ortonormais
p	- Polo da função de Laguerre
b e c	- Parâmetros da função de Kautz
γ	- Parâmetros da GOBF com funções internas de estrutura <i>ladder</i>
Υ	- Conjunto dos parâmetros γ
n_b	- Número de polos na GOBF
θ	- Parâmetros de otimização em um modelo BFO
ς	- Número de variáveis nos antecedentes das regras <i>fuzzy</i>
μ_i	- Valor de ativação de uma regra <i>fuzzy</i> i
$L_{i,j}$	- Função de pertinência dos modelos TS-BFO
K	- Número de regras dos modelos TS-BFO
N	- Número de objetos utilizados para Identificação
$\tilde{y}(k)$	- Saída do modelo local BFO nos modelos TS-BFO
$\hat{y}$	- Saída estimada
y	- Saída medida ou simulada
$\mathcal{L}$	- Espaço das variáveis <i>fuzzy</i> nos antecedentes das regras
$\mathcal{Y}$	- Espaço da(s) saída(s) do modelos <i>fuzzy</i> TS-BFO
$v_{i,j}$	- Protótipo dos grupos de particionamento
F_i	- Matriz de covariância do grupo i
$S(A, B)$	- Valor de similaridade entre duas funções de pertinência A e B
σ_i	- Parâmetros das funções de pertinência
E	- Ordem do modelo de Volterra
ω	- Frequência do sinal
$Y(\omega)$	- Resposta em frequência do sinal do sistema linear
$\hat{Y}(\omega)$	- Resposta em frequência do modelo estimado
Λ_ω	- Fator de ponderação nas frequências de otimização
$\Phi(\omega)$	- Resposta em frequência da BFO
$U(\omega)$	- Transformada de Fourier do sinal de entrada

Capítulo 1

Introdução

1.1 Modelagem por Base de Funções Ortonormais

A obtenção de modelos de sistemas reais é de importância em quase todas as áreas da ciência. Os modelos auxiliam no entendimento e compreensão das características do processo analisado, por exemplo, através de previsões ou simulações de seu comportamento. Em engenharia, modelos são necessários para o projeto de novos processos e para a análise daqueles já existentes. Normalmente, técnicas avançadas de projeto de controladores, otimização e supervisão são baseadas em modelos dos processos, sendo que a qualidade do modelo influencia diretamente na qualidade da solução final do problema.

Com os avanços obtidos nas últimas décadas nas áreas de computação e informática, houve um impulso no desenvolvimento de técnicas eficientes para a modelagem e controle de sistemas dinâmicos, principalmente para os processos não-lineares. Os recursos computacionais atualmente disponíveis permitem a análise de sistemas complexos, possibilitando a elaboração de modelos matemáticos eficientes e técnicas de controle que atendem a especificações mais rígidas, garantindo melhor desempenho dos processos.

Dentre as diversas técnicas de modelagem de sistemas dinâmicos apresentadas na literatura, este trabalho aborda os modelos obtidos a partir da utilização de funções ortonormais que formem uma base completa para o espaço de Lebesgue $L_2[0, \infty)$, as conhecidas bases de funções ortonormais (BFO) de Laguerre, Kautz e Generalizadas (Heuberger, Van den Hof e Wahlberg, 2005). Modelos com BFO's têm apresentado sucesso nas tarefas de modelagem e controle de sistemas dinâmicos (Wahlberg, 1991; Dumont e Fu, 1993a; Wahlberg, 1994; Van den Hof, 1995; Heuberger, 1995; Bokor e Schipp., 1998; Oliveira e Dumont, 2000; Campello e Amaral, 2002; Campello, Meleiro e Amaral, 2004; Heuberger et al., 2005; Campello, Oliveira e Amaral, 2007; Machado, 2007). Esses modelos apresentam características que os tornam atraentes na modelagem de sistemas dinâmicos, como

a ausência de recursão da saída, o que evita a propagação de erro na predição de sinais e resulta em modelos mais precisos (Campello et al., 2007); a não necessidade de se conhecer previamente a estrutura exata do vetor de regressão (ordens e atrasos de transporte), podendo assim tratar atrasos de transporte de maneira explícita ou implícita; se necessário, é possível aumentar a capacidade de representação do modelo aumentando-se o número de funções ortonormais utilizadas, o que pode ser feito adaptativamente (*online*); desacoplamento natural das saídas em modelos multivariáveis; tolerância a dinâmicas não modeladas; garantia de que a representação de um sistema estável é também estável. Estas propriedades fazem da representação em BFO's interessante quando se deseja modelar processos dinâmicos, principalmente se os modelos forem utilizados para o projeto de sistemas de controle (Nelles, 2001; Heuberger et al., 2005).

Uma base de funções ortonormais pode ser construída selecionando-se os parâmetros de projeto livre (polos) de maneira arbitrária, devido à sua completude (Heuberger et al., 2005). Entretanto, se a escolha da base e dos polos a serem utilizados for adequada, obtém-se um modelo com menor número de funções na base. As bases de funções ortonormais mais comumente utilizadas na representação de sistemas lineares e não-lineares são as bases de Laguerre, de Kautz e a base de funções ortonormais generalizadas. A base de Laguerre é a mais apropriada para representar sistemas com polos puramente reais ou com parte imaginária de valor reduzido. A base de Kautz necessita de um número menor de funções para representar sistemas com dinâmica oscilatória, por ser parametrizada por polos complexos conjugados. Quando o sistema em estudo possui múltiplos modos dominantes (amortecidos ou não), utilizam-se modelos formados por base de funções ortonormais generalizadas (GOBF - *Generalized Orthonormal Basis Functions*), em que as funções ortonormais podem apresentar vários polos distintos. A classe das funções ortonormais generalizadas é obtida pela conexão de n filtros *all-pass* de n_b -ésima ordem, ou seja, cada filtro terá n_b polos distintos ou não, podendo ser de qualquer natureza, reais ou complexos (Ninness, 1997; Heuberger et al., 2005). Um caso especial de GOBF's corresponde em se conectar diferentes filtros *all-pass* de primeira e segunda ordem, onde cada *all-pass* pode apresentar polo(s) distinto(s). Estas bases de funções são também conhecidas por funções de Takenaka-Malmquist (Heuberger et al., 2005). Contudo, as funções de Takenaka-Malmquist apresentam algumas características indesejáveis quando não se conhece os polos dos sistemas e deseja-se identificá-los, sendo tais características: a necessidade da definição *a priori* da ordem e do tipo dos polos das funções; a parametrização das funções não é única e depende da combinação dos polos; modelos com polos complexos geram saídas complexas ou devem utilizar transformações de variáveis (complexo para real) para se obter resposta real. Uma alternativa para estes problemas é a utilização de modelos GOBF com funções internas de estrutura *ladder*. Estes modelos apresentam vantagens por serem parametrizados somente por parâmetros reais, independentemente da natureza dos polos das funções internas (Heuberger et al., 2005). Nesta tese é desenvolvida uma metodologia para a

obtenção de modelos GOBF com funções internas de estrutura *ladder* com representação em espaço de estados.

1.1.1 Modelagem de Sistemas Não-lineares

As bases de funções ortonormais podem ser utilizadas para a modelagem de sistemas dinâmicos estáveis lineares ou não-lineares (Oliveira, Campello e Amaral, 1999; Campello, Amaral e Favier, 2001; Campello e Amaral, 2002; Campello, Favier e Amaral, 2003; da Rosa, Amaral e Campello, 2005; Machado, 2007). A identificação de sistemas não-lineares é um campo que tem se desenvolvido consideravelmente nos últimos anos tendo em vista o estudo de uma classe maior de sistemas, o que permite assim o uso de uma mesma abordagem para uma variedade de sistemas não-lineares com diferentes propriedades (Nelles, 2001). Nesse escopo, inserem-se as representações de sistemas não-lineares através de modelos nebulosos, ou modelos *fuzzy*, em especial os modelos *fuzzy* do tipo Takagi-Sugeno, e dos modelos de Volterra, ambos agregando em sua representação a utilização de BFO's e, em especial, as GOBF's.

Os conceitos de conjuntos *fuzzy*, ou conjuntos nebulosos, foram originalmente propostos por Zadeh (1965). Com os trabalhos de Takagi e Sugeno (1985), introduziu-se uma nova metodologia de modelagem baseada em regras, munida de conjuntos *fuzzy*, cujos modelos são capazes de representar o comportamento de sistemas não-lineares graças à propriedade de serem aproximadores universais de funções em um espaço compacto (Wang e Mendel, 1992; Wang, 1998; Kosko, 1994). A estes modelos dá-se a denominação de modelos *fuzzy* Takagi-Sugeno (TS). Estes modelos consistem de interpolações de múltiplos modelos locais (usualmente lineares ou afins), o que favorece a interpretabilidade matemática e o projeto de sistemas de controle, uma vez que localmente se trabalha com sistemas lineares. Os modelos *fuzzy* TS utilizados neste trabalho apresentam BFO's nos consequentes das regras, com relevante interesse nos modelos GOBF com funções internas de estrutura *ladder*. Portanto, esses modelos, denominados TS-BFO (Campello e Amaral, 2002; Campello, 2002; Medeiros, 2006; Medeiros, Amaral e Campello, 2006; Machado, 2007), dispõem de todas as vantagens destacadas anteriormente para as estruturas dinâmicas BFO. Outras propostas de aproximadores universais de funções utilizando BFO's são encontradas na literatura (envolvendo redes neurais artificiais, funções de base radial, etc.); contudo, tais modelos não são matematicamente interpretáveis e, além disso, não são constituídos por modelos locais, como os modelos *fuzzy* TS (Campello, Meleiro e Amaral, 2004; Machado, 2007).

Outra abordagem utilizada na modelagem de sistemas dinâmicos não-lineares é a classe de modelos polinomiais sem realimentação do sinal de saída, conhecidos como modelos de Volterra (Schetzen, 1980; Rugh, 1981; Nelles, 2001; Aguirre, 2007). Estes modelos atraem o interesse na representação matemática de sistemas não-lineares devido à sua estrutura não-linear caracterizada

como uma generalização direta do modelo de resposta ao impulso, possuindo importantes aplicações envolvendo identificação e controle de processos dinâmicos não-lineares (Eykhof, 1974; Schetzen, 1980; Campello e Oliveira, 2007). Os modelos de Volterra são parametrizados por meio de funções multidimensionais conhecidas como *kernels* de Volterra, sendo estes *kernels*, em geral, representados por um elevado número de parâmetros, fato comum na representação impulsiva de sistemas (por exemplo, modelos IIR ou FIR). Contudo, o número de termos necessários para a representação dos modelos pode ser reduzido utilizando-se um modelo de Volterra associado com uma base de funções ortonormais, onde a representação possui uma dinâmica linear constituída por um conjunto de filtros de funções ortonormais seguida de um mapeamento estático não-linear (Schetzen, 1980; Rugh, 1981; Nelles, 2001). Neste sentido, descreve-se matematicamente cada *kernel* do modelo como uma expansão em bases de funções ortonormais, com a possibilidade dos *kernels* serem simétricos ou não simétricos, de tal forma que, o número de bases será único para cada *kernel* no caso simétrico, ou terá um número de bases igual à dimensão do *kernel* para o caso não simétrico. A combinação destas duas abordagens utilizadas na modelagem de sistemas resulta em um modelo conhecido como Volterra-BFO. A combinação de modelos de Volterra com funções ortonormais permite aproximar o sistema com dinâmicas não-lineares utilizando-se um número menor de termos se comparada à abordagem clássica de modelos de Volterra com resposta impulsiva. A utilização de GOBF em conjunto com os modelos de Volterra, ou Volterra-GOBF, permite a incorporação de características de sistemas não-lineares com múltiplos modos dominantes, o que leva a uma maior eficiência destes modelos.

A principal limitação das abordagens citadas anteriormente é que estas requerem o conhecimento prévio sobre as dinâmicas dominantes do sistema a ser modelado. Em alguns trabalhos encontrados na literatura, como (Broome, 1965; Wahlberg e Makila, 1996; Bokor e Schipp., 1998; Heuberger et al., 2005), propõe-se a seleção automática dos polos que caracterizam os conjuntos de funções ortonormais discretas. Em trabalhos como (Clowes, 1965; Masnadi-Shirazi e Ahmed, 1991; Fu e Dumont, 1993; Oliveira e Silva, 1995; Tanguy, Morvan, Vilbé e C., 1995) trata-se o problema da escolha ótima do polo de Laguerre por meio de abordagens analíticas aplicando-a a modelos lineares, enquanto em outros estudos como (Campello et al., 2001; Campello, 2002; Campello et al., 2003; Campello, Favier e Amaral, 2004) aborda-se o caso de modelos não-lineares do tipo Volterra-BFO. A seleção de polos para funções de Kautz encontra-se em alguns trabalhos na literatura como em (Oliveira e Silva, 1995; Tanguy, Morvan, Vilbé e C., 2000; Tanguy, Morvan, Vilbé e C., 2002), onde os modelos abordados são lineares. Estudos encontrados em (da Rosa et al., 2005; Medeiros et al., 2006; da Rosa, Campello e Amaral, 2007; da Rosa, Campello e Amaral, 2008b; da Rosa, 2009; da Rosa, Campello e Amaral, 2009) apresentam soluções para a identificação e otimização de modelos lineares e não-lineares. Em (da Rosa et al., 2005; da Rosa et al., 2007; da Rosa et al., 2008b; da Rosa, 2009;

da Rosa et al., 2009) são otimizados modelos BFO lineares e não-lineares (modelos Volterra-BFO) com funções de Kautz e GOBF com funções de Takenaka-Malmquist, sendo os dois últimos trabalhos baseados em algoritmos de otimização não-linear.

Neste trabalho propõe-se a identificação de sistemas lineares e não-lineares por meio de modelos com BFO (em especial os modelos GOBF com funções internas de estrutura *ladder*) através da seleção numérica dos polos das bases ortonormais e dos demais parâmetros que compõem os modelos baseada nos dados de entrada e saída do sistema. Neste aspecto, utilizam-se algoritmos de otimização não-linear (Luenberger, 2003) e procura-se obter os gradientes das saídas dos modelos BFO com relação aos polos da base e demais parâmetros, sendo estes gradientes utilizados para se obter a direção a ser seguida no espaço de busca dos parâmetros do modelo. Destaca-se como uma das principais contribuições desta tese, a identificação e otimização de modelos GOBF com funções internas de estrutura *ladder* com a obtenção de cálculos analíticos dos gradientes exatos das bases GOBF's com funções internas em relação aos parâmetros a serem otimizados. Além disso, para a otimização dos modelos de Kautz propõe-se um novo método, em comparação ao encontrado em (da Rosa, 2009; da Rosa et al., 2009), em que se calcula o gradiente das saídas das funções de Kautz em batelada. Apresenta-se também a solução para a identificação e otimização de modelos de Laguerre.

Para a obtenção dos modelos *fuzzy* TS-BFO utiliza-se inicialmente um algoritmo de agrupamento *fuzzy* aplicado aos dados de entrada e saída para a determinação do modelo inicial com suas regras formadas pelas funções de pertinência nos antecedentes e pelos modelos locais BFO nos consequentes. Posteriormente, analisa-se o modelo e realiza-se, se necessário, a simplificação deste com a eliminação de possíveis redundâncias nas regras executando-se algoritmos de fusão e eliminação de regras e funções de pertinência desnecessárias (Babuška, 1998; Setnes, Babuska, Kaymak e van Nauta Lemke, 1998; Kaymak e Babuska, 1995). Dentro deste contexto, apresentam-se, nesta tese, algoritmos para a determinação e análise de casos de simplificação dos modelos baseados em critérios de similaridade entre regras e funções de pertinência do modelo nebuloso. Definido o modelo inicial, pode-se então aplicar o processo de otimização não-linear para o ajuste dos polos, dos coeficientes da expansão das funções ortonormais presentes nos consequentes das regras e dos parâmetros das funções de pertinência nos antecedentes das regras. Para a implementação do algoritmo de otimização deve-se calcular os gradientes do modelo com relação aos seus parâmetros de interesse. Estes gradientes responsáveis fornecem a direção de busca dos valores ótimos dos parâmetros. Assim, são desenvolvidos cálculos analíticos para a determinação dos gradientes da saída do modelo *fuzzy* TS-BFO com relação aos polos da BFO (funções de Laguerre, Kautz ou GOBF com funções internas de estrutura *ladder*), aos coeficientes da expansão das funções ortonormais (consequentes das regras) e aos parâmetros das funções de pertinência pertencentes aos antecedentes das regras.

A abordagem de otimização estende-se também a modelos de Volterra-BFO. Em (da Rosa, 2009; da Rosa et al., 2009) é apresentada uma metodologia para a otimização de modelos Volterra-BFO com *kernels* simétricos e BFO representadas por funções de Kautz ou GOBF com funções de Takenaka-Malmquist. Nesta tese, apresenta-se o algoritmo para a obtenção de modelos Volterra-GOBF com funções internas de estrutura *ladder* com *kernels* simétricos ou não simétricos. Para o processo de otimização, são desenvolvidos cálculos analíticos dos gradientes dos modelos de Volterra-GOBF com funções internas com relação aos seus parâmetros, tanto para os modelos com *kernels* simétricos como para os modelos com *kernels* não simétricos.

1.1.2 Modelagem de Sistemas Lineares no Domínio da Frequência

Além da modelagem e identificação de sistemas dinâmicos no domínio do tempo, é possível também realizar a modelagem e identificação de sistemas dinâmicos, paramétrica ou não-paramétrica, no domínio da frequência (Ljung, 1999; Aguirre, 2007; Pintelon e Schoukens, 2001). Tem crescido o interesse de engenheiros e pesquisadores na utilização de modelos no domínio da frequência devido à facilidade de utilização de analisadores de frequência ou de métodos eficazes de determinação da resposta em frequência, como algoritmos computacionais para o cálculo de transformadas de Fourier de sinais amostrados no tempo (Oppenheim, Schaffer e Schaffer, 1999). Em algumas aplicações, como no caso de sistemas de telecomunicações, as medidas do sistema sob análise estão disponíveis diretamente no domínio da frequência. Outras aplicações de destaque desta abordagem de identificação se dão nos campos da engenharia mecânica, com a análise de estruturas e de vibrações, em sistemas elétricos de potência ou em sistemas de controle.

A identificação no domínio da frequência apresenta vantagens quando comparada com os métodos de identificação no domínio do tempo. Uma dessas vantagens é a redução na dimensionalidade dos dados amostrados no tempo, o que permite simplificar a identificação de sistemas que demandam a aquisição de uma grande quantidade de dados. Outras vantagens são citadas na literatura como a redução do efeito de ruídos, combinação de experimentos, entre outros (Pintelon e Schoukens, 2001; Heuberger et al., 2005). A identificação de sistemas dinâmicos no domínio da frequência pode ser benéfica especialmente em situações onde se deve obter um modelo que represente o processo em certa faixa de frequência, como em aplicações voltadas a sistemas de controle.

Diversos trabalhos na literatura abordam o problema de identificação de sistemas dinâmicos no domínio da frequência (Sidman, DeAngelis e Verghese, 1991; Hakvoort e den Hof, 1994; Lin e Wu, 1982; McKelvey e Akcay, 1994; Pintelon, Guillaume, Rolain, Schoukens e Vanhamme, 1994; Ljung, 1999; Aguirre, 2007; Pintelon e Schoukens, 2001). Nesta tese tem-se especial interesse pela identificação, no domínio da frequência, de sistemas dinâmicos por meio de BFO. Em trabalhos como (Vries e Van den Hof, 1998; Akcay e Heuberger, 2001; Heuberger et al., 2005; Deschrijver

e Dhaene, 2006; Reginato e Oliveira, 2008; Reginato, 2008) propõe-se a identificação de modelos BFO no domínio da frequência. Em (Vries e Van den Hof, 1998; Heuberger et al., 2005; Deschrijver e Dhaene, 2006) faz-se a identificação de modelos BFO lineares obtendo-se os coeficientes da expansão da série de funções ortonormais, com os polos fixos determinados a priori. (Akçay e Heuberger, 2001) propõe a identificação de um modelo BFO aproximado através de um algoritmo iterativo. Em (Reginato e Oliveira, 2008; Reginato, 2008) os autores utilizam métodos de otimização baseados em algoritmos de nuvem de partículas para otimizar os polos da GOBF e métodos de otimização não-lineares para obter os coeficientes da expansão. Neste trabalho é apresentado um processo de identificação no domínio da frequência para modelos BFO baseado na minimização de uma função de custo que considera o erro de resposta em frequência entre o sistema analisado e o modelo a ser obtido, otimizando-se os polos do modelo e os coeficientes da expansão da série de funções ortonormais.

O processo de otimização do modelo BFO baseado no domínio da frequência é semelhante ao processo proposto para otimização de modelos BFO no domínio do tempo. Para a identificação dos modelos são empregados métodos de otimização não-linear que utilizam a informação do gradiente da resposta em frequência dos modelos BFO com relação aos parâmetros de interesse. Os gradientes das funções de Kautz e Laguerre no domínio da frequência são calculados analiticamente através de equações desenvolvidas pelo autor. Para os modelos GOBF emprega-se o cálculo numérico dos gradientes com relação aos parâmetros de interesse.

As abordagens e metodologias propostas nesta tese permitem obter modelos para sistemas dinâmicos dispondo-se de pouco ou nenhum conhecimento a priori das características dinâmicas do sistema sob análise, o que é de grande interesse quando se trata da identificação e modelagem de sistemas dinâmicos.

1.2 Organização do Trabalho e Contribuições

O presente trabalho está organizado da seguinte forma:

- **Capítulo 2 - Modelagem por Base de Funções Ortonormais:** Neste capítulo apresentam-se algumas técnicas de modelagem matemática de sistemas dinâmicos lineares estáveis dentre as quais se dá maior destaque à modelagem por meio de base de funções ortonormais, mostrando suas características e vantagens em relação a outras técnicas de modelagem. Dentre as bases de funções ortonormais, empregam-se as bases de Laguerre, Kautz e as bases de funções generalizadas com funções internas de estrutura *ladder*, às quais se tem especial interesse neste trabalho, devido às suas características vantajosas com relação às outras abordagens, como a liberdade de se trabalhar com múltiplos polos de natureza distintas parametrizados somente por

parâmetros reais. Apresentam-se também técnicas de otimização não-linear empregadas na otimização e identificação de modelos BFO a partir de dados de entrada e saída dos sistemas sob análise. Para a execução dos processos de otimização emprega-se o cálculo dos gradientes das funções com relação a seus parâmetros. Os gradientes são obtidos através de cálculos analíticos com uma nova proposta para o cálculo no caso das funções de Kautz e com a apresentação dos cálculos dos gradientes para os modelos GOBF com funções internas. Para ilustrar o desempenho da metodologia proposta, apresentam-se exemplos de modelagem de sistemas dinâmicos utilizando-se BFO.

- **Capítulo 3 - Modelagem de Sistemas Não-lineares por base de funções ortonormais:** No capítulo 3 apresentam-se as propostas de modelagem de sistemas dinâmicos não-lineares estáveis por meio de abordagens que utilizam BFO, sendo eles os modelos *fuzzy* Takagi-Sugeno com BFO nos consequentes das regras, ou modelos *fuzzy* TS-BFO, e modelos de Volterra com BFO nos *kernels*, ou modelos Volterra-BFO. Dá-se especial ênfase ao emprego de base de funções ortonormais generalizadas com funções internas nos modelos não-lineares citados (modelos TS-GOBF e modelos Volterra-GOBF). Desenvolvem-se técnicas para se obter os modelos TS-GOBF e Volterra-GOBF diretamente através de sinais de entrada e saída dos sistemas sob estudo sem conhecimento a priori de suas características dinâmicas. Para a obtenção dos modelos TS-GOBF emprega-se inicialmente técnicas de agrupamento *fuzzy* para a determinação inicial dos modelos. Posteriormente, utilizam-se metodologias de simplificação de modelos *fuzzy* TS para a análise e possível eliminação de funções de pertinência ou regras desnecessárias. Finalmente, utilizam-se algoritmos de otimização não-linear, auxiliados pelo cálculo analítico dos gradientes destes modelos com relação aos seus parâmetros, para se realizar um ajuste fino no modelo otimizando-se os polos da BFO, os coeficientes da expansão da série BFO e os parâmetros das funções de pertinência e, desta maneira, melhorar sua precisão. Para os modelos de Volterra-GOBF com funções internas apresentam-se os cálculos necessários para se otimizar tais modelos empregando-se metodologias de otimização não-linear e os cálculos analíticos dos gradientes de modelos simétricos e não simétricos utilizados pelos algoritmos de otimização. Para finalizar apresentam-se exemplos de aplicações de modelagem de sistemas não-lineares empregando-se as metodologias propostas.
- **Capítulo 4 - Identificação de modelos BFO no domínio da frequência:** No capítulo 4 apresenta-se o procedimento de identificação de modelos BFO lineares no domínio da frequência. Utilizando-se as informações da resposta em frequência do sistema sob estudo, desenvolvem-se métodos de identificação de modelos BFO lineares com a finalidade de se determinar os polos das funções ortonormais presentes na base, assim como os coeficientes

da expansão da BFO . No processo de identificação são utilizadas técnicas de otimização não-linear que minimizam uma função de custo que considera o erro entre a resposta em frequência do sinal amostrado diretamente do processo e a resposta em frequência do modelo estimado. Para a implementação do processo de otimização no domínio da frequência é necessário o cálculo dos gradientes que determinam as direções de busca dos parâmetros de interesse. Assim, as principais contribuições deste capítulo inserem-se no desenvolvimento do cálculo analítico dos gradientes da resposta em frequência, das funções de Laguerre e Kautz com relação aos parâmetros do modelo e a otimização destes modelos no domínio da frequência. As informações dos gradientes são utilizadas no processo de otimização empregado. Também apresenta-se uma abordagem numérica para a obtenção dos gradientes das respostas em frequência dos modelos GOBF com funções internas de estrutura *ladder* e sua posterior otimização. Ao final do capítulo, apresentam-se exemplos da modelagem de sistemas lineares através de suas respostas no domínio da frequência empregando-se as metodologias propostas.

- **Capítulo 5 - Propostas e Conclusões:** Ao decorrer do capítulo 5 apresentam-se as conclusões dos trabalhos realizados, contribuições e perspectivas de trabalhos futuros.

Capítulo 2

Modelagem de Sistemas Lineares por Base de Funções Ortonormais

2.1 Introdução

A modelagem matemática de sistemas físicos reais é importante em muitas áreas da ciência e do conhecimento. Em engenharia, os modelos são utilizados para a análise, otimização, simulação e controle de processos. Aliada à busca por técnicas avançadas de projeto de controladores, otimização e supervisão há também necessidade de técnicas avançadas e eficientes de modelagem e identificação de processos. Uma destas técnicas consiste em utilizar bases de funções ortonormais (BFO) na modelagem de sistemas dinâmicos.

A utilização de bases de funções ortonormais apresenta vantagens como a redução na ordem da representação do modelo, simplificando os problemas de identificação e controle associados (Van den Hof, 1995; Heuberger et al., 2005). Modelos que utilizam a abordagem BFO possibilitam também a incorporação de conhecimento prévio sobre a dinâmica do sistema (Heuberger, 1995; Ninness, 1997). Deste modo, o problema de estimação dos coeficientes dos modelos pode ter uma rápida taxa de convergência, levando a modelos parcimoniosos quando comparados com outras abordagens, como modelos FIR (*Finite Impulsive Response*) (Ljung, 1999).

Outra característica da utilização de modelos BFO em sistemas dinâmicos é a ausência de realimentação da saída, isto é, não existe realimentação de erros de previsão, o que normalmente resulta em modelos mais precisos (Nelles, 2001). Devido à completude da base também é possível aumentar a sua capacidade de representação aumentando-se o número de funções ortonormais da base (Heuberger et al., 2005). Os modelos baseados em BFO's exibem ainda tolerância a dinâmicas não-modeladas e sensibilidade reduzida aos parâmetros estimados, dentre outras que serão apresentadas neste capítulo (Nelles, 2001). As funções ortonormais utilizadas com maior frequência

na representação através de BFO's são as funções de Laguerre, de Kautz e as bases de funções ortonormais generalizadas (GOBF - *Generalized Orthonormal Basis Function*) (Heuberger et al., 2005; Campello et al., 2007). Neste trabalho propõe-se uma nova metodologia para a obtenção de um modelo em espaço de estados para base de funções ortonormais generalizadas (GOBF) com funções internas de estrutura *ladder* (Heuberger et al., 2005). Estas funções apresentam uma vantagem sobre as demais funções ortonormais normalmente utilizadas, que é a flexibilidade de serem parametrizadas somente por parâmetros reais independentemente da natureza e do número de polos das funções, sejam eles reais, complexos conjugados ou ambos, o que não acontece em outros modelos GOBF. Esta representação de modelos com múltiplos modos dominantes possibilita a representação de sistemas com dinâmicas de ordem elevada através de modelos mais simples do que os obtidos normalmente com funções de Laguerre ou Kautz.

Para a obtenção de modelos BFO mais precisos e parcimoniosos é necessário que o(s) polo(s) do modelo esteja(m) próximo(s) do(s) polo(s) do sistema sob estudo. Esta tarefa nem sempre é simples, pois muitas vezes não se tem conhecimento a priori das características dinâmicas do sistema. Uma maneira de se obter os polos do sistema é pelo emprego de métodos de otimização através da minimização dos erros de saída do modelo (Nelles, 2001; Heuberger et al., 2005; da Rosa, 2009). Neste trabalho, apresenta-se o método para o cálculo analítico dos gradientes das GOBF's com relação aos seus parâmetros. Tais gradientes, associados a métodos de otimização não-linear, permitem a obtenção de modelos precisos e que não necessitam de conhecimento a priori do sistema a ser modelado. Apresentam-se também métodos para o cálculo dos gradientes de modelos BFO com funções de Laguerre e Kautz calculados em batelada, diferentemente do método proposto por da Rosa (2009), onde os gradientes são calculados individualmente para cada função de Kautz.

2.2 Base de Funções Ortonormais

Um sistema linear estável e invariante no tempo pode ser completamente caracterizado por sua resposta ao impulso $h(j)$, $j = 0, \dots, \infty$ discreta no tempo (Ljung, 1999). Admitindo que tal sistema seja causal e que $h(j)$ pertença ao espaço das funções quadraticamente somáveis (Espaço de Lebesgue $L_2[0, \infty)$ no caso contínuo), então a resposta impulsiva do processo, $h(j)$, pode ser modelada através de uma série de funções ortonormais:

$$h(j) = \sum_{i=1}^{\infty} g_i \phi_i(j) \quad (2.1)$$

onde $\{\phi_i(j)\}_{i=1}^{\infty}$ é uma base de funções ortonormais (BFO) e g_i são os coeficientes associados ao desenvolvimento em série da resposta ao impulso $h(j)$ através dessa base. As BFO's são bases com-

pletas no espaço das funções quadraticamente somáveis no intervalo $[0, \infty)$ e suas funções possuem a propriedade de aproximarem com precisão arbitrária, através de uma combinação linear de finitas funções da base, qualquer função linear no espaço L_2 (Heuberger et al., 2005; Campello et al., 2007). Além disso, a representação de sistemas dinâmicos lineares e estáveis através de modelos dados por BFO's apresenta vantagens quando comparados com outros modelos normalmente utilizados para a representação dos mesmos sistemas:

- Ausência de realimentação da saída, o que, conseqüentemente, leva a modelos mais precisos e evita a propagação de erro na predição de sinais;
- A utilização BFO's com polo(s) próximo(s) à dinâmica dominante do sistema requer uma menor quantidade de termos para representar o sistema com a mesma precisão de um modelo FIR.
- Não há a necessidade de determinar previamente a ordem do modelo nem o atraso de transporte, eliminando-se assim a etapa (geralmente árdua) de determinação dos regressores;
- É possível aumentar a capacidade de aproximação dos modelos simplesmente aumentando-se o número de funções, o que pode ser feito de forma adaptativa;
- Os modelos BFO apresentam grande tolerância às dinâmicas não modeladas e sensibilidade reduzida aos parâmetros estimados (dinâmicas não dominantes);
- Capacidade de lidar de forma robusta com atrasos de transporte.
- Desacoplamento natural das saídas em sistemas multivariáveis;
- É garantido que a representação de um sistema estável é também estável.

A equação de convolução para um sistema dinâmico linear pode ser escrita definindo-se a resposta ao impulso $h(j)$ pelo desenvolvimento apresentado na equação (2.1) com uma aproximação com n_f funções ortonormais como:

$$\begin{aligned}
 \hat{y}(k) &= \sum_{j=0}^{\infty} h(j)u(k-j) \\
 &= \sum_{j=0}^{\infty} \sum_{i=1}^{n_f} g_i \phi_i(j)u(k-j) \\
 &= \sum_{i=1}^{n_f} g_i \sum_{j=0}^{\infty} \phi_i(j)u(k-j)
 \end{aligned}$$

$$\hat{y}(k) = \sum_{i=1}^{n_f} g_i \psi_i(k) \quad (2.2)$$

onde $\psi_i(k)$ é a convolução da entrada $u(k)$ com a i -ésima função ortonormal ϕ_i no instante k . Associando-se a cada função ortonormal uma representação equivalente em função de transferência em z que define um filtro linear, $\psi_i(k)$ pode ser então visto como a filtragem do sinal de entrada $u(k)$.

A modelagem por base de funções ortonormais admite a representação do modelo em uma realização em espaço de estados. Dado um conjunto de polos, pode-se construir uma realização em espaço de estados para uma BFO dada pela equação (2.3).

$$\begin{bmatrix} x(k+1) \\ \hat{y}(k) \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x(k) \\ u(k) \end{bmatrix} \quad (2.3)$$

A realização (2.3) não é única e admite diversas representações para a mesma BFO. Para que a realização seja ortogonal ela deve satisfazer a seguinte condição (Roberts e Mullis, 1987):

$$AA^T + BB^T = I \quad (2.4)$$

O modelo apresentado na equação (2.2) pode ser representado no espaço de estados como:

$$\begin{aligned} \Psi(k+1) &= A_f \Psi(k) + B_f u(k) \\ \hat{y}(k) &= C_f \Psi(k) \end{aligned} \quad (2.5)$$

onde as matrizes A_f e B_f dependem diretamente dos parâmetros (polos) do modelo, a matriz C_f contém os termos da expansão da série com a saída das funções ortonormais e $\Psi(k) = [\psi_1(k) \dots \psi_{n_f}(k)]^T$ é o vetor de estados ortonormais (denominação usada por conveniência já que a ortonormalidade é uma propriedade da função e não dos estados) de ordem n_f , sendo estes estados coincidentes com as saídas dos filtros correspondentes às funções ortonormais de ordem equivalente. Este modelo pode ainda incluir uma parcela na parte estática da representação que modela o nível DC da saída do sistema sob estudo.

Duas funções utilizadas com muita frequência na representação através de BFO's, e de especial interesse neste trabalho, são as funções de Laguerre¹ e de Kautz. A base de funções ortonormais de Laguerre é caracterizada pela utilização de funções de transferência com apenas um polo real

¹Edmond Nicolas Laguerre - Nasceu 1834 em Bar-le-Duc, França. Graduou-se pela École Polytechnique em Paris em 1854. Publicou seus trabalhos a respeito das funções polinomiais de Laguerre em 1879 e mostrou que uma função arbitrária pode ser expandida em série de funções polinomiais de Laguerre. Faleceu em 1886 em Bar-le-Duc, França.

(Wahlberg, 1991):

$$\Phi_i(z) = \frac{\sqrt{1-p^2}}{z-p} \left(\frac{1-pz}{z-p} \right)^{i-1}, \quad i = 1, 2, \dots \quad (2.6)$$

onde $p = \{p : p \in \mathbb{R} \text{ e } |p| < 1\}$ é o polo que parametriza as funções ortonormais e i a ordem da função dada. Os modelos com funções de Laguerre podem ser caracterizados como modelos equivalentes a modelos FIR com polos fora da origem, sendo o modelo FIR com polos na origem um caso especial das BFO's (Heuberger et al., 2005; Campello et al., 2007). O modelo FIR normalmente requer um número elevado de termos para representar sistemas com dinâmica dominante lenta. Como se pode notar através da equação (2.6), os modelos de Laguerre de ordem i podem ser escritos como modelos em cascata de ordem $i - 1$, conforme mostra a Fig. 2.1.

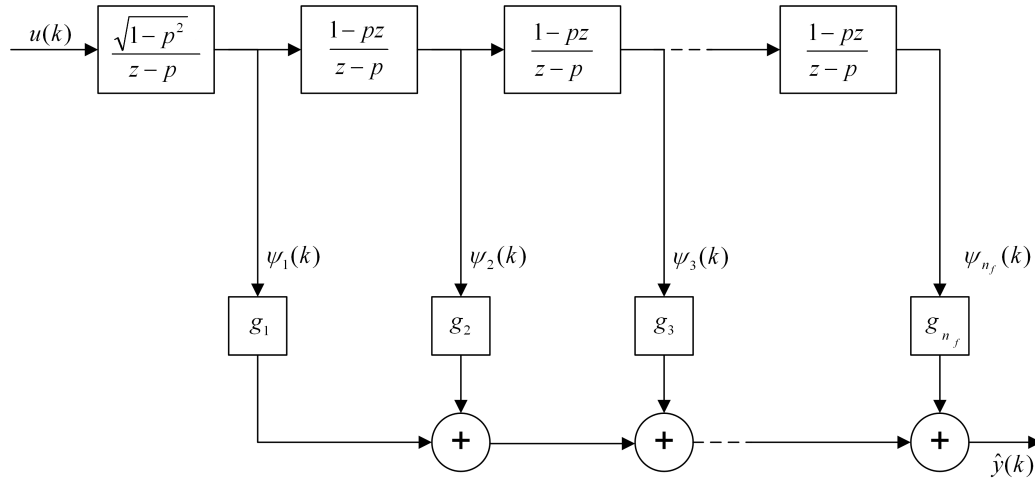


Fig. 2.1: Representação em Blocos das Funções de Laguerre.

A representação em cascata da função de Laguerre lhe confere a propriedade de recursividade já, que a i -ésima função pode ser escrita em função da $(i-1)$ -ésima. Sendo assim, é possível representar a dinâmica de funções de Laguerre através de uma equação de estados, como apresentado na equação (2.5), sendo que as matrizes A_f , B_f e C_f são dadas por:

$$A_f = \begin{bmatrix} p & 0 & 0 & \cdots & 0 \\ 1-p^2 & p & 0 & \cdots & 0 \\ (-p)(1-p^2) & 1-p^2 & p & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (-p)^{n_f-2}(1-p^2) & (-p)^{n_f-3}(1-p^2) & \cdots & p \end{bmatrix}$$

$$B_f = \sqrt{1-p^2} \begin{bmatrix} 1 & -p & (-p)^2 & \cdots & (-p)^{n_f-1} \end{bmatrix}^T$$

$$C_f = \begin{bmatrix} g_1 & g_2 & g_3 & \cdots & g_{n_f} \end{bmatrix}$$

onde p é o polo de Laguerre, n_f é o número de filtros de Laguerre e g_i são os coeficientes da expansão da série. O vetor de estados $\Psi(k)$ é dado por:

$$\Psi(k) = [\psi_1(k) \quad \psi_2(k) \quad \dots \quad \psi_{n_f}(k)]^T$$

com $\psi_i(k)$ sendo a saída da i -ésima função de Laguerre no instante k .

A base de Laguerre é mais apropriada para representar sistemas com polos puramente reais ou com parte imaginária de valor reduzido (Campello et al., 2007). A base de Kautz é mais apropriada para representar sistemas com dinâmica oscilatória, por ser parametrizada por polos complexos conjugados. As funções de Kautz (Wahlberg, 1994) são parametrizadas pelo par de polos complexos conjugados $\beta = \alpha \pm j\vartheta$ segundo a equação:

$$\begin{aligned} \Phi_{2i-1}(z) &= \frac{\sqrt{(1-c^2)}(z-b)}{z^2 + b(c-1)z - c} \left(\frac{-cz^2 + b(c-1)z + 1}{z^2 + b(c-1)z - c} \right)^{i-1} \\ \Phi_{2i}(z) &= \frac{\sqrt{(1-b^2)(1-c^2)}}{z^2 + b(c-1)z - c} \left(\frac{-cz^2 + b(c-1)z + 1}{z^2 + b(c-1)z - c} \right)^{i-1} \end{aligned} \quad (2.7)$$

onde:

$$b = (\beta + \beta^*)/(1 + \beta\beta^*) \text{ e } c = -\beta\beta^* \quad (2.8)$$

sendo $0 < b < 1$ e $-1 < c < 0$. Assim como as funções de Laguerre, as funções de Kautz podem ser descritas em uma representação recursiva e, conseqüentemente, em cascata, como representada na Fig. 2.2.

Analogamente ao desenvolvido para as funções de Laguerre, as funções de Kautz podem ser escritas em espaço de estados. Mbarek, Messaoud e Favier (2003) apresentam uma representação em espaço de estados para as funções de Kautz utilizando duas matrizes A_1 e A_2 que nesta tese foram substituídas pela matriz A_K para facilitar os cálculos do gradiente dos estados com relação aos parâmetros b e c , como será apresentado posteriormente. A representação em espaço de estados da função de Kautz é dada por:

$$\begin{aligned} \mathcal{X}(k+1) &= A_K \mathcal{X}(k) + B_K \mathcal{U}(k) \\ \hat{y}(k) &= C_K \mathcal{X}(k) \end{aligned} \quad (2.9)$$

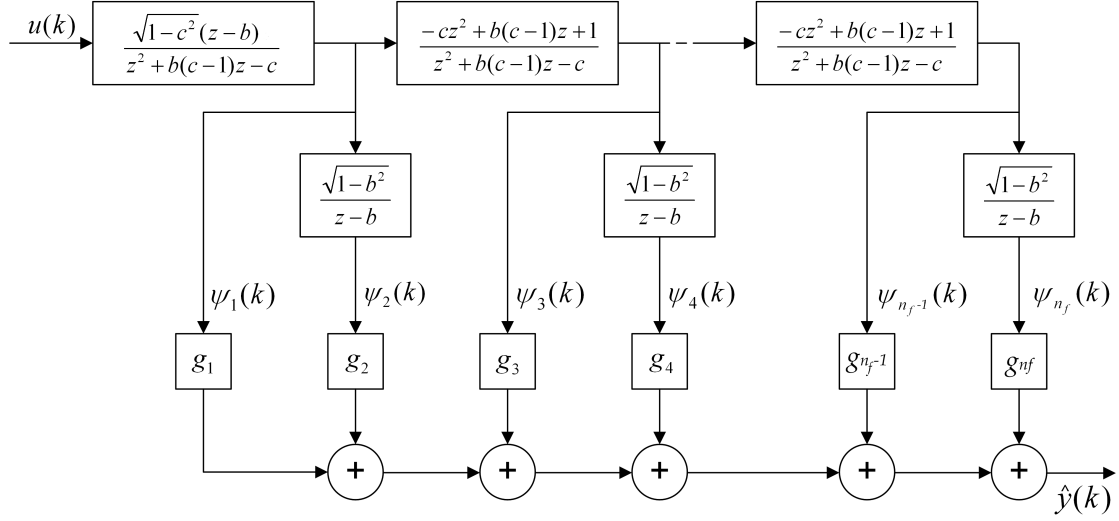


Fig. 2.2: Representação em Blocos das Funções de Kautz.

onde:

$$\mathcal{X}(k) = \begin{bmatrix} \psi_1(k-1) & \psi_1(k) & \cdots & \psi_{n_f-1}(k-1) & \psi_{n_f-1}(k) & \psi_{n_f}(k-1) & \psi_{n_f}(k) \end{bmatrix}^T$$

$$\mathcal{U}(k) = \begin{bmatrix} u(k-1) & u(k) \end{bmatrix}^T$$

$$A_K = \begin{bmatrix} 0 & a_{1,2} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ a_{2,1} & a_{2,2} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a_{3,4} & \cdots & 0 & 0 & 0 & 0 \\ 0 & a_{4,2} & 0 & a_{4,4} & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & a_{2(n_f-1)-1, 2(n_f-1)} & 0 & 0 \\ a_{2(n_f-1), 1} & a_{2(n_f-1), 2} & 0 & 0 & \cdots & a_{2(n_f-1), 2n_f-3} & a_{2(n_f-1), 2n_f-2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & a_{2n_f-1, 2n_f} \\ 0 & 0 & 0 & 0 & \cdots & 0 & a_{2n_f, 2n_f-2} & 0 & a_{2n_f, 2n_f} \end{bmatrix}$$

$$B_K = \sqrt{1 - c^2} \begin{bmatrix} 0 & 0 \\ b_{2,1} & b_{2,2} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ b_{6,1} & b_{6,2} \\ \vdots & \vdots \\ 0 & 0 \\ b_{2(n_f-1),1} & b_{2(n_f-1),2} \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$C_K = \begin{bmatrix} 0 & g_1 & 0 & g_2 & \cdots & 0 & g_{n_f-1} & 0 & g_{n_f} \end{bmatrix}$$

sendo a matriz A_K construída a partir dos seguintes elementos, considerando-se i igual à cardinalidade das linhas da matriz A_K com os termos definidos por $a_{l,m}$, onde l indica a linha e m a coluna:

- Para i ímpar:
 - . $a_{i,2i} = 1$,
 - . demais termos iguais a 0;
- Para $i = 2$:
 - . $a_{i,i} = -b(c - 1)$,
 - . $a_{i,i-1} = c$,
 - . demais termos iguais a 0;
- Para i múltiplo de 4:
 - . $a_{i,i} = b$,
 - . $a_{i,i-2} = \sqrt{1 - b^2}$,
 - . demais termos iguais a 0;
- Para outras linhas i pares os termos serão:
 - . $a_{i,i} = -b(c - 1)$,

- . $a_{i,i-1} = c$,
- . $a_{i,i-j*4} = (-c)^{j-1}(b(c-1) + (-c)(-b(c-1)))$ e
- $a_{i,i-(j*4+1)} = (-c)^{j-1}((-c)(c) + 1)$, para $j = 0, 1, 2, \dots$ tal que $4j < i$,
- . demais termos iguais a 0;

Já os elementos da matriz B_K serão definidos por:

- Para $i = 2 + 4j$, com $j = 0, 1, 2, \dots$ tal que $j \leq n_f/2$
 - . $b_{i,1} = (-c)^j$,
 - . $b_{i,2} = (-c)^j(-b)$;
- Para demais i 's:
 - . $b_{i,1} = b_{i,2} = 0$,

Quando o sistema em estudo necessita de modelos com mais de dois polos distintos, deve-se utilizar modelos formados por base de funções ortonormais generalizadas em que as funções ortonormais podem apresentar n_b polos distintos, como será apresentado na próxima seção.

2.3 Base de Funções Ortonormais Generalizadas

A classe das bases de funções ortonormais generalizadas (GOBF - *Generalized Orthonormal Basis Functions*) é obtida pela conexão de n filtros *all-pass* de n_b -ésima ordem, ou seja, cada filtro terá n_b polos distintos ou não, podendo ser de qualquer natureza, reais ou complexos (Ninness, 1997; Heuberger et al., 2005). A figura 2.3 apresenta a estrutura em blocos de um modelo GOBF, onde os blocos representados por G_b são as funções ortonormais que compõem a base e C_i^T , com $i = 1, \dots, n$, são os vetores de coeficientes que multiplicam a série com as saídas das funções para formar a saída estimada $\hat{y}(k)$.

Um caso particular de GOBF corresponde em conectar diferentes *all-pass* de primeira e segunda ordem, onde cada *all-pass* pode apresentar polo(s) distinto(s). Estas bases de funções são também conhecidas por funções de Takenaka-Malmquist (Heuberger et al., 2005):

$$F_\kappa(z) = \frac{\sqrt{1 - |\xi_\kappa|^2}}{z - \xi_\kappa} \prod_{i=1}^{\kappa-1} \left[\frac{1 - \xi_i z}{z - \xi_i} \right] \quad (2.10)$$

onde $\xi_i, i = 1, \dots, \kappa$ são os polos do modelo GOBF. Modelos com polos complexos conjugados podem ser combinados, correspondendo a funções semelhantes às funções de Kautz. Contudo, este tipo

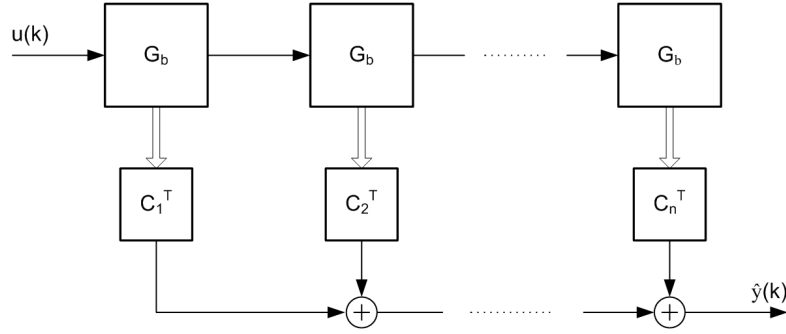


Fig. 2.3: Representação em blocos de um modelo GOBF.

de modelo apresenta algumas desvantagens quando as características dinâmicas do sistema (polos) são desconhecidas e o modelo deve ser obtido por meio de métodos de identificação:

- É necessário se especificar a priori o número de polos reais e polos complexos conjugados;
- A parametrização não é única, uma vez que qualquer permutação de polos levará a modelos idênticos.
- Para a determinação dos parâmetros no caso de polos complexos conjugados (partes reais e imaginárias), há uma interdependência dos parâmetros, pois os polos devem estar dentro do círculo unitário, ou seja, os polos devem ter módulo menor que 1 (Heuberger et al., 2005).
- Saídas dos filtros com polos complexos são complexas (não reais), embora isso possa ser contornado com a transformação proposta por Ninness (1997).

Existe outra representação para GOBF's que não apresenta os problemas descritos acima. Esta representação é conhecida como base de funções ortonormais generalizadas com funções internas de estrutura *ladder*, sendo os modelos BFO de Laguerre e Kautz casos especiais das GOBF's com funções internas (Heuberger et al., 2005).

2.3.1 Funções Internas

Uma função de transferência racional $G(z)$ é denominada “interna” se for estável e satisfizer a seguinte equação (2.11), sendo assim um filtro *all-pass*.

$$G(z)G(1/z) = 1 \quad (2.11)$$

Funções internas bem conhecidas e estudadas são as funções que formam o modelo FIR, i.e., $G(z) = z^{-1}$. Um exemplo de função interna $G(z)$ com polo real ξ_m é dado pela equação (2.12), que claramente satisfaz a condição (2.11):

$$G(z) = \frac{1 - \xi_m z}{z - \xi_m} \quad (2.12)$$

A função interna apresentada na equação (2.12) pode ser escrita na forma de espaço de estados. Sendo $\hat{y}(z) = G(z)u(z)$, a realização em espaço de estados de $G(z)$ é dada através do seguinte desenvolvimento:

$$(z - \xi_m)\hat{y}(z) = (1 - \xi_m z)u(z) \quad (2.13)$$

sendo a equação (2.13) reescrita na forma de equação as diferenças:

$$\hat{y}(k) = \xi_m \hat{y}(k-1) - \xi_m u(k) + u(k-1) \quad (2.14)$$

Definindo-se $x'(k) = \xi_m \hat{y}(k-1) + u(k-1)$, obtém-se:

$$\begin{cases} x'(k+1) &= \xi_m \hat{y}(k) + u(k) \\ \hat{y}(k) &= x'(k) - \xi_m u(k), \end{cases} \quad (2.15)$$

$$\begin{cases} x'(k+1) &= \xi_m x'(k) + (1 - \xi_m^2)u(k) \\ \hat{y}(k) &= x'(k) - \xi_m u(k), \end{cases} \quad (2.16)$$

Definindo-se $x(k) = x'(k)/\sqrt{1 - \xi_m^2}$, tem-se que:

$$\begin{cases} \sqrt{1 - \xi_m^2}x(k+1) &= \xi_m \sqrt{1 - \xi_m^2}x(k) + (1 - \xi_m^2)u(k) \\ \hat{y}(k) &= \sqrt{1 - \xi_m^2}x(k) - \xi_m u(k), \end{cases} \quad (2.17)$$

ou

$$\begin{cases} x(k+1) &= \xi_m x(k) + \sqrt{1 - \xi_m^2}u(k) \\ \hat{y}(k) &= \sqrt{1 - \xi_m^2}x(k) - \xi_m u(k), \end{cases} \quad (2.18)$$

ou equivalentemente,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} \xi_m & \sqrt{1 - \xi_m^2} \\ \sqrt{1 - \xi_m^2} & -\xi_m \end{bmatrix} \quad (2.19)$$

Esta representação de modelos com polo real é equivalente à representação com funções de Laguerre (Heuberger et al., 2005). Modelos de sistemas dinâmicos com par de polos complexos conjugados também possuem representação interna. Sejam as seguintes funções internas com polos complexo

conjugados ξ_m e ξ_m^* :

$$G_1(z) = \frac{1 - \xi_m^* z}{z - \xi_m} \quad (2.20)$$

$$G_2(z) = \frac{1 - \xi_m z}{z - \xi_m^*} \quad (2.21)$$

Ligando-se em cascata as funções (2.20) e (2.21), obtém-se uma função interna de segunda ordem como apresenta-se a seguir:

$$G(z) = \frac{1 - \xi_m^* z}{z - \xi_m} \frac{1 - \xi_m z}{z - \xi_m^*}$$

que pode se reescrever como:

$$G(z) = \frac{1 - (\xi_m + \xi_m^*)z + \xi_m \xi_m^* z^2}{z^2 - (\xi_m + \xi_m^*)z + \xi_m \xi_m^*} \quad (2.22)$$

Como é possível notar por inspeção, a equação (2.22) é uma equação a parâmetros complexos. Para obter-se uma equação a parâmetros reais, introduzir-se-ão os seguintes parâmetros:

$$b = \frac{\xi_m + \xi_m^*}{1 - \xi_m \xi_m^*} \quad \text{e} \quad c = -\xi_m \xi_m^* \quad (2.23)$$

Utilizando-se os parâmetros b e c apresentados em (2.23), pode-se reescrever a equação (2.22) da seguinte maneira com parâmetros reais:

$$G(z) = \frac{-cz^2 + b(c-1)z + 1}{z^2 + b(c-1)z - c} \quad (2.24)$$

esta representação é equivalente às funções de Kautz. Efetuando-se um desenvolvimento semelhante ao realizado para o modelo com polo real, é possível se obter um modelo em espaço de estados onde as matrizes do modelo são dadas por:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \left[\begin{array}{cc|c} b & c\sqrt{1-b^2} & \sqrt{1-b^2}\sqrt{1-c^2} \\ \sqrt{1-b^2} & -bc & -b\sqrt{1-c^2} \\ \hline 0 & \sqrt{1-c^2} & -c \end{array} \right]$$

A conexão em cascata de n_b funções internas gera outras funções internas de ordem n_b , mas quando envolvem polos complexos, seus parâmetros geralmente se tornam complexos ou necessitam de alguma transformação para que sejam reais. Para evitar este problema, há uma representação para modelos GOBF através de funções internas em que os polos são parametrizados sempre por parâmetros reais independentemente da natureza dos polos. Esta representação foi proposta por Gray e Markel (1975) como uma realização em espaço de estados para uma função interna com somente um

polo com duas entradas e duas saídas cujo modelo em espaço de estados, conhecida como estrutura *ladder*, é dada por (Gray e Markel, 1975; Heuberger et al., 2005):

$$\begin{bmatrix} \varphi_i(k+1) \\ y_1(k) \\ y_2(k) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ \sqrt{1-\gamma^2} & -\gamma & 0 \\ \gamma & \sqrt{1-\gamma^2} & 0 \end{bmatrix} \begin{bmatrix} \varphi_i(k) \\ u_1(k) \\ u_2(k) \end{bmatrix} \quad (2.25)$$

com $-1 < \gamma < 1$. O modelo (2.25) pode ser representado através do diagrama de blocos apresentado na figura 2.4:

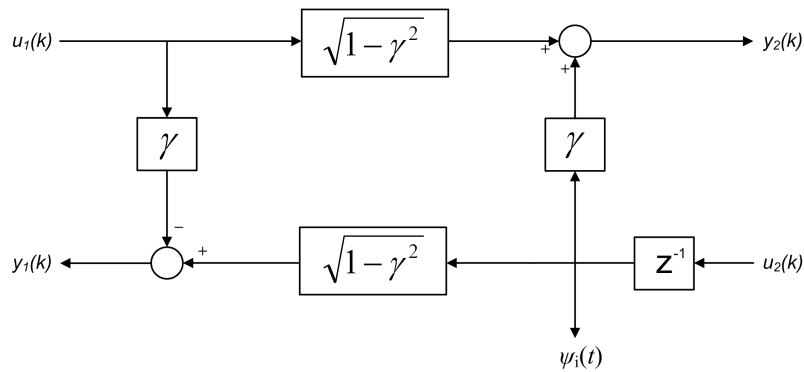


Fig. 2.4: Representação em blocos da estrutura interna parametrizada por γ .

O modelo em espaço de estados apresentado em (2.25) deriva do modelo de primeira ordem apresentado em (2.18), onde a saída $y_2(k)$ corresponde à variável $x(k+1)$ (ou $\psi_i(k+1)$). Com a saída $y_2(k)$ conectada à entrada $u_2(k)$ obtém-se a representação do modelo (2.18). Esta estrutura favorece a conexão em cascata das funções de primeira ordem e fornece uma metodologia para se obter modelos GOBF parametrizados somente por parâmetros reais (Gray e Markel, 1975; Roberts e Mullis, 1987; Heuberger et al., 2005). Por questões de simplificação de nomenclatura os modelos GOBF com funções internas de estrutura *ladder* serão referidos no texto a partir deste ponto somente como modelos GOBF ou modelos GOBF com funções internas.

Esta estrutura é normalizada, isto é, gera uma base ortonormal. A ligação dos n_b blocos que formam uma base ortonormal generalizada com n_b polos, que podem ou não serem distintos, se dá através da estrutura de ligação apresentada na figura 2.5:

As estruturas GOBF's são formadas, portanto, como mostrado na figura 2.3, pela ligação de n blocos G_b formados pela base de funções ortonormais com polos reais e/ou complexos conjugados, reduzindo-se, desta maneira, a quantidade de funções necessárias à representação de um determinado processo. Para representar esta estrutura GOBF, desenvolve-se nesta tese uma representação original em espaço de estados para se obter as saídas das funções na base. Inicialmente será apresentado o

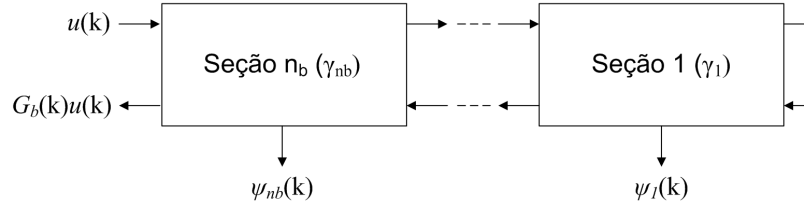


Fig. 2.5: Representação em blocos das funções da base de uma GOBF.

modelo em espaço de estados de um único bloco G_b com n_b funções internas, parametrizadas pelos parâmetros γ_i com $i = 1, \dots, n_b$. O modelo em espaço de estados para G_b é dado por:

$$\begin{aligned}\varphi(k+1) &= A_b \varphi(k) + B_b u(k) \\ \hat{y}(k) &= C_b \varphi(k)\end{aligned}\tag{2.26}$$

onde $\varphi(k) = [\varphi_1(k) \dots \varphi_{n_b}(k)]$ são os estados (ou saída dos filtros) que compõem o bloco G_b da GOBF. As matrizes A_b e B_b dependem diretamente dos parâmetros γ_i , com $i = 1, \dots, n_b$. A matriz A_b do modelo em espaço de estados será definida por:

$$A_b = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n_b} \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots & a_{2,n_b} \\ a_{3,1} & a_{3,2} & a_{3,3} & \cdots & a_{3,n_b} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n_b,1} & a_{n_b,2} & a_{n_b,3} & \cdots & a_{n_b,n_b} \end{bmatrix}\tag{2.27}$$

onde os termos $a_{i,j}$, com i sendo a linha a que o elemento pertence e j a coluna, são definidos por:

- Para $i = 1$, os elementos são:

$$\begin{aligned}a_{1,j} &= \gamma_1, & \text{para } j = 1 \\ a_{1,j} &= \gamma_j \prod_{l=1}^{j-1} \sqrt{1 - \gamma_l^2}, & \text{para } j = 2, \dots, n_b.\end{aligned}$$

- Para $i = 2, \dots, n_b$, os elementos são:

$$\begin{aligned}a_{i,j} &= \sqrt{1 - \gamma_j^2}, & \text{para } j = i - 1; \\ a_{i,j} &= -\gamma_{j-1} \gamma_j, & \text{para } j = i;\end{aligned}$$

$$\begin{aligned}
 a_{i,j} &= -\gamma_{i-1}\gamma_j \prod_{l=i}^{j-1} \sqrt{1 - \gamma_l^2}, \quad \text{para } j = i + 1, \dots, n_b; \\
 a_{i,j} &= 0, \quad \text{para demais elementos.}
 \end{aligned}$$

A matriz B_b do modelo (2.26) é definida por:

$$B_b = \begin{bmatrix} b_{1,1} \\ b_{2,1} \\ b_{3,1} \\ \vdots \\ b_{n_b,1} \end{bmatrix} \quad (2.28)$$

com seus elementos $b_{i,1}$ dados por:

- Para $i = 1$, o elemento será:

$$b_{1,1} = \prod_{l=1}^{n_b} \sqrt{1 - \gamma_l^2}.$$

- Para $i = 2, \dots, n_b$, os elementos são:

$$b_{i,1} = -\gamma_{i-1} \prod_{l=i}^{n_b} \sqrt{1 - \gamma_l^2}$$

A matriz C_b presente no modelo (2.26) é dada por:

$$C_b = \begin{bmatrix} g_1 & g_2 & \dots & g_{n_b} \end{bmatrix}$$

onde g_j com $j = 1, \dots, n_b$ são os coeficientes da expansão em série do modelo GOBF.

Modelos GOBF podem ser constituídos por n funções G_b . Isto é, neste caso há conexão em cascata de n funções generalizadas G_b . O modelo geral da base de funções ortonormais generalizadas em espaço de estados (no formato apresentado em (2.5)) é dado por:

$$\begin{aligned}
 \Psi(k+1) &= A_f \Psi(k) + B_f u(k) \\
 \hat{y}(k) &= C_f \Psi(k)
 \end{aligned} \quad (2.29)$$

onde $\Psi(k)$ é formado por $n_b n$ saídas das funções da base. A matriz A_f do modelo (2.29) é definida

da seguinte forma:

$$A_f = \begin{bmatrix} A_b & 0 & 0 & \cdots & 0 \\ A_I & A_b & 0 & \cdots & 0 \\ (-\gamma_{n_b})A_I & A_I & A_b & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (-\gamma_{n_b})^{n-2}A_I & (-\gamma_{n_b})^{n-3}A_I & (-\gamma_{n_b})^{n-4}A_I & \cdots & A_b \end{bmatrix} \quad (2.30)$$

sendo as submatrizes A_b definidas em (2.27). As submatrizes A_I têm seus elementos $a_{I_{i,j}}$, onde i é a linha e j a coluna, dados por:

- Para $i = 1, \dots, n_b$ e $j \neq n_b$, o elemento $a_{I_{i,j}}$ será:

$$a_{I_{i,j}} = 0;$$

- Para $j = n_b$, então os elementos $a_{i,j}$ serão:

Se $i = 1$

$$a_{I_{i,j}} = (1 - \gamma_{n_b}^2) \prod_{l=1}^{n_b-1} \sqrt{1 - \gamma_l^2}.$$

Se $i = 2, \dots, n_b$

$$a_{I_{i,j}} = (1 - \gamma_{n_b}^2)(-\gamma_{i-1}) \prod_{l=i}^{n_b-1} \sqrt{1 - \gamma_l^2}.$$

A matriz B_f presente no modelo (2.29) é formada como mostrado a seguir:

$$B_f = \begin{bmatrix} B_b \\ (-\gamma_{n_b})B_b \\ (-\gamma_{n_b})^2 B_b \\ \vdots \\ (-\gamma_{n_b})^{n-1} B_b \end{bmatrix} \quad (2.31)$$

com a submatriz B_b definida em (2.28).

A matriz C_f é definida por:

$$C_f = \begin{bmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,n_b} & \cdots & g_{1,n_b n} \end{bmatrix}$$

onde $g_{1,j}$ com $j = 1, \dots, n_b n$ são os coeficientes da expansão em série do modelo GOBF.

A representação em espaço de estados dos modelos GOBF com funções internas de estrutura *ladder* foi obtida a partir da representação gráfica em diagrama de blocos. O modelos escritos em função das entradas e das variáveis de interesse chegando-se à representação apresentada.

A representação em espaço de estados do modelo GOBF com funções internas possibilita que se conheça os polos do modelo através da determinação dos autovalores da matriz A_f . Sendo o modelo GOBF definido pelas matrizes da representação dinâmica em espaço de estados (A_f, B_f, C_f) (equação (2.29)), uma possível realização do sistema para representação através de uma função de transferência é dada por (Heuberger, 1995):

$$G_k(z) = C_f(Iz - A_f)^{-1}B_f$$

onde os polos do sistema podem ser definidos através da obtenção dos autovalores da matriz A_f . Esta propriedade é interessante já que o modelo é parametrizado somente pelos parâmetros reais γ independentemente das características dos polos.

2.4 Ajuste dos parâmetros das BFO's

Os parâmetros a serem determinados na modelagem de sistemas dinâmicos lineares por base de funções ortonormais são os polos das funções ortonormais e os coeficientes da expansão da série de funções. Conforme se discutiu em (Oliveira, 1997; Campello, 2002; Campello et al., 2007; Machado, 2007), uma vez conhecidos os polos, pode-se determinar os termos da expansão da série através do já bem conhecido método de mínimos quadrados (MMQ).

A seleção do(s) polo(s) da base de funções ortonormais que representam a dinâmica do sistema normalmente depende de um conhecimento prévio do sistema ou de métodos de otimização que minimizem o erro do modelo estimado, como apresentado em (Heuberger et al., 2005; Nelles, 2001; da Rosa, 2009). Em trabalhos como (da Rosa, Campello e Amaral, 2008a; da Rosa, 2009; da Rosa et al., 2009) apresenta-se uma metodologia de otimização dos parâmetros de modelos de Kautz e GOBF com funções de Takenaka-Malmquist através de buscas exatas dadas pelos cálculos dos gradientes da saída do modelo estimado com relação aos seus parâmetros. Nesta seção propõem-se a otimização, através de métodos de otimização não-linear (Luenberger, 2003), dos parâmetros de um modelo linear composto por funções ortonormais, sejam elas as funções de Laguerre, de Kautz ou GOBF com funções internas, sendo proposta para as funções de Kautz uma metodologia computacionalmente mais eficiente para o cálculo dos gradientes quando comparado aos resultados apresentados em (da Rosa et al., 2008a). Apresentam-se também métodos para o cálculo analítico dos gradientes das funções de Laguerre e GOBF com funções internas com relação aos parâmetros de interesse para

otimização.

2.4.1 Retropropagação através do tempo

Os modelos são otimizados utilizando-se a informação dos sinais de entrada e saída amostrados em k instantes de tempo. Para o cálculo dos gradientes empregados no processo de otimização utiliza-se a técnica de retropropagação através do tempo, onde todas as amostras nos instantes passados são consideradas no processo de otimização. A técnica de retropropagação através do tempo (BPTT²) foi desenvolvida por Rumelhart, Hinton e Williams (1986) como uma extensão do algoritmo de retropropagação padrão (Haykin, 1999; Nelles, 2001) aplicado à sintonia de parâmetros de modelos recursivos.

Considere um sistema dinâmico linear representado por um mapeamento estático $f(\cdot)$ aplicado sobre os valores passados dos sinais de entrada e saída, ou seja, $\hat{y}(k) = f[u(k-1), \dots, u(k-n_u), \hat{y}(k-1), \dots, \hat{y}(k-n_y)]$, onde $\hat{y}(k)$ denota a saída do modelo. Neste tipo de modelo, conhecido como *Output-Error* (OE) (Nelles, 2001), o mapeamento $f(\cdot)$ possui uma relação dinâmica com a saída do modelo, uma vez que ele está implicitamente e recursivamente incluso nos termos $\hat{y}(k-i)$. Sendo assim, é possível representar a dinâmica do sistema por meio de vários modelos estáticos, o que pode ser feito através de recursões no tempo. Na técnica BPTT, desenvolve-se a recorrência do modelo até as condições iniciais das entradas, estados e saídas de um sistema dinâmico, como mostra a Fig. 2.6 para um sistema do tipo $\hat{y}(k) = f(\theta, u(k-1), \hat{y}(k-1))$, onde θ são os parâmetros do sistema $f(\cdot)$ em questão.

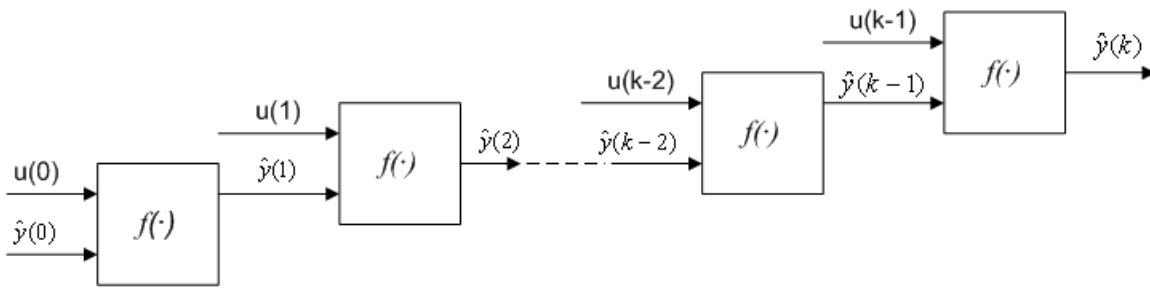


Fig. 2.6: Representação de um modelo com recorrência.

Os parâmetros θ do modelo são obtidos utilizando-se métodos de otimização a partir dos dados de E/S, minimizando-se alguma função de custo que relacione o erro entre os valores de saída obtidos diretamente da planta e os valores estimados através do modelo. Para esta minimização, deve-se calcular o seu gradiente com relação aos parâmetros que se deseja otimizar, necessitando-se, portanto,

²backpropagation-through-time

calcular a derivada da saída $\hat{y}(k)$ com relação aos parâmetros θ . Como o modelo $f(\cdot)$ é assumido linear, pode-se inferir que:

$$\frac{\partial \hat{y}(k)}{\partial \theta} = \frac{\partial f(\cdot)}{\partial \theta} + \frac{\partial f(\cdot)}{\partial \hat{y}(k-1)} \cdot \frac{\partial \hat{y}(k-1)}{\partial \theta} \quad (2.32)$$

onde o primeiro termo da equação (2.32) corresponde à parte estática do modelo, na qual a saída atual depende diretamente dos parâmetros θ . Já o segundo termo da equação (2.32) corresponde à parte dinâmica, onde a saída em $\hat{y}(k-1)$, que também compõe a saída estimada de maneira recursiva, depende dos parâmetros θ que se deseja otimizar. Para modelos em espaço de estados, a saída $\hat{y}(k-1)$ pode ser substituída pelo valor do estado no instante $(k-1)$. O cálculo do segundo termo da equação exige que se conheça a derivada da saída $\hat{y}(k-1)$ com relação aos parâmetros θ , uma vez que a saída pode ser escrita como $\hat{y}(k-1) = f(\theta, u(k-2), \hat{y}(k-2))$, o que corresponde a uma defasagem no tempo para o instante anterior. Da mesma forma que na equação (2.32), o cálculo de $\frac{\partial \hat{y}(k)}{\partial \theta}$ dependerá da saída no instante anterior e assim sucessivamente até $k = 0$, sendo que, neste instante, tem-se $\hat{y}(0) = 0$ e, conseqüentemente, $\partial \hat{y}(0)/\partial \theta = 0$. Em suma, a técnica BPTT decompõe a dinâmica do sistema em uma série de representações estáticas. Esta abordagem permite descrever as derivadas da saída do modelo em termos somente das condições iniciais e do sinal de entrada, fazendo-se uma recursão de k passos no tempo.

Assim, o algoritmo BPTT calcula a derivada exata para as k amostras do sistema utilizadas na identificação do modelo da planta. Conhecidos os gradientes das saídas com relação aos parâmetros θ pode-se, através de alguma técnica de otimização não-linear baseada em gradiente, obter os valores de θ que parametrizam a função $f(\cdot)$.

2.4.2 Métodos de otimização não-linear

Neste trabalho, técnicas de otimização não-lineares são necessárias para se obter os parâmetros dos modelos dos sistemas, o que se dá pela minimização de uma determinada função de custo com relação a um grupo de parâmetros. Os parâmetros a serem otimizados nos modelos de sistemas dinâmicos lineares por BFO's ou GOBF's são os polos das funções ortonormais e os coeficientes das expansões das funções. A otimização de modelos de Kautz se dá através de métodos de otimização que resolvam problemas com restrições não-lineares. Já os modelos de Laguerre e GOBF por funções internas não apresentam restrições não-lineares.

Os problemas de otimização não-lineares podem ser irrestritos ou mais completos, como o apresentado em (2.33), apresentando restrições quanto aos valores que os parâmetros a serem otimizados podem assumir, sendo estas restrições de igualdade ou desigualdade, ou ainda serem restritas a limi-

tes, que de certa forma, podem ser vistas também como restrições de desigualdades.

$$\begin{aligned}
 & \min_{x \in \mathbb{R}^n} f(x) \\
 & \text{sujeito a} \\
 & G_i(x) = 0, \quad i = 1, \dots, m_e \\
 & G_i(x) \leq 0, \quad i = m_e + 1, \dots, m \\
 & x_l \leq x \leq x_u
 \end{aligned} \tag{2.33}$$

onde x é o vetor de variáveis que se deseja otimizar ($x \in \mathbb{R}^n$), $f(x)$ é a função objetivo que retorna um valor escalar ($f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$), $G(x)$ é o vetor de funções que retorna os valores das restrições de igualdade e desigualdade avaliados em x ($G(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m$), x_l e x_u são respectivamente os limites inferiores e superiores das variáveis.

Uma solução eficiente e precisa para problemas como o apresentado em (2.33) depende de diversos fatores, como a correta formulação do problema e da determinação das restrições de forma consistente para não se gerar infactibilidades. Além disso, a escolha do método de otimização a ser utilizado para resolução do problema está diretamente ligada ao problema de otimização que se deseja resolver. Quando o problema trata somente de função objetivo e restrições lineares, este pode ser resolvido por métodos de programação linear. Já problemas não-lineares podem apresentar não-linearidades na função objetivo, nas restrições ou em ambos, podendo também serem restritos ou irrestritos (Nelles, 2001). Para solucionar problemas de otimização não-lineares, existem diversas técnicas, a maioria delas baseada na informação de gradiente que indica a direção de busca dos valores ótimos dos parâmetros que se desejar otimizar, como métodos de gradiente descendente, métodos de Newton, quasi-Newton, entre outros (Luenberger, 2003). Dos métodos que utilizam informação de gradiente, os métodos mais eficientes são os métodos quasi-Newton. Estes métodos constroem a informação de curvatura da função objetivo a cada repetição para formular um problema quadrático da forma

$$\min_x \frac{1}{2} x^T Q x + c^T x + b$$

onde a matriz Hessiana, Q , é definida positiva simétrica, c é um vetor constante, e b é uma constante. A solução ótima para este problema acontece quando as derivadas parciais de x se anulam, i.e.,

$$\nabla f(x^*) = Qx^* + c = 0$$

com o ponto de solução ótima podendo ser reescrito como

$$x^* = -Q^{-1}c$$

Métodos do tipo Newton calculam Q^{-1} diretamente e seguem em uma direção de descida para localizar o mínimo após um determinado número de iterações. Contudo, calcular numericamente a matriz Q^{-1} envolve um grande esforço computacional. Métodos quasi-Newton evitam isto usando o comportamento observado de $f(x)$ e $\nabla f(x)$ para construir a informação de curvatura e assim fazer uma aproximação H para Q^{-1} utilizando uma técnica de atualização apropriada para tal matriz (Luenberger, 2003). Um grande número de métodos de atualização da Hessiana foi desenvolvido, porém, a fórmula de Broyden, Fletcher, Goldfarb, e Shanno (BFGS) se apresenta como o método mais eficiente para a solução de problemas de propósito geral. A fórmula dada por BFGS é (Luenberger, 2003):

$$H_{k+1} = H_k + \frac{q_k q_k^T}{q_k^T s_k} - \frac{H_k^T s_k^T s_k H_k}{s_k^T H_k s_k}$$

onde

$$\begin{aligned} s_k &= x_{k+1} - x_k \\ q_k &= \nabla f(x_{k+1}) - \nabla f(x_k) \end{aligned}$$

Inicialmente, H_0 pode ser fixada como uma matriz definida positiva simétrica qualquer, como por exemplo, a matriz identidade I . Um outro procedimento bem conhecido para se obter a matriz H é a fórmula DFP, de Davidon, Fletcher, e Powell (Scales, 1985). Contudo, resultados comparativos mostram que o método BFGS é superior ao algoritmo DFP (Scales, 1985).

A informação de gradiente é provida pelo cálculo analítico ou através de processos numéricos que utilizam método de diferenças finitas. Uma vez conhecido o gradiente da função objetivo com respeito à(s) variável(eis) a ser(em) otimizada(s), a cada iteração k uma busca linear é executada na direção

$$d = -H_k \cdot \nabla f(x_k)$$

Algumas importantes propriedades dos métodos quasi-Newton são (Nelles, 2001):

- Não há necessidade de cálculo de derivadas de segunda ordem;
- Custo computacional quadrático para multiplicação de matrizes (Métodos de Newton apresentam custo computacional cúbico para o mesmo cálculo);
- Complexidade computacional quadrática de memória uma vez que o valor da Hessiana deve ser armazenado;
- Convergência rápida;

O método BFGS é definido para solução de problemas irrestritos, contudo, a otimização de alguns modelos estudados neste trabalho apresenta restrições, como será visto na sequência, exigindo assim a utilização de algoritmos que resolvam tais problemas. Em otimização restrita, o objetivo geral consiste em transformar o problema restrito em um subproblema mais simples que pode ser resolvido e cuja solução será usada como a base de um processo iterativo. Uma característica comum a uma ampla classe de métodos é a de transformar o problema restrito em um problema básico irrestrito usando uma função de penalidade para as restrições que estão próximas ou entre as restrições. Deste modo o problema restrito é resolvido usando-se uma sucessão de processos de otimização parametrizados e irrestritos que no limite converge para a solução do problema restrito. Porém, tais métodos são considerados atualmente relativamente ineficientes e foram substituídos por métodos que são baseados na solução das equações de Kuhn-Tucker (KT) (Nelles, 2001). As equações de KT são condições necessárias para a otimalidade de um problema de otimização restrita. Se o problema de otimização for um problema convexo, isto é, $f(x)$ e $G_i(x), i = 1, \dots, m$, são funções convexas, então as equações de KT são ambas necessárias e suficientes para que haja um ponto de mínimo global.

Seja o problema de otimização definido em (2.33), as equações de Kuhn-Tucker para tal problema podem ser definidas como:

$$\begin{aligned} \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla G_i(x^*) &= 0 \\ \lambda_i^* G_i(x^*) &= 0 \quad i = 1, \dots, m \\ \lambda_i^* &\geq 0 \quad i = m_e + 1, \dots, m \end{aligned} \quad (2.34)$$

onde a primeira equação indica que o gradiente da função objetivo é igual à combinação linear dos gradientes das restrições ativas no ponto de solução x^* . Além disso, a primeira equação descreve o cancelamento entre o gradiente da função objetivo e as restrições ativas no ponto de solução x^* . Para o gradiente ser cancelado, multiplicadores de Lagrange ($\lambda_i^*, i = 1, \dots, m$) são necessários para balancear as diferenças em magnitude entre o gradiente da função objetivo e o gradiente das restrições. Devido ao fato de somente restrições ativas serem consideradas na operação de cancelamento, os multiplicadores de Lagrange referentes às restrições não ativas devem ser considerados iguais a 0, o que está implícito nas duas últimas equações de (2.34).

A solução das equações de KT forma a base de muitos algoritmos de otimização não-linear, sendo que tais algoritmos tentam obter diretamente os multiplicadores de Lagrange. Métodos de otimização quasi-Newton garantem a convergência superlinear pelo uso de informações de segunda ordem das equações de KT. Estes métodos são comumente conhecidos como métodos de Programação Sequencial Quadrática (SQP - *Sequential Quadratic Programming*), uma vez que subproblemas de programação quadrática são resolvidos a cada iteração. Baseado em trabalhos como

(Powell, 1978b; Powell, 1978a), o método permite imitar métodos quasi-Newton para solução de problemas de otimização restritos como se fossem problemas irrestritos. A cada iteração principal, uma aproximação do lagrangeano ao problema e da Hessiana da função objetivo é feita através de métodos quasi-Newton, sendo estes então utilizados na geração de subproblemas de programação quadrática, cujas soluções serão utilizadas para formar a direção de busca para a solução do problema geral. A seguir apresentar-se-á de forma geral o método, podendo-se encontrar maiores detalhes em (Fletcher, 1980; Powell, 1983). Para a solução do problema de otimização foi utilizada a função no MATLAB[®] `fmincon` que utiliza os métodos aqui discutidos (*Optimization Toolbox User's Guide*, 2008).

Dado um problema geral de otimização restrita, em que a função objetivo $f(x)$ não necessariamente é quadrática ou que se $f(x)$ for quadrática SQP recai em um problema de programação quadrática único, como descrito em (2.33), a idéia principal do método SQP consiste na formulação de subproblemas de programação quadrática baseados na aproximação quadrática da função lagrangeana:

$$L(x, \lambda) = f(x) + \sum_{i=1}^m \lambda_i \cdot g_i(x) \quad (2.35)$$

onde assume-se, para fins de simplificação, que as restrições de limite das variáveis são restrições de desigualdade.

Os subproblemas que se deseja minimizar com a finalidade de se conhecer as direções de busca são dados por:

$$\begin{aligned} \min_{d \in R^n} \quad & \frac{1}{2} d^T H_k d + \nabla f(x_k)^T d \\ \text{s.a.} \quad & \nabla g_i(x_k)^T d + g_i(x_k) = 0, \quad i = 1, \dots, m_e \\ & \nabla g_i(x_k)^T d + g_i(x_k) \leq 0, \quad i = m_e + 1, \dots, m \end{aligned} \quad (2.36)$$

Estes subproblemas podem ser resolvidos usando-se qualquer algoritmo de programação quadrática conhecido (Luenberger, 2003). A solução é então usada para formar o novo processo iterativo:

$$x_{k+1} = x_k + \alpha_k d_k$$

O tamanho do passo, α_k , é determinado por um método de busca linear apropriado tal que um decrescimento suficiente da função de custo seja obtido. A matriz H_k é uma aproximação definida positiva da matriz Hessiana da função lagrangeana (2.35). H_k pode ser atualizada por qualquer um dos métodos quasi-Newton, sendo o método BFGS o algoritmo mais utilizado para tal atualização,

utilizando-se a seguinte adaptação:

$$H_{k+1} = H_k + \frac{q_k q_k^T}{q_k^T s_k} - \frac{H_k^T s_k^T s_k H_k}{s_k^T H_k s_k}$$

onde

$$\begin{aligned} s_k &= x_{k+1} - x_k \\ q_k &= \nabla f(x_{k+1}) + \sum_{i=1}^n \lambda_i \cdot \nabla g_i(x_{k+1}) - \left(\nabla f(x_k) + \sum_{i=1}^n \lambda_i \cdot \nabla g_i(x_k) \right) \end{aligned}$$

com $\lambda_i (i = 1, \dots, m)$ sendo uma estimativa dos multiplicadores de Lagrange. Maiores informações sobre o processo podem ser encontradas em (Powell, 1978b).

Para a otimização de modelos que utilizam funções de Laguerre ou GOBF com funções internas, o algoritmo de Levenberg-Marquardt (Levenberg, 1944; Marquardt, 1963; Nocedal, 1999; Nelles, 2001) se apresenta como alternativa na otimização dos parâmetros, uma vez que, estes modelos não envolvem restrições que não sejam de limitantes superiores e inferiores para os parâmetros. A base deste algoritmo é uma aproximação quadrática da função de custo na vizinhança de seus argumentos. O algoritmo aproxima a matriz Hessiana da função de custo com relação aos parâmetros usando somente informações das derivadas de primeira ordem correspondentes, que são obtidas através do cálculo dos gradientes. Denotando $F(x)$ como o Jacobiano da função de custo $f(x)$, o algoritmo atualiza a variável de otimização x de acordo com a seguinte equação (Nelles, 2001):

$$x_k = x_{k-1} - \left[F_{k-1}^T F_{k-1} + \sigma_{k-1} I \right]^{-1} F_{k-1}^T f(x_{k-1}) \quad (2.37)$$

onde σ_{k-1} representa um fator de regularização e I é uma matriz identidade.

O algoritmo de Levenberg-Marquardt é um método de pseudo-segunda ordem, pois ele utiliza apenas avaliações da função e seu gradiente. Este algoritmo é globalmente convergente, ou seja, a partir de qualquer ponto inicial x_0 , o algoritmo converge para um ponto x_{otimo} satisfazendo as condições de otimalidade necessárias para um minimizador local, ou seja, a função de custo apresenta gradiente nulo. Contudo, não é possível garantir que x_{otimo} seja um minimizador global de $f(x)$, a menos que $f(x)$ seja uma função convexa (Nocedal, 1999). O método Levenberg-Marquardt pode ser visto como uma combinação dos métodos de gradiente de maior descida e de Gauss-Newton (Nelles, 2001): quando a busca está longe da solução ótima, o algoritmo se comporta como o método do gradiente (lento, mas com convergência garantida), enquanto que próximo à solução ótima, o algoritmo se comporta como o método de Gauss-Newton, convergindo assim mais rápido. Para a solução do problema de otimização com o algoritmo de Levenberg-Marquardt, utilizou-se a função no

MATLAB[®] `lsqcurvefit` na qual o algoritmo está implementado (*Optimization Toolbox User's Guide*, 2008).

2.4.3 Estimação e Otimização dos Parâmetros dos modelos por BFO's

Nesta seção apresenta-se a formalização para os problemas de otimização que envolvem a otimização de modelos lineares formados por bases de funções ortonormais. Deseja-se que o modelo identificado apresente o menor erro possível entre a saída do sistema e a saída estimada no mesmo instante de tempo k . Isto é, deseja-se minimizar uma função de custo do erro quadrático da saída, como mostra a equação (2.38).

$$\min_{\theta} J = \frac{1}{2} \sum_{k=0}^N (\hat{y}(k) - y(k))^2 \quad (2.38)$$

com N sendo o número de amostras de entrada e saída do sistema e $\hat{y}(k)$ a saída estimada definida por:

$$\hat{y}(k) = C_f \cdot \Psi(k)$$

sendo $C_f = [g_1 \ g_2 \ \dots \ g_{n_f}]$ os termos da expansão da série das funções ortonormais e $\Psi(k) = [\psi_1(k) \ \dots \ \psi_{n_f}(k)]^T$ a saída das funções ortonormais, onde $\psi_{n_f}(k)$ corresponde à saída da função de mais elevada ordem (n_f -ésima ordem).

A função de custo (2.38) é minimizada com relação ao vetor de parâmetros $\theta = [\Upsilon^T \ C_f^T]^T$, onde Υ representa o polo da função de Laguerre, o par de polos complexos conjugados para as funções de Kautz (ou parâmetros b e c) ou os parâmetros $\gamma_i, i = 1, \dots, n_b$ para as funções internas nos modelos GOBF; C_f representa os coeficientes da expansão da série da BFO conforme apresentado anteriormente. Desta maneira pode-se definir o problema de otimização como um problema de minimização de J com relação aos parâmetros θ do modelo BFO. Para se resolver o problema apresentado, torna-se necessário calcular o gradiente da função de custo com relação aos parâmetros que se deseja otimizar, ou seja, $\nabla_{\theta} J$, onde tal gradiente pode ser escrito como:

$$\nabla_{\theta} J = [\nabla_{\Upsilon}^T J \ \nabla_{C_f}^T J]^T \quad (2.39)$$

Analisando-se individualmente os termos que compõem a equação (2.39), o gradiente da função de custo com relação aos parâmetros C_f da expansão é dado por:

$$\nabla_{C_f} J = \sum_{k=0}^N (y(k) - \hat{y}(k)) \nabla_{C_f} \hat{y}(k)$$

$$= \sum_{k=0}^N (y(k) - \hat{y}(k)) \Psi(k) \quad (2.40)$$

ou, mais especificamente, com relação a cada termo da expansão tem-se:

$$\frac{\partial J}{\partial g_i} = \sum_{k=0}^N (\hat{y}(k) - y(k)) \psi_i(k)$$

Já o gradiente de J com relação aos parâmetros Υ é definido por:

$$\begin{aligned} \nabla_{\Upsilon} J &= \sum_{k=0}^N (y(k) - \hat{y}(k)) \nabla_{\Upsilon} \hat{y}(k) \\ &= \sum_{k=0}^N (y(k) - \hat{y}(k)) \sum_{i=1}^{n_f} g_i \nabla_{\Upsilon} \psi_i(k) \\ &= \sum_{k=0}^N \sum_{i=1}^{n_f} (y(k) - \hat{y}(k)) g_i \nabla_{\Upsilon} \psi_i(k) \end{aligned} \quad (2.41)$$

onde $\nabla_{\Upsilon} \psi_i(k)$ é igual a $\partial \psi_i(k) / \partial p$ para os modelos com funções de Laguerre e $\nabla_{\Upsilon} \psi_i(k) = [\partial \psi_i(k) / \partial b \ \partial \psi_i(k) / \partial c]^T$ para Kautz com b e c sendo os parâmetros das funções de Kautz como definido na equação (2.7). Já para as funções internas nos modelos GOBF, $\nabla_{\Upsilon} \psi_i(k) = [\partial \psi_i(k) / \partial \gamma_1 \dots \partial \psi_i(k) / \partial \gamma_{n_b}]^T$, com os parâmetros $\gamma_j, j = 1, \dots, n_b$, diretamente ligados aos polos do modelo GOBF.

Neste trabalho utilizar-se-á o cálculo analítico destes gradientes da função de custo com relação aos parâmetros Υ e C_f , fornecendo assim o valor exato do gradiente para a otimização do(s) polo(s) e dos coeficientes da expansão das funções ortonormais através de métodos de otimização não-linear. A seguir são apresentadas as soluções para o cálculo do gradiente citado para as funções de Laguerre, Kautz e GOBF com funções internas.

a - Gradiente para funções de Laguerre

Nesta seção são apresentados os cálculos analíticos dos gradientes das funções de Laguerre com relação ao polo p das funções. Para o cálculo do gradiente utilizar-se-á a representação em espaço de estados do modelo de Laguerre como descrito na seção 2.2. Como se deseja conhecer somente as saídas dos filtros, ou seja, os estados nos instantes k para derivá-las com relação ao polo, é necessário se utilizar somente a parte dinâmica do modelo (2.5). Sendo assim, o modelo em espaço de estados

para as funções de Laguerre será escrito como:

$$\Psi(k+1) = A_f \Psi(k) + B_f u(k) \quad (2.42)$$

onde A_f e B_f são as mesmas definidas na seção 2.2 para as funções de Laguerre.

Desenvolvendo-se o modelo (2.42) é possível verificar que a solução para $\Psi(k)$ será dada por:

$$\Psi(k) = A_f^k \Psi(0) + \sum_{i=0}^{k-1} A_f^i B_f u(k-1-i) \quad (2.43)$$

O gradiente de $\Psi(k)$ com relação ao polo p pode então ser calculado através da derivada da equação (2.43) com relação ao polo, como apresentado na equação (2.44):

$$\frac{\partial \Psi(k)}{\partial p} = \frac{\partial A_f^k}{\partial p} \Psi(0) + \sum_{i=0}^{k-1} \left(\frac{\partial A_f^i}{\partial p} B_f + A_f^i \frac{\partial B_f}{\partial p} \right) u(k-1-i) \quad (2.44)$$

sendo $\frac{\partial A_f^k}{\partial p}$ dado por:

$$\frac{\partial A_f^k}{\partial p} = \sum_{j=1}^k A_f^{j-1} \frac{\partial A_f}{\partial p} A_f^{k-j} \quad (2.45)$$

Das equações (2.44) e (2.45) torna-se necessário calcular as derivadas de A_f e B_f com relação ao polo p . O valor destas derivadas é dado por:

$$\frac{\partial A_f}{\partial p} = \begin{bmatrix} dap_{11} & 0 & 0 & \cdots & 0 \\ dap_{21} & dap_{22} & 0 & \cdots & 0 \\ dap_{31} & dap_{32} & dap_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ dap_{n_f 1} & dap_{n_f 2} & dap_{n_f 3} & \cdots & dap_{n_f n_f} \end{bmatrix}$$

onde i é igual à linha e j é igual à coluna do elemento dap_{ij} , sendo este dado por:

- Nos casos em que $j > i$:

$$. \, dap_{ij} = 0;$$

- Com $i = j$

$$. \, dap_{ij} = 1;$$

- Com $j < i$

$$.dap_{ij} = (-1)(i-j-1)(-p)^{(i-j-2)}(1-p^2) + (-p)^{(i-j-1)}(-2p), \text{ tal que } i \leq n \text{ e } (i-j-1) \geq 0;$$

Já para a matriz B_f , tem-se que a sua derivada com relação ao polo p é dada por:

$$\frac{\partial B_f}{\partial p} = \begin{bmatrix} dbp_{11} & dbp_{21} & dbp_{31} & \cdots & dbp_{n_f 1} \end{bmatrix}^T$$

com seus elementos sendo:

$$dbp_{i1} = \frac{-p}{\sqrt{1-p^2}}(-p)^{(i-1)} + \sqrt{1-p^2}(i-1)(-1)(-p)^{(i-2)};$$

para $i = 1, \dots, n_f$, e $0 < p < 1$.

Desta maneira calcula-se analiticamente o gradiente das saídas das funções de Laguerre com relação ao parâmetro p utilizando-se a técnica de retropropagação através do tempo (conforme apresentado na seção 2.2), como apresentada na Seção 2.4.1.

b - Gradiente para funções de Kautz

Nesta seção serão desenvolvidos os cálculos analíticos dos gradientes das funções de Kautz com relação ao par de polos complexos conjugados β e β^* das funções, ou, mais especificamente, com relação aos parâmetros b e c obtidos diretamente da combinação dos polos, como descritos em (2.7). Para o cálculo do gradiente utilizar-se-á o modelo das funções de Kautz no espaço de estados, como descrito em (2.9), sendo que as saídas dos filtros nos instantes k ($\Psi(k)$) são descritas como:

$$\begin{aligned} \mathcal{X}(k+1) &= A_K \mathcal{X}(k) + B_K \mathcal{U}(k) \\ \Psi(k) &= C_{K_m} \mathcal{X}(k) \end{aligned} \tag{2.46}$$

onde:

$$\mathcal{X}(k) = \begin{bmatrix} \psi_1(k-1) & \psi_1(k) & \cdots & \psi_{n_f-1}(k-1) & \psi_{n_f-1}(k) & \psi_{n_f}(k-1) & \psi_{n_f}(k) \end{bmatrix}^T$$

$$\mathcal{U}(k) = \begin{bmatrix} u(k-1) & u(k) \end{bmatrix}^T$$

onde A_K e B_K são as mesmas definidas na seção 2.2 para as funções de Kautz e C_{K_m} é definida por:

$$C_{K_m} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

Desenvolvendo-se a representação em espaço de estados (2.46), pode-se verificar que a expressão para $\Psi(k)$ será dado por:

$$\Psi(k) = C_{K_m} A_K^k \mathcal{X}(0) + \sum_{i=0}^{k-1} C_{K_m} A_K^i B_K \mathcal{U}(k-1-i) \quad (2.47)$$

A derivada parcial de $\Psi(k)$ com relação ao parâmetro b é dada pela equação (2.48):

$$\frac{\partial \Psi(k)}{\partial b} = C_{K_m} \frac{\partial A_K^k}{\partial b} \mathcal{X}(0) + C_{K_m} \sum_{i=0}^{k-1} \left(\frac{\partial A_K^i}{\partial b} B_K + A_K^i \frac{\partial B_K}{\partial b} \right) \mathcal{U}(k-1-i) \quad (2.48)$$

sendo $\frac{\partial A_K^k}{\partial b}$ dado por:

$$\frac{\partial A_K^k}{\partial b} = \sum_{j=1}^k A_K^{j-1} \frac{\partial A_K}{\partial b} A_K^{k-j} \quad (2.49)$$

A derivada da matriz A_K com relação ao parâmetro b é dada por:

$$\frac{\partial A_K}{\partial b} = \begin{bmatrix} 0 & dab_{1,2} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ dab_{2,1} & dab_{2,2} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & dab_{3,4} & \cdots & 0 & 0 & 0 & 0 \\ 0 & dab_{4,2} & 0 & dab_{4,4} & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & dab_{2n_f-1,2n_f} \\ 0 & 0 & 0 & 0 & \cdots & 0 & dab_{2n_f,2n_f-2} & 0 & dab_{2n_f,2n_f} \end{bmatrix}$$

onde os elementos $dab_{i,j}$ pertencentes à matriz $\frac{\partial A_K}{\partial b}$, com i indicando a linha e j a coluna do respectivo elemento, são definidos por:

- Para i ímpar:
 - . Todos os termos iguais a 0;
- Para $i = 2$:
 - . $dab_{i,i} = -(c-1)$,
 - . demais termos iguais a 0;
- Para i múltiplo de 4:
 - . $dab_{i,i} = 1$,

- . $dab_{i,i-2} = \frac{-b}{\sqrt{1-b^2}},$
- . demais termos iguais a 0;
- Para outras linhas i pares os termos serão:
 - . $dab_{i,i} = -(c-1),$
 - . $dab_{i,i-4j} = (-c)^{j-1}[(c-1) + (-c)(-(c-1))], \quad \text{para } j = 0, 1, 2, \dots \text{ tal que } 4j < i,$
 - . demais termos iguais a 0;

Para a matriz B_K tem-se que a sua derivada com relação ao parâmetro b é dada por:

$$\frac{\partial B_K}{\partial b} = \sqrt{1-c^2} \begin{bmatrix} 0 & 0 \\ dbb_{2,1} & dbb_{2,2} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ dbb_{6,1} & dbb_{6,2} \\ \vdots & \vdots \\ 0 & 0 \\ dbb_{2(n_f-1),1} & dbb_{2(n_f-1),2} \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

com seus elementos dados por:

- Para $i = 2 + 4j$, com $j = 0, 1, 2, \dots$ tal que $j \leq n_f/2$
 - . $dbb_{i,1} = 0,$
 - . $dbb_{i,2} = (-1)(-c)^j;$
- Para demais i 's:
 - . $dbb_{i,1} = dbb_{i,2} = 0,$

Para os cálculos das derivadas com relação ao parâmetro c , derivando a equação (2.47) com relação a tal parâmetro, obtém-se a seguinte equação:

$$\frac{\partial \Psi(k)}{\partial c} = C_{K_m} \frac{\partial A_K^k}{\partial c} \mathcal{X}(0) + C_{K_m} \sum_{i=0}^{k-1} \left(\frac{\partial A_K^i}{\partial c} B_K + A_K^i \frac{\partial B_K}{\partial c} \right) \mathcal{U}(k-1-i) \quad (2.50)$$

sendo $\frac{\partial A_K^k}{\partial c}$ é dado por:

$$\frac{\partial A_K^k}{\partial c} = \sum_{j=1}^k A_K^{j-1} \frac{\partial A_K}{\partial c} A_K^{k-j} \quad (2.51)$$

Novamente, como para os cálculos com relação ao parâmetro b , através de (2.50) e (2.51) é necessário o cálculo das derivadas de A_K e B_K com relação ao parâmetro c . O valor da derivada da matriz A_K com relação a c é dada por:

$$\frac{\partial A_K}{\partial c} = \begin{bmatrix} 0 & dac_{1,2} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ dac_{2,1} & dac_{2,2} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & dac_{3,4} & \cdots & 0 & 0 & 0 & 0 \\ 0 & dac_{4,2} & 0 & dac_{4,4} & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & dac_{2n_f-1,2n_f} \\ 0 & 0 & 0 & 0 & \cdots & 0 & dac_{2n,2n-2} & 0 & dac_{2n_f,2n_f} \end{bmatrix}$$

onde os elementos da matriz $\frac{\partial A_K}{\partial c}$ são dados por $dac_{i,j}$ com i indicando a linha e c a coluna do respectivo elemento. Considerando-se i correspondente a i -ésima linha da matriz $\frac{\partial A_K}{\partial c}$, tem-se:

- Para i ímpar:
 - . Todos os termos iguais a 0;
- Para $i = 2$:
 - . $dac_{i,i} = -b$,
 - . $dac_{i,i-1} = 1$,
 - . demais termos iguais a 0;
- Para i múltiplo de 4:
 - . Todos os termos iguais a 0;
- Para outras linhas i pares os termos serão:
 - . $dac_{i,i} = -b$,
 - . $dac_{i,i-1} = 1$,
 - . $dac_{i,i-4j} = (j-1)(-1)(-c)^{j-2}(bc^2 - b) + 2bc(-c)^{j-1}$
 - . $dac_{i,i-(4j+1)} = (j-1)(-1)(-c)^{j-2}(-c^2 + 1) + (-c)^{j-1}(-2c)$, para $j = 0, 1, 2, \dots$ tal que $4j < i$,

. demais termos iguais a 0;

Para a matriz B_K tem-se que a sua derivada com relação ao parâmetro c é dada por:

$$\frac{\partial B_K}{\partial c} = \frac{-c}{\sqrt{1-c^2}} \begin{bmatrix} 0 & 0 \\ dbc_{2,1} & dbc_{2,2} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ dbc_{6,1} & dbc_{6,2} \\ \vdots & \vdots \\ 0 & 0 \\ dbc_{2(n_f-1),1} & dbc_{2(n_f-1),2} \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

com $-1 < c < 0$, i sendo a i -ésima linha da matriz e com os elementos da matriz dados por:

- Para $i = 2 + 4j$, com $j = 0, 1, 2, \dots$ tal que $j \leq n/2$

$$. \text{ } dbc_{i,1} = (-c)^j + \left(\frac{1-c^2}{-c}\right)(-1)(j)(-c)^{j-1},$$

$$. \text{ } dbc_{i,2} = (-b)(-c)^j + \left(\frac{1-c^2}{-c}\right)(-1)(j)(-c)^{j-1};$$

- Para demais i 's:

$$. \text{ } dbc_{i,1} = dbc_{i,2} = 0,$$

Desta maneira pode-se determinar os gradientes das saídas das funções Kautz com relação aos parâmetros b e c , sendo estes relacionados com os polos de tais funções como citado anteriormente. Os parâmetros b e c apresentam restrições quanto aos valores que podem assumir, estando b restrito ao intervalo $(0, 1)$ e c restrito ao intervalo $(-1, 0)$. Além disso, os parâmetros b e c apresentam uma restrição adicional quanto aos valores que podem assumir, já que tais parâmetros devem ser convertidos em par de polos complexos conjugados β e $\bar{\beta}$. As equações em (2.8) podem ser reescritas, como apresentada abaixo, em função da parte real e imaginária, α e ϑ , do par de polos complexos conjugados:

$$b = \frac{2\alpha}{1-c} \quad \text{e} \quad c = -\alpha^2 - \vartheta^2$$

com os parâmetros α e ϑ dados por:

$$\alpha = \frac{b(1-c)}{2} \quad \text{e} \quad \vartheta = \sqrt{-c - \alpha^2} \quad (2.52)$$

Como α e ϑ devem ser valores reais, da equação (2.52) vem a seguinte restrição para a solução do problema:

$$\begin{aligned} -c - \alpha^2 &> 0, \\ \alpha^2 &< -c \end{aligned} \quad (2.53)$$

substituindo-se em (2.53) o valor de α encontrado na equação (2.52) tem-se que:

$$\begin{aligned} \left(\frac{b(1-c)}{2} \right)^2 &< -c, \\ b^2(1-c)^2 &< -4c, \\ b^2(1-c)^2 + 4c &< 0. \end{aligned} \quad (2.54)$$

A equação (2.54) é, portanto, uma restrição adicional aos valores que b e c podem assumir, restringindo assim o espaço de busca do valor ótimo destes parâmetros, como indica a figura 2.7.

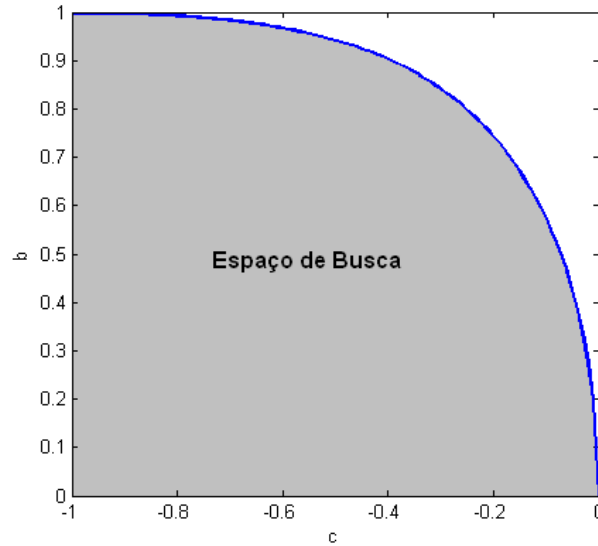


Fig. 2.7: Espaço de busca dos parâmetros b e c

Em da Rosa et al. (2008a) é proposta uma abordagem com o cálculo da derivada de forma individual para cada filtro com relação aos parâmetros b e c , enquanto que neste trabalho o cálculo é feito em batelada, isto é, o cálculo dos gradientes é realizado de maneira simultânea para todos os filtros. Esta nova abordagem se apresentou superior, pois permite uma redução no esforço computacional comparada àquela proposta em (da Rosa et al., 2008a). Em um exemplo em que os sinais de entrada e saída apresentam 1024 pontos e com a utilização de 8 filtros de Kautz houve uma redução de 25% no tempo de processamento para o cálculo das derivadas dos filtros de Kautz com a aplicação da abordagem deste trabalho em comparação ao proposto em (da Rosa et al., 2008a).

c - Gradiente para funções generalizadas

Para as funções generalizadas formadas através de funções de Takenaka-Malmquist a solução dos gradientes utilizados para se otimizar os parâmetros de tais funções foi apresentada por da Rosa (2009). Contudo, como descrito na seção 2.3, estas funções apresentam alguns inconvenientes que são contornados com a utilização de funções internas na representação através de modelos GOBF. Assim como para as funções de Laguerre e Kautz, nesta seção são desenvolvidos os cálculos analíticos dos gradientes das funções generalizadas com relação aos seus parâmetros, com a diferença de que estes parâmetros não são os polos, mas sim os coeficientes reais $\gamma_i, i = 1, \dots, n_b$.

Para o cálculo do gradiente utilizar-se-á o modelo em espaço de estados descrito em (2.29), e reapresentado na equação (2.55):

$$\Psi(k+1) = A_f \Psi(k) + B_f u(k) \quad (2.55)$$

onde $\Psi(k)$ é formado por $n_b n$ saídas das funções da base. Desenvolvendo-se o modelo (2.55) é possível verificar que a solução para $\Psi(k)$ será dada por:

$$\Psi(k) = A_f^k \Psi(0) + \sum_{i=0}^{k-1} A_f^i B_f u(k-1-i) \quad (2.56)$$

O gradiente de $\Psi(k)$ com relação aos parâmetro γ_l é dado pela equação (2.57):

$$\frac{\partial \Psi(k)}{\partial \gamma_l} = \frac{\partial A_f^k}{\partial \gamma_l} \Psi(0) + \sum_{i=0}^{k-1} \left(\frac{\partial A_f^i}{\partial \gamma_l} B_f + A_f^i \frac{\partial B_f}{\partial \gamma_l} \right) u(k-1-i) \quad (2.57)$$

sendo $\frac{\partial A_f^k}{\partial \gamma_l}$ dada por:

$$\frac{\partial A_f^k}{\partial \gamma_l} = \sum_{j=1}^k A_f^{j-1} \frac{\partial A_f}{\partial \gamma_l} A_f^{k-j} \quad (2.58)$$

A derivada da matriz A_f com relação ao parâmetro γ_l , presente na equação (2.58), depende do parâmetro γ_l a que se está relacionada a derivada, uma vez que os cálculos variam para cada um destes. Inicialmente serão apresentadas as derivadas das submatrizes A_b e B_b que compõem respectivamente as matrizes A_f e B_f , cujos resultados serão utilizados nos cálculos de $\frac{\partial A_f}{\partial \gamma_l}$ e $\frac{\partial B_f}{\partial \gamma_l}$. Para a matriz A_b

tem-se que sua derivada com relação ao parâmetro γ_l é dada por:

$$\frac{\partial A_b}{\partial \gamma_l} = \begin{bmatrix} da_{1,1} & da_{1,2} & da_{1,3} & \cdots & da_{1,n_b} \\ da_{2,1} & da_{2,2} & da_{2,3} & \cdots & da_{2,n_b} \\ da_{3,1} & da_{3,2} & da_{3,3} & \cdots & da_{3,n_b} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ da_{n_b,1} & da_{n_b,2} & da_{n_b,3} & \cdots & da_{n_b,n_b} \end{bmatrix} \quad (2.59)$$

onde os termos $da_{i,j}$, com i sendo a linha a que o elemento pertence e j a coluna, são definidos por:

- Para $i = 1$:

- Para $l = 1$:

$$\begin{aligned} da_{1,j} &= 1, & \text{para } j = 1 \\ da_{1,j} &= -\gamma_1(1 - \gamma_1)^{-\frac{1}{2}} \gamma_j \prod_{k=2}^{j-1} \sqrt{1 - \gamma_k^2}, & \text{para } j = 2, \dots, n_b. \end{aligned}$$

- Para $l = 2, \dots, n_b$:

$$\begin{aligned} da_{1,j} &= 0, & \text{para } j = 1 \\ da_{1,j} &= -\gamma_l(1 - \gamma_l)^{-\frac{1}{2}} \gamma_j \prod_{\substack{k=2 \\ k \neq l}}^{j-1} \sqrt{1 - \gamma_k^2}, & \text{para } l < j \text{ e } l \neq 1. \\ da_{1,j} &= \prod_{k=1}^{j-1} \sqrt{1 - \gamma_k^2}, & \text{para } j = l. \\ da_{1,j} &= 0, & \text{para } l > j. \end{aligned}$$

- Para $i = 2, \dots, n_b$:

- Para $l < i - 1$

$$da_{i,j} = 0;$$

- Para $l = i - 1$

$$\begin{aligned} da_{i,j} &= -\gamma_l(1 - \gamma_j^2)^{-\frac{1}{2}}, & \text{para } j = i - 1; \\ da_{i,j} &= -\gamma_j, & \text{para } j = i; \end{aligned}$$

$$\begin{aligned}
da_{i,j} &= -\gamma_j \prod_{k=i}^{j-1} \sqrt{1 - \gamma_k^2}, & \text{para } j = i + 1, \dots, n_b; \\
da_{i,j} &= 0, & \text{para demais elementos.}
\end{aligned}$$

- Para $l = i$ e $l \neq n_b$

$$\begin{aligned}
da_{i,j} &= 0, & \text{para } j = i - 1; \\
da_{i,j} &= -\gamma_{j-1}, & \text{para } j = i; \\
da_{i,j} &= \gamma_{i-1} \gamma_j \gamma_l (1 - \gamma_l^2)^{-\frac{1}{2}} \prod_{\substack{k=i \\ k \neq l}}^{j-1} \sqrt{1 - \gamma_k^2}, & \text{para } j = i + 1, \dots, n_b; \\
da_{i,j} &= 0, & \text{para demais elementos.}
\end{aligned}$$

- Para $l = n_b$

$$\begin{aligned}
da_{i,j} &= 0, & \text{para } j = i - 1; \\
da_{i,j} &= -\gamma_{j-1}, & \text{para } j = i \text{ e } i = n_b; \\
da_{i,j} &= -\gamma_{i-1} \prod_{k=i}^{j-1} \sqrt{1 - \gamma_k^2}, & \text{para } j = n_b \text{ e } i \neq n_b; \\
da_{i,j} &= 0, & \text{para demais elementos.}
\end{aligned}$$

A matriz B_b do modelo (2.26) é definida por:

$$\frac{\partial B_b}{\partial \gamma_l} = \begin{bmatrix} db_{1,1} \\ db_{2,1} \\ db_{3,1} \\ \vdots \\ db_{n_b,1} \end{bmatrix} \quad (2.60)$$

onde seus parâmetros são dados, de acordo com a variável de derivação γ_l , por:

- Para $i = 1$:

- Para qualquer $l = 1, \dots, n_b$:

$$db_{1,1} = \gamma_l(1 - \gamma_l^2)^{-\frac{1}{2}} \prod_{\substack{k=1 \\ k \neq l}}^{n_k} \sqrt{1 - \gamma_k^2}.$$

onde $n_k = n_b$ se $l \neq n_b$, e $n_k = n_b - 1$ se $l = n_b$.

• Para $i = 2, \dots, n_b$:

- Para $l < i - 1$

$$db_{i,1} = 0$$

- Para $l = i - 1$

$$db_{i,1} = - \prod_{k=i}^{n_b} \sqrt{1 - \gamma_k^2}$$

- Para $l \geq i$

$$db_{i,1} = -\gamma_{i-1}\gamma_l(1 - \gamma_l^2)^{-\frac{1}{2}} \prod_{\substack{k=i \\ k \neq l}}^{n_k} \sqrt{1 - \gamma_k^2}$$

onde $n_k = n_b$ se $l \neq n_b$, e $n_k = n_b - 1$ se $l = n_b$.

A matriz A_I também compõe a matriz A_f , presente no modelo GOBF apresentado em (2.29), e assim deve ter calculada suas derivadas com relação aos parâmetros γ_l para poder compor o cálculo de $\frac{\partial A_f}{\partial \gamma_l}$ que aparece na equação (2.58). A matriz $\frac{\partial A_I}{\partial \gamma_l}$ é dada por:

$$\frac{\partial A_I}{\partial \gamma_l} = \begin{bmatrix} dai_{1,1} & dai_{1,2} & dai_{1,3} & \cdots & dai_{1,n_b} \\ dai_{2,1} & dai_{2,2} & dai_{2,3} & \cdots & dai_{2,n_b} \\ dai_{3,1} & dai_{3,2} & dai_{3,3} & \cdots & dai_{3,n_b} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ dai_{n_b,1} & dai_{n_b,2} & dai_{n_b,3} & \cdots & dai_{n_b,n_b} \end{bmatrix} \quad (2.61)$$

com seus elementos $dai_{i,j}$ dados de acordo o γ_l a que se refere a derivada, onde o índice i indica a linha da matriz e o índice j a coluna:

• Para $i = 1, \dots, n_b$ e $j \neq n_b$, o elemento $dai_{i,j}$ será:

$$dai_{i,j} = 0;$$

- Se $j = n_b$, então os elementos $dai_{i,j}$ serão:

Com $i = 1$:

- Para $l = 1, \dots, n_b - 1$

$$dai_{i,j} = -\gamma_l(1 - \gamma_l^2)^{-\frac{1}{2}}(1 - \gamma_{n_b}^2) \prod_{\substack{k=i \\ k \neq l}}^{n_b-1} \sqrt{1 - \gamma_k^2};$$

- Para $l = n_b$

$$dai_{i,j} = -2\gamma_l \prod_{k=i}^{n_b-1} \sqrt{1 - \gamma_k^2};$$

Com $i = 2, \dots, n_b$:

- Para $l = i - 1$

$$dai_{i,j} = -(1 - \gamma_{n_b}^2) \prod_{k=i}^{n_b-1} \sqrt{1 - \gamma_k^2}.$$

- Para $l = i, \dots, n_b - 1$

$$dai_{i,j} = (-\gamma_l)(1 - \gamma_l^2)^{-\frac{1}{2}}(1 - \gamma_{n_b}^2)(-\gamma_{i-1}) \prod_{\substack{k=i \\ k \neq l}}^{n_b-1} \sqrt{1 - \gamma_k^2}.$$

- Para $l = n_b$

$$dai_{i,j} = (-2\gamma_{n_b})(-\gamma_{i-1}) \prod_{l=i+1}^{n_b-1} \sqrt{1 - \gamma_l^2}.$$

- Para demais elementos

$$dai_{i,j} = 0$$

De posse dos cálculos de $\frac{\partial A_b}{\partial \gamma_l}$ e $\frac{\partial A_l}{\partial \gamma_l}$ é possível realizar o cálculo da matriz $\frac{\partial A_f}{\partial \gamma_l}$ e, em sequência, e calcular $\frac{\partial \Psi(k)}{\partial \gamma_l}$. Para se obter a matriz $\frac{\partial A_f}{\partial \gamma_l}$ deve se realizar os seguintes cálculos (derivada da equação (2.30)):

- Se $l \neq n_b$

$$\frac{\partial A_f}{\partial \gamma_l} = \begin{bmatrix} \frac{\partial A_b}{\partial \gamma_l} & 0 & 0 & \dots & 0 \\ \frac{\partial A_I}{\partial \gamma_l} & \frac{\partial A_b}{\partial \gamma_l} & 0 & \dots & 0 \\ (-\gamma_{n_b}) \frac{\partial A_I}{\partial \gamma_l} & \frac{\partial A_I}{\partial \gamma_l} & \frac{\partial A_b}{\partial \gamma_l} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (-\gamma_{n_b})^{n-2} \frac{\partial A_I}{\partial \gamma_l} & (-\gamma_{n_b})^{n-3} \frac{\partial A_I}{\partial \gamma_l} & (-\gamma_{n_b})^{n-4} \frac{\partial A_I}{\partial \gamma_l} & \dots & \frac{\partial A_b}{\partial \gamma_l} \end{bmatrix} \quad (2.62)$$

onde n é o número de blocos G_b .

- Se $l = n_b$

$$\frac{\partial A_f}{\partial \gamma_l} = \begin{bmatrix} \frac{\partial A_b}{\partial \gamma_l} & 0 & 0 & \dots & 0 \\ \frac{\partial A_I}{\partial \gamma_l} & \frac{\partial A_b}{\partial \gamma_l} & 0 & \dots & 0 \\ ta_{3,1} & \frac{\partial A_I}{\partial \gamma_l} & \frac{\partial A_b}{\partial \gamma_l} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ ta_{n,1} & ta_{n,2} & ta_{n,3} & \dots & \frac{\partial A_b}{\partial \gamma_l} \end{bmatrix} \quad (2.63)$$

onde:

$$ta_{i,j} = (-1)(i - (j + 1))(-\gamma_{n_b})^{i-(j+2)} A_I + (-\gamma_{n_b})^{i-(j+1)} \frac{\partial A_I}{\partial \gamma_l}$$

para valores de $i > 2$ e $j < i - 1$.

Da mesma maneira, uma vez que foram calculadas as derivadas $\frac{\partial B_b}{\partial \gamma_l}$ é possível se estabelecer a derivada da matriz B_f com relação ao parâmetro γ_l (derivada da equação (2.31)):

- Se $l \neq n_b$:

$$\frac{\partial B_f}{\partial \gamma_l} = \begin{bmatrix} \frac{\partial B_b}{\partial \gamma_l} \\ (-\gamma_{n_b}) \frac{\partial B_b}{\partial \gamma_l} \\ (-\gamma_{n_b})^2 \frac{\partial B_b}{\partial \gamma_l} \\ \vdots \\ (-\gamma_{n_b})^{n-1} \frac{\partial B_b}{\partial \gamma_l} \end{bmatrix} \quad (2.64)$$

- Se $l = n_b$:

$$\frac{\partial B_f}{\partial \gamma_l} = \begin{bmatrix} \frac{\partial B_b}{\partial \gamma_l} \\ (-1)B_b + (-\gamma_{n_b}) \frac{\partial B_b}{\partial \gamma_l} \\ (-1)(2)(-\gamma_{n_b})B_b + (-\gamma_{n_b})^2 \frac{\partial B_b}{\partial \gamma_l} \\ \vdots \\ (-1)(n_b - 1)(-\gamma_{n_b})^{n-2}B_b + (-\gamma_{n_b})^{n-1} \frac{\partial B_b}{\partial \gamma_l} \end{bmatrix} \quad (2.65)$$

Calculando-se esses valores é possível se obter o gradiente de $\Psi(k)$ com relação aos parâmetros γ_l apresentado na equação (2.57) e utilizar o cálculo exato do gradiente nos métodos de otimização não-linear já apresentados.

2.5 Algoritmo

O método proposto para a otimização dos modelos GOBF pode então ser resumido pelos seguintes passos:

1. Atribua valores iniciais de parâmetros θ^0 (normalmente de maneira aleatória ou baseado em algum conhecimento prévio) e calcule por MMQ os valores dos coeficientes de C_f .
2. Calcule o gradiente $\nabla_{\theta} J = [\nabla_{\Upsilon}^T J \quad \nabla_{C_f}^T J]^T$ utilizando as equações (2.40) e (2.41). Para o cálculo de $\nabla_{\Upsilon} \psi_i(k)$ em (2.40), ou de forma genérica, a derivada de $\Psi(k)$ com relação aos parâmetros $\gamma_l, l = 1, \dots, n_b$, utiliza-se as equações apresentadas de (2.57) até (2.65).
3. Use o valor de $\nabla_{\theta} J$ no algoritmo de otimização de Levenberg-Marquardt para atualizar o valor de parâmetros θ . Nota: inclua o intervalo de factibilidade dos parâmetros do modelo GOBF como restrições dentro do problema de otimização para garantir a factibilidade dos polos e dos modelos otimizados ($-1 < \gamma_l < 1$ com $l = 1, \dots, n_b$);
4. Volte ao passo 2 até que um critério de parada seja atingido.

As restrições citadas no passo 3 do algoritmo acima são necessárias para evitar que o processo de otimização produza polos instáveis ou que o modelo GOBF perca sua capacidade de representação. Para a otimização dos modelos de Laguerre e Kautz, algoritmo semelhante deve ser seguido utilizando-se as equações apropriadas.

2.6 Resultados

Esta seção tem como objetivo ilustrar a aplicação dos processos de identificação de sistemas lineares por meio de modelos BFO's que empregam funções de Laguerre, Kautz ou GOBF com funções internas. O processo de identificação emprega os métodos propostos de otimização dos parâmetros do modelo utilizando os cálculos de gradientes analíticos apresentados anteriormente.

Considere inicialmente um sistema de terceira ordem com um polo real de multiplicidade 3. O sistema em questão é dado pela seguinte função de transferência em z :

$$G(z) = \frac{0,01741}{z^3 - 2,222z^2 + 1,646z - 0,4066} \quad (2.66)$$

com 3 polos em $z = 0,7408$

Para a identificação do sistema utilizam-se amostras de entrada e saída do sistema obtidas através de simulação, como mostra a figura 2.8, na qual a entrada $u(k)$ é um sinal aleatório com distribuição uniforme entre -1 e 1.

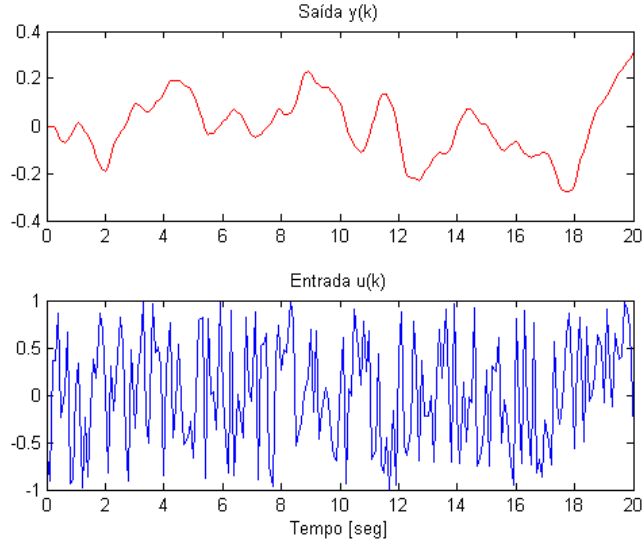


Fig. 2.8: Sinais utilizados na identificação do Modelo de GOBF com 3 polos

Como é comum na modelagem de sistemas dinâmicos a partir de dados de entrada e saída, partiu-se do pressuposto de que as características dinâmicas do sistema sob análise eram desconhecidas. Desta maneira, fez-se a identificação de modelos onde se variou o número de polos e de funções. Conforme descrito na seção 2.4.3, para a execução do processo de otimização, realizou-se o cálculo dos gradientes com relação aos parâmetros dos modelos a serem otimizados a partir das amostras de entrada e saída do sistema. Para este exemplo, identificou-se o sistema por meio de modelos com funções de Laguerre e GOBF com funções internas. A tabela 2.1 apresenta os resultados comparativos entre os modelos de Laguerre otimizados com o erro quadrático médio (EQM) gerado pelos modelos estimados (para dados de validação).

Tab. 2.1: Resultados dos modelos de Laguerre para o primeiro exemplo.

N. de Funções	Param.	Polo Inicial	Polo Final	EQM
1	2	0,5000	0,9222	0,0095
2	3	0,5000	0,8090	$2,2269 \cdot 10^{-4}$
3	4	0,5000	0,7408	$2,0869 \cdot 10^{-26}$
4	5	0,5000	0,7407	$2,4647 \cdot 10^{-16}$

Em seguida, identificou-se o sistema utilizando-se a abordagem de modelos GOBF com funções

internas. Da mesma maneira como se procedeu com os modelos de Laguerre, considera-se que não há nenhum conhecimento a priori sobre as características dinâmicas do sistema. A tabela 2.2 apresenta os resultados para os modelos de GOBF com funções internas. Na tabela 2.2 a coluna NF indica o número de blocos G_b e a coluna NP indica o número de polos do modelo. A coluna Param. indica o número de parâmetros otimizados no modelo GOBF (n_b polos + coeficientes da expansão da GOBF).

Tab. 2.2: Resultados dos modelos GOBF para o primeiro exemplo.

NF	NP	Param.	Polos Finais	EQM
1	1	2	0,9222	0,0095
2	1	3	0,8090	$2,2269 \cdot 10^{-4}$
3	1	4	0,7408	$2,0791 \cdot 10^{-26}$
4	1	5	0,7407	$2,4948 \cdot 10^{-16}$
1	2	4	$0,8486 \pm 0,0806i$	$7,3566 \cdot 10^{-5}$
2	2	6	0,7335 e 0,7427	$7,9837 \cdot 10^{-15}$
3	2	8	0,7045 e 0,7528	$1,4079 \cdot 10^{-15}$
1	3	6	$0,7478 \pm 0,0109i$ 0,7267	$2,3724 \cdot 10^{-12}$
2	3	9	0,7433 e 0,7397 0,7267	$1,0009 \cdot 10^{-14}$
3	3	12	0,7408 $0,1299 \pm 0,8143i$	$9,0340 \cdot 10^{-31}$

Através da análise das tabelas 2.1 e 2.2 pode-se verificar que o método proposto foi eficiente na identificação do sistema sob análise. Os modelos de Laguerre e GOBF com um polo apresentaram resultados semelhantes, sendo que os modelos de Laguerre com 3 funções e GOBF com um polo e 3 blocos G_b apresentaram os melhores resultados quando considerados o número de parâmetros dos modelos e sua precisão. Inicializou-se os modelos GOBF com um polo em 0,5, os modelos com dois polos em $0,5 \pm 0,5i$ e os modelos com três polos foram inicializados nos polos $0,1325 \pm 0,8141i$ e 0,7350 com os parâmetros γ iniciais sendo $[0,5 - 0,5 \ 0,5]$.

A figura 2.9 apresenta o decaimento da função de custo J em relação a cada iteração do algoritmo de otimização para o modelo GOBF com um polo e 3 blocos G_b . A figura 2.10 mostra a comparação, para dados de validação, entre a saída estimada do modelo com polos iniciais em 0,5 e finais em 0,7408 para o mesmo modelo GOBF da figura 2.9.

Verifica-se através da análise da tabela 2.2 que os polos otimizados, nos modelos com 2 ou 3 polos, não são exatamente os polos do sistema como aconteceu quando se restringiu os polos nos modelos com somente um polo. Contudo, os polos se encontram na vizinhança dos polos exatos do sistema e apresentam uma boa precisão. Modelos com reduzido número de funções ou com número de polos inferiores ao valor necessário para se representar adequadamente o sistema podem apresentar

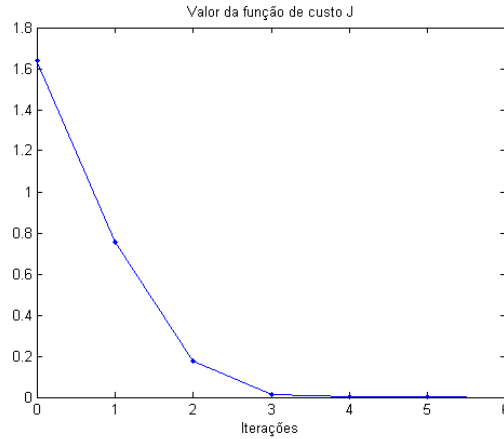


Fig. 2.9: Valor da função de custo

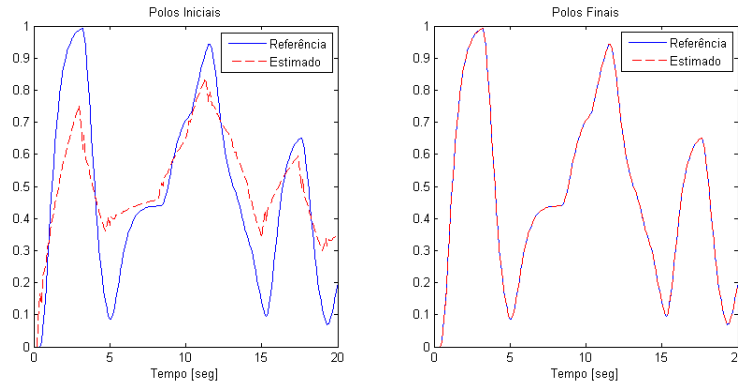


Fig. 2.10: Respostas do modelo com 1 polo com multiplicidade 3

soluções menos precisas por não serem capazes de representar adequadamente a dinâmica do sistema analisado.

A seguir, apresenta-se a simulação e a identificação de um sistema com 4 polos (dois pares de polos complexos conjugados). A função de transferência em z do sistema é dada pela equação (2.67):

$$G(z) = \frac{0,02167z^3 + 0,1226z^2 + 0,06434z + 0,003167}{z^4 - 1,483z^3 + 0,9539z^2 - 0,2995z + 0,04076} \quad (2.67)$$

cujos polos estão localizados em $z = 0,3635 \pm 0,2536i$ com multiplicidade 2.

Para o processo de identificação gerou-se através de simulação uma massa de dados de entrada e saída do sistema proposto conforme apresentado na figura 2.11. Como entrada foi utilizado um sinal aleatório de distribuição uniforme entre -1 e 1.

Assim como no exemplo anterior, considerou-se que não havia nenhum conhecimento a priori do sistema a ser modelado. A tabela 2.3 mostra os resultados dos modelos GOBF onde se variou o

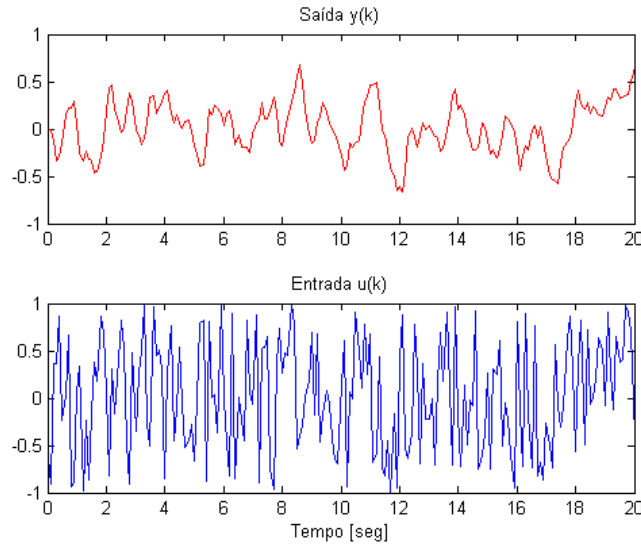


Fig. 2.11: Sinais utilizados na identificação do Modelo de GOBF com 4 polos

número de polos e o número de blocos G_b . Na tabela 2.3 a coluna NF indica o número de blocos G_b e a coluna NP indica o número de polos do modelo. O EQM é referente a dados utilizados para validação do modelo.

O valor inicial do polo para os modelos com somente um parâmetro γ é de 0,5. Para o sistema com dois parâmetros γ , inicializou-se os parâmetros em $[0,5 \ -0,5]$. Os parâmetros γ do sistema com 3 polos foram inicializados em $[0,5 \ 0,5 \ 0,5]$ e o sistema com 4 polos teve seus parâmetros inicializados em $[0,5 \ 0,5 \ 0,5 \ 0,5]$. Como pode-se verificar pela análise da tabela 2.3, o modelo com 2 polos e dois blocos G_b apresentou os melhores resultados quando considerada a precisão e o número de parâmetros dos modelos. Outros modelos, como os modelos com 4 polos, apresentaram uma excelente precisão com polos convergindo para polos na vizinhança dos polos do sistema. A figura 2.12 mostra a resposta do modelo com 2 polos e dois blocos G_b utilizando-se dados de validação.

Para finalizar a aplicação de modelos GOBF à modelagem de sistemas lineares foi simulado e identificado um sistema com 3 polos, sendo um par de polos complexos e um polo real, dado pela seguinte função de transferência em z :

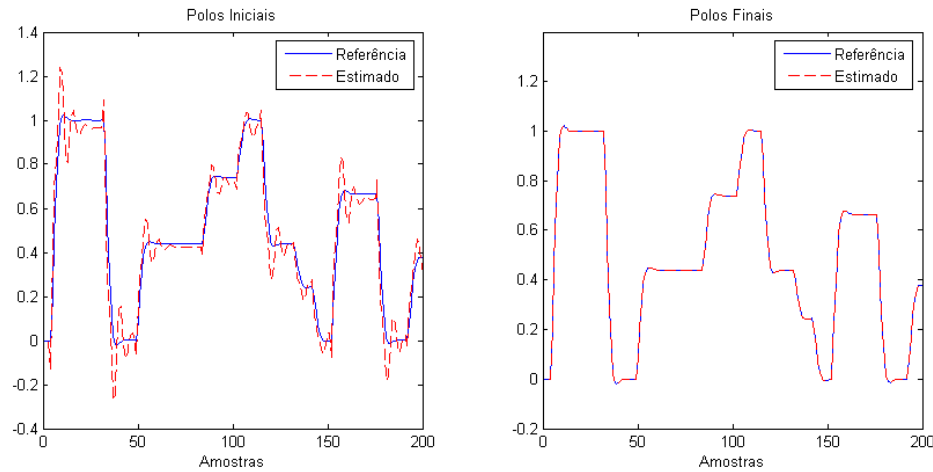
$$G(z) = \frac{0,07287z^2 + 0,1587z + 0,02187}{z^3 - 1,191z^2 + 0,5352z - 0,09072} \quad (2.68)$$

cujos polos estão localizados em $z = 0,3708 \pm 0,2537i$ e $z = 0,4493$.

Assim como para os modelos anteriores utilizou-se um sinal aleatório com distribuição uniforme dentro do intervalo $[-1, 1]$ como sinal de entrada para o sistema sob estudo e, conjuntamente com a resposta do sistema, foi implementado o método de identificação proposto. A tabela 2.4 apresenta

Tab. 2.3: Resultados dos modelos GOBF para o segundo exemplo.

NF	NP	Param.	Polos Finais	EQM
1	1	2	0,7462	0,0037
2	1	3	0,4671	$4,7335 \cdot 10^{-4}$
3	1	4	0,6044	$2,1150 \cdot 10^{-4}$
4	1	5	0,4345	$1,2888 \cdot 10^{-5}$
1	2	4	$0,5977 \pm 0,2549i$	$9,0818 \cdot 10^{-5}$
2	2	6	$0,3635 \pm 0,2536i$	$1,1613 \cdot 10^{-32}$
3	2	8	$0,3633 \pm 0,2533i$	$2,5884 \cdot 10^{-16}$
1	3	6	$0,4537 \pm 0,3364i$ 0,4332	$2,6546 \cdot 10^{-7}$
2	3	9	$0,3635 \pm 0,2536i$ -0,3718	$1,1613 \cdot 10^{-32}$
3	3	12	$0,3638 \pm 0,2532i$ -0,3999	$4,1518 \cdot 10^{-17}$
1	4	8	$0,4047 \pm 0,2815i$ $0,3924 \pm 0,2700i$	$4,9191 \cdot 10^{-17}$
2	4	12	$0,3627 \pm 0,3422i$ $0,3987 \pm 0,2760i$	$1,9628 \cdot 10^{-28}$

Fig. 2.12: Respostas do modelos com 2 polos e dois blocos G_b

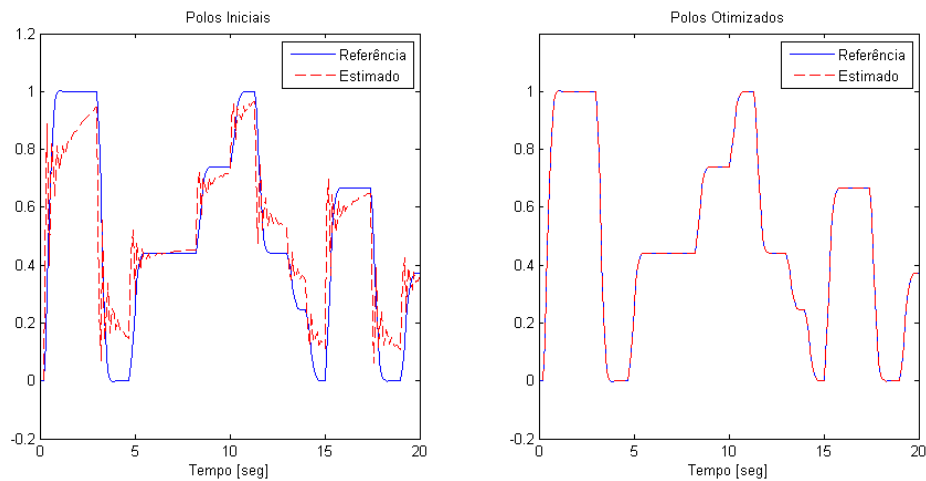
alguns resultados de modelos GOBF com funções internas onde se variou o número de polos e blocos G_b que os compõem.

Inicializou-se os parâmetros γ_l dos modelos presentes na tabela 2.4 em 0,5; $[0,5 \ -0,5]$ e $[0,5 \ 0,5 \ 0,5]$ respectivamente para os modelos de 1, 2 e 3 polos. Através da análise da tabela 2.4 verifica-se que o modelo com 3 polos e um bloco G_b apresenta o melhor resultado dos modelos

Tab. 2.4: Resultados dos modelos GOBF para o terceiro exemplo.

NF	NP	Param.	Polos Finais	EQM
1	1	2	0,8424	0,0141
2	1	3	0,5079	$8,3448 \cdot 10^{-4}$
3	1	4	0,3374	$3,0469 \cdot 10^{-5}$
1	2	4	$0,5750 \pm 0,2644i$	$2,0012 \cdot 10^{-4}$
2	2	6	$0,3297 \pm 0,1571i$	$5,4917 \cdot 10^{-8}$
3	2	8	$0,2141 \pm 0,1479i$	$5,0686 \cdot 10^{-9}$
1	3	6	$0,3710 \pm 0,2538i$ 0,4490	$1,0605 \cdot 10^{-19}$
2	3	9	$0,3636 \pm 0,2621i$ 0,4320	$1,8292 \cdot 10^{-13}$

analisados. Este modelo apresenta uma ótima precisão e teve seus polos convergindo para os polos exatos do sistema. A figura 2.13 apresenta o comparativo entre a saída do modelo de 3 parâmetros γ e 1 bloco G_b com polos iniciais propostos e o modelo com os parâmetros otimizados (dados de validação).

Fig. 2.13: Respostas do modelos com 3 polos distintos e 3 γ 's

2.7 Conclusão

Neste capítulo apresentaram-se métodos de modelagem matemática de sistemas dinâmicos lineares utilizando-se bases de funções ortonormais, com destaque aos modelos GOBF com funções internas de estrutura *ladder*, descritos sob uma representação em espaço de estados obtida através de uma nova metodologia desenvolvida neste trabalho. A modelagem utilizando GOBF com funções internas apresenta vantagens frente às demais representações com BFO, porque possibilita que sistemas dinâmicos com múltiplos modos dominantes com polos reais e/ou complexos sejam representados por modelos parametrizados somente por coeficientes reais.

A modelagem de sistemas dinâmicos por modelos BFO necessita de uma escolha adequada dos polos para que se obtenham modelos com uma boa precisão. Contudo, esta escolha não é simples, tendo em vista que nem sempre se conhece os polos do sistema modelado, ou mesmo a sua característica dinâmica. Para solucionar este problema apresentou-se nesta tese um processo de otimização para os parâmetros dos modelos. Este processo se mostra vantajoso em diversos aspectos, já que nenhuma ou quase nenhuma informação *a priori* das características do sistema se torna necessária para a sua identificação. A otimização dos modelos se dá através de técnicas bem conhecidas de otimização não-linear que utilizam a informação de gradiente em seus cálculos. Neste capítulo apresentaram-se os cálculos necessários para a obtenção analítica e em batelada dos gradientes das funções ortonormais. Esta abordagem apresenta, conforme observado em análises preliminares, um menor esforço computacional exigido para o cálculo dos gradientes quando comparado a outros métodos existentes na literatura. A representação em espaço de estados para os modelos GOBF com funções internas possibilitou o desenvolvimento do cálculo analítico do gradiente das funções GOBF com relação aos seus parâmetros e permitiu obter a direção de busca para a otimização destes parâmetros. Também se apresentou uma nova realização em espaço de estados para as funções de Kautz que contribui para facilitar o cálculo dos gradientes destas funções. Outra contribuição deste capítulo refere-se à imposição de restrições nos valores que os parâmetros da função de Kautz podem assumir, para resultar em soluções factíveis.

Resultados de simulações ilustraram a eficácia dos métodos propostos para a otimização dos modelos com base de funções ortonormais na modelagem de sistemas dinâmicos lineares. Os resultados mostraram que, sem nenhum conhecimento prévio sobre a dinâmica dos sistemas identificados, foi possível obter modelos com precisão e seus polos convergem ou se aproximam bastante dos polos exatos do sistema sob análise.

Capítulo 3

Modelagem de Sistemas Não-lineares por Base de Funções Ortonormais

3.1 Introdução

Quando a utilização de modelos lineares não é suficiente para representar com qualidade o sistema dinâmico sob estudo, torna-se necessária a utilização de modelos não-lineares. Este capítulo apresenta a modelagem de sistemas dinâmicos não-lineares através da utilização de abordagens com representações em tempo discreto e monovariável, ou *SISO - Single Input Single Output*, extensivas a representações multivariáveis. Os modelos não-lineares analisados são descritos por:

$$\hat{y}(k) = \mathcal{H}(\Phi(k)) \quad (3.1)$$

onde $\hat{y}(k)$ é a saída estimada ou prevista no instante k e $\Phi(k)$ é o vetor de regressores, que nos modelos deste capítulo são as saídas dos filtros da BFO. O operador $\mathcal{H}$ é um operador estático não-linear genérico, de modo que, dado que $\mathcal{H}$ é estático, a dinâmica se concentra de maneira exclusiva no vetor $\Phi(k)$. Uma ampla gama de sistemas não-lineares reais admite modelos como o descrito pela expressão (3.1) (Eykhof, 1974).

Nesta tese tem-se especial interesse nos modelos não-lineares do tipo apresentado na equação (3.1) e que empreguem BFO, como os modelos *fuzzy* Takagi-Sugeno (TS) com modelos BFO nos consequentes das regras (TS-BFO) e os modelos de Volterra com BFO na composição dos *kernels* dos modelos (Volterra-BFO) (Campello et al., 2007; Campello e Oliveira, 2007).

Usualmente, modelos *fuzzy* de sistemas dinâmicos são construídos com base em modelos de topologia do tipo NARX¹, em que a saída do sistema em um dado instante de tempo é representada através

¹*Nonlinear AutoRegressive with eXogenous inputs*

da combinação de sinais passados da entrada e da saída (Babuška, 1998; Nelles, 2001). Modelos *fuzzy* apresentam a característica de descrever sistemas de forma aproximada, porém com precisão arbitrária, ou seja, como aproximadores universais de funções numa dada região compacta, apresentando a capacidade de mapear de forma eficiente relações entre entradas e saídas de sistemas em geral (Wang e Mendel, 1992; Kosko, 1994; Wang, 1998; Campello e Amaral, 2006). Modelos *fuzzy* do tipo Takagi-Sugeno com funções ortonormais nos consequentes das regras tem sido empregados com sucesso na modelagem de sistemas dinâmicos não-lineares e controladores baseados em modelos para a mesma classe de sistemas (Oliveira et al., 1999; Campello e Amaral, 2002; Machado, Campello e Amaral, 2007). Esta abordagem de modelos *fuzzy* Takagi-Sugeno com funções ortonormais tem se mostrado tão eficiente na modelagem de sistemas não-lineares quanto outras abordagens que empregam tais funções, como modelos de Volterra (da Rosa et al., 2005) ou redes neurais artificiais, com algumas vantagens como possibilitar uma interpretabilidade dos modelos, o que não acontece com modelos de Volterra ou com redes neurais.

Em trabalhos como (Oliveira et al., 1999; Campello e Amaral, 2002) utiliza-se algum conhecimento prévio da dinâmica do sistema modelado para determinar o(s) polo(s) das funções ortonormais que irão compor os consequentes das regras do modelo *fuzzy* TS-BFO. Em (Medeiros et al., 2006) é proposta uma otimização dos modelos *fuzzy* TS-BFO através de técnicas de algoritmos genéticos (Bäck, Fogel e Michalewicz, 2000) exigindo-se assim pouco ou quase nenhum conhecimento a priori da dinâmica do sistema sob análise. Em (Machado, 2007) apresenta-se uma metodologia para se obter as regras e conjuntos *fuzzy* do modelo TS-BFO de maneira automática através de técnicas de agrupamento *fuzzy* (Babuška, Van der Veen e Kaymak, 2002) em conjunto com critérios de avaliação dos grupos obtidos, sendo necessário um conhecimento a priori da dinâmica do sistema a ser modelado.

Neste capítulo propõe-se uma nova abordagem, baseada em técnicas de otimização não-linear, para obterem-se modelos *fuzzy* TS-BFO sem nenhum conhecimento a priori das características dinâmicas do sistema a ser modelado. Em (Medeiros et al., 2006), os modelos *fuzzy* TS-BFO são obtidos em uma abordagem em que se otimizam os parâmetros dos modelos da população utilizados no algoritmo genético (cada modelo da população é avaliado de acordo com uma função de *fitness* que leva em consideração os dados de entrada e saída - abordagem *top-down*) enquanto que neste trabalho o modelo é obtido exclusivamente, em uma abordagem *bottom-up*, a partir dos dados de entrada e saída amostrados do sistema sob análise. Primeiramente, como proposto em (Machado, 2007), deve-se obter um modelo inicial TS-BFO através de técnicas de agrupamento *fuzzy* e critérios de avaliação dos agrupamentos. Em seguida, como contribuição deste trabalho, se necessário, deve-se eliminar os parâmetros redundantes dos antecedentes do modelo utilizando-se técnicas de fusão de grupos e de funções de pertinência que apresentam semelhanças estruturais de maneira a reduzir a complexidade dos modelos obtidos. Também se realiza a análise e eventual exclusão de conjuntos *fuzzy* similares

ao universo de discurso das variáveis das premissas. A seguir, através de técnicas de otimização não-linear, otimizam-se os polos das funções ortonormais, os coeficientes da expansão da série das BFO's presentes nos consequentes das regras e os parâmetros das funções de pertinência presentes nos antecedentes das regras de maneira a se obter modelos mais precisos a partir das amostras de entrada e saída do sistema estudado. Para a otimização destes parâmetros utilizando-se técnicas de otimização não-linear, assim como para os modelos BFO's lineares, há a necessidade do cálculo dos gradientes da saída estimada com relação aos parâmetros de interesse indicando a direção de busca destes. A presente tese apresenta os cálculos analíticos necessários para se obter os valores destes gradientes.

Aborda-se nesta tese outra metodologia para a modelagem de sistemas não-lineares, os modelos de Volterra (Nelles, 2001; Aguirre, 2007). Estes modelos apresentam uma variante em tempo discreto e podem ser vistos como uma generalização de E -ésima ordem do modelo linear tipo FIR, com o *kernel* do modelo sendo uma generalização de ϱ -ésima dimensão da função de resposta ao impulso de um modelo FIR. Boyd e Chua (1985) mostraram que o modelo de Volterra com resposta ao impulso pode aproximar com precisão arbitrária qualquer sistema que admita descrição entrada-saída através de um operador $\mathcal{H}$ contínua e com *fading memory*², desde que a entrada u seja limitada inferior e superiormente.

A modelagem de sistemas dinâmicos não-lineares através de modelos de Volterra envolve diretamente o problema de estimação dos *kernels* do modelo, que tem sido investigado por décadas (Eykhof, 1974; Billings, 1980; Schetzen, 1980; Rugh, 1981; Doyle III, Pearson e Ogunnaike, 2002). A principal dificuldade na solução deste problema é que os *kernels* são, em princípio, funções não parametrizadas, e só podem ser mensuradas se suas contribuições individuais puderem ser separadas da resposta total do sistema (Schetzen, 1980). Uma abordagem utilizada para se obter o *kernel* do modelo de Volterra consiste em tratar cada um dos parâmetros do *kernel* como um parâmetro independente a ser estimado (Doyle III et al., 2002). Nesse caso, o modelo de Volterra de resposta ao impulso torna-se linear nesses parâmetros e algoritmos de estimação clássicos podem ser aplicados. Essa abordagem, no entanto, usualmente torna o modelo sobre-parametrizado. Logo, é importante reduzir sua complexidade antes do procedimento de estimação para se melhorar o seu condicionamento numérico e reduzir a variância do estimador. Uma estratégia de particular interesse para lidar com esse problema, originalmente sugerida por Wiener (1958), é o desenvolvimento dos *kernels* de Volterra através de base de funções ortonormais, conforme discutido em (Campello, 2002; Campello et al., 2003; Campello et al., 2007; da Rosa, 2009). Nesta tese apresentam-se modelos de Volterra que utilizam GOBF com funções internas em sua estrutura, incorporando as características apresentadas para os modelos GOBF na seção 2.3. Apresenta-se também uma metodologia para se obter, através

²Sistemas com fading memory são aqueles para os quais a resposta a uma dada excitação de entrada se anula ou decresce de forma rápida o suficiente para que não seja significativa após um dado período finito de tempo, o que basicamente remete a sistemas estáveis com estado estacionário único.

de processos de otimização não-linear, modelos de Volterra-GOBF com funções internas a partir de medidas de entrada e saída do sistema sob análise. Para implementação do processo de otimização dos modelos Volterra-GOBF, desenvolve-se o cálculo analítico dos gradientes da saída dos modelos, de *kernels* simétricos ou não simétricos, com relação aos parâmetros de interesse.

3.2 Modelos *Fuzzy* TS-BFO

O conceito de conjuntos *fuzzy*, ou teoria de conjuntos nebulosos, foi originalmente proposto por Zadeh (1965) como uma generalização da idéia de um conjunto ordinário ou *crisp* (não *fuzzy*). Um conjunto *fuzzy* pode ser visto como um conjunto em que o valor do predicado pode pertencer ao intervalo $I = [0, 1]$, ao contrário dos conjuntos ordinários em que o valor pertence ao conjunto $\{0, 1\}$, ou seja, para conjuntos ordinários um dado elemento $\in$ ou $\notin$ a um dado conjunto enquanto que nos conjuntos *fuzzy* um elemento pertence a um conjunto com valor de verdade entre 0 e 1. Desta maneira, os conjuntos *fuzzy* estão sujeitos a uma lógica com abordagem multivalorada. Os modelos *fuzzy* são aproximadores universais de funções numa região compacta, apresentando assim a capacidade de mapear de forma eficiente relações entre entradas e saídas de sistemas em geral (Wang e Mendel, 1992; Kosko, 1994; Wang, 1998; Campello e Amaral, 2006).

Um dos modelos baseados em regras mais usado para aproximação de sistemas dinâmicos não-lineares foi introduzido por Takagi e Sugeno (1985) e, na literatura referente à modelagem *fuzzy*, é citado como modelo Takagi-Sugeno (TS). Tais modelos consistem essencialmente de uma interpolação de múltiplos modelos locais (usualmente lineares ou afins), o que é favorável ao projeto de controladores e de aproximadores universais de funções. A forma mais comum de um modelo TS baseado em regras é dada por:

$$\underbrace{\text{Se } x_1 \text{ é } A_{i,1} \text{ e } \dots \text{ e } x_j \text{ é } A_{i,j}}_{\text{antecedente}} \text{ então } \underbrace{\check{y}_i = f_i(x)}_{\text{consequente}} \quad (3.2)$$

A primeira parte da regra, chamada de antecedente, é definida como uma proposição (ou uma combinação de proposições) *fuzzy* “ $x_1 \text{ é } A_{i,1} \text{ e } \dots \text{ e } x_j \text{ é } A_{i,j}$ ” onde x_j é uma variável real exata (não *fuzzy*) e $A_{i,j}$ é um conjunto *fuzzy* definido pela função de pertinência

$$A_{i,j}(x) : \mathbb{R} \rightarrow [0, 1]$$

onde o indexador $i = 1, \dots, K$ denota a i -ésima regra, sendo K o número de regras na base e j o indexador de variável $j = 1, \dots, \varsigma$, com ς sendo o número de variáveis nos antecedentes das regras. O grau de desempenho (grau de verdade) $A_{i,j}(x)$ dos antecedentes para um dado valor do vetor de

regressão x é avaliado como o grau de ativação da função de pertinência no conjunto $A_{i,j}$. O antecedente é usualmente descrito pela combinação dos conjuntos *fuzzy* em cada subespaço univariável definido para cada componente de x . Os antecedentes são combinados usando conjunções, isto é, conectivos lógicos **E**. O grau de ativação da i -ésima regra, $\mu_i(x)$, calculado como:

$$\mu_i(x) = A_{i,1}(x_1) \wedge A_{i,2}(x_2) \wedge \dots \wedge A_{i,\varsigma}(x_\varsigma) \quad (3.3)$$

onde o operador $(\wedge)$ é usado para conjunção. Para o operador produto obtém-se:

$$\mu_i(x) = A_{i,1}(x_1) \cdot A_{i,2}(x_2) \cdot \dots \cdot A_{i,\varsigma}(x_\varsigma) \quad (3.4)$$

Note que $\mu_i(x)$ em (3.3) e (3.4) é uma função de pertinência *fuzzy* com entrada multivariável constituída pela intersecção dos conjuntos univariáveis *fuzzy* das entradas x . O formato e a característica do conjunto *fuzzy* resultante depende do operador de intersecção utilizado. A Fig. 3.1 mostra um exemplo de intersecção realizado com uso do operador mínimo.

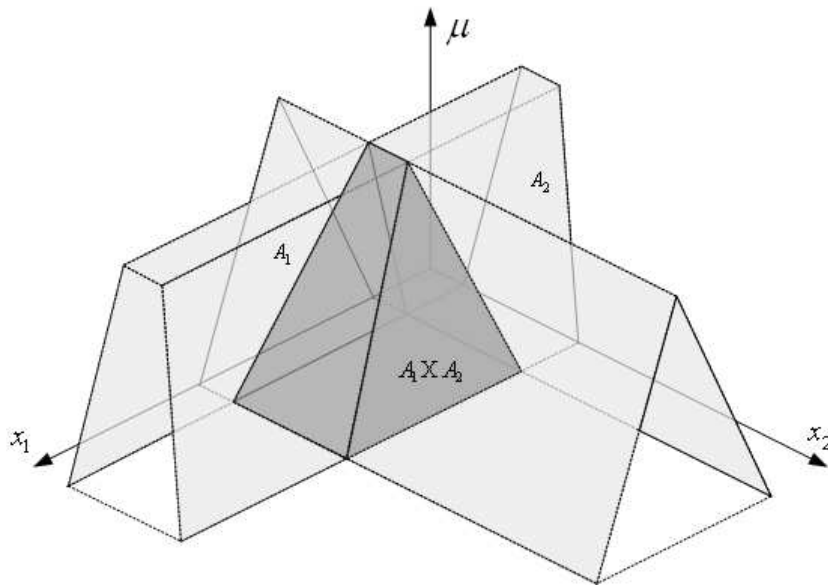


Fig. 3.1: Intersecção de dois conjuntos *fuzzy* A_1 e A_2 (em escuro).

A segunda parte da regra (3.2), chamada de consequente, é, em geral, uma função das variáveis da premissa. As regras normalmente se diferenciam pelos conjuntos *fuzzy* dos antecedentes e também pelas funções $f_i(x)$ dos consequentes. No entanto, na prática, $f_i(x)$ são funções parametrizadas, cuja estrutura não muda independentemente da regra, variando somente os parâmetros. Uma estrutura

simples e prática muito utilizada é a “função afim”, que resulta na seguinte regra:

$$\text{Se } x \text{ é } A_i \text{ então } \tilde{y}_i = a_i^T x + b_i \quad (3.5)$$

onde a_i é um vetor de parâmetros e b_i é o *offset*. Este modelo é denominado TS afim. Quando $b_i = 0$, $i = 1, \dots, K$, o modelo é chamado TS linear. Os antecedentes definem em que região do espaço de regressão o hiperplano do consequente do modelo é válido. Neste sentido, os modelos TS afins definem hiperplanos que podem aproximar a “função-alvo dentro da região de interpolação”. Um caso especial de (3.5) é aquele formado por funções constantes $\tilde{y}_i = b_i$, $i = 1, \dots, K$, sendo este modelo conhecido como modelo singleton (Yager e Filev, 1994).

Para a descrição anterior do modelo utilizaram-se funções lineares e afins estáticas. Porém, o modelo *fuzzy* TS pode ter qualquer tipo de função no consequente, como funções não-lineares ou relações dinâmicas, como equações diferenciais, equações a diferenças, etc. Neste trabalho as funções nos consequentes serão modelos dinâmicos BFO, já apresentados na seção 2.2. A utilização de modelos com bases de funções ortonormais nos consequentes das regras apresenta todas as vantagens inerentes à modelagem de sistemas dinâmicos com BFO, como apresentado na seção 2.2. Para casos em que se utiliza uma mesma base BFO para todos os modelos locais, faz-se necessário somente o cálculo da parte estática da formulação para cada modelo uma vez que a parte dinâmica é a mesma para todos os modelos (Oliveira et al., 1999; Campello, 2002; Campello e Amaral, 2002; Campello, Meleiro e Amaral, 2004). Sendo assim, as formas de A_f e B_f da representação em espaço de estados das funções ortonormais, equação (2.5), são únicas e dependem exclusivamente da base de funções ortonormais adotada para a implementação do modelo. Os modelos *fuzzy* TS são compostos de um conjunto de K regras *fuzzy* no seguinte formato:

$$\begin{aligned} \text{Se } \psi_1(k) \text{ é } L_{i,1} \text{ e } \dots \text{ e } \psi_{n_f}(k) \text{ é } L_{i,n_f} \\ \text{Então } \tilde{y}_i(k) = C_{f_i}^T \Psi(k) \end{aligned} \quad (3.6)$$

onde as entradas do modelo TS são dadas pelas saídas do conjunto de filtros ortonormais, isto é, $x_1 = \psi_1(k), \dots, x_{n_f} = \psi_{n_f}(k)$; $L_{i,j}$ é o conjunto *fuzzy* da i -ésima regra associada com a j -ésima variável da premissa e $C_{f_i}^T$ é a matriz de coeficientes da expansão em espaço de estados da BFO do i -ésimo modelo local $\tilde{y}_i(k)$, como apresentado na seção 2.2. A saída global $\hat{y}(k)$ é obtida usando-se a interpolação clássica TS da saída dos modelos locais (Babuška, 1998; Nelles, 2001):

$$\hat{y}(k) = \frac{\sum_{i=1}^K \mu_i(\Psi(k)) \tilde{y}_i(k)}{\sum_{i=1}^K \mu_i(\Psi(k))} \quad (3.7)$$

onde $\mu_i(\Psi(k))$ é o valor de ativação da i -ésima regra como definido em (3.4), $\tilde{y}_i(k)$ o valor da saída

de cada modelo local BFO e K o número de regras ou modelos locais.

Pode-se notar que, a princípio, todos os n_f estados BFO são considerados na premissa das regras. Frequentemente, no entanto, somente um subconjunto de estados é suficiente para determinar as regiões ativas do sistema. Considerando-se somente um subconjunto de $\vars \leq n_f$ estados como variáveis na premissa, há uma simplificação na estrutura das regras, o que contribui para a parcimônia do modelo obtido (Medeiros, 2006; Machado, 2007).

Para a determinação dos modelos *fuzzy* TS-BFO é necessário definir as regras do modelo com os seus respectivos antecedentes e consequentes. No que diz respeito aos consequentes das regras, uma propriedade já conhecida e muito utilizada de modelos *fuzzy* TS é que a saída de um modelo TS é linear com relação aos parâmetros das funções do consequente de suas regras se essas funções são lineares ou afins com relação aos seus argumentos, i.e., $\tilde{y}_i(k) = g_{i,0} + g_{i,1}\psi_1(k) + \dots + g_{i,n_f}\psi_{n_f}(k)$. As funções $\tilde{y}_i(k)$ podem ser determinadas a partir de medidas de entrada e saída do sistema pela estimação numérica dos parâmetros $g_{i,j}$, utilizando-se, e.g., um algoritmo de mínimos quadrados. As principais dificuldades na identificação de sistemas dinâmicos por modelos *fuzzy* TS-BFO estão em se determinar o número K de regras *fuzzy*, ou de modelos locais, com seus respectivos antecedentes e consequentes, a sintonia dos conjuntos *fuzzy* $L_{i,j}$ nas premissas das regras e a otimização do(s) polo(s) da BFO. Uma metodologia sistemática para abordar esse problema é apresentada na sequência. Para a determinação do número de regras *fuzzy* e dos conjuntos que compõem os antecedentes das regras utiliza-se o algoritmo de agrupamento de Gustafson-Kessel juntamente com critérios de avaliação de agrupamentos em que é possível se determinar qual modelo *fuzzy* apresenta melhor resultado para representar os dados de entrada e saída. Posteriormente são aplicados no modelo obtido critérios de avaliação que analisam se existem regras ou funções de pertinência redundantes ou desnecessárias e realizam as operações necessárias para a simplificações dos modelos, como a eliminação de regras e funções de pertinência ou a fusão de funções quando necessário. Para concluir, é realizada a otimização do modelo TS-BFO com o cálculo, através de algoritmos de otimização não-linear, dos polos das funções ortonormais, os coeficientes da expansão da série das BFO's presentes nos consequentes das regras e os parâmetros das funções de pertinência presentes nos antecedentes das regras de maneira a se obter modelos mais precisos. Todo este processo de obtenção do modelo *fuzzy* TS-BFO é descrito nas seções a seguir.

3.3 Projeto de modelos *Fuzzy* TS-BFO

3.3.1 Premissa das Regras

Um conjunto de regras *fuzzy* pode ser obtido através de informações fornecidas por um especialista ou então diretamente extraído de dados obtidos do processo, como, por exemplo, a resposta do processo a determinado sinal de excitação. Esta última abordagem, a qual se trata de uma abordagem *bottom-up*, será a utilizada neste trabalho e será apresentada a seguir.

Uma forma de se obter as regras a partir dos dados do sistema é utilizando algoritmos de agrupamento *fuzzy* (*fuzzy clustering*). Agrupamento de dados numéricos forma a base de muitos algoritmos de classificação e modelagem. O intuito do agrupamento é identificar grupos naturais de dados a partir de uma grande massa de dados para produzir uma representação concisa do comportamento do sistema. Os principais algoritmos de agrupamento *fuzzy* podem ser classificados nos três seguintes grupos (Hellendoorn e Driankov, 1997):

1. Algoritmos que utilizam uma medida de distância adaptativa, tais como Gustafson-Kessel (Gustafson e Kessel, 1979) ou algoritmos de estimação de máxima verossimilhança *fuzzy* (Gath e Geva, 1989).
2. Algoritmos baseados em protótipos hiperplanares, também conhecidos como variedade linear *fuzzy* e *c*-elipsóides *fuzzy* (Bezdek, 1981a; Bezdek, 1981b; Bezdek, 1981c).
3. Algoritmos de *c*-regressão, os quais utilizam protótipos definidos por funções de regressão (Hathaway e Bezdek, 1993).

Todas as três classes de algoritmos podem ser vistas como uma extensão do algoritmo *c-means* (Bezdek, 1981a), que minimiza a variância dos dados para os protótipos (centros) dos grupos. A principal diferença entre estes algoritmos está na definição de distância utilizada para estabelecer os grupos, o que influencia na sua estrutura e formato. Neste trabalho dar-se-á maior atenção ao algoritmo de Gustafson-Kessel, que será apresentado em detalhes na seção a seguir. As informações obtidas a partir dos grupos serão, então, utilizadas para obter modelos *fuzzy* TS que representem de maneira satisfatória o sistema sob estudo, com o menor número de grupos (i.e. regras/ modelos locais) possível.

Para a definição das regras *fuzzy* e assim dos antecedentes das regras com os conjuntos *fuzzy* $L_{i,j}$ em (3.6) considera-se que os dados de entrada e saída disponíveis são suficientemente representativos da hipersuperfície não-linear $\hat{y}(k) = \mathcal{H}(\Psi(k))$ (pelo menos na região onde se espera encontrar trajetórias de operação do sistema). Considerando-se um mapeamento desconhecido $\mathcal{H}$ que representa uma hipersuperfície não-linear no subespaço de dimensão $\varsigma + 1$ definido a partir do vetor de regressão

$\Psi(k)$ (domínio $\mathcal{L}$) e da saída y (co-domínio $\mathcal{Y}$), i.e., $\mathcal{L} \times \mathcal{Y} \subset \mathbb{R}^{\varsigma+1}$ pode-se, através de um modelo *fuzzy* TS, aproximar esta hipersuperfície através da combinação de hiperplanos, definidos através de modelos locais presentes nos consequentes das regras em (3.6). O problema é como determinar os limites *fuzzy* de cada hiperplano, que são delimitados pelos conjuntos *fuzzy* $L_{i,j}$ na premissa das regras correspondentes. Uma solução eficaz para este problema (Babuška, 1998; Espinosa, Vandewalle e Wertz, 2004) consiste em se aproximar os hiperplanos por hiperelipsoides planas através do bem conhecido algoritmo de agrupamento *fuzzy* de Gustafson-Kessel (GK) (Gustafson e Kessel, 1979), resultando em uma partição *fuzzy* dos dados em grupos (*clusters*) *fuzzy* de forma hiperelipsoidal. Os conjuntos *fuzzy* de cada regra podem ser então obtidos pela projeção dos valores de ativação dos objetos com relação aos grupos *fuzzy* no(s) eixo(s) do espaço de entrada (espaço unidimensional das variáveis da premissa - estados BFO no presente trabalho). Estes antecedentes serão compostos pelas funções de pertinência que aproximam a casca convexa das projeções por funções convexas, podendo estas, serem funções trapezoidais, triangulares, gaussianas, etc. Desta maneira, cada grupo define uma regra do modelo *fuzzy*. Neste trabalho são utilizadas funções trapezoidais, conforme ilustrado na figura 3.2. Este interesse se deve ao fato das funções trapezoidais, ao contrário das triangulares e gaussianas, possuírem intervalos de pertinência unitária (*core* não *singleton*) que permite a um único modelo local (regra) representar sozinho uma determinada região do espaço de variáveis *fuzzy*.

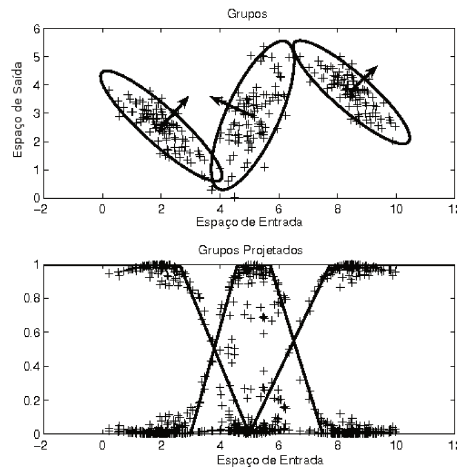


Fig. 3.2: Exemplo de agrupamento GK e projeção.

3.3.2 Algoritmo de Gustafson-Kessel

Seja um conjunto de dados de entrada-saída medidos do sistema (x_j, y_j) , $j = 1, \dots, N$, onde $x \in \mathbb{R}^{\varsigma}$ são as ς -variáveis nas premissas, y a saída do sistema e N o número de medidas. Denotando

$z_j = [x_j, y_j]^T$, pode-se então escrever o conjunto de dados de forma matricial:

$$Z = [z_1, \dots, z_N], \quad z_j \in \mathbb{R}^{s+1} \quad (3.8)$$

Os vetores z_j serão particionados em K grupos com seus protótipos (centros) dados por:

$$v_i = [v_{i,1} \ \dots \ v_{i,s+1}]^T \in \mathbb{R}^{s+1}, \quad i = 1, \dots, K. \quad (3.9)$$

Ao conjunto dos protótipos dos K grupos pode-se dar a notação: $V = [v_1, \dots, v_K]$. O particionamento dos dados é definido pela matriz de particionamento *fuzzy* dada por:

$$U = [\varpi_{i,j}]_{K \times N}$$

onde $\varpi_{i,j} \in [0, 1]$ representa o grau de pertinência da j -ésima medida z_j ao i -ésimo grupo (*fuzzy cluster*) com protótipo v_i . O algoritmo de Gustafson-Kessel (GK) é um dentre aqueles que encontram a matriz de particionamento e os protótipos dos grupos pela minimização da seguinte função objetivo (Bezdek, 1981a; Babuška et al., 2002):

$$J(Z, V, U) = \sum_{i=1}^K \sum_{j=1}^N \varpi_{i,j}^m d^2(z_j, v_i) \quad (3.10)$$

sujeito a

$$\sum_{i=1}^K \varpi_{i,j} = 1, \quad j = 1, \dots, N, \quad (3.11)$$

$$0 < \sum_{j=1}^N \varpi_{i,j} < N, \quad i = 1, \dots, K. \quad (3.12)$$

onde $m > 1$ é um parâmetro que controla a *fuzzificação* do grupo, onde valores maiores de m acarretam uma maior sobreposição dos grupos. Tipicamente o valor $m = 2$ é utilizado (Babuška e Verbruggen, 1997). A função $d(z_j, v_i)$ é a distância do vetor z_j com relação ao protótipo v_i do grupo. A restrição (3.11) impede uma solução trivial $U = [0]_{K \times N}$ e a restrição (3.12) visa garantir que os grupos não sejam vazios e nem contenham todas as medidas com grau 1. Esta é uma definição padrão de agrupamento com uma função objetivo *fuzzy* em que se deve fazer a minimização da função objetivo através de um processo iterativo (Bezdek, 1981a). O formato geométrico do grupo é determinado pelo tipo de função $d(z_j, v_i)$ utilizada. Gustafson e Kessel (1979) generalizaram o algoritmo

de agrupamento para um método de distância adaptativa:

$$d^2(z_j, v_i) = (z_j - v_i)^T M_i (z_j - v_i) \quad (3.13)$$

onde M_i é uma matriz definida positiva adaptada de acordo com a forma atual do grupo, que pode ser descrita em função da matriz de covariância do grupo, F_i :

$$F_i = \frac{\sum_{j=1}^N \varpi_{i,j}^m (z_j - v_i)(z_j - v_i)^T}{\sum_{j=1}^N \varpi_{i,j}^m} \quad (3.14)$$

A matriz M_i utilizada em (3.13) é calculada como a matriz inversa normalizada da matriz de covariância do grupo, dada por:

$$M_i = \det(F_i)^{\frac{1}{d+1}} F_i^{-1} \quad (3.15)$$

A normalização dada pelo determinante de F_i é utilizada para impedir uma solução trivial de (3.10) para $M_i = [0]$. Uma vez calculadas as matrizes M_i calculam-se as distâncias quadráticas $d^2(z_j, v_i)$ como dado em (3.13) e atualizam-se as matrizes de particionamento U , com seus elementos dados por:

$$\varpi_{i,j} = \frac{d^2(z_j, v_i)^{\frac{-1}{m-1}}}{\sum_{l=1}^K d^2(z_j, v_l)^{\frac{-1}{m-1}}}$$

onde, se $d^2(z_j, v_i) = 0$ para algum $j = k$, faça $\varpi_{i,k} = 1$ e $\varpi_{i,j} = 0, \forall j \neq k$.

Os procedimentos do algoritmo acima são repetidos até que a convergência seja obtida, isto é, $\|U_l - U_{l-1}\| < \epsilon$ onde $\|\cdot\|$ é uma norma matricial, l é o passo de iteração e ϵ é uma tolerância adotada para o término da otimização, tipicamente $\epsilon = 0,01$ ou $\epsilon = 0,001$ para dados normalizados (Babuška e Verbruggen, 1997). Após a convergência do algoritmo de GK, obtém-se a matriz de particionamento U , os protótipos v_i dos grupos e as matrizes de covariância $F_i, i = 1, \dots, K$.

3.3.3 Número de Regras

Algoritmos de agrupamento *fuzzy* usualmente requerem que o número de grupos K (número de regras do modelo TS-BFO) seja previamente definido pelo usuário (Höppner, Klawonn, Kruse e Runkler, 1999). Um procedimento utilizado na literatura consiste em se executar o algoritmo de agrupamento diversas vezes para vários números de grupos e, ao final, selecionar o número que apresenta o melhor resultado de acordo com um dado critério (Babuška, 1998; Höppner et al., 1999). Diferentes critérios de avaliação para agrupamentos *fuzzy* são encontrados na literatura, cada um apresentando características particulares que levam a avaliações menos ou mais precisas, dependendo da classe (ou instância) específica do problema em questão. Por esta razão, sugere-se utilizar um

conjunto destes critérios para auxiliar na tomada de decisão de quantos grupos melhor representam os dados em estudo (Bezdek e Pal, 1998; Campello e Hruschka, 2006). Os critérios adotados aqui para fins de ilustração são o *Fuzzy Hyper-Volume (FHV)*, o *Average Partition Density (APD)*, o *Fuzzy Silhouette (FS)*, e uma versão variacional do *Average Within-Cluster Distance (AWCD)*, os quais serão apresentados a seguir.

Inspirado no conceito de que um bom agrupamento é aquele que não apresenta grupos muito dispersos, Gath e Geva (1989) propuseram medidas de validação baseadas no critério de hipervolume ou densidade do agrupamento *fuzzy*. Mais especificamente, o critério de avaliação conhecido como *Fuzzy Hypervolume (FHV)* é dado por:

$$FHV = \sum_{i=1}^K [\det(F_i)]^{-1/2} \quad (3.16)$$

onde F_i é a matriz de covariância do i -ésimo grupo *fuzzy*, dada por:

$$F_i = \frac{\sum_{j=1}^N h(i|z_j)(z_j - v_i)(z_j - v_i)^T}{\sum_{j=1}^N h(i|z_j)} \quad (3.17)$$

onde $h(i|z_j)$ é a probabilidade de seleção do i -ésimo grupo dada a j -ésima amostra z_j . Quando se escolhe o valor do parâmetro *fuzzificador* sendo $m = 2$, o grau de ativação da j -ésima amostra ao i -ésimo grupo, $\varpi_{i,j}$, se aproxima da probabilidade $h(i|z_j)$ (Gath e Geva, 1989). Portanto, a matriz de covariância em (3.17) pode ser reescrita como:

$$F_i = \frac{\sum_{j=1}^N \varpi_{i,j}(z_j - v_i)(z_j - v_i)^T}{\sum_{j=1}^N \varpi_{i,j}} \quad (3.18)$$

Como os autovalores de F_i estão diretamente relacionados à variância do i -ésimo grupo (Campello e Hruschka, 2006), o determinante em (3.16), dado pelo produto dos autovalores de F_i , fornece a medida da dispersão n -dimensional (hipervolume) do grupo. Bons agrupamentos *fuzzy* são aqueles que apresentam menor valor de FHV , ou seja, aqueles que apresentam grupos *fuzzy* mais compactos.

A Densidade média de particionamento (*APD - Average Partition Density*) é outro critério de qualidade de agrupamento *fuzzy* que pode ser utilizado para determinação do número ideal de grupos. Este critério é definido através da seguinte equação:

$$APD = \frac{1}{K} \sum_{i=1}^K \frac{R_i}{[\det(F_i)]^{1/2}} \quad (3.19)$$

onde K é o número de grupos e R_i é a "soma dos números centrais", definida por:

$$R_i = \sum_j \varpi_{i,j}, \forall j \text{ tal que } (z_j - v_i)^T F_i^{-1} (z_j - v_i) < 1 \quad (3.20)$$

que é a soma dos valores de pertinência das amostras j localizados dentro de um raio de proximidade do grupo i . Como para grupos mais compactos o valor do determinante será menor e R_i será maior, se conclui pela análise de (3.19) que os melhores agrupamentos serão aqueles que apresentarem maior *APD*.

Os critérios *FHV* e *APD* são critérios que, de maneira geral, não são apropriados quando se trabalha com muitos atributos (variáveis e/ou dimensões). Este fato acontece porque estes critérios exigem o cálculo do determinante e/ou inversa (ver (3.16), (3.19) e (3.20)) da matriz de covariância F_i . Os cálculos do determinante e da inversão demandam um grande esforço computacional ($O(n_1^3)$ com n_1 sendo o número de parâmetros) e uma dimensão elevada da matriz F_i pode causar problemas numéricos no cálculo de tal determinante e/ou inversa (Campello e Hruschka, 2006), além de ser computacionalmente custoso.

Outro critério que pode ser utilizado para avaliação do número ideal de grupos *fuzzy* é a medida de distância média interna de cada grupo (*AWCD* - *Average Within-Cluster Distance*), dada pela equação:

$$AWCD = \frac{1}{K} \sum_{i=1}^K \frac{\sum_{j=1}^N \varpi_{i,j}^m \|x_j - v_i\|^2}{\sum_{j=1}^N \varpi_{i,j}^m} \quad (3.21)$$

onde $\|\cdot\|$ e m se referem à mesma norma e ao parâmetro *fuzzificador* utilizados no algoritmo de agrupamento, respectivamente. Este método utiliza como avaliação o valor médio das distâncias internas dos grupos. A distância interna de um determinado grupo é dada pela média ponderada das distâncias entre todas as amostras e o protótipo (centro) do grupo com cada distância ponderada pelo grau de pertinência da amostra ao grupo (ϖ_{ij}). Os valores de *AWCD* decrescem monotonicamente, no caso ideal, conforme o número de grupos aumenta, com o valor ideal de grupos indicado no ponto onde o gráfico referente aos valores de *AWCD* descreve um “joelho” mais acentuado.

O quarto critério de determinação do número ideal de grupos considerado neste trabalho é conhecido como *Fuzzy Silhouette* - *FS* (Campello e Hruschka, 2006) o qual é uma generalização *Fuzzy* do critério *Average Silhouette Width Criterion* ou *Crisp Silhouette* (Kaufman e Rousseeuw, 1990; Everitt, Landau e Leese, 2001). Para definir este critério, considere uma amostra $j \in \{1, 2, \dots, N\}$ pertencente ao grupo $i \in \{1, \dots, K\}$. Sob o ponto de vista do contexto de agrupamentos *crisp*, isto significa que a amostra j está mais próxima ao protótipo do grupo i do que do protótipo de qualquer outro grupo. No entanto, sob o ponto de vista do agrupamento *fuzzy*, isso significa que a amostra j pertence ao grupo i com grau $\varpi_{i,j}$ maior do que pertence a outro grupo

fuzzy, ou seja, $\varpi_{i,j} > \varpi_{q,j} \forall q \in \{1, \dots, K\}, q \neq i$.

Seja a distância média da amostra j para todas as outras amostras pertencentes ao grupo i denotada por a_{ij} e seja a distância média desta amostra para todos as amostras pertencentes a outro grupo q , com $q \neq i$, denotada por d_{qj} . Finalmente, seja b_{ij} o menor d_{qj} calculado sobre os grupos $q = 1, \dots, K, q \neq i$, que representa a dissimilaridade da amostra j para o grupo vizinho mais próximo do seu grupo. Então, a silhueta da amostra j é definida como:

$$s_j = \frac{b_{ij} - a_{ij}}{\max(a_{ij}, b_{ij})} \quad (3.22)$$

Levando-se em consideração o grau de ativação da amostra j dentro dos grupos, pode-se construir um critério generalizado chamado *fuzzy silhouette*, ou silhueta *fuzzy*, definido por:

$$FS = \frac{\sum_{j=1}^N (\varpi_{i,j} - \varpi_{q,j})^\alpha s_j}{\sum_{j=1}^N (\varpi_{i,j} - \varpi_{q,j})^\alpha} \quad (3.23)$$

onde s_j é a silhueta da amostra j de acordo com a equação (3.22), $\varpi_{i,j}$ e $\varpi_{q,j}$ são respectivamente o primeiro e o segundo maior elementos da j -ésima coluna da matriz de partição *fuzzy* e $\alpha \geq 0$ é um coeficiente que dá maior ou menor peso à diferença $(\varpi_{i,j} - \varpi_{q,j})$. Quando $\alpha = 0$ a equação (3.23) passa a ser uma média aritmética dos termos s_j , não levando em consideração as informações do agrupamento *fuzzy*, o que corresponde ao critério da silhueta *crisp* original. O valor típico de α para o caso *fuzzy* é 1 (Campello e Hruschka, 2006).

A diferença $(\varpi_{i,j} - \varpi_{q,j})$, que corresponde à diferença entre o primeiro e o segundo maiores valores de pertinência da amostra j com relação aos grupos, tem o objetivo de dar maior peso aos dados pertencentes a determinado grupo com maior proximidade do protótipo do grupo do que àqueles em regiões de mais acentuada sobreposição entre os grupos. O valor ideal de grupos é sugerido pelo maior valor de FS .

• Determinação do número de grupos

Para otimizar o número de grupos e, portanto, a base de regras no modelo TS-BFO, com base em múltiplos critérios de avaliação, utiliza-se um esquema de votação. Este esquema, no entanto, despreza toda a informação quantitativa (valores) dos critérios, usando apenas a informação de qual número de grupos apresentou maior valor na maioria dos critérios. Alternativamente, pode-se combinar o resultado de múltiplos critérios em um único índice médio, e.g., através da média aritmética ou pelo produto dos resultados, como discutido em (Machado, 2007).

A metodologia completa para se obter um modelo mais parcimonioso é a que se segue. Primeiramente, o algoritmo de GK é sequencialmente aplicado ao conjunto de dados no espaço $\mathcal{L} \times \mathcal{Y}$ dos

produtos $\Psi(k) \times y(k)$, com o número de grupos variando em um intervalo de valores aceitáveis, pré-definido pelo usuário. Ao final de cada iteração, as partições *fuzzy* obtidas dos dados são analisadas pelos critérios de avaliação e os respectivos resultados são armazenados. Como apresentado em (Machado, 2007), é utilizada uma versão modificada do critério *AWCD*, uma vez que o critério original decresce monotonicamente com o número de grupos e seu resultado de melhor agrupamento é indicado pelo ponto onde há maior variação na curva. Por esta razão, propõe-se a utilização da informação de primeira ordem (taxa de variação) da curva *AWCD*. Dessa forma o critério passa a indicar o melhor agrupamento no seu ponto de máximo. O resultado do critério *FHV* deve ser adequado com a inversão do sinal, para que o valor máximo passe a indicar o melhor agrupamento. Feito isso e posto que todos os demais critérios são naturalmente de máximo, o resultado de cada um é individualmente normalizado para que todos fiquem compreendidos no intervalo $[0, 1]$ (o pior resultado em 0 e o melhor em 1). Normalizados todos os critérios, calcula-se a média aritmética ponto a ponto e obtém-se um resultado geral. O melhor agrupamento *fuzzy* é indicado pelo ponto de maior valor no resultado geral obtido. Outras abordagens, como o produto entre os resultados dos critérios, podem ser adotadas como complemento na tomada de decisão.

A seguir apresenta-se um exemplo que ilustra a aplicação do método de GK no agrupamento de dados, do método de determinação do número de grupos e de modelagem de um sistema através das técnicas discutidas neste trabalho. No exemplo a seguir deseja-se realizar a construção de um modelo *fuzzy* TS que irá modelar uma dada função não-linear apresentada em (Setnes et al., 1998). Considere função dada pela equação (3.24):

$$y(x) = 3e^{-x^2} \cdot \text{sen}(\pi x) + \eta \quad (3.24)$$

onde η é um ruído gaussiano com média zero e desvio padrão dado $\sigma = 0,1$. Como entrada, foram gerados 300 números aleatórios x distribuídos uniformemente no intervalo $[-3, 3]$. A partir da equação (3.24), com as entradas geradas, foram obtidas as saídas $y(x)$ como indicado na Fig. 3.3.

Os objetos foram agrupados utilizando-se o algoritmo de GK e, analisando-se entre um número de 2 a 10 grupos para o agrupamento dos objetos, pode-se verificar que 7 grupos é o valor de grupos que melhor agrupa os dados, conforme indicado tanto na análise pela média dos critérios de avaliação de agrupamento quanto pelo produto dos mesmos, conforme indicado na Fig. 3.4.

Como pode ser verificado na Fig. 3.5 o algoritmo de agrupamento conseguiu agrupar os dados de forma satisfatória, mesmo os dados apresentando ruídos.

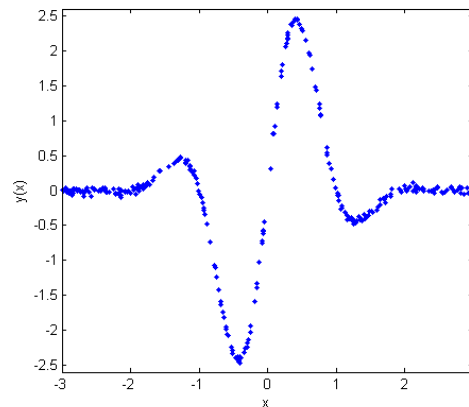


Fig. 3.3: Dados de identificação contaminados com ruído.

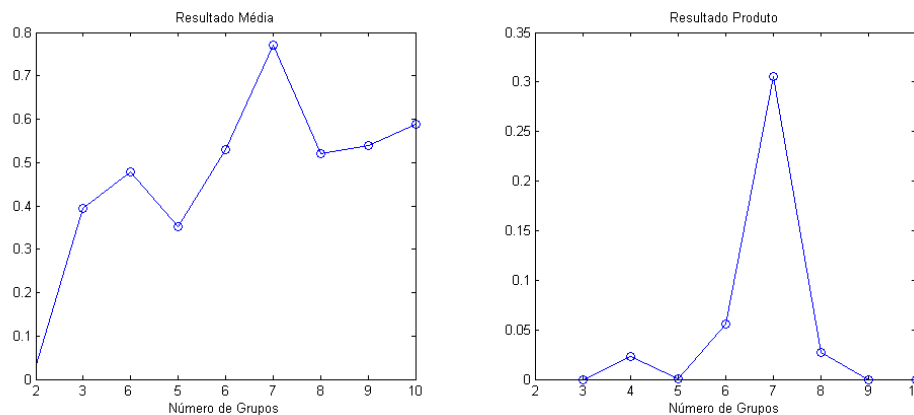


Fig. 3.4: Resultado dado pela combinação dos critérios de análise de agrupamento.

A Fig. 3.6 mostra as funções de pertinência para os antecedentes das regras obtidas a partir das projeções dos valores de ativação dos objetos, nos respectivos grupos, no espaço de entrada. As funções de pertinência cobrem todo o universo de discurso, garantindo que em todo o universo de discurso haverá pelo menos uma regra ativada.

Uma vez determinados os antecedentes das regras o próximo passo então consiste em se determinar os consequentes das mesmas regras. Para este modelo TS serão utilizadas funções afins do tipo $y(x) = a.x + b$, que irão aproximar a função não-linear localmente. Determinadas as funções dos consequentes das regras, via mínimos quadrados, foi então feita a estimação da saída do modelo, como apresentado na Fig. 3.7.

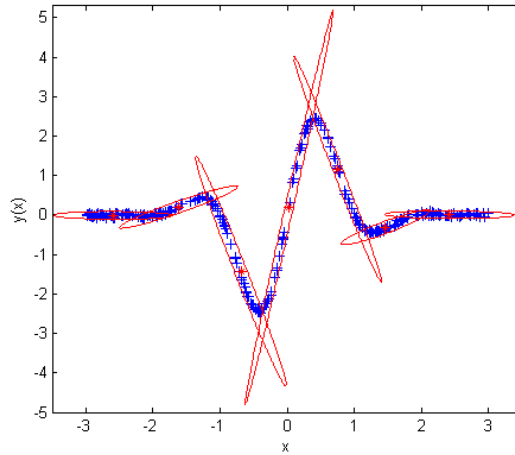


Fig. 3.5: Resultado do agrupamento dos objetos por GK.

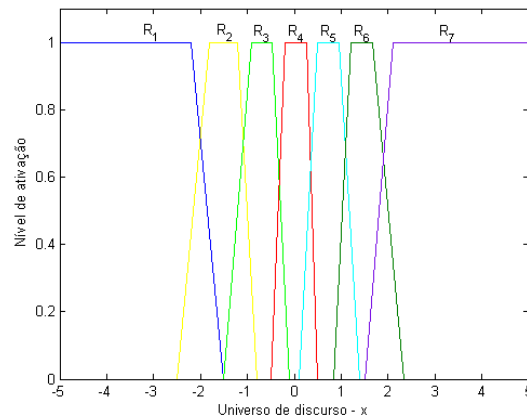


Fig. 3.6: Funções de pertinência dos antecedentes das regras.

Como pode ser verificado visualmente no gráfico da Fig. 3.7, obteve-se um bom modelo para a função que se desejava mapear. O erro quadrático médio - EQM entre o valor estimado pelo modelo e o valor calculado através da função analítica dada em (3.24) é de 0,0047.

3.4 Simplificação da base de Regras dos Modelos *Fuzzy*

O processo de obtenção dos modelos *fuzzy* por agrupamento baseado nos dados de entrada e saída, como apresentado, nem sempre gera modelos parcimoniosos. Podem ocorrer redundâncias de regras (figura 3.8.a com regras 1 e 3 iguais), de funções de pertinência entre si ou com relação ao universo de discurso (figura 3.8.b) ou mesmo com grupos que podem ser fundidos no espaço $\mathcal{L} \times \mathcal{Y}$

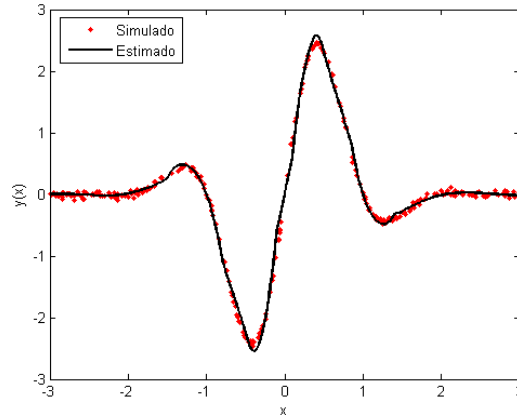


Fig. 3.7: Comparação entre os valores simulados e os estimados.

dos produtos $\Psi(k) \times y(k)$ (figura 3.8.c). Para estes casos de redundâncias existem métodos para realizar a simplificação da base de regras através da eliminação ou fusão de regras e/ou funções de pertinência similares (Babuška, 1998; Setnes et al., 1998; Kaymak e Babuska, 1995). Estes métodos serão discutidos a seguir para cada um dos casos de redundâncias apresentados.

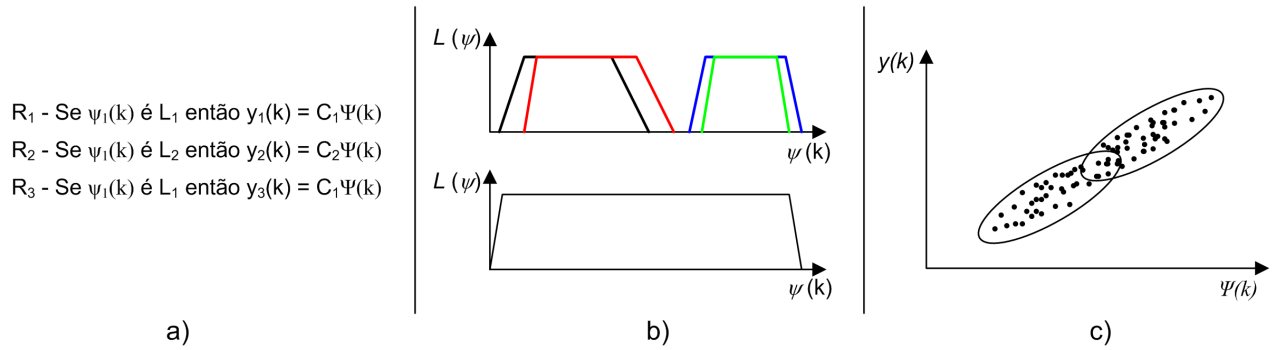


Fig. 3.8: Casos de redundâncias do modelo *fuzzy*.

. Modelos com grupos compatíveis no espaço $\mathcal{L} \times \mathcal{Y}$ (*compatible cluster merging* - CCM)

Grupos compatíveis e que são candidatos a serem fundidos são, em geral, grupos com protótipos próximos e com os hiperelipsoides que os definem alinhados no mesmo hiperplano. O algoritmo CCM que resolve este problema foi inicialmente proposto por Krishnapuram e Freg (1992), sendo que originalmente ele trabalha com o agrupamento de dados por um limitante superior de grupos escolhido pela pessoa que está analisando o problema e aplicando-se o algoritmo sucessivamente se reduz o número de grupos similares até que se convirja para um valor no qual não há mais fusões

a serem realizadas. No presente trabalho realiza-se a avaliação prévia do número ideal de grupos utilizando-se critérios de avaliação apresentados na seção 3.3.3, sendo o valor obtido pelos critérios utilizado como limitante superior.

Os grupos são definidos através de matrizes quadradas da mesma dimensão do espaço de dados $\mathcal{L} \times \mathcal{Y}$ em que os autovetores da matriz definem as direções principais dos hiperelipsóides associados aos grupos e os respectivos autovalores estabelecem o tamanho dos eixos principais, além de um vetor que indica o centro, ou protótipo, de cada grupo, como indica a figura 3.9.

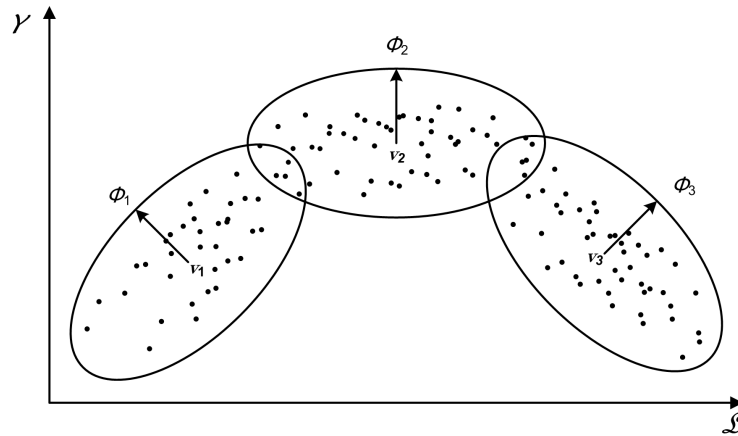


Fig. 3.9: Grupos de dados definidos por hiperelipsóides.

O algoritmo CCM infere o grau de compatibilidade entre os grupos baseado em parâmetros obtidos através de cálculos com os autovalores e os autovetores da matriz que define os grupos. Em (Kaymak e Babuska, 1995) é proposto o cálculo de dois parâmetros que serão utilizados para definir a fusão dos grupos, como apresentado a seguir:

$$c_{1ij} = |\phi_{i,n_i} \cdot \phi_{j,n_j}| \quad (3.25)$$

$$c_{2ij} = \|v_i - v_j\| \quad (3.26)$$

onde i e j indicam os índices relativos aos grupos com $i = 1, \dots, K$ e $j = 1, \dots, K$; ϕ_{i,n_i} e ϕ_{j,n_j} são, respectivamente, os autovetores referentes aos menores autovalores dos grupos i e j ; v_i e v_j os centros dos grupos i e j . A equação (3.25) avalia se dois grupos são paralelos no mesmo hiperplano, ou seja, se c_{1ij} apresenta um valor próximo de 1 os dois grupos são altamente paralelos. A equação (3.26), com o parâmetro c_{2ij} , determina se os centros dos dois grupos são suficientemente próximos para que tais grupos sejam fundidos; desta maneira, c_{2ij} deve apresentar um valor próximo de 0 para que dois grupos estejam relativamente próximos. Através destes critérios são então formadas duas matrizes $\mathcal{C}_1[c_{1ij}]$ e $\mathcal{C}_2[c_{2ij}]$ cujos elementos c_{1ij} indicam o módulo do produto interno entre os autovetores e

c_{2ij} o grau de proximidade entre os grupos i e j de acordo com o critério correspondente. Contudo, os dois critérios não são isoladamente suficientes para estabelecer quais grupos serão fundidos sendo necessária a utilização de um critério de decisão que englobe as informações presentes nas matrizes $\mathcal{C}_1$ e $\mathcal{C}_2$. Kaymak e Babuska (1995) propõem um critério de avaliação e um algoritmo de decisão para se determinar se dois grupos devem ou não serem fundidos.

O algoritmo proposto em (Kaymak e Babuska, 1995) constroi uma matriz S que tem seus elementos pertencentes ao intervalo $[0, 1]$ indicando o grau de similaridade entre os grupos i e j . Para a construção da matriz S normalizam-se os elementos das matrizes $\mathcal{C}_1$ e $\mathcal{C}_2$ de maneira que grupos i j paralelos tenham valores de $\bar{c}_{1ij}$ próximos de 1 e grupos i j com centros próximos tenham valores de $\bar{c}_{2ij}$ próximos de 1. Estabelecidos esses novos valores normalizados, forma-se a matriz S combinando-se os termos $\bar{c}_{1ij}$ e $\bar{c}_{2ij}$ através de uma média geométrica $s_{ij} = \sqrt{\bar{c}_{1ij} \bar{c}_{2ij}}$. Como a fusão dos grupos é uma decisão baseada no valor da similaridade entre os grupos estabelece-se então uma nova matriz S^0 que é composta a partir dos elementos s_{ij}^0 definidos pela função (3.27):

$$s_{ij}^0 = \begin{cases} 1, & \text{se } s_{ij} \geq \varepsilon \\ 0, & \text{se } s_{ij} < \varepsilon \end{cases} \quad (3.27)$$

onde $\varepsilon \in [0, 1]$ representa um valor limite acima do qual deve-se fundir os grupos referentes a posição destes valores na matriz S^0 . Valores de ε pequenos levam a fusão de grupos não muito similares e provavelmente a modelos com um poder de representação prejudicado, já valores elevados levam a modelos sem fusão de grupos similares e menos parcimoniosos. Na literatura é sugerido que o valor de ε que leva a simplificações com bons resultados deva estar dentro do intervalo $[0, 50, 75]$, dependendo do problema a ser tratado (Kaymak e Babuska, 1995). A figura 3.10 mostra um exemplo de como será a composição do novo agrupamento após a fusão e como serão as funções de pertinência nas premissas das regras. Na figura 3.10.a mostra-se o agrupamento inicial com os grupos 2 e 3 com uma grande similaridade e a figura 3.10.b apresenta como ficará o novo agrupamento e as novas funções de pertinência após a aplicação do algoritmo CCM.

Além da similaridade existente entre grupos, outras similaridades devem ser analisadas, como a similaridade entre uma função de pertinência e o universo de discurso e a similaridade entre funções de pertinência obtidas pelas projeções dos grupos.

. Similaridade entre uma função de pertinência e o universo de discurso

A similaridade entre uma função de pertinência A e o universo de discurso se dá quando:

$$A(x) \cong 1, \quad \forall x \in X$$

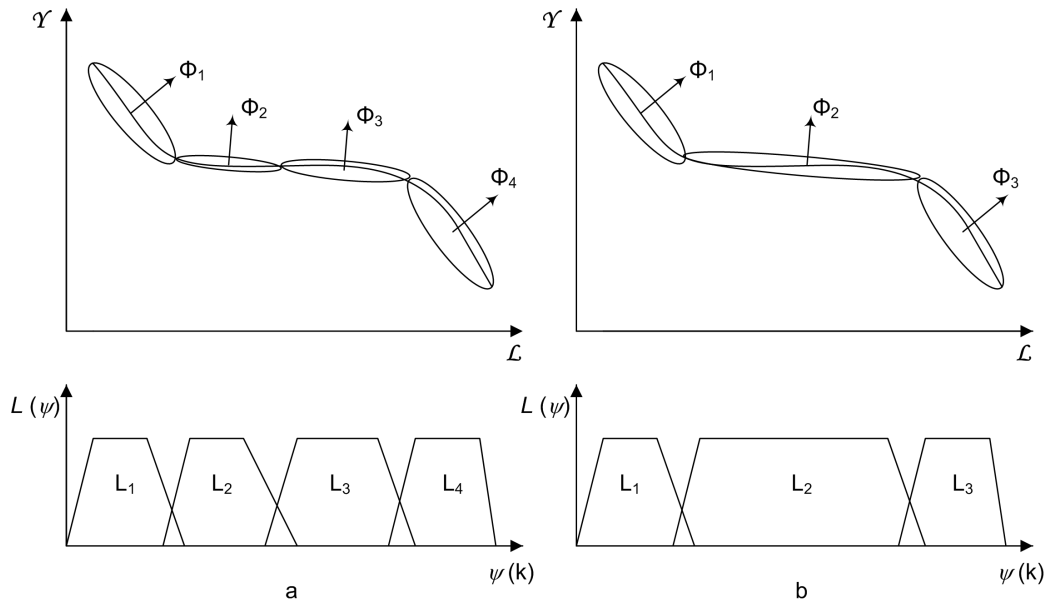


Fig. 3.10: Fusão de grupos similares

onde X representa o universo de discurso da variável de interesse x . Quando este tipo de função de pertinência aparece em um modelo *fuzzy* estimado, ela deve ser excluída do modelo para fins de simplificação, uma vez que não acrescenta informação de relevância ao modelo. Como será abordado a seguir, conjuntos com similaridade com relação ao universo de discurso X superior a um dado limite ρ podem ser eliminados uma vez que a função de pertinência pode cobrir quase a totalidade do universo de discurso.

. Similaridade entre funções de pertinência

Para a classificação da similaridade entre funções de pertinência pode-se estabelecer um grau de similaridade entre duas funções, uma vez que duas funções dificilmente serão estritamente idênticas.

$$s_p = S(A, B) = \text{grau}(A = B), s_p \in [0, 1]$$

sendo que duas funções altamente similares possuem um valor de s_p próximo de 1 e funções não-similares com valor de s_p próximo de 0.

A simplificação da base de regras utiliza medidas de similaridade que devem satisfazer a 4 critérios apresentados a seguir, considerando duas funções de pertinência $A(x)$ e $B(x)$ relativas aos subconjuntos *fuzzy* no universo de discurso X :

1. Funções de pertinência cujos suportes não apresentam superposição devem ser consideradas totalmente desiguais, $s_p = 0$

$$S(A, B) = 0 \Leftrightarrow A(x)B(x) = 0, \quad \forall x \in X.$$

2. Funções de pertinência com superposição devem apresentar valor de similaridade $s_p > 0$

$$S(A, B) > 0 \Leftrightarrow \exists x \in X, \text{ tal que } A(x)B(x) \neq 0.$$

3. Somente funções de pertinência iguais apresentam valor de similaridade $s = 1$

$$S(A, B) = 1 \Leftrightarrow A(x) = B(x), \quad \forall x \in X.$$

4. A similaridade entre duas funções de pertinência não deve ser influenciada por escalonamento ou deslocamento do domínio em que elas estão definidas

$$S(A, B) = S(A', B'), \quad A'(l + kx) = A(x), \quad B'(l + kx) = B(x), \quad k, l \in \mathbb{R}, k > 0.$$

Encontram-se na literatura diferentes métodos de medida de similaridade, sendo eles classificados em similaridades geométricas *fuzzy* e similaridades baseadas na teoria de conjuntos *fuzzy* (Setnes et al., 1998). Como discutido por Zwick, Carlstein e Budescu (1987), as medidas geométricas de similaridade são melhor aplicáveis na medida de similaridade (ou dissimilaridade) entre conjuntos *fuzzy* distintos (representam a similaridade como a proximidade entre conjuntos *fuzzy* e não como a igualdade entre conjuntos), enquanto que as medidas de similaridade baseadas na teoria de conjuntos nebulosos são melhor aplicáveis à análise entre conjuntos que se sobrepõem. As medidas baseadas na teoria *fuzzy* se mostram mais adequadas nas medidas em que se deseja analisar quão iguais são as funções de pertinência. Estas medidas utilizam operações de união e intersecção, não sendo assim influenciadas pela escala das variáveis, como pode ocorrer em medidas geométricas. Nesta tese usa-se uma medida de similaridade baseada na teoria de conjuntos *fuzzy* que satisfaz os 4 critérios apresentados, isto é:

$$S(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (3.28)$$

onde $|\cdot|$ representa a cardinalidade do conjunto, e $\cap$ e $\cup$ representam os operadores intersecção e união respectivamente. Reescrevendo a equação (3.28) em termos de funções de pertinência tem-se:

$$S(A, B) = \frac{\sum_{j=1}^r [A(x_j) \wedge B(x_j)]}{\sum_{j=1}^r [A(x_j) \vee B(x_j)]} \quad (3.29)$$

em um universo de discurso discreto $X = \{x_j | j = 1, 2, \dots, r\}$, com $\wedge$ e $\vee$ representando respectivamente os operadores de mínimo e máximo.

Quando dois conjuntos A e B são considerados muito similares há três possibilidades de simplificar os conjuntos:

- Substituir A por B
- Substituir B por A
- Substituir ambos (A e B) por um novo conjunto C

Dois aspectos importantes devem ser verificados no processo de simplificação de funções de pertinência: a cobertura do universo de discurso das variáveis relativas aos conjuntos sob análise e a melhora ou manutenção da precisão do modelo. Considerando a cobertura do universo de discurso das variáveis *fuzzy*, medidas devem ser tomadas para se garantir que em todo o universo de discurso haja pelo menos uma função de pertinência ativa. Ocorrendo algum ponto do universo de discurso que não tenha nenhuma função de pertinência ativa deve-se ampliar a área de uma função de pertinência adjacente a fim de garantir a cobertura total do universo. Sob o ponto de vista da precisão, assumindo como critério de avaliação do modelo uma função de custo J (e.g. a soma do erro quadrático) os efeitos sobre J ao se substituir A ou B por C devem ser tais que a função J seja sempre otimizada ou pelo menos tenha seu valor preservado, sendo, para isto, analisado qual conjunto tem maior influência em J quando substituído. Quando o modelo é mais sensível a A do que a B , o conjunto A deve substituir o conjunto B , ou no caso de um novo conjunto C substituir ambos, o mesmo deve se aproximar mais de A do que de B .

Para o processo de composição da nova função de pertinência C que irá substituir as funções A e B pode-se utilizar a operação de união $A \cup B$. Como neste trabalho utilizam-se funções de pertinência trapezoidais, definidas como:

$$A(x; \sigma_{1_A}, \sigma_{2_A}, \sigma_{3_A}, \sigma_{4_A}) = \begin{cases} 0, & x \leq \sigma_{1_A} \text{ ou } x \geq \sigma_{4_A} \\ 1, & \sigma_{2_A} \leq x \leq \sigma_{3_A} \\ \alpha, & \alpha \in (0, 1), \text{ no conj. complementar} \end{cases} \quad (3.30)$$

com $\sigma_{1_A} \leq \sigma_{2_A} \leq \sigma_{3_A} \leq \sigma_{4_A}$, a operação de união $A \cup B$ será definida por:

- Sendo $A(x; \sigma_{1_A}, \sigma_{2_A}, \sigma_{3_A}, \sigma_{4_A})$ e $B(x; \sigma_{1_B}, \sigma_{2_B}, \sigma_{3_B}, \sigma_{4_B})$ as funções de pertinência relativas aos conjuntos A e B respectivamente, o conjunto C resultante da fusão de A e B , com sua função de pertinência dada por $C(x; \sigma_{1_C}, \sigma_{2_C}, \sigma_{3_C}, \sigma_{4_C})$, tem seus parâmetros definidos por:

$$\sigma_{1_C} = \min(\sigma_{1_A}, \sigma_{1_B})$$

$$\begin{aligned}
\sigma_{2_C} &= \lambda_2 \sigma_{2_A} + (1 - \lambda_2) \sigma_{2_B} \\
\sigma_{3_C} &= \lambda_3 \sigma_{3_A} + (1 - \lambda_3) \sigma_{3_B} \\
\sigma_{4_C} &= \max(\sigma_{4_A}, \sigma_{4_B})
\end{aligned} \tag{3.31}$$

com os parâmetros $\lambda_2, \lambda_3 \in [0, 1]$ determinados de acordo a influência do conjunto na precisão do modelo, utilizando-se neste trabalho os valores de $\lambda_2 = \lambda_3 = 0,5$ para se ponderar de maneira semelhante as duas funções de pertinência fundidas. A operação descrita acima garante que no lugar das duas funções de pertinência primitivas não haverá região sem cobertura. A figura 3.11 mostra como se dá a fusão de funções e finalmente a função resultante.

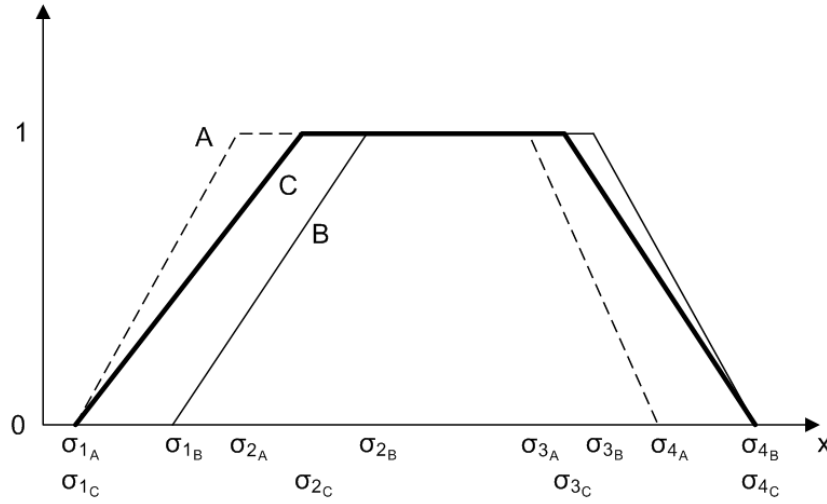


Fig. 3.11: Resultado de fusão entre conjuntos similares

A simplificação da base de regras por similaridade dos conjuntos *fuzzy* se dá pela fusão de conjuntos similares e/ou pela eliminação de conjuntos similares ao universo de discurso. Dois níveis limites λ e ρ determinam, respectivamente, se o conjunto deve ser fundido a outro conjunto ou eliminado quando comparado ao universo de discurso. O algoritmo se inicia verificando a similaridade entre os conjuntos através da equação (3.29) e os pares que apresentam um valor de similaridade superior ao valor de λ devem ser fundidos. Depois de realizadas as fusões nesta fase, novamente analisa-se a similaridade entre os conjuntos até que não haja mais conjuntos a serem fundidos. Uma vez concluída esta fase, analisa-se se há algum conjunto com similaridade com relação ao universo de discurso X superior ao valor de *threshold* ρ . Faz-se esta análise de maneira similar a que é feita com relação a duas funções de pertinência, contudo, ao se comparar a similaridade de uma função de pertinência $A(x_j)$ com universo de discurso X considera-se que o universo de discurso é equivalente a uma função de pertinência com valor de ativação igual a 1 para todo $x_j \in X$. Havendo conjuntos nessa

condição, estes devem ser excluídos da base de regras *fuzzy*. A escolha dos valores de ρ e λ depende de cada aplicação, sendo que, valores reduzidos destes parâmetros levarão a modelos com poucos conjuntos, o que geralmente também reduz a precisão do modelo. Contudo, valores elevados de tais limites levam a modelos redundantes, não parcimoniosos e sobredeterminados. Adota-se empiricamente, neste trabalho, os valores limites para determinação da simplificação da similaridade como $\rho = 0,85$ e $\lambda = 0,6$, baseando-se em experimentos preliminares de simulações realizadas.

3.5 Ajuste global dos parâmetros do modelo final

Uma vez executadas as operações de fusão e eliminação de funções de pertinência e regras por similaridade, realiza-se o ajuste fino do modelo obtido através da otimização dos parâmetros do modelo TS-BFO, minimizando-se o erro quadrático entre a saída do modelo e a saída medida em todos os instantes de tempo. A equação (3.32) mostra a função de custo a ser minimizada:

$$\min_{\theta} J = \frac{1}{2} \sum_{k=1}^N (\hat{y}(k) - y(k))^2 \quad (3.32)$$

com

$\hat{y}(k)$ - saída global estimada

$y(k)$ - saída medida do sistema

N - número de amostras do sistema

θ - parâmetros de otimização

A minimização da função de custo se dá em relação aos parâmetros $\theta = [\Upsilon \ C_{f_i} \ L_{i,j}]^T$, onde Υ é o vetor com os parâmetros relativos ao(s) polo(s) do modelo, C_{f_i} a matriz com os coeficientes da expansão em série dos modelos BFO locais presentes nos consequentes das regras, definida em (3.6), e $L_{i,j}$ as funções de pertinência presentes nos antecedentes das regras do modelo *fuzzy* TS-BFO. Analisando-se individualmente os parâmetros a serem otimizados, com relação ao(s) polo(s) tem-se que para os modelos GOBF com funções internas Υ será $\Upsilon = [\gamma_1 \ \cdots \ \gamma_{n_b}]$ como apresentado em (2.26). Já com relação às funções de Laguerre, $\Upsilon = p$ (eq. (2.6)), e referente às funções de Kautz tem-se que o parâmetro $\Upsilon = [b \ c]$ (equações (2.7) e (2.8)).

A seguir, analisa-se a otimização para modelos TS-GOBF com funções internas nos consequentes das regras, sendo que a mesma análise se estende a modelos com funções de Laguerre ou modelos com funções de Kautz (veja Apêndice B). Para se determinar a direção de busca dos valores ótimos referentes aos parâmetros $\Upsilon = [\gamma_1 \ \gamma_2 \ \cdots \ \gamma_{n_b}]^T$ é necessário se conhecer o gradiente de J com

relação a tais parâmetros:

$$\nabla_{\gamma} J = \left[\frac{\partial J}{\partial \gamma_1} \quad \cdots \quad \frac{\partial J}{\partial \gamma_{n_b}} \right]^T \quad (3.33)$$

Inicialmente obtém-se o gradiente da função de custo com relação ao parâmetro γ_l :

$$\frac{\partial J}{\partial \gamma_l} = \sum_{k=1}^N (y(k) - \hat{y}(k)) \frac{\partial \hat{y}(k)}{\partial \gamma_l} \quad (3.34)$$

onde $\hat{y}(k)$ para o modelo *fuzzy* TS-GOBF é dado por:

$$\hat{y}(k) = \frac{\sum_{i=1}^K \mu_i(\Psi(k)) \check{y}_i(k)}{\sum_{i=1}^K \mu_i(\Psi(k))}$$

com

$\mu_i(\Psi(k))$ - ativação da i -ésima regra ($i = 1, \dots, K$) - eq. (3.3);

$\check{y}_i(k)$ - saída do i -ésimo modelo local;

$\Psi(k)$ - saída dos filtros da GOBF no instante k ;

K - número de regras no modelo.

A saída $\hat{y}(k)$ pode ser reescrita como:

$$\hat{y}(k) = \frac{\sum_{i=1}^K \left(\prod_{j=1}^{\varsigma} L_{ij}(\psi_j(k)) \right) C_{f_i}^T \Psi(k)}{\sum_{i=1}^K \left(\prod_{j=1}^{\varsigma} L_{ij}(\psi_j(k)) \right)} \quad (3.35)$$

onde $L_{ij}(\psi_j(k))$ são os valores de ativação das funções de pertinência, $\psi_j(k)$ é a saída do j -ésimo filtro no instante k e ς o número de variáveis no antecedente das regras com $\varsigma \leq n$. A figura 3.12 mostra o modelo TS-BFO com duas funções ortonormais nos antecedentes e nos consequentes das regras com duas regras onde fica explícita a dependência entre as variáveis do modelo.

Definindo-se o numerador e o denominador da equação (3.35) como:

$$v = \sum_{i=1}^K \left(\prod_{j=1}^{\varsigma} L_{ij}(\psi_j(k)) \right) C_{f_i}^T \Psi(k) \quad \text{e} \quad u = \sum_{i=1}^K \left(\prod_{j=1}^{\varsigma} L_{ij}(\psi_j(k)) \right) \quad (3.36)$$

tem-se que:

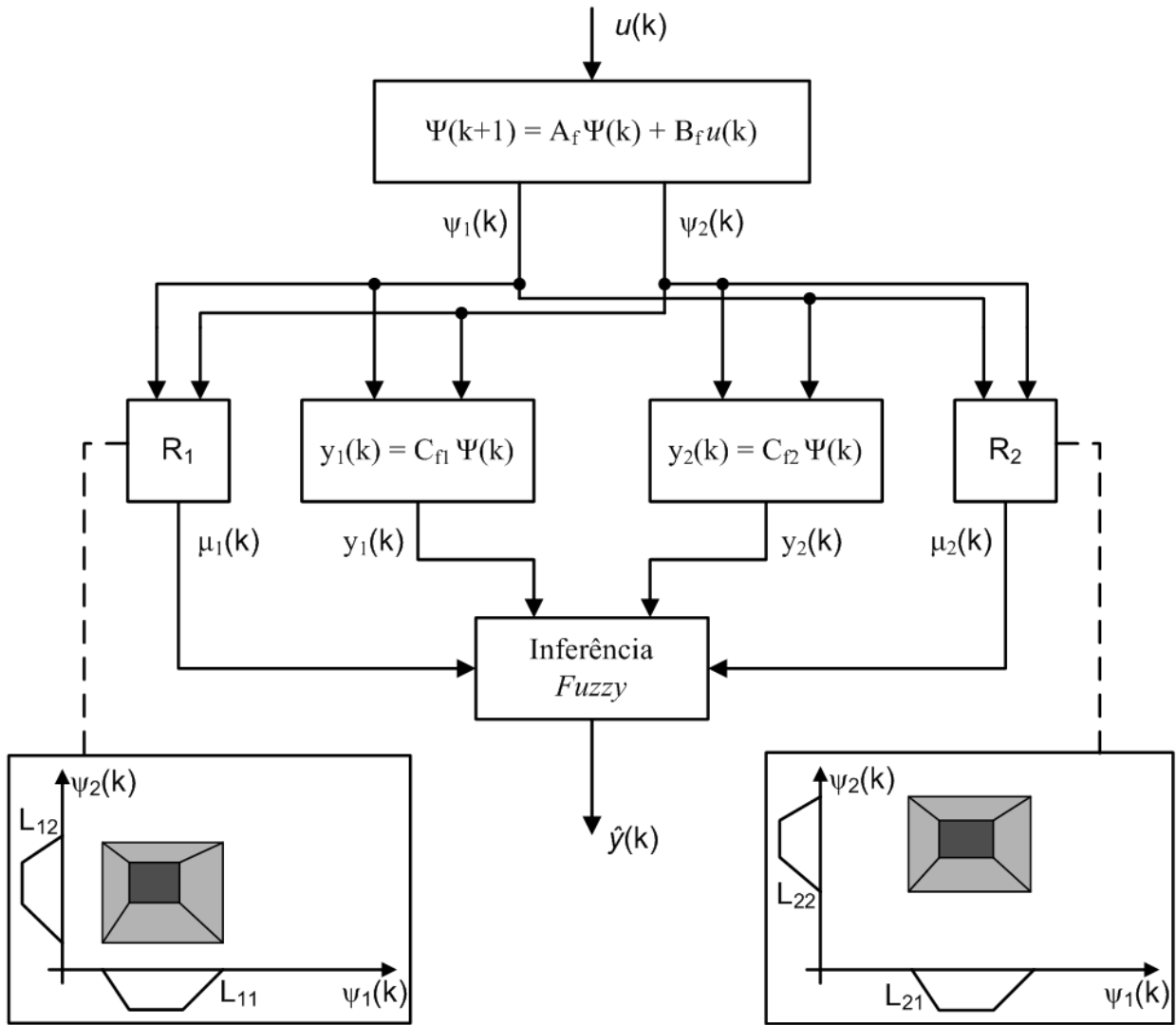


Fig. 3.12: Modelo TS-BFO

$$\hat{y}(k) = \frac{v}{u} \quad \Rightarrow \quad \frac{\partial \hat{y}(k)}{\partial \gamma_l} = \frac{\frac{\partial v}{\partial \gamma_l} \cdot u - v \cdot \frac{\partial u}{\partial \gamma_l}}{u^2} \quad (3.37)$$

Calculando-se o valor da derivada parcial do denominador (u) com relação ao parâmetro γ_l tem-se:

$$\frac{\partial u}{\partial \gamma_l} = \frac{\partial \psi_1(k)}{\partial \gamma_l} \left[\sum_{i=1}^K \frac{\partial L_{i1}(\psi_1(k))}{\partial \psi_1(k)} \cdot L_{i2}(\psi_2(k)) \cdot \dots \cdot L_{is}(\psi_s(k)) \right] +$$

$$\begin{aligned} & \frac{\partial \psi_2(k)}{\partial \gamma_l} \left[\sum_{i=1}^K L_{i1}(\psi_1(k)) \cdot \frac{\partial L_{i2}(\psi_2(k))}{\partial \psi_2(k)} \cdot \dots \cdot L_{i\zeta}(\psi_\zeta(k)) \right] + \\ & + \dots + \frac{\partial \psi_\zeta(k)}{\partial \gamma_l} \left[\sum_{i=1}^K L_{i1}(\psi_1(k)) \cdot L_{i2}(\psi_2(k)) \cdot \dots \cdot \frac{\partial L_{i\zeta}(\psi_\zeta(k))}{\partial \psi_\zeta(k)} \right] \end{aligned} \quad (3.38)$$

onde $\frac{\partial \psi_j(k)}{\partial \gamma_l}$ é definido através dos resultados apresentados na seção 2.4.3 e $\frac{\partial L_{ij}(\psi_j(k))}{\partial \psi_j(k)}$ é definida a seguir.

No modelo TS-BFO abordado, como já citado anteriormente, são utilizadas funções de pertinência trapezoidais. Contudo, a função trapezoidal não é diferenciável em todo o universo de discurso (não apresenta derivada nos pontos $\sigma_1, \sigma_2, \sigma_3, \sigma_4$) o que inviabiliza o cálculo do gradiente da função de pertinência com relação aos estados das BFO's. Para contornar este problema dos pontos não diferenciáveis aproxima-se a função neste pontos por outra função que a torna diferenciável utilizando-se a seguinte função de suavização (Gopal e Biegler, 1999):

$$E_i(x) = \eta \cdot \frac{x + \sqrt{x^2 + \delta^2}}{2}$$

onde η é igual ao coeficiente angular da reta, x é a variável independente e δ o parâmetro que determina a aproximação entre a curva aproximada e a ideal no ponto de vértice.

Para representar o trapézio realiza-se uma composição de funções como mostra a figura 3.13:

Definindo-se as funções como:

$$\begin{aligned} E_1(x) &= \frac{1}{\sigma_2 - \sigma_1} \left(\frac{(x - \sigma_1) + \sqrt{(x - \sigma_1)^2 + \delta^2}}{2} \right) \\ E_2(x) &= \frac{-1}{\sigma_2 - \sigma_1} \left(\frac{(x - \sigma_2) + \sqrt{(x - \sigma_2)^2 + \delta^2}}{2} \right) \\ E_3(x) &= \frac{-1}{\sigma_4 - \sigma_3} \left(\frac{(x - \sigma_3) + \sqrt{(x - \sigma_3)^2 + \delta^2}}{2} \right) \\ E_4(x) &= \frac{1}{\sigma_4 - \sigma_3} \left(\frac{(x - \sigma_4) + \sqrt{(x - \sigma_4)^2 + \delta^2}}{2} \right) \end{aligned}$$

$$E(x) = E_1(x) + E_2(x) + E_3(x) + E_4(x) \quad (3.39)$$

Derivando-se a função de pertinência $E(x)$ com relação à variável independente x obtém-se:

$$\frac{\partial E(x)}{\partial x} = \frac{\partial E_1}{\partial x} + \frac{\partial E_2}{\partial x} + \frac{\partial E_3}{\partial x} + \frac{\partial E_4}{\partial x}$$

com

$$\frac{\partial E_1}{\partial x} = \frac{1}{2(\sigma_2 - \sigma_1)} \left(1 + \frac{(x - \sigma_1)}{\sqrt{(x - \sigma_1)^2 + \delta^2}} \right)$$

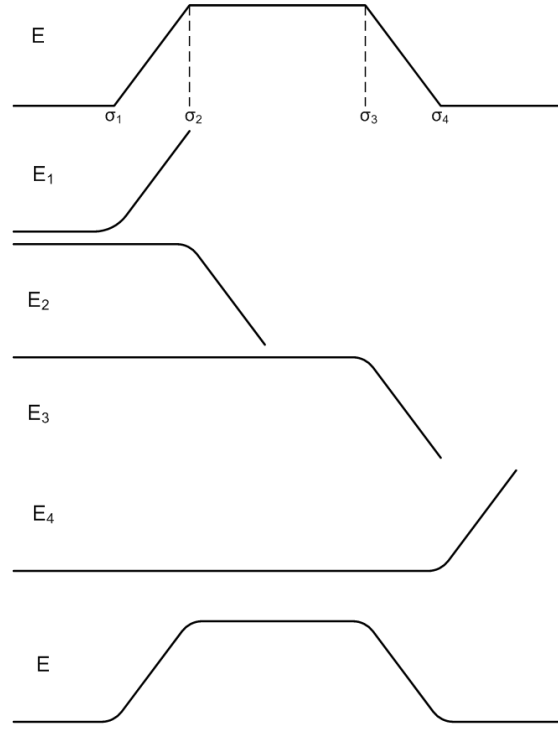


Fig. 3.13: Função de pertinência suavizada

$$\begin{aligned}\frac{\partial E_2}{\partial x} &= \frac{-1}{2(\sigma_2 - \sigma_1)} \left(1 + \frac{(x - \sigma_2)}{\sqrt{(x - \sigma_2)^2 + \delta^2}} \right) \\ \frac{\partial E_3}{\partial x} &= \frac{-1}{2(\sigma_4 - \sigma_3)} \left(1 + \frac{(x - \sigma_3)}{\sqrt{(x - \sigma_3)^2 + \delta^2}} \right) \\ \frac{\partial E_4}{\partial x} &= \frac{1}{2(\sigma_4 - \sigma_3)} \left(1 + \frac{(x - \sigma_4)}{\sqrt{(x - \sigma_4)^2 + \delta^2}} \right)\end{aligned}$$

A mesma análise desenvolvida para o cálculo $\frac{\partial E(x)}{\partial x}$ é utilizada para se calcular a derivada da função de pertinência $L_{ij}(\psi_j(k))$ com relação à $\psi_j(k)$. Com o cálculo de $\frac{\partial L_{ij}(\psi_j(k))}{\partial \psi_j(k)}$ define-se o valor de $\frac{\partial u}{\partial \gamma_l}$.

Para o cálculo $\frac{\partial v}{\partial \gamma_l}$ pode-se reescrever a matriz $C_{f_i}^T$ presente na equação (3.36) em função dos parâmetros da expansão da BFO:

$$C_{f_i} = [g_{1,i} \ g_{2,i} \ \dots \ g_{n_f,i}]^T, \quad (3.40)$$

onde n_f é a maior ordem filtro da BFO e $g_{q,i}$ são os coeficientes da expansão da BFO. Assim, o

numerador da função (3.35) apresentado em (3.36) pode ser redefinido como:

$$v = \sum_{i=1}^K \left[\left(\prod_{j=1}^{\varsigma} L_{i,j}(\psi_j(k)) \right) \sum_{q=1}^{n_f} g_{q,i} \psi_q(k) \right]$$

Analisando-se a derivada parcial do numerador v com relação ao parâmetro γ_l do modelo de GOBF, $\frac{\partial v}{\partial \gamma_l}$, tem-se a fórmula generalizada dada por:

$$\begin{aligned} \frac{\partial v}{\partial \gamma_l} = & \sum_{i=1}^K \sum_{q=1}^{n_f} g_{q,i} \left(\frac{\partial L_{i1}(\psi_1(k))}{\partial \psi_1(k)} \frac{\partial \psi_1(k)}{\partial \gamma_l} \cdot L_{i2}(\psi_2(k)) \cdot \dots \cdot L_{i\varsigma}(\psi_{\varsigma}(k)) \psi_q(k) + \right. \\ & L_{i1}(\psi_1(k)) \cdot \frac{\partial L_{i2}(\psi_2(k))}{\partial \psi_2(k)} \frac{\partial \psi_2(k)}{\partial \gamma_l} \cdot \dots \cdot L_{i\varsigma}(\psi_{\varsigma}(k)) \psi_q(k) + \dots + \\ & L_{i1}(\psi_1(k)) \cdot L_{i2}(\psi_2(k)) \cdot \dots \cdot \frac{\partial L_{i\varsigma}(\psi_{\varsigma}(k))}{\partial \psi_{\varsigma}(k)} \frac{\partial \psi_{\varsigma}(k)}{\partial \gamma_l} \psi_q(k) + \\ & \left. L_{i1}(\psi_1(k)) \cdot L_{i2}(\psi_2(k)) \cdot \dots \cdot L_{i\varsigma}(\psi_{\varsigma}(k)) \frac{\partial \psi_q(k)}{\partial \gamma_l} \right) \end{aligned} \quad (3.41)$$

Determinados os valores de $\frac{\partial u}{\partial \gamma_l}$, $\frac{\partial v}{\partial \gamma_l}$, u e v , substituem-se estes valores em (3.37) e obtém-se o valor de $\frac{\partial \hat{y}(k)}{\partial \gamma_l}$. Substituindo-se os valores de $\frac{\partial \hat{y}(k)}{\partial \gamma_l}$ na equação (3.34) é possível se calcular o gradiente da função de custo com relação ao parâmetro γ_l , sendo o mesmo feito para se calcular o gradiente da função J com relação aos demais parâmetros γ . Os parâmetros γ_l , com $l = 1, \dots, n_b$ presente no modelo GOBF com funções internas apresentam valores restritos ao intervalo $(-1, 1)$.

Como citado anteriormente, além de se desejar minimizar a função de custo J com relação aos parâmetros referentes aos polos das BFO deseja-se também otimizar J com relação aos coeficientes da expansão da série BFO presente nos consequentes e também com relação aos antecedentes das regras do modelo *fuzzy* TS-BFO. Otimizando-se J com relação aos coeficientes $g_{q,i}$ presentes em C_{f_i} (consequentes das regras, eq.(3.40)), tem-se que a derivada da função de custo com relação a tais parâmetros é dada por:

$$\frac{\partial J}{\partial g_{q,i}} = \sum_{k=1}^N (y(k) - \hat{y}(k)) \frac{\partial \hat{y}(k)}{\partial g_{q,i}} \quad (3.42)$$

Reescrevendo a equação (3.35) com C_{f_i} apresentado como na equação (3.40) tem-se:

$$\hat{y}(k) = \frac{\sum_{i=1}^K \left[\left(\prod_{j=1}^{\varsigma} L_{ij}(\psi_j(k)) \right) \sum_{q=1}^{n_f} g_{q,i} \psi_q(k) \right]}{\sum_{i=1}^K \left(\prod_{j=1}^{\varsigma} L_{ij}(\psi_j(k)) \right)} \quad (3.43)$$

cujas derivadas com relação a $g_{q,i}$ é dada por:

$$\frac{\partial \hat{y}}{\partial g_{q,i}} = \frac{\psi_q(k) \cdot \mu_i(\Psi(k))}{\sum_{i=1}^K \left(\prod_{j=1}^s L_{ij}(\psi_j(k)) \right)} \quad (3.44)$$

com

$$\mu_i(\Psi(k)) = \prod_{j=1}^s L_{ij}(\psi_j(k)) \quad (3.45)$$

Para o cálculo de otimização da função de custo J com relação aos parâmetros das funções de pertinência presentes nos antecedentes das regras *fuzzy* tem-se:

$$\frac{\partial J}{\partial \sigma_{ih}} = \sum_{k=1}^N (y(k) - \hat{y}(k)) \frac{\partial \hat{y}(k)}{\partial \sigma_{ih}} \quad (3.46)$$

onde σ_{ih} pode assumir qualquer um dos parâmetros da função de pertinência L_{ih} definidos por $\sigma_{1_{ih}}, \sigma_{2_{ih}}, \sigma_{3_{ih}}$ ou $\sigma_{4_{ih}}$, com tais parâmetros pertencentes à uma função de pertinência trapezoidal como mostra a figura 3.14.

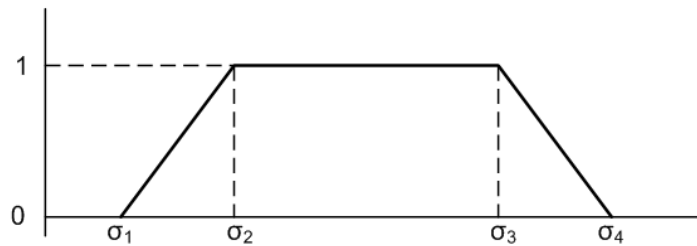


Fig. 3.14: Função de pertinência trapezoidal

No cálculo da derivada parcial $\frac{\partial \hat{y}(k)}{\partial \sigma_{ih}}$ deve-se considerar novamente a equação da saída estimada $\hat{y}(k)$ do modelo *fuzzy* TS-BFO apresentada em (3.35), obtendo-se a seguinte relação para o cálculo de $\frac{\partial \hat{y}(k)}{\partial \sigma_{ih}}$:

$$\hat{y}(k) = \frac{v}{u} \Rightarrow \frac{\partial \hat{y}(k)}{\partial \sigma_{ih}} = \frac{\frac{\partial v}{\partial \sigma} \cdot u - v \cdot \frac{\partial u}{\partial \sigma}}{u^2} \quad (3.47)$$

Analisando-se o denominador u na equação (3.47), obtém-se a seguinte fórmula para o cálculo da derivada desta parcela com relação aos parâmetros σ_{ih} da função de pertinência que se deseja otimizar:

$$u = \sum_{i=1}^K \left(\prod_{j=1}^s L_{ij}(\psi_j(k)) \right)$$

com sua derivada dada por:

$$\frac{\partial u}{\partial \sigma_{ih}} = \frac{\partial L_{ih}(\psi_h(k))}{\partial \sigma_{ih}} \prod_{\substack{j=1 \\ j \neq h}}^s L_{ij}(\psi_j(k))$$

onde L_{ih} representa uma dada função de pertinência h pertencente a regra i .

Assim como no cálculo da derivada das funções de pertinência com relação aos estados da BFO, utiliza-se a aproximação dada na equação (3.39) e, desta maneira, as derivadas das funções de pertinência com relação à σ_{ih} ou equivalentemente aos parâmetros σ_{1ih} , σ_{2ih} , σ_{3ih} ou σ_{4ih} são dadas por:

- com relação ao parâmetro σ_{1ih}

$$\frac{\partial L_{ih}(\psi_h(k))}{\sigma_{1ih}} = \frac{\partial L_{1ih}}{\partial \sigma_{1ih}} + \frac{\partial L_{2ih}}{\partial \sigma_{1ih}}$$

com

$$\begin{aligned} \frac{\partial L_{1ih}}{\partial \sigma_{1ih}} &= \frac{(\psi_h(k) - \sigma_{2ih})}{2(\sigma_{2ih} - \sigma_{1ih})^2} - \frac{(\psi_h(k) - \sigma_{1ih})}{2\sqrt{(\psi_h(k) - \sigma_{1ih})^2 + \delta^2} \cdot (\sigma_{2ih} - \sigma_{1ih})} + \frac{\sqrt{(\psi_h(k) - \sigma_{1ih})^2 + \delta^2}}{2(\sigma_{2ih} - \sigma_{1ih})^2} \\ \frac{\partial L_{2ih}}{\partial \sigma_{1ih}} &= -\frac{(\psi_h(k) - \sigma_{2ih}) + \sqrt{(\psi_h(k) - \sigma_{2ih})^2 + \delta^2}}{2(\sigma_{2ih} - \sigma_{1ih})^2} \end{aligned}$$

- com relação ao parâmetro σ_{2ih}

$$\frac{\partial L_{ih}(\psi_h(k))}{\sigma_{2ih}} = \frac{\partial L_{1ih}}{\partial \sigma_{2ih}} + \frac{\partial L_{2ih}}{\partial \sigma_{2ih}}$$

com

$$\begin{aligned} \frac{\partial L_{1ih}}{\partial \sigma_{2ih}} &= -\frac{(\psi_h(k) - \sigma_{1ih}) + \sqrt{(\psi_h(k) - \sigma_{1ih})^2 + \delta^2}}{2(\sigma_{2ih} - \sigma_{1ih})^2} \\ \frac{\partial L_{2ih}}{\partial \sigma_{2ih}} &= \frac{(\psi_h(k) - \sigma_{1ih})}{2(\sigma_{2ih} - \sigma_{1ih})^2} + \frac{(\psi_h(k) - \sigma_{1ih})}{2\sqrt{(\psi_h(k) - \sigma_{1ih})^2 + \delta^2} \cdot (\sigma_{2ih} - \sigma_{1ih})} + \frac{\sqrt{(\psi_h(k) - \sigma_{1ih})^2 + \delta^2}}{2(\sigma_{2ih} - \sigma_{1ih})^2} \end{aligned}$$

- com relação ao parâmetro σ_{3ih}

$$\frac{\partial L_{ih}(\psi_h(k))}{\sigma_{3ih}} = \frac{\partial L_{3ih}}{\partial \sigma_{3ih}} + \frac{\partial L_{4ih}}{\partial \sigma_{3ih}}$$

com

$$\frac{\partial L_{3_{ih}}}{\partial \sigma_{3_{ih}}} = \frac{(\sigma_{4_{ih}} - \psi_h(k))}{2(\sigma_{4_{ih}} - \sigma_{3_{ih}})^2} + \frac{(\psi_h(k) - \sigma_{3_{ih}})}{2\sqrt{(\psi_h(k) - \sigma_{3_{ih}})^2 + \delta^2} \cdot (\sigma_{4_{ih}} - \sigma_{3_{ih}})} - \frac{\sqrt{(\psi_h(k) - \sigma_{3_{ih}})^2 + \delta^2}}{2(\sigma_{4_{ih}} - \sigma_{3_{ih}})^2}$$

$$\frac{\partial L_{4_{ih}}}{\partial \sigma_{3_{ih}}} = \frac{(\psi_h(k) - \sigma_{4_{ih}}) + \sqrt{(\psi_h(k) - \sigma_{4_{ih}})^2 + \delta^2}}{2(\sigma_{4_{ih}} - \sigma_{3_{ih}})^2}$$

- com relação ao parâmetro $\sigma_{4_{ih}}$

$$\frac{\partial L_{ih}(\psi_h(k))}{\sigma_{4_{ih}}} = \frac{\partial L_{3_{ih}}}{\partial \sigma_{4_{ih}}} + \frac{\partial L_{4_{ih}}}{\partial \sigma_{4_{ih}}}$$

com

$$\frac{\partial L_{3_{ih}}}{\partial \sigma_{4_{ih}}} = \frac{(\psi_h(k) - \sigma_{3_{ih}}) + \sqrt{(\psi_h(k) - \sigma_{3_{ih}})^2 + \delta^2}}{2(\sigma_{4_{ih}} - \sigma_{3_{ih}})^2}$$

$$\frac{\partial L_{4_{ih}}}{\partial \sigma_{4_{ih}}} = \frac{\sigma_{3_{ih}} - \psi_h(k)}{2(\sigma_{4_{ih}} - \sigma_{3_{ih}})^2} - \frac{(\psi_h(k) - \sigma_{4_{ih}})}{2\sqrt{(\psi_h(k) - \sigma_{4_{ih}})^2 + \delta^2} \cdot (\sigma_{4_{ih}} - \sigma_{3_{ih}})} - \frac{\sqrt{(\psi_h(k) - \sigma_{4_{ih}})^2 + \delta^2}}{2(\sigma_{4_{ih}} - \sigma_{3_{ih}})^2}$$

Analisando-se o numerador v presente na equação (3.47), obtém-se o resultado da derivada do numerador com relação aos parâmetros σ_{ih} da função de pertinência que se deseja otimizar:

$$v = \sum_{i=1}^K \left(\prod_{j=1}^{\varsigma} L_{ij}(\psi_j(k)) \right) \check{y}_i(k)$$

$$\check{y}_i(k) = C_i^T \Psi(k)$$

cujas derivadas são dadas por:

$$\frac{\partial v}{\partial \sigma_{ih}} = \frac{\partial L_{ih}(\psi_h(k))}{\partial \sigma_{ih}} \left(\prod_{\substack{j=1 \\ j \neq h}}^{\varsigma} L_{ij}(\psi_j(k)) \right) \check{y}_i(k) \quad (3.48)$$

Obtidos os valores de u , v , $\frac{\partial u}{\partial \sigma_{ih}}$ e $\frac{\partial v}{\partial \sigma_{ih}}$ e os substituindo na equação (3.47) é possível se determinar o gradiente da função de custo J com relação aos parâmetros $\sigma_{1_{ih}}$, $\sigma_{2_{ih}}$, $\sigma_{3_{ih}}$ ou $\sigma_{4_{ih}}$ que compõem as funções de pertinência presentes nos antecedentes das regras.

Como citado anteriormente, a otimização da função de custo (3.32) exige restrições quanto aos valores dos parâmetros γ presentes nas funções internas dos modelos GOBF. Quanto aos demais parâmetros a serem otimizados existem restrições referentes à localização dos parâmetros das funções

de pertinência trapezoidais em que se deve respeitar:

$$\sigma_{1_{ih}} \leq \sigma_{2_{ih}} \leq \sigma_{3_{ih}} \leq \sigma_{4_{ih}}$$

Estas restrições exigem a utilização de algoritmos de otimização que solucionem problemas não-lineares com restrições. Para solucionar tal problema utilizou-se o método de segunda ordem quasi-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS) com busca da solução restrita dada por programação quadrática sequencial conforme discutido na seção 2.4.2 (Nelles, 2001). Este método, assim como demais métodos de Newton, não apresenta garantias de convergência a não ser que o processo seja iniciado nas proximidades da solução (Luenberger, 2003).

3.6 Algoritmo de Otimização dos modelos *fuzzy* TS-BFO

A obtenção de modelos *fuzzy* Takagi-Sugeno - BFO a partir de dados de entrada e saída de sistemas dinâmicos lineares e não-lineares segue, conforme apresentado neste capítulo, uma sequência de passos desde a obtenção inicial do modelo através de um processo de agrupamento *fuzzy* por algoritmo de Gustafson-Kessel até a otimização dos parâmetros do modelo por técnicas de otimização não-linear. O algoritmo a seguir descreve o procedimento global a ser realizado.

1. Atribua valores iniciais de parâmetros γ_l (normalmente de maneira aleatória ou baseado em algum conhecimento prévio) e calcule as saídas dos filtros ortonormais.
2. Dado um conjunto de amostras de entrada (saídas dos filtros ortonormais) e saída do sistema dinâmico, aplica-se o algoritmo de agrupamento de Gustafson-Kessel com o método de determinação da melhor número de grupos a fim de se obter o modelo *fuzzy* TS-BFO inicial como apresentado na seção 3.3.
3. Analisar a similaridade entre os grupos obtidos conforme algoritmo proposto na seção 3.4.
 - **Se** a similaridade for maior que um dado valor limite:
 Realizar o agrupamento dos grupos similares, atualizar o agrupamento *fuzzy* e retornar ao item 3.
 - **Senão** seguir para o procedimento 4.
4. Analisar a similaridade entre os conjuntos *fuzzy* (funções de pertinência) resultantes conforme algoritmo proposto na seção 3.4.

- Se a similaridade for maior que um dado valor limite:
 Funde-se os conjuntos similares como descrito na seção 3.4, atualiza-se a base de regras *fuzzy* e retorna-se ao item 4.
 - Senão segue para o item 5.
5. Analisar a similaridade entre os conjuntos *fuzzy* e o universo de discurso.
- Se o valor de similaridade for maior que um dado limitante:
 Excluir os conjuntos similares ao universo de discurso, atualizar a base de regras do modelo *fuzzy* e voltar à linha 5.
 - Senão segue para a linha 6.
6. Realizado o processo de análise e eliminação das similaridades do modelo inicialmente obtido conforme descrito nas etapas anteriores realiza-se o processo de otimização dos parâmetros do modelo *fuzzy* TS-GOBF:
7. Calcule o gradiente $\nabla_{\theta} J = [\nabla_{\Upsilon}^T J \quad \nabla_{C_{f_i}}^T J \quad \nabla_{L_{i,j}}^T J]^T$ utilizando as equações (3.33), (3.42) e (3.46), onde Υ é o vetor com os parâmetros relativos ao(s) polo(s) do modelo, C_{f_i} a matriz com os coeficientes da expansão em série dos modelos BFO locais presentes nos consequentes das regras, definida em (3.6), e $L_{i,j}$ as funções de pertinência presentes nos antecedentes das regras do modelo *fuzzy* TS-BFO. Para o cálculo de $\frac{\partial J}{\partial \gamma_l}$ em (3.33), com $l = 1, \dots, n_b$, utiliza-se as equações apresentadas de (3.34) até (3.41). Para se calcular $\nabla_{C_{f_i}} J$, deve-se obter os resultados das equações de (3.42) até (3.45) para todos os termos da expansão dos modelos locais. O valor de $\nabla_{L_{i,j}} J$ é obtido através dos cálculos das equações apresentadas de (3.46) a (3.48) para cada parâmetros de todas as funções de pertinência.
8. Use o valor de $\nabla_{\theta} J$ no algoritmo de otimização não-linear apresentado na seção 2.4.2 para atualizar o valor de parâmetros θ . Nota: inclua o intervalo de factibilidade dos parâmetros do modelo GOBF como restrições dentro do problema de otimização para garantir a factibilidade dos polos e dos modelos otimizados; além disso inclua as restrições quanto aos valores dos parâmetros das funções de pertinência.
9. Volte ao passo 7 até que um critério de parada seja atingido.

Este algoritmo garante que os modelos obtidos serão otimizados para melhor representar os sistemas não-lineares sob análise.

3.7 Modelos de Volterra

Os modelos de Volterra de tempo discreto são descritos por uma representação polinomial:

$$\hat{y}(k) = h_0 + \sum_{\varrho=1}^{\infty} \sum_{k_1=1}^{\infty} \cdots \sum_{k_{\varrho}=1}^{\infty} h_{\varrho}(k_1, k_2, \dots, k_{\varrho}) \prod_{j=1}^{\varrho} u(k - k_j) \quad (3.49)$$

onde $u(k)$, $\hat{y}(k)$ e h_{ϱ} são, respectivamente, a entrada, a saída estimada e o *kernel* (núcleo) de ϱ -ésimo grau do modelo apresentado na equação (3.49). Os *kernels* são definidos por meio de funções multidimensionais que geralmente representam a resposta impulsiva de sistemas e que apresentam um elevado número de parâmetros. O modelo de Volterra é composto por uma soma de infinitos termos onde necessariamente deve haver uma convergência dos termos do modelo. Através da utilização de uma série truncada, de ordem E , na composição dos modelos de Volterra, que desprezam os termos desnecessários de ordem superior, é possível se obter modelos mais simples, que utilizam menos parâmetros sem que haja perda significativa de qualidade na precisão na representação de sistemas reais (Eykhof, 1974):

$$\hat{y}(k) = h_0 + \sum_{\varrho=1}^E \sum_{k_1=1}^{\epsilon_{\varrho}} \cdots \sum_{k_{\varrho}=1}^{\epsilon_{\varrho}} h_{\varrho}(k_1, k_2, \dots, k_{\varrho}) \prod_{j=1}^{\varrho} u(k - k_j) \quad (3.50)$$

sendo ϵ_{ϱ} a quantidade de elementos a partir da qual o *kernel* h_{ϱ} é assumido nulo em qualquer dimensão. Como é possível se verificar, a representação (3.50) é claramente uma generalização de E -ésima ordem do modelo linear tipo FIR, sendo o *kernel* h_{ϱ} uma generalização de ϱ -ésima dimensão da função de resposta ao impulso de um modelo FIR. O modelo de Volterra é também uma realização específica de entrada-saída do tipo $y(k) = \mathcal{G}(\{u(\tau)\}_{\tau=-\infty}^k)$, onde $\mathcal{G}$ é um operador não-linear.

Os *kernels* do modelo de Volterra podem apresentar três importantes propriedades (Eykhof, 1974):

- $h_{\varrho}(k_1, k_2, \dots, k_{\varrho}) = 0, \forall k_l < 0$, tal que $l = 1, 2, \dots, \varrho$;
- $\lim_{k_l \rightarrow \infty} h_{\varrho}(k_1, k_2, \dots, k_{\varrho}) = 0$, para $l = 1, 2, \dots, \varrho$;
- $h_{\varrho}(k_1, k_2, \dots, k_{\varrho})$ podem ser considerados simétricos na construção do modelo;

A primeira propriedade implica que o sistema é causal e tem condições iniciais nulas, ou seja, o sistema físico realizável pelo modelo de Volterra não depende de valores futuros do sinal de entrada. A segunda propriedade está relacionada ao fato de sistemas modelados por modelos de Volterra serem do tipo *fading memory*, onde a resposta do sistema a uma excitação na entrada desaparece com o tempo até não mais contribuir significativamente com a saída. Esta propriedade permite o truncamento da série de Volterra considerando-se $h_{\varrho}(k_1, k_2, \dots, k_{\varrho}) = 0$ para

$k_l > \epsilon_\varrho \forall l \in \{1, 2, \dots, \varrho\}$ onde o escalar ϵ_ϱ indica o limite de tempo além do qual a contribuição do $(l+1)$ -ésimo *kernel* é desprezível. Relacionada à terceira propriedade, tem-se que os termos do *kernel* $h_\varrho(k_1, k_2, \dots, k_\varrho)$ referentes a qualquer permutação dos índices $k_1, k_2, \dots, k_\varrho$ multiplicam os mesmo valores de $u(k - k_1) \dots u(k - k_\varrho)$ e com isso pode-se admitir que os termos do *kernel* apresentam simetria sem perda da capacidade de aproximação, implicando em uma simplificação do processo de identificação.

Uma abordagem de especial interesse neste trabalho se refere ao desenvolvimento dos *kernels* de Volterra através de base de funções ortonormais, conforme discutido em (Campello, 2002; Campello et al., 2003; Campello et al., 2007; da Rosa, 2009). Estes modelos têm a capacidade de simplificar a representação de sistemas não-lineares através de modelos de Volterra com a incorporação da base de funções ortonormais em seu *kernel*. A seguir apresentar-se-ão modelos de Volterra que utilizam GOBF em seu desenvolvimento incorporando as características dos modelos GOBF com funções internas, apresentadas na seção 2.3, como também uma metodologia para se obter tais modelos através de técnicas de otimização não-linear.

3.7.1 Modelo Volterra-GOBF

Como discutido anteriormente, os modelos de Volterra apresentam uma desvantagem por, em geral, necessitarem de um número elevado de parâmetros em seus *kernels*. Para a redução do número de termos e uma simplificação do modelo de Volterra propõe-se a utilização de modelos de Volterra compostos por base de funções ortonormais (Schetzen, 1980; Rugh, 1981; Heuberger et al., 2005). Para a construção do modelo de Volterra com BFO considerar-se-á, por hipótese, que os *kernels* h_ϱ em (3.50) são tais que $h_\varrho(k_1, \dots, k_\varrho) = 0$ para $k_i > \epsilon_\varrho (\forall i \in \{1, \dots, \varrho\})$. Os *kernels* são, portanto, absolutamente somáveis em $[0, \infty)$, o que significa que o modelo é estável e atende à segunda propriedade dos modelos Volterra apresentada anteriormente podendo então ser truncado. No desenvolvimento do modelo de Volterra com BFO assume-se, por simplicidade e sem perda de generalidade, que para um dado *kernel*, em todas as direções emprega-se uma mesma base de funções, de tal forma que o *kernel* ϱ -dimensional qualquer é dado por (Schetzen, 1980):

$$h_\varrho(k_1, \dots, k_\varrho) = \sum_{i_1=1}^{\infty} \dots \sum_{i_\varrho=1}^{\infty} g_{i_1, \dots, i_\varrho} \prod_{j=1}^{\varrho} \phi_{\varrho, i_j}(k_j) \quad (3.51)$$

onde ϕ_{ϱ, i_j} é a j -ésima função ortonormal da i -ésima base e $g_{(\cdot)}$ são os coeficientes do desenvolvimento.

Pode-se verificar que, para o *kernel* de 1ª ordem ($\varrho = 1$), o desenvolvimento acima é equivalente ao desenvolvimento da resposta ao impulso do modelo linear de convolução visto na seção 2.2. Substituindo-se os *kernels* com BFO dados pela equação (3.51) no modelo apresentado na equação

(3.50), e lembrando que $h_\varrho(k_1, \dots, k_\varrho) = 0$ para $k_i > \epsilon_\varrho$, pode-se reescrever o modelo de Volterra como:

$$\hat{y}(k) = g_0 + \sum_{m=1}^E \sum_{i_1=1}^{n_{m1}} \sum_{i_2=1}^{n_{m2}} \cdots \sum_{i_m=1}^{n_{mm}} g_{i_1, \dots, i_m} \prod_{j=1}^m \psi_{m, i_j}(k) \quad (3.52)$$

onde $g_0 = h_0$ e ψ_{m, i_j} é a saída do i -ésimo filtro ortonormal da j -ésima base pertencente ao *kernel* de ordem m (i.e. o i_j -ésimo estado com uma base distinta em cada direção do *kernel* de ordem m), dado por $\psi_{m, i_j}(k) = \sum_{\tau=0}^{\infty} \phi_{m, i_j}(\tau) u(k - \tau)$. Nos exemplos apresentados nesta tese, assim como em trabalhos (Billings, 1980; Dumont e Fu, 1993b; Oliveira e Amaral, 2000), utilizam-se modelos de segunda ordem ($E = 2$). Problemas mais complexos, como modelos de telecomunicações podem exigir representações que empreguem *kernels* de ordem superior (Khouaja. e Favier, 2004; Fernandes, Favier e Mota, 2007). Os desenvolvimentos dos *kernels* de primeiro e segundo graus são truncados em quantidades n_1 e n_2 de funções ortonormais, respectivamente, com o modelo (3.52) reescrito como:

$$\hat{y}(k) = g_0 + \sum_{i_1=1}^{n_1} g_{i_1} \psi_{1, i_1}(k) + \sum_{i_1=1}^{n_2} \sum_{i_2=1}^{i_1} g_{i_1, i_2} \psi_{2, i_1}(k) \psi_{2, i_2}(k) \quad (3.53)$$

Para fins computacionais, consideram-se os coeficientes $g_{(\cdot)}$ como parâmetros a serem estimados numericamente. Considerando-se modelos que atendam à propriedade de simetria, isto é, quando utilizada uma mesma BFO em todas as direções em um dado *kernel*, os coeficientes $g_{i_1, \dots, i_m}$ multiplicam o mesmo fator $\psi_{m, i_1}(k), \dots, \psi_{m, i_m}(k)$ independentemente da permutação dos termos no produto das funções ortonormais, o que possibilita uma simplificação no modelo. Desta maneira, para os modelos simétricos, a representação (3.52) pode ser reescrita como:

$$\hat{y}(k) = g_0 + \sum_{m=1}^E \sum_{i_1=1}^{n_{m1}} \sum_{i_2=1}^{i_1} \cdots \sum_{i_m=1}^{i_{m-1}} g_{i_1, \dots, i_m} \prod_{j=1}^m \psi_{m, i_j}(k) \quad (3.54)$$

Os vetores regressores do modelo de Volterra são definidos como G e ζ , onde G e ζ são, respectivamente, o vetor de coeficientes dos *kernels* a serem identificados e o vetor com as saídas das BFO's, dados por:

$$G = [g_0 \ g_1 \ \cdots \ g_{n_{11}} \ g_{1,1} \ g_{2,1} \ g_{2,2} \ \cdots \ g_{n_{21},1} \ g_{n_{21},2} \ \cdots \ g_{n_{21},n_{22}} \ \cdots]^T \quad (3.55)$$

$$\begin{aligned} \zeta = [& 1 \ \psi_{1,1}(k) \ \cdots \ \psi_{1,n_{11}}(k) \ \psi_{2,1}(k)^2 \ \psi_{2,1}(k)\psi_{2,2}(k) \ \psi_{2,2}(k)^2 \\ & \cdots \ \psi_{2,n_{21}}(k)\psi_{2,1}(k) \ \psi_{2,n_{21}}(k)\psi_{2,2}(k) \ \cdots \ \psi_{2,n_{21}}(k)^2 \ \cdots]^T \end{aligned} \quad (3.56)$$

Para os modelos de segunda ordem ($E = 2$) com *kernel* simétrico utilizados nos exemplos apresentados a seguir tem-se que o número de parâmetros do modelo se dá por $\chi = n_{11} + n_{b_1} +$

$n_{2_1}(n_{2_1} + 1)/2 + n_{b_2}$, onde, n_{b_1} é o número de polos e n_{1_1} o número de funções, ambos pertencentes ao *kernel* de primeira ordem. Os parâmetros n_{b_2} e n_{2_1} são, respectivamente, o número de polos e o número de funções pertencentes ao *kernel* de segunda ordem.

Existem modelos de Volterra com BFO em que a simetria dos *kernels* não é válida. Isto acontece quando os *kernels* de segunda ordem ou superior apresentam BFO independentes com polos distintos em cada uma das direções do *kernel* (Schetzen, 1980; Campello, Favier e Amaral, 2004; Campello, Amaral e Favier, 2006) como mostra a equação (3.52). Este modelo melhora a capacidade de modelagem do modelo de Volterra por incorporar outras dinâmicas ao modelo. Para os modelos de segunda ordem ($E = 2$) com *kernel* não simétrico tem-se que o número de parâmetros é dado por $\chi = n_{1_1} + n_{b_1} + n_{2_1}n_{2_2} + n_{b_{2,1}} + n_{b_{2,2}}$, onde n_{b_1} é o número de polos e n_{1_1} o número de funções, ambos pertencentes ao *kernel* de primeira ordem. Os parâmetros $n_{b_{2,1}}$, $n_{b_{2,2}}$, n_{2_1} e n_{2_2} são, respectivamente, o número de polos e o número de funções pertencentes ao *kernel* de segunda ordem em cada direção do *kernel*.

3.7.2 Otimização dos modelos de Volterra-GOBF

Assim como para os modelos BFO's lineares e *fuzzy* TS-BFO é possível se otimizar os modelos de Volterra com GOBF através de métodos de otimização não-linear. Nesta tese apresentam-se duas contribuições para a otimização dos modelos Volterra-GOBF que consistem do cálculo exato de gradiente de modelos Volterra-GOBF simétrico com funções internas e o cálculo do gradiente dos modelos de Volterra-GOBF com funções internas com direções independentes nos *kernels* de ordem superior.

Primeiramente, apresentar-se-á a otimização de modelos de Volterra-GOBF com funções internas com *kernels* de ordem superior simétricos. Para a otimização é necessário se realizar o cálculo dos gradientes do modelo com relação aos seus parâmetros, ou seja, os polos das BFO's e os coeficientes dos *kernels* do modelo (da Rosa, 2009). Para a otimização do modelo deve se minimizar a função de custo apresentada em (2.38):

$$\min_{\theta} J = \frac{1}{2} \sum_{k=0}^N (\hat{y}(k) - y(k))^2$$

onde θ representa os parâmetros do modelo de Volterra-GOBF, $y(k)$ a saída medida do sistema e $\hat{y}(k)$ a saída estimada, sendo esta a mesma apresentada em (3.52). A minimização da função de custo se dará com relação aos parâmetros $\theta = [\Upsilon \ G]^T$, onde Υ é o vetor com os parâmetros relativos aos polos das GOBF's presentes no modelo de Volterra e G o vetor com os coeficientes dos *kernels* dos modelos de Volterra-GOBF, como definido em (3.55).

Analisando-se individualmente os parâmetros a serem otimizados tem-se, com relação aos polos

dos modelos Volterra-GOBF com funções internas, que $\Upsilon = [\gamma_{1,1} \cdots \gamma_{1,n_{b_1}} \cdots \gamma_{m,1} \cdots \gamma_{m,n_{b_m}}]$. Para se determinar a direção de busca dos valores ótimos referentes aos parâmetros Υ faz-se necessário se conhecer o gradiente de J com relação a tais parâmetros:

$$\nabla_{\Upsilon} J = \sum_{k=1}^N (y(k) - \hat{y}(k)) \nabla_{\Upsilon} \hat{y}(k) \quad (3.57)$$

Ignorando-se o termo de ordem zero g_0 no modelo de Volterra-BFO em (3.52), define-se a saída estimada $\hat{y}(k)$ do modelo de Volterra como:

$$\hat{y}(k) = \sum_{m=1}^E \check{y}_m(k) \quad (3.58)$$

onde cada m -ésimo termo $\check{y}_m(k)$ do modelo pode ser escrito como:

$$\check{y}_m(k) = \sum_{i_1=1}^{n_{m_1}} \cdots \sum_{i_m=1}^{n_{m_m}} g_{i_1, \dots, i_m} \prod_{j=1}^m \psi_{m,i_j}(k) \quad (3.59)$$

Como os parâmetros da GOBF de um *kernel* não interferem nas saídas das funções ortonormais dos demais *kernels* o Jacobiano $\nabla_{\Upsilon} \hat{y}(k)$ para o modelo Volterra-GOBF com *kernels* simétricos dá-se por:

$$\nabla_{\Upsilon} \hat{y}(k) = \left[\frac{\partial \check{y}_1}{\partial \gamma_{1,1}} \cdots \frac{\partial \check{y}_1}{\partial \gamma_{1,n_{b_1}}} \cdots \frac{\partial \check{y}_m}{\partial \gamma_{m,1}} \cdots \frac{\partial \check{y}_m}{\partial \gamma_{m,n_{b_m}}} \right]^T \quad (3.60)$$

Para cada m -ésimo termo $\check{y}_m(k)$ tem-se que a sua derivada com relação aos parâmetros $\gamma_{m,\iota}$ dá-se por:

$$\begin{aligned} \frac{\partial \check{y}_m(k)}{\partial \gamma_{m,\iota}} &= \sum_{i_1=1}^{n_{m_1}} \cdots \sum_{i_m=1}^{n_{m_m}} g_{i_1, \dots, i_m} \left[\sum_{l=1}^m \left(\frac{\partial \psi_{m,i_l}(k)}{\partial \gamma_{m,\iota}} \prod_{\substack{j=1 \\ j \neq l}}^m \psi_{m,i_j}(k) \right) \right] \\ &= \sum_{l=1}^m \left(\sum_{i_1=1}^{n_{m_1}} \cdots \sum_{i_m=1}^{n_{m_m}} g_{i_1, \dots, i_m} \prod_{\substack{j=1 \\ j \neq l}}^m \psi_{m,i_j}(k) \right) \frac{\partial \psi_{m,i_l}(k)}{\partial \gamma_{m,\iota}} \end{aligned} \quad (3.61)$$

onde $\frac{\partial \psi_{m,i_l}(k)}{\partial \gamma_{m,\iota}}$ é a derivada parcial da i_l -ésima função da GOBF do m -ésimo termo do modelo de Volterra com relação ao parâmetro $\gamma_{m,\iota}$ com $\iota = 1, \dots, n_{b_m}$. O cálculo das derivadas dos termos

$\frac{\partial \psi_{m,i_l}(k)}{\partial \gamma_{m,\epsilon}}$ é o mesmo apresentado na seção 2.4.3.

Considerando-se a otimização do modelo Volterra-GOBF com relação à matriz de coeficientes dos *kernels* G , tem-se que o Jacobiano da função de custo J com relação estes coeficientes é dada por:

$$\nabla_G J = \sum_{k=1}^N (y(k) - \hat{y}(k)) \nabla_G \hat{y}(k) \quad (3.62)$$

Assim como se procedeu no desenvolvimento do cálculo dos gradientes da saída estimada com relação aos parâmetros $\gamma_{m,\epsilon}$, dividindo-se a saída em parcelas relativas à ordem do *kernel*, adota-se o mesmo procedimento na análise do gradiente da saída estimada com relação aos coeficientes $g_{i_1, \dots, i_m}$ dos *kernels*. Sendo assim, o Jacobiano $\nabla_G \hat{y}(k)$ pode ser escrito como:

$$\nabla_G \hat{y}(k) = \left[\frac{\partial \check{y}_1}{\partial g_1} \quad \dots \quad \frac{\partial \check{y}_1}{\partial g_{n_{11}}} \quad \dots \quad \frac{\partial \check{y}_m}{\partial g_{1, \dots, 1}} \quad \dots \quad \frac{\partial \check{y}_m}{\partial g_{n_{m1}, \dots, n_{mm}}} \right]^T \quad (3.63)$$

O gradiente da saída estimada do m -ésimo termo do modelo de Volterra $\check{y}_m(k)$ com relação aos coeficientes $g_{i_1, \dots, i_m}$ das expansões das funções GOBF nos *kernels* é dado por:

$$\frac{\partial \check{y}_m(k)}{\partial g_{i_1, \dots, i_m}} = \psi_{m,i_1}(k) \cdot \dots \cdot \psi_{m,i_m}(k) \quad (3.64)$$

Especificamente para os modelos de segunda ordem utilizados nesta tese tem-se que os cálculos de suas derivadas com relação aos parâmetros $\gamma_{m,i}$ e aos coeficientes $g_{i,j}$ dão-se por:

- Para os termos do *kernel* de primeira ordem tem-se:

$$\frac{\partial \check{y}_1(k)}{\partial \gamma_{1,i}} = \sum_{i_1=1}^{n_{11}} g_{i_1} \frac{\partial \psi_{1,i_1}(k)}{\partial \gamma_{1,i}}$$

e

$$\frac{\partial \check{y}_1(k)}{\partial g_{i_1}} = \psi_{1,i_1}(k)$$

- Para os termos do *kernel* de segunda ordem:

$$\frac{\partial \check{y}_2(k)}{\partial \gamma_{2,i}} = \sum_{i_1=1}^{n_{21}} \sum_{i_2=1}^{i_1} g_{i_1, i_2} \left[\psi_{2,i_1}(k) \frac{\partial \psi_{2,i_2}(k)}{\partial \gamma_{2,i}} + \psi_{2,i_2}(k) \frac{\partial \psi_{2,i_1}(k)}{\partial \gamma_{2,i}} \right]$$

e

$$\frac{\partial \check{y}_2(k)}{\partial g_{i_1, i_2}} = \psi_{2,i_1}(k) \psi_{2,i_2}(k)$$

Outra contribuição que se desenvolveu nesta tese consiste na otimização de modelos de Volterra-GOBF com funções independentes nos *kernels* de ordem superior do modelo, ou seja, modelos com *kernels* não simétricos. Para o processo de otimização apresentam-se os cálculos analíticos dos gradientes do modelo, com relação aos parâmetros de interesse, que fornecem a direção de busca aos algoritmos de otimização utilizados. Os cálculos dos gradientes com relação aos parâmetros dos polos do modelo e aos coeficientes dos *kernels* se assemelham aos apresentados nas equações (3.34) e (3.62) utilizando a propriedade apresentada nas equações (3.52) e (3.59) em que a saída estimada $\hat{y}(k)$ é dividida em parcelas $\tilde{y}_m(k)$ relativas a cada *kernel* do modelo de Volterra. O gradiente entre $\partial\tilde{y}_m(k)/\partial\gamma_{m,l}$ se diferenciam dos modelos com *kernels* simétricos. Para modelos de Volterra com funções GOBF independentes em cada dimensão nos *kernels* de ordem superior o gradiente da parcela $\tilde{y}_m(k)$ (eq. 3.52) do modelo com relação aos parâmetros $\gamma_{m,j,l}$ (onde m é a ordem do *kernel*, j indica a base a que o parâmetro pertence (direção no *kernel*) e l é a ordem do parâmetro) dá-se por:

$$\frac{\partial\tilde{y}_m(k)}{\partial\gamma_{m,j,l}} = \sum_{i_1=1}^{n_{m_1}} \cdots \sum_{i_m=1}^{n_{m_m}} g_{i_1,\dots,i_m} \psi_{m,i_1}(k) \frac{\partial\psi_{m,i_j}(k)}{\partial\gamma_{m,j,l}} \cdots \psi_{m,i_m}(k) \quad (3.65)$$

onde $\psi_{m,i_j}(k)$ com m igual à ordem do *kernel* e i_j indica a ordem da função ortonormal na direção j do *kernel*.

Os gradientes dos modelos Volterra-GOBF com *kernels* não simétricos com relação aos coeficientes $g_{i_1,\dots,i_m}$ são obtidos de maneira similar ao procedimento realizado com os modelos de Volterra-GOBF com *kernels* simétricos (equações (3.62) e (3.63)) com o diferencial que nos modelos não simétricos há maior número de coeficientes a se otimizar e a equação equivalente à equação (3.64) é dada por:

$$\frac{\partial\tilde{y}_m(k)}{\partial g_{i_1,\dots,i_m}} = \psi_{m,i_1}(k) \cdots \psi_{m,i_m}(k) \quad (3.66)$$

Para os modelos com *kernel* de segunda ordem utilizados neste trabalho o gradiente apresentado na equação genericamente na equação (3.65) é dado por:

- Para o *kernel* de primeira ordem:

$$\frac{\partial\tilde{y}_1(k)}{\partial\gamma_{1,1,l}} = \sum_{i_1=1}^{n_{1_1}} g_{i_1} \frac{\partial\psi_{1,i_1}(k)}{\partial\gamma_{1,1,l}} \quad (3.67)$$

- Para o *kernel* de segunda ordem:

$$\frac{\partial\tilde{y}_2(k)}{\partial\gamma_{2,1,l}} = \sum_{i_1=1}^{n_{2_1}} \sum_{i_2=1}^{n_{2_2}} g_{i_1,i_2} \frac{\partial\psi_{2,i_1}(k)}{\partial\gamma_{2,1,l}} \psi_{2,i_2}(k) \quad (3.68)$$

e

$$\frac{\partial \check{y}_2(k)}{\partial \gamma_{2,2,l}} = \sum_{i_1=1}^{n_{21}} \sum_{i_2=1}^{n_{22}} g_{i_1,i_2} \psi_{2,i_1}(k) \frac{\partial \psi_{2,i_2}(k)}{\partial \gamma_{2,2,l}} \quad (3.69)$$

3.8 Algoritmo de Otimização dos modelos Volterra GOBF

Para se obter modelos Volterra GOBF a partir de dados de entrada e saída de sistemas dinâmicos não-lineares deve-se seguir, conforme apresentado neste capítulo, uma sequência de passos para se otimizar os parâmetros do modelo por técnicas de otimização não-linear. O algoritmo a seguir descreve o procedimento global a ser realizado:

1. Partindo de um vetor inicial de parâmetros θ^0 (normalmente estabelecido de maneira aleatória ou baseado em algum conhecimento prévio), faça:
2. Calcule o gradiente $\nabla_{\theta} J = [\nabla_{\Upsilon}^T J \quad \nabla_G^T J]^T$ utilizando as equações (3.57) e (3.62), onde Υ é o vetor com os parâmetros relativos ao(s) polo(s) do modelo e G o vetor com os coeficientes dos *kernels* dos modelos de Volterra-GOBF, como definido em (3.55). Para o cálculo de $\nabla_{\Upsilon} \hat{y}(k)$ em (3.57), com $l = 1, \dots, n_b$, utiliza-se as equações apresentadas de (3.58) até (3.61). Para se calcular $\nabla_G \hat{y}(k)$, deve-se obter os resultados das equações (3.63) e (3.64). Este procedimento é válido para a otimização de modelos de Volterra GOBF com *kernels* simétricos. Para o caso de modelos com *kernels* não simétricos, $\frac{\partial \hat{y}_m(k)}{\partial \gamma_{m,j,l}}$ deve ser calculado através da equação (3.65) e $\frac{\partial \hat{y}_m(k)}{\partial g_{i_1, \dots, i_m}}$ através da equação (3.66). Estes valores serão utilizados respectivamente nos cálculos de $\nabla_{\Upsilon} J$ e $\nabla_G J$
3. Use o valor de $\nabla_{\theta} J$ no algoritmo de otimização não-linear apresentado na seção 2.4.2 para atualizar o valor de parâmetros θ . Nota: inclua o intervalo de factibilidade dos parâmetros do modelo GOBF como restrições dentro do problema de otimização para garantir a factibilidade dos polos e dos modelos otimizados.
4. Volte ao passo 2 até que um critério de parada seja atingido.

Este algoritmo resume os procedimentos apresentados para a otimização de modelos de Volterra GOBF com *kernels* simétricos ou não simétricos.

A utilização destes algoritmos de otimização baseados em gradientes não garante que os valores obtidos para os parâmetros de interesse sejam ótimos globais. Assim, se o algoritmo convergir para algum mínimo local, e se o modelo não tem a precisão desejada, os seguintes procedimentos podem ser adotados para se melhorar a qualidade do modelo e do processo de otimização (especialmente aplicado aos modelos de Volterra-GOBF):

- i. Inicie o modelo com poucas funções na base, reduzindo desta maneira a capacidade de aproximação do modelo. Esta metodologia permite uma melhor e mais rápida aproximação dos polos iniciais com relação aos polos exatos do sistema. O número de funções na base pode ser incrementado posteriormente de maneira a melhorar a precisão do modelo;

- ii. Decomponha os *kernels* multidimensionais em um conjunto de bases GOBF independentes de maneira que em cada dimensão do *kernel* se tenha bases distintas que contribuam para a melhoria das características de representação do modelo final;
- iii. Utilize algum método clássico de perturbação para se fugir de mínimos locais em busca de outros mínimos locais com melhores resultados de otimização;
- iv. Implemente um método de otimização iterativo que irá otimizar primeiramente a parte linear dos modelos de Volterra-GOBF (*kernels* de primeira ordem). Obtida a convergência da parte linear, otimiza-se posteriormente a parcela não-linear do modelo (*kernels* de ordem superior do modelo de Volterra-GOBF).

3.9 Resultados

Esta seção tem como objetivo apresentar resultados e simulações que ilustram o funcionamento do processo de identificação de sistemas dinâmicos não-lineares por meio de modelos *fuzzy* TS-BFO e de Volterra-GOBF.

A identificação dos modelos *fuzzy* TS-BFO ou TS-GOBF se inicia pela obtenção dos modelos através do agrupamento dos dados de entrada e saída e posteriormente passa por um processo de otimização, como o descrito na seção 3.5. Como visto anteriormente, ao longo deste capítulo, o processo total de otimização consiste, primeiramente, de uma avaliação da similaridade entre grupos, entre conjuntos *fuzzy* e entre os conjuntos e o universo de discurso da variável de interesse. Realizadas as avaliações de similaridade, efetuam-se em sequência as eventuais fusões entre grupos, conjuntos e a eliminação de conjuntos quando estes são desnecessários. Após este processo, com o auxílio de técnicas de otimização não-linear o modelo é refinado através do ajuste de seus parâmetros minimizando-se a soma do erro quadrático entre as saídas estimadas pelo modelo e as saídas obtidas de amostras do sistema sob estudo. Já a identificação de modelos de Volterra-GOBF consiste, como já apresentado, da escolha do número de polos e de funções que se utilizará no modelo. Posteriormente, aplica-se um processo de otimização não-linear ao modelo obtido a fim de melhorá-lo e otimizar os coeficientes dos *kernels* e os parâmetros das GOBF's.

Os resultados a seguir ilustram o desempenho da metodologia apresentada, aplicada na modelagem de sistemas dinâmicos não-lineares reais e simulados. A eficácia dos métodos pode ser verificada uma vez que os modelos resultantes do processo de otimização são computacionalmente mais simples e com uma precisão superior aos modelos originais.

3.9.1 Sistemas de Volterra Simulados

Nesta seção realiza-se a identificação de modelos de Volterra de segunda ordem com a finalidade de se ilustrar a capacidade dos processos de identificação e otimização de recuperar o modelo original sem que se tenha conhecimento a priori do sistema sob análise.

Considere inicialmente um modelo de Volterra-GOBF de segunda ordem, com *kernel* de primeira ordem formado por uma GOBF com 3 polos e 3 funções e um *kernel* de segunda ordem simétrico formado por uma GOBF com 2 polos e 4 funções, com o modelo de Volterra definido como:

$$\hat{y}(k) = G_1 \Psi_1(k) + \Psi_2(k)^T G_2 \Psi_2(k)$$

onde G_1 e G_2 são as matrizes com os coeficientes dos *kernels* e $\Psi_1(k)$ e $\Psi_2(k)$ as saídas dos filtros das GOBF's. Os parâmetros Υ e os coeficientes dos *kernels* são definidos a seguir:

$$\begin{aligned} p_{11} &= 0,8454 & \text{e} & & p_{12} &= 0,2583 \pm 0,7261i \\ \Upsilon_1 &= [0,7700 \quad -0,4583 \quad 0,5000] \\ G_1 &= [0,4329 \quad 0,2259 \quad 0,5798] \\ p_{21} &= 0,5615 \pm 0,3783i \\ \Upsilon_2 &= [0,7700 \quad -0,4583] \\ G_2 &= \begin{bmatrix} 0,7604 & 0,5298 & 0,6405 & 0,2091 \\ 0 & 0,3798 & 0,7833 & 0,6808 \\ 0 & 0 & 0,4611 & 0,5678 \\ 0 & 0 & 0 & 0,7942 \end{bmatrix} \end{aligned} \tag{3.70}$$

Para a realização dos processos de identificação e posterior validação do modelo utilizaram-se dois conjuntos de dados com cada um deles dispondo de 400 pares de amostras de entrada e saída com as amostras de entrada variando entre 0 e 1 em degraus aleatórios com distribuição uniforme de duração de 20 amostras. Na figura 3.15 apresentam-se os dados de entrada e saída utilizados na estimação dos modelos.

Para a modelagem do sistema sob análise partiu-se do pressuposto de que as características deste eram desconhecidas. Desta maneira, fez-se a identificação de modelos variando-se o número de polos e de funções nos *kernels* de primeira e segunda ordem. Para a implementação do processo de otimização dos modelos realizou-se o cálculo dos gradientes com relação aos parâmetros a serem otimizados a partir das amostras de entrada e saída do sistema. Neste exemplo, identificou-se o sistema por meio de modelos de Volterra-GOBF com funções internas com *kernels* simétricos. A tabela 3.1 apresenta alguns resultados comparativos entre diversos modelos de Volterra-GOBF otimizados com

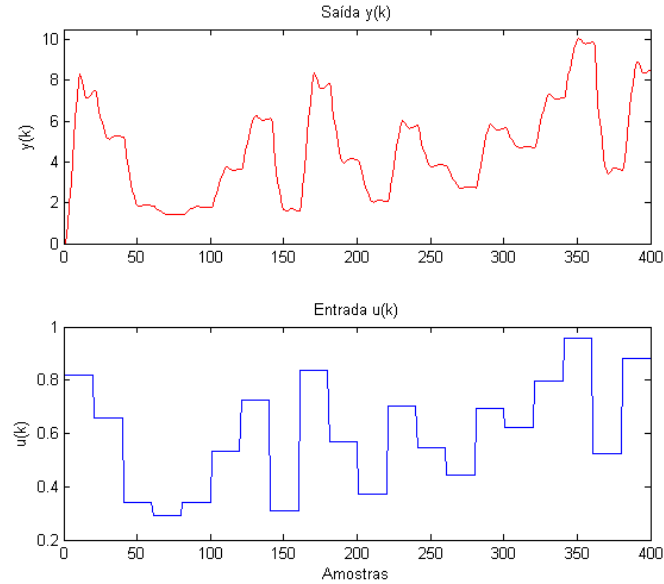


Fig. 3.15: Sinais utilizados na identificação do modelo de Volterra de segunda ordem.

os respectivos erros quadráticos médios (EQM) obtidos pela comparação entre a saída dos modelos estimados e a saída amostrada do sistema sob análise para dados de validação dos modelos, onde, NF e NP indicam, respectivamente, o número de funções e de polos em cada *kernel*.

Tab. 3.1: Resultados dos modelos Volterra-GOBF

NP ₁	NP ₂	NF ₁	NF ₂	EQM
1	1	1	1	0,26052
1	2	2	2	0,16448
2	2	2	6	3,8701.10 ⁻³
2	2	6	6	3,1010.10 ⁻³
3	2	3	4	6,7003.10 ⁻⁷
3	2	6	6	1,3882.10 ⁻³
3	2	9	6	1,7873.10 ⁻³

Através da análise da tabela 3.1 pode-se verificar que o método proposto foi eficiente na identificação do sistema sob análise. Mesmo os modelos com número diferente de polos e funções do sistema original apresentaram bons resultados mas o modelo com melhor resultado foi o modelo com 3 e 2 polos e 3 e 4 funções nos *kernels* de primeira e segunda ordem, respectivamente, que equivale exatamente ao mesmo número de polos e funções do sistema sob análise. O modelo com melhor resultado teve seus parâmetros Υ inicializados com valores $\Upsilon_1 = [0, 5 - 0, 5 \ 0, 5]$ e $\Upsilon_2 = [0, 5 - 0, 5]$ para os respectivos *kernels* e os coeficientes dos *kernels* inicializados com valores obtidos através de

regressão por mínimos quadrados entre a saída e o vetor de entrada dado pelas saídas das funções ortonormais. Os valores dos parâmetros do melhor modelo otimizado são apresentados a seguir:

$$\begin{aligned}
 p_{11} &= 0,8454 & \text{e} & & p_{12} &= 0,2530 \pm 0,7264i \\
 \Upsilon_1 &= [0,7696 \quad -0,4581 \quad 0,5002] \\
 G_1 &= [0,4335 \quad 0,2294 \quad 0,5769] \\
 p_{21} &= 0,5615 \pm 0,3783i \\
 \Upsilon_2 &= [0,7700 \quad -0,4584] \\
 G_2 &= \begin{bmatrix} 0,7658 & 0,4497 & 0,6895 & 0,1207 \\ 0 & 0,2704 & 0,8982 & 0,7945 \\ 0 & 0 & 0,4114 & 0,6018 \\ 0 & 0 & 0 & 0,8091 \end{bmatrix}
 \end{aligned} \tag{3.71}$$

A figura 3.16 apresenta o decaimento do valor da função de custo J em relação a cada iteração do algoritmo de otimização para o modelo GOBF com melhor resultado na tabela 3.1. A função de custo J apresenta um valor inicial de 55,193 e um valor final de $4,7984 \cdot 10^{-5}$ após 213 iterações do algoritmo de otimização. A figura 3.17 mostra a comparação entre a saída estimada do modelo com polos iniciais já apresentados e com os polos finais apresentados no conjunto de valores em (3.71) (dados de validação).

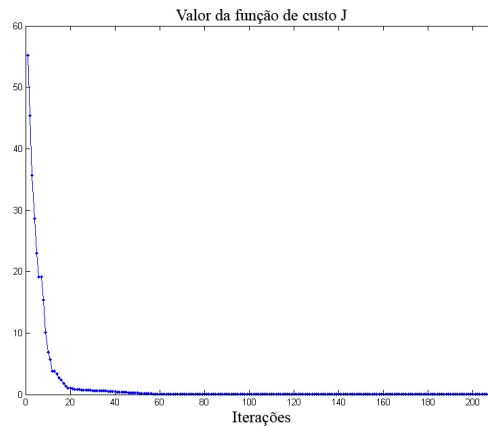


Fig. 3.16: Valor da função de custo

Verifica-se através da análise dos parâmetros finais, apresentados em (3.71), que o modelo otimizado não apresenta exatamente os mesmos parâmetros do sistema original, presentes em (3.70), mas que estes se encontram na vizinhança dos parâmetros originais do sistema o que garante ao modelo otimizado uma boa precisão, indicando a eficiência dos métodos propostos de identificação de

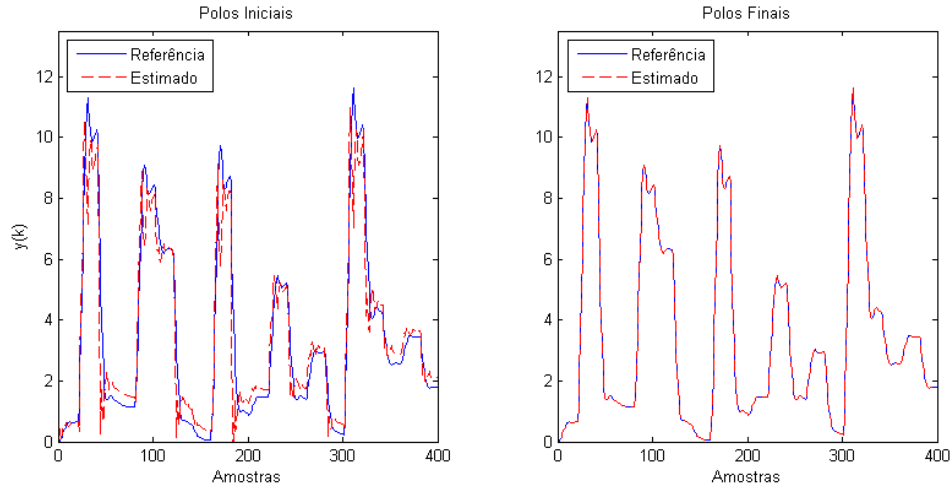


Fig. 3.17: Respostas do modelo de Volterra com *Kernel* simétrico

modelos não-lineares. A seguir, apresenta-se um outro exemplo onde o sistema simulado é do tipo Volterra-GOBF com *kernel* de segunda ordem não simétrico.

Considere agora um modelo de Volterra-GOBF de segunda ordem, com *kernel* de primeira ordem formado por uma GOBF com 1 polo e 2 funções e um *kernel* de segunda ordem não simétrico formado por uma GOBF com 2 polos e 2 funções em uma direção e uma GOBF com 1 polo e 3 funções em outra direção, como o modelo definido a seguir:

$$\hat{y}(k) = G_1 \Psi_1(k) + \Psi_{21}(k)^T G_2 \Psi_{22}(k)$$

onde G_1 e G_2 são as matrizes com os coeficientes dos *kernels* e $\Psi_1(k)$, $\Psi_{21}(k)$ e $\Psi_{22}(k)$ são as saídas dos filtros das GOBF's. Os parâmetros Υ , os polos e os coeficientes dos *kernels* são definidos a seguir:

$$\begin{aligned} p_1 &= 0,4300 \\ \Upsilon_1 &= [0,4300] \\ G_1 &= [0,8681 \quad 0,7756] \\ p_{21} &= 0,4285 \pm 0,6043i \\ \Upsilon_{21} &= [0,5533 \quad -0,5488] \\ p_{22} &= 0,7731 \\ \Upsilon_{22} &= [0,7731] \\ G_2 &= \begin{bmatrix} 0,4319 & 0,4361 & 0,2243 \\ 0,0132 & 0,3760 & 0,8386 \end{bmatrix} \end{aligned} \tag{3.72}$$

Para os processos de identificação e validação dos modelos utilizou-se, assim como para o caso anterior, dois conjuntos de dados com cada um deles dispondo de 400 pares de amostras de entrada e saída com as amostras de entrada variando entre 0 e 1 em degraus aleatórios com distribuição uniforme de duração de 20 amostras. Na figura 3.18 apresentam-se os dados de entrada e saída utilizados na estimação dos modelos.

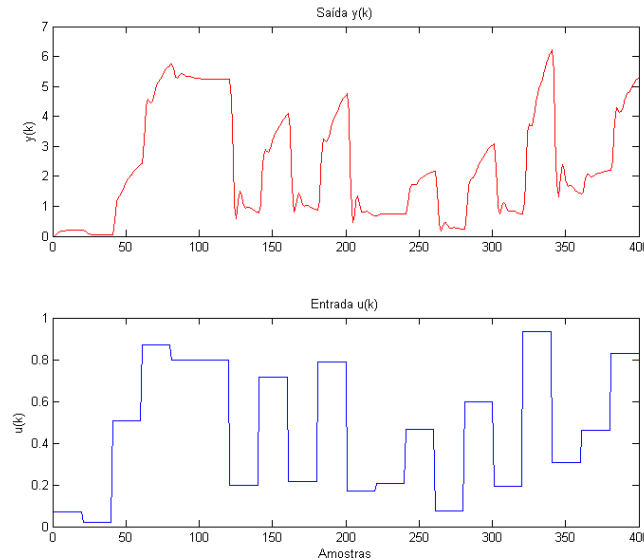


Fig. 3.18: Sinais de identificação - Modelo de Volterra de segunda ordem, *kernel* não simétrico

Novamente, partiu-se do pressuposto de que as características do sistema a ser identificado eram desconhecidas. Desta maneira, fez-se a identificação de diversos modelos variando-se o número de polos e de funções nos *kernels* de primeira e segunda ordem. Para se identificar o modelo do sistema realizou-se o processo de otimização proposto com o cálculo dos gradientes da saída do modelo com relação aos parâmetros a serem otimizados a partir das amostras de entrada e saída do sistema e com a utilização de algoritmos de otimização não-linear obtém-se os modelos desejados. Neste exemplo identificou-se o sistema por meio de modelos de Volterra-GOBF com funções internas com *kernels* não simétricos. Na tabela 3.2 têm-se alguns resultados comparativos entre diversos modelos de Volterra-GOBF otimizados com seus EQM's dados pela comparação entre a saída dos modelos e a saída obtida do sistema sob análise para dados de validação dos modelos, onde, NF e NP indicam, respectivamente, o número de funções e de polos em cada *kernel*.

Através da análise da tabela 3.2 pode-se verificar que o método proposto foi eficiente na identificação do sistema sob análise com o melhor modelo sendo aquele que apresenta o mesmo número de parâmetros do sistema original. Além deste modelo, outros modelos com número de polos e funções diferentes do sistema analisado foram capazes de aproximar o sistema com boa precisão.

Tab. 3.2: Resultados dos modelos Volterra-GOBf não simétrico

NP ₁	NP ₂₁	NP ₂₂	NF ₁	NF ₂₁	NF ₂₂	EQM
1	1	1	1	1	1	0,06969
1	1	2	2	2	6	1,5105.10 ⁻⁴
1	2	1	2	2	3	2,1560.10 ⁻¹⁸
1	1	2	3	3	2	9,4415.10 ⁻¹¹
1	1	3	2	2	6	8,2728.10 ⁻⁵
1	1	3	3	2	6	8,2183.10 ⁻⁵
1	3	1	2	6	2	3,0725.10 ⁻²
1	3	2	2	6	2	3,6836.10 ⁻¹⁰
2	1	2	4	3	2	6,9530.10 ⁻¹¹
3	2	3	6	2	9	1,3348.10 ⁻⁹

O modelo com melhor resultado teve seus parâmetros Υ inicializados com valores $\Upsilon_1 = [0, 5]$, $\Upsilon_{21} = [0, 5 - 0, 5]$ e $\Upsilon_{22} = [0, 5]$ para os respectivos *kernels* e os coeficientes dos *kernels* inicializados com valores obtidos através de regressão por mínimos quadrados entre a saída e o vetor de entrada dado pelas saídas das funções ortonormais. Os valores dos parâmetros do melhor modelo otimizado são apresentados a seguir:

$$\begin{aligned}
 p_1 &= 0,4300 \\
 \Upsilon_1 &= [0,4300] \\
 G_1 &= [0,8681 \ 0,7756] \\
 p_{21} &= 0,4285 \pm 0,6043i \\
 \Upsilon_{21} &= [0,5533 \ -0,5488] \\
 p_{22} &= 0,7731 \\
 \Upsilon_{22} &= [0,7731] \\
 G_2 &= \begin{bmatrix} 0,4319 & 0,4361 & 0,2243 \\ 0,0132 & 0,3760 & 0,8386 \end{bmatrix}
 \end{aligned} \tag{3.73}$$

A figura 3.19 mostra a comparação entre a saída estimada do modelo com polos iniciais já apresentados e com os polos e parâmetros finais destacados no conjunto de valores (3.73) (dados de validação).

Verifica-se através da análise dos parâmetros finais, apresentados em (3.73), que o modelo otimizado tem os mesmos parâmetros do sistema original, apresentados em (3.72), possivelmente se diferenciando somente na quinta ou superiores casas decimais o que faz com que o EQM entre a saída do modelo e do sistema não seja nulo. Os resultados comprovam a eficiência dos métodos

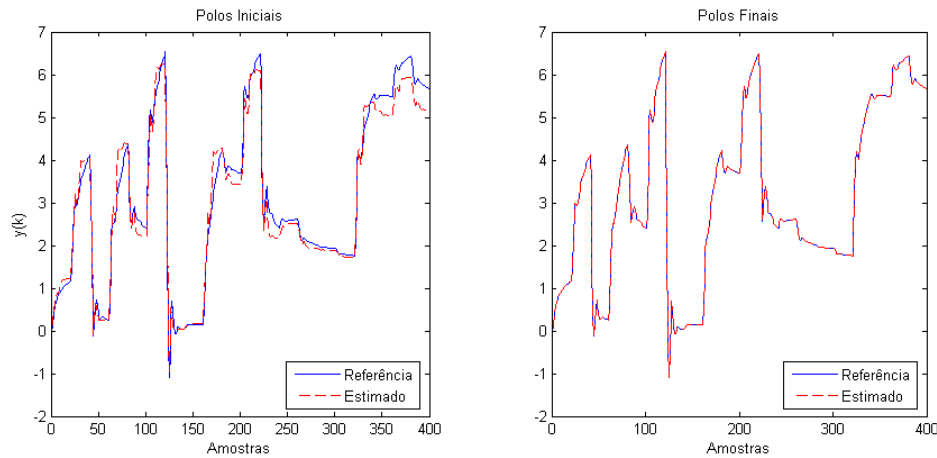


Fig. 3.19: Respostas do modelo de Volterra com *kernel* não simétrico

propostos de identificação de modelos não-lineares de Volterra com *kernels* não simétricos e também mostram que modelos com mais polos e funções são capazes de aproximar com precisão a saída do sistema, pois dessa maneira aumenta-se a capacidade de aproximação da base ortonormal.

3.9.2 Modelagem de nível líquido em um tanque

No exemplo apresentado a seguir faz-se a identificação e modelagem de um processo real de transferência de líquidos entre dois tanques como apresentado na figura 3.20. O processo é descrito em (Lindskog, 1996; Hellendoorn e Driankov, 1997) e a saída $y(k)$ é o nível do líquido no tanque superior (dado em centímetros) e a entrada $u(k)$ é a tensão (em volts) aplicada ao sistema que bombeia líquido para o tanque superior.

Para a identificação e posterior validação do modelo utilizou-se dois conjuntos de dados reais obtidos por (Lindskog, 1996). Cada conjunto possui 1000 pares de dados de entrada e saída amostrados a cada 1 segundo, em um total de cerca de 16 minutos de operação real do processo. A figura 3.21 mostra os dados de entrada e saída utilizados na identificação do processo. As principais causas de não-linearidades apresentadas pelo processo são o fluxo da saída de líquido que é aproximadamente proporcional a $\sqrt{y}$, segundo a lei de Bernoulli para pequenos orifícios, e a saturação existente em y devido à posição do orifício acima da base do tanque.

Na modelagem deste sistema utilizaram-se 3 tipos de modelos, um modelo de Volterra-GOBF de segunda ordem com funções iguais nas duas direções do *kernel* de segunda ordem, um modelo de Volterra-GOBF de segunda ordem com *kernel* de segunda ordem não simétrico e um modelo *fuzzy* TS-GOBF. Em todos os modelos empregou-se GOBF's com funções internas conforme apresentado na seção 2.3. Como não há conhecimento a priori sobre a dinâmica do sistema, utilizou-se GOBF's com

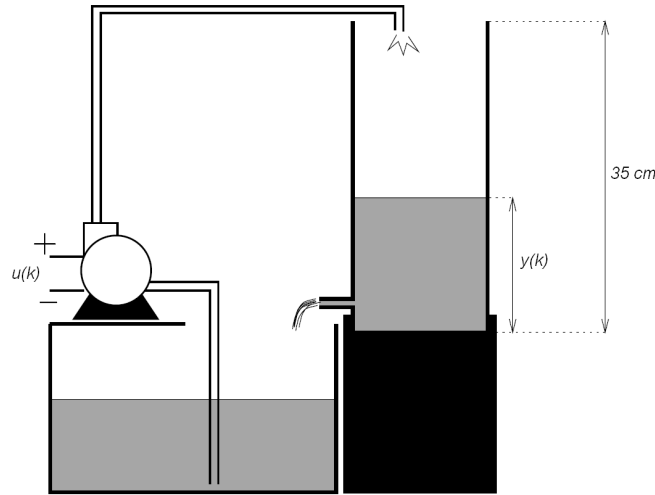


Fig. 3.20: Estrutura esquemática do tanque de líquido.

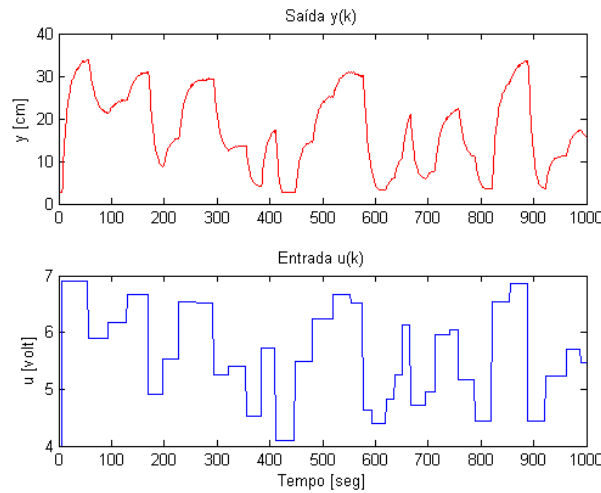


Fig. 3.21: Dados utilizados na estimação do modelo.

diferentes números de polos e de funções para se encontrar o modelo que melhor represente o sistema em cada abordagem empregada. A tabela 3.3 apresenta os melhores resultados encontrados para os modelos nas diversas abordagens com os respectivos polos empregados inicialmente no processo de estimação do modelo, o EQM do modelo inicial (EQMi), os polos do modelo otimizado e o EQM dos modelos otimizados (EQMf) (os valores de EQM foram obtidos com dados de validação). Os coeficientes dos *kernels* foram inicializados com valores obtidos através de regressão por mínimos quadrados entre a saída e o vetor de entrada dado pelas saídas das funções ortonormais.

A tabela 3.4 apresenta, para os melhores resultados encontrados, o número de funções ortonormais (NF) utilizadas em cada *kernel* para os modelos de Volterra-GOBF ou nos consequentes das regras

Tab. 3.3: Valores iniciais e finais do EQM e dos polos dos modelos empregados - Sistema de Tanques.

Tipo de modelo	Polos iniciais	EQMi	Polos finais	EQMf
Volterra-GOBF	$p_1 = 0,5000$ $p_2 = 0,5000$	1,8917	$p_1 = 0,9122$ $p_2 = 0,9272$	0,8870
Volterra-GOBF Direções Independentes	$p_1 = 0,5000$ $p_{21} = 0,5000$ $p_{22} = 0,5000$	1,8917	$p_1 = 0,9327$ $p_{21} = 0,9004$ $p_{22} = 0,8904$	0,7793
Fuzzy TS-GOBF Funções int.	$p_1 = -0,4748 \pm 0,5488i$ $p_2 = 0,9495$	1,2918	$p_1 = -0,4212 \pm 0,9384i$ $p_2 = 0,9384$	0,1635

dos modelos *fuzzy* TS-GOBF, o número de polos (NP), o número de regras no modelo *fuzzy* (NR) e o número de parâmetros estimados e otimizados que se emprega em cada modelo. A abreviação ke_{ij} indica o número de funções presentes da j -ésima direção do i -ésimo *kernel* para o caso.

Tab. 3.4: Informação dos modelos do Sistema de Tanques.

Tipo de modelo	NF	NP	NR	Nº de Param.
Volterra-GOBF	$ke_1 = 8$ $ke_2 = 4$	1 1	-	20
Volterra-GOBF Direções Independentes	$ke_{11} = 8$ $ke_{21} = 4$ $ke_{22} = 4$	1 1 1	-	27
Fuzzy TS-GOBF Funções int.	12	3	3	24

Como se pode verificar na tabela 3.3 os resultados indicam que houve uma diminuição no valor do EQM entre a saída do sistema real e a saída estimada pelos modelos empregados após o processo de otimização, o que ilustra a eficácia da metodologia proposta. O número de polos dos modelos apresentados e o número de funções foram obtidos a partir de resultados de simulação onde variou-se o número de polos e funções. Dos modelos analisados, aquele que melhor se adequou a representação do sistema foi o modelo TS-GOBF funções internas com 3 polos e 12 funções. Os valores dos polos iniciais para o modelos *fuzzy* TS-BFO correspondem aos parâmetros iniciais de γ_i , com $i = 1, 2, 3$, iguais a $[0,5 \ 0,5 \ 0,5]$ e os valores finais dos polos apresentados na tabela 3.3 correspondem aos γ_i , iguais a $[0,5372 \ 0,4518 \ 0,4892]$. Através da análise da tabela 3.4 pode-se verificar que os modelos nas três abordagens empregadas apresentam um número similar de parâmetros a serem otimizados. No modelo *fuzzy* TS-GOBF para cada função de pertinência presente no modelo tem-se quatro parâmetros que as definem. O modelo apresentado tem 3 regras com uma variável no antecedentes das regras.

A figura 3.22 apresenta a saída do modelo *fuzzy* TS-GOBF da tabela 3.3 com uma comparação entre as amostras obtidas da saída do sistema e a saída do modelo inicial identificado através do processo de agrupamento descrito na seção 3.3 aplicados a dados de validação. Também se apresenta na mesma figura uma comparação entre as amostras obtidas da saída do sistema e a saída obtida com o modelo final otimizado de acordo com o processo apresentado nas seções 3.4 e 3.5. As figuras mostram a melhoria na precisão da saída do modelo otimizado.

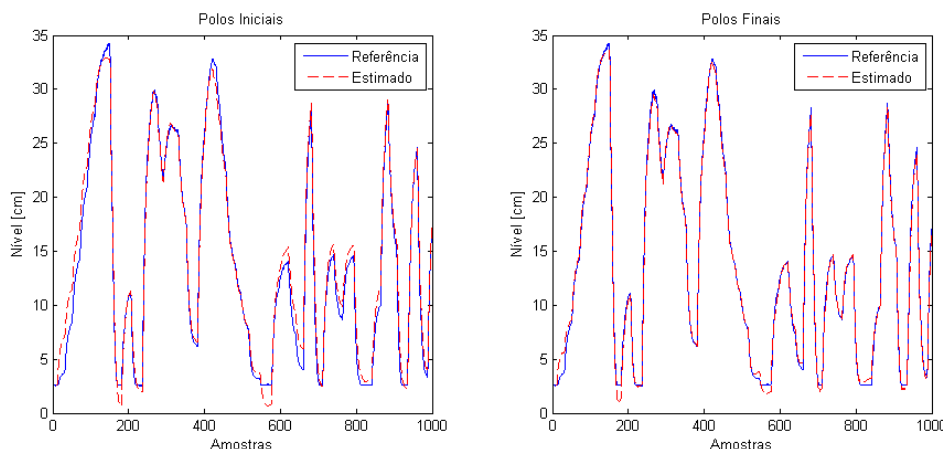


Fig. 3.22: Saídas do modelo *Fuzzy* TS-GOBF - pólos iniciais e otimizados.

3.9.3 Planta para produção de Etanol

Como segundo exemplo de aplicação para os métodos de modelagem propostos utiliza-se uma planta industrial para a produção de etanol. As medidas de entrada e saída da planta em funcionamento foram obtidas por meio de um simulador cujos parâmetros cinéticos foram já validados na planta real (Andrietta, 1994; Andrietta e Maugeri, 1994). Este simulador consiste basicamente de um conjunto de equações diferenciais ordinárias não-lineares, que representam as reações bioquímicas presentes no processo real.

O processo de fermentação para produção de etanol é ilustrado na figura 3.23. O sistema é um processo típico de grande escala industrial, composto de quatro reatores (tanques de fermentação), dispostos em série e que funcionam como uma pilha de reciclagem para produzir etanol a partir de xarope de cana-de-açúcar. O processo é alimentado com uma mistura composta de açúcares (Açúcares redutores totais - TRS), bem como fontes de nitrogênio e sais minerais, chamada meio de alimentação. O meio de alimentação é convertido em etanol por um processo de fermentação realizado com a levedura *Saccharomyces cerevisiae*. Um conjunto de centrífugas separa o meio fermentado que é formado por uma mistura de água, CO_2 , açúcares, microorganismos (30-45 g/l de células) e álcool,

em duas fases. A fase pesada contém a maioria das células (160-200 g/l), enquanto que a fase leve contém cerca de 3 g/l de células e é 9-12% de álcool. A fase leve é então enviada para a unidade de destilação, onde o álcool é extraído.

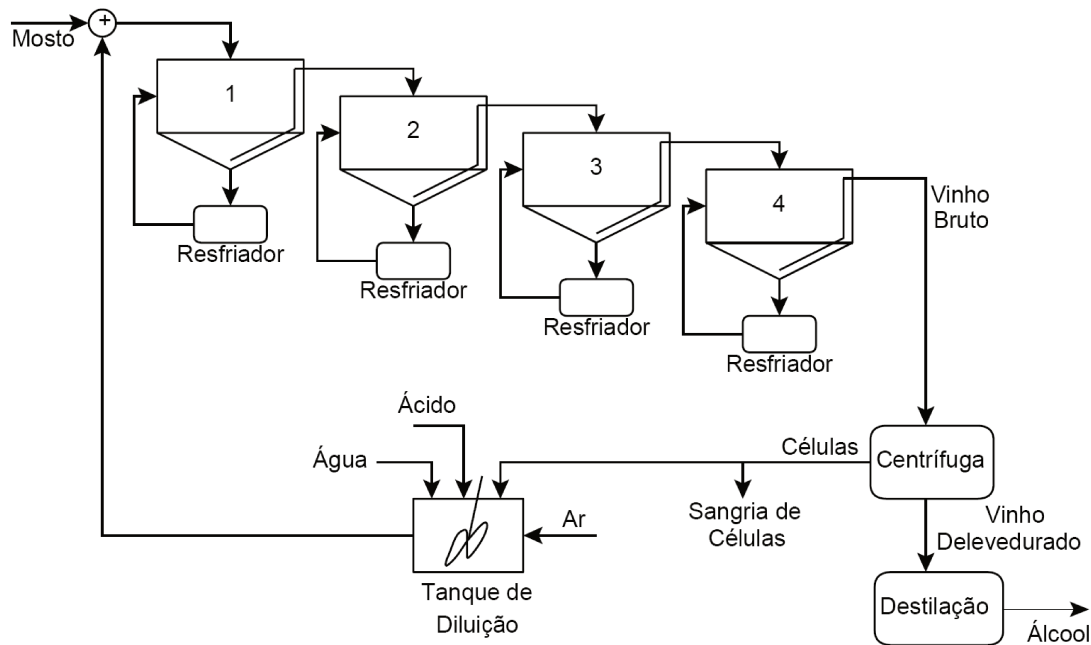


Fig. 3.23: Figura esquemática da planta industrial para produção de etanol.

Como o comportamento dos microorganismos é muito sensível às condições ambiente, alguns deles são eliminados e os restantes são submetidos a um tratamento ácido e diluição antes de serem reciclados no primeiro reator. Além disso, cada reator tem um sistema externo de troca de calor com sistema independente de controle que têm como objetivo manter a temperatura dos reagentes (caldo de fermentação) constante a um nível ideal para o processo de fermentação. Tanto a temperatura de reator quanto a taxa de reciclagem das células de microorganismos são fixados para maximizar a eficiência das reações da planta industrial (Andrietta, 1994; Andrietta e Maugeri, 1994). Uma descrição detalhada do processo pode ser encontrada em (Campello, 2002).

Para a identificação e validação dos modelos propostos usa-se um conjunto de medidas de entrada e saída do sistema correspondente a 20 dias operacionais da planta, onde a primeira metade é utilizada para a estimação do modelo e a segunda metade para a validação do modelo. Nestes dados, o sinal de entrada é formado por uma sequência de degraus com período de 10 h (longo bastante para que o processo possa chegar a quase o estado fixo) e amplitude distribuída uniformemente dentro do intervalo operacional $[50, 150 \text{ m}^3/\text{h}]$. O sinal de saída é dado pela concentração de álcool na saída do quarto tanque $[\text{g/l}]$. As medidas de entrada e saída têm um tempo de amostragem $T = 30 \text{ min}$ que

leva em conta a dinâmica do processo. A figura 3.24 mostra os sinais de entrada e saída utilizados no processo de identificação do modelo.

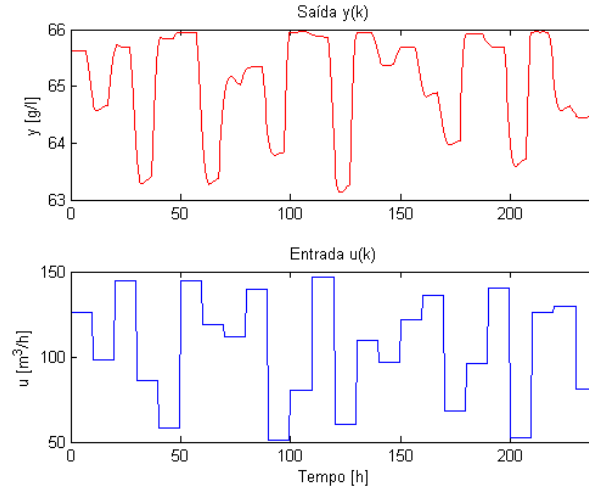


Fig. 3.24: Dados de identificação na estimação do processo de Etanol.

As medidas utilizadas na identificação foram normalizadas entre $[-1, 1]$ para se evitar problemas numéricos na solução computacional do problema de identificação e otimização dos modelos. Realizou-se a modelagem com 3 tipos de modelos sendo um modelo *fuzzy* TS-GOBF com funções internas e dois modelos de Volterra-GOBF de segunda ordem, sendo um deles com funções semelhantes nas duas direções do *kernel* (*kernel* simétrico) e o outro um modelo com funções independentes no *kernel* de segunda ordem. Para a identificação variou-se o número de polos e funções nas bases ortonormais presentes nos modelos de Volterra e no modelo *fuzzy*. Inicializou-se os coeficientes dos *kernels* com valores obtidos através de regressão por mínimos quadrados entre a saída e o vetor de entrada dado pelas saídas das funções ortonormais. Na tabela 3.5 encontram-se os melhores resultados para os modelos com as diversas abordagens propostas nesta tese com os respectivos EQM's obtidos com dados de validação, além dos polos iniciais e dos polos finais obtidos através do processo de otimização.

Na tabela 3.6 encontram-se para os melhores resultados obtidos no processo de identificação o número de funções ortonormais (NF) presentes em cada *kernel* nos modelos de Volterra-GOBF ou que estão nos consequentes das regras dos modelos *fuzzy* TS-GOBF, o número de polos (NP) das funções, o número de regras no modelo *fuzzy* (NR) e o número de parâmetros estimados e otimizados que se emprega em cada modelo. A abreviação ke_{ij} indica o número de funções presentes da j -ésima direção do i -ésimo *kernel*.

Os resultados presentes na tabela 3.5 mostram que houve uma melhoria na precisão dos modelos otimizados com a diminuição no valor do EQM entre a saída do sistema real e a saída estimada

Tab. 3.5: Valores iniciais e finais do EQM e dos polos dos modelos empregados - Sistema de Etanol.

Tipo de modelo	Polos iniciais	EQMi	Polos finais	EQMf
Volterra-GOBF	$p_1 = 0,5000$ $p_2 = 0,5000 \pm 0,5000i$	0,0427	$p_1 = 0,8739$ $p_2 = 0,7007 \pm 0,1479i$	0,0068
Volterra-GOBF Direções Independentes	$p_1 = 0,5000 \pm 0,5000i$ $p_{21} = 0,5000 \pm 0,5000i$ $p_{22} = 0,5000 \pm 0,5000i$	0,0136	$p_1 = 0,7112 \pm 0,2269i$ $p_{21} = 0,8572 \pm 0,5044i$ $p_{22} = 0,8082 \pm 0,6843i$	0,0063
Fuzzy TS-GOBF Funções int.	$p_1 = 0,5000 \pm 0,5000i$	0,1712	$p_1 = 0,7931 \pm 0,0994i$	0,0038

Tab. 3.6: Informação dos modelos do Sistema de Etanol.

Tipo de modelo	NF	NP	NR	Nº de Param.
Volterra-GOBF	$ke_1 = 3$ $ke_2 = 2$	1 2	-	9
Volterra-GOBF Direções Independentes	$ke_{11} = 2$ $ke_{21} = 4$ $ke_{22} = 4$	2 2 2	-	24
Fuzzy TS-GOBF Funções int.	6	2	5	28

comprovando a eficácia da metodologia proposta. O número de polos dos modelos apresentados e o número de funções foram obtidos empiricamente como os melhores resultados dentre os testes realizados sem que se houvesse conhecimento a priori do sistema. Dos modelos analisados, aquele que melhor se adequou a representação do sistema foi o modelo *fuzzy* TS-GOBF com funções internas e 2 polos e 5 regras com uma variável na premissa. A tabela 3.7 apresenta os valores dos parâmetros γ_i para cada um dos modelos identificados e apresentados na tabela 3.5.

Tab. 3.7: Valores iniciais e finais dos Parâmetros dos modelos GOBF - Sistema de Etanol.

Tipo de modelo	Param. iniciais	Param. finais
Volterra-GOBF	$\gamma_1 = 0,5000$ $\Upsilon_2 = [0,6667 \quad -0,5000]$	$\gamma_1 = 0,8739$ $\Upsilon_2 = [0,9263 \quad -0,5177]$
Volterra-GOBF Direções Independentes	$\Upsilon_1 = [0,6667 \quad -0,5000]$ $\Upsilon_{21} = [0,6667 \quad -0,5000]$ $\Upsilon_{22} = [0,6667 \quad -0,5000]$	$\Upsilon_2 = [0,9134 \quad -0,5573]$ $\Upsilon_{21} = [0,9506 \quad -0,4324]$ $\Upsilon_{22} = [0,9610 \quad -0,5530]$
Fuzzy TS-GOBF Funções int.	$\Upsilon = [0,6667 \quad -0,5000]$	$\Upsilon = [0,7531 \quad 0,0788]$

A figura 3.25 apresenta a saída do modelo *fuzzy* TS-GOBF da tabela 3.5 obtida com dados de

validação do modelo. Apresenta-se na figura uma comparação entre as amostras obtidas da saída do sistema e a saída obtida com o modelo inicial identificado através do processo de agrupamento *fuzzy*. Encontra-se também na mesma figura uma comparação entre as amostras obtidas da saída do sistema e as saídas obtidas com o modelo final otimizado de acordo com o processo apresentado nas seções 3.4 e 3.5.

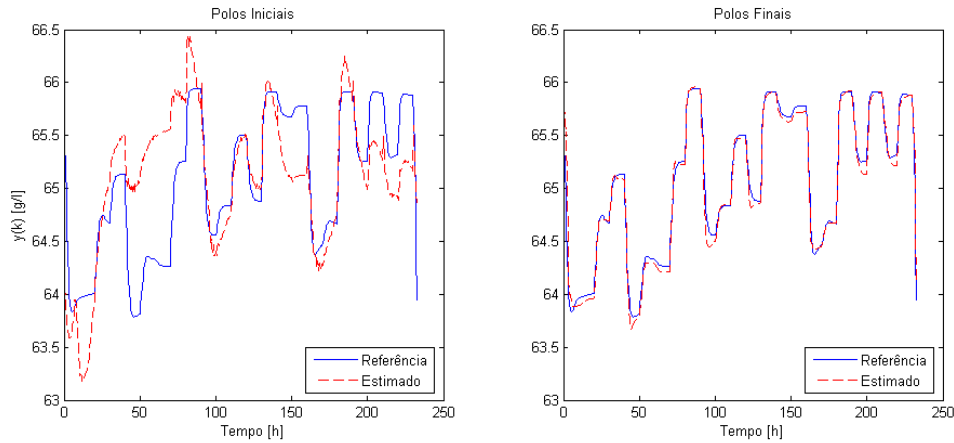


Fig. 3.25: Saídas do modelo *Fuzzy TS-GOBF* - pólos iniciais e otimizados - Sistema de Etanol.

3.9.4 Modelo Racional NARX

Neste exemplo utiliza-se um sistema não-linear simulado que exhibe comportamentos dinâmicos distintos em diferentes regiões do intervalo de operação (Lazar, 2001). Este sistema é dado por um modelo não-linear representado na equação (3.74) que relaciona entradas e saídas em instantes passados de tempo:

$$y(k) = \frac{2,5y(k-1)y(k-2)}{1 + y(k-1)^2 + y(k-2)^2} + 0,3 \cos(0,5(y(k-1) + y(k-2))) + 1,2u(k-1) \quad (3.74)$$

Para a realização do processo de identificação e posterior validação do modelo utilizou-se dois conjuntos de dados com cada um deles dispondo de 300 pares de amostras de entrada e saída com as amostras de entrada variando entre 0 e 2 em degraus aleatórios com distribuição uniforme de duração de 20 amostras. A figura 3.26 mostra os dados de entrada e saída utilizados na estimação dos modelos.

Conforme se procedeu nos exemplos anteriores, utilizou-se 3 tipos de modelos para se efetuar a modelagem do sistema não-linear apresentado na equação (3.74), sendo eles, um modelo *fuzzy TS-GOBF*, um modelo de Volterra-GOBF de segunda ordem com *kernel* de segunda ordem simétrico e

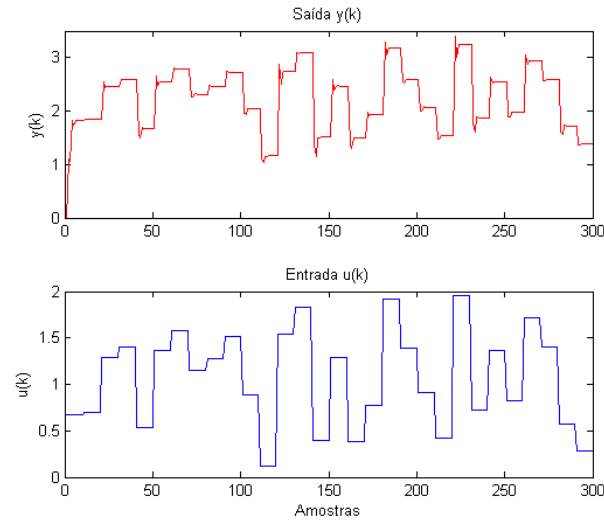


Fig. 3.26: Dados utilizados na estimação do Sistema NARX.

um modelo de Volterra-GOBF de segunda ordem com *kernel* não simétrico empregando em todos os modelos bases de funções ortonormais generalizadas com funções internas conforme apresentado na seção 2.3. Para a modelagem do sistema sob análise variou-se, para os modelos empregados, o número de polos e de funções ortonormais utilizadas a fim de se obter o melhor modelo sem conhecimento a priori da dinâmica deste. Inicializou-se os coeficientes dos *kernels* com valores obtidos através de regressão por mínimos quadrados entre a saída e o vetor de entrada dado pelas saídas das funções ortonormais. Na tabela 3.8 encontram-se os melhores resultados para os modelos com as abordagens propostas, onde, apresentam-se o tipo dos modelos com os polos empregados inicialmente no processo de estimação, o EQM do modelo inicial (EQMi), os polos do modelo otimizado e seus respectivos EQM's (EQMf) (EQM's obtidos a partir de dados de validação).

Tab. 3.8: Valores iniciais e finais dos EQM's e polos dos modelos empregados - Sistema NARX.

Tipo de modelo	Polos iniciais	EQMi	Polos finais	EQMf
Volterra-GOBF	$p_1 = 0,5000$ $p_2 = 0,50 \pm 0,50i$	$7,703 \cdot 10^{-3}$	$p_1 = 0,1454$ $p_2 = -0,5192 \pm 0,5050i$	$1,073 \cdot 10^{-3}$
Volterra-GOBF Direções Independentes	$p_1 = 0,5000$ $p_{21} = 0,5000i$ $p_{22} = 0,50 \pm 0,50i$	$7,1817 \cdot 10^{-3}$	$p_1 = 0,1458$ $p_{21} = 0,4659$ $p_{22} = -0.5103 \pm 0.6884i$	$8,3425 \cdot 10^{-4}$
Fuzzy TS-BFO GOBF int.	$p_1 = 0,50 \pm 0,50i$	$4,805 \cdot 10^{-3}$	$p_1 = 0,0447 \pm 0,1094i$	$1,6827 \cdot 10^{-4}$

A tabela 3.9 apresenta os valores dos parâmetros γ_i para cada um dos modelos identificados e apresentados na tabela 3.8.

Tab. 3.9: Valores iniciais e finais dos Parâmetros dos modelos GOBF - Sistema NARX.

Tipo de modelo	Param. iniciais	Param. finais
Volterra-GOBF	$\gamma_1 = 0,5000$ $\Upsilon_2 = [0,6667 \quad -0,5000]$	$\gamma_1 = 0,1454$ $\Upsilon_2 = [0,6811 \quad -0,5246]$
Volterra-GOBF Direções Independentes	$\gamma_1 = 0,5000$ $\Upsilon_{21} = 0,5000$ $\Upsilon_{22} = [0,6667 \quad -0,5000]$	$\gamma_1 = 0,1458$ $\Upsilon_{21} = 0,4659$ $\Upsilon_{22} = [-0,5885 \quad -0,7343]$
Fuzzy TS-GOBF Funções int.	$\Upsilon = [0,6667 \quad -0,5000]$	$\Upsilon = [0,7531 \quad 0,0788]$

Na tabela 3.10 encontram-se o número de funções ortonormais (NF) utilizadas em cada *kernel* para os modelos de Volterra-GOBF ou nos consequentes das regras dos modelos *fuzzy* TS-GOBF, o número de polos (NP), o número de regras no modelo *fuzzy* (NR) e o número de parâmetros estimados e otimizados que se emprega em cada modelo apresentado na tabela 3.8. A abreviação ke_{ij} indica o número de funções presentes da j -ésima direção do i -ésimo *kernel*.

Tab. 3.10: Informação dos modelos do Sistema NARX.

Tipo de modelo	NF	NP	NR	Nº de Param.
Volterra-GOBF	$ke_1 = 2$ $ke_2 = 2$	1 2	-	9
Volterra-GOBF Direções Independentes	$ke_{11} = 2$ $ke_{21} = 2$ $ke_{22} = 2$	1 1 2	-	10
Fuzzy TS-BFO GOBF int.	4	2	4	22

Analisando-se a tabela 3.8 pode-se verificar que os resultados indicam uma melhoria na precisão dos modelos otimizados com a diminuição no valor dos EQM's entre a saída do sistema real e a saída estimada pelos modelos, o que mostra a eficácia da metodologia proposta. Dentre os modelos analisados, o que melhor se adequou a representação do sistema foi o modelo TS-GOBF com funções internas com 2 polos e 4 funções nos consequentes das regras. Através da análise da tabela 3.10 pode-se verificar que o modelo *fuzzy* TS-BFO apresenta um maior número de parâmetros que os dois modelos de Volterra que apresentam um número equivalente de parâmetros. O modelo *fuzzy* empregado apresenta 4 regras com uma variável no antecedentes das regras.

A figura 3.27 apresenta a saída do modelo *fuzzy* TS-GOBF da tabela 3.8 para os dados de validação do modelo com uma comparação entre as amostras obtidas da saída do sistema e aquela obtida com o modelo inicialmente identificado através do processo de agrupamento *fuzzy* descrito na seção 3.3.

Apresenta-se na mesma figura uma comparação entre as amostras da saída do sistema e do modelo final otimizado de acordo com o processo apresentado nas seções 3.4 e 3.5. As figuras mostram a melhoria na resposta do modelo otimizado, comprovando a eficiência dos métodos de otimização propostos.

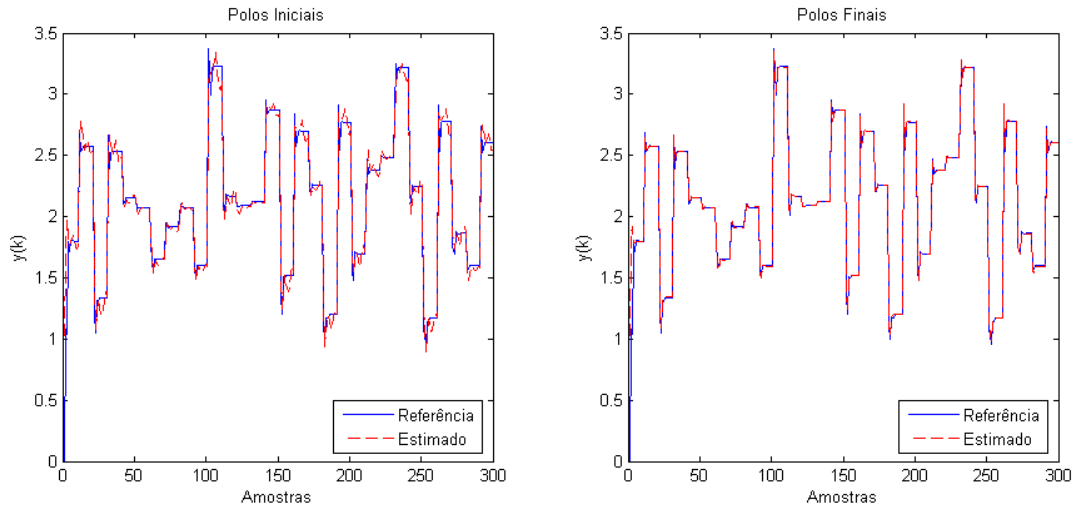


Fig. 3.27: Saídas do modelo *Fuzzy TS-GOBF* - pólos iniciais e otimizados - Sistema NARX.

3.10 Conclusão

Este capítulo teve como principal objetivo apresentar conceitos fundamentais do projeto de modelos com base de funções ortonormais generalizadas empregados na modelagem de sistemas não-lineares. Para a modelagem de sistemas não-lineares com GOBF's empregou-se modelos de Volterra-GOBF com *kernels* simétricos e não simétricos e modelos *fuzzy* Takagi-Sugeno com funções ortonormais generalizadas, com funções internas, nos consequentes das regras (TS-GOBF). As funções ortonormais generalizadas com funções internas são as mesmas apresentadas na seção 2.3. Os modelos propostos agregam as vantagens estruturais apresentadas para modelos lineares GOBF com funções internas (ver seção 2.3), além de apresentarem as características de aproximadores de funções, inerentes aos modelos de Volterra ou *fuzzy TS*. Assim, são eficazes na modelagem de sistemas dinâmicos não-lineares, pois são modelos mais simples estruturalmente quando comparados a metodologias de modelagem de sistemas dinâmicos não-lineares hoje utilizadas, como os modelos polinomiais NARX e NARMAX, ou modelos tradicionais de Volterra. Para as duas representações apresentou-se uma abordagem *bottom-up* em que os modelos são obtidos utilizando-se amostras de entrada e saída do sistema sob análise sem a necessidade de informação *a priori* da dinâmica do sistema.

A técnica de obtenção do modelo *fuzzy* TS-GOBF consistiu de uma abordagem *bottom-up* onde primeiramente se determinou um modelo inicial a partir do agrupamento *fuzzy* dos dados utilizados nos antecedentes das regras do modelo e da determinação, através de critérios de avaliação, do número apropriado de grupos (regras) a serem utilizados. Destacam-se como contribuição neste trabalho os processos de otimização utilizados para se simplificar os modelos inicialmente obtidos, eliminando-se e fundindo-se grupos, conjuntos e regras *fuzzy* redundantes além da sintonia fina dos parâmetros do modelo final obtido. A sintonia fina do modelo TS-GOBF final se deu através de métodos de otimização não-linear, que minimizam uma função custo onde se avaliou o somatório do erro quadrático da saída do modelo com relação à saída obtida diretamente do sistema. Para a implementação dos métodos de otimização, desenvolveu-se o cálculo analítico dos gradientes da saída dos modelos com relação aos parâmetros de interesse. Desta maneira obtiveram-se modelos mais eficientes do ponto de vista da precisão otimizando-se os polos do sistema, as funções de pertinência nos antecedentes das regras e os coeficientes da expansão do modelo GOBF presentes nos consequentes das regras.

Para a identificação de modelos Volterra-GOBF empregou-se técnicas de otimização não-linear que minimizam uma função de custo que leva em consideração o erro quadrático da saída do modelo em comparação à saída real do sistema sob análise. Para a determinação da direção de busca dos algoritmos de otimização utilizou-se o gradiente da saída do modelo com relação aos parâmetros de interesse. Nesta tese apresentou-se como contribuição o cálculo analítico dos gradientes da saída dos modelos Volterra-GOBF com funções internas com relação aos parâmetros de interesse tanto para modelos com *kernels* simétricos como para modelos com *kernels* não simétricos.

Para ilustrar a eficácia dos métodos propostos para a modelagem de sistemas não-lineares e dos processos de otimização apresentou-se a aplicação na modelagem de 5 sistemas, dois reais e três simulados. Os resultados mostraram como, sem nenhum conhecimento prévio sobre a dinâmica dos sistemas identificados, foi possível obter modelos com grande precisão, modelando de maneira eficiente os sistemas analisados e especialmente para os dois primeiros casos, recuperar a informação dos polos dos sistemas e de suas dinâmicas. Analisando-se os resultados para os 3 últimos estudos de caso verificou-se uma maior eficiência dos modelos *fuzzy* TS-GOBF. Contudo, a partir de tais resultados não se pode afirmar que estes modelos são sempre superiores, pois muitas vezes a qualidade do modelo está diretamente ligada à não-linearidade característica do sistema analisado.

Capítulo 4

Identificação de modelos BFO no domínio da frequência

4.1 Introdução

A identificação de modelos no domínio da frequência é uma área que tem crescido e atraído cada vez mais pesquisadores e engenheiros. Devido à maior disponibilidade de analisadores comerciais de frequência, tem-se empregado a determinação experimental da resposta em frequência de processos dinâmicos no projeto de modelos ou de sistemas de controle para estes processos. A identificação de sistemas baseada no domínio da frequência se destaca, em relação à identificação no domínio do tempo, por apresentar vantagens, como a redução de ruído, redução do número de dados, entre outros que serão abordados na próxima seção (Schoukens e Pintelon, 1991). Esta abordagem possui aplicação na identificação de modelos em áreas da engenharia como sistemas de controle, sistemas de comunicação e sistemas de potência. O objetivo do processo de otimização na frequência é desenvolver modelos com desempenho garantido em determinadas regiões do espectro de frequências, abordagem útil para modelos para controle (Schoukens e Pintelon, 1991).

Este capítulo tem como objetivo apresentar o processo de identificação de modelos BFO lineares no domínio da frequência. Este problema já foi abordado na literatura em trabalhos como (Vries e Van den Hof, 1998; Akçay e Heuberger, 2001; Heuberger et al., 2005; Deschrijver e Dhaene, 2006; Reginato e Oliveira, 2008). Nestes trabalhos os autores abordam o problema de forma a determinar os coeficientes da expansão da série de funções ortonormais ou os parâmetros de realizações em funções de transferência de modelos BFO. Nesta tese, deseja-se determinar os polos das funções ortonormais presentes na base, assim como os coeficientes da expansão da BFO utilizando somente informações da resposta do sistema e do modelo no domínio da frequência.

Neste processo de modelagem são utilizadas, assim como para a análise temporal, técnicas de

otimização não-linear que minimizam uma função de custo que considera o erro entre a resposta em frequência do sinal amostrado diretamente do processo e a resposta em frequência do sinal do modelo. Para a implementação do processo de otimização no domínio da frequência é necessário o cálculo das direções de busca dos parâmetros de interesse. Assim, as contribuições deste capítulo inserem-se no desenvolvimento do cálculo analítico dos gradientes das funções de Laguerre e Kautz, no domínio da frequência, utilizados no processo de otimização dos modelos. A metodologia empregada para se realizar a otimização dos modelos BFO no domínio da frequência são apresentados a seguir.

4.2 Identificação na frequência

Além da identificação de sistemas dinâmicos no domínio do tempo, outra abordagem empregada para a modelagem de sistemas dinâmicos é realizar sua identificação no domínio da frequência (Ljung, 1993; Ljung, 1999; Pintelon e Schoukens, 2001). A abordagem de identificação na frequência apresenta algumas características importantes quando comparadas à identificação no tempo (Ljung, 1993):

- Pré-filtragem - Este é um procedimento conhecido e também utilizado na identificação no tempo para se eliminar ruídos indesejáveis nos dados usados na identificação dos modelos. Para dados no domínio da frequência este procedimento é bem simples, basta garantir a banda de frequência em que se deseja que o modelo tenha uma performance garantida ao identificá-lo ou estabelecer pesos diferentes para as frequências nas quais o modelo será identificado.
- Redução na dimensionalidade dos dados - Existem sistemas que demandam a aquisição de uma grande massa de dados para a sua identificação no domínio do tempo. Quando estes dados são convertidos para o domínio da frequência eles podem ser facilmente reduzidos em número de frequências normalizadas (ω_k) menor do que o número de dados.
- Combinação de experimentos - Como a identificação na frequência é feita em diferentes frequências (ω_k), pode-se combinar diferentes experimentos que envolvam as diferentes frequências de interesse.

Para a identificação na frequência, o sinal aplicado como entrada no sistema deve apresentar persistência nas frequências onde deseja-se conhecer a resposta do sistema. Sendo assim, alguns sinais são tipicamente aplicados ao sistema para que seja realizada sua identificação no domínio da frequência:

- **Ruído branco**

Uma escolha simples é gerar um sinal do tipo ruído branco gaussiano, e se necessário, filtrá-lo com filtro linear para, através de uma escolha adequada do filtro, atingir os espectros de sinal adequados.

- **Sinal binário aleatório**

Um sinal binário aleatório deriva de um processo aleatório que assume apenas dois valores. A maneira mais simples de se obter tal sinal é gerando-se um sinal tipo ruído branco com média zero e utilizar apenas o sinal (positivo ou negativo) dos dados obtidos. Posteriormente pode-se adequar a amplitude de sinal conforme desejado. Porém, tomando-se somente sinal (positivo ou negativo) do sinal obtido, o espectro de frequência é alterado.

- **Sinal binário pseudo-aleatório - PRBS**

Um sinal binário pseudo-aleatório é um sinal periódico, determinístico, com propriedades de ruído branco. É obtido através do seguinte algoritmo:

$$u(k) = \text{res}(A(z)u(k), 2) = \text{res}(a_1 u(k-1) + \dots + a_n u(k-n), 2) \quad (4.1)$$

onde $\text{res}(x, 2)$ é o resíduo de x dividido por 2 e $u(k)$ assume valores entre 0 e 1. O vetor de entradas passadas $[u(k-1) \dots u(k-n)]$ apresenta apenas 2^n valores diferentes e a sequência $u(k)$ é periódica com período máximo $M_n = 2^n - 1$.

- **Múltiplas senoides**

Um sinal do tipo múltiplas senoides é formado por uma soma de senoides de diferentes frequências:

$$u(k) = \sum_{i=1}^{\mathcal{N}} a_i \text{sen}(\omega_i + \nu_i) \quad (4.2)$$

A escolha adequada de $\mathcal{N}$, a_i e ω_i permite definir a energia do sinal precisamente nas frequências desejadas.

- **Sinal *chirp***

O sinal *chirp* é uma senoide com frequência variável em uma determinada faixa Ω : $\omega_1 \leq \omega \leq \omega_2$ sobre um certo período de tempo

Na literatura existem várias metodologias para se identificar sistemas dinâmicos na frequência obtendo-se modelos paramétricos e/ou não paramétricos (Pintelon e Schoukens, 2001; Aguirre, 2007). Nesta tese tem-se especial interesse pela identificação de sistemas dinâmicos por meio de BFO utilizando-se a abordagem no domínio da frequência. Em trabalhos como (Vries e Van den Hof, 1998; Akçay e Heuberger, 2001; Heuberger et al., 2005; Deschrijver e Dhaene, 2006; Reginato e Oliveira, 2008; Reginato, 2008) propõe-se a identificação de modelos BFO no domínio da frequência, onde se faz a identificação de modelos lineares obtendo-se, em geral, os coeficientes da expansão da série de funções ortonormais, com o polos fixos. Em (Reginato e Oliveira, 2008; Reginato, 2008) os autores utilizam métodos de otimização baseados em algoritmos de nuvem de partículas para se otimizar os polos da GOBF e métodos de otimização não-lineares para se obter os coeficientes da expansão.

Nesta tese deseja-se utilizar modelos paramétricos lineares de BFO's e otimizá-los no domínio da frequência, obtendo-se simultaneamente os valores ótimos dos polos e dos coeficientes da expansão da série de funções ortonormais. Assim, conforme feito para o processo de otimização dos parâmetros dos modelos com funções ortonormais no tempo, deseja-se realizar, no domínio da frequência, a minimização de uma determinada função de custo J que envolva o erro quadrático entre os módulos das respostas em frequência da saída observada e da saída prevista, para se otimizar os parâmetros de modelos BFO (polos e coeficientes da expansão):

$$\min_{\theta} J = \frac{1}{2} \sum_{\omega \in W} \left[\Lambda_{\omega} (\|Y(\omega)\| - \|\hat{Y}(\omega)\|) \right]^2 \quad (4.3)$$

onde $W = \{\omega_1, \dots, \omega_N\}$ é um conjunto ordenado de N valores de frequências bem definidos e nos quais se deseja otimizar o modelo; Λ_{ω} é um fator de ponderação para as frequências que se deseja dar maior ou menor ênfase na otimização, $\theta = [\Upsilon^T \ C_f^T]^T$, onde Υ representa o polo p dos modelos de Laguerre, os parâmetros b e c dos modelos de Kautz ou os parâmetros $\gamma_i, i = 1, \dots, n_b$ para as funções internas nos modelos GOBF e C_f representa os coeficientes da expansão da série da BFO. Os termos $Y(\omega)$ e $\hat{Y}(\omega)$ representam, respectivamente, a resposta em frequência do sistema analisado e do modelo que se deseja otimizar. A resposta em frequência da saída observada pode ser obtida diretamente através da aplicação de uma transformada rápida de Fourier (Oppenheim et al., 1999) (FFT - Fast Fourier Transform) ao sinal $y(k)$ amostrado no tempo.

A saída estimada no tempo $\hat{y}(k)$ pode ser representada como:

$$\hat{y}(k) = C_f \phi(k) * u(k)$$

onde $C_f = [g_1 \ g_2 \ \dots \ g_{n_f}]$, $\phi(k)$ é a resposta impulsiva de um filtro ortonormal, como definida na equação (2.1), $u(k)$ é o sinal de entrada do sistema e o operador “*” define a operação de convolução no tempo. Como este é um processo de convolução no tempo, seu equivalente no domínio da frequência pode ser visto como um processo de multiplicação:

$$\hat{Y}(\omega) = C_f \Phi(\omega) U(\omega) \quad (4.4)$$

Considerando-se $u(k)$ como o sinal de entrada amostrado diretamente do sistema, sua representação em frequência pode ser obtida diretamente pelo cálculo de sua transformada de Fourier - $U(\omega)$. Com a utilização da transformada de Fourier do sinal $u(k)$ e com a resposta em frequência do modelo, dada pela equação (4.4), o problema formulado inicialmente em (4.3) pode ser reescrito como:

$$\min_{\theta} J = \frac{1}{2} \sum_{\omega \in W} [\Lambda_{\omega}(\|Y(\omega)\| - \|C_f \Phi(\omega) U(\omega)\|)]^2 \quad (4.5)$$

onde $\Phi(\omega)$ é a resposta em frequência do modelo BFO.

A norma de um número complexo x é dada por:

$$\|x\| = \sqrt{x \bar{x}}$$

onde $\bar{x}$ é o complexo conjugado de x . Desta maneira, $\|\hat{Y}(\omega)\|$ pode ser reescrito como:

$$\begin{aligned} \|\hat{Y}(\omega)\| &= \|C_f \Phi(\omega) U(\omega)\| \\ \|\hat{Y}(\omega)\| &= \sqrt{(C_f \Phi(\omega) U(\omega)) \overline{(C_f \Phi(\omega) U(\omega))}} \\ \|\hat{Y}(\omega)\| &= \sqrt{(C_f \Phi(\omega) U(\omega)) (C_f \overline{\Phi(\omega)} \overline{U(\omega)})} \end{aligned} \quad (4.6)$$

Para a solução do problema de otimização apresentado em (4.3) deve-se obter a derivada da função de custo com relação aos parâmetros de interesse. Calculando-se a derivada de (4.3) com relação aos parâmetros θ pode-se encontrar:

$$\nabla_{\theta} J = \sum_{\omega \in W} \Lambda_{\omega}(\|\hat{Y}(\omega)\| - \|Y(\omega)\|) \nabla_{\theta} \|\hat{Y}(\omega)\| \quad (4.7)$$

onde

$$\nabla_{\theta} \|\hat{Y}(\omega)\| = \left[\nabla_{\Upsilon}^T \|\hat{Y}(\omega)\| \ \nabla_{C_f}^T \|\hat{Y}(\omega)\| \right]^T$$

Como definido anteriormente, os parâmetros de interesse com relação aos quais se deseja otimizar a função de custo são $\theta = [\Upsilon^T \ C_f^T]^T$. Com $\|\hat{Y}(\omega)\|$ definido em (4.6) a sua derivada com relação ao

parâmetro Υ , correspondente ao(s) polo(s) da BFO, é dada por:

$$\nabla_{\Upsilon} \|\hat{Y}(\omega)\| = \text{Re} \left[\frac{C_f(\nabla_{\Upsilon} \Phi(\omega)) U(\omega) \bar{\hat{Y}}(\omega) + \hat{Y}(\omega) C_f(\nabla_{\Upsilon} \bar{\Phi}(\omega)) \bar{U}(\omega)}{2\sqrt{Y(\omega) \bar{\hat{Y}}(\omega)}} \right] \quad (4.8)$$

Note que somente interessa conhecer a parcela real da derivada $\nabla_{\Upsilon} \|\hat{Y}(\omega)\|$, já que Υ é formada somente por parâmetros reais e a função de custo também é uma função real. O cálculo da equação (4.8) é comum para todas as BFO's diferindo-se somente pelo cálculo de $\nabla_{\Upsilon} \Phi(\omega)$, onde, $\Phi(\omega) = [\phi_1(\omega) \ \phi_2(\omega) \ \dots \ \phi_{n_f}(\omega)]^T$.

Para as funções de Laguerre (eq. (2.6)) o gradiente $\nabla_{\Upsilon} \Phi(\omega)$, onde $\Upsilon = p$, é dado por:

$$\nabla_p \Phi(\omega) = \left[\frac{\partial \phi_1(\omega)}{\partial p} \ \dots \ \frac{\partial \phi_{n_f}(\omega)}{\partial p} \right]^T$$

onde $\frac{\partial \phi_i(\omega)}{\partial p}$ é dado por:

- Para $i = 1$;

$$\frac{\partial \phi_1(\omega)}{\partial p} = \frac{1}{e^{j\omega} - p} \left(\frac{-p}{\sqrt{1-p^2}} + \frac{\sqrt{1-p^2}}{e^{j\omega} - p} \right) \quad (4.9)$$

- Para $i \geq 2$;

$$\begin{aligned} \frac{\partial \phi_i(\omega)}{\partial p} = & \frac{1}{e^{j\omega} - p} \left(\frac{-p}{\sqrt{1-p^2}} + \frac{\sqrt{1-p^2}}{e^{j\omega} - p} \right) \left(\frac{1 - pe^{j\omega}}{e^{j\omega} - p} \right)^{i-1} + \\ & (i-1) \left(\frac{\sqrt{1-p^2}}{e^{j\omega} - p} \right) \left(\frac{-e^{j\omega}(e^{j\omega} - p) + (1 - pe^{j\omega})}{(e^{j\omega} - p)^2} \right) \left(\frac{1 - e^{j\omega}p^2}{e^{j\omega} - p} \right)^{i-2} \end{aligned} \quad (4.10)$$

Para o cálculo de $\nabla_p \bar{\Phi}(\omega)$ utiliza-se as equações (4.9) e (4.10) e substitui-se $e^{j\omega}$ por $e^{-j\omega}$.

Para os modelos de Kautz (equações (2.7)), tem-se que $\Upsilon = [b \ c]$, onde b e c são os parâmetros que definem os polos das funções de Kautz. Os gradientes $\nabla_b \Phi(\omega)$ e $\nabla_c \Phi(\omega)$ são dados por:

$$\nabla_b \Phi(\omega) = \left[\frac{\partial \phi_1(\omega)}{\partial b} \ \dots \ \frac{\partial \phi_{n_f}(\omega)}{\partial b} \right]^T \quad \text{e} \quad \nabla_c \Phi(\omega) = \left[\frac{\partial \phi_1(\omega)}{\partial c} \ \dots \ \frac{\partial \phi_{n_f}(\omega)}{\partial c} \right]^T$$

onde $\frac{\partial \phi_i(\omega)}{\partial b}$ é dado por:

- Para $i = 1$;

$$\frac{\partial \phi_{2i-1}(\omega)}{\partial b} = \frac{\partial \phi_1(\omega)}{\partial b} = -\frac{\sqrt{1-c^2}}{e^{2j\omega} + b(c-1)e^{j\omega} - c} - \frac{\sqrt{1-c^2}(e^{j\omega} - b)(c-1)e^{j\omega}}{(e^{2j\omega} + b(c-1)e^{j\omega} - c)^2} \quad (4.11)$$

e

$$\frac{\partial \phi_{2i}(\omega)}{\partial b} = -\frac{b(1-c^2)}{\sqrt{(1-b^2)(1-c^2)}(e^{2j\omega} + b(c-1)e^{j\omega} - c)} - \frac{\sqrt{(1-b^2)(1-c^2)}(c-1)e^{j\omega}}{(e^{2j\omega} + b(c-1)e^{j\omega} - c)^2} \quad (4.12)$$

- Para $i \geq 2$;

$$\frac{\partial \phi_{2i-1}(\omega)}{\partial b} = \frac{\partial \phi_1(\omega)}{\partial b} \phi(\omega)^{i-1} + \phi_1(\omega) \frac{\partial \phi(\omega)}{\partial b} (i-1) \phi^{i-2}(\omega) \quad (4.13)$$

e

$$\frac{\partial \phi_{2i}(\omega)}{\partial b} = \frac{\partial \phi_2(\omega)}{\partial b} \phi(\omega)^{i-1} + \phi_2(\omega) \frac{\partial \phi(\omega)}{\partial b} (i-1) \phi^{i-2}(\omega) \quad (4.14)$$

onde

$$\phi(\omega) = \frac{-ce^{2j\omega} + b(c-1)e^{j\omega} + 1}{e^{2j\omega} + b(c-1)e^{j\omega} - c} \quad (4.15)$$

e

$$\frac{\partial \phi(\omega)}{\partial b} = -\frac{(c-1)e^{j\omega}}{e^{2j\omega} + b(c-1)e^{j\omega} - c} - \frac{(-ce^{2j\omega} + b(c-1)e^{j\omega} + 1)(c-1)e^{j\omega}}{(e^{2j\omega} + b(c-1)e^{j\omega} - c)^2} \quad (4.16)$$

As derivadas $\frac{\partial \phi_i(\omega)}{\partial c}$ são dadas por:

- Para $i = 1$;

$$\frac{\partial \phi_{2i-1}(\omega)}{\partial c} = \frac{\partial \phi_1(\omega)}{\partial c} = -\frac{c(e^{j\omega} - b)}{\sqrt{1-c^2}(e^{2j\omega} + b(c-1)e^{j\omega} - c)} - \frac{\sqrt{1-c^2}(e^{j\omega} - b)(be^{j\omega} - 1)}{(e^{2j\omega} + b(c-1)e^{j\omega} - c)^2} \quad (4.17)$$

e

$$\frac{\partial \phi_2(\omega)}{\partial c} = -\frac{c(1-b^2)}{(\sqrt{(1-b^2)(1-c^2)})(e^{2j\omega} + b(c-1)e^{j\omega} - c)} - \frac{\sqrt{(1-b^2)(1-c^2)}(be^{j\omega} - 1)}{(e^{2j\omega} + b(c-1)e^{j\omega} - c)^2} \quad (4.18)$$

- Para $i \geq 2$;

$$\frac{\partial \phi_{2i-1}(\omega)}{\partial c} = \frac{\partial \phi_1(\omega)}{\partial c} \phi(\omega)^{i-1} + \phi_1(\omega) \frac{\partial \phi(\omega)}{\partial c} (i-1) \phi^{i-2}(\omega) \quad (4.19)$$

e

$$\frac{\partial \phi_{2i}(\omega)}{\partial c} = \frac{\partial \phi_2(\omega)}{\partial c} \phi(\omega)^{i-1} + \phi_2(\omega) \frac{\partial \phi(\omega)}{\partial c} (i-1) \phi^{i-2}(\omega) \quad (4.20)$$

onde $\phi(\omega)$ é a mesma apresentada em (4.15) e:

$$\frac{\partial \phi(\omega)}{\partial c} = -\frac{-e^{2j\omega} + be^{j\omega}}{e^{2j\omega} + b(c-1)e^{j\omega} - c} - \frac{(-ce^{2j\omega} + b(c-1)e^{j\omega} + 1)(be^{j\omega} - 1)}{(e^{2j\omega} + b(c-1)e^{j\omega} - c)^2} \quad (4.21)$$

Para os cálculos de $\frac{\partial \overline{\phi_{2i-1}}(\omega)}{\partial b}$, $\frac{\partial \overline{\phi_{2i}}(\omega)}{\partial b}$, $\frac{\partial \overline{\phi_{2i-1}}(\omega)}{\partial c}$, $\frac{\partial \overline{\phi_{2i}}(\omega)}{\partial c}$ utilizam-se as equações de (4.11) a (4.21) e substitui-se $e^{j\omega}$ por $e^{-j\omega}$.

Com relação aos parâmetros dos coeficientes da expansão da série de funções ortonormais, o gradiente de $\|\hat{Y}(\omega)\|$ é dado por:

$$\nabla_{C_f} \|\hat{Y}(\omega)\| = \Lambda_\omega \|\Phi(\omega)U(\omega)\| \quad (4.22)$$

Os modelos GOBF com funções internas têm suas derivadas com relação aos parâmetros de interesse calculadas numericamente. Isto se deve ao fato de que ainda não foi possível se estabelecer de maneira simples e genérica uma realização em função de transferência para modelos GOBF com funções internas de qualquer ordem. Desta maneira, serão utilizados, para a otimização em frequência dos modelos GOBF com funções internas, os algoritmos implementados no programa MATLAB para o cálculo numérico das derivadas e solução dos algoritmos de otimização não-linear (*Optimization Toolbox User's Guide*, 2008). Para a solução dos problemas de otimização no domínio da frequência (eq. 4.3) são utilizados os algoritmos de otimização não-linear apresentados na seção 2.4.2 e implementados através da *toolbox* de otimização do MATLAB.

4.3 Algoritmo

O método proposto para a otimização dos modelos BFO no domínio da frequência pode ser resumido pelos seguintes passos:

1. Atribua valores iniciais de parâmetros θ^0 (normalmente de maneira aleatória ou baseando-se em algum conhecimento prévio) e calcule por MMQ os valores dos coeficientes de C_f .
2. Calcule o gradiente $\nabla_{\theta} J = [\nabla_{\Upsilon}^T J \quad \nabla_{C_f}^T J]^T$ utilizando as equações (4.7), (4.8) e (4.22). Para o cálculo de $\nabla_{\Upsilon} \|\hat{Y}(\omega)\|$ é necessário se calcular $\nabla_p \psi_i(\omega)$ em (4.9) e (4.10) para as funções de Laguerre ou $\nabla_b \psi_i(\omega)$ e $\nabla_c \psi_i(\omega)$ em (4.11) até (4.21) para os modelos de Kautz. Para os modelos GOBF, $\nabla_{\gamma_l} \psi_i(\omega)$ com $\gamma_l, l = 1, \dots, n_b$, deve ser calculado numericamente.
3. Use o valor de $\nabla_{\theta} J$ no algoritmo de otimização de Levenberg-Marquardt para atualizar o valor de parâmetros θ . Nota: inclua o intervalo de factibilidade dos parâmetros do modelo GOBF como restrições dentro do problema de otimização para garantir a factibilidade dos polos e dos modelos otimizados ($0 < p < 1$ para os modelos de Laguerre, a restrição apresentada na

equação (2.54) para os modelos de Kautz ou $-1 < \gamma_l < 1$ com $l = 1, \dots, n_b$ para os modelos GOBF);

4. Volte ao passo 2 até que um critério de parada seja atingido.

As restrições citadas no passo 3 do algoritmo acima são necessárias para evitar que o processo de otimização produza polos instáveis ou que o modelo GOBF perca sua capacidade de representação. Para a otimização dos modelos de Laguerre e Kautz, algoritmo semelhante deve ser seguido utilizando-se as equações apropriadas.

4.4 Resultados

Esta seção tem como objetivo ilustrar a aplicação dos processos de identificação, no domínio da frequência, de sistemas lineares por meio de modelos BFO's que empregam funções de Laguerre, Kautz e GOBF com funções internas. O processo de identificação emprega os métodos propostos de otimização dos parâmetros do modelo utilizando os cálculos de gradientes analíticos apresentados anteriormente para os modelos de Laguerre e Kautz, e cálculos numéricos para os gradientes das GOBF's. Os modelos empregados para o processo de identificação na frequência são os mesmos apresentados na seção 2.6, onde se fez a identificação de modelos lineares empregando modelos com BFO no tempo.

Considere inicialmente um sistema de terceira ordem com um polo real de multiplicidade 3. O sistema em questão é dado pela seguinte função de transferência em z :

$$G(z) = \frac{0,01741}{z^3 - 2,222z^2 + 1,646z - 0,4066} \quad (4.23)$$

com 3 polos em $z = 0,7408$

Para a identificação do sistema utilizam-se amostras temporais de entrada e saída do sistema obtidas através de simulação, como mostra a figura 4.1, onde o sinal de entrada $u(k)$ é um sinal binário pseudo-aleatório - PRBS.

O sinal de saída tem sua transformada de Fourier (para frequências normalizadas) apresentada na figura 4.2.

Para a modelagem e identificação do sistema dinâmico analisado, partiu-se do pressuposto de que as características dinâmicas do sistema eram desconhecidas. Desta maneira, fez-se a identificação de modelos variando-se o número de funções de Laguerre utilizadas. Como os sinais são amostrados no tempo, aplicou-se a FFT aos sinais de entrada e saída para a realização da otimização na frequência. Realizou-se a identificação do sistema em todo o espectro de frequência com $\Lambda_\omega = 1$ para $\omega \in W$.

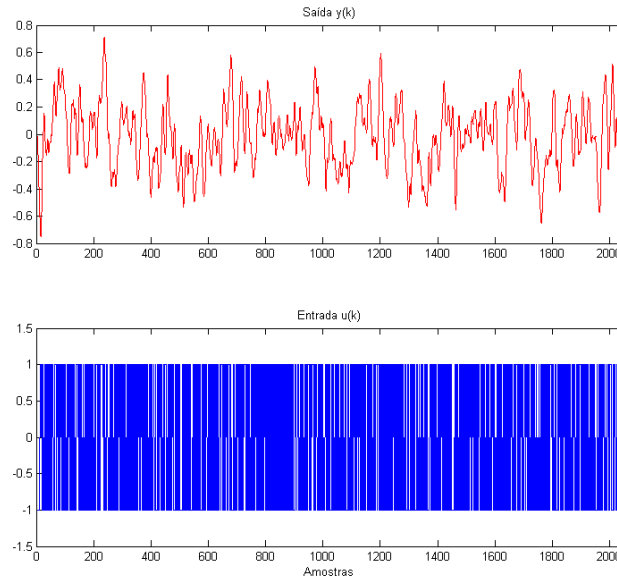


Fig. 4.1: Sinais utilizados na identificação do Modelo de Terceira Ordem.

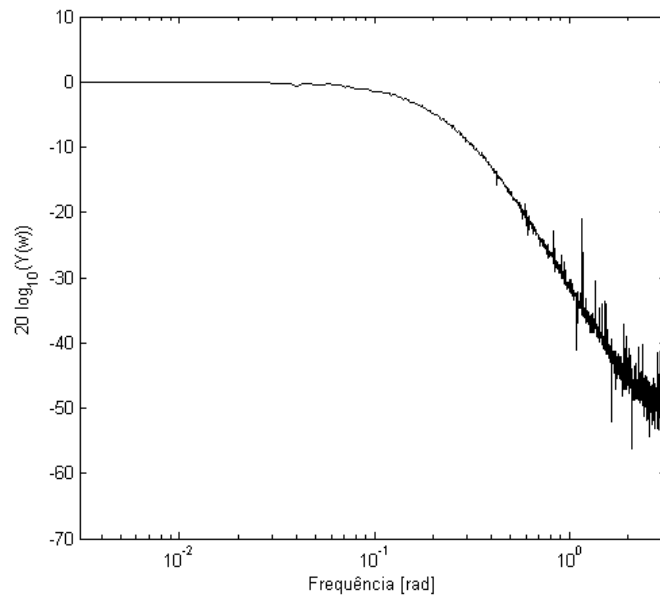


Fig. 4.2: Relação dos módulos da transformada de Fourier discreta dos sinais de saída e entrada.

Para a execução do processo de otimização, realizou-se o cálculo analítico dos gradientes com relação aos parâmetros dos modelos a serem otimizados e os utilizou em algoritmos de otimização não-linear. Para este exemplo, identificou-se o sistema por meio de modelos com funções de Laguerre. A tabela 4.1 apresenta os resultados comparativos entre os modelos de Laguerre otimizados com o erro

quadrático médio - EQM gerado pelos modelos estimados analisados levando-se em conta dados de validação.

Tab. 4.1: Resultados dos modelos de Laguerre para Ident. na frequência.

N. de Funções	Param.	Polo Inicial	Polo Final	EQM
1	2	0,5000	0,9301	12,8404
2	3	0,5000	0,8163	2,0849
3	4	0,5000	0,7410	0,0079
4	5	0,5000	0,7376	0,0112

Também realizou-se a modelagem do mesmo sistema utilizando-se modelos GOBF com funções internas no domínio da frequência. Conforme foi realizado com os modelos de Laguerre, variou-se o número de polos e de funções nos modelos GOBF para se obter o modelo que melhor representasse a dinâmica do sistema. A Tabela 4.2 apresenta os resultados dos modelos GOBF com os valores dos polos iniciais, finais e valores de EQM para saída do modelo no domínio da frequência com dados de validação.

Tab. 4.2: Resultados dos modelos GOBF para Ident. na frequência - polos reais.

N. de polos	N. Blocos G_b	Param.	Polo Inicial	Polo Final	EQM
1	1	2	0,5000	0,9301	12,8206
1	2	3	0,5000	0,8219	2,3111
1	3	4	0,5000	0,7408	$7,9090 \cdot 10^{-12}$
1	4	2	0,5000	0,7406	$5,6512 \cdot 10^{-11}$
3	1	4	$-0,4748 \pm 0,5488i$ 0,9495	$0,7482 \pm 0,0117i$ 0,7260	$4,2934 \cdot 10^{-8}$

Através da análise das tabelas 4.1 e 4.2 pode-se verificar que os métodos propostos foram eficientes na identificação do sistema sob análise. Para os modelos de Laguerre, aquele que apresentou melhor resultado foi o modelo com 3 funções, com o polo do modelo convergindo para um valor próximo do polo do sistema sob análise. Para os modelos GOBF tem-se que a representação com 1 polo e 3 funções (blocos G_b) apresentou melhor resultado com o polo convergindo para um o valor do polo do sistema sob análise. As diferenças entre os resultados entre os modelos de Laguerre e GOBF podem se dar por eventuais problemas numéricos. O modelo GOBF com 3 polos e um bloco G_b apresentou um bom resultado com a sua inicialização sendo realizada nos polos equivalentes aos parâmetros $\gamma_i, i = 1, 2, 3$, com $\Upsilon = [0,5 \ 0,5 \ 0,5]$.

A figura 4.3 mostra a comparação entre a saída estimada do modelo com polos iniciais em 0,5 e finais em 0,7408 para o modelo GOBF com 3 funções aplicado a dados de validação.

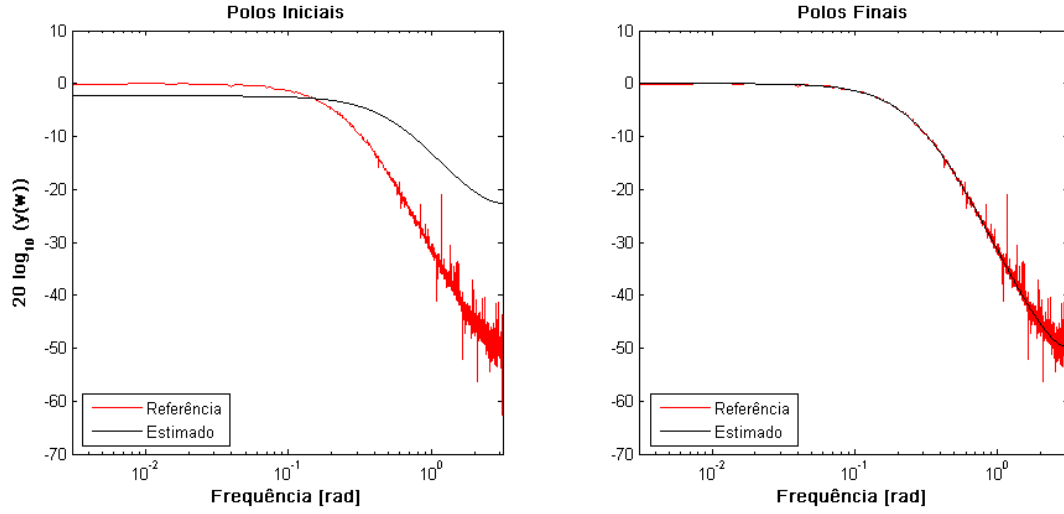


Fig. 4.3: Relação dos módulos da transformada de Fourier discreta dos sinais de saída e entrada do modelo com 1 polo com multiplicidade 3.

Verifica-se através da análise das tabelas 4.1 e 4.2 que os polos otimizados, nos modelos com 4 funções, não são exatamente os polos do sistema como aconteceu quando se otimizou o modelo com 3 funções. Contudo, os polos se encontram na vizinhança dos polos exatos do sistema e apresentam uma boa precisão devido a melhoria na capacidade de aproximação do modelo com mais funções na base.

A seguir, apresenta-se a simulação e a identificação de um sistema com 4 polos, sendo eles dois pares de polos complexos conjugados. A função de transferência em z do sistema é dada pela equação (4.24):

$$G(z) = \frac{0,009119z^2 + 0,01396z + 0,00534}{z^4 - 2,562z^3 + 2,539z^2 - 1,151z + 0,2019} \quad (4.24)$$

cujos polos estão localizados em $z = 0,6404 \pm 0,1981i$ com multiplicidade 2.

Para o processo de identificação gerou-se, através de simulação, uma massa de dados de entrada e saída do sistema proposto, conforme apresentado na figura 4.4. Como entrada foi utilizado um sinal PRBS com variação $[-1, 1]$.

A figura 4.5 apresenta a transformada de Fourier (para frequências normalizadas) do sinal amostrado no tempo da saída do sistema analisado (eq. (4.24)), $y(k) \rightarrow Y(\omega)$.

Assim como no exemplo anterior considerou-se que não havia nenhum conhecimento a priori do sistema que se deseja identificar e realizou-se a identificação do sistema em todo o espectro de frequência com $\Lambda_\omega = 1$ para $\omega \in W$. A tabela 4.3 apresenta os resultados dos modelos com funções de Kautz, onde se variou o número de funções empregadas nos modelos. Na tabela 4.3, por se tratar de funções de Kautz, o número de funções varia em números pares. Como pode-se verificar pela

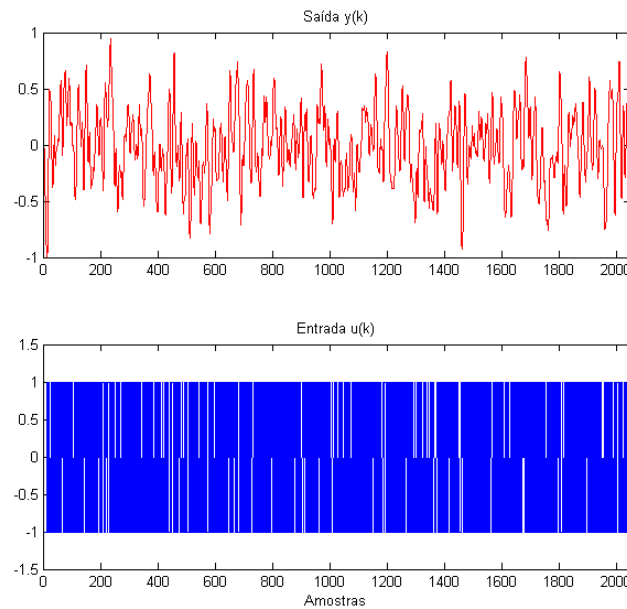


Fig. 4.4: Sinais utilizados na identificação do sistema de quarta ordem.

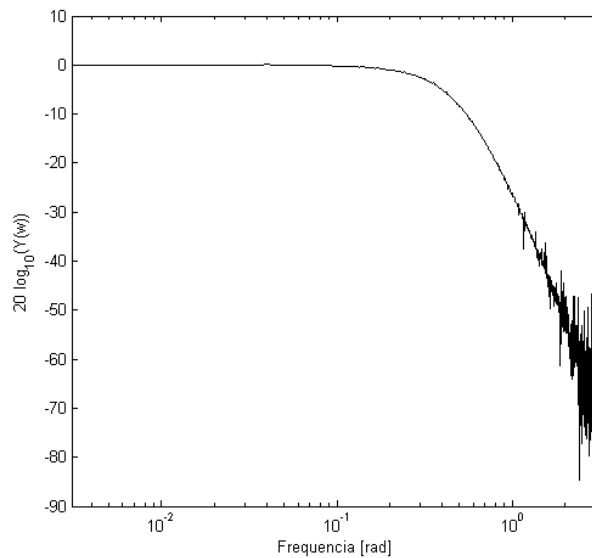


Fig. 4.5: Relação dos módulos da transformada de Fourier discreta dos sinais de saída e entrada do sistema de quarta ordem.

análise da tabela 4.3, o modelo com 4 funções de Kautz apresentou os melhores resultados quando considerada a precisão e o número de parâmetros dos modelos, com os polos do modelo convergindo para valores próximos aos valores dos polos reais do sistema analisado. Além disso, um modelo de

Kautz com 4 funções de Kautz equivale a um modelo de quarta ordem com par de polos complexos conjugados de multiplicidade 2.

Tab. 4.3: Resultados dos modelos Kautz para o segundo exemplo.

N. de Funções	Param.	Polos Iniciais	Polos Finais	EQM
2	4	$0,5 \pm 0,1i$	$0,7512 \pm 0,1660i$	0,6695
4	6	$0,5 \pm 0,1i$	$0,6398 \pm 0,1980i$	0,0023
6	8	$0,5 \pm 0,1i$	$0,6208 \pm 0,2100i$	0,0027

Realizou-se também a modelagem do sistema com 4 polos utilizando-se modelos GOBF com funções internas no domínio da frequência. Para a identificação, variou-se o número de polos e de funções nos modelos GOBF para se obter o modelo que melhor representasse a dinâmica do sistema. A Tabela 4.4 apresenta os resultados dos modelos GOBF com os valores dos polos iniciais, finais e valores de EQM para saída do modelo no domínio da frequência para dados de validação. O modelo com 4 polos teve sua inicialização nos polos equivalentes aos parâmetros $\Upsilon = [0,5 \ 0,5 \ 0,5 \ 0,5]$.

Tab. 4.4: Resultados dos modelos GOBF para Ident. na frequência - polos complexos.

N. de polos	N. Blocos G_b	Param.	Polos Iniciais	Polo Finais	EQM
2	1	4	$0,5 \pm 0,1$	$0,8181 \pm 0,1751i$	6,8830
2	2	6	$0,5 \pm 0,1$	$0,6405 \pm 0,1980i$	$1,2985 \cdot 10^{-4}$
2	3	8	$0,5 \pm 0,1$	$0,6218 \pm 0,2143i$	$1,3997 \cdot 10^{-4}$
4	1	8	$-0,2353 \pm 0,7817i$ $0,9834$ e $-0,7629$	$0,6651 \pm 0,2213i$ $0,6094 \pm 0,1620i$	$1,2774 \cdot 10^{-4}$

Analisando-se também a tabela 4.4, pode-se verificar que o modelo GOBF que apresentou melhores resultados foi o modelo com dois blocos G_b com 2 polos em cada função. Com este modelo foi possível obter polos muito próximos dos polos exatos do sistema sob estudo. Também é possível se verificar que o modelo GOBF com 4 polos distintos e um bloco G_b apresentou bons resultados com os polos otimizados próximos aos polos do sistema analisado.

A figura 4.6 apresenta o resultado da resposta em frequência do modelo Kautz para os polos iniciais e os polos finais após o processo de otimização aplicados a dados de validação.

Para finalizar a aplicação do processo de otimização de modelos lineares por BFO no domínio da frequência, foi simulado e identificado um sistema com 3 polos, sendo um par de polos complexos e um polo real, dado pela seguinte função de transferência em z :

$$G(z) = \frac{0,07287z^2 + 0,1587z + 0,02187}{z^3 - 1,191z^2 + 0,5352z - 0,09072} \quad (4.25)$$

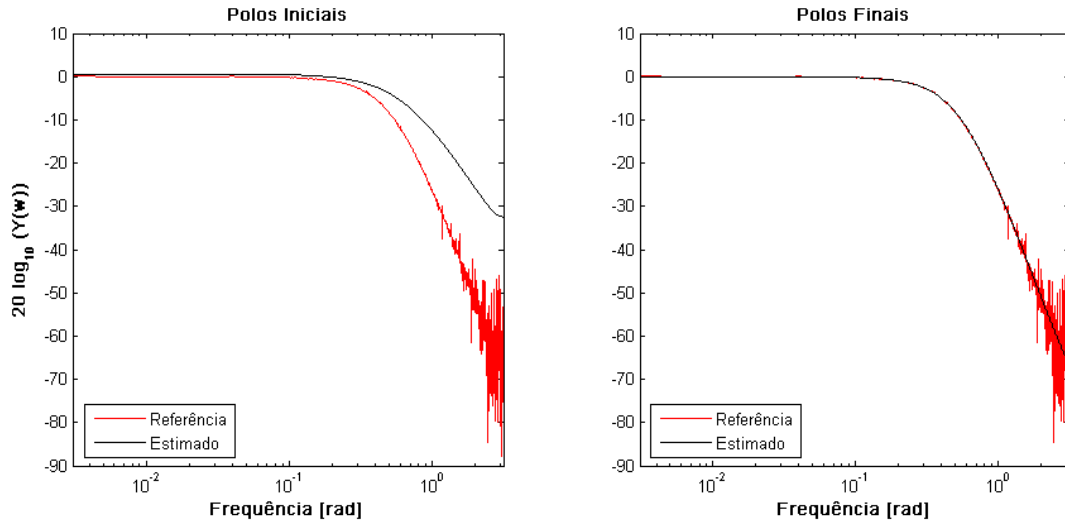


Fig. 4.6: Relação dos módulos da transformada de Fourier discreta dos sinais de saída e entrada dos modelos com 4 funções de Kautz.

cujos polos estão localizados em $z = 0,3708 \pm 0,2537i$ e $z = 0,4493$.

Assim como para os modelos anteriores, para a realização do processo de identificação gerou-se através de simulação uma massa de dados de entrada e saída do sistema proposto. Como entrada foi utilizado um sinal PRBS com variação $[-1, 1]$. A figura 4.7 apresenta a transformada de Fourier (para frequências normalizadas) do sinal amostrado no tempo da saída do sistema analisado (eq. (4.25)), $y(k) \rightarrow Y(\omega)$.

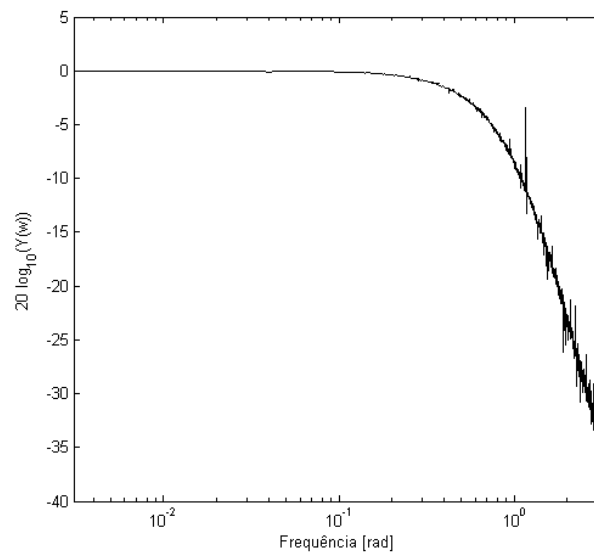


Fig. 4.7: Relação dos módulos da transformada de Fourier discreta dos sinais de saída e entrada do sistema de terceira ordem: Polos complexos + polo real.

Para a identificação do sistema sob análise, variou-se o número de polos e de funções nos modelos GOBF e aplicou-se o processo de otimização apresentado para se obter o modelo que melhor representa a dinâmica do sistema. No processo de identificação considerou-se todas as frequências do espectro de frequência ($\Lambda_\omega = 1$ para $\omega \in W$). A Tabela 4.5 apresenta os resultados dos modelos GOBF com os valores dos respectivos polos iniciais, finais e valores de EQM para saída dos modelos no domínio da frequência aplicados a dados de validação. O modelo GOBF com 3 polos e 1 bloco G_b teve inicialização realizada nos polos equivalentes aos parâmetros $\gamma_i, i = 1, 2, 3$, com $\Upsilon = [0,5 \ 0,5 \ 0,5]$. O modelo com 4 polos teve sua inicialização nos polos equivalentes aos parâmetros $\Upsilon = [0,5 \ 0,5 \ 0,5 \ 0,5]$.

Tab. 4.5: Resultados dos modelos GOBF para Ident. na frequência - múltiplos modos.

N. de polos	N. Blocos G_b	Param.	Polos Iniciais	Polo Finais	EQM
1	1	2	0,5000	0,7661	34,4722
2	1	4	$0,5 \pm 0,5$	$0,5647 \pm 0,2525i$	0,7264
3	1	6	$-0,4748 \pm 0,5488i$ 0,9495	$0,3710 \pm 0,2541i$ 0,4483	$9,0372 \cdot 10^{-6}$
4	1	8	$-0,2353 \pm 0,7817i$ 0,9834 e -0,7629	$0,3709 \pm 0,2531i$ 0,4483 e -0,7479	$8,9023 \cdot 10^{-5}$

Analisando-se também a tabela 4.5, verifica-se que o modelo GOBF que apresentou melhores resultados foi o modelo com 3 polos e um bloco G_b . Este modelo apresentou polos, após o processo de otimização, muito próximos dos polos exatos do sistema sob estudo, ilustrando a eficiência do método de identificação e dos modelos GOBF em representar sistemas com múltiplos modos dominantes. A figura 4.8 apresenta o resultado da resposta em frequência do modelo GOBF para os polos iniciais e os polos finais após o processo de otimização aplicados a dados de validação.

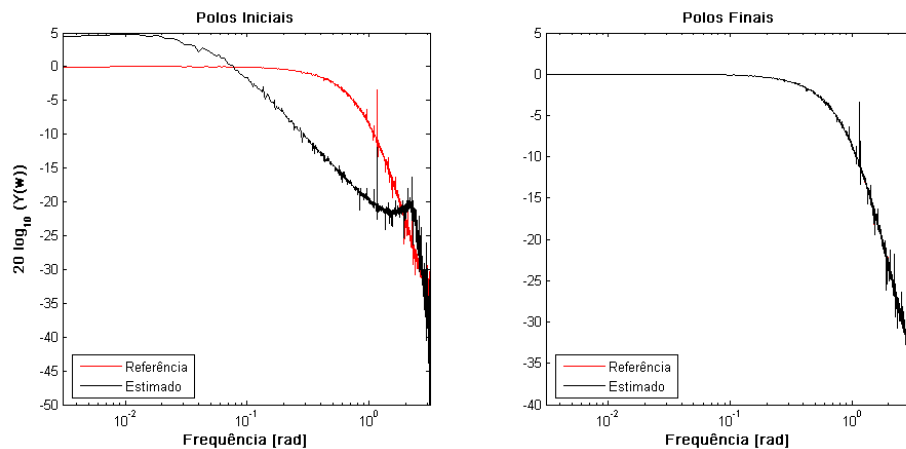


Fig. 4.8: Relação dos módulos da transformada de Fourier discreta dos sinais de saída e entrada do modelo GOBF com 3 polos.

Considere o seguinte sistema dinâmico de quinta ordem apresentado em (Vries e Van den Hof, 1998):

$$G(z) = \frac{0,2530z^4 - 0,9724z^3 + 1,4283z^2 - 0,9493z + 0,2410}{z^5 - 4,1500z^4 + 6,8831z^3 - 5,6871z^2 + 2,3333z - 0,3787} \quad (4.26)$$

cujos polos estão localizados em $z = 0,9509 \pm 0,1982i$, $z = 0,8497 \pm 0,0964i$ e $z = 0,5490$.

A figura 4.10 apresenta a transformada discreta do sinal de saída do modelo 4.26 (em módulo):

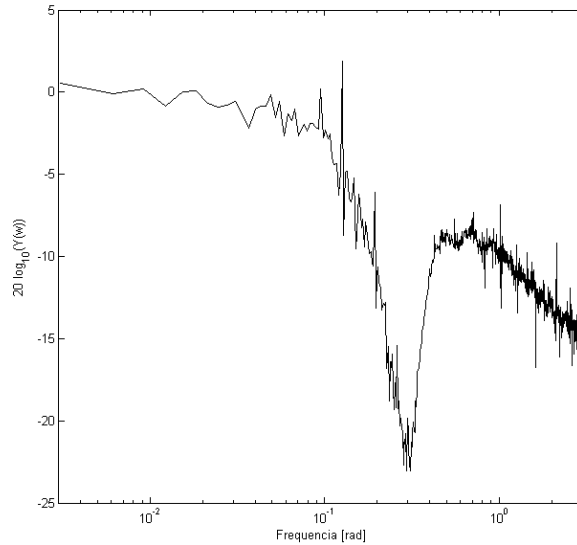


Fig. 4.9: Relação dos módulos da transformada de Fourier discreta dos sinais de saída e entrada do modelo GOBF com 5 polos.

Neste exemplo deseja-se otimizar o modelo que representa o sistema apresentado na equação (4.26) em uma determinada faixa do espectro de frequências. Para a identificação e otimização do modelo de acordo com os parâmetros de projeto considerando-se as frequências de interesse faz-se $\Lambda_\omega = 1$ para as frequências no intervalo entre $0,15 \text{ rad}$ e $0,50 \text{ rad}$ que equivale à região do espectro de frequências que está em um “vale” no módulo da resposta em frequência do sistema analisado. Otimizou-se modelos GOBF variando-se o número de polos com os resultados apresentados na tabela 4.6). O EQM apresentado na tabela 4.6 foi obtido levando-se em consideração todo o espectro de frequência

Como pode-se observar na tabela 4.6 o modelo com 3 polos já foi suficiente para modelar o sistema na região de interesse do espectro de frequência. Os polos iniciais equivalem aos parâmetros $\Upsilon = [0,5 \quad -0,5 \quad 0,5]$ e os polos finais aos parâmetros $\Upsilon = [0,9806 \quad -0,8727 \quad 0,4703]$. A figura 4.10 apresenta o resultado para o modelo otimizado nas frequências pré-estabelecidas utilizando-se dados de validação. Analisando-se o gráfico pode-se verificar que o modelo apresenta resultados com maior precisão nas frequências de interesse do que naquelas fora do intervalo. Pode-se observar

Tab. 4.6: Resultados do modelo GOBF para Ident. na frequência em uma área do espectro de freq.

N. de polos	N. Blocos G_b	Polos Iniciais	Polo Finais	EQMi	EQMf
1	1	0,5000	0,4066	0,0130	0,0125
2	1	$0,5 \pm 0,5i$	0,9455 0,1836	0,0159	0,0043
3	1	$0,1325 \pm 0,8141i$ 0,7350	$0,6799 \pm 0,2606i$ 0,8871	0,0153	$1,2528 \cdot 10^{-4}$
4	1	$-0,2353 \pm 0,7817i$ $0,9834 e -0,7629$	$0,6706 \pm 0,1427i$ $0,8482 e -0,9880$	0,0224	$1,2525 \cdot 10^{-4}$
5	1	$-0,7022 \pm 0,4048i$ $-0,0450 \pm 0,8737i$ 0,9945	$0,6709 \pm 0,1470i$ $-0,8346 \pm 0,4815i$ 0,8494	0,0306	$1,2363 \cdot 10^{-4}$

que nas frequências de interesse o valor do módulo da resposta em frequência tem valor inferior aos valores nas demais frequências e mesmo assim o método conseguiu obter um bom modelo para a aproximação do sistema comprovado a sua eficiência.

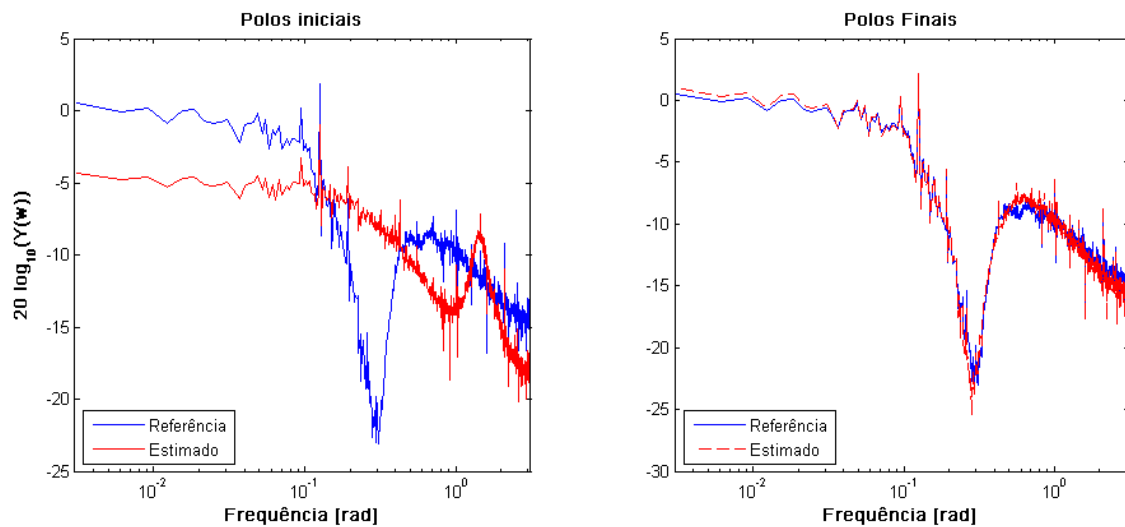


Fig. 4.10: Relação dos módulos da transformada de Fourier discreta dos sinais de saída e entrada do modelo GOBF com 3 polos; Frequência específica 0,15 rad a 0,50 rad.

4.5 Conclusão

Neste capítulo apresentaram-se métodos de otimização de modelos de sistemas dinâmicos lineares por meio de BFO no domínio da frequência. Esta abordagem apresenta vantagens sobre a otimização no tempo, pois possibilita a redução da influência de ruído na estimação dos parâmetros, redução do número de dados e permite desenvolver modelos com desempenho garantido em regiões especificadas do espectro de frequências, sendo assim útil no desenvolvimento de modelos voltados para controle. Destacou-se como principal contribuição o cálculo analítico dos gradientes das funções de Kautz e Laguerre, com relação aos seus parâmetros, no domínio da frequência. Estes cálculos permitiram determinar a direção de busca de tais parâmetros no processo de otimização empregando-se métodos de otimização não-linear. Também foram otimizados os modelos GOBF com funções internas, porém com o cálculo numérico dos gradientes destes modelos, pois ainda não foi possível estabelecer uma fórmula genérica para o cálculo analítico dos gradientes dos modelos GOBF, de qualquer ordem, em relação aos parâmetros de interesse. O desenvolvimento de métodos para o cálculo analítico destes gradientes para modelos GOBF de qualquer ordem será tema de estudo em futuros trabalhos.

Para ilustrar a eficiência do método proposto e dos cálculos desenvolvidos para a otimização dos modelos com base de funções ortonormais na modelagem de sistemas dinâmicos lineares no domínio da frequência apresentou-se a aplicação na modelagem de três sistemas dinâmicos lineares. Os resultados mostraram que, sem nenhum conhecimento prévio sobre a dinâmica dos sistemas identificados, foi possível obter modelos com precisão e nos quais os polos convergem ou se aproximam bastante dos polos exatos do sistema sob análise, assim como acontece com a metodologia apresentada para a otimização no domínio do tempo. Os resultados apresentados foram obtidos otimizando-se tanto modelos em todo o espectro de frequência, quanto modelos que garantem a resposta em frequência somente no intervalo de resposta em frequência de interesse.

Capítulo 5

Comentários Finais, Contribuições e Perspectivas

Neste trabalho desenvolveu-se a modelagem de sistemas dinâmicos estáveis através de base de funções ortonormais generalizadas (GOBF) com funções internas. Esta abordagem modela sistemas dinâmicos lineares ou não-lineares com múltiplos modos através de uma parametrização que utiliza somente valores reais, sejam os polos do sistema reais e/ou complexos. Para a modelagem de sistemas não-lineares apresentou-se duas abordagens: os modelos *fuzzy* Takagi-Sugeno com funções ortonormais nos consequentes das regras - TS-BFO (em especial os modelos *fuzzy* TS-GOBF) e os modelos de Volterra com *kernels* formados por base de funções ortonormais generalizadas - Volterra-GOBF. Também apresentou-se uma nova abordagem para a otimização de modelos lineares BFO no domínio da frequência.

Uma das principais contribuições desta tese consistiu em apresentar uma proposta de otimização e ajuste fino dos parâmetros dos modelos não-lineares constituídos por bases de funções ortonormais, como os modelos *fuzzy* TS-BFO e os modelos de Volterra-GOBF. Modelos *fuzzy* TS-BFO já foram apresentados em outros trabalhos (Oliveira et al., 1999; Campello, 2002; Campello e Amaral, 2002; Campello, Meleiro e Amaral, 2004; Medeiros et al., 2006; Machado, 2007) e a eficácia demonstrada por esta abordagem motivou sua utilização neste trabalho. Os modelos *fuzzy* TS-BFO, além das características inerentes à modelagem por dinâmica BFO, são aproximadores universais de funções em um espaço compacto e agregam as propriedades de interpretabilidade, úteis no projeto de sistemas de controle. Nesta tese utilizou-se modelos *fuzzy* TS com GOBF's com funções internas nos consequentes das regras.

Realizou-se a identificação dos modelos *fuzzy* TS-BFO utilizando-se medidas dos sinais de entrada e saída do sistema a ser modelado. As regras do modelo *fuzzy* foram determinadas utilizando-se uma técnica de agrupamento *fuzzy* (*fuzzy clustering*) proposta por Gustafson e Kessel (1979), a partir

da qual definem-se os antecedentes das regras. O número de modelos locais que compõem o modelo TS-GOBF inicial são obtidos utilizando-se a informação fornecida por critérios de avaliação dos agrupamentos (Machado, 2007). Este procedimento pode gerar modelos que apresentem redundâncias de regras ou de funções de pertinência pertencentes ao modelo *fuzzy* TS. Para solucionar este problema, utilizou-se algoritmos que simplificam os modelos quando há redundâncias (Kaymak e Babuska, 1995). Os demais parâmetros do modelo TS-GOBF foram calculados utilizando-se técnicas de otimização não-linear. Para realizar-se este procedimento desenvolveu-se o cálculo analítico dos gradientes da saída do modelo TS-BFO com relação aos parâmetros do modelo (polos da BFO, coeficientes da expansão da BFO e parâmetros das funções de pertinência). Esta metodologia possibilitou a sintonia fina dos parâmetros dos modelos inicialmente obtidos.

Para os modelos de Volterra-GOBF, propostos em trabalhos como (Nelles, 2001; Campello et al., 2007; da Rosa, 2009), desenvolveu-se uma nova abordagem utilizando-se base de funções ortonormais generalizadas com funções internas nos *kernels* dos modelos. Esta representação possibilitou modelar sistemas não-lineares de forma mais parcimoniosa quando comparada aos modelos de Volterra clássicos com *kernels* baseados em modelos de resposta ao impulso, além de modelar dinâmicas de múltiplos modos somente por parâmetros reais. Obteve-se os gradientes analíticos da saída do modelo de Volterra-GOBF, seja com *kernels* simétricos ou não simétricos, com relação aos parâmetros a serem determinados. Estes valores foram utilizados em algoritmos de otimização que possibilitaram a obtenção de modelos mais precisos do sistema sem nenhum conhecimento a priori de suas características.

Resultados da modelagem de sistemas utilizando-se as técnicas apresentadas mostraram que é possível obter-se modelos com excelente precisão utilizando-se somente amostras de entrada e saída dos sistemas analisados. Os resultados mostraram também uma considerável melhoria na precisão dos modelos após o processo de otimização.

Além da identificação de sistemas não-lineares por modelos BFO, abordou-se também, nesta tese, a modelagem de sistemas lineares por BFO no domínio do tempo e da frequência. No contexto de modelos lineares GOBF com funções internas no domínio do tempo desenvolveu-se uma nova representação em espaço de estados. Esta representação permitiu o desenvolvimento dos cálculos analíticos dos gradientes dos modelos GOBF com funções internas com relação aos seus parâmetros. Também abordou-se a modelagem de sistemas dinâmicos lineares com BFO's no domínio da frequência. A identificação dos modelos no domínio da frequência apresenta algumas vantagens sobre a identificação no tempo, pois possibilitam a redução da influência de ruído na estimação dos parâmetros; redução do número de dados e o desenvolvimento de modelos com desempenho garantido em regiões especificadas do espectro de frequências. Neste contexto, pode-se destacar como principal contribuição o desenvolvimento do cálculo analítico dos gradientes da resposta em frequência das

funções de Kautz e Laguerre, com relação aos seus parâmetros, no domínio da frequência. O cálculo dos gradientes permitiu determinar a direção de busca dos parâmetros dos modelos em processos de otimização não-linear. Também foram otimizados os modelos GOBF com funções internas, porém com o cálculo numérico dos seus gradientes, pois ainda não foi possível estabelecer uma fórmula genérica para o cálculo analítico dos gradientes dos modelos GOBF, de qualquer ordem, em relação aos parâmetros a serem determinados.

Exemplos de aplicação dos métodos de identificação de modelos lineares BFO nos domínios do tempo e da frequência foram realizados. Os resultados ilustraram a eficiência dos métodos propostos e dos cálculos desenvolvidos para o processo de otimização dos modelos BFO, possibilitando a obtenção de modelos precisos através de dados de entrada e saída dos sistemas analisados.

5.1 Contribuições

Neste trabalho destacam-se algumas contribuições realizadas no sentido de tornar possível as implementações propostas ou o aperfeiçoamento de técnicas já utilizadas. Algumas destas contribuições são:

- Desenvolvimento de uma representação em espaço de estados para modelos GOBF com funções internas, possibilitando a modelagem de sistemas dinâmicos de qualquer ordem exclusivamente através de modelos parametrizados por coeficiente reais.
- Proposta, cálculo e implementação dos procedimentos para otimização dos parâmetros dos modelos lineares GOBF e dos modelos não-lineares *fuzzy* TS-BFO (*fuzzy* TS-GOBF) e Volterra-GOBF. Desenvolveram-se os cálculos necessários para a obtenção dos gradientes da saída do modelo GOBF com funções internas em relação aos parâmetros a serem otimizados. Apresentou-se os cálculos desenvolvidos para a obtenção dos gradientes da saída dos modelos *fuzzy* TS-BFO com relação aos parâmetros de interesse (parâmetros das funções de pertinência, polos das funções ortonormais e coeficientes da expansão da série da BFO). Para os modelos *fuzzy* TS-BFO propôs-se métodos de simplificação dos modelos iniciais e soluções pontuais como a suavização de funções de pertinência trapezoidais, o que impede a ocorrência de pontos onde não se há derivada definida. Para os modelos Volterra-GOBF desenvolveu-se os cálculos analíticos da saída dos modelos com *kernels* não-simétricos com relação aos parâmetros do modelo.
- Desenvolvimento do processo de otimização de modelos BFO lineares no domínio da frequência. Destacam-se os cálculos desenvolvidos para a obtenção analítica dos gradientes da resposta em frequência dos modelos BFO lineares (Laguerre e Kautz) com relação aos parâmetros dos

modelos. Ainda neste tema, apresentou-se também o processo de otimização de modelos lineares GOBF com o cálculo dos gradientes feito de maneira numérica.

- Cálculo dos gradientes das saídas das funções de Kautz com relação aos parâmetros de projeto apresentando-se uma nova metodologia em que os gradientes são calculados em batelada, o que reduziu o esforço computacional no processo de otimização dos modelos. Além disso, verificou-se a necessidade do estabelecimento de restrições aos parâmetros otimizados de forma a garantir a factibilidade dos modelos obtidos.

5.2 Perspectivas e Atividades Futuras

As perspectivas para pesquisas futuras relacionadas aos temas abordados pelo presente trabalho são apresentadas a seguir:

- i. Estudo para o desenvolvimento do cálculo analítico dos gradientes em relação aos parâmetros otimizados para modelos lineares GOBF no domínio da frequência;
- ii. Estudo do horizonte de retropropagação para o cálculo do gradiente da saída do modelo com relação aos parâmetros;
- iii. Estudo da possibilidade da obtenção de modelos BFO de fase mínima através de processo de otimização;
- iv. Estimativa adaptativa, em tempo real, dos modelos de BFO's para utilização em sistemas de controle;

Com relação ao item *i*) pretende-se continuar os estudos iniciados no capítulo 4 e desenvolver, assim como realizado para os modelos de Laguerre e Kautz, o cálculo analítico do gradiente da resposta em frequência do modelo GOBF linear com funções internas de qualquer ordem com relação aos parâmetros do modelo. Estes gradientes serão utilizados, conforme apresentado no capítulo 4 para a otimização dos modelos GOBF linear com funções internas no domínio da frequência.

A respeito do item *ii*), em (Nelles, 2001) é sugerido que nem sempre há necessidade de se realizar a retropropagação até as condições iniciais dos sinais para obter-se o valor exato dos gradientes. Pretende-se em trabalhos futuros mostrar que existe um limiar a partir do qual a retropropagação para o cálculo dos gradientes de modelos no domínio do tempo, apresentados neste trabalho, pode ser finalizado sem que se utilize todos os sinais até as condições iniciais. Este procedimento resultará na redução de esforço computacional. Estudos preliminares têm mostrado que é possível se obter tal limiar e garantir a precisão nos cálculos dos gradientes.

Com relação ao item *iii*) sabe-se que os filtro *all-pass* que constituem a base das funções ortogonais estudadas (Heuberger et al., 2005) são sistemas dinâmicos de fase não-mínima. A utilização de modelos de fase não-mínima no projeto de sistemas de controle pode levar a projeto de controladores instáveis. Para evitar este tipo de problema pretende-se investigar métodos de se garantir que os modelos obtidos a partir do processo de otimização empregado sejam sempre modelos de fase mínima.

Referente ao item *iv*) pretende-se estabelecer uma técnica recursiva de obtenção dos modelos BFO's em tempo real para a modelagem de sistemas variantes no tempo e possível aplicação destes modelos em sistemas de controle adaptativos baseados em modelos (Camacho e Bordons, 2007; Åström e Wittenmark, 2008).

É importante destacar que os itens *ii*) a *iii*) serão estudados tanto para os casos de modelos lineares, quanto para os casos de modelos não-lineares de Volterra e *fuzzy* TS com a parte dinâmica dos modelos formada por BFO's.

Bibliografia

- Aguirre, L. A. (2007). *Introdução à Identificação de Sistemas*, 3rd edn, Editora UFMG, Belo Horizonte.
- Akçay, H. e Heuberger, P. (2001). A frequency-domain iterative identification algorithm using general orthonormal basis functions, *Automatica* **37**(5): 663–674.
- Andrietta, S. R. (1994). *Modelagem, Simulação e Controle de Fermentação Alcoólica Contínua em Escala Industrial*, PhD thesis, FEA/UNICAMP, Campinas/SP/Brazil.
- Andrietta, S. R. e Maugeri, F. (1994). Optimum design of a continuous fermentation unit of an industrial plant for alcohol production, in E. Galindo e O. T. Ramirez (eds), *Advances in Bioprocess Engineering*, Kluwer Academic Publishers, pp. 47–52.
- Åström, K. J. e Wittenmark, B. (2008). *Adaptive Control*, 2nd edn, Dover Publications, Inc., Mineola, New York.
- Babuška, R. (1998). *Fuzzy Modeling for Control*, Kluwer.
- Babuška, R., Van der Veen, P. e Kaymak, U. (2002). Improved covariance estimation for gustafson-kessel clustering, *Proceedings of the 2002 IEEE International Conference on Fuzzy Systems*, USA, pp. 1081 – 1085.
- Babuška, R. e Verbruggen, H. B. (1997). Fuzzy set methods for local modelling and identification, in R. Murray-Smith e T. A. Johansen (eds), *Multiple Model Approaches to Modelling and Control*, Taylor and Francis, London, chapter 2.
- Bäck, T., Fogel, D. B. e Michalewicz, Z. (2000). *Evolutionary Computation 1: Basic Algorithms and Operators*, 1st edn, Institute of Physics Publishing, Bristol and Philadelphia.
- Bezdek, J. (1981a). *Pattern Recognition with Fuzzy Objective Function*, 1st edn, Plenum Press, New York.

- Bezdek, J. C.; Coray, C. G. R. W. J. (1981b). Detection and characterization of cluster substructure, i. linear structure: Fuzzy c-lines., *SIAM Journal on Applied Mathematics* **40**(2): 339–357.
- Bezdek, J. C.; Coray, C. G. R. W. J. (1981c). Detection and characterization of cluster substructure ii. fuzzy c- varieties and convex combinations thereof, *SIAM Journal on Applied Mathematics* **40**(2): 358–372.
- Bezdek, J. C. e Pal, N. R. (1998). Some new indexes of cluster validity, *IEEE Trans. on Systems, Man and Cybernetics – B* **28**(3): 301–315.
- Billings, S. A. (1980). Identification of nonlinear systems - a survey, *IEE Proc. Pt D* **127**(6): 272–285.
- Bokor, J. e Schipp., F. (1998). Approximate identification in laguerre and kautz bases, *Automatica* **34**(4): 463–468.
- Boyd, S. e Chua, L. O. (1985). Fading memory and the problem of approximating nonlinear operators with Volterra series, *IEEE Trans. on Circuits and Systems* **32**(11): 1150–1161.
- Broome, P. W. (1965). Discrete orthonormal sequences, *Journal of the Association for Computing Machinery* **12**(2): 151–168.
- Camacho, E. F. e Bordons, C. (2007). *Model Predictive Control*, 2nd edn, Springer Verlag.
- Campello, R. e Amaral, W. (2006). Hierarchical fuzzy relational models: Linguistic interpretation and universal approximation, *IEEE Transactions on Fuzzy Systems* **14**: 446–453.
- Campello, R. J. G. B. (2002). *Arquiteturas e Metodologias para Modelagem e Controle de Sistemas Complexos Utilizando Ferramentas Clássicas e Modernas*, Tese de Doutorado, DCA/FEEC/UNICAMP.
- Campello, R. J. G. B. e Amaral, W. C. (2002). Takagi-sugeno fuzzy models within orthonormal basis function framework and their application to process control, *IEEE Int. Conf. Fuzzy Systems*, Honolulu, USA, pp. 1399–1404.
- Campello, R. J. G. B., Amaral, W. C. e Favier, G. (2001). Optimal laguerre series expansion of discrete volterra models, *Proc. European Control Conference*, Porto, Portugal.
- Campello, R. J. G. B., Amaral, W. C. e Favier, G. (2006). A note on the optimal expansion of volterra models using laguerre functions, *Automatica* **42**(4): 689–693.

- Campello, R. J. G. B., Favier, G. e Amaral, W. C. (2003). Optimal expansions of discrete-time volterra models using laguerre functions, *Proc. 13th IFAC Symposium on System Identification*, Rotterdam, Holanda.
- Campello, R. J. G. B., Favier, G. e Amaral, W. C. (2004). Optimal expansions of discrete-time Volterra models using Laguerre functions, *Automatica* **40**: 815–822.
- Campello, R. J. G. B. e Hruschka, E. R. (2006). A fuzzy extension of the silhouette width criterion for cluster analysis, *Fuzzy Sets and Systems* **157**: 2858–2875.
- Campello, R. J. G. B., Meleiro, L. A. C. e Amaral, W. C. (2004). Control of a bioprocess using orthonormal basis function fuzzy models, *IEEE Int. Conf. Fuzzy Systems*, Budapest, Hungary, pp. 801–806.
- Campello, R. J. G. B. e Oliveira, G. H. C. (2007). Modelos não-lineares, in L. A. Aguirre (ed.), *Enciclopédia de Automática*, vol. 3, Blucher, São Paulo, chapter 4.
- Campello, R. J. G. B., Oliveira, G. H. C. e Amaral, W. C. (2007). Identificação e Controle de Processos via Desenvolvimentos em Séries Ortonormais: Partes A (Identificação) e B (Controle), *Controle & Automação* **18**(3): 298–332.
- Clowes, G. J. (1965). Choice of time-scaling factor for linear system approximation using orthonormal laguerre functions, *IEEE Transactions on Automatic Control* **10**(4): 487–489.
- da Rosa, A. (2009). *Identificação de Sistemas Não-Lineares Usando Modelos de Volterra Baseados em Funções Ortonormais de Kautz e Generalizadas*, Tese de Doutorado, DCA/FEEC/UNICAMP.
- da Rosa, A., Amaral, W. C. e Campello, R. J. G. B. (2005). Choice of free parameters in expansions of discrete-time volterra models using kautz functions, *Proc. 16th IFAC World Congress*, Vol. 1, Praga, República Tcheca.
- da Rosa, A., Campello, R. J. G. B. e Amaral, W. C. (2007). Choice of free parameters in expansions of discrete-time volterra models using kautz functions, *Automatica* **43**(6): 1084–1091.
- da Rosa, A., Campello, R. J. G. B. e Amaral, W. C. (2008a). Cálculo de direções de busca exatas para otimização de filtros de laguerre e de kautz, *XVII Congresso Brasileiro de Automática*, Vol. 1.
- da Rosa, A., Campello, R. J. G. B. e Amaral, W. C. (2008b). An optimal expansion of volterra models using independent kautz bases for each kernel dimension, *International Journal of Control* **81**(6): 962–975.

- da Rosa, A., Campello, R. J. G. B. e Amaral, W. C. (2009). Exact Search Directions for Optimization of Linear and Nonlinear Models Based on Generalized Orthonormal Functions, *IEEE Transactions on Automatic Control* **54**(12): 2757–2772.
- Deschrijver, D. e Dhaene, T. (2006). Parametric identification of frequency domain systems using orthonormal rational bases, *14 th IFAC Symposium on System Identification*, pp. 837–842.
- Doyle III, F. J., Pearson, R. K. e Ogunnaike, B. A. (2002). *Identification and Control using Volterra Models*, Springer-Verlag.
- Dumont, G. A. e Fu, Y. (1993a). Non-linear adaptive control via Laguerre expansion of Volterra kernels, *Int. J. Adaptive Control and Signal Processing* **7**(5): 367–382.
- Dumont, G. A. e Fu, Y. (1993b). Non-linear adaptive control via Laguerre expansion of Volterra kernels, *Int. J. Adaptive Control and Signal Processing* **7**: 367–382.
- Espinosa, J., Vandewalle, J. e Wertz, V. (2004). *Fuzzy Logic, Identification and Predictive Control*, Springer-Verlag.
- Everitt, B. S., Landau, S. e Leese, M. (2001). *Cluster analysis*, 4rd edn, Arnold.
- Eykhof, P. (1974). *System Identification, Parameter and State Estimation*, John Willey, New York.
- Fernandes, C. A. R., Favier, G. e Mota, J. C. M. (2007). Input orthogonalization methods for third-order mimo volterra channel identification, *GRETSI - Actes de Colloque*, Troyes, France, pp. 1021–1024.
- Fletcher, R. (1980). *Practical Methods of Optimization, Vol. 1, Unconstrained Optimization, e Vol. 2, Constrained Optimization*, Vol. 1, John Wiley and Sons.
- Fu, Y. e Dumont, G. A. (1993). An optimum time scale for discrete Laguerre network, *IEEE Transactions on Automatic Control* **38**(6): 934–938.
- Gath, I. e Geva, A. B. (1989). Unsupervised optimal fuzzy clustering, *IEEE Transactions on pattern analysis and machine intelligence* **2**(7): 773–781.
- Gopal, V. e Biegler, L. T. (1999). Smoothing methods for complementarity problems in process engineering, *AIChE Journal* **45**(7): 1535–1547.
- Gray, A., J. e Markel, J. (1975). A normalized digital filter structure, *Acoustics, Speech and Signal Processing, IEEE Transactions on* **23**(3): 268–277.

- Gustafson, D. e Kessel, W. C. (1979). Fuzzy clustering with a fuzzy covariance matrix, *Proc. IEEE CDC* pp. 761–766.
- Hakvoort, R. G. e den Hof, P. M. J. V. (1994). Frequency domain curve fitting with maximum amplitude criterion and guaranteed stability, *EURASIP Journal on Applied Signal Processing* **60**: 809–825.
- Hathaway, R. e Bezdek, J. (1993). Switching regression models and fuzzy clustering, *IEEE Transactions on Fuzzy Systems* **1**(3): 195–204.
- Haykin, S. (1999). *Neural Networks, A Comprehensive Foundation*, 2nd edn, Prentice Hall.
- Hellendoorn, H. e Driankov, D. (1997). *Fuzzy Model Identification: Selected Approaches*, Springer-Verlag.
- Heuberger, P. S. C., V. d. H. P. M. J. e. B. O. H. (1995). A generalized orthonormal basis for linear dynamical systems, *IEEE Transactions on Automatic Control* **40**(3): 451–465.
- Heuberger, P. S. C., Van den Hof, P. M. J. e Wahlberg, B. (2005). *Modelling and Identification with Rational Orthogonal Basis Functions*, Springer.
- Höppner, F., Klawonn, F., Kruse, R. e Runkler, T. (1999). *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*, John Wiley & Sons.
- Kaufman, L. e Rousseeuw, P. J. (1990). *Finding Groups in Data: an Introduction to Cluster Analysis*, 1st edn, John Wiley and Sons.
- Kaymak, U. e Babuska, R. (1995). Compatible cluster merging for fuzzy modelling, *International Joint Conference of the Fourth IEEE International Conference on Fuzzy Systems and The Second International Fuzzy Engineering Symposium., Proceedings of 1995 IEEE International Conference on Fuzzy Systems, 1995*. **2**: 897–904.
- Khouaja., A. e Favier, G. (2004). Identification of parafac-volterra cubic models using an alternating recursive least squares algorithm, *EUSIPCO 2004 - XII European Signal Processing Conference*, Vienna, Austria, pp. 1903–1906.
- Kosko, B. (1994). Fuzzy systems as universal approximators, *IEEE Transaction on Computers* **43**(11): 1329–1333.
- Krishnnapuram, R. e Freg, C.-P. (1992). Fitting an unknown number of lines and planes to image data through compatible cluster merging, *Pattern Recognition* **25**(4): 385–400.

- Lazar, M. (2001). *Non-linear controler based on the ep-sac approach*, Master's thesis, University of Ghent, Departament of Control, Engineering and Automation, Flanders, Belgium.
- Levenberg, K. (1944). A method for the solution of certain nonlinear problems in least square, *Quarterly of Applied Mathematics* **2**(2): 164–168.
- Lin, P. L. e Wu, Y. C. (1982). Identification of multi-input multi-output linear systems from frequency response data, *J. Dyn. Syst., Meas. Contr.* **4**: 37–43.
- Lindskog, P. (1996). *Methods, Algorithms and Tools for System Identification Based on Prior Knowledge*, PhD thesis, Linkoping University, Sweden.
- Ljung, L. (1993). Some results on identifying linear systems using frequency domain data, *32nd Conference on Decision and Control*, San Antonio, Texas, EUA, pp. 3534–3538.
- Ljung, L. (1999). *System identification: Theory for the user*, 2nd edn, Prentice Hall.
- Luenberger, D. G. (2003). *Linear and Nonlinear Programming*, 2nd edn, Springer.
- Machado, J. B. (2007). *Modelagem e Controle Preditivo utilizando Multimodelos*, Dissertação de Mestrado, DCA/FEEC/UNICAMP.
- Machado, J. B., Campello, R. J. G. B. e Amaral, W. C. (2007). Design of OBF-TS fuzzy models based on multiple clustering validity criteria, *Proc. IEEE Int. Conf. on Tools with Artificial Intelligence*, Patras, Grécia, pp. 336–339.
- Marquardt, D. W. (1963). An algorithm for the least-square estimation of nonlinear parameters, *SIAM Journal of Applied Mathematics* **11**(2): 431 – 441.
- Masnadi-Shirazi, M. A. e Ahmed, N. (1991). Optimum laguerre networks for a class of discrete-time systems, *IEEE Transactions on Signal Processing* **39**(9): 2104–2108.
- Mbarek, A., Messaoud, H. e Favier, G. (2003). Robust predictive control using kautz model, *ICECS - 10th IEEE International Conference on Electronics, Circuits and Systems*, pp. 184–187.
- McKelvey, T. e Akcay, H. (1994). An efficient frequency domain state-space identification algorithm: Robustness and stochastic analysis, Lake Buena Vista, FL, pp. 3348–3353.
- Medeiros, A., Amaral, W. e Campello, R. (2006). GA Optimization of Generalized OBF TS Fuzzy Models with Global and Local Estimation Approaches, *2006 IEEE International Conference on Fuzzy Systems*, Vol. 5, pp. 1835 – 1842.

- Medeiros, A. V. (2006). *Modelagem de Sistemas Dinâmicos Não Lineares Utilizando Sistemas Fuzzy, Algoritmos Genéticos e Funções de Base Ortonormal*, Dissertação de Mestrado, DCA/FEEC/UNICAMP.
- Nelles, O. (2001). *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*, Springer-Verlag.
- Ninness, B. e Gustafsson, F. (1997). A unifying construction of orthonormal bases for system identification, *IEEE Transactions on Automatic Control* **42**(4): 515–521.
- Nocedal, J. e Wright, S. J. (1999). *Numerical Optimization*, Springer-Verlag.
- Oliveira e Silva, T. A. M. (1995). On the determination of the optimal pole position of laguerre filters, *IEEE Transactions on Signal Processing* **43**(9): 2079–2087.
- Oliveira, G. H. C. (1997). *Controle Preditivo para Processos com incertezas estruturadas baseado em séries de funções ortonormais*, Tese de Doutorado, DCA/FEEC/UNICAMP, Campinas.
- Oliveira, G. H. C., A. W. C. F. G. e Dumont, G. A. (2000). Constrained robust predictive controller for uncertain processes modeled by orthonormal series functions, *Automatica* **36**(4): 563–571.
- Oliveira, G. H. C. e Amaral, W. (2000). Identificação e controle de processos não-lineares utilizando séries de volterra e bases de funções ortonormais, *XIII Congresso Brasileiro de Automática*, Vol. 1, pp. 2174–2179.
- Oliveira, G. H. C., Campello, R. e Amaral, W. C. (1999). Fuzzy models within orthonormal basis function framework, *Proc. 8th IEEE Internat. Conference on Fuzzy Systems*, Seoul, Korea, pp. 957–962.
- Oppenheim, A. V., Schafer, R. W. e Schafer, R. W. (1999). *Discrete-Time Signal Processing*, 2nd edn, Prentice-Hall.
- Optimization Toolbox User's Guide* (2008). The MathWorks, Inc.
- Pintelon, R., Guillaume, P., Rolain, Y., Schoukens, J. e Vanhamme, H. (1994). Parametric identification of transfer functions in the frequency domain - a survey, *IEEE Trans. Automat. Contr.* **39**(11): 2245–2260.
- Pintelon, R. e Schoukens, J. (2001). *System identification: A Frequency Domain Approach*, 1st edn, IEEE Press, New York.

- Powell, M. (1978a). The convergence of variable metric methods for nonlinearly constrained optimization calculations, in R. M. O.L. Mangasarian e S. Robinson (eds), *Nonlinear Programming 3*, Academic Press.
- Powell, M. (1978b). A fast algorithm for nonlinearly constrained optimization calculations, in G.A.Watson (ed.), *Numerical Analysis*, Springer Verlag.
- Powell, M. (1983). Variable metric methods for constrained optimization, mathematical programming: The state of the art, in M. G. A. Bachem e B. Korte (eds), *Nonlinear Programming 3*, Springer Verlag, pp. 288–311.
- Reginato, B. C. (2008). *Ambiente computacional para identificação de sistemas nos domínios do tempo e da frequência usando bases de funções ortonormais generalizadas*, Master's thesis, Pontifícia Universidade Católica do Paraná, Curitiba, Brasil.
- Reginato, B. C. e Oliveira, G. H. C. (2008). Algoritmo híbrido para seleção de pólos em modelos formados por bases de funções ortonormais generalizadas, *Congresso Brasileiro de Automática*, Juiz de Fora, Brasil.
- Roberts, R. A. e Mullis, C. T. (1987). *Digital Signal Processing*, Addison-Wesley Publishing Company, Inc.
- Rugh, W. J. (1981). *Nonlinear System Theory: The Volterra / Wiener Approach*, The Johns Hopkins University Press.
- Rumelhart, D. E., Hinton, G. E. e Williams, R. J. (1986). Learning internal representation by error propagation, in D. E. Rumelhart e J. L. McClelland (eds), *Parallel Distributed Processing: Explorations in the Microstructure of cognition*, MIT Press, Cambridge.
- Scales, L. E. (1985). *Introduction to Nonlinear Optimization*, Macmillan, Londres.
- Schetzen, M. (1980). *The Volterra and Wiener Theories of Nonlinear Systems*, John Wiley & Sons.
- Schoukens, J. e Pintelon, R. (1991). *Identification of Linear Systems: A Practical Guideline to Accurate Modeling*, Pergamon, New York.
- Setnes, M., Babuska, R., Kaymak, U. e van Nauta Lemke, H. (1998). Similarity measures in fuzzy rule base simplification, *IEEE Transactions on Systems, Man, and Cybernetics, Part B* **28**(3): 376–386.

- Sidman, M. D., DeAngelis, F. E. e Verghese, G. C. (1991). Parametric system identification on logarithmic frequency response data, *IEEE Transactions on Automatic Control* **36**(9): 1065–1070.
- Takagi, T. e Sugeno, M. (1985). Fuzzy identification of systems and its application to modeling and control, *IEEE Transaction on Systems, Man and Cybernetics* **15**(1): 116–132.
- Tanguy, N., Morvan, R., Vilbé, P. e C., C. L. (1995). Optimum choice of free parameter in orthonormal approximations, *IEEE Transactions on Automatic Control* **40**(10): 1811–1813.
- Tanguy, N., Morvan, R., Vilbé, P. e C., C. L. (2000). Pertinent parameters for kautz approximation, *Eletronic Letters* **36**(8): 769–771.
- Tanguy, N., Morvan, R., Vilbé, P. e C., C. L. (2002). Pertinent choice of parameters for discrete kautz approximation, *IEEE Transactions on Automatic Control* **47**(5): 783–787.
- Van den Hof, P. M. J., H. P. S. C. e. B. J. (1995). System identification with generalized orthonormal basis functions, *Automatica* **31**(12): 1821–1834.
- Vries, D. K. e Van den Hof, P. M. J. (1998). Frequency domain identification with generalized orthonormal basis functions, *IEEE Transactions on Automatic Control* **43**(5): 656–669.
- Wahlberg, B. (1991). System identification using Laguerre models, *IEEE Transactions on Automatic Control* **36**(5): 551–562.
- Wahlberg, B. (1994). System identification using Kautz models, *IEEE Transactions on Automatic Control* **39**(6): 1276–1282.
- Wahlberg, B. e Makila, P. M. (1996). Approximation of stable linear dynamical systems using laguerre e kautz functions, *Automatica* **32**(5): 693–708.
- Wang, L. X. (1998). Universal aproximation by hierarchical fuzzy systems, *Fuzzy Sets and Systems* **93**(2): 223–230.
- Wang, L.-X. e Mendel, J. M. (1992). Fuzzy basis functions, universal approximation, and orthogonal least-square learning, *IEEE Transaction on Neural Network* **43**(5): 807–814.
- Wiener, N. (1958). *Nonlinear Problems in Random Theory*, Wiley.
- Yager, R. R. e Filev, D. P. (1994). *Essentials of Fuzzy Modeling and Control*, John Wiley and Sons.
- Zadeh, L. A. (1965). Fuzzy sets, *Information and Control* **8**: 338–353.

- Zwick, R., Carlstein, E. e Budescu, D. V. (1987). Measures of similarity among fuzzy concepts: A comparative analysis, *Int. J. Approx. Reas.* **1**: 221–242.

Apêndice A

Modelo TS-BFO com modelos de Kautz

A.1 Ajuste global dos parâmetros do modelo final

Uma vez realizadas as operações de fusão e eliminação de funções de pertinência e regras por similaridade, deseja-se realizar o ajuste fino do modelo obtido através da otimização dos parâmetros do modelo TS-BFO, minimizando-se o erro quadrático entre a saída estimada pelo modelo e a saída medida em todos os instantes de tempo. A equação (A.1) mostra a função de custo a ser minimizada:

$$\min_{\theta} J = \frac{1}{2} \sum_{k=1}^N (\hat{y}(k) - y(k))^2 \quad (\text{A.1})$$

com

$\hat{y}(k)$ - saída global estimada

$y(k)$ - saída medida do sistema

N - número de amostras do sistema

θ - Parâmetros de otimização

A minimização da função de custo se dará com relação aos parâmetros $\theta = [p \ C_i \ L_{i,j}]^T$, onde p são os parâmetros b e c das funções de Kautz (equações (2.7) e (2.8)), C_i é a matriz de expansão dos modelos BFO locais presentes nos consequentes das regras, definida em (3.6), e $L_{i,j}$ as funções de pertinência presentes nos antecedentes das regras do modelo *fuzzy* TS-BFO.

A seguir, analisa-se a otimização para modelos TS-BFO com funções de Kautz nos consequentes das regras. Para a determinação da direção de busca dos valores ótimos para os parâmetros referentes

aos pólos é necessário se conhecer o gradiente de J com relação aos parâmetros b e c :

$$\nabla_p J = \left[\frac{\partial J}{\partial b} \quad \frac{\partial J}{\partial c} \right]^T \quad (\text{A.2})$$

Analisando primeiramente o gradiente da função de custo com relação ao parâmetro b tem-se:

$$\frac{\partial J}{\partial b} = \sum_{k=1}^N (y(k) - \hat{y}(k)) \frac{\partial \hat{y}(k)}{\partial b} \quad (\text{A.3})$$

onde $\hat{y}(k)$ para o modelo *fuzzy* TS-BFO é dado por:

$$\hat{y}(k) = \frac{\sum_{i=1}^K \mu_i(\Psi(k)) \check{y}_i(k)}{\sum_{i=1}^K \mu_i(\Psi(k))}$$

com

$\mu_i(\Psi(k))$ - ativação da i -ésima regra ($i = 1, \dots, K$) - eq. (3.3);

$\check{y}_i(k)$ - saída do i -ésimo modelo local;

$\Psi(k)$ - saída dos filtros das BFO's no instante k .

A saída $\hat{y}(k)$ pode ser reescrita como:

$$\hat{y}(k) = \frac{\sum_{i=1}^K \left(\prod_{j=1}^{\varsigma} L_{i,j}(\psi_j(k)) \right) C_i^T \Psi(k)}{\sum_{i=1}^K \left(\prod_{j=1}^{\varsigma} L_{i,j}(\psi_j(k)) \right)} \quad (\text{A.4})$$

onde $L_{i,j}(\psi_j(k))$ são os valores de ativação das funções de pertinência, $\psi_j(k)$ é a saída do j -ésimo filtro no instante k e ς o número de variáveis no antecedente das regras com $\varsigma \leq n$. A figura 3.12 mostra o modelo TS-BFO com duas funções BFO nos antecedentes e nos consequentes das regras com duas regras onde fica explícito a dependência entre as variáveis do modelo.

Definindo-se o numerador e o denominador da equação (A.4) como:

$$v = \sum_{i=1}^K \left(\prod_{j=1}^{\varsigma} L_{i,j}(\psi_j(k)) \right) C_i^T \Psi(k) \quad \text{e} \quad u = \sum_{i=1}^K \left(\prod_{j=1}^{\varsigma} L_{i,j}(\psi_j(k)) \right) \quad (\text{A.5})$$

tem-se que:

$$\hat{y}(k) = \frac{v}{u} \Rightarrow \frac{\partial \hat{y}(k)}{\partial b} = \frac{\frac{\partial v}{\partial b} \cdot u - v \cdot \frac{\partial u}{\partial b}}{u^2} \quad (\text{A.6})$$

o mesmo valendo para $\frac{\partial \hat{y}(k)}{\partial c}$

Calculando-se o valor da derivada parcial do denominador (u) com relação ao parâmetro b tem-se:

$$\begin{aligned} \frac{\partial u}{\partial b} = & \frac{\partial \psi_1(k)}{\partial b} \left[\sum_{i=1}^K \frac{\partial L_{i1}(\psi_1(k))}{\partial \psi_1(k)} \cdot L_{i2}(\psi_2(k)) \cdot \dots \cdot L_{i\varsigma}(\psi_\varsigma(k)) \right] + \\ & \frac{\partial \psi_2(k)}{\partial b} \left[\sum_{i=1}^K L_{i1}(\psi_1(k)) \cdot \frac{\partial L_{i2}(\psi_2(k))}{\partial \psi_2(k)} \cdot \dots \cdot L_{i\varsigma}(\psi_\varsigma(k)) \right] + \\ & + \dots + \frac{\partial \psi_\varsigma(k)}{\partial b} \left[\sum_{i=1}^K L_{i1}(\psi_1(k)) \cdot L_{i2}(\psi_2(k)) \cdot \dots \cdot \frac{\partial L_{i\varsigma}(\psi_\varsigma(k))}{\partial \psi_\varsigma(k)} \right] \end{aligned}$$

onde $\frac{\partial \psi_j(k)}{\partial b}$ é definido através dos resultados apresentados na seção 2.4.3 e $\frac{\partial L_{ij}}{\partial \psi_j(k)}$ é definida a seguir.

No modelo TS-BFO abordado, como já citado anteriormente, são utilizadas funções de pertinência trapezoidais. Para que os cálculos dos gradientes das funções de pertinência seja possível é necessário realizar uma suavização das funções de pertinência como se realizou na seção 3.5. Os cálculos para as derivadas das funções de pertinência com relação aos seus parâmetros são os mesmos apresentados na seção 3.5. Com o cálculo de $\frac{\partial L_{ij}(x)}{\partial \psi_j(k)}$ defini-se o valor de $\frac{\partial u}{\partial b}$. A mesma abordagem pode ser aplicada para o cálculo da derivada do denominador com relação ao parâmetro c , $\frac{\partial u}{\partial c}$.

Para o cálculo do numerador v , de (A.5), com relação ao parâmetro b a matriz C_i^T presente na equação (A.5) pode ser reescrita em função dos parâmetros da expansão da BFO:

$$C_{fi} = [g_{1,i} \ g_{2,i} \ \dots \ g_{n_f,i}]^T, \quad (\text{A.7})$$

onde n_f é a maior ordem filtro da base de Kautz e $g_{q,i}$ são os coeficientes da expansão da BFO. Assim, o numerador da função (A.4) apresentado em (A.5) pode ser redefinido como:

$$v = \sum_{i=1}^K \left[\left(\prod_{j=1}^{\varsigma} L_{i,j}(\psi_j(k)) \right) \sum_{q=1}^{n_f} g_{q,i} \psi_q(k) \right]$$

Analisando-se agora a derivada parcial do numerador v com relação ao parâmetro b do modelo de Kautz, $\frac{\partial v}{\partial b}$, tem-se a fórmula generalizada dada por:

$$\frac{\partial v}{\partial b} = \sum_{i=1}^K \sum_{q=1}^{n_f} g_{q,i} \left(\frac{\partial L_{i1}(\psi_1(k))}{\partial \psi_1(k)} \frac{\partial \psi_1(k)}{\partial b} \cdot L_{i2}(\psi_2(k)) \cdot \dots \cdot L_{i\varsigma}(\psi_\varsigma(k)) \psi_q(k) + \right.$$

$$\begin{aligned}
& L_{i1}(\psi_1(k)) \cdot \frac{\partial L_{i2}(\psi_2(k))}{\partial \psi_2(k)} \frac{\partial \psi_2(k)}{\partial b} \dots L_{i\varsigma}(\psi_\varsigma(k)) \psi_q(k) + \dots + \\
& L_{i1}(\psi_1(k)) \cdot L_{i2}(\psi_2(k)) \dots \frac{\partial L_{i\varsigma}(\psi_\varsigma(k))}{\partial \psi_\varsigma(k)} \frac{\partial \psi_\varsigma(k)}{\partial b} \psi_q(k) + \\
& L_{i1}(\psi_1(k)) \cdot L_{i2}(\psi_2(k)) \dots L_{i\varsigma}(\psi_\varsigma(k)) \frac{\partial \psi_q(k)}{\partial b}
\end{aligned}$$

O mesmo cálculo pode ser realizado para se determinar $\frac{\partial v}{\partial c}$, obtendo-se a seguinte equação:

$$\begin{aligned}
\frac{\partial v}{\partial c} = & \sum_{i=1}^K \sum_{q=1}^{n_f} g_{q,i} \left(\frac{\partial L_{i1}(\psi_1(k))}{\partial \psi_1(k)} \frac{\partial \psi_1(k)}{\partial c} \cdot L_{i2}(\psi_2(k)) \dots L_{i\varsigma}(\psi_\varsigma(k)) \psi_q(k) + \right. \\
& L_{i1}(\psi_1(k)) \cdot \frac{\partial L_{i2}(\psi_2(k))}{\partial \psi_2(k)} \frac{\partial \psi_2(k)}{\partial c} \dots L_{i\varsigma}(\psi_\varsigma(k)) \psi_q(k) + \dots + \\
& L_{i1}(\psi_1(k)) \cdot L_{i2}(\psi_2(k)) \dots \frac{\partial L_{i\varsigma}(\psi_\varsigma(k))}{\partial \psi_\varsigma(k)} \frac{\partial \psi_\varsigma(k)}{\partial c} \psi_q(k) + \\
& \left. L_{i1}(\psi_1(k)) \cdot L_{i2}(\psi_2(k)) \dots L_{i\varsigma}(\psi_\varsigma(k)) \frac{\partial \psi_q(k)}{\partial c} \right)
\end{aligned}$$

Depois de determinados os valores de $\frac{\partial u}{\partial b}$, $\frac{\partial v}{\partial b}$, u e v , estes são substituídos em (A.6) e então é obtido o valor de $\frac{\partial \hat{y}(k)}{\partial b}$, sendo o procedimento análogo adotado para o cálculo de $\frac{\partial \hat{y}(k)}{\partial c}$. Substituindo-se os valores de $\frac{\partial \hat{y}(k)}{\partial b}$ na equação (A.3) é possível se calcular o gradiente da função de custo com relação ao parâmetro b , sendo o mesmo feito para se calcular o gradiente da função J com relação ao parâmetro c . Os parâmetros b e c apresentam restrições quanto aos valores que podem assumir, estando b restrito ao intervalo $(0, 1)$ e c restrito ao intervalo $(-1, 0)$. Além disso, os parâmetros b e c apresentam uma restrição adicional quanto aos valores que podem assumir, já que tais parâmetros devem ser convertidos em par de pólos complexos conjugados β e $\bar{\beta}$ conforme descrito na seção 2.4.3.

Como citado anteriormente, além de se desejar minimizar a função de custo J com relação aos parâmetros b e c referentes aos polos das funções de Kautz deseja-se também otimizar J com relação aos parâmetros dos consequentes e antecedentes das regras do modelo *fuzzy* TS-BFO com funções de Kautz. Os cálculos para estes gradientes são os mesmos apresentados na seção 3.5.

Apêndice B

Modelo Não-lineares com GOBF's

B.1 Tabelas com os modelos identificados

Este apêndice tem como objetivo apresentar as tabelas com os diversos modelos obtidos durante o processo de identificação empregados na seção 3.9 em que se variou o número de polos e de funções nas GOBF's empregadas nos modelos não-lineares. As tabelas se referem aos modelos dos sistemas nível líquido em um tanque, da planta para produção de etanol e modelo racional NARX e apresentam resultados de alguns modelos de Volterra-GOBF com *kernels* simétricos e não-simétricos e modelos de *fuzzy* TS-GOBF.

B.1.1 Modelagem de nível líquido em um tanque

A tabela B.1 a seguir apresenta alguns modelos obtidos através da variação do número de polos e de funções empregando-se as abordagens de modelagem de sistemas não-lineares com GOBF's utilizadas nesta tese. Nas tabelas a seguir tem-se o número de funções ortonormais (NF) presentes em cada *kernel* nos modelos de Volterra-GOBF ou número de funções nos consequentes das regras dos modelos TS-GOBF, o número de polos (NP) das funções, o número de regras no modelo *fuzzy* (NR) e o EQM final da saída de cada modelo. A abreviação ke_i indica a qual *i*-*kernel* o número de funções se refere.

Tab. B.1: Modelos para Sistema de nível de líquidos em um tanque.

Tipo de modelo	NF	NP	NR	EQMf
Volterra-GOBF	$ke_1 = 1$	1	-	2,0192
	$ke_2 = 1$	1	-	
	$ke_1 = 3$	1	-	1,0835
	$ke_2 = 4$	2	-	
	$ke_1 = 4$	2	-	1,0791
	$ke_2 = 4$	2	-	
	$ke_1 = 9$	3	-	1,0821
	$ke_2 = 6$	3	-	
	$ke_1 = 8$	1	-	0,8870
	$ke_2 = 4$	1	-	
Volterra-GOBF Direções Independentes	$ke_1 = 8$	1	-	0,8379
	$ke_{21} = 4$	1	-	
	$ke_{22} = 4$	1	-	
	$ke_1 = 8$	2	-	0,8835
	$ke_{21} = 4$	1	-	
	$ke_{22} = 4$	1	-	
	$ke_1 = 12$	2	-	1,1071
	$ke_{21} = 4$	1	-	
	$ke_{22} = 4$	1	-	
	$ke_1 = 8$	2	-	1,0378
	$ke_{21} = 4$	2	-	
	$ke_{22} = 4$	2	-	
Fuzzy TS-GOBF Funções int.	12	1	6	0,7463
	8	2	5	0,4521
	12	3	3	0,1635

B.1.2 Planta para produção de Etanol

A tabela a seguir mostra alguns modelos obtidos variando-se o número de polos e de funções para as abordagens de modelagem de sistemas não-lineares com GOBF's propostas nesta tese.

Tab. B.2: Modelos para Sistema de Etanol.

Tipo de modelo	NF	NP	NR	EQMf
Volterra-GOBF	$ke_1 = 1$	1	-	0,01293
	$ke_2 = 1$	1	-	0,01293
	$ke_1 = 4$	1	-	0,00882
	$ke_2 = 4$	1	-	0,00882
	$ke_1 = 2$	2	-	0,00791
	$ke_2 = 1$	1	-	0,00791
	$ke_1 = 4$	2	-	0,007381
	$ke_2 = 3$	1	-	0,007381
	$ke_1 = 3$	1	-	0,006784
	$ke_2 = 2$	2	-	0,006784
Volterra-GOBF Direções Independentes	$ke_1 = 1$	1	-	0,01293
	$ke_{21} = 1$	1	-	0,01293
	$ke_{22} = 1$	1	-	0,01293
	$ke_1 = 1$	1	-	0,007913
	$ke_{21} = 2$	1	-	0,007913
	$ke_{22} = 2$	1	-	0,007913
	$ke_1 = 3$	1	-	0,007135
	$ke_{21} = 3$	1	-	0,007135
	$ke_{22} = 3$	1	-	0,007135
	$ke_1 = 4$	2	-	0,006643
	$ke_{21} = 2$	1	-	0,006643
	$ke_{22} = 2$	1	-	0,006643
	$ke_1 = 2$	2	-	0,006322
	$ke_{21} = 4$	2	-	0,006322
	$ke_{22} = 4$	2	-	0,006322
Fuzzy TS-GOBF Funções int.	4	1	7	0,00713
	6	1	6	0,005144
	4	2	5	0,004641
	6	2	5	0,003815

B.1.3 Modelo Racional NARX

A tabela B.3 apresenta modelos obtidos através da variação do número de polos e de funções empregando-se as abordagens de modelagem de sistemas não-lineares com GOBF's propostas nesta tese.

Tab. B.3: Modelos para Sistema Racional NARX.

Tipo de modelo	NF	NP	NR	EQMf
Volterra-GOBF	$ke_1 = 1$	1	-	0,002358
	$ke_2 = 1$	1	-	
	$ke_1 = 2$	1	-	0,001846
	$ke_2 = 3$	3	-	
	$ke_1 = 2$	2	-	0,001073
	$ke_2 = 2$	2	-	
	$ke_1 = 4$	2	-	0,001574
	$ke_2 = 4$	2	-	
	$ke_1 = 6$	3	-	0,001719
	$ke_2 = 4$	2	-	
Volterra-GOBF Direções Independentes	$ke_1 = 1$	1	-	0,002423
	$ke_{21} = 1$	1	-	
	$ke_{22} = 1$	1	-	
	$ke_1 = 3$	1	-	0,001955
	$ke_{21} = 2$	1	-	
	$ke_{22} = 3$	1	-	
	$ke_1 = 2$	1	-	$8,3425 \cdot 10^{-4}$
	$ke_{21} = 2$	1	-	
	$ke_{22} = 2$	2	-	
	$ke_1 = 3$	1	-	0,001245
	$ke_{21} = 2$	2	-	
	$ke_{22} = 2$	2	-	
	$ke_1 = 2$	2	-	0,001290
	$ke_{21} = 1$	1	-	
	$ke_{22} = 3$	3	-	
Fuzzy TS-GOBF Funções int.	2	2	5	$2,0152 \cdot 10^{-3}$
	4	2	4	$1,6827 \cdot 10^{-4}$
	6	2	4	$1,7949 \cdot 10^{-4}$
	3	3	5	$1,8259 \cdot 10^{-3}$

Apêndice C

Trabalhos publicados e em desenvolvimento

Os trabalhos a seguir foram desenvolvidos e publicados durante o doutorado:

- Machado, J. B., Campello, R. J. G. B. e Amaral, W. C. (2007). Design of OBF-TS fuzzy models based on multiple clustering validity criteria, Proc. IEEE Int. Conf. on Tools with Artificial Intelligence, Patras, Grécia, pp. 336-339.
- Machado, J. B., Amaral, W. C. e Campello, R. J. G. B. (2008). Controlador Preditivo Multi-modelos com Modelagem Fuzzy TS-FBO, XVII Congresso Brasileiro de Automática, Juiz de Fora - MG, Brasil.
- Machado, J. B., Campello, R. J. G. B. e Amaral, W. C. (2010). Modelos lineares GOBF com funções internas, XVIII Congresso Brasileiro de Automática, Bonito - MS, Brasil.
- OLIVEIRA, Gustavo Henrique da Costa ; ROSA, Alex da ; CAMPELLO, R. J. G. B. ; MACHADO, Jeremias Barbosa ; AMARAL, W. C. . An Introduction to Models based on Laguerre, Kautz and Other Related Orthonormal Functions - Part I: Linear and Uncertain Models. International Journal of Modelling, Identification and Control (Aceito para publicação) , 2011
- OLIVEIRA, Gustavo Henrique da Costa ; ROSA, Alex da ; CAMPELLO, R. J. G. B. ; MACHADO, Jeremias Barbosa ; AMARAL, W. C. . An Introduction to Models based on Laguerre, Kautz and Other Related Orthonormal Functions - Part II: Nonlinear Models. International Journal of Modelling, Identification and Control (Aceito para publicação) , 2011.

Os trabalhos a seguir estão em fase de elaboração e submissão:

- Machado, J. B., Campello, R. J. G. B. e Amaral, W. C. . Nonlinear Fuzzy Models within Generalized Orthonormal Basis Functions Framework. IEEE Transactions on Systems, Man, and Cybernetics: Part B (Submetido).

- Márcio Feliciano Braga, Jeremias Barbosa Machado, Ricardo José Gabrielli Barreto Campello e W.C. Amaral. Optimization of Volterra Models with Asymmetrical Kernels based on Generalized Orthonormal Functions. 19th Mediterranean Conference on Control and Automation, Corfu, Greece, 2011 (Submetido).
- Machado, J. B., Campello, R. J. G. B. e Amaral, W. C. . Volterra Models within Ladder Structured Generalized Orthonormal Basis Functions Framework. (Em elaboração).
- Machado, J. B., Campello, R. J. G. B. e Amaral, W. C. . Optimization of Orthonormal Basis Functions Models in Frequency Domain. (Em elaboração).

Durante o doutorado publicou-se ainda alguns trabalhos oriundos de atividades desenvolvidas em disciplinas cursadas:

- Artigos completos publicados em periódicos:
Simões, F. O. ; Neto, M. U.; Machado, J. B. ; Nagle, E. J. ; Runstein, F. O. ; Gomes, L. C. T.; Compressão de fala utilizando quantização vetorial e redes neurais não supervisionadas. Cadernos CPQD tecnologia, v. 5, p. 33-48, 2009.
- Capítulos de livros publicados:
Simões, F. O. ; Neto, M. U.; Machado, J. B. ; Nagle, E. J. ; Runstein, F. O. ; Gomes, L. C. T.; *Vector Quantization of Speech Frames Based on Self-Organizing Maps*. Em: Hussain, A.; Aleksander, I.; Smith, L.S.; Barros, A.K.; Chrisley, R.; Cutsuridis, V.. (Org.). Brain Inspired Cognitive Systems 2008 - Series: Advances in Experimental Medicine and Biology. 1 ed. Springer: New York, 2010, v. 657, p. 201-216.
- Trabalhos completos publicados em anais de congressos:
Simões, F. O. ; Neto, M. U.; Machado, J. B. ; Nagle, E. J. ; Runstein, F. O. ; Gomes, L. C. T.; Quantização de Sinais de Fala Utilizando Redes Neurais Não-Supervisionadas. Em: I Congresso Tecnológico InfoBrasil TI e Telecom, 2008, Fortaleza - CE. Congresso Tecnológico InfoBrasil 2008, 2008.
Simões, F. O. ; Neto, M. U.; Machado, J. B. ; Nagle, E. J. ; Runstein, F. O. ; Gomes, L. C. T.; *Speech Compression using Vector Quantization and Unsupervised Neural Networks*. Em: BICS 2008 - Brain Inspired Cognitive Systems 2008, 2008, São Luís, Maranhão - Brasil. Proceedings BICS 2008, 2008.