

Universidade Estadual de Campinas – UNICAMP
Faculdade de Engenharia Elétrica e de Computação – FEEC
Depto. de Eng. Computação e Automação Industrial – DCA

Identificação de Sistemas Não-Lineares Usando Modelos de Volterra Baseados em Funções Ortonormais de Kautz e Generalizadas

Autor: MSc. Alex da Rosa

Orientador: Prof. Dr. Wagner Caradori do Amaral

Co-orientador: Prof. Dr. Ricardo José Gabrielli Barreto Campello

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

Campinas-SP
2009

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

R71i Rosa, Alex da
Identificação de sistemas não-lineares usando modelos de Volterra baseados em funções ortonormais de Kautz e generalizadas / Alex da Rosa – Campinas, SP: [s.n.], 2009.

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. Identificação de sistemas. 2. Sistemas não-lineares. 3. Volterra, Séries de. 4. Otimização. 5. Métodos de gradiente conjugado. 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.

Universidade Estadual de Campinas – UNICAMP
Faculdade de Engenharia Elétrica e de Computação – FEEC
Depto. de Eng. Computação e Automação Industrial – DCA

Identificação de Sistemas Não-Lineares Usando Modelos de Volterra Baseados em Funções Ortonormais de Kautz e Generalizadas

Tese de Doutorado apresentada em 12 de março de 2009

Autor: MSc. Alex da Rosa

Orientador: Prof. Dr. Wagner Caradori do Amaral

Co-orientador: Prof. Dr. Ricardo J. G. B. Campello (SCC/ICMC/USP)

BANCA EXAMINADORA

Prof. Dr. Wagner Caradori do Amaral (FEEC/UNICAMP)

Prof. Dr. Gérard Favier (I3S/CNRS/Universidade de Nice, França)

Prof. Dr. Gustavo Henrique da Costa Oliveira (PUC-PR)

Prof. Dr. Rafael Santos Mendes (FEEC/UNICAMP)

Prof. Dr. Paulo Augusto Valente Ferreira (FEEC/UNICAMP)

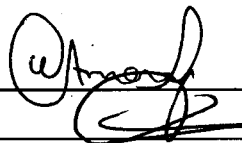
COMISSÃO JULGADORA - TESE DE DOUTORADO

Candidato: Alex da Rosa

Data da Defesa: 12 de março de 2009

Título da Tese: "Identificação de Sistemas Não-Lineares Usando Modelos de Volterra Baseados em Funções Ortonormais de Kautz e Generalizadas"

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



Prof. Dr. Gerard Favier:

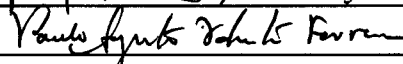
Prof. Dr. Gustavo Henrique da Costa Oliveira:



Prof. Dr. Rafael Santos Mendes:



Prof. Dr. Paulo Augusto Valente Ferreira:



Este trabalho é dedicado
aos meus pais, Ciro da Rosa
e Vanda da Rosa.

Agradecimentos

Eu gostaria primeiramente de agradecer ao meu orientador, professor Dr. Wagner Caradori do Amaral, por um total de seis anos de supervisão, desde o mestrado. Agradeço especialmente pelo seu comprometimento em me orientar neste trabalho de doutorado, bem como pelo tempo dispensado por ele ao ler os rascunhos desta tese. Seus comentários críticos sobre o meu trabalho foram muito úteis para o conteúdo e apresentação de meus argumentos e discussões.

Estendo meus agradecimentos ao professor Dr. Ricardo Campello, pela co-orientação deste trabalho. Suas sugestões e comentários permitiram aprimorar a qualidade da tese.

Eu também gostaria de agradecer a todos os amigos e professores da Unicamp. Particularmente, agradeço ao professor Dr. Paulo Valente pelas inúmeras sugestões e leituras de versões iniciais deste trabalho.

Sou grato ainda aos professores que contribuíram para a minha vida acadêmica, desde minha graduação na Universidade Federal de Goiás, e também a alguns colegas de Goiânia que me apoiaram todos esses anos.

Agradeço infinitamente aos meus pais, Ciro e Vanda, aos meus irmãos Alan, Alessandra e Andréia, e à minha noiva Danielle. Mesmo a mais de 800 quilômetros de distância, todos sempre me deram muita força para que eu atingisse meus objetivos.

Ao CNPq, pelo apoio financeiro durante o desenvolvimento deste trabalho.

A Deus, acima de tudo.

Resumo

Este trabalho enfoca a modelagem de sistemas não-lineares usando modelos de Volterra com funções de base ortonormal (*Orthonormal Basis Functions* – OBF). Os modelos de Volterra representam uma generalização do modelo de resposta ao impulso para a descrição de sistemas não-lineares e, em geral, exigem um elevado número de termos para representar os kernels de Volterra. Esta desvantagem pode ser superada representando-se os kernels usando um conjunto de funções ortonormais. O modelo resultante, conhecido como modelo OBF-Volterra, pode ser truncado em um número menor de termos se as funções da base forem projetadas adequadamente. O problema central é como selecionar os pólos livres que completamente parametrizam estas funções, particularmente as funções de Kautz e as funções ortonormais generalizadas (*Generalized Orthonormal Basis Functions* – GOBF). Uma das abordagens adotadas para resolver este problema envolve a minimização de um limitante superior para o erro resultante do truncamento da expansão do kernel. Cada kernel multidimensional é decomposto em um conjunto de bases de Kautz independentes, em que cada base é parametrizada por um par individual de pólos complexos conjugados com a intenção de representar a dinâmica dominante do kernel ao longo de uma dimensão particular. Obtém-se uma solução analítica para um dos parâmetros de Kautz, válida para modelos de Volterra de qualquer ordem. Outra abordagem envolve a otimização numérica das bases de funções ortonormais usadas para a aproximação de sistemas dinâmicos. Esta estratégia é baseada no cálculo de expressões analíticas para os gradientes da saída dos filtros ortonormais com relação aos pólos da base. Estes gradientes fornecem direções de busca exatas para otimizar os pólos de uma dada base ortonormal. As direções de busca, por sua vez, podem ser usadas como parte de um procedimento de otimização para obter o mínimo de uma função de custo que leva em consideração o erro de estimação da saída do sistema. As expressões relativas à base de Kautz e à base GOBF são obtidas. A metodologia proposta conta somente com dados entrada-saída medidos do sistema a ser modelado, isto é, não se exige nenhuma informação prévia sobre os kernels de Volterra. Exemplos de simulação ilustram a aplicação desta abordagem para a modelagem de sistemas lineares e não-lineares, incluindo um sistema real de levitação magnética com comportamento oscilatório. Por último, estuda-se a representação de sistemas dinâmicos incertos baseada em modelos com incerteza estruturada. A incerteza de um conjunto de kernels de Volterra é mapeada em intervalos de pertinência que definem os coeficientes da expansão ortonormal. Condições adicionais são propostas para garantir que todos os kernels do processo sejam representados pelo modelo, o que permite estimar os limites das incertezas.

Palavras-chave: Identificação de sistemas, Sistemas não-lineares, Séries de Volterra, Funções de base ortonormal, Otimização, Métodos baseados no gradiente, Sistemas dinâmicos incertos.

Abstract

This work is concerned with the modeling of nonlinear systems using Volterra models with orthonormal basis functions (OBF). Volterra models represent a generalization of the impulse response model for the description of nonlinear systems and, in general, require a large number of terms for representing the Volterra kernels. Such a drawback can be overcome by representing the kernels using a set of orthonormal functions. The resulting model, so-called OBF-Volterra model, can be truncated into fewer terms if the basis functions are properly designed. The underlying problem is how to select the free-design poles that fully parameterize these functions, particularly the two-parameter Kautz functions and the Generalized Orthonormal Basis Functions (GOBF). One of the approaches adopted to solve this problem involves minimizing an upper bound for the error resulting from the truncation of the kernel expansion. Each multidimensional kernel is decomposed into a set of independent Kautz bases, in which every basis is parameterized by an individual pair of complex conjugate poles intended to represent the dominant dynamic of the kernel along a particular dimension. An analytical solution for one of the Kautz parameters, valid for Volterra models of any order, is derived. Other approach involves the numerical optimization of orthonormal bases of functions used for approximation of dynamic systems. This strategy is based on the computation of analytical expressions for the gradients of the output of the orthonormal filters with respect to the basis poles. These gradients provide exact search directions for optimizing the poles of a given orthonormal basis. Such search directions can, in turn, be used as part of an optimization procedure to locate the minimum of a cost-function that takes into consideration the error of estimation of the system output. The expressions relative to the Kautz basis and to the GOBF are addressed. The proposed methodology relies solely on input-output data measured from the system to be modeled, i.e., no previous information about the Volterra kernels is required. Simulation examples illustrate the application of this approach to the modeling of linear and nonlinear systems, including a real magnetic levitation system with oscillatory behavior. At last, the representation of uncertain systems based on models having structured uncertainty is studied. The uncertainty of a set of Volterra kernels is mapped on to intervals defining the coefficients of the orthonormal expansion. Additional conditions are proposed to guarantee that all the process kernels to be represented by the model, which allows estimating the uncertainty bounds.

Keywords: System identification, Nonlinear systems, Volterra series, Orthonormal basis functions, Optimization, Gradient-based methods, Uncertain dynamic systems.

Sumário

Lista de Figuras	xv
Lista de Tabelas	xix
1 Introdução	1
1.1 Histórico sobre o uso das bases de Laguerre, Kautz e GOBF	3
1.2 Organização do trabalho e contribuições	7
2 Modelos de Volterra Baseados em Funções Ortonormais	11
2.1 Modelos de Volterra	12
2.1.1 Exemplo	16
2.1.2 Simetrização dos kernels de Volterra	17
2.2 Modelos baseados em funções ortonormais	18
2.2.1 Motivação inicial	18
2.2.2 Representação matemática	20
2.3 Funções de base ortonormal	22
2.4 Resumo	27
3 Desenvolvimento Ótimo de Modelos de Volterra Usando Bases Múltiplas de Kautz	29
3.1 Formulação e solução do problema	30
3.2 Resultados de simulação	37
3.2.1 Comparação: base única <i>versus</i> bases múltiplas	43
3.3 Estimação simultânea dos kernels de Volterra e dos pólos de Kautz	44
3.3.1 Modelagem de um sistema de levitação magnética	45
3.3.2 Simulação computacional para o levitador	46
3.4 Resumo e contribuições	50
4 Direções de Busca Exatas Para Otimização dos Pólos de Bases Ortonormais	53
4.1 Formulação do problema	54
4.1.1 Técnica <i>back-propagation-through-time</i>	54
4.1.2 Otimização dos pólos	56

4.2	Modelo baseado em funções de Kautz	59
4.2.1	Funções de Kautz com índice par	59
4.2.2	Funções de Kautz com índice ímpar	64
4.2.3	Algoritmo	67
4.3	Modelo baseado em funções ortonormais generalizadas (GOBF)	68
4.3.1	Caso GOBF com pólos reais	69
4.3.2	Caso GOBF com um par de pólos complexos	72
4.3.3	Caso GOBF com dois pares de pólos complexos	75
4.3.4	Caso GOBF com qualquer número de pares de pólos complexos	79
4.4	Generalização para modelos de Volterra de ordem superior	79
4.5	Exemplos ilustrativos	82
4.5.1	Exemplo 4.1: modelo linear	82
4.5.2	Exemplo 4.2: modelo não-linear	86
4.6	Aplicação na modelagem do sistema de levitação magnética	90
4.6.1	Modelo com funções de Kautz	90
4.6.2	Modelo com funções GOBF	95
4.7	Resumo e contribuições	97
5	Modelagem de Sistemas Não-Lineares Incertos Usando Funções Ortonormais	99
5.1	Formulação de modelos com incertezas	100
5.1.1	Modelos OBF lineares incertos	100
5.1.2	Modelos OBF não-lineares incertos	101
5.2	Cálculo de limitantes de incertezas em modelos OBF	104
5.2.1	Modelos lineares	104
5.2.2	Modelos não-lineares	105
5.3	Exemplo ilustrativo	107
5.4	Resumo e contribuições	115
6	Conclusões, Contribuições e Perspectivas Futuras	117
6.1	Conclusões e contribuições	117
6.2	Perspectivas futuras	119
A	Provas e Demonstrações	121
A.1	Demonstração da equação (3.2)	121
A.2	Demonstração do Teorema 3.1.1	122
A.3	Demonstração da factibilidade da solução (3.17)	124
A.4	Demonstração do Teorema 3.1.2	124
A.5	Demonstração do Teorema 3.1.4	126
B	Análise do Custo Computacional do Método Proposto no Capítulo 4	129
B.1	Modelo de Kautz linear	129
B.2	Modelo GOBF linear	130

B.3 Modelos de Volterra	131
C Trabalhos Publicados	133
Referências Bibliográficas	135

Lista de Figuras

2.1	Representação esquemática de um sistema representado por uma série de Volterra.	14
2.2	Sistema não-linear com kernel de Volterra de segunda ordem.	17
2.3	Diagrama de blocos de um sistema linear que gera as funções ortonormais $\psi_{l,n}(k)$ em (2.23).	24
3.1	Kernel de segunda ordem com dinâmicas dominantes diferentes em cada direção.	35
3.2	Planos transversais do kernel de Volterra ilustrado na Figura 3.1, mostrando as dinâmicas dominantes em cada direção.	36
3.3	Kernel em (3.21) para $\rho_1 = 0,45$, $\rho_2 = 0,7$, $\omega_1 = 100$ e $\omega_2 = 1$	38
3.4	Aproximação do kernel em (3.21) usando $M = 8$ funções de Kautz com os parâmetros sub-ótimos $(b_1; c_{\text{otimo},1}) = (0,6; -0,2321)$ e $(b_2; c_{\text{otimo},2}) = (0,5; -0,3058)$	40
3.5	Superfície de erro de aproximação do kernel em (3.21) usando $M = 8$ funções de Kautz com os parâmetros sub-ótimos $(b_1; c_{\text{otimo},1}) = (0,6; -0,2321)$ e $(b_2; c_{\text{otimo},2}) = (0,5; -0,3058)$	41
3.6	Superfície de erro de aproximação do kernel em (3.21) usando $M = 8$ funções de Kautz com os parâmetros ótimos $(b_1; c_{\text{otimo},1}) = (0,770; -0,4583)$ e $(b_2; c_{\text{otimo},2}) = (0,345; -0,2473)$	42
3.7	Diagrama do levitador magnético.	45
3.8	Dados entrada/saída do levitador magnético. Acima: corrente através da Bobina 2 (sinal de entrada u). Abaixo: posição do Disco 1 (sinal de saída y).	47
3.9	Kernel de primeira ordem estimado para o modelo do levitador magnético.	49
3.10	Kernel de segunda ordem estimado para o modelo do levitador magnético.	49
3.11	Saída medida (em linha contínua) do sistema de levitação magnética e saída do modelo (em pontilhado).	50
4.1	Esquema da técnica <i>back-propagation-through-time</i>	56
4.2	Estrutura em que a saída do modelo é utilizada na recursão temporal.	56

4.3	Estrutura geral de um modelo OBF: a saída é gerada por um mapeamento estático aplicado a um conjunto de funções ortonormais.	57
4.4	Diagrama de blocos de um modelo dinâmico baseado em funções de Kautz.	60
4.5	Diagrama de blocos de um modelo dinâmico baseado em funções GOBF com pólos reais.	70
4.6	Evolução de um dos pólos ao longo das iterações, para o sistema linear com resposta ao impulso (4.87) com $\zeta = 0, 7 + i0, 5$	83
4.7	Validação do modelo. Acima: saída real do sistema linear com resposta ao impulso (4.87) com $\zeta = 0, 7 + i0, 5$. Abaixo: modelo correspondente usando os pólos ótimos obtidos.	84
4.8	Erro de estimação da saída do sistema linear com resposta ao impulso dada por (4.87).	85
4.9	Kernel de primeira ordem original (à esquerda) e reproduzido (à direita) com a base GOBF parametrizada pelos pólos ótimos obtidos. .	88
4.10	Kernel de segunda ordem original (à esquerda) e reproduzido (à direita) com a base GOBF parametrizada pelos pólos ótimos obtidos. .	89
4.11	Saídas real, do modelo e sinal de erro obtidos com a base GOBF parametrizada pelos pólos ótimos obtidos.	90
4.12	Evolução, ao longo das iterações, do pólo de primeira ordem para o modelo do sistema de levitação magnética, mostrada juntamente com algumas curvas de nível para EQS.	92
4.13	Evolução, ao longo das iterações, do pólo de segunda ordem para o modelo do sistema de levitação magnética, mostrada juntamente com algumas curvas de nível para EQS.	92
4.14	<i>Zoom</i> na região próxima ao pólo final obtido para o kernel de primeira ordem $\beta_{\text{ótimo}} = 0, 7328 \pm i0, 2956$	93
4.15	<i>Zoom</i> na região próxima ao pólo final obtido para o kernel de segunda ordem $\beta_{\text{ótimo}} = 0, 8125 \pm i0, 2863$	93
4.16	Validação do modelo do levitador magnético. Saída real ($y(k)$ em azul) e saída do modelo ($\hat{y}(k)$ em vermelho) usando os pólos ótimos obtidos.	94
4.17	Erro de estimação entre a saída real e a saída do modelo do levitador magnético.	95
4.18	Evolução, ao longo das iterações, dos pólos do termo de primeira ordem para o modelo do levitador magnético usando funções GOBF. .	96
4.19	Evolução, ao longo das iterações, dos pólos do termo de segunda ordem para o modelo do levitador magnético usando funções GOBF. . .	96
5.1	Conjunto de kernels de primeira ordem gerado por (5.18) (linhas pontilhadas), mostrado juntamente com os limites inferior e superior deste conjunto (linhas contínuas).	108

5.2	Limites inferior (azul) e superior (vermelho) do conjunto de kernels de segunda ordem $\{h_2(k_1, k_2)\}$ gerado por (5.19).	109
5.3	Máxima variação sobre o kernel de primeira ordem central $\bar{h}_1(k_1)$. . .	110
5.4	Máxima variação sobre o kernel de segunda ordem central $\bar{h}_2(k_1, k_2)$. .	111
5.5	Comparação entre os limitantes inferior e superior do kernel incerto de primeira ordem calculados utilizando a abordagem da seção 5.1 e a abordagem da seção 5.2.	112
5.6	<i>Zoom</i> na região $6 \leq k_1 \leq 20$	113
5.7	<i>Zoom</i> na região $16 \leq k_1 \leq 32$	113
5.8	Comparação entre os limitantes inferior e superior do kernel incerto de segunda ordem calculados utilizando a abordagem da seção 5.1 e a abordagem da seção 5.2.	114
5.9	<i>Zoom</i> na região $8 \leq k_2 \leq 20$	115

Lista de Tabelas

3.1	Valores sub-ótimos dos parâmetros e dos pólos de Kautz para cada direção do kernel $h_2(k_1, k_2)$ em (3.21).	39
3.2	Erro de aproximação do kernel h_2 em (3.21) de acordo com o número de funções de Kautz utilizadas na expansão ortonormal.	40
3.3	Valores ótimos dos parâmetros e dos pólos de Kautz para o kernel $h_2(k_1, k_2)$ em (3.21).	41
3.4	Erro de aproximação do kernel h_2 em (3.21) usando base única de Kautz.	43
3.5	Valores finais dos parâmetros de Kautz para os kernels de primeira e segunda ordem do modelo do levitador magnético.	48
4.1	Valores finais dos parâmetros e pólos de Kautz para o sistema com resposta ao impulso (4.87) com $\zeta = 0, 7 + i0, 5$	83
4.2	Pólos ótimos de Kautz obtidos para diferentes valores de ζ em (4.87), com a precisão mantida em 10^{-5}	85
4.3	Pólos ótimos obtidos para a base GOBF do modelo (4.91) usando $M = 4$ funções.	88
4.4	Comparação dos erros obtidos usando diferentes abordagens para otimizar os parâmetros do modelo (4.91).	89
4.5	Valores ótimos dos parâmetros e pólos de Kautz obtidos para o sistema de levitação magnética.	91
4.6	Pólos obtidos para o modelo do sistema de levitação magnética com GOBF.	97
4.7	Comparação dos erros obtidos usando diferentes abordagens para otimizar os parâmetros do modelo do sistema de levitação magnética.	97

Capítulo 1

Introdução

O processo de modelagem de sistemas reais apresenta uma grande importância em quase todas as áreas da ciência. Os modelos são úteis para análise e projeto dos sistemas, ou seja, auxilia no entendimento e fornece maior compreensão das características dos mesmos. Neste aspecto, os modelos tornam possível prever ou simular o comportamento de um determinado sistema. Em engenharia, modelos são necessários para o projeto de novos processos e para a análise de processos existentes. Normalmente, técnicas avançadas de projeto de controladores, otimização e supervisão são baseadas em modelos dos processos. A qualidade do modelo tipicamente determina a qualidade da solução final do problema, dando assim à modelagem uma relevante importância no desenvolvimento do projeto. Como consequência, surge uma demanda por esquemas avançados de modelagem e identificação.

O estudo de áreas envolvendo projetos de identificação, modelagem e controle de sistemas dinâmicos tem crescido em importância devido a várias razões. Antes dos computadores digitais, os cálculos usados nessas aplicações consumiam muito tempo, e os erros provenientes de simulações eram consideráveis. Hoje em dia o uso de computadores se expandiu, permitindo a análise de sistemas de controle através de modelos mais detalhados para o seu projeto. Com o aumento do poder computacional, pode-se impor especificações de desempenho mais rígidas nos projetos de engenharia, levando à necessidade de modelos mais detalhados do processo em estudo, especialmente no que diz respeito a previsão de seu comportamento transitório.

Há algumas décadas, identificação de sistemas não-lineares era um campo de abordagens superficiais, cada uma delas aplicável a uma classe restrita de sistemas. Com a introdução de técnicas modernas de otimização, uma classe maior de sistemas tem sido estudada, permitindo assim o uso de uma mesma abordagem para uma variedade de sistemas não-lineares com diferentes propriedades. Nesse escopo, inserem-se os modelos de Volterra, comumente utilizados para a representação de processos não-lineares.

Este trabalho aborda o problema de modelagem de sistemas não-lineares utilizando modelos de Volterra (Schetzen 1980, Rugh 1981, Nelles 2001, Doyle III *et*

al. 2002, Aguirre 2007). Os modelos de Volterra são uma classe de modelos polinomiais sem realimentação do sinal de saída e possuem importantes aplicações em estudos envolvendo identificação e controle de processos dinâmicos. Estes modelos têm grande relevância na representação matemática de sistemas não-lineares devido à sua estrutura geral não-linear, além de serem uma generalização direta do modelo de resposta ao impulso, que relaciona a entrada e a saída de sistemas lineares (Eykhoff 1974, Schetzen 1980, Aguirre 2007, Campello e Oliveira 2007c).

Os modelos de Volterra são parametrizados por funções multidimensionais conhecidas como kernels de Volterra. Em geral, um número elevado de parâmetros é necessário para representar os kernels, que são respostas ao impulso generalizadas. Entretanto, o número de termos necessários para a representação dos modelos pode ser reduzido usando-se uma estrutura do tipo Wiener (Wiener 1958) ortogonalizada em uma base de funções apropriada. Esta 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). Nesta abordagem, cada kernel do modelo é descrito matematicamente por meio de uma expansão em uma base de funções ortonormais (*Orthonormal Basis Functions* – OBF). O modelo resultante, também conhecido como modelo OBF-Volterra, permite aproximar o sistema em estudo usando um número menor de termos se a base de funções for projetada adequadamente. As funções ortonormais são geralmente caracterizadas por parâmetros de projeto livre (pólos). A parametrização de modelos de Volterra por meio de uma estrutura baseada em funções ortonormais é atrativa, dentre outras razões, porque o modelo resultante é linear nos parâmetros (Eykhoff 1974, Schetzen 1980). Esta característica permite usar técnicas de regressão linear para identificar o sistema a partir de medições realizadas. Modelos que utilizam bases de funções ortonormais na sua representação também são conhecidos como modelos OBF.

Em princípio, um conjunto de funções ortonormais é capaz de modelar qualquer sistema estável, independentemente de sua dinâmica (Nelles 2001). A utilização de funções ortonormais permite uma redução na ordem da representação do modelo, simplificando assim os problemas de identificação e controle associados (Van den Hof *et al.* 1995, Heuberger *et al.* 2005). Uma abordagem com bases ortonormais possibilita também incorporar no processo de identificação um conhecimento prévio sobre a dinâmica do sistema (Heuberger *et al.* 1995, Ninness e Gustafsson 1997). Assim, o problema de estimação dos coeficientes do modelo se beneficia da rápida taxa de convergência, implicando em um menor número de termos a serem determinados para modelar o sistema com uma precisão satisfatória. Esta condição provoca uma consequente redução na variância da estimação do modelo.

Outras características desejáveis podem ser obtidas com o uso de funções ortonormais em modelos de sistemas dinâmicos. Dentre elas, pode-se citar a ausência de realimentação da saída, o que por sua vez não provoca realimentação de erros de previsão de horizonte longo, normalmente levando a modelos mais precisos (Nelles 2001). Além disso, não é necessário conhecer os termos passados relevantes

dos sinais de entrada/saída do sistema. É possível também aumentar a capacidade de representação dos modelos simplesmente incrementando o número de funções ortonormais da base. Modelos baseados em funções ortonormais exibem ainda tolerância a dinâmicas não-modeladas e sensibilidade reduzida aos parâmetros estimados (Dumont e Fu 1993).

Potenciais aplicações das funções ortonormais residem no campo de identificação de sistemas e processamento de sinais (Oliveira e Silva 1995a, Heuberger *et al.* 2005). Outras aplicações destas funções incluem também teorias de circuitos, de controle e de otimização de sistemas. A decomposição de sinais e sistemas em termos de componentes ortonormais pode ter um papel importante em algoritmos de otimização correspondentes, uma vez que a escolha de estruturas ortonormais pode ser feita dependente de conhecimento prévio do sistema a ser descrito (Heuberger *et al.* 1995, Van den Hof *et al.* 1995).

Modelos lineares e não-lineares baseados em conjuntos de funções ortonormais têm sido usados em estudos envolvendo identificação e controle de processos dinâmicos (Schetzen 1980, Dumont e Fu 1993, Heuberger *et al.* 1995, Bokor e Schipp 1998, Oliveira *et al.* 2000, Doyle III *et al.* 2002, Heuberger *et al.* 2005, Campello *et al.* 2007a, Campello *et al.* 2007b). Em princípio, uma base ortonormal pode ser construída selecionando-se os parâmetros de projeto livre (pólos) de maneira arbitrária. Entretanto, se a escolha da base for adequada para o sistema em estudo, uma boa aproximação pode ser gerada com um número reduzido de parâmetros. Muitos trabalhos recentes têm estudado a seleção ótima dos pólos que caracterizam os conjuntos de funções ortonormais discretas mais comumente utilizadas na representação de sistemas lineares e não-lineares, que são as bases de Laguerre, de Kautz, e as funções ortonormais generalizadas (*Generalized Orthonormal Basis Functions* – GOBF) (Broome 1965, Wahlberg e Mäkilä 1996, Bokor e Schipp 1998, Heuberger *et al.* 2005).

1.1 Histórico sobre o uso das bases de Laguerre, Kautz e GOBF

Nas últimas décadas, tem havido grande interesse no uso de funções ortonormais (discretas e contínuas) de Laguerre para o estudo de aproximação, modelagem, filtragem, identificação e controle de sistemas (Clowes 1965, Schetzen 1980, Masnadi-Shirazi e Ahmed 1991, Wahlberg 1991, Fu e Dumont 1993, Oliveira e Silva 1994, Tanguy *et al.* 1995, Campello *et al.* 2004, Kibangou *et al.* 2005b, Campello *et al.* 2007a, Campello *et al.* 2007b). Uma importante vantagem das funções de Laguerre é que elas possuem transformadas que são funções racionais com uma forma repetitiva simples, por serem parametrizadas por um único pólo real. Na área de modelagem, quando um modelo truncado usando funções de Laguerre é utilizado, um problema de projeto consiste da seleção ótima dos pólos que minimizam o erro de aproximação.

A seleção de pólos da base de Laguerre já foi estudada exhaustivamente na literatura. Um dos primeiros trabalhos nesse sentido abordou a aproximação de sistemas contínuos lineares (Clowes 1965). Investigou-se também a seleção ótima de pólos de Laguerre para uma classe particular de sistemas lineares (Masnadi-Shirazi e Ahmed 1991), propondo-se métodos *off-line* para otimizar a parametrização de modelos lineares (modelos de Volterra de primeira ordem) baseados na minimização da energia do erro de aproximação da resposta ao impulso. Estes métodos fornecem o valor ótimo do pólo de Laguerre, porém eles requerem um custo computacional relativamente alto, pois há a necessidade de se calcular raízes reais de polinômios de ordem elevada.

Métodos que fornecem uma solução analítica global para a escolha do pólo ótimo de Laguerre na representação de sistemas lineares são apresentados na literatura (Fu e Dumont 1993, Tanguy *et al.* 1995). Em um destes trabalhos (Fu e Dumont 1993), otimiza-se uma função de custo quadrática que força uma rápida convergência da série ortonormal através de um aumento linear do custo atribuído a cada coeficiente adicional. Em outro trabalho (Tanguy *et al.* 1995), propõe-se um método que não exige conhecimento completo do sinal a ser aproximado, sendo assim adequado para lidar com dados experimentais. No que diz respeito ao estudo de sistemas não-lineares discretos, foram estudados inicialmente os modelos OBF-Volterra de segunda ordem (Campello *et al.* 2001), e posteriormente os modelos de qualquer ordem (Campello 2002, Campello *et al.* 2003, Campello *et al.* 2004, Campello *et al.* 2006, Campello *et al.* 2007a), apresentando-se uma solução analítica global para a escolha ótima do pólo de Laguerre. Nestes trabalhos, adotou-se uma função de custo que força a convergência da série aumentando-se linearmente o custo designado a cada coeficiente da expansão ortonormal. Foram obtidas expressões analíticas para os pólos ótimos de Laguerre quando os kernels de Volterra são parametrizados por uma base única (Campello *et al.* 2003, Campello *et al.* 2004) e também por bases distintas (Campello *et al.* 2006). Estes resultados representam extensões para o contexto de sistemas não-lineares de um estudo pioneiro que estudou modelos lineares (Fu e Dumont 1993). Estudos envolvendo a seleção de pólos da base de Laguerre inclui ainda outros trabalhos (Wahlberg 1991, Oliveira e Silva 1995a, Ngia 2001).

Dinâmicas pouco amortecidas, entretanto, são difíceis de aproximar com um número pequeno de funções de Laguerre. As funções de Kautz conseguem aproximar sinais com comportamento oscilatório com uma maior precisão. Por este motivo, estas funções têm recebido maior atenção em trabalhos recentes (Wahlberg 1994, Oliveira e Silva 1995b, den Brinker *et al.* 1996, Tanguy *et al.* 2002, da Rosa *et al.* 2007, da Rosa *et al.* 2008a). O problema de identificação de sistemas lineares e invariantes no tempo por meio de uma expansão finita usando funções de Kautz no domínio da frequência já foram abordados (Wahlberg 1994). Condições de otimalidade para a aproximação de sistemas SISO – *Single Input-Single Output* – lineares por meios de uma expansão truncada utilizando funções de Kautz também já foram estudados (Oliveira e Silva 1995b, den Brinker *et al.* 1996). Estas condições de otimalidade são de grande interesse teórico, entretanto elas podem resultar em

cálculos complicados em casos práticos (Tanguy *et al.* 2000). A relação entre as bases de funções de Laguerre e de Kautz foi estudada através de uma análise de mapeamentos no domínio complexo z (Wahlberg e Mäkilä 1996). De acordo com as discussões apresentadas, essa análise fornece uma intuição sobre que tipo de funções ortonormais deveriam ser utilizadas dado um certo conhecimento prévio a respeito do sistema linear a ser modelado.

No contexto da seleção dos pólos da base de Kautz, um procedimento para escolher adequadamente os parâmetros de Kautz no caso de sistemas lineares com um número finito de funções utilizado na expansão dos modelos foi proposto na literatura (Tanguy *et al.* 2000). Uma expressão analítica para a seleção sub-ótima dos pólos de Kautz na representação de sistemas lineares discretos também já foi apresentada (Tanguy *et al.* 2002). Estes resultados foram posteriormente estendidos para o contexto de modelos de Volterra (da Rosa 2005, da Rosa *et al.* 2005, da Rosa *et al.* 2007). Nesta extensão, a estratégia apresentada é válida quando os kernels do modelo são previamente conhecidos, além de adotar a mesma base de Kautz para representar todas as direções do kernel. A solução consiste de uma solução analítica global para a escolha ótima de um dos parâmetros de Kautz na representação de modelos de Volterra. Esta solução é baseada na minimização de um limitante superior resultante da aproximação truncada dos kernels de Volterra usando bases de funções ortonormais de Kautz.

Apesar dos importantes resultados teóricos e dos avanços atingidos com relação às abordagens analíticas para otimizar os pólos das bases ortonormais, todas as soluções que foram obtidas até hoje possuem pelo menos uma das seguintes desvantagens:

1. Os kernels do sistema precisam ser conhecidos *a priori*. Esta é uma difícil imposição na prática, pois os kernels são funções não-parametrizadas cuja medição só se torna possível se suas contribuições individuais puderem ser separadas da resposta total do sistema;
2. A solução é apenas sub-ótima, pois minimiza um limitante superior do erro de aproximação do modelo e/ou otimiza somente um sub-conjunto dos parâmetros da base;
3. A solução é restrita às bases de Laguerre ou de Kautz. Não se tem conhecimento de trabalhos na literatura que tenham obtido soluções analíticas para otimização dos pólos de bases GOBF.

Levando-se em consideração estas desvantagens relacionadas com a otimização analítica dos pólos das bases ortonormais, uma outra abordagem apresentada na literatura para a seleção de pólos está relacionada com procedimentos numéricos de otimização (Bodin *et al.* 1997, Sarroukh *et al.* 2001, Hacioğlu *et al.* 2001, Kibangou *et al.* 2003, Kibangou *et al.* 2005a, Kibangou *et al.* 2005c, Patwardhan *et al.* 2005). Pelas razões descritas acima, estratégias numéricas são mandatórias em problemas

de otimização usando modelos baseados em GOBF, ou quando não se conhece os kernels do sistema previamente. Foram propostos métodos iterativos de seleção de pólos para representação de modelos OBF-Volterra de segunda ordem (Kibangou *et al.* 2003, Kibangou *et al.* 2005c). A desvantagem destes métodos é que eles possuem alta demanda computacional, pois são baseados em uma busca exaustiva dos pólos da base. Em cada passo do procedimento, uma nova função é selecionada a partir de um conjunto de funções GOBF que possui pólos reais e distintos entre si. O pólo associado com cada nova função é escolhido como um elemento de um conjunto candidato de pólos que resultam da discretização de seu intervalo de factibilidade. Esta estratégia foi posteriormente melhorada ao se reduzir a complexidade computacional, e ainda estendendo-se os resultados para modelos de Volterra de qualquer ordem (Kibangou *et al.* 2005a). Conforme comentado pelos autores desses trabalhos (veja e.g. (Kibangou *et al.* 2005b)), não se tem nenhuma análise de convergência desses métodos.

Uma teoria geral das funções ortonormais foi desenvolvida para a modelagem de sistemas lineares usando GOBF (Heuberger *et al.* 1995, Van den Hof *et al.* 1995), explorando-se a conexão entre o comportamento dinâmico de um sistema e as funções de base ortonormal. Por meio de métodos de identificação do tipo mínimos quadrados, observou-se que um número reduzido de coeficientes precisa ser estimado quando a base de funções é escolhida adequadamente. Resultados mostraram que esta escolha apropriada leva a modelos aproximados mais exatos com poucos coeficientes estimados. Mostrou-se ainda para o caso linear que a qualidade de um modelo baseado em funções ortonormais aumenta rapidamente quando a sua dinâmica se aproxima da dinâmica do sistema real.

Conforme será discutido no Capítulo 2, a expressão geral das funções GOBF possui a desvantagem de gerar funções com respostas ao impulso complexas. Para superar esta propriedade indesejável, um novo método para a construção de GOBF foi desenvolvido (Ninness e Gustafsson 1997). Esta metodologia fornece um número infinito de novas bases ortonormais que permitem incorporar um conhecimento prévio sobre qualquer número e tipo de pólos estáveis. No aspecto da seleção de pólos em modelos lineares usando GOBF, condições de otimalidade para séries truncadas também foram apresentadas (Oliveira e Silva 1997, Heuberger *et al.* 2005), discutindo-se aspectos teóricos sobre a estacionariedade de uma função objetivo na seleção ótima de pólos das funções ortonormais. Esses resultados são uma extensão de estudos anteriores realizados para a base de Laguerre (Oliveira e Silva 1994) e para a base de Kautz (Oliveira e Silva 1995b).

Outros trabalhos recentes (Nalbantoğlu *et al.* 2003, Ziaei e Wang 2006) utilizaram funções GOBF para identificação de sistemas físicos reais. O primeiro destes trabalhos estudou a alocação de pólos estimados a partir de funções de transferência não-paramétricas usando técnicas de identificação no domínio da frequência. Já o segundo abordou a identificação de sistemas baseados em GOBF aplicada a um sistema onde a função de transferência do modelo possui pólos reais e complexos. Uma estratégia de otimização global foi implementada para obter a alocação de pólos

de funções generalizadas. Ainda no contexto do uso de funções GOBF na identificação de sistemas, abordou-se a aplicação destas funções em processos estocásticos (Ninness *et al.* 1999).

Uma formulação geral de otimização numérica na representação de sistemas lineares que envolve a decomposição do problema de identificação do modelo completo em um procedimento iterativo de dois passos é apresentada na literatura (Patwardhan *et al.* 2005): uma atualização nos parâmetros da base ortonormal (pólos) seguido por uma atualização dos parâmetros restantes do modelo mantendo-se os pólos fixos. Esta decomposição reduz a complexidade do problema de identificação, porém nenhuma estratégia foi fornecida para se determinar a direção de busca a ser seguida pelo algoritmo de otimização. A principal dificuldade neste aspecto é devido ao fato de que os pólos da base estão relacionados com a saída do modelo através de equações dinâmicas. Um trabalho complementar (Patwardhan *et al.* 2006) abordou o controle preditivo baseado em funções ortonormais. O clássico método de descida do gradiente já foi aplicado para a seleção dos pólos das bases de Laguerre, de Kautz e GOBF na descrição de sistemas não-lineares (Hacıoğlu *et al.* 2001). Entretanto, a formulação do gradiente e as correspondentes direções de busca foram somente aproximadas por meio de estimativas instantâneas (estáticas) de suas derivadas parciais.

Os objetivos deste trabalho envolvem apresentar resultados a respeito de modelos ótimos que simplifiquem suas implementações computacionais. As metodologias abordadas nesse sentido envolvem a representação de sistemas dinâmicos não-lineares por meio de modelos de Volterra baseados em funções ortonormais.

1.2 Organização do trabalho e contribuições

O presente trabalho está organizado da seguinte forma:

- **Capítulo 2 – Modelos de Volterra Baseados em Funções Ortonormais:** Discute os modelos discretos de Volterra e sua relevância na modelagem de sistemas dinâmicos, além dos aspectos matemáticos dessa representação. O capítulo apresenta as séries de Volterra como uma generalização do modelo de resposta ao impulso, bem conhecido no estudo de sistemas lineares. O número elevado de parâmetros geralmente necessários para representar os kernels de Volterra pode ser reduzido ao se descrever cada kernel usando uma base de funções ortonormais. O modelo resultante, conhecido como modelo OBF-Volterra, consiste de uma dinâmica linear composta por um conjunto de filtros ortonormais, seguida por um mapeamento estático, aqui representado pela série de Volterra. Esta nova representação requer um número reduzido de termos se a base ortonormal utilizada for escolhida adequadamente. Apresentam-se ainda as bases de funções ortonormais mais comumente utilizadas na modelagem de sistemas dinâmicos. Dentre elas, destacam-se as bases de Laguerre e de Kautz, que são um caso especial das funções ortonor-

mais generalizadas (GOBF). As funções ortonormais geralmente são caracterizadas por pólos livres, o que motiva o estudo da sua seleção ótima. Este problema é tema de estudo dos Capítulos 3 e 4.

- **Capítulo 3 – Desenvolvimento Ótimo de Modelos de Volterra Usando Bases Múltiplas de Kautz:** Estuda a seleção dos pólos livres que parametrizam as bases de funções ortonormais de Kautz. Por meio de um teorema, obtém-se um limitante superior para o erro resultante do truncamento da expansão dos kernels de Volterra em um número finito de funções da base de Kautz. O problema central do capítulo é abordado minimizando-se esse limitante superior do erro de aproximação, obtendo-se assim uma solução analítica sub-ótima para a parametrização das bases de Kautz utilizadas na expansão dos kernels de Volterra. Nesta abordagem, que é válida para modelos de Volterra de qualquer ordem, cada direção dos kernels é parametrizado por uma base de Kautz individual. Este resultado, que também é apresentado em (da Rosa *et al.* 2008a), representa uma generalização de trabalhos anteriores que utilizaram uma única base de Kautz para todas as direções dos kernels (da Rosa 2005, da Rosa *et al.* 2005, da Rosa *et al.* 2006, da Rosa *et al.* 2007). Esta abordagem também pode ser vista como uma extensão, para o contexto da base de Kautz, de outro trabalho que investigou esse problema no escopo da base de Laguerre (Campello *et al.* 2006). Deseja-se assim fazer uma comparação entre as diferentes abordagens. Desta forma, é esperado que o uso de bases múltiplas reduza o erro de truncamento quando o kernel apresenta dinâmicas diferentes entre si em cada uma de suas direções. Propõe-se ainda um procedimento iterativo para a estimação simultânea dos kernels de Volterra e dos pólos de Kautz, que é aplicado na modelagem de um sistema não-linear de levitação magnética com comportamento oscilatório.
- **Capítulo 4 – Direções de Busca Exatas Para Otimização dos Pólos de Bases Ortonormais:** Apresenta resultados referentes à aproximação de modelos OBF-Volterra por meio da seleção numérica dos pólos das bases ortonormais. O problema central abordado, também apresentado em (da Rosa *et al.* 2008b, da Rosa *et al.* aceito), envolve a minimização de uma função de custo quadrática que leva em conta o erro de estimação da saída do sistema. Particular ênfase é dada para modelos baseados em funções de Kautz e para modelos baseados em GOBF, embora um estudo relativo às funções de Laguerre pode ser facilmente adaptado como um caso particular. Calcula-se analiticamente os gradientes das saídas dos filtros de Kautz e GOBF em relação aos parâmetros (pólos) destas funções usando uma técnica *back-propagation-through-time*. Estes gradientes podem indicar uma direção de busca a ser seguida por um algoritmo de otimização não-linear para a seleção ótima dos pólos. Em particular, adota-se aqui o algoritmo de Levenberg-Marquardt (Levenberg 1944, Marquardt 1963, Nelles 2001) que vem a ser uma escolha usual em problemas de otimização não-linear de porte médio (Nocedal

e Wright 1999, Nelles 2001). Os resultados teóricos obtidos são aplicados na modelagem de sistemas lineares e não-lineares, inclusive do sistema de levitação magnética descrito no Capítulo 3.

- **Capítulo 5 – Modelagem de Sistemas Não-Lineares Incertos Usando Funções Ortonormais:** Aborda o cálculo de limitantes de incertezas de sistemas não-lineares incertos usando modelos de Volterra baseados em funções ortonormais. Estas incertezas são formuladas sobre os kernels de Volterra do modelo e são mapeadas em intervalos de pertinência que definem os coeficientes da expansão ortonormal. Uma série de problemas de otimização não-linear são propostos para estimar os limitantes das incertezas, permitindo assegurar que os comportamentos do processo incerto estejam representados no modelo. Os resultados teóricos apresentados representam uma extensão, para o contexto dos modelos não-lineares de Volterra, de estudos anteriores que abordaram esse problema para os modelos lineares (Oliveira 1997, Oliveira *et al.* 2000, Moreira 2006) . A metodologia proposta é aplicada para obter a representação de sistemas não-lineares incertos por meio de modelos baseados em funções ortonormais. No exemplo ilustrativo apresentado, o modelo adotado utiliza funções de Kautz.
- **Capítulo 6 – Conclusões, Contribuições e Perspectivas Futuras:** Apresenta as conclusões deste trabalho, bem como as contribuições na área de modelagem de sistemas lineares e não-lineares. O capítulo discute ainda as perspectivas para trabalhos futuros.

Capítulo 2

Modelos de Volterra Baseados em Funções Ortonormais

A modelagem de sistemas apresentada neste capítulo é baseada na representação em modelos de Volterra. Estes modelos constituem uma classe de modelos polinomiais não-recursivos, isto é, sem realimentação da saída. Embora seja uma ferramenta matemática, a sua aplicação à representação entrada/saída de sistemas pode ser discutida inicialmente sem a necessidade de um desenvolvimento matemático mais elaborado. Serão discutidas as principais vantagens e desvantagens dos modelos de Volterra na modelagem de sistemas lineares e não-lineares. Aborda-se também o conceito de simetrização dos kernels de Volterra, que se apresenta como uma ferramenta matemática para se obter uma representação de sistemas via modelos de Volterra com menor complexidade.

Este capítulo tem o objetivo de apresentar a terminologia, a notação e as manipulações básicas relacionadas com as representações de sistemas não-lineares por modelos de Volterra. Esses modelos são parametrizados por funções multidimensionais, chamadas kernels de Volterra, e representam uma generalização do modelo de resposta ao impulso para a descrição de sistemas não-lineares. Em princípio, um número elevado de parâmetros é necessário para representar os kernels de Volterra (respostas ao impulso generalizadas) pois, como os modelos de Volterra não possuem realimentação do sinal de saída, eles geralmente necessitam de um grande número de termos (Schetzen 1980, Nelles 2001, Doyle III *et al.* 2002, Aguirre 2007, Campello e Oliveira 2007c). Esta desvantagem pode ser contornada descrevendo-se cada kernel por meio de uma expansão em bases de funções ortonormais. O modelo resultante em geral exige um número menor de termos para representar o sistema.

Apresenta-se também os conjuntos de funções ortonormais mais comumente utilizados na aproximação de sinais e de sistemas, que são as bases de Laguerre, de Kautz e GOBF (Broome 1965, Wahlberg e Mäkilä 1996, Bokor e Schipp 1998, Heuberger *et al.* 2005). Posteriormente, uma maior atenção será dada para as funções de Kautz, abordando-se sua relevância na aproximação de sistemas estáveis invariantes no tempo. Conforme será visto, as funções de Kautz

são parametrizadas por um par de pólos complexos conjugados, permitindo assim uma melhor representação de sistemas com comportamento oscilatório quando comparado com a base de Laguerre, por exemplo.

2.1 Modelos de Volterra

Dentro do contexto dos sistemas lineares, o conhecimento de uma função temporal que caracterize um sistema físico do mundo real (por exemplo, a resposta ao impulso) é suficiente para determinar a saída do sistema para qualquer sinal de entrada (Eykhoff 1974, Schetzen 1980). Uma aplicação comumente utilizada é a previsão da saída de sistemas para projeto de controladores preditivos. Na descrição de um processo não-linear, também seria desejável obter uma função temporal que o caracterize, daí o interesse pelo estudo dos modelos de Volterra.

Os modelos de Volterra são essencialmente uma expansão funcional entrada-saída de um sistema não-linear cuja estrutura é dada por uma generalização direta do modelo de resposta ao impulso (Eykhoff 1974, Nelles 2001, Aguirre 2007). A conexão destes modelos com a teoria de sistemas não-lineares e a sua capacidade de representar uma grande classe de sistemas não-lineares os tornam bastante atrativos.

A representação de sistemas por meio de modelos de Volterra foi introduzida pelo italiano Vito Volterra, no final do século XIX. Entretanto, o primeiro uso da série de Volterra para o estudo de sistemas não-lineares é creditado à Norbert Wiener, já no século XX. A utilização destes modelos permite (Eykhoff 1974):

- obter uma relação explícita entre os sinais de entrada e saída;
- estudar sistemas excitados por sinais aleatórios;
- descrever um sistema não-linear como uma generalização do caso linear;
- estender a modelos não-lineares conceitos definidos somente para os lineares.

Uma importante vantagem dos modelos de Volterra é que eles representam uma extensão direta dos modelos lineares de resposta ao impulso. Descrever um sistema não-linear por meio de funções temporais que o caracterizem completamente é uma vantagem que pode ser utilizada na solução de problemas referentes a esses sistemas. Em outras palavras, as séries de Volterra tornam possível usar alguns conceitos, definidos somente para os sistemas lineares, para representar também os processos não-lineares (Eykhoff 1974).

No domínio do tempo discreto, a descrição matemática dos modelos de Volterra relaciona a saída $y(k)$ de um processo físico causal com sua entrada $u(k)$ como

(Schetzen 1980, Rugh 1981, Doyle III *et al.* 2002):

$$\begin{aligned}
 y(k) = & h_0 + \sum_{\tau_1=0}^{\infty} h_1(\tau_1)u(k - \tau_1) + \sum_{\tau_1=0}^{\infty} \sum_{\tau_2=0}^{\infty} h_2(\tau_1, \tau_2)u(k - \tau_1)u(k - \tau_2) + \\
 & + \cdots + \sum_{\tau_1=0}^{\infty} \cdots \sum_{\tau_\eta=0}^{\infty} h_\eta(\tau_1, \dots, \tau_\eta)u(k - \tau_1) \cdots u(k - \tau_\eta) + \cdots
 \end{aligned} \tag{2.1}$$

onde as funções multivariáveis $h_\eta(\tau_1, \dots, \tau_\eta)$ são os kernels de Volterra de ordem η , e h_0 é um termo adicional inserido apenas para a representação de eventual nível constante não-nulo na saída do sistema. Um modelo de Volterra descreve a saída de um sistema não-linear como a soma das respostas de operadores de primeira ordem, de segunda ordem, de terceira ordem, e assim por diante.

A série de Volterra dada na equação (2.1) representa uma generalização do modelo de resposta ao impulso. Este modelo, tradicionalmente utilizado para análise de sistemas lineares, descreve a saída do sistema como a convolução de sua resposta ao impulso $h(k)$ com a entrada do sistema, i.e.:

$$y(k) = \sum_{\tau=0}^{\infty} h(\tau)u(k - \tau) \tag{2.2}$$

A equação (2.2) descreve os sistemas dinâmicos lineares invariantes no tempo, de forma que a dinâmica do sistema fica completamente determinada por sua resposta ao impulso (Eykhoff 1974, Doyle III *et al.* 2002, Heuberger *et al.* 2005, Aguirre 2007).

Desconsiderando o termo h_0 no modelo de Volterra em (2.1), este modelo pode ser reescrito da seguinte forma:

$$y(k) = \sum_{\eta=1}^{\infty} \sum_{\tau_1=0}^{\infty} \cdots \sum_{\tau_\eta=0}^{\infty} h_\eta(\tau_1, \dots, \tau_\eta) \prod_{l=1}^{\eta} u(k - \tau_l) \tag{2.3}$$

Ou ainda:

$$y(k) = \mathcal{H}_1[u(k)] + \mathcal{H}_2[u(k)] + \cdots + \mathcal{H}_\eta[u(k)] + \cdots \tag{2.4}$$

onde

$$\mathcal{H}_\eta[u(k)] = \sum_{\tau_1=0}^{\infty} \cdots \sum_{\tau_\eta=0}^{\infty} h_\eta(\tau_1, \dots, \tau_\eta)u(k - \tau_1) \cdots u(k - \tau_\eta) \tag{2.5}$$

Como o termo $\mathcal{H}_\eta[u(k)]$ em (2.5) contém contribuições de grau η para a entrada $u(k)$, nota-se que o modelo de Volterra em (2.1) representa uma generalização não-linear polinomial do modelo linear (2.2). Em outras palavras, as funções

$h_\eta(\tau_1, \dots, \tau_\eta)$ são uma generalização de dimensão η (com $\eta = 1, 2, \dots$) da função de resposta ao impulso de um sistema linear. Neste contexto, η representa o grau de não-linearidade do sistema.

A série de Volterra pode ser ilustrada de uma forma esquemática como apresentado na Figura 2.1. Nesta representação, a saída $y(k)$ é descrita como uma soma das contribuições de cada funcional \mathcal{H}_j (com $j = 1, 2, \dots, \eta$) aplicado à entrada $u(k)$, onde os termos \mathcal{H}_j são dados pela equação (2.5).

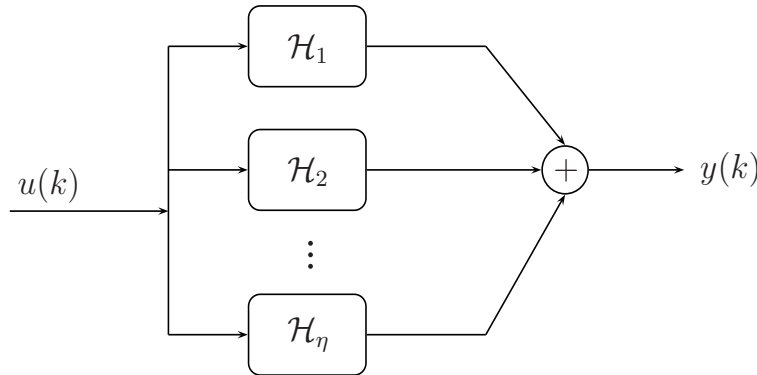


Figura 2.1: Representação esquemática de um sistema representado por uma série de Volterra.

A representação do modelo (2.1) é do tipo

$$y(k) = \mathcal{O}(\{u(\tau)\}_{\tau=-\infty}^{\tau=k}) \quad (2.6)$$

onde \mathcal{O} é um operador não-linear que representa o sistema a ser modelado. Em termos matemáticos, \mathcal{O} pode ser um funcional, ou seja, é uma operação sobre uma função cujo resultado é um número.

Como visto acima, os modelos de Volterra possuem somas de infinitos termos. Consequentemente, devem haver condições de convergência associadas que precisam ser satisfeitas para garantir que a representação de um sistema por tal modelo tenha significado físico. Em alguns casos, é possível investigar a convergência da série de Volterra usando uma série de potências (Schetzen 1980). Entretanto, raramente a série é obtida em uma forma fechada (expressão analítica). Na maioria dos casos, somente uma série truncada pode ser usada, desde que forneça uma estimativa satisfatória da representação quando são desprezados os termos de ordem superior. Se a série de Taylor do operador \mathcal{O} tiver convergência garantida, é possível truncá-la a partir de um certo número de termos. A partir do último termo considerado na aproximação da série, as parcelas de maior ordem podem ser desprezadas. Essa simplificação reduz o esforço computacional requerido nas simulações, sendo capaz de representar com boa precisão grande parte dos operadores \mathcal{O} não-lineares (Boyd e Chua 1985). Detalhes formais sobre a convergência da série de Volterra podem ser encontrados na literatura (Eykhoff 1974, Rugh 1981).

Na modelagem de sistemas dinâmicos estáveis, os kernels de Volterra podem apresentar três propriedades importantes (Eykhoff 1974):

- $h_\eta(\tau_1, \tau_2, \dots, \tau_\eta)$ podem ser construídos de forma a serem simétricos
- $\lim_{\tau_l \rightarrow \infty} h_\eta(\tau_1, \tau_2, \dots, \tau_\eta) = 0 \quad l = 1, 2, \dots, \eta$
- $h_\eta(\tau_1, \tau_2, \dots, \tau_\eta) = 0 \quad \forall \tau_l < 0 \quad l = 1, 2, \dots, \eta$

Quanto à primeira propriedade, tem-se que os termos $h_\eta(\tau_1, \tau_2, \dots, \tau_\eta)$ referentes a qualquer permutação dos índices $\tau_1, \tau_2, \dots, \tau_\eta$ multiplicam o mesmo valor $u(k - \tau_1) \cdots u(k - \tau_\eta)$ na equação (2.3). Logo, pode-se arbitrar a simetria dos kernels sem nenhuma consequência para a capacidade de aproximação do modelo. Neste caso, o problema de estimação dos kernels simplifica-se à estimação referente apenas a uma das metades do domínio. Já a segunda propriedade surge apenas na modelagem de sistemas em que a resposta a uma excitação na entrada desaparece com o tempo até não contribuir mais com a saída significativamente, o que remete aos sistemas estáveis (Boyd e Chua 1985). Sistemas com esta característica são também conhecidos como sistemas com *fading memory*. Esta propriedade é importante, dentre outros fatores, porque permite um truncamento em cada série de Volterra. O truncamento do termo de ordem η , por exemplo, é realizado considerando que $h_\eta(\tau_1, \dots, \tau_\eta) = 0$ para $\tau_l > \epsilon_\eta$ ($\forall l \in \{1, \dots, \eta\}$), onde o escalar ϵ_η representa o período de tempo além do qual a contribuição do respectivo kernel é desprezível. A terceira propriedade advém do fato de o sistema ser considerado causal e com condições iniciais nulas, isto é, a resposta de um sistema físico realizável não depende de valores futuros do sinal de entrada.

Embora uma série de Volterra, como aquela em (2.1), seja capaz de modelar sistemas dinâmicos não-lineares, ela possui algumas desvantagens. A principal delas é a dificuldade de obter um número necessário de kernels para aproximar com precisão um dado sistema. O número de kernels necessários para aproximar um sistema não-linear dá uma ideia do grau de não-linearidade do sistema. Assim, um sistema que é suavemente não-linear pode requerer somente um modelo de Volterra de segunda ordem, enquanto que um sistema altamente não-linear pode necessitar de um número maior de kernels para fornecer resultados satisfatórios.

Dado um sistema dinâmico descrito por um modelo de Volterra como em (2.1), o problema de identificação consiste em estimar o conjunto de kernels $h_\eta(\tau_1, \dots, \tau_\eta)$ do modelo. Este problema é, em princípio, extremamente complicado, uma vez que os kernels são funções multidimensionais não-parametrizadas (Schetzen 1980, Rugh 1981). Uma abordagem intuitiva seria generalizar a ideia dos modelos FIR (*Finite Impulse Response*) (Eykhoff 1974, Heuberger *et al.* 2005, Aguirre 2007), substituindo o valor de cada componente do kernel como um parâmetro individual a ser estimado. Esta abordagem seria possível para a modelagem de sistemas com *fading memory*, em que, com base na segunda propriedade dos kernels de Volterra

discutida anteriormente, o conjunto de parâmetros poderia ser finito através do truncamento dos termos da série. Porém, mesmo no caso dos modelos FIR que envolvem apenas o kernel de primeira ordem, a quantidade de parâmetros necessários para uma aproximação satisfatória pode ser elevada, principalmente se o sistema possuir dinâmicas lentas. A inserção de kernels de ordens superiores provocaria um aumento exponencial da quantidade de parâmetros, o que tornaria essa abordagem infactível para grande parte das aplicações práticas (Billings 1980).

De uma forma geral, os modelos de Volterra são úteis para representação de sinais e sistemas devido a sua estrutura geral não-linear. Entretanto, a principal desvantagem desses modelos é que eles necessitam de um número considerável de termos para a representação dos kernels de Volterra (Eykhoff 1974, Rugh 1981, Nelles 2001, Doyle III *et al.* 2002, Aguirre 2007, Campello e Oliveira 2007c). Em princípio, a medição dos kernels só se torna possível se suas contribuições individuais puderem ser separadas da resposta total do sistema (Schetzen 1980). Nenhum método exato de isolar um operador de Volterra individual existe para modelos que não são de ordem finita. Outra desvantagem reside na convergência da série de Volterra. A representação de um sistema físico por meio de uma série de Volterra converge apenas em um intervalo limitado da amplitude do sinal de entrada.

Uma abordagem interessante para contornar os problemas discutidos acima foi apresentada inicialmente por Wiener (1958), e consiste na expansão dos kernels de Volterra usando conjuntos de funções ortonormais. Com esta abordagem, o número de termos necessários para a representação dos modelos de Volterra pode ser drasticamente reduzido (Schetzen 1980, Rugh 1981, Heuberger *et al.* 2005). Estes modelos serão estudados posteriormente na seção 2.2.

2.1.1 Exemplo

A tarefa de identificar um sistema utilizando uma série de Volterra consiste em estimar os kernels $h_\eta(k_1, k_2, \dots, k_\eta)$ do modelo (2.1). Entretanto este problema pode ser complexo, já que os kernels são funções multidimensionais não-parametrizadas. Este exemplo ilustra a descrição de um sistema não-linear usando um modelo de Volterra com kernel de segunda ordem.

Considere a conexão não-linear de dois sub-sistemas, conforme mostrado na Figura 2.2.

Os dois sub-sistemas desta figura são lineares, caracterizados por suas respectivas respostas ao impulso, $h_a(k)$ e $h_b(k)$. Eles são descritos pela conhecida operação de convolução:

$$y_i(k) = \sum_{\tau=0}^k h_i(\tau)u(k-\tau) \quad i = a, b \quad (2.7)$$

De acordo com o diagrama ilustrado na Figura 2.2, o sistema completo é descrito

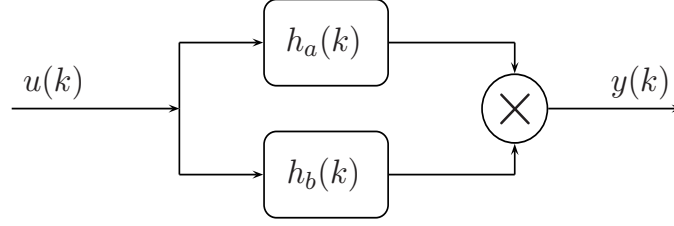


Figura 2.2: Sistema não-linear com kernel de Volterra de segunda ordem.

pelo produto das saídas dos dois sub-sistemas lineares, i.e.:

$$\begin{aligned}
 y(k) &= y_a(k)y_b(k) \\
 &= \sum_{\tau_1=0}^k h_a(\tau_1)u(k-\tau_1) \sum_{\tau_2=0}^k h_b(\tau_2)u(k-\tau_2) \\
 &= \sum_{\tau_1=0}^k \sum_{\tau_2=0}^k h_a(\tau_1)u(k-\tau_1)h_b(\tau_2)u(k-\tau_2) \quad (2.8)
 \end{aligned}$$

Observando-se a equação acima, nota-se que o kernel de segunda ordem do sistema é dado por:

$$h_2(\tau_1, \tau_2) = h_a(\tau_1)h_b(\tau_2) \quad (2.9)$$

Portanto, o sistema da Figura 2.2 pode ser representado por meio de um modelo de Volterra de segunda ordem, conforme visto acima. O kernel de Volterra de segunda ordem deste sistema é o produto das respostas ao impulso individuais de cada um dos dois sub-sistemas que compõem o sistema completo.

2.1.2 Simetrização dos kernels de Volterra

Na análise dos kernels de Volterra, pode haver a necessidade de se realizar operações que envolvam reordenamento dos argumentos τ 's em $h_\eta(\tau_1, \dots, \tau_\eta)$. Qualquer kernel assimétrico pode ser simetrizado, de maneira que a análise pode ser realizada considerando-se somente sua versão simétrica, sem qualquer perda de generalidade (Schetzen 1980). O principal objetivo da simetrização é que considerar que um determinado kernel seja simétrico exige a obtenção de um menor número de parâmetros. Esta propriedade permite simplificar os cálculos computacionais necessários à estimação dos kernels do modelo.

A regra geral para simetrizar um kernel assimétrico é genericamente escrita como (Rugh 1981):

$$h_\eta^{\text{sim}}(\tau_1, \dots, \tau_\eta) = \frac{1}{\eta!} \sum_{\pi(\cdot)} h_\eta(\tau_{\pi(1)}, \dots, \tau_{\pi(\eta)}) \quad (2.10)$$

onde $\pi(\cdot)$ denota qualquer permutação dos inteiros $1, 2, \dots, \eta$. A soma indicada na equação (2.10) é sobre todas as $\eta!$ permutações dos inteiros de 1 a η . Usando essa definição, um kernel de segunda ordem por exemplo é considerado simétrico se $h_2(\tau_1, \tau_2) = h_2(\tau_2, \tau_1)$.

Para ilustrar o processo de simetrização de um kernel de Volterra, considere que o kernel de terceira ordem de um determinado sistema seja dado por:

$$h_3(\tau_1, \tau_2, \tau_3) = g(\tau_1)g(\tau_2)g(\tau_3)f(\tau_1 + \tau_2) \quad (2.11)$$

Aplicando a definição dada em (2.10) à equação (2.11), tem-se que:

$$\begin{aligned} h_3^{\text{sim}}(\tau_1, \tau_2, \tau_3) &= \frac{1}{6} \left[2g(\tau_2)g(\tau_1)g(\tau_3)f(\tau_2 + \tau_1) + 2g(\tau_3)g(\tau_2)g(\tau_1)f(\tau_2 + \tau_3) + \right. \\ &\quad \left. + 2g(\tau_1)g(\tau_3)g(\tau_2)f(\tau_1 + \tau_3) \right] \\ &= \frac{1}{3} g(\tau_2)g(\tau_1)g(\tau_3) \left[f(\tau_2 + \tau_1) + f(\tau_2 + \tau_3) + f(\tau_1 + \tau_3) \right] \quad (2.12) \end{aligned}$$

Embora a versão simétrica do kernel usualmente seja caracterizada por expressões matemáticas mais complexas do que uma versão assimétrica, a simetrização pode ser um procedimento importante. Isso se justifica pelo fato de que, em muitos casos, as propriedades dos sistemas podem estar relacionadas de forma mais simples com as propriedades do kernel simétrico do que com as propriedades de sua versão assimétrica. Além disso, a determinação de apenas uma das partes simétricas é suficiente para a caracterização completa do kernel (Schetzen 1980, Rugh 1981).

2.2 Modelos baseados em funções ortonormais

Tendo em vista as dificuldades já discutidas neste capítulo para o problema de aproximação de sinais e sistemas, uma abordagem usando desenvolvimento ortonormal se apresenta como uma importante alternativa. Esta abordagem consiste em representar um determinado sistema usando uma base ortonormal para o espaço de interesse¹. Esse tema é um importante tópico de pesquisa nas áreas de automação, controle e processamento de sinais, e vem se constituindo um objeto de estudo amplo nas últimas décadas.

2.2.1 Motivação inicial

O número de parâmetros necessários para representar os modelos de Volterra pode ser reduzido se uma base de funções ortonormais for utilizada para descrever os

¹Uma diferente abordagem com bases ortonormais envolve o uso de *wavelets*, mas este assunto não será estudado aqui.

kernels do modelo. Esta nova representação contorna a desvantagem do número elevado de termos necessário no modelo original de Volterra, incorporando conhecimento prévio sobre a dinâmica do sistema dentro do processo de identificação (Van den Hof *et al.* 1995, Ninness e Gustafsson 1997, Nelles 2001). Esse conhecimento pode ser obtido, por exemplo, através de intuições físicas sobre o sistema, ou então a partir de medições experimentais.

A ideia do uso de funções ortonormais em modelagem para controle e processamento de sinais é clássica, e foi apresentada inicialmente por Wiener (1958). Esta representação permite que o problema de aproximação correspondente tenha sua solução simplificada. As propriedades de ortonormalidade dessas funções resultam em um modelo geral mais simples (Schetzen 1980, Doyle III *et al.* 2002, Heuberger *et al.* 2005). A representação de um sistema estável em termos de uma base de funções ortonormais se torna extremamente simples se a dinâmica das funções do modelo se aproxima da dinâmica do processo. Uma estrutura com bases ortonormais permite incorporar no processo de identificação o conhecimento sobre a dinâmica do sistema (Heuberger *et al.* 1995). Além disso, o problema de estimação dos coeficientes da expansão se beneficia de uma rápida taxa de convergência, implicando em um menor número de coeficientes a serem determinados para modelar o sistema com uma boa precisão (Van den Hof *et al.* 1995, Heuberger *et al.* 2005). Outra vantagem é que as funções ortonormais são geradas por filtros passa-todas, sendo assim robustas para implementações numéricas (Wahlberg 1994).

Modelos baseados em conjuntos de funções ortonormais têm sido usados em estudos envolvendo identificação e controle de processos dinâmicos nas últimas décadas (Dumont e Fu 1993, Heuberger *et al.* 1995, Bokor e Schipp 1998, Oliveira *et al.* 2000, Doyle III *et al.* 2002, Heuberger *et al.* 2005, Campello *et al.* 2007a, Campello *et al.* 2007b). Muitos problemas de controle podem ser formulados como a otimização de uma certa função de custo sobre uma classe de sistemas estáveis, e as funções ortonormais fornecem boas parametrizações desta classe de sistemas (Wahlberg e Mäkilä 1996). Modelos que utilizam bases de funções ortonormais possuem propriedades atrativas também em termos de variância da sua estimação. Sabe-se que, em modelos com um grande número de parâmetros, a possibilidade de se obter estimativas erradas é maior, aumentando assim a variância da estimação dos parâmetros. Porém, uma seleção adequada dos parâmetros/pólos da base ortonormal exige um número menor de termos para representar os modelos que conseguem descrever com precisão a dinâmica do processo.

Como os pólos das funções ortonormais são parâmetros de projeto livre dentro de uma região de estabilidade, um problema diretamente relacionado é como fazer uma boa seleção destes parâmetros (Tanguy *et al.* 1995, Oliveira e Silva 1995b, Heuberger *et al.* 2005). Em termos de projeto de engenharia, parece razoável escolher os pólos das funções próximos aos pólos do sistema. Este problema será tema de estudo dos Capítulos 3 e 4.

2.2.2 Representação matemática

Esta seção apresenta a descrição matemática dos modelos de Volterra baseados em funções ortonormais (OBF), levando à representação conhecida como modelos OBF-Volterra. Esta representação é possível para sistemas dinâmicos em que os kernels do sistema são absolutamente somáveis no espaço $[0, \infty)$, i.e.:

$$\sum_{k_1=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} |h_\eta(k_1, \dots, k_\eta)| < \infty \quad (2.13)$$

A condição acima implica que o sistema é estável. Assim, a representação dos kernels de Volterra em uma base ortonormal permite que eles sejam truncados, conforme as discussões sobre as propriedades dos kernels apresentadas na seção 2.1.

Nos modelos OBF-Volterra, os kernels h_η são descritos matematicamente como uma expansão numa base de funções ortonormais $\{\psi_{l,n}\}$, da seguinte forma (Schetzen 1980, Doyle III *et al.* 2002):

$$h_\eta(k_1, \dots, k_\eta) = \sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} \alpha_{i_1, \dots, i_\eta} \prod_{l=1}^{\eta} \psi_{l, i_l}(k_l) \quad (2.14)$$

A expansão dada em (2.14) também pode ser expressa no domínio da frequência complexa z . Neste caso, esta expansão é dada em termos da transformada Z das funções ortonormais, como a seguir:

$$H_\eta(z_1, \dots, z_\eta) = \sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} \alpha_{i_1, \dots, i_\eta} \prod_{l=1}^{\eta} \Psi_{l, i_l}(z_l) \quad (2.15)$$

onde $\psi_{l, i_l}(k)$ é a representação das funções ortonormais no domínio do tempo, ou seja, $\Psi_{l, i_l}(z) = \mathcal{Z}[\psi_{l, i_l}(k)]$.

Devido à propriedade de ortonormalidade do conjunto de funções $\{\psi_{l, i_l}\}$, os coeficientes $\alpha_{(\cdot)}$ podem ser calculados a partir de (2.14) como:

$$\alpha_{i_1, \dots, i_\eta} = \sum_{k_1=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} h_\eta(k_1, \dots, k_\eta) \prod_{l=1}^{\eta} \psi_{l, i_l}(k_l) \quad (2.16)$$

Para fazer uma comparação com os modelos lineares, considere um dado kernel de primeira ordem. Neste caso, tem-se que $\eta = 1$ em (2.14), e portanto pode-se escrever:

$$h_1(k_1) = \sum_{i_1=1}^{\infty} \alpha_{i_1} \psi_{1, i_1}(k_1) \quad (2.17)$$

A equação (2.17) descreve um kernel de primeira ordem, que é a resposta ao impulso $h_1(k)$ do sistema, como uma combinação linear dos elementos da base de funções ortonormais $\{\psi_{1,i_1}\}$. Os coeficientes desta expansão são calculados como:

$$\alpha_{i_1} = \sum_{k_1=0}^{\infty} h_1(k_1) \psi_{1,i_1}(k_1) \quad (2.18)$$

Se o kernel em questão for de segunda ordem, sua expansão em uma base de funções ortonormais e os coeficientes desta expansão são, respectivamente, dados por:

$$h_2(k_1, k_2) = \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{\infty} \alpha_{i_1, i_2} \psi_{1,i_1}(k_1) \psi_{2,i_2}(k_2) \quad (2.19)$$

$$\alpha_{i_1, i_2} = \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} h_2(k_1, k_2) \psi_{1,i_1}(k_1) \psi_{2,i_2}(k_2) \quad (2.20)$$

Para sistemas com memória finita, a equação geral (2.14) é utilizada limitando-se o número de termos nas somas infinitas, fazendo com que os termos de ordem superior do kernel $h_\eta(k_1, \dots, k_\eta)$ sejam considerados nulos. Em outras palavras, as somas infinitas em (2.16) podem ser truncadas se a estabilidade do kernel for garantida. Isto representa um truncamento prático para os kernels de Volterra, i.e., uma constante $\epsilon < \infty$ tal que $h_\eta(k_1, \dots, k_\eta)$ seja assumido nulo para $k_l > \epsilon$, $\forall l \in \{1, \dots, \eta\}$. Este valor pode ser selecionado de acordo com a dinâmica do sistema a ser modelado. Neste caso, observa-se o tempo de acomodação do sistema para se fazer uma seleção adequada do truncamento. Por razões computacionais, na prática a quantidade de funções utilizadas é limitada a um número finito M que seja o suficiente para garantir uma boa aproximação. Desta forma, pode-se escrever:

$$\tilde{h}_\eta(k_1, \dots, k_\eta) = \sum_{i_1=1}^M \cdots \sum_{i_\eta=1}^M \alpha_{i_1, \dots, i_\eta} \prod_{l=1}^{\eta} \psi_{l, i_l}(k_l) \quad (2.21)$$

Sob certas condições impostas ao sinal de entrada $u(k)$ no modelo de Volterra em (2.3), kernels de ordem superior podem ser ignorados de forma que o modelo de Volterra resultante seja truncado em uma ordem finita N (Eykhoff 1974). Basicamente, estas condições exigem que a entrada seja limitada tal que $|u(k)| < 1$ (Dumont e Fu 1993, Schetzen 1980), independentemente do tipo de não-linearidade. Além disso, se for assumido que $u(k) = 0$ para $k < 0$, então a equação (2.3) pode ser reescrita da seguinte forma:

$$\tilde{y}(k) = \sum_{\eta=1}^N \left[\sum_{i_1=1}^M \cdots \sum_{i_\eta=1}^M \alpha_{i_1, \dots, i_\eta} \prod_{l=1}^{\eta} \left(\sum_{\tau_l=0}^k \psi_{l, i_l}(\tau_l) u(k - \tau_l) \right) \right] \quad (2.22)$$

O número de funções a serem utilizadas no modelo (2.22), ou seja M , representa apenas um compromisso entre a capacidade de aproximação e a quantidade de parâmetros do modelo. Assim, o número total de parâmetros do modelo depende da quantidade de funções ortonormais utilizadas nos desenvolvimentos dos kernels. Modelos truncados de Volterra, como aquele em (2.22) podem aproximar qualquer sistema não-linear invariante no tempo, com memória finita e entrada limitada (Boyd e Chua 1985). A vantagem prática da representação de modelos de Volterra em uma base de funções ortonormais é a redução da complexidade do modelo original.

Note que as equações (2.14)–(2.22) implicitamente assumem que todos os kernels do modelo de Volterra são representados na mesma base ortonormal $\{\psi_{l,i_l}\}$. Entretanto, cada kernel h_η pode ser expandido usando uma base independente de funções $\{\psi_{\eta,l,i_l}\}$ (Schetzen 1980, Campello *et al.* 2004), conforme será mostrado nos exemplos ilustrativos do Capítulo 3. Por questão de simplicidade e sem nenhuma perda de generalidade, no entanto, a maior parte dos desenvolvimentos teóricos será apresentada seguindo a mesma representação adotada nesta seção, i.e., assumindo que todos os kernels de Volterra são expandidos usando a mesma base ortonormal.

É importante observar também que o modelo em (2.22) é linear nos parâmetros $\alpha_{i_1,\dots,i_\eta}$. Assim, o problema de estimação dos kernels de Volterra resume-se à estimação linear desses parâmetros, o que pode ser feito através de algoritmos do tipo mínimos quadrados, por exemplo.

A dinâmica linear dada pelo respectivo conjunto de filtros ortonormais representa uma incorporação de conhecimento sobre a dinâmica do sistema dentro do processo de identificação (Van den Hof *et al.* 1995, Ninness e Gustafsson 1997, Nelles 2001, Heuberger *et al.* 2005). Esta característica permite obter uma aproximação satisfatória utilizando um número reduzido de funções. Porém, o número de funções necessárias para proporcionar uma dada representação depende da complexidade do problema. Sistemas dinâmicos com múltiplos modos dominantes, por exemplo, requerem uma quantidade maior de funções. A quantidade de funções necessárias para se obter uma precisão pré-determinada também depende da base utilizada.

A maior limitação dos modelos OBF-Volterra é decorrente de seu truncamento explícito em uma quantidade finita de termos (ordem). Este truncamento é adotado na prática até o termo de segunda ordem ($N = 2$ em (2.22)), como usualmente feito na literatura (Billings 1980, Dumont e Fu 1993). Apesar de limitar a quantidade de parâmetros a serem estimados, um truncamento pode também limitar a capacidade de representação dos modelos.

Os principais conjuntos de funções ortonormais usados na representação de sinais e sistemas são apresentados a seguir.

2.3 Funções de base ortonormal

O uso de conjuntos de funções ortonormais com o objetivo de representar sinais e sistemas tem sido estudado há várias décadas (Takenaka 1925, Wiener 1958).

Trabalhos pioneiros estudaram a construção de um conjunto ortonormal de funções contínuas (Kautz 1954) e discretas (Broome 1965).

As funções de Laguerre e de Kautz (Broome 1965, Wahlberg e Mäkilä 1996, Bokor e Schipp 1998) são as bases ortonormais mais utilizadas em identificação e controle de sistemas. Estas funções são apropriadas para modelar sistemas que possuam kernels com dinâmica dominante de primeira ou segunda ordem, respectivamente. Para modelar dinâmicas dominantes mais complexas, é apropriado utilizar uma base de funções parametrizada por um conjunto variado de pólos, como a Base de Funções Ortonormais Generalizadas (*Generalized Orthonormal Basis Functions* – GOBF)². A vantagem do uso de GOBF, quando comparada com as bases de funções de Laguerre e de Kautz, é a capacidade de incorporar uma variedade de pólos reais e complexos no modelo do sistema. No entanto, as funções GOBF envolvem uma parametrização mais complexa da base ortonormal.

O conjunto das funções ortonormais discretas pode ser gerado a partir do cascadeamento de diferentes filtros passa-todas de ordem um ou dois, como a seguir (Ninness e Gustafsson 1997, Heuberger *et al.* 2005):

$$\Psi_{l,n}(z) = \frac{z\sqrt{1-|\beta_{l,n}|^2}}{z-\beta_{l,n}} \prod_{j=1}^{n-1} \left(\frac{1-\bar{\beta}_{l,j}z}{z-\beta_{l,j}} \right) \quad n = 1, 2, \dots \quad (2.23)$$

onde $\beta_{l,n}$ são os pólos estáveis da base ortonormal ($\beta_{l,n} \in \mathbb{C} : |\beta_{l,n}| < 1$), e $\bar{\beta}_{l,n}$ denota o complexo conjugado de $\beta_{l,n}$. As funções dadas pela equação (2.23) são também conhecidas como funções de Takenaka-Malmquist (Heuberger *et al.* 2005).

O conjunto de funções $\{\Psi_{l,n}\}$ satisfaz a propriedade de ortonormalidade

$$\frac{1}{2\pi i} \oint_{\mathbb{T}} \Psi_{l,q}(z) \bar{\Psi}_{l,r}(\bar{z}^{-1}) \frac{dz}{z} = \begin{cases} 0 & \text{se } q \neq r \\ 1 & \text{se } q = r \end{cases}$$

onde a curva \mathbb{T} é o círculo unitário: $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$. Analogamente, para as funções no domínio do tempo $\psi_{l,n}(k)$, que são dadas pela transformada Z inversa de (2.23), i.e. $\psi_{l,n}(k) = \mathcal{Z}^{-1}[\Psi_{l,n}(k)]$, vale o seguinte:

$$\sum_{k=0}^{\infty} \psi_{l,q}(k) \psi_{l,r}(k) = \begin{cases} 0 & \text{se } q \neq r \\ 1 & \text{se } q = r \end{cases}$$

A partir da definição dada em (2.23), pode-se obter um diagrama de blocos equivalente, considerando cada filtro ortonormal como um sistema físico. Este diagrama está ilustrado na Figura 2.3.

Em geral as funções $\psi_{l,n}(k)$ serão complexas. Esta propriedade é fisicamente indesejável em problemas do mundo real, com exceção de alguns problemas na

²Embora o termo funções ortonormais generalizadas tenha sido originalmente adotado em Ninness e Gustafsson (1997), este termo será utilizado genericamente neste trabalho para se referir às bases de funções ortonormais com múltiplos modos.

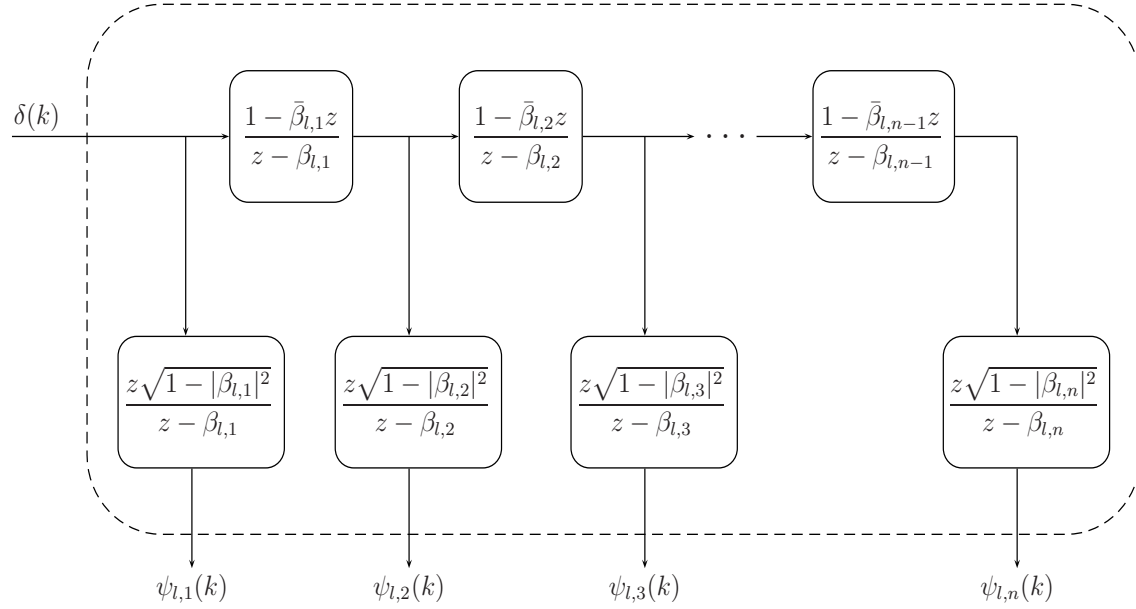


Figura 2.3: Diagrama de blocos de um sistema linear que gera as funções ortonormais $\psi_{l,n}(k)$ em (2.23).

área de telecomunicações. Um método de construção de GOBF que contorna esta desvantagem ao se obter uma nova base de funções ortonormais com respostas ao impulso reais foi apresentado em Ninness e Gustafsson (1997). Para obter esta nova representação, considera-se que as $m - 1$ primeiras funções ortonormais $\{\Psi_{l,1}, \Psi_{l,2}, \dots, \Psi_{l,m-1}\}$ possuem seus respectivos pólos reais $\{\beta_1, \beta_2, \dots, \beta_{m-1}\}$. A inclusão de um pólo complexo $\beta_m \in \mathbb{C}$ exige que duas novas funções com resposta ao impulso reais sejam formadas como uma combinação linear de $\Psi_{l,m}(z)$ e $\Psi_{l,m+1}(z)$ em (2.23) (Ninness e Gustafsson 1997). A nova base formada contém apenas funções com elementos reais e preserva a propriedade de ortonormalidade. Esta nova representação das funções ortonormais generalizadas é mais apropriada de ser utilizada em problemas na área de identificação de sistemas, e será estudada posteriormente na seção 4.3 do Capítulo 4.

As funções dadas em (2.23) formam um conjunto completo no espaço $\ell^2[0, \infty)$ se e somente se (Ninness e Gustafsson 1997, Heuberger *et al.* 2005):

$$\sum_{n=1}^{\infty} (1 - |\beta_{l,n}|) = \infty \quad (2.24)$$

Com esta condição, qualquer sinal de energia finita pode ser aproximado com uma dada precisão fazendo-se o truncamento de uma expansão infinita em bases de funções ortonormais.

O caso particular das funções em (2.23) em que todos os pólos da base são reais e iguais, i.e. $\beta_{l,n} = \bar{\beta}_{l,n} = c_l \quad \forall n$, resulta na base de funções de Laguerre. As

funções de Laguerre são expressas no domínio complexo z como (Wahlberg 1991, Fu e Dumont 1993, Oliveira e Silva 1994):

$$\Psi_{l,n}(z) = \frac{z\sqrt{1-c_l^2}}{z-c_l} \left(\frac{1-c_l z}{z-c_l} \right)^{n-1} \quad (2.25)$$

com c_l denotando os pólos da l -ésima base de Laguerre e satisfazendo $|c_l| < 1$. Quando $c_l = 0$, as funções de transferência (2.25) são dadas por $\Psi_{l,n}(z) = z^{-(n-1)}$, e então o modelo (2.22) é reduzido a um modelo de Volterra NFIR (*Nonlinear Finite Impulse Response*) (Nelles 2001, Heuberger *et al.* 2005), isto é, uma versão truncada do modelo (2.3).

As funções de Laguerre possuem transformadas que são funções racionais com uma forma repetitiva simples, por serem parametrizadas por um único pólo real. Estas funções são adequadas para representar sistemas bem amortecidos com uma constante de tempo dominante. Alocando os pólos c_l próximos aos pólos dominantes do sistema, o número de termos necessários para representá-lo adequadamente pode ser reduzido consideravelmente. Apesar disso, as funções de Laguerre possuem as seguintes desvantagens (Wahlberg 1994):

- sistemas com pólos dominantes situados dispersamente no plano complexo não são muito bem descritos usando-se um número reduzido de funções;
- a presença de pólos que produzam oscilação pode causar problemas computacionais em termos da velocidade de convergência, já que eles sempre ocorrem aos pares de números complexos conjugados.

A aproximação de uma função cuja transformada possua um ou mais pares de pólos complexos dominantes por meio de uma série truncada de Laguerre requer um número elevado de termos (Oliveira e Silva 1995b, Tanguy *et al.* 2000, Nelles 2001). Sistemas sub-amortecidos geralmente não são possíveis de serem representados usando um número pequeno de funções de Laguerre. O número necessário de termos da base ortonormal pode ser elevado, uma vez que os pólos reais que parametrizam a base de Laguerre em (2.25) estão longe dos pólos complexos conjugados sub-amortecidos que descrevem o sistema. Uma forma de contornar esta desvantagem é utilizar uma base de funções ortonormais que seja caracterizada por pólos complexos conjugados. O caso particular das funções geradas em (2.23) quando o conjunto de pólos $\{\beta_{l,n}\}$ é $\{\beta_l, \bar{\beta}_l, \beta_l, \bar{\beta}_l, \dots\}$, com $\beta_l, \bar{\beta}_l \in \mathbb{C}$, resulta na base de funções ortonormais de Kautz. As funções de Kautz são definidas como (Wahlberg 1994, Tanguy *et al.* 2002, Heuberger *et al.* 2005):

$$\begin{aligned} \Psi_{l,2n}(z) &= \frac{z\sqrt{(1-c_l^2)(1-b_l^2)}}{z^2+b_l(c_l-1)z-c_l} \left[\frac{-c_l z^2 + b_l(c_l-1)z + 1}{z^2 + b_l(c_l-1)z - c_l} \right]^{n-1} \\ \Psi_{l,2n-1}(z) &= \frac{z(z-b_l)\sqrt{1-c_l^2}}{z^2+b_l(c_l-1)z-c_l} \left[\frac{-c_l z^2 + b_l(c_l-1)z + 1}{z^2 + b_l(c_l-1)z - c_l} \right]^{n-1} \end{aligned} \quad (2.26)$$

onde os parâmetros reais b_l e c_l estão relacionadas com o par de pólos de Kautz $(\beta_l, \bar{\beta}_l)$ da seguinte forma:

$$b_l = \frac{\beta_l + \bar{\beta}_l}{1 + \beta_l \bar{\beta}_l} \quad (2.27)$$

$$c_l = -\beta_l \bar{\beta}_l \quad (2.28)$$

e satisfazendo $|b_l| < 1$ e $|c_l| < 1$.

As funções de Kautz têm recebido maior atenção em recentes trabalhos envolvendo modelagem e identificação, porque permitem aproximar eficientemente sinais e sistemas com comportamento oscilatório (Wahlberg 1994, Oliveira e Silva 1995b, Tanguy *et al.* 2000, Tanguy *et al.* 2002, da Rosa *et al.* 2005, da Rosa *et al.* 2007, da Rosa *et al.* 2008a, da Rosa *et al.* aceito). A utilização destas funções para representar sistemas que possuam esta característica permite ajustar não somente o fator de amortecimento do modelo, mas também a sua frequência de oscilação.

Conforme já foi visto anteriormente, a base de funções de Laguerre, definida na equação (2.25), é completamente caracterizada pela presença de um pólo real em suas funções de transferência. Já a base de funções de Kautz, definida em (2.26), é completamente caracterizada pela presença de um par de pólos complexos conjugados. Outras configurações podem ainda ser implementadas, combinando estes dois tipos de pólos. Por exemplo, para uma determinada base ortonormal que contempla dinâmicas de primeira e de segunda ordem, caracterizadas pelos pólos $p_l \in \mathbb{R}$ e $\beta_l, \bar{\beta}_l \in \mathbb{C}$ respectivamente, a função interna da base será dada por:

$$\begin{aligned} G_l(z) &= \left(\frac{1 - \bar{\beta}_l z}{z - \beta_l} \right) \cdot \left(\frac{1 - \beta_l z}{z - \bar{\beta}_l} \right) \cdot \left(\frac{1 - p_l z}{z - p_l} \right) \\ &= \frac{1 - (\beta_l + \bar{\beta}_l + p_l)z + [p_l(\beta_l + \bar{\beta}_l) + \beta_l \bar{\beta}_l]z^2 - \beta_l \bar{\beta}_l p_l z^3}{z^3 - (\beta_l + \bar{\beta}_l + p_l)z^2 + [p_l(\beta_l + \bar{\beta}_l) + \beta_l \bar{\beta}_l]z - \beta_l \bar{\beta}_l p_l} \end{aligned} \quad (2.29)$$

Filtros com funções de transferência como a equação (2.29) permitem aproximar com melhor precisão os sistemas que possuem dinâmica dominante superior a dois. Outras combinações envolvendo diferentes números de pólos reais e complexos são possíveis de serem implementadas, devendo-se escolher adequadamente a ordem da dinâmica do filtro.

Uma propriedade importante das funções ortonormais estudadas nesta seção é que elas incorporam o comportamento dinâmico do sistema em estudo (Van den Hof *et al.* 1995, Ninness e Gustafsson 1997, Nelles 2001, Heuberger *et al.* 2005). Neste caso, pólos complexos conjugados podem ser combinados com pólos reais em uma mesma função de transferência. A utilização de bases de funções ortonormais pode ter vantagens importantes em problemas de identificação e de aproximação de modelos. Se a dinâmica da base que gera o sistema estiver próxima da dinâmica do sistema a ser modelado, então a convergência da série aumenta (Heuberger *et al.* 1995, Van

den Hof *et al.* 1995). Em outras palavras, reduz-se o número de parâmetros a serem determinados para modelar o sistema, conseqüentemente aumentando-se sua precisão.

Os modelos de Volterra baseados em funções ortonormais apresentam características importantes no contexto de identificação de sistemas não-lineares. O problema central abordado no desenvolvimento destes modelos é a seleção dos pólos das funções ortonormais e do número de funções necessárias para se obter uma boa aproximação da dinâmica do sistema. Em princípio, uma base ortonormal pode ser construída selecionando-se os pólos de maneira arbitrária. Entretanto, usar um conjunto arbitrário de funções da base pode exigir um número elevado de termos na expansão ortonormal (2.14), e portanto o modelo resultante pode conter muitos parâmetros. Assim, se a escolha da base de funções for adequada para o sistema em estudo, uma boa aproximação pode ser gerada com um número reduzido de coeficientes. Este tema será objeto de estudo nos próximos capítulos deste trabalho.

2.4 Resumo

Este capítulo apresentou alguns conceitos matemáticos referentes ao problema de aproximação de sistemas não-lineares usando modelos de Volterra. As principais vantagens destes modelos residem no fato de que eles representam uma generalização direta do modelo linear de resposta ao impulso, o que permite estender aos sistemas não-lineares alguns conceitos definidos somente para os lineares. Entretanto, uma grande dificuldade na estimação dos kernels de um sistema reside no fato de que os kernels são funções multidimensionais não-parametrizadas. Assim, um elevado número de parâmetros é necessário para representar os kernels de Volterra, especialmente porque os modelos de Volterra não possuem realimentação do sinal de saída. O preço a pagar pela falta da realimentação é que a ordem do modelo pode ser alta para descrever adequadamente a dinâmica do processo.

O número de parâmetros dos modelos de Volterra pode ser reduzido se uma base de funções ortonormais for utilizada para descrever os kernels do modelo. Esta nova representação, conhecida como modelo OBF-Volterra, consiste em descrever um modelo de Volterra por meio de um conjunto de funções ortonormais. Nas últimas décadas, estes modelos têm apresentado potenciais aplicações em áreas como automação, teorias de controle, otimização de sistemas e processamento de sinais. A utilização de um modelo baseado em funções ortonormais contorna a desvantagem da exigência de um número elevado de termos necessário no modelo de Volterra original, pois incorpora um conhecimento prévio sobre a dinâmica do processo dentro das bases ortonormais. Uma importante característica é que o modelo é linear nos parâmetros, permitindo o uso de algoritmos de mínimos quadrados para a estimação destes parâmetros.

Apresentou-se ainda as principais funções ortonormais utilizadas na modelagem e controle de sistemas. Foram discutidos alguns dos aspectos mais importantes sobre

a escolha da base de funções a ser utilizada. Dentre eles, está o fato de ser vantajoso usar o menor número possível de funções para se garantir um modelo de menor ordem. Nesse aspecto, destacam-se as funções de Laguerre e de Kautz, que representam um caso particular das funções ortonormais generalizadas (GOBF). Conforme as discussões apresentadas, estas funções são representadas por parâmetros livres (pólos) que permitem caracterizá-las como sistemas físicos reais. A seleção dos pólos que parametrizam as funções ortonormais abordadas será objeto de estudo dos capítulos seguintes.

Capítulo 3

Desenvolvimento Ótimo de Modelos de Volterra Usando Bases Múltiplas de Kautz

Este capítulo aborda o problema de modelagem e identificação de sistemas dinâmicos não-lineares utilizando modelos baseados em funções ortonormais de Kautz. Estuda-se a representação de modelos de OBF-Volterra por meio da seleção ótima dos pólos da base de Kautz. Dentro deste contexto, propõe-se uma metodologia de projeto da base através da otimização dos pólos das funções de Kautz.

O objetivo deste capítulo é desenvolver uma metodologia que permita reduzir o número de funções de Kautz associadas com um dado erro de truncamento da série de Volterra em (2.3). O truncamento dos kernels de Volterra, como na equação (2.21), possui dimensão finita e permite reduzir a complexidade da sua representação. Este problema foi inicialmente abordado no contexto de sistemas lineares ao se minimizar um limitante superior para o erro resultante da expansão truncada de modelos de resposta ao impulso utilizando funções de Kautz (Tanguy *et al.* 2002). Mais recentemente, obteve-se uma generalização destes resultados para o contexto dos modelos não-lineares de Volterra (da Rosa 2005, da Rosa *et al.* 2005, da Rosa *et al.* 2007). A abordagem utilizada permitiu obter uma solução analítica para a seleção de um dos parâmetros das funções de Kautz. Esta solução foi obtida a partir da minimização de um limitante superior do erro de aproximação dos kernels de Volterra, e pode ser interpretada como uma escolha sub-ótima dos pólos de Kautz.

O problema central abordado neste capítulo é como fazer a seleção dos pólos livres que parametrizam as funções de Kautz. Nesta abordagem, cada kernel multi-dimensional $h_\eta(k_1, \dots, k_\eta)$ é decomposto em um conjunto finito de funções de Kautz independentes para cada direção $l = 1, \dots, \eta$ do kernel. Cada conjunto de funções de Kautz, por sua vez, é parametrizado por um pólo individual, representando a dinâmica dominante do kernel em uma direção particular. A principal contribuição é a obtenção de uma solução analítica global para uma expansão ótima dos kernels de Volterra usando funções de Kautz. Com o uso de uma base independente para

cada direção, espera-se assim reduzir o erro de truncamento quando as dinâmicas dominantes do kernel ao longo das múltiplas direções são diferentes entre si. Este problema também é apresentado em um trabalho encontrado na literatura (da Rosa *et al.* 2008a), onde os resultados são semelhantes aos discutidos aqui. A metodologia proposta representa uma generalização de um método anterior, que adota um único pólo para todas as direções dos kernels (da Rosa 2005, da Rosa *et al.* 2005, da Rosa *et al.* 2007). Os resultados a serem apresentados aqui podem ser vistos ainda como uma extensão de outro trabalho para o contexto da base de Kautz, onde um problema análogo foi investigado no escopo da base de Laguerre (Campello *et al.* 2006).

Para ilustrar os resultados teóricos obtidos, apresenta-se um exemplo de simulação envolvendo a aproximação de um sistema não-linear com comportamento oscilatório por meio de uma parametrização adequada das funções de Kautz. Uma aplicação dos resultados obtidos também é apresentada para a modelagem de um sistema real de levitação magnética.

3.1 Formulação e solução do problema

Conforme discutido no Capítulo 1, o problema da escolha ótima do pólo das funções discretas de Laguerre utilizando uma abordagem analítica já foi extensivamente abordado na literatura. As funções de Laguerre possuem uma propriedade que torna este tipo de solução mais simples do que no caso de Kautz. Esta propriedade se deve ao fato de que as funções discretas de Laguerre satisfazem uma equação a diferenças apropriada (ou equação diferencial, no caso contínuo), permitindo assim uma redução na complexidade matemática associada com o desenvolvimento da solução para o pólo ótimo de Laguerre. Fez-se uso desta vantagem em trabalhos anteriores que lidaram com a seleção analítica de pólos ótimos de Laguerre (Fu e Dumont 1993, Oliveira e Silva 1995a, Tanguy *et al.* 1995, Campello *et al.* 2001, Campello 2002, Campello *et al.* 2003, Campello *et al.* 2004, Campello *et al.* 2006, Campello *et al.* 2007a).

No entanto, uma equação a diferenças para a base de Kautz ainda não foi estabelecida na literatura, não permitindo assim utilizar a estratégia usada no caso Laguerre. O desenvolvimento apresentado nesta seção é baseado no fato de que é possível relacionar no domínio z as funções de Laguerre e de Kautz. Esta propriedade permite que grande parte do desenvolvimento necessário para se obter a solução desejada para a base de Kautz pode ser realizada utilizando-se a base de funções de Laguerre. Desta forma, a estratégia apresentada aqui consiste da adaptação do problema original (Kautz) em um problema transformado (Laguerre), que possui solução conhecida. O método proposto nesta seção é baseado, em sua maior parte, nas funções de Laguerre, que são parametrizadas por apenas um parâmetro livre.

As funções de Kautz, definidas na equação (2.26), dependem de dois parâmetros reais: b_l e c_l . A seleção destes parâmetros tem influência direta no cálculo dos coeficientes $\alpha_{i_1, \dots, i_\eta}$ em (2.16). Apesar da otimização simultânea de b_l e c_l em ter-

mos analíticos ainda ser um problema em aberto, é possível considerar um destes parâmetros como fixo para obter a escolha ótima para o outro. Este problema já foi abordado para o caso em que uma única base de funções de Kautz é adotada para representar todas as direções de cada kernel de Volterra (da Rosa 2005, da Rosa *et al.* 2005, da Rosa *et al.* 2007). Isso implica que o mesmo pólo de Kautz é calculado para todas as múltiplas direções do kernel. Neste capítulo, estes resultados são estendidos de forma que bases de Kautz independentes são adotadas para cada uma das direções dos kernels. Desta forma, pólos distintos são utilizados para representar a dinâmica de cada direção $l = 1, \dots, \eta$ do η -ésimo kernel $h_\eta(k_1, \dots, k_\eta)$. A seguir, esta abordagem é desenvolvida em detalhes.

Inicialmente, defina a seguinte norma quadrática:

$$\|h_\eta\|^2 = \sum_{k_1=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} h_\eta^2(k_1, \dots, k_\eta) \quad (3.1)$$

Seja o Erro Quadrático Normalizado (EQN) da aproximação do kernel h_η definido como $\text{EQN} \triangleq (\|h_\eta - \tilde{h}_\eta\|^2) / \|h_\eta\|^2$, com h_η e \tilde{h}_η dados de acordo com as equações (2.14) e (2.21), respectivamente. Usando a propriedade de ortonormalidade do conjunto $\{\psi_{l,n}\}$, pode-se provar¹ que:

$$\sum_{k_1=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} h_\eta^2(k_1, \dots, k_\eta) = \sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} \alpha_{i_1, \dots, i_\eta}^2 \quad (3.2)$$

onde $\alpha_{i_1, \dots, i_\eta}$ são os coeficientes da expansão de $h_\eta(k_1, \dots, k_\eta)$ de acordo com a equação (2.16). Assim, é possível reescrever EQN como:

$$\text{EQN} = \frac{\sum_{i_1=M+1}^{\infty} \cdots \sum_{i_\eta=M+1}^{\infty} \alpha_{i_1, \dots, i_\eta}^2}{\sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} \alpha_{i_1, \dots, i_\eta}^2} \quad (3.3)$$

De acordo com a equação (3.3), note que quanto maior o número de funções ortonormais utilizadas na aproximação do kernel, menor será o erro resultante desta aproximação, ou seja, $\text{EQN} \rightarrow 0$ a medida que $M \rightarrow \infty$. Esta propriedade se deve ao fato de que as bases ortonormais $\{\psi_{l,n}\}$ são completas no espaço das funções quadraticamente somáveis (estáveis) (Heuberger *et al.* 1995, Ninness e Gustafsson 1997, Heuberger *et al.* 2005).

Pode-se obter um limitante superior de EQN em (3.3) quando o conjunto ortonormal é a base de funções de Kautz. Inicialmente sejam $\{\phi_{l,n}\}$ bases de Laguerre no

¹Veja detalhes na seção A.1 do Apêndice A.

domínio do tempo, i.e., a transformada Z inversa de (2.25). Considere $\alpha_{i_1, \dots, i_\eta}$ os coeficientes da expansão do kernel de Volterra $h_\eta(k_1, \dots, k_\eta)$ usando bases de funções de Kautz independentes em cada direção, de acordo com a equação (2.14). Defina agora as seguintes funções auxiliares, para $l = 1, 2, \dots, \eta$:

$$g_{\text{par},l}(k_1, \dots, k_\eta) \triangleq \sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} \alpha_{2i_1, \dots, 2i_\eta} \prod_{l=1}^{\eta} \phi_{l,i_l}(k_l) \quad (3.4)$$

$$g_{\text{impar},l}(k_1, \dots, k_\eta) \triangleq \sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} \alpha_{2i_1-1, \dots, 2i_\eta-1} \prod_{l=1}^{\eta} \phi_{l,i_l}(k_l) \quad (3.5)$$

Com base na definição de $g_{\text{par},l}(k_1, \dots, k_\eta)$ e $g_{\text{impar},l}(k_1, \dots, k_\eta)$ em (3.4) e (3.5), respectivamente, um limitante superior de EQN para um modelo baseado em funções de Kautz pode ser obtido por meio do seguinte teorema.

Teorema 3.1.1. *O erro de aproximação truncada do kernel h_η usando uma expansão truncada com M funções de bases de Kautz (uma base independente para cada direção do kernel) é limitado por:*

$$\text{EQN} \leq \frac{2}{\eta(M+1)\|h_\eta\|^2} \sum_{l=1}^{\eta} \left[\frac{m_{2,l}c_l^2 - 2m_{1,l}c_l + m_{3,l}}{1 - c_l^2} \right] \quad (3.6)$$

onde os termos $m_{p,l}$ ($p = 1, 2, 3$) são calculados como:

$$m_{1,l} = \mu_{1,l}(g_{\text{par},l}) + \mu_{1,l}(g_{\text{impar},l}) \quad (3.7)$$

$$m_{2,l} = \mu_{2,l}(g_{\text{par},l}) + \mu_{2,l}(g_{\text{impar},l}) \quad (3.8)$$

$$m_{3,l} = \mu_{2,l}(g_{\text{par},l}) + \mu_{2,l}(g_{\text{impar},l}) + \eta\mu_{3,l}(g_{\text{par},l}) + \eta\mu_{3,l}(g_{\text{impar},l}) \quad (3.9)$$

e os momentos $\mu_{1,l}(x), \mu_{2,l}(x), \mu_{3,l}(x)$ são dados por:

$$\mu_{1,l}(x) = \sum_{k_1=0}^{\infty} \cdots \sum_{k_l=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} k_l x(k_1, \dots, k_l, \dots, k_\eta) x(k_1, \dots, k_l - 1, \dots, k_\eta) \quad (3.10)$$

$$\mu_{2,l}(x) = \sum_{k_1=0}^{\infty} \cdots \sum_{k_l=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} k_l x^2(k_1, \dots, k_l, \dots, k_\eta) \quad (3.11)$$

$$\mu_{3,l}(x) = (1/\eta) \sum_{k_1=0}^{\infty} \cdots \sum_{k_l=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} x^2(k_1, \dots, k_l, \dots, k_\eta) \quad (3.12)$$

Demonstração. Veja a seção A.2 do Apêndice A. □

Usando o resultado do Teorema 3.1.1, uma seleção ótima dos parâmetros c_l das bases de Kautz pode ser obtida pela solução do seguinte problema de otimização:

$$\min_{c_1, \dots, c_\eta} L(c_1, \dots, c_\eta) \triangleq \frac{2}{\eta(M+1)\|h_\eta\|^2} \sum_{l=1}^{\eta} \left[\frac{m_{2,l}c_l^2 - 2m_{1,l}c_l + m_{3,l}}{1 - c_l^2} \right] \quad (3.13)$$

A solução do problema de otimização dado em (3.13) pode ser obtida levando-se em conta que as condições de otimalidade de primeira ordem (derivadas) são necessárias e suficientes, pois a função $L(c_1, \dots, c_\eta)$ é pseudo-convexa. Esta propriedade pode ser demonstrada através do seguinte lema (Bazaraa *et al.* 1993):

Lema 3.1.1. *Seja uma função $f : \mathcal{X} \rightarrow \mathbb{R}$ definida como*

$$f(x) = \frac{\varpi(x)}{v(x)} \quad (3.14)$$

onde \mathcal{X} é um conjunto convexo aberto, e:

- ϖ é uma função convexa, diferenciável e não-negativa para todo $x \in \mathcal{X}$;
- v é uma função côncava, diferenciável e positiva para todo $x \in \mathcal{X}$.

Tem-se que f é uma função pseudo-convexa em \mathcal{X} .

A partir das equações (3.8) e (3.11), é direto verificar que $m_{2,l} > 0$. Consequentemente, a função $\varpi(c_l) \triangleq m_{2,l}c_l^2 - 2m_{1,l}c_l + m_{3,l}$ é convexa. Ela também é diferenciável. Além disso, $\varpi(c_l)$ é não-negativa para todo $c_l \in (-1, 1)$, caso contrário EQN seria negativo de acordo com a equação (3.6), o que não é possível por definição. A função $v(c_l) \triangleq 1 - c_l^2$, por sua vez, é diferenciável, côncava e positiva para $c_l \in (-1, 1)$. Então, $L(c_1, \dots, c_\eta)$ é uma função pseudo-convexa para $|c_l| < 1$, o que implica que as condições de otimalidade de primeira ordem são necessárias e suficientes para resolver o problema (3.13).

Dado que $\|h_\eta\|$ é uma constante não-nula para um dado sistema, a condição necessária para resolver (3.13) é, para todo $l \in \{1, \dots, \eta\}$, dada por:

$$\begin{aligned} \frac{\partial L(c_1, \dots, c_\eta)}{\partial c_l} &= \\ &= \frac{2[(2m_{2,l}c_l - 2m_{1,l})(1 - c_l^2) - (-2c_l)(m_{2,l}c_l^2 - 2m_{1,l}c_l + m_{3,l})]}{\eta(M+1)\|h_\eta\|^2(1 - c_l^2)^2} = 0 \end{aligned} \quad (3.15)$$

A condição (3.15) é satisfeita se e somente se

$$m_{1,l}c_l^2 - (m_{2,l} + m_{3,l})c_l + m_{1,l} = 0 \quad l = 1, \dots, \eta \quad (3.16)$$

Então, definindo $\xi_l \triangleq (m_{2,l} + m_{3,l})/(2m_{1,l})$, a solução de (3.16) resulta em:

$$c_{\text{otimo},l} = \begin{cases} \xi_l - \sqrt{\xi_l^2 - 1} & \text{se } \xi_l > 1 \\ \xi_l + \sqrt{\xi_l^2 - 1} & \text{se } \xi_l < -1 \end{cases} \quad l = 1, \dots, \eta \quad (3.17)$$

Para garantir que a equação (3.17) seja realmente solução do problema (3.13), é necessário demonstrar que a solução obtida atenda à restrição $|c_l| < 1$. A prova desta condição está apresentada na seção A.3 do Apêndice A.

A solução dada por (3.17) é uma expressão analítica para a seleção ótima dos parâmetros c_l das bases de Kautz para b_l fixos, segundo o critério (3.13). O desenvolvimento apresentado pode ser visto como uma escolha sub-ótima para o par de parâmetros de Kautz (b_l, c_l) que minimizam um limitante superior do erro de aproximação truncada do kernel de Volterra, quando bases independentes de funções são utilizadas para representar cada direção do kernel.

Finalmente, verifica-se que a condição $-1 \leq \xi_l \leq 1$ é sempre infactível, conforme o seguinte teorema:

Teorema 3.1.2. *A variável ξ_l , definida anteriormente como $\xi_l \triangleq (m_{2,l} + m_{3,l})/(2m_{1,l})$, possui valor absoluto sempre maior do que 1.*

Demonstração. Veja a seção A.4 do Apêndice A. □

Teorema 3.1.3. *Se h_η é simétrico (veja seção 2.1.2) e se os parâmetros b_l são iguais para todo $l = 1, \dots, \eta$, então a solução dada por (3.17) se reduz ao caso particular em que a expansão do kernel utiliza uma base única de Kautz, i.e. (da Rosa et al. 2005, da Rosa et al. 2007):*

$$c_{\text{otimo}} = \begin{cases} \xi - \sqrt{\xi^2 - 1} & \text{se } \xi > 1 \\ \xi + \sqrt{\xi^2 - 1} & \text{se } \xi < -1 \end{cases} \quad (3.18)$$

Demonstração. Quando o kernel h_η é simétrico, e com os parâmetros b_l de Kautz satisfazendo $b = b_1 = \dots = b_\eta$, tem-se que os termos $\mu_{1,l}, \mu_{2,l}, \mu_{3,l}, m_{1,l}, m_{2,l}, m_{3,l}$ nas equações (3.7)–(3.12) são iguais para todo $l = 1, \dots, \eta$. Neste caso, resulta que os valores de ξ_l também são iguais para todo $l = 1, \dots, \eta$, i.e. $\xi = \xi_1 = \dots = \xi_\eta$. Assim, a solução única dada pela equação (3.18) claramente se torna equivalente a cada solução individual em (3.17), ou seja $c_{\text{otimo}} = c_{\text{otimo},1} = \dots = c_{\text{otimo},\eta}$. □

Note que, sob a perspectiva da saída do modelo de Volterra em (2.3), é possível substituir qualquer kernel não-simétrico por um equivalente simétrico através de um simples procedimento de simetrização. Entretanto, kernels simétricos são equivalentes a suas partes não-simétricas correspondentes somente em termos da saída do modelo, ou seja, eles não são equivalentes entre si como funções multidimensionais a serem descritas por meio de uma expansão em série. Isto significa que a simetrização de um dado kernel não garante que sua expansão convencional usando uma base única de Kautz seja equivalente à expansão original do kernel com bases distintas para cada direção. A vantagem de se usar uma base independente para cada direção $l = 1, \dots, \eta$, como resolvido na equação (3.17), é que ela oferece flexibilidade e espera-se reduzir o erro de truncamento quando o kernel possui dinâmicas

dominantes diferentes ao longo de suas múltiplas dimensões. Por exemplo, bases independentes podem fornecer uma melhor aproximação de um kernel de segunda ordem com dinâmica bem amortecida em uma das dimensões e sub-amortecida na outra, como ilustrado na Figura 3.1. Neste caso, é mais apropriado utilizar bases com pólos distintos em cada uma das direções do kernel, como sugerem as Figuras 3.2(a) e 3.2(b), que mostram os planos transversais do kernel de Volterra ilustrado na Figura 3.1 para explicitar a dinâmica em cada direção.

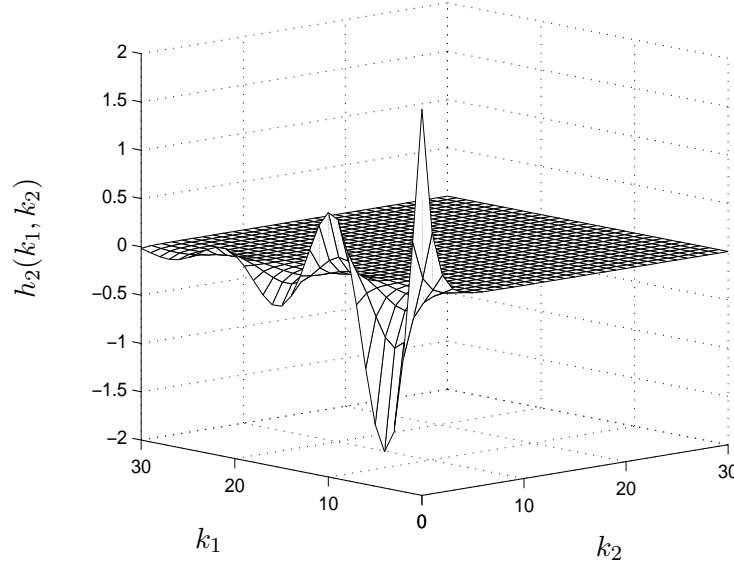


Figura 3.1: Kernel de segunda ordem com dinâmicas dominantes diferentes em cada direção.

Portanto, para descrever um sistema usando um modelo de Volterra simétrico, como usual na prática, não há necessidade de se usar múltiplas bases de Kautz para representar cada kernel. Contudo, ao se adotar bases múltiplas, o uso de kernels assimétricos pode fornecer melhores resultados de aproximação.

Observação 3.1.1. *Conforme discutido na seção 2.2, deve-se garantir ao kernel h_η a condição de ser absolutamente somável. Esta condição é necessária para possibilitar descrevê-lo como uma expansão em bases ortonormais, de acordo com (2.14). Esta mesma imposição também deve ser feita para as funções $g_{\text{par},l}$ e $g_{\text{impar},l}$, uma vez que elas podem ser vistas como uma expansão em bases ortonormais, de acordo com as equações (3.4) e (3.5), respectivamente. Embora a condição de ser absolutamente somável implica que o kernel é estável, isto não ocorre para os momentos $\mu_{1,l}(x)$, $\mu_{2,l}(x)$, $\mu_{3,l}(x)$ em (3.10)–(3.12). A convergência destes termos é necessária para que o critério (3.13) seja bem definido e possa ser otimizado. Esta convergência*

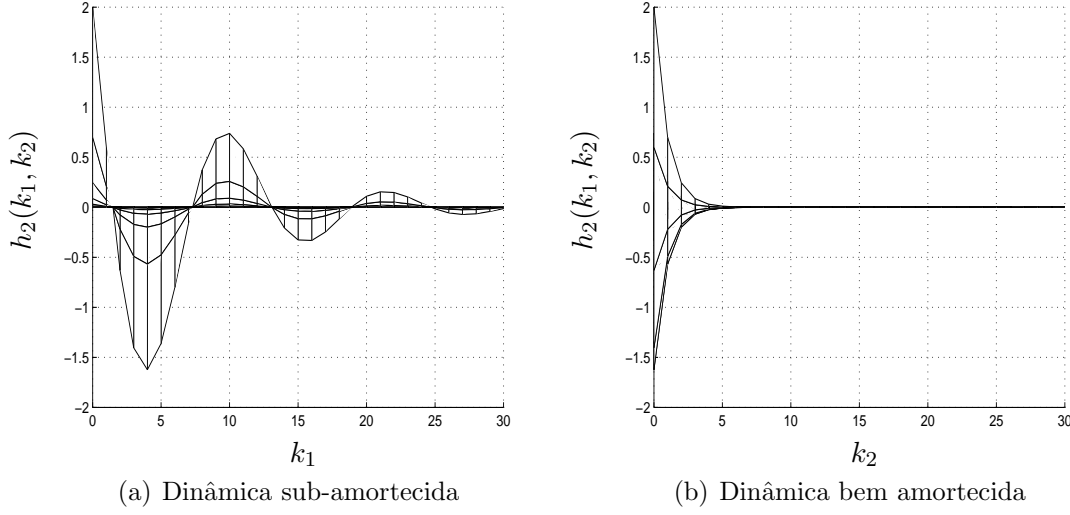


Figura 3.2: Planos transversais do kernel de Volterra ilustrado na Figura 3.1, mostrando as dinâmicas dominantes em cada direção.

depende da taxa de decaimento do kernel, isto é, a taxa com que $h_\eta(k_1, \dots, k_\eta)$ estavel tende a zero conforme k_l ($\forall l \in \{1, \dots, \eta\}$) tende ao infinito. Uma condição suficiente para a convergência é assumir que o kernel, além de ser não-nulo ($\|h_\eta\| \neq 0$), possua memória finita, i.e. $h_\eta(k_1, \dots, k_\eta) = 0$ para um dado $k_l > \epsilon_l$, com $\epsilon_l < \infty$. Essa hipótese leva à usual representação truncada do kernel, adotada nas implementações computacionais dos modelos de Volterra.

Os cálculos necessários para a obtenção das funções $g_{\text{par},l}(k_1, \dots, k_\eta)$ e $g_{\text{impar},l}(k_1, \dots, k_\eta)$ por meio das equações (3.4) e (3.5) podem ser complexos. A principal dificuldade está na necessidade de se calcular explicitamente os coeficientes $\alpha_{(\cdot)}$ dados por (2.16). Para contornar este problema, propõe-se um método alternativo para calcular $g_{\text{par},l}$ e $g_{\text{impar},l}$ independentemente de $\alpha_{(\cdot)}$, de acordo com o teorema a seguir.

Teorema 3.1.4. *As funções $g_{\text{par},l}(k_1, \dots, k_\eta)$ e $g_{\text{impar},l}(k_1, \dots, k_\eta)$, definidas nas equações (3.4) e (3.5), respectivamente, podem ser expressas em termos do η -ésimo kernel h_η e das funções de Kautz como*

$$g_{\text{par},l}(k_1, \dots, k_\eta) = \sum_{\tau_1=0}^{\infty} \cdots \sum_{\tau_\eta=0}^{\infty} h_\eta(\tau_1, \dots, \tau_\eta) \prod_{l=1}^{\eta} \hat{\psi}_{l,2(k_l+1)}(\tau_l) \quad (3.19)$$

$$g_{\text{impar},l}(k_1, \dots, k_\eta) = \sum_{\tau_1=0}^{\infty} \cdots \sum_{\tau_\eta=0}^{\infty} h_\eta(\tau_1, \dots, \tau_\eta) \prod_{l=1}^{\eta} \hat{\psi}_{l,2(k_l+1)-1}(\tau_l) \quad (3.20)$$

em que $\hat{\psi}_{l,n}(\tau)$ denota a n -ésima função de Kautz com $c_l = 0$.

Demonstração. Veja a seção A.5 do Apêndice A. □

De acordo com o Teorema 3.1.4, as funções $g_{\text{par},l}$ e $g_{\text{impar},l}$ dependem unicamente do parâmetro b_l da base de Kautz. Uma consequência direta é que esta propriedade também vale para os momentos $\mu_{1,l}, \mu_{2,l}, \mu_{3,l}$ e também para $m_{1,l}, m_{2,l}, m_{3,l}$, como é facilmente concluído a partir das equações (3.10)–(3.12), e (3.7)–(3.9), respectivamente. Desta forma, a expressão analítica para os parâmetros $c_{\text{otimo},l}$ de Kautz, dada pela equação (3.17), também depende somente de b_l .

Para ter um entendimento mais claro sobre o método proposto, este pode ser resumido pelos seguintes passos. Para cada direção $l = 1, \dots, \eta$ do kernel $h_\eta(k_1, \dots, k_\eta)$, faça:

1. Escolher um valor arbitrário para o parâmetro $b_l \in (-1, 1)$;
2. Uma vez que o kernel h_η é conhecido, calcular as funções $g_{\text{par},l}$ e $g_{\text{impar},l}$ usando as equações (3.19) e (3.20), respectivamente;
3. Calcular os momentos μ_1, μ_2, μ_3 usando (3.10)–(3.12) e os termos m_1, m_2, m_3 a partir das equações (3.7)–(3.9);
4. Calcular $c_{\text{otimo},l}$ usando a equação (3.17).

Assim, os parâmetros de Kautz $(b_l; c_{\text{otimo},l})$ representam aqueles que minimizam o limitante superior $L(c_1, \dots, c_\eta)$ do erro resultante da expansão truncada do kernel de Volterra h_η em bases de funções de Kautz. Este resultado pode ser visto como uma seleção sub-ótima dos pólos de Kautz. Os pólos ótimos podem ser obtidos fazendo-se uma busca exaustiva para o parâmetro b_l dentro do intervalo de factibilidade discretizado $(-1, 1)$, encontrando-se o valor de b_l para o qual o par de parâmetros $(b_l; c_{\text{otimo},l})$ forneça o menor erro de aproximação (EQN).

Os resultados matemáticos obtidos anteriormente são importantes na aproximação ortonormal de sinais e sistemas. A seção seguinte ilustra estes resultados por meio de simulações computacionais.

3.2 Resultados de simulação

Conforme as discussões apresentadas no Capítulo 2, sistemas pouco amortecidos são melhor representados por modelos baseados em funções ortonormais de Kautz. Nesta seção, algumas simulações são apresentadas para ilustrar os resultados matemáticos deduzidos na seção 3.1. Os resultados teóricos serão aplicados para se obter uma boa representação do kernel de segunda ordem de um determinado sistema.

Para efeitos de comparação, considere o mesmo kernel de Volterra de segunda ordem simulado em um trabalho anterior (da Rosa *et al.* 2007):

$$h_2(k_1, k_2) = (k_1 - 2k_2) \exp(-\rho_1 k_1 - \rho_2 k_2) \cos(\omega_1 k_1 + \omega_2 k_2) \quad (3.21)$$

para $k_1, k_2 \geq 0$. Considera-se também que $h_2(k_1, k_2) = 0$ para valores negativos de k_1 ou k_2 (sistema causal).

Os elementos do kernel com atraso de memória maior que 30 são considerados nulos. Em outras palavras, as somas infinitas em (2.16) são truncadas em $k_1, k_2 = 30$. Para o kernel da equação (3.21), ρ_l e ω_l ($l = 1, 2$) são constantes reais que determinam a dinâmica do sistema. Note que deve-se ter $\rho_l > 0$ para que h_2 seja estável. As constantes ρ_l podem ser vistas como o fator de decaimento do kernel, ou seja, elas dão a ideia da velocidade com que $h_2(k_1, k_2)$ tende a zero à medida que $k_l \rightarrow \infty$. As constantes ω_l podem ser interpretadas como as frequências com que o kernel oscila em cada direção axial l .

A Figura 3.3 ilustra o kernel $h_2(k_1, k_2)$ para $\rho_1 = 0,45$, $\rho_2 = 0,7$, $\omega_1 = 100$ e $\omega_2 = 1$.

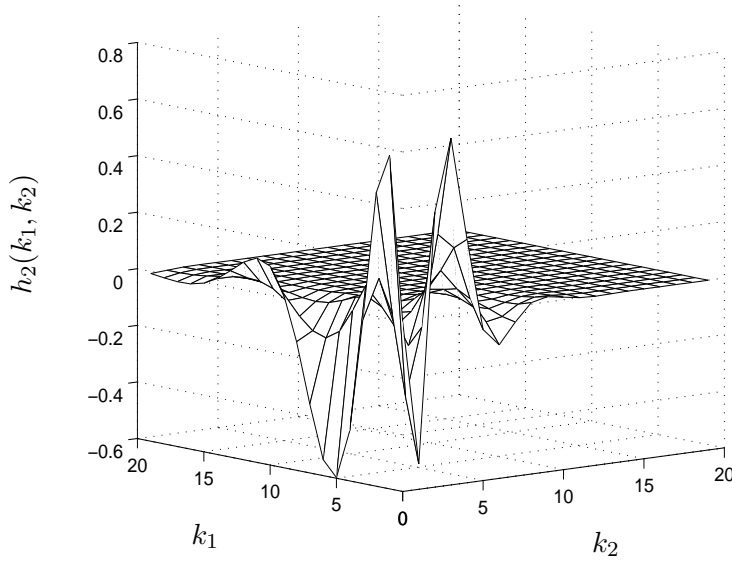


Figura 3.3: Kernel em (3.21) para $\rho_1 = 0,45$, $\rho_2 = 0,7$, $\omega_1 = 100$ e $\omega_2 = 1$.

A estratégia proposta neste capítulo será utilizada com o objetivo de se determinar os pólos de Kautz para cada uma das duas direções do kernel h_2 em (3.21). Seguindo a sequência de passos deste procedimento, conforme apresentado no final da seção 3.1, simulações computacionais forneceram os resultados a seguir:

1. Escolheu-se, por exemplo, os valores de $b_1 = 0,6$ para a primeira direção axial e $b_2 = 0,5$ para a segunda direção.
2. Utilizando a expressão para o kernel dado na equação (3.21), com $h_2(k_1, k_2) = 0$ para valores de k_1, k_2 superiores a 30, calcula-se $g_{\text{par},l}$ e $g_{\text{impar},l}$

(para $l = 1, 2$) usando as equações (3.19) e (3.20), respectivamente. Tem-se assim:

$$g_{\text{par},l}(k_1, k_2) = \sum_{\tau_1=0}^{30} \sum_{\tau_2=0}^{30} h_2(\tau_1, \tau_2) \hat{\psi}_{1,2(k_1+1)}(\tau_1) \hat{\psi}_{2,2(k_2+1)}(\tau_2)$$

$$g_{\text{impar},l}(k_1, k_2) = \sum_{\tau_1=0}^{30} \sum_{\tau_2=0}^{30} h_2(\tau_1, \tau_2) \hat{\psi}_{1,2(k_1+1)-1}(\tau_1) \hat{\psi}_{2,2(k_2+1)-1}(\tau_2)$$

3. Calcula-se os momentos $\mu_{1,l}, \mu_{2,l}, \mu_{3,l}$ usando (3.10)–(3.12) e os termos $m_{1,l}, m_{2,l}, m_{3,l}$ a partir das equações (3.7)–(3.9);
4. Por meio da equação (3.17), finalmente obtém-se: $c_{\text{otimo},1} = -0,2321$ e $c_{\text{otimo},2} = -0,3058$.

Os resultados obtidos utilizando a metodologia proposta podem ser resumidos na Tabela 3.1. Os pólos de Kautz β_1 e β_2 foram calculados usando as relações (2.27) e (2.28).

Tabela 3.1: Valores sub-ótimos dos parâmetros e dos pólos de Kautz para cada direção do kernel $h_2(k_1, k_2)$ em (3.21).

Direção de h_2	b_l	$c_{\text{otimo},l}$	β_l
$l = 1$	0,6	-0,2321	$0,3696 \pm i0,3090$
$l = 2$	0,5	-0,3058	$0,3264 \pm i0,4463$

Com os valores sub-ótimos dos parâmetros de Kautz mostrados na Tabela 3.1, a aproximação do kernel em (3.21) tem coeficientes $\alpha_{(\cdot)}$ calculados conforme a equação (2.16), que fica:

$$\alpha_{i_1, i_2} = \sum_{k_1=0}^{30} \sum_{k_2=0}^{30} h_2(k_1, k_2) \psi_{1, i_1}(k_1) \psi_{2, i_2}(k_2) \quad (3.22)$$

onde as funções $\psi_{1, i_1}(k)$ estão parametrizadas com $(b_1; c_{\text{otimo},1})$ e as funções $\psi_{2, i_2}(k)$ estão parametrizadas com $(b_2; c_{\text{otimo},2})$. Utilizando-se então a equação (2.21), obtém-se a aproximação correspondente do kernel h_2 em função do número de termos M da base de Kautz, i.e.:

$$\tilde{h}_2(k_1, k_2) = \sum_{i_1=1}^M \sum_{i_2=1}^M \alpha_{i_1, i_2} \psi_{1, i_1}(k_1) \psi_{2, i_2}(k_2)$$

A Figura 3.4 ilustra esta aproximação do kernel h_2 para o caso em que $M = 8$. O valor do EQN associado a esta aproximação, calculado pela equação (3.3), está

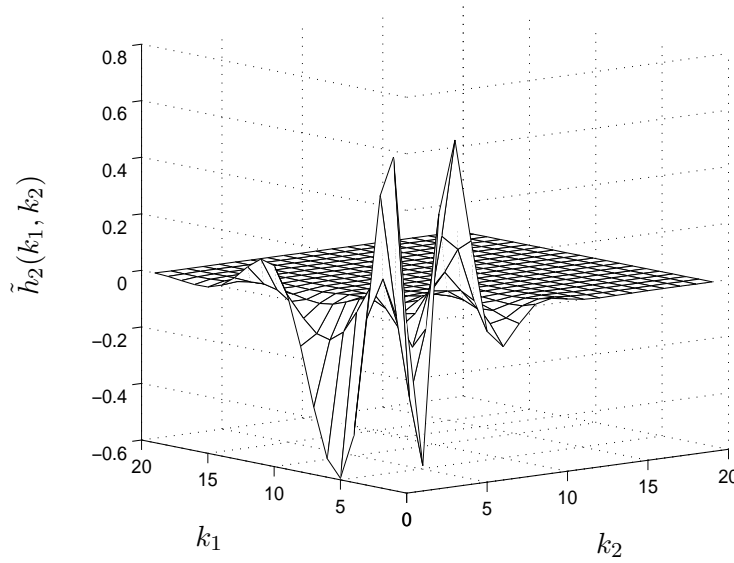


Figura 3.4: Aproximação do kernel em (3.21) usando $M = 8$ funções de Kautz com os parâmetros sub-ótimos $(b_1; c_{\text{otimo},1}) = (0, 6; -0, 2321)$ e $(b_2; c_{\text{otimo},2}) = (0, 5; -0, 3058)$.

Tabela 3.2: Erro de aproximação do kernel h_2 em (3.21) de acordo com o número de funções de Kautz utilizadas na expansão ortonormal.

Número de funções (M)	EQN = $\frac{\ h_2 - \tilde{h}_2\ ^2}{\ h_2\ ^2}$
2	0,8110
4	0,1036
6	0,0077
8	$8,1136 \times 10^{-4}$

mostrado na Tabela 3.2 para diferentes números de funções de Kautz. As funções de Kautz estão parametrizadas com os pólos mostrados na Tabela 3.1.

A Tabela 3.2 apresenta resultados quantitativos a respeito da aproximação do kernel (3.21) usando bases de Kautz, em que cada direção de $h_2(k_1, k_2)$ é representada por um pólo de Kautz individual. Estes resultados indicam que a seleção ótima dos pólos permite obter uma qualidade satisfatória de aproximação utilizando um número reduzido de funções de Kautz. Pode-se ainda representar graficamente o erro proveniente da aproximação do kernel h_2 , como ilustrado na Figura 3.5. Este erro representa a diferença entre as superfícies mostradas das Figuras 3.3 e 3.4.

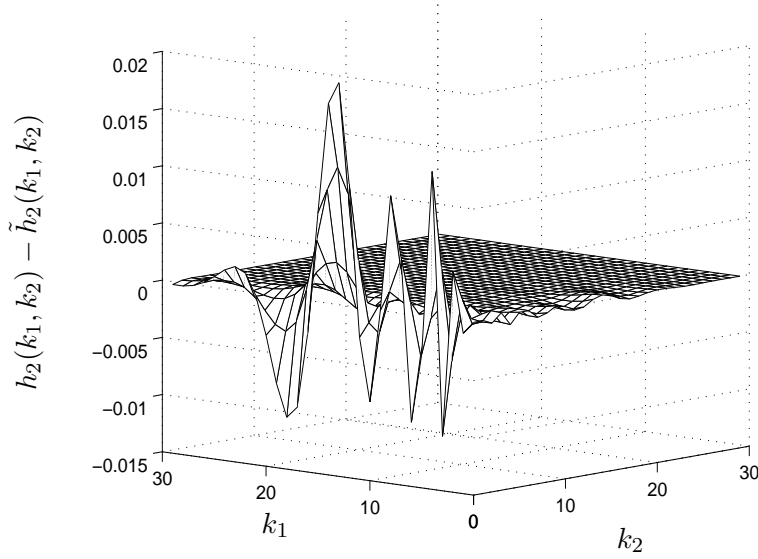


Figura 3.5: Superfície de erro de aproximação do kernel em (3.21) usando $M = 8$ funções de Kautz com os parâmetros sub-ótimos $(b_1; c_{\text{otimo},1}) = (0, 6; -0, 2321)$ e $(b_2; c_{\text{otimo},2}) = (0, 5; -0, 3058)$.

O pólo ótimo de Kautz para o kernel h_2 pode ser obtido fazendo-se uma busca em todo o espaço de factibilidade dos pólos. Variando os parâmetros $(b_1, b_2) \in (-1, 1) \times (-1, 1)$, obtém-se os valores de $c_{\text{otimo},1}$ e $c_{\text{otimo},2}$, respectivamente, para os quais os pares de parâmetros de Kautz $(b_1, c_{\text{otimo},1})$ e $(b_2, c_{\text{otimo},2})$ resultam em um determinado valor de EQN. O menor valor deste erro para o kernel h_2 dado na equação (3.21) é obtido escolhendo-se $b_1 = 0,770$ e $b_2 = 0,345$. Os valores dos parâmetros e dos pólos de Kautz nesta condição estão apresentados na Tabela 3.3. A Figura 3.6 mostra a superfície do erro de aproximação do kernel h_2 usando-se $M = 8$ funções da base de Kaut. O valor do erro desta aproximação é $\text{EQN} = 3,6082 \times 10^{-7}$.

Tabela 3.3: Valores ótimos dos parâmetros e dos pólos de Kautz para o kernel $h_2(k_1, k_2)$ em (3.21).

Direção de h_2	b_l	$c_{\text{otimo},l}$	β_l
$l = 1$	0,770	-0,4583	$0,5615 \pm i0,3783$
$l = 2$	0,345	-0,2473	$0,2152 \pm i0,4483$

Os coeficientes da expansão ortonormal em (3.22) são apresentados a seguir, para

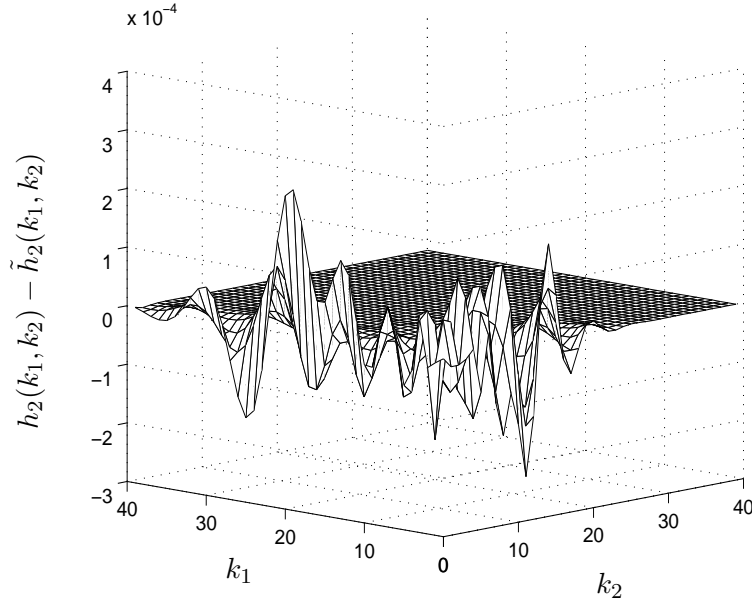


Figura 3.6: Superfície de erro de aproximação do kernel em (3.21) usando $M = 8$ funções de Kautz com os parâmetros ótimos $(b_1; c_{\text{otimo},1}) = (0,770; -0,4583)$ e $(b_2; c_{\text{otimo},2}) = (0,345; -0,2473)$.

os pólos mostrados na Tabela 3.3, para diferentes números de funções da base.

$$M = 2 \quad \Rightarrow \quad \alpha_{i_1, i_2} = \begin{bmatrix} 0,1653 & -0,4698 \\ 0,5401 & -0,1488 \end{bmatrix}$$

$$M = 4 \quad \Rightarrow \quad \alpha_{i_1, i_2} = \begin{bmatrix} 0,1653 & -0,4698 & 0,8225 & 0,4091 \\ 0,5401 & -0,1488 & 0,1092 & 0,3104 \\ -0,7618 & 0,0941 & 0,0087 & 0,0544 \\ -0,7487 & -0,2360 & -0,0669 & 0,0094 \end{bmatrix}$$

$$M = 6 \quad \Rightarrow$$

$$\alpha_{i_1, i_2} = \begin{bmatrix} 0,1653 & -0,4698 & 0,8225 & 0,4091 & 0,1182 & 0,0162 \\ 0,5401 & -0,1488 & 0,1092 & 0,3104 & 0,0665 & 0,0391 \\ -0,7618 & 0,0941 & 0,0087 & 0,0544 & 0,0088 & 0,0051 \\ -0,7487 & -0,2360 & -0,0669 & 0,0094 & -0,0005 & 0,0034 \\ -0,1284 & -0,0330 & -0,0082 & 0,0018 & 0,0000 & 0,0004 \\ -0,0507 & -0,0444 & -0,0082 & -0,0023 & -0,0005 & 0,0001 \end{bmatrix}$$

3.2.1 Comparação: base única *versus* bases múltiplas

A solução dada na equação (3.17) permite utilizar diferentes pólos (parâmetros) de Kautz em cada direção $l = 1, \dots, \eta$ do kernel $h_\eta(k_1, \dots, k_\eta)$. Esta solução apresenta maior flexibilidade em relação a um método anterior (da Rosa 2005, da Rosa *et al.* 2005, da Rosa *et al.* 2007), que permitia somente a utilização de um pólo de Kautz para todas as direções do kernel. Para efeitos de comparação, esta seção objetiva aplicar ambos os métodos ao kernel h_2 em (3.21), a fim de se tirar conclusões a respeito da melhoria do presente método para com o anterior.

Considere então novamente o kernel na equação (3.21). Os mesmos valores para os parâmetros b_1 e b_2 escolhidos na simulação anterior serão utilizados agora. Em outras palavras, a escolha de um mesmo valor de b_l , $\forall l \in \{1, \dots, \eta\}$, representa exatamente o caso particular encontrado na literatura (da Rosa *et al.* 2005, da Rosa *et al.* 2007). A Tabela 3.4 mostra os resultados obtidos em termos do número de funções de Kautz utilizadas no modelo.

Tabela 3.4: Erro de aproximação do kernel h_2 em (3.21) usando base única de Kautz.

Número de funções (M)	EQN para $b_1 = b_2 = 0,6$	EQN para $b_1 = b_2 = 0,5$
2	0,8422	0,7878
4	0,1081	0,2029
6	0,0086	0,0133
8	0,0008	0,0046

Conforme pode ser observado ao se comparar os resultados mostrados nas Tabelas 3.2 e 3.4, em geral o erro de aproximação de h_2 é menor quando pólos distintos são utilizados para representar cada direção do kernel. O valor do erro não é necessariamente menor porque os parâmetros b_l não são ótimos, uma vez que estes são escolhidos arbitrariamente pelo usuário. Uma metodologia proposta em trabalhos anteriores (da Rosa *et al.* 2005, da Rosa *et al.* 2007) como a melhor seleção dos parâmetros de Kautz fornece o par $(b; c_{\text{ótimo}}) = (0,593; -0,2594)$, para as duas direções de h_2 . O erro de aproximação deste kernel quando usa-se $M = 8$ funções da base de Kautz parametrizadas em $(b_1; c_{\text{ótimo},1}) = (b_2; c_{\text{ótimo},2}) = (0,593; -0,2594)$ é igual a $\text{EQN} = 1,242 \times 10^{-3}$. De acordo com os resultados apresentados na seção anterior, considerando-se os parâmetros ótimos obtidos $(b_1; c_{\text{ótimo},1}) = (0,770; -0,4583)$ e $(b_2; c_{\text{ótimo},2}) = (0,345; -0,2473)$, viu-se que o erro correspondente foi $\text{EQN} = 3,6082 \times 10^{-7}$. Para se atingir um erro desta ordem usando o outro método (da Rosa *et al.* 2005, da Rosa *et al.* 2007), seriam necessárias 14 funções de Kautz.

Nota-se que o kernel em (3.21) é assimétrico, i.e., $h_2(k_1, k_2) \neq h_2(k_2, k_1)$. Isto significa que o kernel h_2 possui dinâmicas diferentes em cada direção, sugerindo assim a utilização de pólos distintos para representar cada uma de suas direções axiais. Analisando os resultados numéricos apresentados nesta seção, verifica-se que o uso de bases múltiplas (η bases independentes) reduz o erro de truncamento quando o

kernel apresenta dinâmicas dominantes distintas em cada direção.

3.3 Estimação simultânea dos kernels de Volterra e dos pólos de Kautz

A aplicação do resultado teórico descrito na seção 3.1, em especial a utilização da equação (3.17) para seleção dos parâmetros de Kautz, requer um conhecimento prévio dos kernels do sistema. No entanto, em geral esta informação não está disponível, e portanto os kernels precisam ser estimados a partir de um conjunto de medições experimentais de entrada/saída do sistema. Propõe-se a seguir um algoritmo iterativo que consiste em recuperar os kernels de Volterra a partir de uma estimação via mínimos quadrados dos coeficientes da expansão em bases de Kautz, seguidos pelo cálculo dos pólos ótimos correspondentes. A ideia central deste algoritmo foi proposta originalmente para modelos de Volterra baseados em funções de Laguerre (Campello 2002), sendo posteriormente adaptado para modelos baseados em funções de Kautz (da Rosa 2005, da Rosa *et al.* 2006).

Os passos do algoritmo estão detalhados a seguir. Partindo de um pólo de Kautz inicial para cada direção $l = 1, \dots, \eta$ de cada kernel $h_\eta(k_1, \dots, k_\eta)$, faça:

1. Selecionar a ordem N do modelo de Volterra a ser utilizado, como na equação (2.1). Modelos de ordem superior a 2 podem possuir elevado custo computacional.
2. Adotar um truncamento dos kernels a serem estimados, ou seja, o valor de k_1, k_2, \dots, k_η acima dos quais os kernels de Volterra são considerados nulos. Estes valores geralmente podem ser selecionados observando-se o tempo de resposta do sistema.
3. Escolher o número M de funções de Kautz que serão utilizadas na aproximação dos kernels $h_\eta(k_1, \dots, k_\eta)$, bem como um valor inicial para os pólos destas funções.
4. A partir dos dados de entrada/saída disponíveis, estimar via mínimos quadrados os coeficientes da expansão ortonormal $\alpha_{(\cdot)}$ em (2.22).
5. Usando os coeficientes $\alpha_{(\cdot)}$, calcular os kernels de Volterra usando a equação (2.21).
6. Para cada direção, encontrar os parâmetros de Kautz b_l e c_l que fornecem o menor erro de aproximação do kernel (EQN) variando-se b_l no intervalo $(-1, 1)$, e calculando os respectivos parâmetros ótimos c_l por meio da equação (3.17).

7. Retornar ao passo 4 até que um dos critérios seja alcançado: ou a diferença entre os kernels em iterações sucessivas seja menor do que um limitante, ou que um número máximo de iterações seja excedido.

Para ilustrar o desempenho desse algoritmo, foram utilizados dados medidos em laboratório de um sistema de levitação magnética, descrito a seguir.

3.3.1 Modelagem de um sistema de levitação magnética

O sistema de levitação magnética considerado aqui, esquematicamente descrito na Figura 3.7, consiste de uma bobina superior e uma inferior que produzem um campo magnético em resposta a uma corrente DC. Dois discos magnéticos movimentam-se por uma haste de vidro. Energizando-se a bobina inferior, um dos discos é levantado por uma força magnética repulsiva. A medida que a corrente na bobina aumenta, também aumentam a intensidade do campo e a altura do disco levantado. Dois sensores a laser medem as posições dos discos. Os discos são feitos de um material altamente rígido (NeBFe) e projetados para oferecer grandes deslocamentos verticais, para demonstrar claramente princípios de levitação e controle de movimento (ECP 1999).

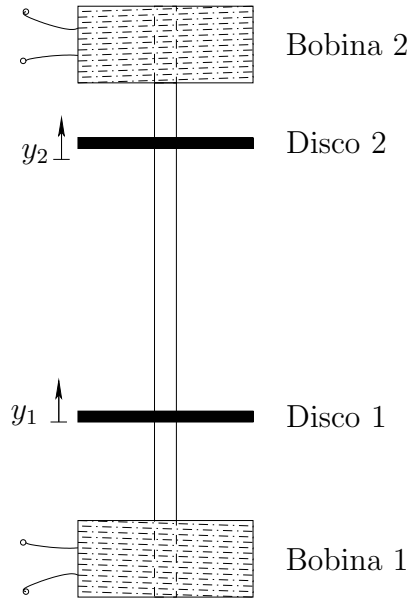


Figura 3.7: Diagrama do levitador magnético.

Considere então o movimento do disco magnético inferior (Disco 1). Seja m_1 a massa do Disco 1, ϑ_1 o coeficiente de atrito viscoso entre este disco e o ar, e g a aceleração da gravidade. A equação que descreve o movimento do Disco 1 é dada por:

$$m_1 \ddot{y}_1 + \vartheta_1 \dot{y}_1 + F_{m12} = F_{u11} - F_{u21} - m_1 g$$

onde y_1 é a posição do Disco 1, F_{u11} é a força magnética de interação entre a Bobina 1 e o Disco 1, F_{u21} é a força magnética de interação da Bobina 2 com o Disco 1, e F_{m12} é a força magnética mútua entre ambos os discos magnéticos. Estas forças são descritas pelas seguintes equações não-lineares:

$$\begin{aligned} F_{m12} &= \frac{v_3}{(y_c + y_2 - y_1 + v_4)^4} \\ F_{u11} &= \frac{i_1}{v_1(k_s y_1 + v_2)^4} \\ F_{u21} &= \frac{i_2}{v_1(y_c + k_s y_1 + v_2)^4} \end{aligned}$$

com:

- m_1 : massa do Disco 1;
- ϑ_1 : coeficiente de atrito viscoso do Disco 1 com o ar;
- y_1 : altura do Disco 1;
- y_2 : altura do Disco 2;
- i_1 : corrente elétrica na Bobina 1;
- i_2 : corrente elétrica na Bobina 2;
- y_c : distância entre as Bobinas 1 e 2;
- v_1, v_2, v_3, v_4 : constantes reais.

A seguir, apresenta-se resultados experimentais com o objetivo de modelar uma planta do sistema de levitação magnética descrito acima.

3.3.2 Simulação computacional para o levitador

Os dados experimentais foram adquiridos mantendo-se constante a corrente elétrica aplicada à Bobina 1 da planta, enquanto variava-se a corrente através da Bobina 2. O sinal de entrada u (corrente na Bobina 2) foi projetado como sendo uma sequência de degraus com amplitudes diferentes de forma a excitar os diferentes modos do sistema. O sinal de saída medido y foi tomado como sendo a posição do Disco 1 (y_1). A Figura 3.8 mostra os dados de entrada e saída disponíveis para estimação do modelo. Antes da estimação, estes dados foram amostrados usando-se um período de amostragem de 0,017 segundos, e normalizados dentro do intervalo $[-1, 1]$ para evitar problemas numéricos. Um outro conjunto similar de dados também foi adquirido e guardado para posterior validação do modelo.

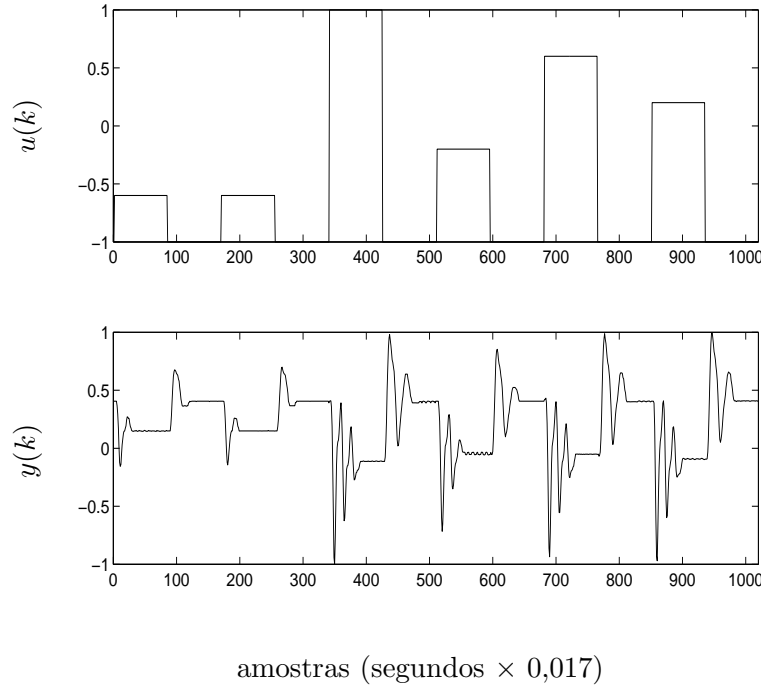


Figura 3.8: Dados entrada/saída do levitador magnético. Acima: corrente através da Bobina 2 (sinal de entrada u). Abaixo: posição do Disco 1 (sinal de saída y).

O objetivo desta simulação é obter um modelo que relacione a entrada u e a saída y , mostradas na Figura 3.8, por meio de uma representação de Volterra de segunda ordem ($N = 2$), como usual na literatura (Billings 1980, Dumont e Fu 1993):

$$\hat{y}(k) = \sum_{\tau_1=0}^{\epsilon} h_1(\tau_1)u(k - \tau_1) + \sum_{\tau_1=0}^{\epsilon} \sum_{\tau_2=0}^{\epsilon} h_2(\tau_1, \tau_2)u(k - \tau_1)u(k - \tau_2)$$

em que os kernels h_1 e h_2 serão representados usando-se uma expansão em base de Kautz $\{\psi_{l,n}\}$, de acordo com a equação (2.21). O modelo resultante será então como a seguir:

$$\begin{aligned} \hat{y}(k) &= \sum_{\tau_1=0}^{\epsilon} \sum_{i_1=1}^M \alpha_{i_1} \psi_{1,i_1}(k_1) u(k - \tau_1) \\ &+ \sum_{\tau_1=0}^{\epsilon} \sum_{\tau_2=0}^{\epsilon} \sum_{i_1=1}^M \sum_{i_2=1}^M \alpha_{i_1, i_2} \psi_{1,i_1}(k_1) \psi_{2,i_2}(k_2) u(k - \tau_1) u(k - \tau_2) \end{aligned} \quad (3.23)$$

No modelo (3.23), apenas os sinais de entrada e saída são conhecidos. Neste caso, nenhuma informação prévia sobre os kernels do modelo está disponível; então é

necessário estimá-los a partir de dados medidos de entrada/saída do sistema para se obter os pólos ótimos de Kautz.

Baseado no tempo de resposta do sistema de levitação magnética, os termos do kernel de memória longa – maior do que 50 elementos – são considerados nulos. Em outras palavras, $\epsilon = 50$ é adotado em (3.23). Nesta simulação, admite-se que o kernel de segunda ordem é simétrico, i.e. $h_2(k_1, k_2) = h_2(k_2, k_1)$. Além disso, $M = 8$ funções de Kautz foram utilizadas. Escolhendo-se, por exemplo, os pólos iniciais de Kautz como $0,5 \pm i0,5$ para os kernel de primeira ordem, bem como para as duas direções do kernel de segunda ordem, os valores iniciais correspondentes para os parâmetros de Kautz, de acordo com as equações (2.27) e (2.28), são $b_1 = b_2 = 0,6667$ e $c_1 = c_2 = -0,5$. A Tabela 3.5 apresenta os valores dos pólos de Kautz após a convergência do procedimento acima, que levou 17 passos. O critério de parada adotado foi que a diferença entre as normas quadráticas de ambos os kernels em passos sucessivos seja menor do que 10^{-3} . Simulações computacionais com diferentes pólos iniciais forneceram resultados similares.

Tabela 3.5: Valores finais dos parâmetros de Kautz para os kernels de primeira e segunda ordem do modelo do levitador magnético.

Ordem do kernel (η)	Direção axial de h_η	b_l	c_l	$(\beta_l, \bar{\beta}_l)$
1	$l = 1$	0,9047	-0,6832	$0,7614 \pm i0,3216$
2	$l = 1$	0,9323	-0,7604	$0,8206 \pm i0,2949$
	$l = 2$	0,9434	-0,6521	$0,7793 \pm i0,2116$

Usando-se os valores dos pólos de Kautz mostrados na Tabela 3.5 e os correspondentes coeficientes da expansão ortonormal (não mostrados), os kernels $h_1(k_1)$ e $h_2(k_1, k_2)$ podem ser calculados de acordo com a equação (2.21). Estes kernels estão ilustrados nas Figuras 3.9 e 3.10, respectivamente.

A Figura 3.11 ilustra a previsão do modelo da saída do sistema, $\hat{y}(k)$, com respeito a um outro conjunto de dados (dados de validação), juntamente com o valor real (medido) da saída $y(k)$. Pode-se notar a boa qualidade do modelo obtido, uma vez que o gráfico da saída do modelo sobrepõe quase perfeitamente a saída medida do sistema.

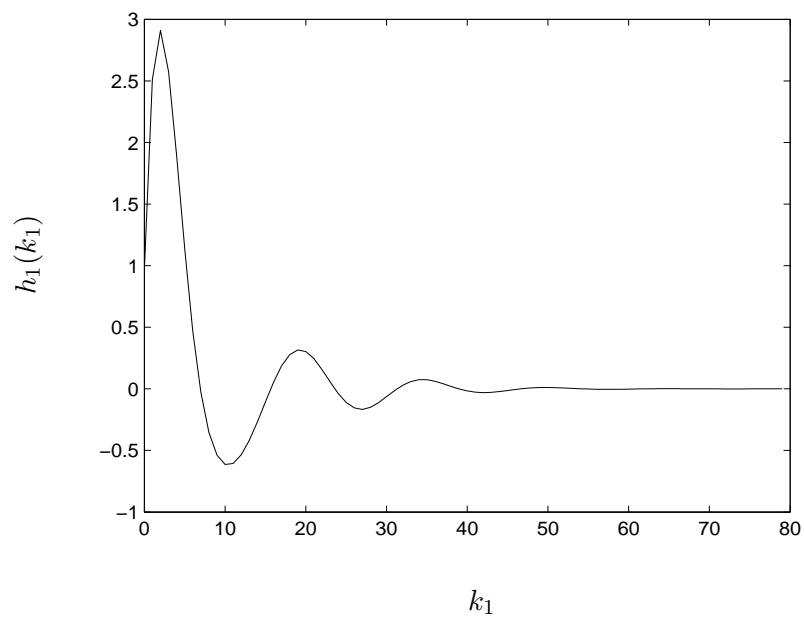


Figura 3.9: Kernel de primeira ordem estimado para o modelo do levitador magnético.

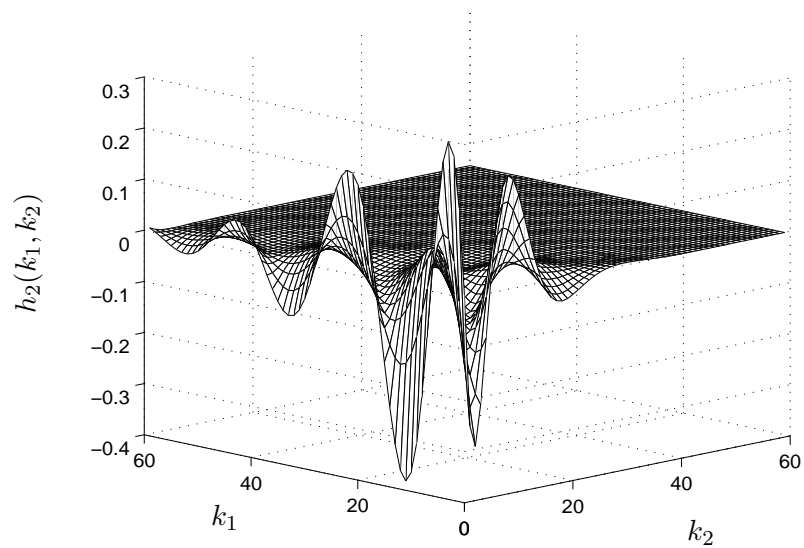


Figura 3.10: Kernel de segunda ordem estimado para o modelo do levitador magnético.

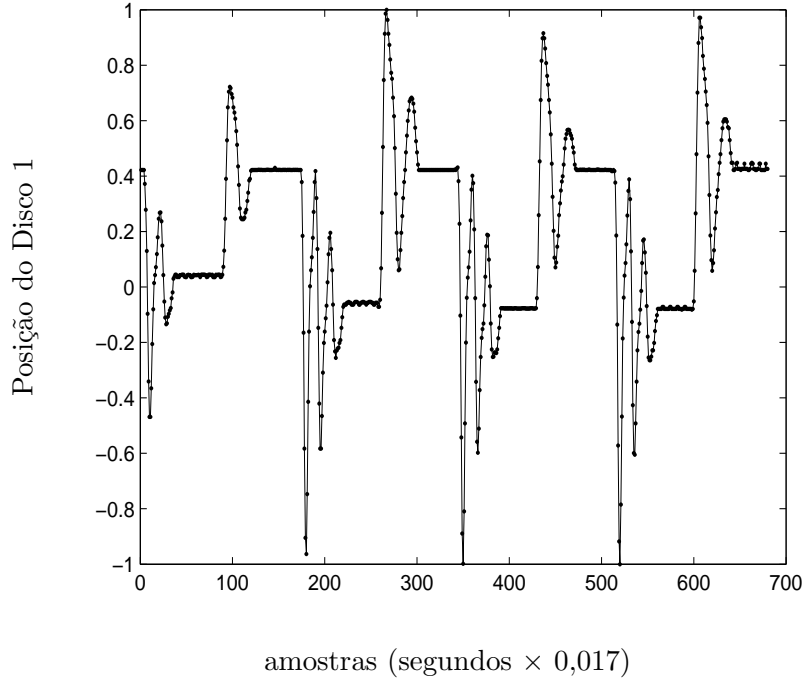


Figura 3.11: Saída medida (em linha contínua) do sistema de levitação magnética e saída do modelo (em pontilhado).

3.4 Resumo e contribuições

Este capítulo apresentou resultados teóricos e práticos referentes ao problema da aproximação de modelos de Volterra baseados em funções ortonormais de Kautz. Discutiu-se o problema da seleção ótima de um dos parâmetros de Kautz em termos do erro de aproximação dos kernels de Volterra. A otimização simultânea dos dois parâmetros de Kautz usando uma abordagem analítica ainda é um problema em aberto. É possível, porém, abordar este problema minimizando-se um limitante superior do erro de aproximação dos kernels com relação a um dos parâmetros, quando o outro é mantido fixo.

A contribuição central foi a obtenção de uma expressão analítica para um limitante superior do erro de aproximação truncada dos kernels de Volterra quando estes são desenvolvidos em base de Kautz. Através da minimização deste limitante superior, obteve-se uma solução analítica global para a seleção ótima de um dos parâmetros de Kautz em modelos de Volterra de qualquer ordem. Esta estratégia permite calcular analiticamente um dos parâmetros ótimos de Kautz, o que pode ser interpretado como uma expressão analítica sub-ótima para os pólos de Kautz. Na abordagem apresentada, um conjunto de funções ortonormais de Kautz independentes é utilizado para representar cada direção axial do kernel de Volterra. Os

resultados apresentados representam uma generalização de trabalhos anteriores (da Rosa 2005, da Rosa *et al.* 2005, da Rosa *et al.* 2007), em que todas as direções do kernel adotam um único pólo de Kautz. Esta nova metodologia também representa uma extensão de um problema análogo investigado no contexto da base de Laguerre (Campello *et al.* 2006).

Por meio de um exemplo ilustrativo, verificou-se que o uso de bases independentes para representar cada direção dos kernels reduz o erro de truncamento quando o kernel apresenta dinâmicas distintas em cada uma de suas direções. Resultados de simulação mostraram que a metodologia proposta forneceu uma boa aproximação de sistemas não-lineares com comportamento oscilatório. A seleção ótima dos pólos de Kautz permite obter uma maior qualidade de aproximação utilizando um menor número de funções da base, reduzindo a ordem de representação do modelo de Volterra resultante.

Apresentou-se também um algoritmo iterativo para a estimação simultânea dos kernels de Volterra e dos pólos de Kautz, a partir da medição de dados entrada/saída. Este algoritmo representa uma adaptação de algoritmos que já foram apresentados na literatura para a base de Laguerre (Campello 2002) e para a base de Kautz com pólo único (da Rosa 2005). A implementação computacional deste procedimento permitiu obter um modelo OBF-Volterra de segunda ordem que representa adequadamente as dinâmicas de um sistema não-linear de levitação magnética.

Capítulo 4

Direções de Busca Exatas Para Otimização dos Pólos de Bases Ortonormais

O problema central abordado no Capítulo 3 envolve a minimização de um limitante superior do erro de aproximação dos kernels de Volterra, quando bases independentes de funções de Kautz são utilizadas para representar cada direção dos kernels. A principal limitação desta abordagem é que ela requer um conhecimento prévio sobre os kernels do sistema a ser modelado. Esta otimização é também apenas sub-ótima, uma vez que somente um dos parâmetros de Kautz é otimizado, e um limitante do erro é considerado como função de custo.

Para eliminar as desvantagens discutidas acima, este capítulo apresenta resultados referentes à aproximação de modelos OBF-Volterra por meio da seleção numérica dos pólos das bases ortonormais. Neste aspecto, obtém-se os gradientes das saídas dos filtros ortonormais com relação aos pólos da base. Estes gradientes podem ser utilizados juntamente com um algoritmo de otimização para obter uma direção a ser seguida no espaço de busca dos pólos ótimos da base ortonormal.

Os resultados iniciais apresentados são válidos para modelos lineares, sendo posteriormente estendidos para modelos de Volterra de qualquer ordem. Além da obtenção das expressões analíticas para os gradientes de modelos baseados em funções de Kautz, obtém-se também as equações para modelos baseados em funções ortonormais generalizadas (GOBF). A abordagem proposta é ainda aplicada na modelagem do sistema de levitação magnética descrito no Capítulo 3.

Neste capítulo, o problema abordado consiste na minimização de uma função de custo quadrática que leva em conta o erro de estimação da saída do sistema. Particular ênfase é dada para modelos baseados em funções de Kautz, porém uma análise similar para as funções de Laguerre pode ser facilmente obtida como um caso particular. As derivadas obtidas são utilizadas como parte de um método de otimização não-linear, para obtenção da direção de busca para os pólos da base. Adota-se aqui o algoritmo de otimização de Levenberg-Marquardt

(Levenberg 1944, Marquardt 1963, Nocedal e Wright 1999, Nelles 2001), que vem a ser uma escolha usual em otimização não-linear, uma vez que ele pode representar o método de descida do gradiente e o método do gradiente conjugado em problemas de porte médio. O método de Levenberg-Marquardt é um algoritmo de pseudo-segunda ordem, pois ele trabalha somente com avaliações da função de custo e com informações do gradiente. A matriz Hessiana da função de custo é estimada usando estas informações. Parte dos resultados apresentados aqui também podem ser encontrados em outros trabalhos, para o caso de modelos lineares (da Rosa *et al.* 2008b) e para modelos não-lineares de Volterra (da Rosa *et al.* aceito).

Os trabalhos que abordaram a seleção de pólos das bases ortonormais em geral não apresentam garantia de otimalidade, ou de convergência, ou de ambos. O objetivo deste capítulo é propor uma nova metodologia para seleção dos pólos das bases ortonormais que garanta otimalidade e convergência.

4.1 Formulação do problema

Esta seção formula um problema de se ajustar a relação dinâmica entre os parâmetros de modelos OBF e a saída do modelo. Primeiramente, a técnica *back-propagation-through-time* (Nelles 2001) é brevemente revisada, cuja ideia básica é recuperar a dinâmica de um dado modelo de forma recursiva até que o modelo seja descrito por relações estáticas. Posteriormente discute-se a otimização dos parâmetros de modelos OBF lineares baseada nessa técnica. Para um melhor entendimento a respeito desta técnica, ela é inicialmente apresentada no contexto dos sistemas lineares. A generalização deste estudo para os modelos não-lineares será estudada posteriormente na seção 4.4, dentro do contexto dos modelos de Volterra.

4.1.1 Técnica *back-propagation-through-time*

Considere um sistema dinâmico linear representado por um mapeamento estático genérico \mathcal{H} aplicado a um conjunto de termos de regressão da entrada e saída, como a seguir:

$$\hat{y}(k) = \mathcal{H}[u(k-1), \dots, u(k-n_u), \hat{y}(k-1), \dots, \hat{y}(k-n_y)] \quad (4.1)$$

onde $\hat{y}(k)$ representa a saída estimada do modelo. Neste tipo de modelo, conhecido como modelo *Output-Error* (OE) (Nelles 2001), o mapeamento estático \mathcal{H} se relaciona dinamicamente com a saída do modelo, uma vez que ele está presente implicitamente e recursivamente nos termos $\hat{y}(k-i)$. Assim, os parâmetros do mapeamento \mathcal{H} não podem ser estimados diretamente a partir dos dados usando uma abordagem padrão de estimação, como o algoritmo de mínimos quadrados. Em vez disso, a abordagem consiste em descrever a dinâmica do sistema por meio de uma série de modelos estáticos representados através de relações explícitas no tempo. Para ilustrar esta discussão, considere por exemplo o modelo em (4.1) com $n_u = n_y = 1$,

i.e.:

$$\hat{y}(k) = \mathcal{H}[u(k-1), \hat{y}(k-1)] \quad (4.2)$$

Suponha que o mapeamento estático \mathcal{H} seja parametrizado por um vetor de parâmetros $\boldsymbol{\theta}$ do modelo. Este vetor pode ser otimizado se o gradiente de uma função de custo com relação a todos os elementos do vetor $\boldsymbol{\theta}$ estiver disponível. Este gradiente depende da derivada da saída do modelo $\hat{y}(k)$ com relação aos elementos de $\boldsymbol{\theta}$. Particularmente, como o modelo é assumido linear, pode-se inferir a partir da equação (4.2) que:

$$\frac{\partial \hat{y}(k)}{\partial \theta_i} = \frac{\partial \mathcal{H}_1[\cdot]}{\partial \theta_i} + \frac{\partial \mathcal{H}_2[\cdot]}{\partial \hat{y}(k-1)} \cdot \frac{\partial \hat{y}(k-1)}{\partial \theta_i} \quad (4.3)$$

onde \mathcal{H}_2 e \mathcal{H}_1 são as porções complementares de \mathcal{H} que, respectivamente, dependem e não dependem de $\hat{y}(k-1)$, enquanto que θ_i é o i -ésimo elemento do vetor de parâmetros $\boldsymbol{\theta}$. O primeiro termo em (4.3) é o gradiente da saída do modelo com relação aos parâmetros e representa a parte estática, enquanto que o segundo termo representa a parte dinâmica. Se $\hat{y}(k-1)$ for um sinal externo, como por exemplo a saída medida do processo $y(k-1)$, então não haverá nenhuma dependência com os parâmetros do modelo, e portanto este termo deve ser igual a zero. Em modelos OE, contudo, $\hat{y}(k-1)$ é a saída do modelo em um instante de tempo anterior, que depende de $\boldsymbol{\theta}$.

O cálculo do segundo termo da equação (4.3) requer a derivada da saída anterior do modelo com relação aos parâmetros. Esta pode ser calculada a partir de:

$$\hat{y}(k-1) = \mathcal{H}[u(k-2), \hat{y}(k-2)] \quad (4.4)$$

Assim, tem-se que a derivada em (4.3) torna-se:

$$\frac{\partial \hat{y}(k-1)}{\partial \theta_i} = \frac{\partial \mathcal{H}_1[\cdot]}{\partial \theta_i} + \frac{\partial \mathcal{H}_2[\cdot]}{\partial \hat{y}(k-2)} \cdot \frac{\partial \hat{y}(k-2)}{\partial \theta_i} \quad (4.5)$$

Novamente, o segundo termo em (4.5) requer as derivadas da saída do modelo num instante de tempo anterior. Este procedimento continua até que as condições iniciais em $k = 0$ sejam atingidas, para o qual se tem $\hat{y}(0) = y(0)$ e portanto:

$$\frac{\partial \hat{y}(0)}{\partial \theta_i} = 0$$

Em resumo, o algoritmo *back-propagation-through-time* decompõe a dinâmica de um sistema em uma série de representações estáticas. Esta abordagem permite descrever as derivadas da saída de um modelo em termos somente das condições iniciais e do sinal de entrada, voltando k passos no tempo. A técnica *back-propagation-through-time* desenvolve um modelo de forma recursiva até se atingir as condições iniciais em $k = 0$, como ilustrado na Figura 4.1.

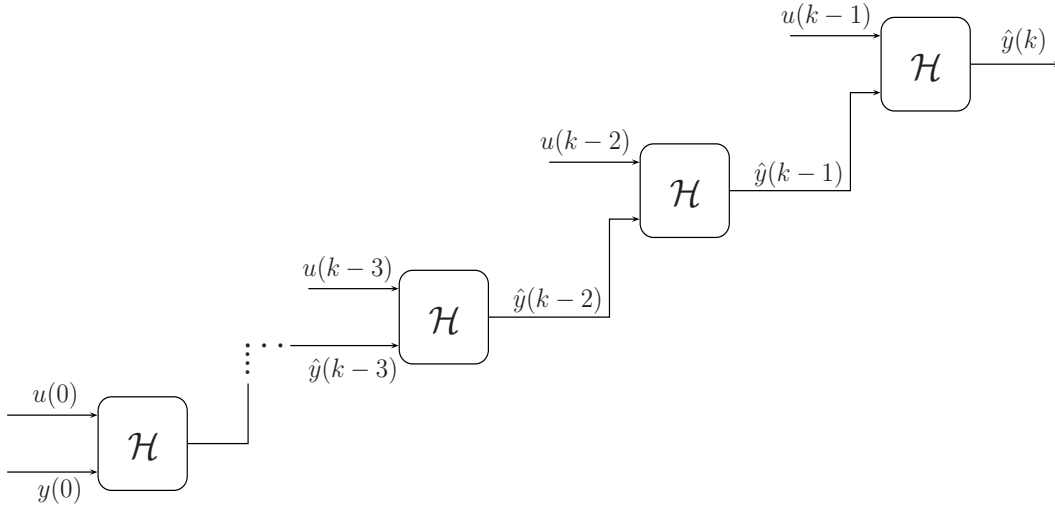


Figura 4.1: Esquema da técnica *back-propagation-through-time*.

Se o modelo é linear, então o mapeamento estático \mathcal{H} mostrado na Figura 4.1 é uma função linear. Se o modelo é uma representação de Volterra, então \mathcal{H} é a série de Volterra.

De acordo com o esquema da técnica *back-propagation-through-time* ilustrado na Figura 4.1, nota-se que a saída do próprio modelo é utilizada na recursão temporal. A estrutura desta abordagem é como mostrado na Figura 4.2.

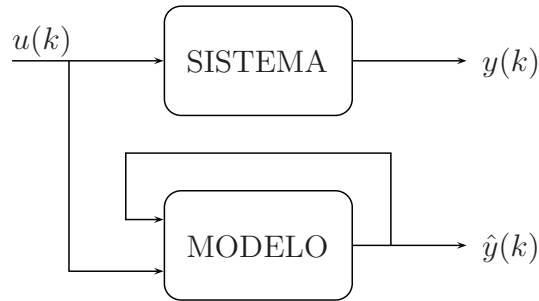


Figura 4.2: Estrutura em que a saída do modelo é utilizada na recursão temporal.

4.1.2 Otimização dos pólos

Uma representação esquemática de modelos baseados em funções ortonormais está ilustrada na Figura 4.3.

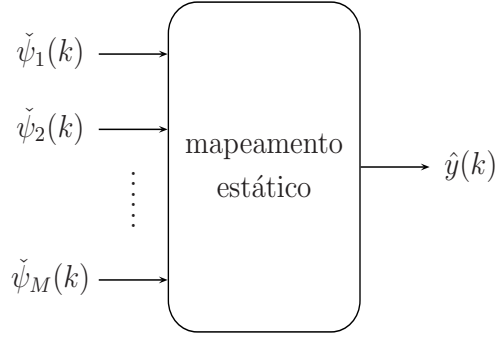


Figura 4.3: Estrutura geral de um modelo OBF: a saída é gerada por um mapeamento estático aplicado a um conjunto de funções ortonormais.

Nesse contexto, $\check{\psi}_m(k)$ representa a resposta do filtro $\psi_m(k)$ à entrada $u(k)$, i.e.:

$$\check{\psi}_m(k) = \sum_{\tau=0}^k \psi_m(\tau) u(k-\tau) \quad (4.6)$$

$$\check{\Psi}_m(z) = \Psi_m(z) U(z) \quad (4.7)$$

Considere inicialmente um modelo OBF-Volterra de primeira ordem (modelo linear) com M funções de uma determinada base ortonormal $\{\psi_m\}$. Neste caso, o mapeamento \mathcal{H} mostrado na Figura 4.3 é uma função linear. Tendo como caso geral a equação (2.22), esse modelo linear pode ser expresso como:

$$\hat{y}(k) = \sum_{m=1}^M \alpha_m \check{\psi}_m(k) \quad (4.8)$$

onde $\check{\psi}_m(k)$ foi definido na equação (4.6).

A estratégia adotada aqui consiste na otimização simultânea do vetor de pólos \mathbf{p} que parametriza a base ortonormal $\{\psi_m\}$ e do vetor de coeficientes correspondente $\boldsymbol{\alpha} \triangleq [\alpha_1 \ \alpha_2 \ \cdots \ \alpha_M]^T$. Para este fim, seja o seguinte problema de otimização:

$$\min_{\boldsymbol{\theta}} J \triangleq \frac{1}{2} \sum_{k=1}^{N_d} [y(k) - \hat{y}(k)]^2 \quad (4.9)$$

onde $\boldsymbol{\theta} = [\mathbf{p}^T \ \boldsymbol{\alpha}^T]^T$ e N_d é o número de dados entrada/saída disponíveis. O gradiente do critério quadrático J em (4.9) com relação ao vetor de pólos \mathbf{p} pode

ser calculado da seguinte forma:

$$\begin{aligned}
 \nabla_{\mathbf{p}} J &= \sum_{k=1}^{N_d} [\hat{y}(k) - y(k)] \nabla_{\mathbf{p}} \hat{y}(k) \\
 &= \sum_{k=1}^{N_d} [\hat{y}(k) - y(k)] \sum_{m=1}^M \alpha_m \nabla_{\mathbf{p}} \check{\psi}_m(k) \\
 &= \sum_{k=1}^{N_d} \sum_{m=1}^M [\hat{y}(k) - y(k)] \alpha_m \nabla_{\mathbf{p}} \check{\psi}_m(k)
 \end{aligned} \tag{4.10}$$

O gradiente de J com relação a α é facilmente obtido a partir de (4.8) como:

$$\begin{aligned}
 \nabla_{\alpha} J &= \sum_{k=1}^{N_d} [\hat{y}(k) - y(k)] \nabla_{\alpha} \hat{y}(k) \\
 &= \sum_{k=1}^{N_d} [\hat{y}(k) - y(k)] \check{\psi}(k)
 \end{aligned} \tag{4.11}$$

onde $\check{\psi}(k) = [\check{\psi}_1(k) \quad \check{\psi}_2(k) \quad \cdots \quad \check{\psi}_M(k)]^T$.

A contribuição central deste capítulo é a obtenção de equações para o cálculo de $\nabla_{\mathbf{p}} \check{\psi}_m(k)$ e a sua aplicação na otimização dos pólos do modelo (4.8) usando o bem conhecido algoritmo de Levenberg-Marquardt (Levenberg 1944, Marquardt 1963, Nocedal e Wright 1999, Nelles 2001). A base deste algoritmo é uma aproximação quadrática de J na vizinhança de seus argumentos (parâmetros do modelo). O algoritmo aproxima a matriz Hessiana de J com relação aos parâmetros usando somente informação das derivadas de primeira ordem correspondentes, que são obtidos através do cálculo dos gradientes. Especificamente, ele estima a matriz Hessiana usando a soma de produtos dos gradientes em (4.10) e (4.11). Denotando $F(\theta)$ com o Jacobiano da função de custo $J(\theta)$, o algoritmo atualiza a variável de otimização θ de acordo com a seguinte equação (Nelles 2001):

$$\theta_t = \theta_{t-1} - \left[F_{t-1}(\theta) F_{t-1}^T(\theta) + \sigma_{t-1} I \right]^{-1} F_{t-1}(\theta) J_{t-1}(\theta) \tag{4.12}$$

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

O algoritmo de Levenberg-Marquardt é um método de pseudo-segunda ordem, pois ele utiliza apenas avaliações da função e de seu gradiente. Este algoritmo é globalmente convergente, ou seja, a partir de qualquer ponto inicial θ^0 , o algoritmo converge para um ponto θ_{otimo} satisfazendo as condições de otimalidade necessárias para um minimizador local de $J(\theta)$, i.e., $\nabla_{\theta} J = \mathbf{0}$. Obviamente, não é possível garantir que θ_{otimo} seja um minimizador global de $J(\theta)$, a menos que $J(\theta)$ seja uma função convexa (Nocedal e Wright 1999). O método de Levenberg-Marquardt pode ser visto como uma combinação dos métodos do gradiente de maior descida e de

Gauss-Newton (Nelles 2001): quando o resultado está longe da solução ótima, o algoritmo se comporta como o método do gradiente (lento, mas com convergência garantida); próximo da solução ótima, o algoritmo se comporta como o método de Gauss-Newton, convergindo assim mais rápido.

Estuda-se inicialmente modelos OBF, como aquele em (4.8), quando a base ortonormal utilizada é o conjunto de funções de Kautz, como detalhado a seguir na seção 4.2. O caso com a base de funções generalizadas (GOBF) será discutido posteriormente na seção 4.3.

4.2 Modelo baseado em funções de Kautz

Nesta seção, um esquema como aquele mostrado na Figura 4.1 é desenvolvido para obter expressões que descrevem a saída dos filtros ortonormais em termos do sinal de entrada e das condições iniciais dos filtros. A partir destas expressões, calcula-se os gradientes das saídas dos filtros com relação aos pólos da base ortonormal, i.e., $\nabla_{\mathbf{p}}\check{\psi}_m(k)$ em (4.10).

4.2.1 Funções de Kautz com índice par

Em vez de considerar o vetor \mathbf{p} contendo os próprios pólos (complexos) da base de Kautz, ele será convenientemente adotado como um vetor contendo os parâmetros (reais) que descrevem os pólos de Kautz. Tem-se assim:

$$\mathbf{p} \triangleq \begin{bmatrix} b \\ c \end{bmatrix} \implies \nabla_{\mathbf{p}}\check{\psi}_m(k) = \begin{bmatrix} \frac{\partial\check{\psi}_m(k)}{\partial b} \\ \frac{\partial\check{\psi}_m(k)}{\partial c} \end{bmatrix} \quad (4.13)$$

Desta forma, o gradiente das saídas dos filtros de Kautz se resume a calcular individualmente as derivadas $\partial\check{\psi}_m(k)/\partial b$ e $\partial\check{\psi}_m(k)/\partial c$.

Inicialmente, considere a definição das funções de Kautz dadas na equação (2.26). A partir desta equação, é possível construir um diagrama de blocos correspondente a um modelo dinâmico baseado nas funções de Kautz. A Figura 4.4 ilustra este diagrama.

Note que, de acordo com o diagrama mostrado na Figura 4.4, pode-se escrever a saída do primeiro filtro par de Kautz $\check{\Psi}_2(z)$ da seguinte forma:

$$\check{\Psi}_2(z) = \frac{\sqrt{(1-c^2)(1-b^2)} z}{z^2 + b(c-1)z - c} U(z)$$

No domínio do tempo, tem-se que:

$$\check{\psi}_2(k+2) + b(c-1)\check{\psi}_2(k+1) - c\check{\psi}_2(k) = \sqrt{(1-c^2)(1-b^2)} u(k+1) \quad (4.14)$$

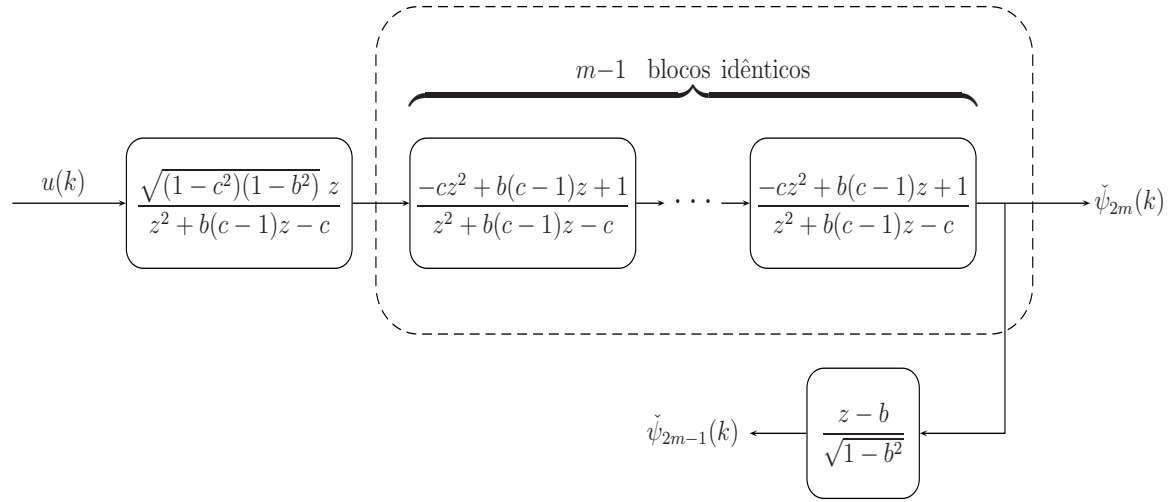


Figura 4.4: Diagrama de blocos de um modelo dinâmico baseado em funções de Kautz.

Para obter uma expressão matemática para $\check{\psi}_2(k)$ em termos somente de uma condição inicial ($\check{\psi}_2(0)$ e $\check{\psi}_2(1)$) e dos valores passados do sinal de entrada $u(k)$, a abordagem utilizada consiste em representar o modelo (4.14) na forma de espaço de estados. Neste caso, o vetor de estados possuirá dimensão igual a 2, uma vez que a equação a diferenças dada em (4.14) é de segunda ordem. Sejam então os estados $x_a(k) = \check{\psi}_2(k)$ e $x_b(k) = x_a(k+1) = \check{\psi}_2(k+1)$. Logo, é possível escrever o seguinte:

$$\begin{cases} x_a(k+1) &= x_b(k) \\ x_b(k+1) &= -b(c-1)x_b(k) + cx_a(k) + \sqrt{(1-c^2)(1-b^2)} u(k+1) \end{cases} \quad (4.15)$$

Ou equivalentemente, na forma matricial:

$$\begin{cases} \begin{bmatrix} x_a(k+1) \\ x_b(k+1) \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ c & -b(c-1) \end{bmatrix} \begin{bmatrix} x_a(k) \\ x_b(k) \end{bmatrix} \\ &+ \begin{bmatrix} 0 \\ \sqrt{(1-c^2)(1-b^2)} \end{bmatrix} u(k+1) \\ \check{\psi}_2(k) &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_a(k) \\ x_b(k) \end{bmatrix} \end{cases} \quad (4.16)$$

O modelo (4.16) possui então uma clássica representação em espaço de estados dada por:

$$\begin{cases} \mathbf{x}_2(k+1) &= \mathcal{A}\mathbf{x}_2(k) + \mathcal{B}u(k+1) \\ \check{\psi}_2(k) &= \mathcal{C}\mathbf{x}_2(k) \end{cases} \quad (4.17)$$

A solução de uma equação em espaço de estados do tipo (4.17) é bem conhecida na literatura, sendo escrita para $\check{\psi}_2(k)$ como (Gabel e Roberts 1987):

$$\check{\psi}_2(k) = \mathcal{C}\mathcal{A}^k\mathbf{x}_2(0) + \sum_{i=0}^{k-1} \mathcal{C}\mathcal{A}^i\mathcal{B}u(k-i) \quad (4.18)$$

onde $\mathbf{x}_2(0) = [\check{\psi}_2(0) \quad \check{\psi}_2(1)]^T$ e as matrizes de estado $\mathcal{A}, \mathcal{B}, \mathcal{C}$ são como segue:

$$\mathcal{A} = \begin{bmatrix} 0 & 1 \\ c & -b(c-1) \end{bmatrix} \quad (4.19)$$

$$\mathcal{B} = \begin{bmatrix} 0 \\ \sqrt{(1-c^2)(1-b^2)} \end{bmatrix} \quad (4.20)$$

$$\mathcal{C} = \begin{bmatrix} 1 & 0 \end{bmatrix} \quad (4.21)$$

O gradiente $\nabla_{\mathbf{p}}\check{\psi}_m(k)$ em (4.10) é então calculado para $m = 1$ em termos das derivadas de (4.18) com relação aos parâmetros b e c que completamente parametrizam a base de Kautz, i.e. $\nabla_{\mathbf{p}}\check{\psi}_2(k) = [\partial\check{\psi}_2(k)/\partial b \quad \partial\check{\psi}_2(k)/\partial c]^T$. Estas derivadas são dadas por:

$$\frac{\partial\check{\psi}_2(k)}{\partial b} = \mathcal{C}\frac{\partial(\mathcal{A}^k)}{\partial b}\mathbf{x}_2(0) + \mathcal{C}\sum_{i=0}^{k-1} \left[\frac{\partial(\mathcal{A}^i)}{\partial b}\mathcal{B} + \mathcal{A}^i\frac{\partial\mathcal{B}}{\partial b} \right] u(k-i) \quad (4.22)$$

$$\frac{\partial\check{\psi}_2(k)}{\partial c} = \mathcal{C}\frac{\partial(\mathcal{A}^k)}{\partial c}\mathbf{x}_2(0) + \mathcal{C}\sum_{i=0}^{k-1} \left[\frac{\partial(\mathcal{A}^i)}{\partial c}\mathcal{B} + \mathcal{A}^i\frac{\partial\mathcal{B}}{\partial c} \right] u(k-i)$$

onde as matrizes $\partial(\mathcal{A}^k)/\partial b$ e $\partial(\mathcal{A}^k)/\partial c$ podem ser calculadas da seguinte forma:

$$\frac{\partial(\mathcal{A}^k)}{\partial b} = \sum_{j=1}^k \mathcal{A}^{j-1} \frac{\partial\mathcal{A}}{\partial b} \mathcal{A}^{k-j} \quad (4.23)$$

$$\frac{\partial(\mathcal{A}^k)}{\partial c} = \sum_{j=1}^k \mathcal{A}^{j-1} \frac{\partial\mathcal{A}}{\partial c} \mathcal{A}^{k-j} \quad (4.24)$$

Portanto, para se obter o gradiente dado pela equação (4.22), é necessário inicialmente calcular a matriz \mathcal{A}^k . Neste trabalho, por exemplo, este cálculo é feito utilizando-se a representação da matriz \mathcal{A} na base de seus autovetores. Lembrando que os pólos de Kautz β e $\bar{\beta}$ estão relacionados com os parâmetros de Kautz b e c por meio das relações (2.27) e (2.28), não é difícil deduzir que:

- a matriz \mathcal{A} possui um autovalor igual a β , com autovetor correspondente $[1 \quad \beta]^T$;
- a matriz \mathcal{A} possui um autovalor igual a $\bar{\beta}$, com autovetor correspondente $[1 \quad \bar{\beta}]^T$.

Baseado nestas observações, segue que a base formada pelos autovetores de \mathcal{A} e a matriz diagonal contendo os autovalores de \mathcal{A} são dados pelas seguintes matrizes Q e Λ , respectivamente:

$$Q = \begin{bmatrix} 1 & 1 \\ \beta & \bar{\beta} \end{bmatrix} \quad \Lambda = \begin{bmatrix} \beta & 0 \\ 0 & \bar{\beta} \end{bmatrix}$$

Assim, reescreve-se a matriz \mathcal{A} como $\mathcal{A} = Q\Lambda Q^{-1}$, levando a seguinte expressão:

$$\mathcal{A}^k = Q\Lambda^k Q^{-1} = \frac{1}{\bar{\beta} - \beta} \begin{bmatrix} \bar{\beta}\beta^k - \beta\bar{\beta}^k & \bar{\beta}^k - \beta^k \\ \bar{\beta}\beta^{k+1} - \beta\bar{\beta}^{k+1} & \bar{\beta}^{k+1} - \beta^{k+1} \end{bmatrix} \quad (4.25)$$

A transformação de similaridade acima, e consequentemente a equação (4.25), são válidas somente se $\beta \neq \bar{\beta}$. Caso contrário, se $\beta = \bar{\beta}$, então a matriz \mathcal{A}^k pode ser obtida como segue:

$$\mathcal{A}^k = \begin{bmatrix} (1-k)\beta^k & k\beta^{k-1} \\ -k\beta^{k+1} & (k+1)\beta^k \end{bmatrix} \quad (4.26)$$

Para calcular o gradiente em (4.22), necessita-se ainda obter as derivadas de matrizes $\partial\mathcal{A}/\partial b$, $\partial\mathcal{A}/\partial c$, $\partial\mathcal{B}/\partial b$ e $\partial\mathcal{B}/\partial c$. Elas são calculadas conforme mostrado a seguir:

$$\begin{aligned} \frac{\partial\mathcal{A}}{\partial b} &= \begin{bmatrix} 0 & 0 \\ 0 & 1-c \end{bmatrix} & \frac{\partial\mathcal{A}}{\partial c} &= \begin{bmatrix} 0 & 0 \\ 1 & -b \end{bmatrix} \\ \frac{\partial\mathcal{B}}{\partial b} &= \begin{bmatrix} 0 \\ \frac{-b\sqrt{1-c^2}}{\sqrt{1-b^2}} \end{bmatrix} & \frac{\partial\mathcal{B}}{\partial c} &= \begin{bmatrix} 0 \\ \frac{-c\sqrt{1-b^2}}{\sqrt{1-c^2}} \end{bmatrix} \end{aligned}$$

Portanto o gradiente $\nabla_{\mathbf{p}}\check{\psi}_2(k) = [\partial\check{\psi}_2(k)/\partial b \quad \partial\check{\psi}_2(k)/\partial c]^T$, que é composto pelas derivadas em (4.22), está agora completamente determinado. A equação (4.22) representa as derivadas da primeira função par de Kautz $\check{\psi}_2(k)$ com relação aos parâmetros de Kautz b e c , onde $\check{\psi}_2(k)$ é a função par de Kautz em (2.26) para $m = 1$.

Resta agora obter expressões equivalentes para os filtros ímpares de Kautz com $m \geq 2$. Voltando ao diagrama de blocos mostrado na Figura 4.4, pode-se escrever:

$$\check{\Psi}_{2m}(z) = \frac{-cz^2 + b(c-1)z + 1}{z^2 + b(c-1)z - c} \check{\Psi}_{2(m-1)}(z) \quad (4.27)$$

No domínio do tempo, tem-se que:

$$\begin{aligned} \check{\psi}_{2m}(k+2) + b(c-1)\check{\psi}_{2m}(k+1) - c\check{\psi}_{2m}(k) = \\ -c\check{\psi}_{2(m-1)}(k+2) + b(c-1)\check{\psi}_{2(m-1)}(k+1) + \check{\psi}_{2(m-1)}(k) \end{aligned} \quad (4.28)$$

Usando um raciocínio análogo ao que foi feito anteriormente para o filtro $\check{\psi}_2(k)$, pode-se reescrever o modelo (4.28) da seguinte forma:

$$\begin{cases} \mathbf{x}_{2m}(k+1) = \mathcal{A}\mathbf{x}_{2m}(k) + \mathcal{B}_1\check{\psi}_{2(m-1)}(k+2) + \mathcal{B}_2\check{\psi}_{2(m-1)}(k+1) \\ \quad + \mathcal{B}_3\check{\psi}_{2(m-1)}(k) \\ \check{\psi}_{2m}(k) = \mathcal{C}\mathbf{x}_{2m}(k) \end{cases} \quad (4.29)$$

onde $\mathbf{x}_{2m}(0) = [\check{\psi}_{2m}(0) \quad \check{\psi}_{2m}(1)]^T$ e as matrizes $\mathcal{A}, \mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3, \mathcal{C}$ são dadas por:

$$\mathcal{A} = \begin{bmatrix} 0 & 1 \\ c & -b(c-1) \end{bmatrix} \quad \mathcal{B}_1 = \begin{bmatrix} 0 \\ -c \end{bmatrix} \quad \mathcal{B}_2 = \begin{bmatrix} 0 \\ b(c-1) \end{bmatrix} \quad (4.30)$$

$$\mathcal{B}_3 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \mathcal{C} = [1 \quad 0] \quad (4.31)$$

Comparando-se o modelo em espaço de estados dado pela equação (4.29) com aquele dado por (4.17), nota-se que o princípio da superposição pode ser utilizado. Isto é feito substituindo o termo da entrada $\mathcal{B}u(k+1)$ em (4.17) individualmente pelos termos $\mathcal{B}_1\check{\psi}_{2(m-1)}(k+2)$, $\mathcal{B}_2\check{\psi}_{2(m-1)}(k+1)$, e $\mathcal{B}_3\check{\psi}_{2(m-1)}(k)$ em (4.29). Desta forma, a solução dada em (4.18) pode ser utilizada para obter a solução de (4.29), chegando-se a:

$$\begin{aligned} \check{\psi}_{2m}(k) = \mathcal{C}\mathcal{A}^k\mathbf{x}_{2m}(0) + \sum_{i=0}^{k-1} \mathcal{C}\mathcal{A}^i \left[\mathcal{B}_1\check{\psi}_{2(m-1)}(k+1-i) + \mathcal{B}_2\check{\psi}_{2(m-1)}(k-i) \right. \\ \left. + \mathcal{B}_3\check{\psi}_{2(m-1)}(k-1-i) \right] \end{aligned} \quad (4.32)$$

Tomando as derivadas de (4.32) em relação aos parâmetros b e c leva à seguinte

equação:

$$\begin{aligned} \frac{\partial \check{\psi}_{2m}(k)}{\partial b} &= c \frac{\partial(\mathcal{A}^k)}{\partial b} \mathbf{x}_{2m}(0) + c \sum_{i=0}^{k-1} \left[\mathcal{A}^i \left(\mathcal{B}_1 \frac{\partial}{\partial b} \check{\psi}_{2(m-1)}(k+1-i) \right. \right. \\ &+ \frac{\partial \mathcal{B}_2}{\partial b} \check{\psi}_{2(m-1)}(k-i) + \mathcal{B}_2 \frac{\partial}{\partial b} \check{\psi}_{2(m-1)}(k-i) + \mathcal{B}_3 \frac{\partial}{\partial b} \check{\psi}_{2(m-1)}(k-1-i) \Big) \\ &+ \frac{\partial(\mathcal{A}^i)}{\partial b} \left(\mathcal{B}_1 \check{\psi}_{2(m-1)}(k+1-i) + \mathcal{B}_2 \check{\psi}_{2(m-1)}(k-i) \right. \\ &\quad \left. \left. + \mathcal{B}_3 \check{\psi}_{2(m-1)}(k-1-i) \right) \right] \end{aligned} \quad (4.33)$$

$$\begin{aligned} \frac{\partial \check{\psi}_{2m}(k)}{\partial c} &= c \frac{\partial(\mathcal{A}^k)}{\partial c} \mathbf{x}_{2m}(0) + c \sum_{i=0}^{k-1} \left[\mathcal{A}^i \left(\mathcal{B}_3 \frac{\partial}{\partial c} \check{\psi}_{2(m-1)}(k-1-i) \right. \right. \\ &+ \frac{\partial \mathcal{B}_1}{\partial c} \check{\psi}_{2(m-1)}(k+1-i) + \mathcal{B}_1 \frac{\partial}{\partial c} \check{\psi}_{2(m-1)}(k+1-i) \\ &+ \frac{\partial \mathcal{B}_2}{\partial c} \check{\psi}_{2(m-1)}(k-i) + \mathcal{B}_2 \frac{\partial}{\partial c} \check{\psi}_{2(m-1)}(k-i) \Big) \\ &+ \frac{\partial(\mathcal{A}^i)}{\partial c} \left(\mathcal{B}_1 \check{\psi}_{2(m-1)}(k+1-i) + \mathcal{B}_2 \check{\psi}_{2(m-1)}(k-i) \right. \\ &\quad \left. \left. + \mathcal{B}_3 \check{\psi}_{2(m-1)}(k-1-i) \right) \right] \end{aligned}$$

onde os termos $\partial(\mathcal{A}^i)/\partial b$, $\partial(\mathcal{A}^i)/\partial c$, e \mathcal{A}^i são dados por (4.23), (4.24) e (4.25), respectivamente. As derivadas $\partial \mathcal{B}_1/\partial c$, $\partial \mathcal{B}_2/\partial b$, e $\partial \mathcal{B}_2/\partial c$ podem ser calculadas a partir de (4.30) como:

$$\frac{\partial \mathcal{B}_1}{\partial c} = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \quad \frac{\partial \mathcal{B}_2}{\partial b} = \begin{bmatrix} 0 \\ c-1 \end{bmatrix} \quad \frac{\partial \mathcal{B}_2}{\partial c} = \begin{bmatrix} 0 \\ b \end{bmatrix}$$

Em resumo, obteve-se nesta seção expressões analíticas para os gradientes das saídas dos filtros de Kautz com relação aos parâmetros de Kautz b e c . Estes gradientes, dados por $\nabla_{\mathbf{p}} \check{\psi}_{2m}(k) = [\partial \check{\psi}_{2m}(k)/\partial b \quad \partial \check{\psi}_{2m}(k)/\partial c]^T$ (para $m = 1, 2, \dots$), são descritos pelas equações (4.22) e (4.33). Apresenta-se a seguir um estudo análogo relacionado com os gradientes das saídas dos filtros ímpares de Kautz.

4.2.2 Funções de Kautz com índice ímpar

Esta seção tem como objetivo apresentar resultados similares aos obtidos na seção anterior, adaptando-os para as funções de Kautz com índice ímpar. Muitos dos desenvolvimentos matemáticos serão omitidos pelo fato de que o raciocínio utilizado é similar àquele já mostrado anteriormente.

Voltando à definição das funções de Kautz na equação (2.26), pode-se escrever a saída do primeiro filtro ímpar $\check{\Psi}_1(z)$ da seguinte forma:

$$\check{\Psi}_1(z) = \frac{\sqrt{1-c^2} z(z-b)}{z^2 + b(c-1)z - c} U(z)$$

No domínio do tempo, tem-se que:

$$\check{\psi}_1(k+2) + b(c-1)\check{\psi}_1(k+1) - c\check{\psi}_1(k) = \sqrt{1-c^2} [u(k+2) - bu(k+1)]$$

Após reescrever a equação a diferenças acima usando uma representação em espaço de estados, pode-se obter sua solução para $\check{\psi}_1(k)$ utilizando o resultado dado na equação (4.18). Tem-se então que:

$$\check{\psi}_1(k) = \mathcal{C}\mathcal{A}^k \mathbf{x}_1(0) + \sum_{i=0}^{k-1} \mathcal{C}\mathcal{A}^i [\mathcal{B}_4 u(k+1-i) + \mathcal{B}_5 u(k-i)] \quad (4.34)$$

onde $\mathbf{x}_1(0) = [\check{\psi}_1(0) \quad \check{\psi}_1(1)]^T$ e as matrizes de estado $\mathcal{A}, \mathcal{B}_4, \mathcal{B}_5, \mathcal{C}$ são dadas por:

$$\begin{aligned} \mathcal{A} &= \begin{bmatrix} 0 & 1 \\ c & -b(c-1) \end{bmatrix} & \mathcal{B}_4 &= \begin{bmatrix} 0 \\ \sqrt{1-c^2} \end{bmatrix} \\ \mathcal{B}_5 &= \begin{bmatrix} 0 \\ -b\sqrt{1-c^2} \end{bmatrix} & \mathcal{C} &= [1 \quad 0] \end{aligned}$$

O gradiente da equação (4.34) resulta em:

$$\begin{aligned} \frac{\partial \check{\psi}_1(k)}{\partial b} &= \mathcal{C} \frac{\partial (\mathcal{A}^k)}{\partial b} \mathbf{x}_1(0) + \mathcal{C} \sum_{i=0}^{k-1} \left[\frac{\partial (\mathcal{A}^i)}{\partial b} (\mathcal{B}_4 u(k+1-i) + \mathcal{B}_5 u(k-i)) \right. \\ &\quad \left. + \mathcal{A}^i \frac{\partial \mathcal{B}_5}{\partial b} u(k-i) \right] \end{aligned} \quad (4.35)$$

$$\begin{aligned} \frac{\partial \check{\psi}_1(k)}{\partial c} &= \mathcal{C} \frac{\partial (\mathcal{A}^k)}{\partial c} \mathbf{x}_1(0) + \mathcal{C} \sum_{i=0}^{k-1} \left[\frac{\partial (\mathcal{A}^i)}{\partial c} (\mathcal{B}_4 u(k+1-i) + \mathcal{B}_5 u(k-i)) \right. \\ &\quad \left. + \mathcal{A}^i \frac{\partial \mathcal{B}_4}{\partial c} u(k+1-i) + \mathcal{A}^i \frac{\partial \mathcal{B}_5}{\partial c} u(k-i) \right] \end{aligned}$$

onde as matrizes $\partial \mathcal{B}_4 / \partial c$, $\partial \mathcal{B}_5 / \partial b$, e $\partial \mathcal{B}_5 / \partial c$ são calculadas da seguinte forma:

$$\begin{aligned} \frac{\partial \mathcal{B}_4}{\partial c} &= \begin{bmatrix} 0 \\ \frac{-c}{\sqrt{1-c^2}} \end{bmatrix} & \frac{\partial \mathcal{B}_5}{\partial b} &= \begin{bmatrix} 0 \\ -\sqrt{1-c^2} \end{bmatrix} & \frac{\partial \mathcal{B}_5}{\partial c} &= \begin{bmatrix} 0 \\ \frac{bc}{\sqrt{1-c^2}} \end{bmatrix} \end{aligned}$$

Resta agora obter uma expressão equivalente para os filtros ímpares de Kautz com $m \geq 2$. De acordo com o diagrama de blocos mostrado na Figura 4.4, pode-se escrever:

$$\check{\Psi}_{2m-1}(z) = \frac{-cz^2 + b(c-1)z + 1}{z^2 + b(c-1)z - c} \check{\Psi}_{2(m-1)-1}(z) \quad (4.36)$$

Ou, equivalentemente, no domínio do tempo:

$$\begin{aligned} \check{\psi}_{2m-1}(k+2) + b(c-1)\check{\psi}_{2m-1}(k+1) - c\check{\psi}_{2m-1}(k) = \\ -c\check{\psi}_{2(m-1)-1}(k+2) + b(c-1)\check{\psi}_{2(m-1)-1}(k+1) + \check{\psi}_{2(m-1)-1}(k) \end{aligned} \quad (4.37)$$

Após reescrever a equação a diferenças (4.37) usando uma representação em espaço de estados, pode-se obter sua solução para $\check{\psi}_{2m-1}(k)$ utilizando o resultado dado na equação (4.18). Logo, tem-se que:

$$\begin{aligned} \check{\psi}_{2m-1}(k) = \mathcal{C}\mathcal{A}^k\mathbf{x}_{2m-1}(0) + \sum_{i=0}^{k-1} \mathcal{C}\mathcal{A}^i \left[\mathcal{B}_1\check{\psi}_{2(m-1)-1}(k+1-i) \right. \\ \left. + \mathcal{B}_2\check{\psi}_{2(m-1)-1}(k-i) + \mathcal{B}_3\check{\psi}_{2(m-1)-1}(k+1-i) \right] \end{aligned} \quad (4.38)$$

onde $\mathbf{x}_{2m-1}(0) = [\check{\psi}_{2m-1}(0) \quad \check{\psi}_{2m-1}(1)]^T$ e as matrizes $\mathcal{A}, \mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3, \mathcal{C}$ são aquelas já apresentadas na seção 4.2.1.

Portanto, as equações analíticas para as derivadas que compõem os gradientes das saídas dos filtros ímpares de Kautz com relação aos parâmetros b e c , para

$m \geq 2$, são calculados a partir de (4.38). Obtém-se assim:

$$\begin{aligned}
 \frac{\partial \check{\psi}_{2m-1}(k)}{\partial b} &= c \frac{\partial(\mathcal{A}^k)}{\partial b} \mathbf{x}_{2m-1}(0) + c \sum_{i=0}^{k-1} \left[\mathcal{A}^i \left(\mathcal{B}_1 \frac{\partial}{\partial b} \check{\psi}_{2(m-1)-1}(k+1-i) \right. \right. \\
 &\quad + \frac{\partial \mathcal{B}_2}{\partial b} \check{\psi}_{2(m-1)-1}(k-i) + \mathcal{B}_2 \frac{\partial}{\partial b} \check{\psi}_{2(m-1)-1}(k-i) \\
 &\quad \left. \left. + \mathcal{B}_3 \frac{\partial}{\partial b} \check{\psi}_{2(m-1)-1}(k-1-i) \right) \right. \\
 &\quad + \frac{\partial(\mathcal{A}^i)}{\partial b} \left(\mathcal{B}_1 \check{\psi}_{2(m-1)-1}(k+1-i) + \mathcal{B}_2 \check{\psi}_{2(m-1)-1}(k-i) \right. \\
 &\quad \left. \left. + \mathcal{B}_3 \check{\psi}_{2(m-1)-1}(k-1-i) \right) \right] \quad (4.39)
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial \check{\psi}_{2m-1}(k)}{\partial c} &= c \frac{\partial(\mathcal{A}^k)}{\partial c} \mathbf{x}_{2m-1}(0) + c \sum_{i=0}^{k-1} \left[\mathcal{A}^i \left(\mathcal{B}_3 \frac{\partial}{\partial c} \check{\psi}_{2(m-1)-1}(k-1-i) \right. \right. \\
 &\quad + \frac{\partial \mathcal{B}_1}{\partial c} \check{\psi}_{2(m-1)-1}(k+1-i) + \mathcal{B}_1 \frac{\partial}{\partial c} \check{\psi}_{2(m-1)-1}(k+1-i) \\
 &\quad + \frac{\partial \mathcal{B}_2}{\partial c} \check{\psi}_{2(m-1)-1}(k-i) + \mathcal{B}_2 \frac{\partial}{\partial c} \check{\psi}_{2(m-1)-1}(k-i) \Big) \\
 &\quad + \frac{\partial(\mathcal{A}^i)}{\partial c} \left(\mathcal{B}_1 \check{\psi}_{2(m-1)-1}(k+1-i) + \mathcal{B}_2 \check{\psi}_{2(m-1)-1}(k-i) \right. \\
 &\quad \left. \left. + \mathcal{B}_3 \check{\psi}_{2(m-1)-1}(k-1-i) \right) \right]
 \end{aligned}$$

4.2.3 Algoritmo

O método proposto pode então ser resumido pelos seguintes passos. Partindo de um vetor inicial de parâmetros $\boldsymbol{\theta}^0$, faça:

1. Calcule o gradiente $\nabla_{\boldsymbol{\theta}}^T J = [\nabla_{\mathbf{p}}^T J \quad \nabla_{\boldsymbol{\alpha}}^T J]$ usando as equações (4.10) e (4.11).
Nota: use as equações (4.22), (4.33), (4.35) e (4.39) para calcular o termo $\nabla_{\mathbf{p}} \check{\psi}_m(k) = [\partial \check{\psi}_m(k)/\partial b \quad \partial \check{\psi}_m(k)/\partial c]^T$ em (4.10);
2. Use $\nabla_{\boldsymbol{\theta}} J$ para atualizar o vetor de parâmetros $\boldsymbol{\theta}$ de acordo com a política de atualização do método de Levenberg-Marquardt. Nota: inclua o intervalo de factibilidade dos parâmetros de Kautz b e c como restrições dentro do modelo de otimização para garantir estabilidade/factibilidade dos pólos;
3. Volte para o passo 1 até que um critério de parada tenha sido atingido.

As restrições mencionadas no passo 2 do algoritmo acima são necessárias para evitar que o processo de otimização produza pólos instáveis, ou seja, fora do círculo unitário. É necessário garantir também que, se o algoritmo produzir pólos com parte

imaginária nula (pólos puramente reais), então os dois pólos devem ser necessariamente iguais. Esta imposição vem da propriedade da base de Kautz, que possui sempre pólos complexos conjugados, por construção.

Para garantir a estabilidade dos pólos de Kautz, muitos trabalhos na literatura (Wahlberg 1994, Ninness e Gustafsson 1997, Nelles 2001, Tanguy *et al.* 2002, Heuberger *et al.* 2005) consideram as restrições sobre os parâmetros de Kautz como aquelas apresentadas no Capítulo 2, i.e., $|b| < 1$ e $|c| < 1$. Porém, uma única condição para garantir a estabilidade dos pólos precisa ser imposta, como deduzido a seguir.

$$\begin{array}{ll}
 \text{Da estabilidade dos pólos:} & 0 < |\beta| < 1 \\
 \text{Elevando-se ao quadrado:} & 0 < |\beta|^2 < 1 \\
 \text{Reescrevendo:} & 0 < \beta\bar{\beta} < 1 \\
 \text{Usando a equação (2.28):} & 0 < -c < 1 \\
 \text{Invertendo-se o sinal da desigualdade:} & -1 < c < 0
 \end{array} \quad (4.40)$$

Para garantir que os pólos de Kautz sejam sempre conjugados, outra restrição necessita ser imposta. Usando as equações (2.27) e (2.28), pode-se reescrever o par de pólos de Kautz $(\beta, \bar{\beta})$ em termos dos parâmetros reais b e c da seguinte forma:

$$(\beta, \bar{\beta}) = \frac{b(1-c)}{2} \pm i\sqrt{\frac{-b^2(1-c)^2}{4} - c} \quad (4.41)$$

Os pólos em (4.41) são complexos conjugados se:

$$\frac{-b^2(1-c)^2}{4} - c \geq 0 \quad (4.42)$$

Caso contrário, os dois pólos de Kautz β e $\bar{\beta}$ em (4.41) seriam reais e distintos, o que não é possível por hipótese.

As restrições sobre os parâmetros dadas pelas inequações (4.40) e (4.42) devem ser incluídas no passo 2 do algoritmo de otimização para garantir estabilidade e factibilidade dos pólos de Kautz, respectivamente.

4.3 Modelo baseado em funções ortonormais generalizadas (GOBF)

As bases de funções de Laguerre e de Kautz são preferíveis quando se deseja modelar sistemas com dinâmica dominante de primeira e segunda ordem, respectivamente. Sistemas que possuem dinâmicas mais complexas são melhor representados por modelos baseados em funções ortonormais generalizadas (GOBF), uma vez que a descrição matemática destas funções envolve pólos associados com múltiplas

dinâmicas (amortecidas e sub-amortecidas). Por esta razão, esta seção tem como objetivo apresentar expressões analíticas para os gradientes das saídas da base GOBF em relação aos seus pólos.

4.3.1 Caso GOBF com pólos reais

Inicialmente, considere a base de funções ortonormais generalizadas na equação (2.23) parametrizadas com pólos reais e distintos entre si, i.e., $\beta_m = \bar{\beta}_m \triangleq p_m \in \mathbb{R}$. Neste caso, as funções de transferência dos filtros são dadas como a seguir:

$$\Psi_m(z) = \frac{z\sqrt{1-p_m^2}}{z-p_m} \prod_{j=1}^{m-1} \left(\frac{1-p_j z}{z-p_j} \right) \quad (4.43)$$

Um diagrama de blocos de um sistema linear baseado nestas funções pode ser construído a partir de (4.43). A Figura 4.5 ilustra este diagrama, onde $\check{\psi}_m(k)$ é a resposta do filtro $\Psi_m(z)$ à entrada $u(k)$, de acordo com as equações (4.6) e (4.7).

Considere um modelo contendo M funções da base ortonormal $\{\psi_m\}$ em (4.43). Todos os parâmetros livres desta base a serem otimizados podem ser dispostos em um único vetor contendo todos os pólos da base, i.e., $\mathbf{p} = [p_1 \ p_2 \ \cdots \ p_M]^T$. Deve-se notar que a m -ésima função da base depende apenas dos m primeiros pólos $\{p_1, \dots, p_m\}$, e portanto não depende dos demais pólos $\{p_{m+1}, \dots, p_M\}$. Levando em consideração que $\partial\check{\psi}_m(k)/\partial p_l = 0$ para $l > m$, o gradiente $\nabla_{\mathbf{p}}\check{\psi}_m(k)$ é dado por:

$$\begin{aligned} \nabla_{\mathbf{p}}\check{\psi}_m(k) &= \begin{bmatrix} \frac{\partial\check{\psi}_m(k)}{\partial p_1} & \cdots & \frac{\partial\check{\psi}_m(k)}{\partial p_m} & \cdots & \frac{\partial\check{\psi}_m(k)}{\partial p_M} \end{bmatrix}^T \\ &= \begin{bmatrix} \frac{\partial\check{\psi}_m(k)}{\partial p_1} & \cdots & \frac{\partial\check{\psi}_m(k)}{\partial p_m} & \underbrace{0 \ \cdots \ 0}_{(M-m) \text{ vezes}} \end{bmatrix}^T \end{aligned} \quad (4.44)$$

O procedimento para calcular as componentes do gradiente (4.44) segue as mesmas linhas de raciocínio utilizada para a obtenção dos gradientes para a base de Kautz. Detalhes deste procedimento são apresentados a seguir.

A saída do primeiro filtro GOBF ($m = 1$ em (4.43)) pode ser escrita a partir da Figura 4.5 da seguinte forma:

$$\check{\Psi}_1(z) = \frac{z\sqrt{1-p_1^2}}{z-p_1} U(z)$$

ou equivalentemente no domínio do tempo:

$$\check{\psi}_1(k+1) - p_1\check{\psi}_1(k) = \sqrt{1-p_1^2} u(k+1) \quad (4.45)$$

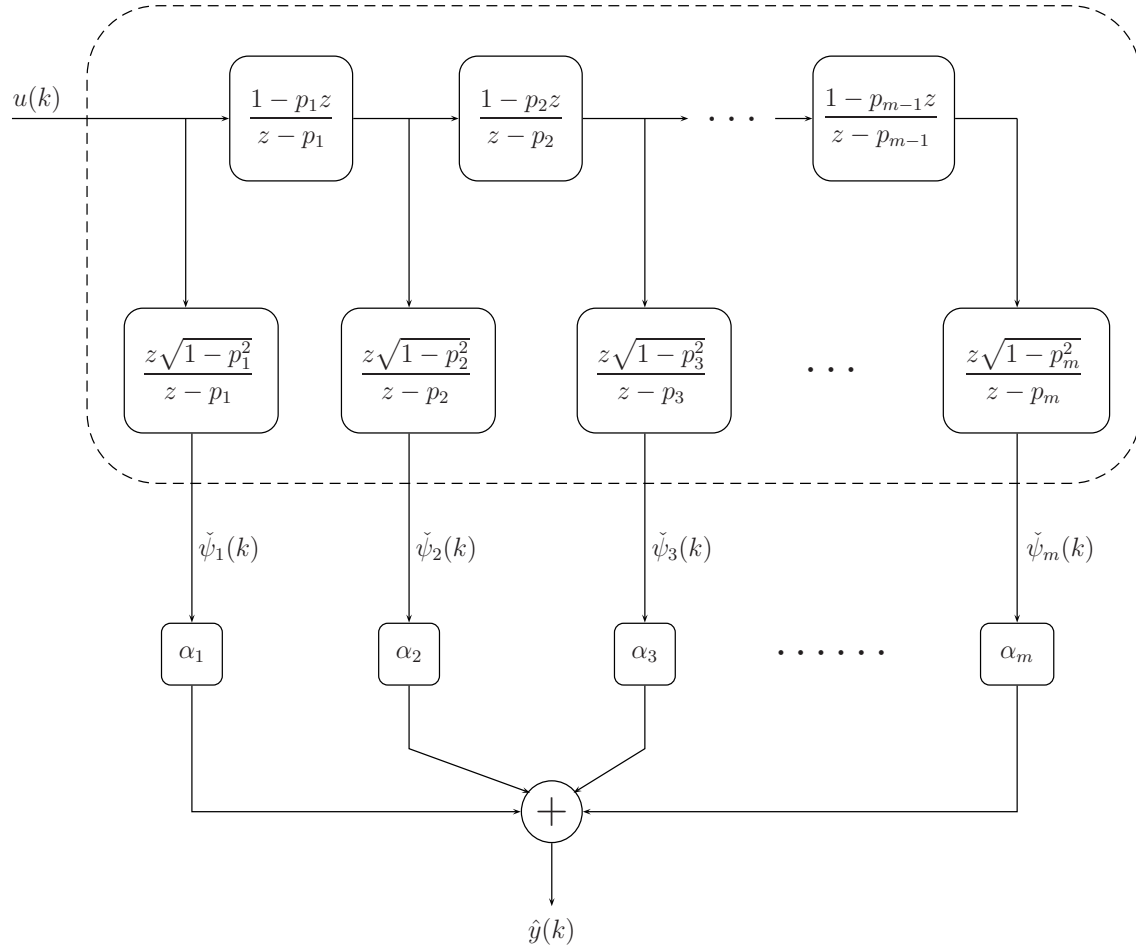


Figura 4.5: Diagrama de blocos de um modelo dinâmico baseado em funções GOBF com pólos reais.

A solução de (4.45) para $\check{\psi}_1(k)$ é facilmente obtida como:

$$\check{\psi}_1(k) = p_1^k \check{\psi}_1(0) + \sqrt{1 - p_1^2} \sum_{i=0}^{k-1} p_1^i u(k - i) \quad (4.46)$$

e a derivada da saída do primeiro filtro GOBF em (4.46) com relação ao pólo p_1 é dada por:

$$\frac{\partial \check{\psi}_1(k)}{\partial p_1} = k p_1^{k-1} \check{\psi}_1(0) + \frac{1}{\sqrt{1 - p_1^2}} \sum_{i=0}^{k-1} [i p_1^{i-1} - (i + 1) p_1^{i+1}] u(k - i) \quad (4.47)$$

Como $\check{\psi}_1(k)$ depende somente do pólo p_1 (e não dos demais pólos p_2, \dots, p_M , então resulta que $\partial \check{\psi}_1(k) / \partial p_l = 0$ para $l = 2, \dots, M$.

Para $m \geq 2$, tem-se a partir das equações (4.43) e (4.7) que:

$$\check{\Psi}_m(z) = \sqrt{\frac{1-p_m^2}{1-p_{m-1}^2}} \left(\frac{1-p_{m-1}z}{z-p_m} \right) \check{\Psi}_{m-1}(z)$$

Equivalentemente, no domínio do tempo tem-se a seguinte equação a diferenças:

$$\check{\psi}_m(k+1) - p_m \check{\psi}_m(k) = \sqrt{\frac{1-p_m^2}{1-p_{m-1}^2}} \left[\check{\psi}_{m-1}(k) - p_{m-1} \check{\psi}_{m-1}(k+1) \right] \quad (4.48)$$

cujá solução para $\check{\psi}_m(k)$ é dada por:

$$\begin{aligned} \check{\psi}_m(k) &= p_m^k \check{\psi}_m(0) + \\ &+ \sqrt{\frac{1-p_m^2}{1-p_{m-1}^2}} \sum_{i=0}^{k-1} p_m^i \left[\check{\psi}_{m-1}(k-1-i) - p_{m-1} \check{\psi}_{m-1}(k-i) \right] \end{aligned} \quad (4.49)$$

Para obter o gradiente $\nabla_{\mathbf{p}} \check{\psi}_m(k)$ em (4.44), é necessário calcular as derivadas de $\check{\psi}_m(k)$ em (4.49) com relação a p_l para $l = 1, \dots, m$. A derivada de (4.49) com relação ao pólo p_m é dado por:

$$\begin{aligned} \frac{\partial \check{\psi}_m(k)}{\partial p_m} &= k p_m^{k-1} \check{\psi}_m(0) + \frac{1}{\sqrt{(1-p_m^2)(1-p_{m-1}^2)}} \cdot \\ &\cdot \sum_{i=0}^{k-1} \left[i p_m^{i-1} - (i+1) p_m^{i+1} \right] \cdot \left[\check{\psi}_{m-1}(k-1-i) - p_{m-1} \check{\psi}_{m-1}(k-i) \right] \end{aligned} \quad (4.50)$$

onde foi usado o fato de que $\partial \check{\psi}_{m-1}(k)/\partial p_m = 0$, pois a função $\check{\psi}_{m-1}(k)$ depende somente do conjunto de pólos $\{p_1, p_2, \dots, p_{m-1}\}$.

A derivada de (4.49) em relação a p_{m-1} é calculada como segue:

$$\begin{aligned} \frac{\partial \check{\psi}_m(k)}{\partial p_{m-1}} &= \sum_{i=0}^{k-1} \left[\sqrt{\frac{1-p_m^2}{(1-p_{m-1}^2)^3}} p_{m-1} p_m^i \left(\check{\psi}_{m-1}(k-1-i) - p_{m-1} \check{\psi}_{m-1}(k-i) \right) \right. \\ &+ \sqrt{\frac{1-p_m^2}{1-p_{m-1}^2}} p_m^i \left(\frac{\partial}{\partial p_{m-1}} \check{\psi}_{m-1}(k-1-i) - \check{\psi}_{m-1}(k-1-i) \right. \\ &\left. \left. - p_{m-1} \frac{\partial}{\partial p_{m-1}} \check{\psi}_{m-1}(k-i) \right) \right] \end{aligned} \quad (4.51)$$

Finalmente, as derivadas de (4.49) com relação a p_l , $l = 1, \dots, m-2$, podem ser calculadas usando a seguinte equação:

$$\frac{\partial \check{\psi}_m(k)}{\partial p_l} = \sqrt{\frac{1-p_m^2}{1-p_{m-1}^2}} \sum_{i=0}^{k-1} p_m^i \left[\frac{\partial}{\partial p_l} \check{\psi}_{m-1}(k-1-i) - p_{m-1} \frac{\partial}{\partial p_l} \check{\psi}_{m-1}(k-i) \right] \quad (4.52)$$

Obteve-se assim equações analíticas que permitem obter o gradiente da função $\check{\psi}_m(k)$ com relação ao conjunto de pólos $\{p_1, p_2, \dots, p_M\}$, dadas pelas equações (4.50), (4.51) e (4.52). Os passos do método proposto, quando a base GOBF é usada no modelo (4.8), são essencialmente os mesmos do que aqueles descritos na seção 4.2.3. A única diferença está relacionada com os gradientes $\nabla_{\mathbf{p}} \check{\psi}_m(k)$, que agora são calculados conforme descrito acima.

É importante ressaltar que a formulação apresentada nesta seção foi desenvolvida considerando que cada pólo da base GOBF é independente dos demais. Isto significa que o número de pólos a serem otimizados coincide com o número de funções da base, i.e. M . Na prática, porém, pode-se exigir que somente um número limitado de pólos independentes seja repetido múltiplas vezes, de tal forma que o número M de funções seja um múltiplo do número de pólos a serem otimizados. A formulação relativa a esse caso particular pode ser obtida de maneira similar ao que foi descrito anteriormente.

O desenvolvimento apresentado nesta seção foi obtido assumindo que as funções GOBF são parametrizadas somente por pólos reais. O caso em que a base ortonormal possui pólos complexos será estudado na próxima seção.

4.3.2 Caso GOBF com um par de pólos complexos

Conforme foi discutido na seção 2.3 do Capítulo 2, em geral as funções ortonormais generalizadas definidas em (2.23) possuem transformada Z inversa que possui coeficientes complexos. Esta desvantagem pode ser superada construindo-se uma nova base de funções ortonormais com respostas ao impulso reais (Ninness e Gustafsson 1997). Parte do desenvolvimento matemático mostrando como esta nova base é construída será reproduzido a seguir.

Considere então um modelo baseado em funções ortonormais em que as $M - 1$ primeiras funções da base $\{\Psi_1(z), \Psi_2(z), \dots, \Psi_{M-1}(z)\}$ são parametrizadas por seus respectivos $M - 1$ pólos reais, dados no conjunto $\{p_1, p_2, \dots, p_{M-1}\}$, assim como na equação (4.43). Suponha que se deseja incluir um pólo complexo β_M nesse conjunto de pólos. De acordo com (Ninness e Gustafsson 1997), duas novas funções $\Psi'_M(z)$ e $\Psi''_M(z)$ com resposta ao impulso reais são construídas a partir de uma combinação linear de $\Psi_M(z)$ e $\Psi_{M+1}(z)$ geradas por (2.23). Neste caso, o conjunto de funções será dado por $\{\Psi_1(z), \dots, \Psi_{M-1}(z), \Psi'_M(z), \Psi''_M(z)\}$, com o correspondente conjunto de pólos $\{p_1, \dots, p_{M-1}, \beta_M, \bar{\beta}_M\}$.

A combinação linear que gera as funções reais é dada como a seguir:

$$\begin{bmatrix} \Psi'_M(z) \\ \Psi''_M(z) \end{bmatrix} = \begin{bmatrix} \kappa'_0 & \kappa'_1 \\ \kappa''_0 & \kappa''_1 \end{bmatrix} \cdot \begin{bmatrix} \Psi_M(z) \\ \Psi_{M+1}(z) \end{bmatrix} \quad (4.53)$$

onde $\kappa'_0, \kappa'_1, \kappa''_0, \kappa''_1$ são escalares complexos arbitrários. Estes escalares podem ser feitos relacionados entre si para assegurar a propriedade de ortonormalidade das novas funções geradas.

De acordo com a equação (2.23), pode-se também escrever o seguinte:

$$\Psi_M(z) = \sqrt{\frac{1 - |\beta_M|^2}{1 - |\beta_{M-1}|^2}} \left(\frac{1 - \bar{\beta}_{M-1}z}{z - \beta_M} \right) \Psi_{M-1}(z) \quad (4.54)$$

Substitui-se então a equação (4.54) em (4.53) e, considerando que pólos complexos são escolhidos aos pares conjugados, após algumas manipulações algébricas, chega-se que a descrição matemática da base GOBF quando um pólo β_m é incorporado no modelo é dada por (Ninness e Gustafsson 1997):

$$\Psi'_M(z) = \frac{z\sqrt{1 - |\beta_M|^2} (\lambda'z + \gamma')}{z^2 - (\beta_M + \bar{\beta}_M)z + |\beta_M|^2} \prod_{j=1}^{M-1} \left(\frac{1 - p_jz}{z - p_j} \right) \quad (4.55)$$

$$\Psi''_M(z) = \frac{z\sqrt{1 - |\beta_M|^2} (\lambda''z + \gamma'')}{z^2 - (\beta_M + \bar{\beta}_M)z + |\beta_M|^2} \prod_{j=1}^{M-1} \left(\frac{1 - p_jz}{z - p_j} \right)$$

onde $\beta_M, \bar{\beta}_M \in \mathbb{C}$ é um par de pólos complexos conjugados incorporado na base ortonormal GOBF, e $\lambda', \gamma', \lambda'', \gamma''$ são parâmetros reais que estão relacionados com β_M da seguinte forma:

$$\begin{bmatrix} \lambda' & \gamma' \end{bmatrix} \begin{bmatrix} 1 + |\beta_M|^2 & \beta_M + \bar{\beta}_M \\ \beta_M + \bar{\beta}_M & 1 + |\beta_M|^2 \end{bmatrix} \begin{bmatrix} \lambda' \\ \gamma' \end{bmatrix} = |1 - \beta_M^2|^2 \quad (4.56)$$

$$\begin{bmatrix} \lambda'' \\ \gamma'' \end{bmatrix} = \frac{1}{\sqrt{1 - \xi^2}} \begin{bmatrix} \xi & 1 \\ -1 & -\xi \end{bmatrix} \begin{bmatrix} \lambda' \\ \gamma' \end{bmatrix} \quad (4.57)$$

$$\xi = \frac{\beta_M + \bar{\beta}_M}{1 + |\beta_M|^2} \quad (4.58)$$

Existem infinitas soluções para a escolha dos parâmetros reais $\lambda', \gamma', \lambda'', \gamma''$ que satisfazem as equações (4.56), (4.57) e (4.58). A princípio, qualquer escolha destes parâmetros pode ser adotada, pois todos valores satisfazendo (4.56)–(4.58) garantem a ortonormalidade da base (Ninness e Gustafsson 1997, Ziaei e Wang 2006).

O cálculo dos gradientes das funções GOBF que contemplam pólos complexos pode então ser obido usando-se a representação dada na equação (4.55). Assim, de acordo com as equações (4.7), (4.43) e (4.55), a seguinte relação pode ser deduzida:

$$\check{\Psi}'_M(z) = \sqrt{\frac{1 - |\beta_M|^2}{1 - p_{M-1}^2}} \cdot \frac{(\lambda'z + \gamma')(1 - p_{M-1}z)}{z^2 - (\beta_M + \bar{\beta}_M)z + |\beta_M|^2} \cdot \check{\Psi}_{M-1}(z) \quad (4.59)$$

Equivalentemente, no domínio do tempo:

$$\begin{aligned} \check{\psi}'_M(k+2) - (\beta_M + \bar{\beta}_M)\check{\psi}'_M(k+1) + |\beta_M|^2\check{\psi}'_M(k) = \\ \sqrt{\frac{1-|\beta_M|^2}{1-p_{M-1}^2}} \cdot \left[-\lambda' p_{M-1}\check{\psi}_{M-1}(k+2) + (\lambda' - \gamma' p_{M-1})\check{\psi}_{M-1}(k+1) + \gamma'\check{\psi}_{M-1}(k) \right] \end{aligned} \quad (4.60)$$

Para resolver a equação a diferenças em (4.60), ela pode ser reescrita usando uma descrição em espaço de estados. Adotando-se uma estratégia semelhante àquela apresentada na seção 4.2 para o caso Kautz, a representação da equação (4.60) em espaço de estados resulta em:

$$\begin{cases} \mathbf{x}'_M(k+1) = \mathcal{A}_1 \mathbf{x}'_M(k) + \mathcal{B}_6 \check{\psi}_{M-1}(k+2) + \mathcal{B}_7 \check{\psi}_{M-1}(k+1) + \mathcal{B}_8 \check{\psi}_{M-1}(k) \\ \check{\psi}'_M(k) = \mathcal{C} \mathbf{x}'_M(k) \end{cases} \quad (4.61)$$

cuja solução para $\check{\psi}'_M(k)$ é dada por:

$$\begin{aligned} \check{\psi}'_M(k) = \mathcal{C} \mathcal{A}_1^k \mathbf{x}'_M(0) + \mathcal{C} \sum_{i=0}^{k-1} \mathcal{A}_1^i \left[\mathcal{B}_6 \check{\psi}_{M-1}(k+1-i) \right. \\ \left. + \mathcal{B}_7 \check{\psi}_{M-1}(k-i) + \mathcal{B}_8 \check{\psi}_{M-1}(k-1-i) \right] \end{aligned} \quad (4.62)$$

onde $\mathbf{x}'_M(0) = [\check{\psi}'_M(0) \quad \check{\psi}'_M(1)]^T$, e as matrizes de estado $\mathcal{A}_1, \mathcal{B}_6, \mathcal{B}_7, \mathcal{B}_8, \mathcal{C}$ são como segue:

$$\begin{aligned} \mathcal{A}_1 = \begin{bmatrix} 0 & 1 \\ -|\beta_M|^2 & \beta_M + \bar{\beta}_M \end{bmatrix} \quad \mathcal{B}_6 = K_M \begin{bmatrix} 0 \\ -\lambda' p_{M-1} \end{bmatrix} \\ \mathcal{B}_7 = K_M \begin{bmatrix} 0 \\ \lambda' - \gamma' p_{M-1} \end{bmatrix} \quad \mathcal{B}_8 = K_M \begin{bmatrix} 0 \\ \gamma' \end{bmatrix} \\ \mathcal{C} = \begin{bmatrix} 1 & 0 \end{bmatrix} \quad K_M = \sqrt{\frac{1-|\beta_M|^2}{1-p_{M-1}^2}} \end{aligned}$$

Para calcular o gradiente de (4.62) em relação a β_M , este pólo pode ser considerado meramente como um vetor bidimensional real, i.e. $\beta_M \triangleq [\mathbf{Re}(\beta_M) \quad \mathbf{Im}(\beta_M)]^T$. Neste caso, o gradiente é definido como $\nabla_{\beta_M} \check{\psi}'_M(k) \triangleq [\partial \check{\psi}'_M(k) / \partial \mathbf{Re}(\beta_M) \quad \partial \check{\psi}'_M(k) / \partial \mathbf{Im}(\beta_M)]^T$, cujos elementos são dados por:

$$\begin{aligned} \frac{\partial \check{\psi}'_M(k)}{\partial \mathbf{Re}(\beta_M)} = \mathcal{C} \frac{\partial \mathcal{A}_1^k}{\partial \mathbf{Re}(\beta_M)} \mathbf{x}'_M(0) + \mathcal{C} \sum_{i=0}^{k-1} \frac{\partial \mathcal{A}_1^i}{\partial \mathbf{Re}(\beta_M)} \left[\mathcal{B}_6 \check{\psi}_{M-1}(k+1-i) \right. \\ \left. + \mathcal{B}_7 \check{\psi}_{M-1}(k-i) + \mathcal{B}_8 \check{\psi}_{M-1}(k-1-i) \right] \\ + \mathcal{A}_1^i \left[\frac{\partial \mathcal{B}_6}{\partial \mathbf{Re}(\beta_M)} \check{\psi}_{M-1}(k+1-i) + \frac{\partial \mathcal{B}_7}{\partial \mathbf{Re}(\beta_M)} \check{\psi}_{M-1}(k-i) + \frac{\partial \mathcal{B}_8}{\partial \mathbf{Re}(\beta_M)} \check{\psi}_{M-1}(k-1-i) \right] \end{aligned} \quad (4.63)$$

$$\begin{aligned}
 \frac{\partial \check{\psi}'_M(k)}{\partial \mathbf{Im}(\beta_M)} &= c \frac{\partial \mathcal{A}_1^k}{\partial \mathbf{Im}(\beta_M)} \mathbf{x}'_M(0) + c \sum_{i=0}^{k-1} \frac{\partial \mathcal{A}_1^i}{\partial \mathbf{Im}(\beta_M)} \left[\mathcal{B}_6 \check{\psi}_{M-1}(k+1-i) \right. \\
 &\quad \left. + \mathcal{B}_7 \check{\psi}_{M-1}(k-i) + \mathcal{B}_8 \check{\psi}_{M-1}(k-1-i) \right] \\
 &+ \mathcal{A}_1^i \left[\frac{\partial \mathcal{B}_6}{\partial \mathbf{Im}(\beta_M)} \check{\psi}_{M-1}(k+1-i) + \frac{\partial \mathcal{B}_7}{\partial \mathbf{Im}(\beta_M)} \check{\psi}_{M-1}(k-i) + \frac{\partial \mathcal{B}_8}{\partial \mathbf{Im}(\beta_M)} \check{\psi}_{M-1}(k-1-i) \right]
 \end{aligned} \tag{4.64}$$

Finalmente, as derivadas de (4.62) com relação a p_l ($l = 1, \dots, M-1$) podem ser calculadas usando a seguinte equação:

$$\begin{aligned}
 \frac{\partial \check{\psi}'_M(k)}{\partial p_l} &= c \sum_{i=0}^{k-1} \mathcal{A}_1^i \left[\mathcal{B}_6 \frac{\partial \check{\psi}_{M-1}(k+1-i)}{\partial p_l} + \frac{\partial \mathcal{B}_6}{\partial p_l} \check{\psi}_{M-1}(k+1-i) \right. \\
 &\quad + \mathcal{B}_7 \frac{\partial \check{\psi}_{M-1}(k-i)}{\partial p_l} + \frac{\partial \mathcal{B}_7}{\partial p_l} \check{\psi}_{M-1}(k-i) \\
 &\quad \left. + \mathcal{B}_8 \frac{\partial \check{\psi}_{M-1}(k-1-i)}{\partial p_l} + \frac{\partial \mathcal{B}_8}{\partial p_l} \check{\psi}_{M-1}(k-1-i) \right]
 \end{aligned} \tag{4.65}$$

onde as derivadas das matrizes $\mathcal{B}_6, \mathcal{B}_7, \mathcal{B}_8$ são como segue:

$$\begin{aligned}
 \frac{\partial \mathcal{B}_6}{\partial p_{M-1}} &= \begin{bmatrix} 0 \\ -\frac{\lambda \sqrt{1-|\beta_{M-1}|^2}}{(1-p_{M-1}^2)^{3/2}} \end{bmatrix} & \frac{\partial \mathcal{B}_7}{\partial p_{M-1}} &= \begin{bmatrix} 0 \\ \frac{(\lambda' p_{M-1} - \gamma') \sqrt{1-|\beta_{M-1}|^2}}{(1-p_{M-1}^2)^{3/2}} \end{bmatrix} \\
 \frac{\partial \mathcal{B}_8}{\partial p_{M-1}} &= \begin{bmatrix} 0 \\ \frac{\lambda' p_{M-1} \sqrt{1-|\beta_{M-1}|^2}}{(1-p_{M-1}^2)^{3/2}} \end{bmatrix}
 \end{aligned}$$

As derivadas $\partial \mathcal{B}_6 / \partial p_l$, $\partial \mathcal{B}_7 / \partial p_l$, e $\partial \mathcal{B}_8 / \partial p_l$ resultam em matrizes nulas se $l \neq M-1$, pois as matrizes $\mathcal{B}_6, \mathcal{B}_7, \mathcal{B}_8$ não dependem de p_l para $l \neq M-1$.

O desenvolvimento matemático necessário para obter as derivadas das funções complementares $\check{\Psi}''_M(z)$ é análogo ao apresentado acima, bastando substituir λ' e γ' por λ'' e γ'' , respectivamente.

4.3.3 Caso GOBF com dois pares de pólos complexos

A ideia do desenvolvimento apresentado a seguir é a construção de novas funções ortonormais com resposta ao impulso reais a partir das funções definidas em (2.23). Este desenvolvimento foi apresentado originalmente em Ninness e Gustafsson (1997) com a inclusão de um par de pólos complexos à uma base com pólos reais. Conforme

sintetizado na seção 4.3.2, partindo-se de uma base ortonormal com $M - 1$ pólos reais, $\{\Psi_1(z), \Psi_2(z), \dots, \Psi_{M-1}(z)\}$, obtém-se duas novas funções ortonormais $\Psi'_M(z)$ e $\Psi''_M(z)$, com pólos $\beta_M, \bar{\beta}_M$.

Se for desejado incluir mais um par de pólos complexos $\beta_{M+1}, \bar{\beta}_{M+1}$ na base GOBF, mais duas funções $\Psi'_{M+1}(z)$ e $\Psi''_{M+1}(z)$ com respostas ao impulso reais são construídas como combinações lineares de $\Psi_{M+2}(z)$ e $\Psi_{M+3}(z)$ em (2.23), como sugerido por (Ninness e Gustafsson 1997). Neste caso, o novo conjunto de funções será $\{\Psi_1(z), \dots, \Psi_{M-1}(z), \Psi'_M(z), \Psi''_M(z), \Psi'_{M+1}(z), \Psi''_{M+1}(z)\}$, com o correspondente conjunto de pólos $\{p_1, \dots, p_{M-1}, \beta_M, \bar{\beta}_M, \beta_{M+1}, \bar{\beta}_{M+1}\}$. Após alguma manipulação algébrica, tem-se que as duas novas funções geradas são dadas por:

$$\Psi'_{M+1}(z) = \frac{z\sqrt{1-|\beta_{M+1}|^2}(\rho'z + \mu')}{z^2 - (\beta_{M+1} + \bar{\beta}_{M+1})z + |\beta_{M+1}|^2} \cdot \left(\frac{1 - \bar{\beta}_M z}{z - \beta_M}\right) \cdot \left(\frac{1 - \beta_M z}{z - \bar{\beta}_M}\right) \prod_{j=1}^{M-1} \left(\frac{1 - p_j z}{z - p_j}\right) \quad (4.66)$$

$$\Psi''_{M+1}(z) = \frac{z\sqrt{1-|\beta_{M+1}|^2}(\rho''z + \mu'')}{z^2 - (\beta_{M+1} + \bar{\beta}_{M+1})z + |\beta_{M+1}|^2} \cdot \left(\frac{1 - \bar{\beta}_M z}{z - \beta_M}\right) \cdot \left(\frac{1 - \beta_M z}{z - \bar{\beta}_M}\right) \prod_{j=1}^{M-1} \left(\frac{1 - p_j z}{z - p_j}\right)$$

onde $\rho', \mu', \rho'', \mu''$ são parâmetros reais que estão relacionados com o pólo β_{M+1} como segue:

$$\begin{bmatrix} \rho' & \mu' \end{bmatrix} \begin{bmatrix} 1 + |\beta_{M+1}|^2 & \beta_{M+1} + \bar{\beta}_{M+1} \\ \beta_{M+1} + \bar{\beta}_{M+1} & 1 + |\beta_{M+1}|^2 \end{bmatrix} \begin{bmatrix} \rho' \\ \mu' \end{bmatrix} = |1 - \beta_{M+1}^2|^2 \quad (4.67)$$

$$\begin{bmatrix} \rho'' \\ \mu'' \end{bmatrix} = \frac{1}{\sqrt{1 - \xi^2}} \begin{bmatrix} \xi & 1 \\ -1 & -\xi \end{bmatrix} \begin{bmatrix} \rho' \\ \mu' \end{bmatrix} \quad (4.68)$$

$$\xi = \frac{\beta_{M+1} + \bar{\beta}_{M+1}}{1 + |\beta_{M+1}|^2} \quad (4.69)$$

Assim como já foi discutido na seção 4.3.2 no que diz respeito à base GOBF com um par de pólos complexos, existem também infinitas soluções para a escolha dos parâmetros reais $\rho', \mu', \rho'', \mu''$ satisfazendo as equações (4.67)–(4.69) para garantir a ortonormalidade da base.

A seguir, apresenta-se o desenvolvimento utilizado para obter os gradientes desta base com relação aos seus pólos. De acordo com as equações (4.7), (4.55) e (4.66), a seguinte relação pode ser deduzida:

$$\check{\Psi}'_{M+1}(z) = \sqrt{\frac{1 - |\beta_{M+1}|^2}{1 - |\beta_M|^2}} \cdot \left(\frac{\rho'z + \mu'}{\lambda'z + \gamma'}\right) \cdot \left(\frac{1 - (\beta_M + \bar{\beta}_M)z + |\beta_M|^2 z^2}{z^2 - (\beta_{M+1} + \bar{\beta}_{M+1})z + |\beta_{M+1}|^2}\right) \check{\Psi}'_M(z) \quad (4.70)$$

Com a inclusão do par de pólos $\beta_{M+1}, \bar{\beta}_{M+1} \in \mathbb{C}$, o vetor de pólos agora é dado por $\mathbf{p} = [p_1 \ \cdots \ p_{M-1} \ \beta_M \ \bar{\beta}_M \ \beta_{M+1} \ \bar{\beta}_{M+1}]^T$. Neste caso, os elementos do gradiente $\nabla_{\mathbf{p}} \check{\psi}'_{M+1}(k)$ são como segue:

$$\begin{aligned} \frac{\partial \check{\psi}'_{M+1}(k)}{\partial p_l} = & \ c_1 \sum_{i=0}^{k-1} \mathcal{A}_2^i \left[\mathcal{B}_9 \frac{\partial \check{\psi}'_M(k+2-i)}{\partial p_l} + \mathcal{B}_{10} \frac{\partial \check{\psi}'_M(k+1-i)}{\partial p_l} + \mathcal{B}_{11} \frac{\partial \check{\psi}'_M(k-i)}{\partial p_l} \right. \\ & \left. + \mathcal{B}_{12} \frac{\partial \check{\psi}'_M(k-1-i)}{\partial p_l} \right] \quad l = 1, \dots, M-1 \end{aligned} \quad (4.74)$$

$$\begin{aligned} \frac{\partial \check{\psi}'_{M+1}(k)}{\partial \mathbf{Re}(\beta_M)} = & \ c_1 \frac{\partial(\mathcal{A}_2^k)}{\partial \mathbf{Re}(\beta_M)} \mathbf{x}'_{M+1}(0) + c_1 \sum_{i=0}^{k-1} \frac{\partial(\mathcal{A}_2^i)}{\partial \mathbf{Re}(\beta_M)} \left[\mathcal{B}_9 \check{\psi}'_M(k+2-i) \right. \\ & + \mathcal{B}_{10} \check{\psi}'_M(k+1-i) + \mathcal{B}_{11} \check{\psi}'_M(k-i) + \mathcal{B}_{12} \check{\psi}'_M(k-1-i) \Big] \\ & + \mathcal{A}_2^i \left[\frac{\partial \mathcal{B}_9}{\partial \mathbf{Re}(\beta_M)} \check{\psi}'_M(k+2-i) + \mathcal{B}_9 \frac{\partial \check{\psi}'_M(k+2-i)}{\partial \mathbf{Re}(\beta_M)} \right. \\ & + \frac{\partial \mathcal{B}_{10}}{\partial \mathbf{Re}(\beta_M)} \check{\psi}'_M(k+1-i) + \mathcal{B}_{10} \frac{\partial \check{\psi}'_M(k+1-i)}{\partial \mathbf{Re}(\beta_M)} \\ & + \frac{\partial \mathcal{B}_{11}}{\partial \mathbf{Re}(\beta_M)} \check{\psi}'_M(k-i) + \mathcal{B}_{11} \frac{\partial \check{\psi}'_M(k-i)}{\partial \mathbf{Re}(\beta_M)} \\ & \left. + \frac{\partial \mathcal{B}_{12}}{\partial \mathbf{Re}(\beta_M)} \check{\psi}'_M(k-1-i) + \mathcal{B}_{12} \frac{\partial \check{\psi}'_M(k-1-i)}{\partial \mathbf{Re}(\beta_M)} \right] \end{aligned} \quad (4.75)$$

$$\begin{aligned} \frac{\partial \check{\psi}'_{M+1}(k)}{\partial \mathbf{Re}(\beta_{M+1})} = & \ c_1 \frac{\partial(\mathcal{A}_2^k)}{\partial \mathbf{Re}(\beta_{M+1})} \mathbf{x}'_{M+1}(0) + c_1 \sum_{i=0}^{k-1} \frac{\partial(\mathcal{A}_2^i)}{\partial \mathbf{Re}(\beta_{M+1})} \left[\mathcal{B}_9 \check{\psi}'_M(k+2-i) \right. \\ & + \mathcal{B}_{10} \check{\psi}'_M(k+1-i) + \mathcal{B}_{11} \check{\psi}'_M(k-i) + \mathcal{B}_{12} \check{\psi}'_M(k-1-i) \Big] \\ & + \mathcal{A}_2^i \left[\frac{\partial \mathcal{B}_9}{\partial \mathbf{Re}(\beta_{M+1})} \check{\psi}'_M(k+2-i) + \frac{\partial \mathcal{B}_{10}}{\partial \mathbf{Re}(\beta_{M+1})} \check{\psi}'_M(k+1-i) \right. \\ & \left. + \frac{\partial \mathcal{B}_{11}}{\partial \mathbf{Re}(\beta_{M+1})} \check{\psi}'_M(k-i) + \frac{\partial \mathcal{B}_{12}}{\partial \mathbf{Re}(\beta_{M+1})} \check{\psi}'_M(k-1-i) \right] \end{aligned} \quad (4.76)$$

As derivadas com relação a parte imaginária dos pólos β_M e β_{M+1} são obtidas a partir de (4.75) e (4.76), substituindo-se $\mathbf{Re}(\beta_M)$ e $\mathbf{Re}(\beta_{M+1})$ por $\mathbf{Im}(\beta_M)$ e $\mathbf{Im}(\beta_{M+1})$, respectivamente.

4.3.4 Caso GOBF com qualquer número de pares de pólos complexos

Para a inclusão de qualquer número de pólos complexos, repete-se o procedimento apresentado nas seções 4.3.2 e 4.3.3. Suponha uma base de funções ortonormais com $r - 1$ pólos reais e c pares de pólos complexos. Como já foi estudado anteriormente, as funções com pólos reais são dadas pela equação (4.43). Generalizando o método de construção apresentado em Ninness e Gustafsson (1997) e nas seções anteriores, o c -ésimo par de funções da base com pólos complexos é dado por:

$$\Psi'_c(z) = \frac{z\sqrt{1-|\beta_c|^2}(\rho'_c z + \mu'_c)}{z^2 - (\beta_c + \bar{\beta}_c)z + |\beta_c|^2} \cdot \prod_{j=1}^{c-1} \left(\frac{1 - \bar{\beta}_j z}{z - \beta_j} \right) \left(\frac{1 - \beta_j z}{z - \bar{\beta}_j} \right) \cdot \prod_{j=1}^{r-1} \left(\frac{1 - p_j z}{z - p_j} \right) \quad (4.77)$$

$$\Psi''_c(z) = \frac{z\sqrt{1-|\beta_c|^2}(\rho''_c z + \mu''_c)}{z^2 - (\beta_c + \bar{\beta}_c)z + |\beta_c|^2} \cdot \prod_{j=1}^{c-1} \left(\frac{1 - \bar{\beta}_j z}{z - \beta_j} \right) \left(\frac{1 - \beta_j z}{z - \bar{\beta}_j} \right) \cdot \prod_{j=1}^{r-1} \left(\frac{1 - p_j z}{z - p_j} \right)$$

onde $\rho'_c, \mu'_c, \rho''_c, \mu''_c$ são parâmetros reais que estão relacionados com o pólo β_c como segue:

$$\begin{bmatrix} \rho'_c & \mu'_c \end{bmatrix} \begin{bmatrix} 1 + |\beta_c|^2 & \beta_c + \bar{\beta}_c \\ \beta_c + \bar{\beta}_c & 1 + |\beta_c|^2 \end{bmatrix} \begin{bmatrix} \rho'_c \\ \mu'_c \end{bmatrix} = |1 - \beta_c^2|^2 \quad (4.78)$$

$$\begin{bmatrix} \rho''_c \\ \mu''_c \end{bmatrix} = \frac{1}{\sqrt{1 - \xi_c^2}} \begin{bmatrix} \xi_c & 1 \\ -1 & -\xi_c \end{bmatrix} \begin{bmatrix} \rho'_c \\ \mu'_c \end{bmatrix} \quad (4.79)$$

$$\xi_c = \frac{\beta_c + \bar{\beta}_c}{1 + |\beta_c|^2} \quad (4.80)$$

O gradiente das novas funções geradas $\Psi'_c(z)$ e $\Psi''_c(z)$ em (4.77) pode ser calculado usando o mesmo raciocínio já apresentado anteriormente neste capítulo.

4.4 Generalização para modelos de Volterra de ordem superior

Os gradientes da função de custo J em (4.9), que são dados pelas equações (4.10) e (4.11), são válidos somente para modelos OBF-Volterra de primeira ordem (modelos lineares). Esta seção estende aqueles resultados com relação aos modelos de ordem superior. A principal modificação nesta extensão envolve a parte estática dos

modelos, que é correspondente à série de Volterra em (2.22) para $N > 1$. A parte dinâmica destes modelos, representada pelo conjunto de funções ortonormais, já foi explorada nas seções anteriores.

A generalização dos resultados referentes aos modelos de Volterra de ordem η é obtida diretamente. Por questão de simplicidade, considere que uma única base ortonormal é utilizada para todas as direções de cada kernel de Volterra. Com isto, o η -ésimo termo do modelo OBF-Volterra em (2.22) pode ser escrito separadamente como:

$$\begin{aligned}\hat{y}_\eta(k) &= \sum_{i_1=1}^M \cdots \sum_{i_\eta=1}^M \alpha_{i_1, \dots, i_\eta} \prod_{l=1}^{\eta} \left(\sum_{\tau_l=0}^k \psi_{i_l}(\tau_l) u(k - \tau_l) \right) \\ &= \sum_{i_1=1}^M \cdots \sum_{i_\eta=1}^M \alpha_{i_1, \dots, i_\eta} \prod_{l=1}^{\eta} \check{\psi}_{i_l}(k)\end{aligned}\quad (4.81)$$

O gradiente de (4.81) em relação ao vetor de pólos \mathbf{p} da base $\{\psi_m\}$ é dado por:

$$\begin{aligned}\nabla_{\mathbf{p}} \hat{y}_\eta(k) &= \sum_{i_1=1}^M \cdots \sum_{i_\eta=1}^M \alpha_{i_1, \dots, i_\eta} \left[\sum_{l=1}^{\eta} \nabla_{\mathbf{p}} \check{\psi}_{i_l}(k) \prod_{\substack{j=1 \\ j \neq l}}^{\eta} \check{\psi}_{i_j}(k) \right] \\ &= \sum_{l=1}^{\eta} \left[\sum_{i_1=1}^M \cdots \sum_{i_\eta=1}^M \alpha_{i_1, \dots, i_\eta} \nabla_{\mathbf{p}} \check{\psi}_{i_l}(k) \prod_{\substack{j=1 \\ j \neq l}}^{\eta} \check{\psi}_{i_j}(k) \right]\end{aligned}\quad (4.82)$$

Os gradientes das saídas das funções ortonormais $\nabla_{\mathbf{p}} \check{\psi}_{i_l}(k)$ em (4.82) são obtidos da mesma forma como apresentado na seção 4.2 para modelos baseados em funções de Kautz, e na seção 4.3 para modelos baseados em GOBF.

Note que a saída total do modelo (2.22) pode ser reescrita usando (4.81) como $\hat{y}(k) = \sum_{\eta=1}^N \hat{y}_\eta(k)$, o que implica que $\nabla_{\mathbf{p}} \hat{y}(k) = \sum_{\eta=1}^N \nabla_{\mathbf{p}} \hat{y}_\eta(k)$. Assim, o gradiente da função de custo J para o problema de otimização (4.9) pode agora ser calculado como:

$$\begin{aligned}\nabla_{\mathbf{p}} J &= \sum_{k=1}^{N_d} [\hat{y}(k) - y(k)] \nabla_{\mathbf{p}} \hat{y}(k) \\ &= \sum_{k=1}^{N_d} [\hat{y}(k) - y(k)] \sum_{\eta=1}^N \nabla_{\mathbf{p}} \hat{y}_\eta(k) \\ &= \sum_{k=1}^{N_d} \sum_{\eta=1}^N [\hat{y}(k) - y(k)] \nabla_{\mathbf{p}} \hat{y}_\eta(k)\end{aligned}\quad (4.83)$$

onde $\nabla_{\mathbf{p}}\hat{y}_\eta(k)$ é dado por (4.82). Usando um raciocínio análogo, tem-se que $\nabla_{\boldsymbol{\alpha}}\hat{y}(k) = \sum_{\eta=1}^N \nabla_{\boldsymbol{\alpha}}\hat{y}_\eta(k)$, onde $\nabla_{\boldsymbol{\alpha}}\hat{y}_\eta(k)$ pode ser facilmente obtido a partir de (4.81), e o gradiente de J com relação a $\boldsymbol{\alpha}$ pode ser calculado como:

$$\begin{aligned}\nabla_{\boldsymbol{\alpha}}J &= \sum_{k=1}^{N_d} [\hat{y}(k) - y(k)] \nabla_{\boldsymbol{\alpha}}\hat{y}(k) \\ &= \sum_{k=1}^{N_d} [\hat{y}(k) - y(k)] \sum_{\eta=1}^N \nabla_{\boldsymbol{\alpha}}\hat{y}_\eta(k) \\ &= \sum_{k=1}^{N_d} \sum_{\eta=1}^N [\hat{y}(k) - y(k)] \nabla_{\boldsymbol{\alpha}}\hat{y}_\eta(k)\end{aligned}\quad (4.84)$$

Apenas para se ter um maior entendimento a respeito da generalização apresentada nesta seção, considere um modelo de Volterra de segunda ordem. Neste caso, tem-se que $\eta = 2$. O termo de segunda ordem do modelo (2.22) é dado por:

$$\begin{aligned}\hat{y}_2(k) &= \sum_{i_1=1}^M \sum_{i_2=1}^M \alpha_{i_1, i_2} \sum_{\tau_1=0}^k \sum_{\tau_2=0}^k \psi_{i_1}(\tau_1) u(k - \tau_1) \psi_{i_2}(\tau_2) u(k - \tau_2) \\ &= \sum_{i_1=1}^M \sum_{i_2=1}^M \alpha_{i_1, i_2} \check{\psi}_{i_1}(k) \check{\psi}_{i_2}(k)\end{aligned}\quad (4.85)$$

O gradiente de $\hat{y}_2(k)$ com relação ao vetor \mathbf{p} que parametriza a base ortonormal $\{\psi_m\}$ resulta em:

$$\nabla_{\mathbf{p}}\hat{y}_2(k) = \sum_{i_1=1}^M \sum_{i_2=1}^M \alpha_{i_1, i_2} \left[\check{\psi}_{i_1}(k) \nabla_{\mathbf{p}}\check{\psi}_{i_2}(k) + \check{\psi}_{i_2}(k) \nabla_{\mathbf{p}}\check{\psi}_{i_1}(k) \right] \quad (4.86)$$

onde $\mathbf{p} \triangleq [b \quad c]^T$ quando o conjunto de funções ortonormais adotado é a base de Kautz (seção 4.2), ou $\mathbf{p} \triangleq [p_1 \quad \dots \quad p_M]^T$ quando o conjunto adotado é a base GOBF com pólos reais (seção 4.3.1), ou $\mathbf{p} \triangleq [\beta_1 \quad \dots \quad \beta_M]^T$ quando o conjunto adotado é a base GOBF com pólos complexos (seção 4.3.2).

Os principais passos do algoritmo proposto na seção 4.2.3 permanecem inalterados. A principal diferença é que os gradientes da função de custo J em (4.9) são agora calculados usando as equações (4.83) e (4.84) no lugar das equações (4.10) e (4.11), respectivamente. Assim, para modelos de Volterra de qualquer ordem, partindo de um vetor inicial de parâmetros $\boldsymbol{\theta}^0$, faça:

1. Calcule o gradiente $\nabla_{\boldsymbol{\theta}}^T J = [\nabla_{\mathbf{p}}^T J \quad \nabla_{\boldsymbol{\alpha}}^T J]$ usando as equações (4.83) e (4.84);

2. Use $\nabla_{\theta} J$ para atualizar o vetor de parâmetros θ de acordo com a política de atualização do método de Levenberg-Marquardt. Nota: inclua intervalos de factibilidade dos pólos da base adotada no modelo como restrições do problema de otimização para garantir estabilidade/factibilidade;
3. Volte para o passo 1 até que um critério de parada tenha sido atingido.

A seguir, apresentam-se experimentos computacionais que ilustram a aplicação da formulação discutida até agora.

4.5 Exemplos ilustrativos

Para ilustrar os resultados teóricos apresentados neste capítulo, simulações computacionais que demonstram a aplicação destes resultados serão apresentados a seguir. Um exemplo ilustrativo é relativo a um modelo linear, e o outro é referente a um modelo não-linear de Volterra.

4.5.1 Exemplo 4.1: modelo linear

Seja um sistema linear com resposta ao impulso (kernel de Volterra de primeira ordem) dada por:

$$h_s(k) = \zeta^k + \bar{\zeta}^k, \quad k \geq 0 \quad (4.87)$$

com $\zeta, \bar{\zeta} \in \mathbb{C}$ e a saída é dada por $y(k) = \sum_{\tau=0}^k h_s(\tau)u(k-\tau)$. Uma vez que a dinâmica em (4.87) é caracterizada por um par de pólos complexos conjugados ζ e $\bar{\zeta}$, espera-se então que os pólos que melhor representam este sistema sejam ζ e $\bar{\zeta}$.

Deseja-se aplicar os resultados matemáticos obtidos neste capítulo para otimização dos pólos de Kautz. O modelo utilizado é aquele na equação (4.8), onde $\{\psi_m\}$ é a base de funções de Kautz. Para o cálculo dos gradientes $\nabla_{\mathbf{p}} J$ e $\nabla_{\alpha} J$, utiliza-se as equações (4.10) e (4.11), respectivamente.

Considere um sistema linear regido pela dinâmica dada na equação (4.87) com $\zeta = 0,7 + i0,5$. Suponha que este sistema seja excitado por um sinal de entrada $u(k)$ aleatório com média nula e variância unitária. Os parâmetros reais correspondentes a este valor de ζ , calculados pelas equações (2.27) e (2.28), são dados por $b = 0,8046$ e $c = -0,74$. A metodologia proposta foi implementada utilizando-se um pólo inicial $\beta^0 = 0,1 + i0,1$, que corresponde a parâmetros de Kautz iniciais $\mathbf{p}^0 = [b^0 \quad c^0]^T = [0,1961 \quad -0,02]^T$.

Os valores ótimos dos parâmetros e pólos de Kautz, obtidos utilizando a metodologia proposta, estão resumidos na Tabela 4.1.

Devido à dinâmica pura de segunda ordem (dinâmica sub-amortecida) em (4.87), duas funções de Kautz com pólos $\beta = \zeta$ e $\bar{\beta} = \bar{\zeta}$ seriam suficientes para se obter uma expansão exata de h_s . Apesar disso, considerando

Tabela 4.1: Valores finais dos parâmetros e pólos de Kautz para o sistema com resposta ao impulso (4.87) com $\zeta = 0,7 + i0,5$.

b_{otimo}	c_{otimo}	$(\beta_{\text{otimo}}, \bar{\beta}_{\text{otimo}})$	J_{otimo}	Iterações
0,8045	-0,7399	$0,6999 \pm i0,5000$	$2,1999 \times 10^{-5}$	10
0,8046	-0,7399	$0,6999 \pm i0,5000$	$1,6177 \times 10^{-6}$	11
0,8046	-0,7400	$0,7000 \pm i0,5000$	$1,4689 \times 10^{-7}$	12
0,8046	-0,7400	$0,7000 \pm i0,5000$	$1,1807 \times 10^{-8}$	13

que o sistema seja desconhecido e apenas observações de entrada/saída estejam disponíveis, pode-se, na prática, testar um modelo com $M > 2$. Considere então que deseja-se representar este sistema por meio de um modelo que utiliza $M = 6$ funções de Kautz. Neste caso, os pólos ótimos obtidos são $(\beta, \bar{\beta}) = 0,7000 \pm i0,5000$, e os coeficientes da expansão ortonormal são calculados como sendo $\alpha_{\text{otimo}} = [2,9735 \quad 0,5238 \quad 0,0000 \quad 0,0000 \quad 0,0000 \quad 0,0000]^T$. Isto significa que apenas os coeficientes correspondentes às duas primeiras funções da base são levados em conta neste modelo, sendo que os coeficientes das demais funções são nulos.

A Figura 4.6 ilustra a evolução dos pólos para o sistema com dinâmica governada pela resposta ao impulso unitário em (4.87). Pode ser visto que o método iterativo convergiu para o pólo ótimo esperado $\zeta = 0,7 + i0,5$.

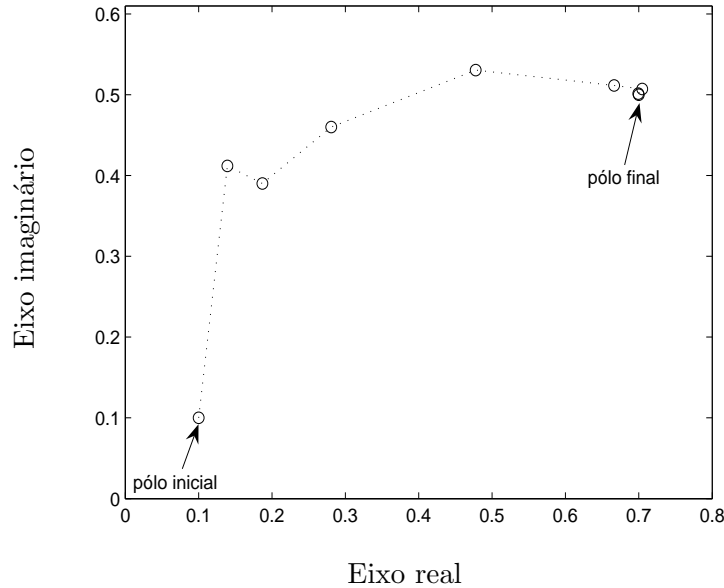


Figura 4.6: Evolução de um dos pólos ao longo das iterações, para o sistema linear com resposta ao impulso (4.87) com $\zeta = 0,7 + i0,5$.

Tendo obtido os pólos de Kautz, a saída real do sistema com kernel dado por (4.87) pode ser comparada com a saída do modelo obtido. A Figura 4.7 ilustra a saída medida $y(k)$ do sistema, juntamente com a saída $\hat{y}(k)$ do modelo, calculada usando a equação (4.8). As funções $\check{\psi}_m(k)$ do modelo são parametrizadas pelos pólos ótimos de Kautz apresentados na Tabela 4.1, com os dados correspondentes a dados de validação.

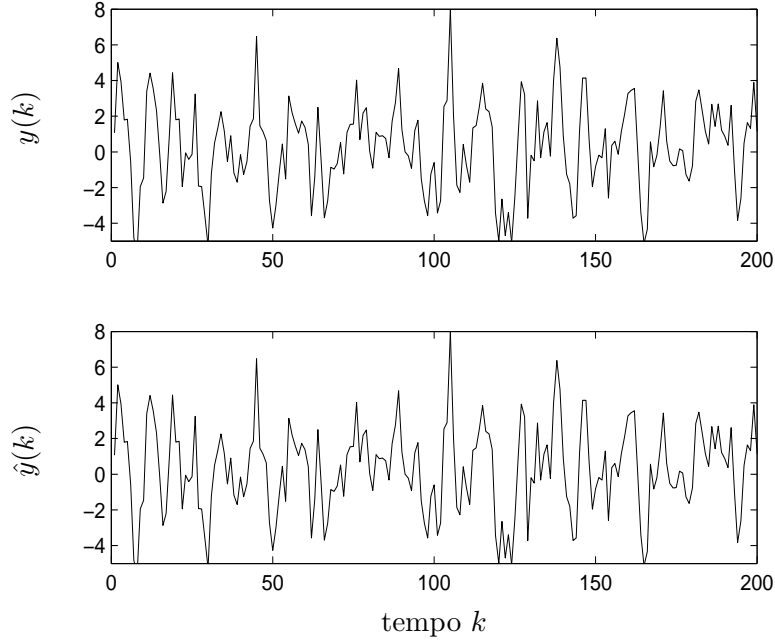


Figura 4.7: Validação do modelo. Acima: saída real do sistema linear com resposta ao impulso (4.87) com $\zeta = 0,7 + i0,5$. Abaixo: modelo correspondente usando os pólos ótimos obtidos.

Uma forma de verificar a boa representação do modelo é observando-se o erro de estimação do modelo, conforme mostrado na Figura 4.8.

A qualidade do modelo obtido pode também ser avaliada calculando-se o erro entre a saída real do sistema e a saída do modelo. Então, define-se o Erro Quadrático da Saída (EQS) como:

$$\text{EQS} \triangleq 10 \log \frac{\sum_{k=1}^{N_d} [y(k) - \hat{y}(k)]^2}{\sum_{k=1}^{N_d} [y(k)]^2} \quad (4.88)$$

onde N_d é o número de dados entrada/saída disponíveis. Neste exemplo, utilizou-se $N_d = 200$. O erro proveniente da aproximação do modelo do processo, obtido com o maior nível de precisão mostrado na Tabela 4.1, é $\text{EQS} = -110,2\text{dB}$.

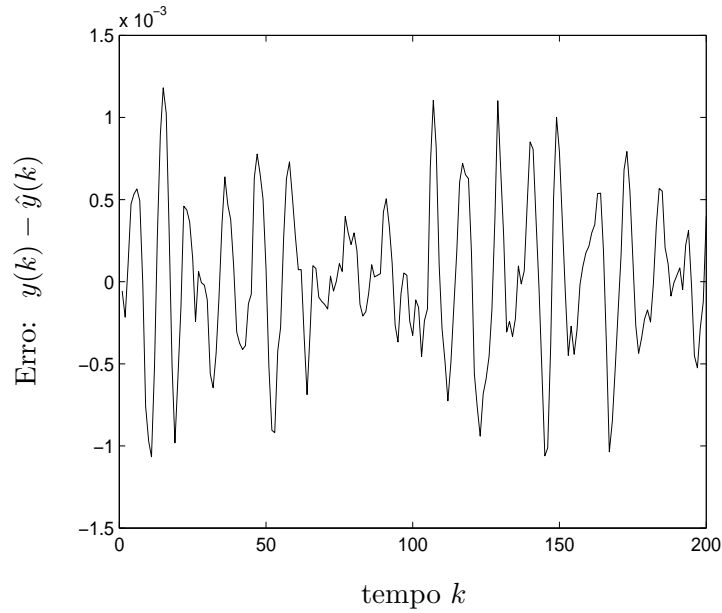


Figura 4.8: Erro de estimação da saída do sistema linear com resposta ao impulso dada por (4.87).

Em uma outra implementação computacional, o pólo ótimo de Kautz foi calculado variando-se a velocidade do sistema regido pela resposta ao impulso em (4.87). A Tabela 4.2 mostra os pólos obtidos para diferentes valores de ζ , correspondendo a dinâmicas rápida, média e lenta. O número de iterações necessárias para se atingir uma precisão de $J_{\text{ótimo}} = 10^{-5}$ também é apresentado. Os pólos iniciais são os mesmos que aqueles usados anteriormente: $0, 1 \pm i0, 1$, ou equivalentemente $\mathbf{p}^0 = [0, 1961 \quad -0, 02]^T$.

Tabela 4.2: Pólos ótimos de Kautz obtidos para diferentes valores de ζ em (4.87), com a precisão mantida em 10^{-5} .

ζ	Pólos ótimos $(\beta_{\text{ótimo}}, \beta_{\text{ótimo}})$	Iterações
$0, 07 + i0, 05$	$0, 0697 \pm i0, 0500$	6
$0, 5 + i0, 3$	$0, 4999 \pm i0, 3001$	10
$0, 9 + i0, 4$	$0, 9015 \pm i0, 3999$	21

Outras simulações foram realizadas com diferentes pólos iniciais, obtendo-se resultados semelhantes aos apresentados. A principal diferença neste sentido está relacionada com o número de iterações necessárias para se atingir a mesma precisão, i.e., quanto mais longe de $(\zeta, \bar{\zeta})$ estiver o pólo de inicialização do método, mais iterações são necessárias para se atingir uma dada precisão.

4.5.2 Exemplo 4.2: modelo não-linear

Para comparar a metodologia proposta com outras abordagens existentes na literatura, o método desenvolvido aqui será usado para selecionar os pólos de um sistema não-linear estudado em um trabalho recente (Kibangou *et al.* 2005c). Naquele trabalho, considera-se a representação de um modelo OBF-Volterra de segunda ordem, com os kernels de primeira e segunda ordem dados por:

$$h_1(k_1) = \mathcal{Z}^{-1} \left[\frac{z + 0,5}{(z - 0,3)(z - 0,2)} \right] \quad (4.89)$$

$$h_2(k_1, k_2) = 0,25g(k_1)g(k_2) \quad \text{onde} \quad g(k) = \mathcal{Z}^{-1} \left[\frac{z + 1}{(z - 0,8)(z + 0,8)} \right] \quad (4.90)$$

em que \mathcal{Z}^{-1} denota a transformada Z inversa unilateral. O cálculo das sequências discretas (4.89) e (4.90) resulta em:

$$\begin{aligned} h_1(k_1) &= \frac{80}{3} \cdot (0,3)^{k_1} - 35 \cdot (0,2)^{k_1} \\ h_2(k_1, k_2) &= 0,25 \cdot \left[\frac{45}{32} \cdot (0,8)^{k_1} + \frac{5}{32} \cdot (-0,8)^{k_1} \right] \cdot \left[\frac{45}{32} \cdot (0,8)^{k_2} + \frac{5}{32} \cdot (-0,8)^{k_2} \right] \end{aligned}$$

para $k_1, k_2 \geq 1$. A maior parte da discussão apresentada naquele trabalho (Kibangou *et al.* 2005c) é focada na comparação do erro da saída do sistema com kernels (4.89) e (4.90) para um número variado de funções da base, bem como o uso de diferentes bases ortonormais (Laguerre e GOBF).

Seguindo as linhas de Kibangou *et al.* (2005c), o modelo utilizado aqui também é uma representação OBF-Volterra de segunda ordem, em que bases GOBF $\{\psi_{1,m}\}$ e $\{\psi_{2,m}\}$ são adotadas para a expansão dos kernels de primeira e de segunda ordem, respectivamente. Uma hipótese inicial feita sobre o modelo é que o kernel de segunda ordem seja simétrico (ver seção 2.1.2), o que pode ser feito sem perda de generalidade. Neste caso, tem-se que $h_2(k_1, k_2) = h_2(k_2, k_1)$, e portanto os coeficientes da expansão ortonormal satisfazem $\alpha_{i_1, i_2} = \alpha_{i_2, i_1}$. Assim, o modelo resultante é o seguinte:

$$\hat{y}(k) = \sum_{i_1=1}^M \alpha_{i_1} \check{\psi}_{1,i_1}(k) + \sum_{i_1=1}^M \sum_{i_2=1}^{i_1} \alpha_{i_1, i_2} \check{\psi}_{2,i_1}(k) \check{\psi}_{2,i_2}(k) \quad (4.91)$$

com $\check{\psi}_{1,i}(k)$ e $\check{\psi}_{2,i}(k)$ denotando o resultado da filtragem do sinal de entrada $u(k)$ pelas funções ortonormais generalizadas $\psi_{1,i}(k)$ e $\psi_{2,i}(k)$, respectivamente. As bases $\{\psi_{1,m}\}$ e $\{\psi_{2,m}\}$ são projetadas de tal forma a ter M pólos reais, o que significa que M funções de cada base devem ser utilizadas. Usando a equação (4.43), defina a seguinte base ortonormal:

$$\Psi_{\eta,m}(z) = \frac{z \sqrt{1 - p_{\eta,m}^2}}{z - p_{\eta,m}} \prod_{j=1}^{m-1} \left(\frac{1 - p_{\eta,j} z}{z - p_{\eta,j}} \right) \quad m = 1, 2, \dots, M \quad \eta = 1, 2 \quad (4.92)$$

Cada base ortonormal em (4.92) é parametrizada por um vetor individual de pólos. Em outras palavras, as funções que descrevem o termo de primeira ordem do modelo (4.91) são parametrizadas por um vetor de pólos $\mathbf{p}_1 \triangleq [p_{1,1} \ p_{1,2} \ \cdots \ p_{1,M}]^T$, enquanto que as funções que descrevem o termo de segunda ordem do modelo serão parametrizadas por $\mathbf{p}_2 \triangleq [p_{2,1} \ p_{2,2} \ \cdots \ p_{2,M}]^T$, onde $p_{\eta,m}$ é definido como o pólo real da m -ésima função da η -ésima base. O cálculo do gradiente de (4.91) com relação a $\mathbf{p} \triangleq [\mathbf{p}_1^T \ \mathbf{p}_2^T]^T$ é então dado por:

$$\begin{aligned} \nabla_{\mathbf{p}} \hat{y}(k) &= \sum_{i_1=1}^M \alpha_{i_1} \nabla_{\mathbf{p}} \check{\psi}_{1,i_1}(k) \\ &+ \sum_{i_1=1}^M \sum_{i_2=1}^{i_1} \alpha_{i_1,i_2} \left[\check{\psi}_{2,i_1}(k) \nabla_{\mathbf{p}} \check{\psi}_{2,i_2}(k) + \check{\psi}_{2,i_2}(k) \nabla_{\mathbf{p}} \check{\psi}_{2,i_1}(k) \right] \end{aligned} \quad (4.93)$$

Obviamente, em (4.93) as componentes do gradiente relativas a uma ordem particular (por exemplo a primeira ordem), que são calculadas com respeito ao vetor de pólos de outra ordem (a segunda ordem), são nulos. Isso é verdadeiro porque as funções ortonormais no termo de primeira ordem dependem unicamente do seu próprio vetor de pólos (\mathbf{p}_1), mas não do vetor de pólos do termo de segunda ordem (\mathbf{p}_2), e vice-versa.

Para efeitos de simulação numérica, assume-se que o sistema em questão é excitado por um sinal de entrada aleatório com média zero e variância unitária. Um experimento com $M = 4$ é realizado usando-se os vetores de pólos iniciais $\mathbf{p}_1^0 = \mathbf{p}_2^0 = [0 \ 0 \ 0 \ 0]^T$. Após a implementação do método de otimização, uma série de pólos é gerada até a convergência do algoritmo, que levou 18 iterações. Os pólos finais obtidos para as bases ortonormais do modelo em (4.91) estão resumidos na Tabela 4.3. Os coeficientes correspondentes à esta expansão são dados por:

$$\boldsymbol{\alpha}_{1,\text{otimo}} = [1,1954 \quad 0,8901 \quad 0,0006 \quad 0,0005]^T \quad (4.94)$$

$$\boldsymbol{\alpha}_{2,\text{otimo}} = \begin{bmatrix} 0,8362 & * & * & * \\ 0,1833 & 0,0117 & * & * \\ -0,0083 & -0,0021 & -0,0001 & * \\ -0,0066 & -0,0005 & -0,0005 & 0,0003 \end{bmatrix} \quad (4.95)$$

onde os elementos denotados por * indicam que a matriz em (4.95) é simétrica.

Uma comparação qualitativa entre os kernels original e recuperado para o modelo (4.91) é apresentada a seguir. As Figuras 4.9 e 4.10 ilustram os kernels originais dados por (4.89) e (4.90), juntamente com os kernels correspondentes do modelo obtido. Estes kernels são calculados a partir dos coeficientes da expansão ortonormal em (4.94) e (4.95), respectivamente.

Tabela 4.3: Pólos ótimos obtidos para a base GOBF do modelo (4.91) usando $M = 4$ funções.

Ordem do kernel (η)	Pólos finais
1	$p_{1,1} = 0,2011$ $p_{1,2} = 0,2985$ $p_{1,3} = 0,0057$ $p_{1,4} = 0,0016$
2	$p_{2,1} = 0,8027$ $p_{2,2} = -0,7982$ $p_{2,3} = 0,0047$ $p_{2,4} = 0,0025$

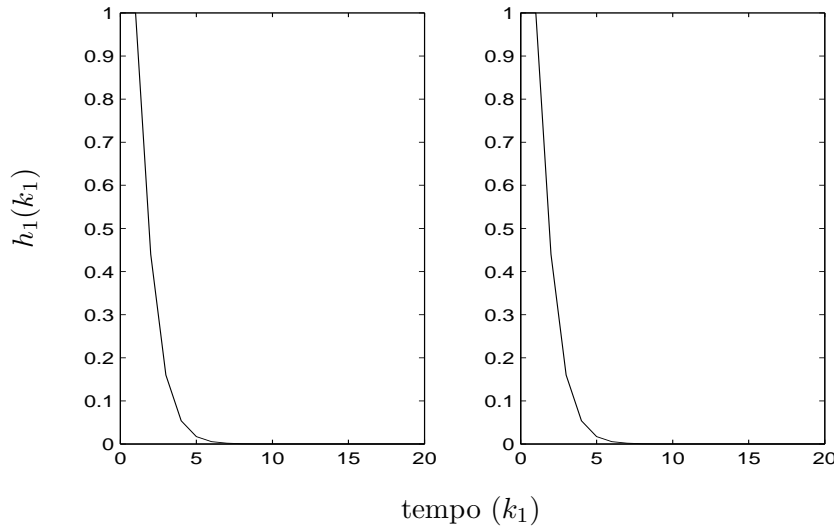


Figura 4.9: Kernel de primeira ordem original (à esquerda) e reproduzido (à direita) com a base GOBF parametrizada pelos pólos ótimos obtidos.

A qualidade desta aproximação também pode ser verificada comparando-se a saída do modelo com a saída do sistema original com relação a um outro conjunto de dados (dados de validação), como ilustrado na Figura 4.11.

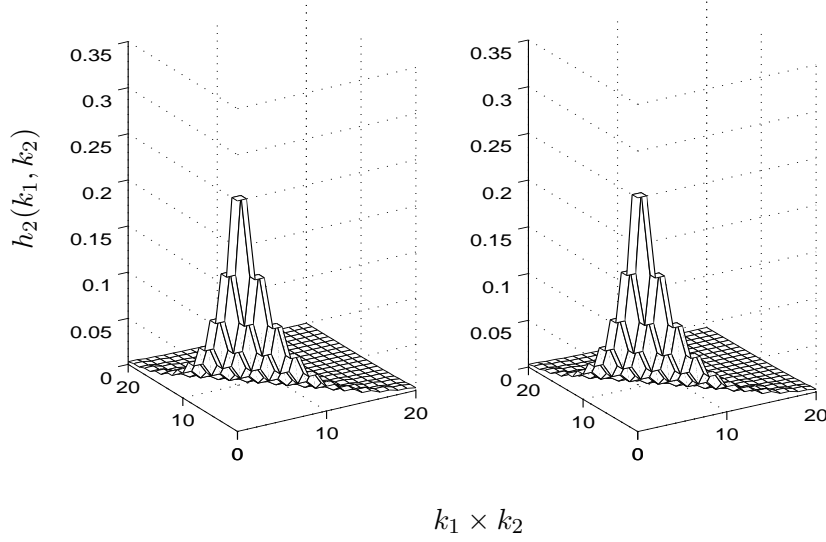


Figura 4.10: Kernel de segunda ordem original (à esquerda) e reproduzido (à direita) com a base GOBF parametrizada pelos pólos ótimos obtidos.

Uma comparação mostrando o erro obtido para algumas bases ortonormais é apresentada na Tabela 4.4. O número de funções adotadas em cada caso é $M = 4$, e o número de dados entrada/saída utilizado é $N_d = 200$. Para uma comparação justa com o trabalho anterior (Kibangou *et al.* 2005c), os resultados obtidos aqui são considerados quando um ruído Gaussiano com relação sinal/ruído (SR) igual a 30dB é adicionado à saída do modelo (4.91).

Tabela 4.4: Comparação dos erros obtidos usando diferentes abordagens para otimizar os parâmetros do modelo (4.91).

Método	EQS (em dB) com SR = 30dB
Em (Kibangou <i>et al.</i> 2005c) (com Laguerre)	-21,5
Em (Kibangou <i>et al.</i> 2005c) (com GOBF)	-27,0
Proposto aqui (com GOBF)	-36,4

Nota-se na Tabela 4.4 que o erro resultante da metodologia proposta sofreu uma atenuação maior do que o erro obtido anteriormente (Kibangou *et al.* 2005c). Como esperado, modelos baseados em GOBF fornecem resultados melhores do que aqueles envolvendo bases mais simples (Laguerre nesse caso). De fato, conforme já discutido neste trabalho, se o sistema possui uma dinâmica dominante mais complexa, então o uso de GOBF é preferível pois pode incorporar uma variedade de tipos de pólos.

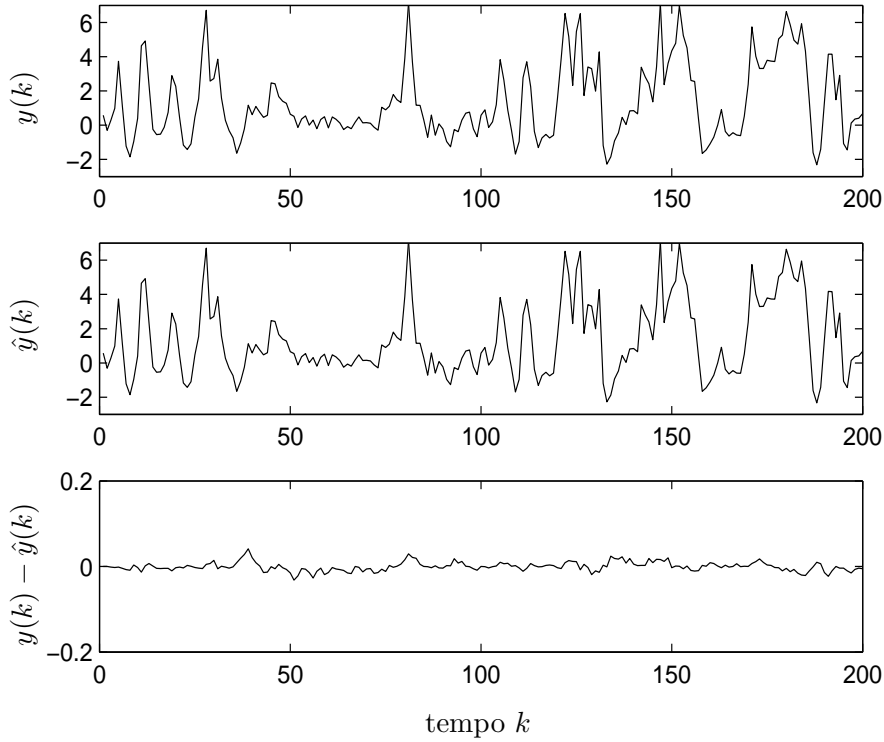


Figura 4.11: Saídas real, do modelo e sinal de erro obtidos com a base GOBF parametrizada pelos pólos ótimos obtidos.

4.6 Aplicação na modelagem do sistema de levitação magnética

Esta seção tem o objetivo de aplicar a metodologia proposta para a modelagem do sistema de levitação magnética descrito no Capítulo 3. Dois modelos OBF-Volterra de segunda ordem serão utilizados: o primeiro modelo é baseado em funções de Kautz, e o segundo é baseado em GOBF.

4.6.1 Modelo com funções de Kautz

O modelo matemático será similar àquele na equação (3.23), com a diferença de que aqui se utilizará uma única base para descrever as duas direções do kernel de segunda ordem. Uma suposição feita sobre o modelo é que o kernel de segunda ordem seja simétrico. Expressando-se então os termos de primeira a segunda ordem do modelo de Volterra como uma expansão em bases de funções de Kautz, tem-se o

seguinte modelo:

$$\hat{y}(k) = \sum_{i_1=1}^M \alpha_{i_1} \check{\psi}_{i_1}(k) + \sum_{i_1=1}^M \sum_{i_2=1}^{i_1} \alpha_{i_1, i_2} \check{\psi}_{i_1}(k) \check{\psi}_{i_2}(k) \quad (4.96)$$

com $\check{\psi}_{i_1}(k)$ e $\check{\psi}_{i_2}(k)$ denotando, respectivamente, as funções de Kautz $\psi_{i_1}(k)$ e $\psi_{i_2}(k)$ filtradas pelo sinal de entrada $u(k)$. Cada base ortonormal será parametrizada por um pólo individual, i.e., as funções que descrevem o termo de primeira ordem do modelo (4.96) serão parametrizadas pelo vetor $\mathbf{p}_1 \triangleq [b_1 \quad c_1]^T$, enquanto que as funções que descrevem o termo de segunda ordem do modelo serão parametrizadas pelo vetor $\mathbf{p}_2 \triangleq [b_2 \quad c_2]^T$. Assim, o gradiente da saída do modelo com relação a $\mathbf{p} \triangleq [\mathbf{p}_1 \quad \mathbf{p}_2]^T$ é dado por:

$$\begin{aligned} \nabla_{\mathbf{p}} \hat{y}(k) &= \sum_{i_1=1}^M \alpha_{i_1} \nabla_{\mathbf{p}} \check{\psi}_{i_1}(\tau_1) \\ &+ \sum_{i_1=1}^M \sum_{i_2=1}^{i_1} \alpha_{i_1, i_2} \left[\check{\psi}_{i_1}(k) \nabla_{\mathbf{p}} \check{\psi}_{i_2}(k) + \check{\psi}_{i_2}(k) \nabla_{\mathbf{p}} \check{\psi}_{i_1}(k) \right] \end{aligned} \quad (4.97)$$

Nesta simulação computacional, uma boa escolha dos pólos de Kautz iniciais pode ser obtida na prática ao se observar a resposta temporal do sistema (Figura 3.8). Apesar disso, adota-se aqui os pólos de Kautz iniciais como sendo $(\beta_1^0, \bar{\beta}_1^0) = (\beta_2^0, \bar{\beta}_2^0) = 0,5 \pm i0,5$ para os termos de primeira e segunda ordem do modelo (4.96), por eles estarem localizados no “centro” do espaço de busca factível para as componentes reais e imaginárias dos pólos. Estes valores fornecem os vetores de parâmetros reais $\mathbf{p}_1^0 = \mathbf{p}_2^0 = [0,6667 \quad -0,5]^T$. O número de funções de Kautz adotadas é $M = 8$. As equações (4.22), (4.33), (4.35) e (4.39) são utilizadas para calcular $\nabla_{\mathbf{p}} \check{\psi}_m(k)$ e, conseqüentemente, para se obter a direção de busca dos pólos.

As Figuras 4.12 e 4.13 ilustram, respectivamente, a evolução dos pólos para os termos de primeira e segunda ordem do modelo do sistema de levitação magnética. Nessas figuras, são também mostradas algumas curvas de nível correspondentes a linhas de igual valor de EQS. A Tabela 4.5 apresenta os valores finais dos parâmetros e dos pólos de Kautz após a convergência do algoritmo, que levou 39 iterações. O critério de parada adotado foi que a diferença entre o valor da função de custo J em (4.9) em iterações sucessivas seja menor do que 10^{-5} .

Tabela 4.5: Valores ótimos dos parâmetros e pólos de Kautz obtidos para o sistema de levitação magnética.

Ordem do kernel (η)	b_{otimo}	c_{otimo}	Pólos ótimos ($\beta_{\text{otimo}}, \bar{\beta}_{\text{otimo}}$)
1	0,9023	-0,6245	$0,7328 \pm i0,2956$
2	0,9328	-0,7421	$0,8125 \pm i0,2863$

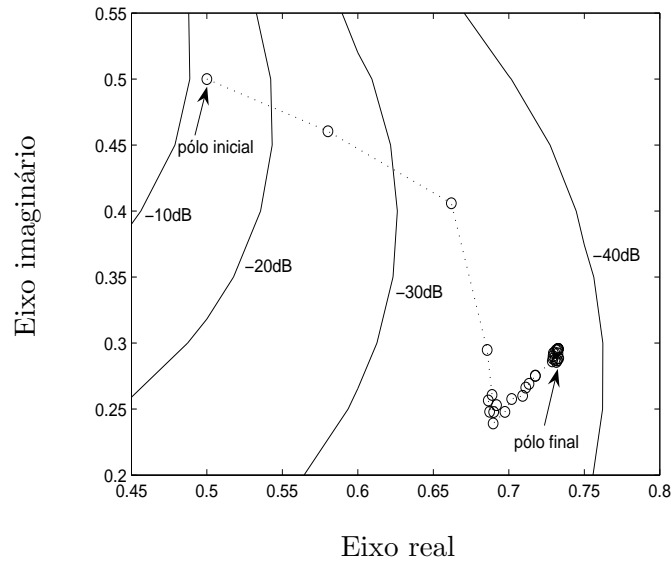


Figura 4.12: Evolução, ao longo das iterações, do pólo de primeira ordem para o modelo do sistema de levitação magnética, mostrada juntamente com algumas curvas de nível para EQS.

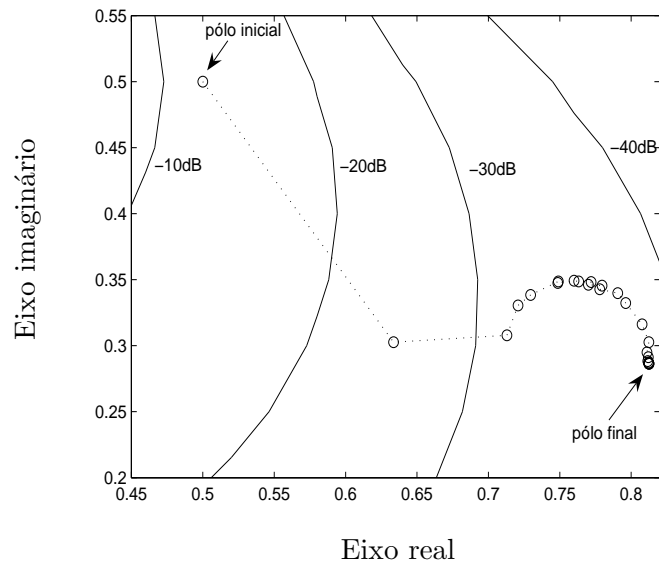


Figura 4.13: Evolução, ao longo das iterações, do pólo de segunda ordem para o modelo do sistema de levitação magnética, mostrada juntamente com algumas curvas de nível para EQS.

Para ilustrar o comportamento da evolução dos pólos na região próxima à convergência do método proposto, as Figuras 4.14 e 4.15 mostram um *zoom* das Figuras 4.12 e 4.13, respectivamente.

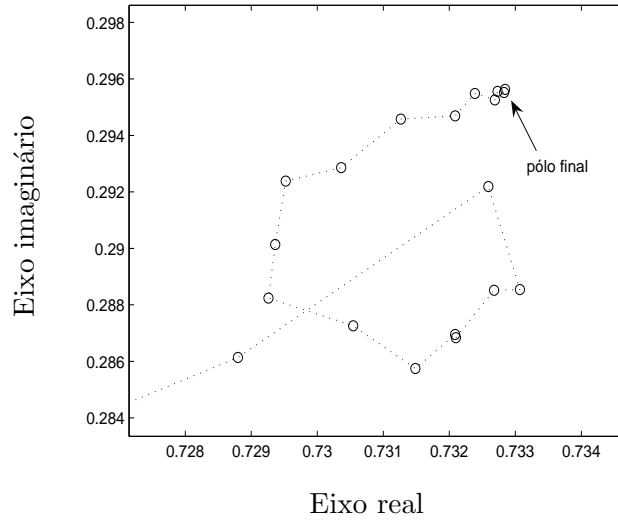


Figura 4.14: *Zoom* na região próxima ao pólo final obtido para o kernel de primeira ordem $\beta_{\text{otimo}} = 0,7328 \pm i0,2956$.

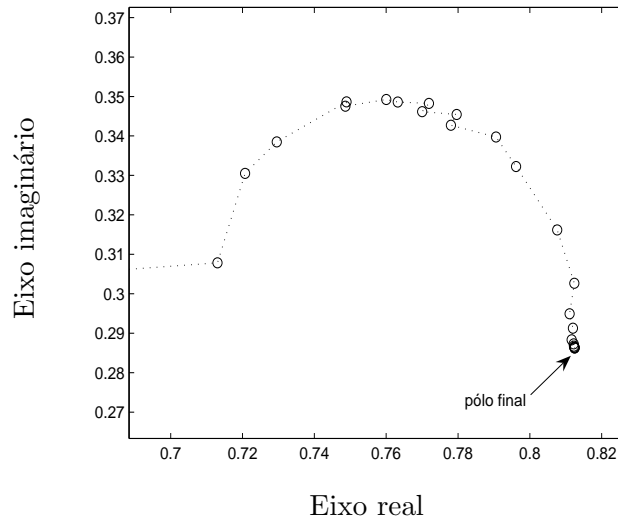


Figura 4.15: *Zoom* na região próxima ao pólo final obtido para o kernel de segunda ordem $\beta_{\text{otimo}} = 0,8125 \pm i0,2863$.

Após ter obtido os pólos ótimos para o sistema de levitação megnética, a saída do modelo pode então ser comparada com a saída do sistema. A Figura 4.16 ilustra a estimação da saída do sistema, $\hat{y}(k)$, com respeito a um outro conjunto de dados (dados de validação), juntamente com a saída real (medida) $y(k)$. O erro resultante desta estimação está mostrado na Figura 4.17.

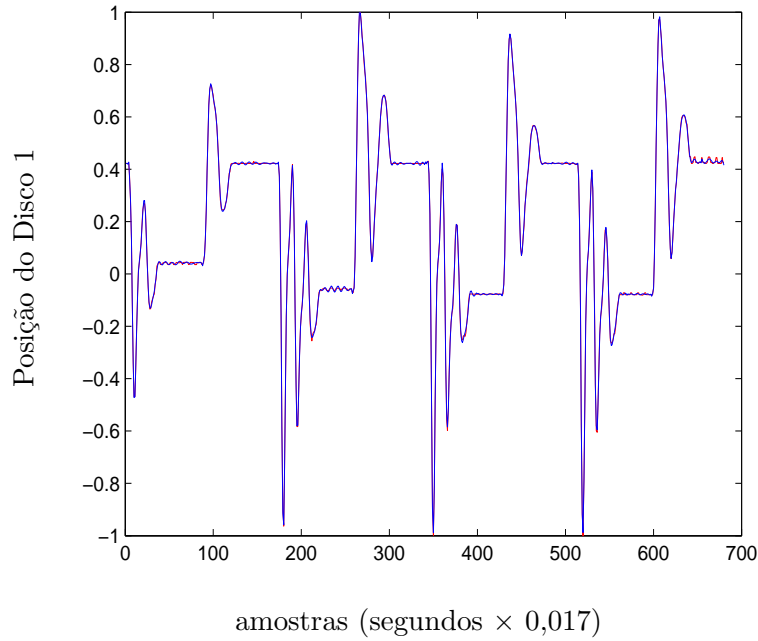


Figura 4.16: Validação do modelo do levitador magnético. Saída real ($y(k)$ em azul) e saída do modelo ($\hat{y}(k)$ em vermelho) usando os pólos ótimos obtidos.

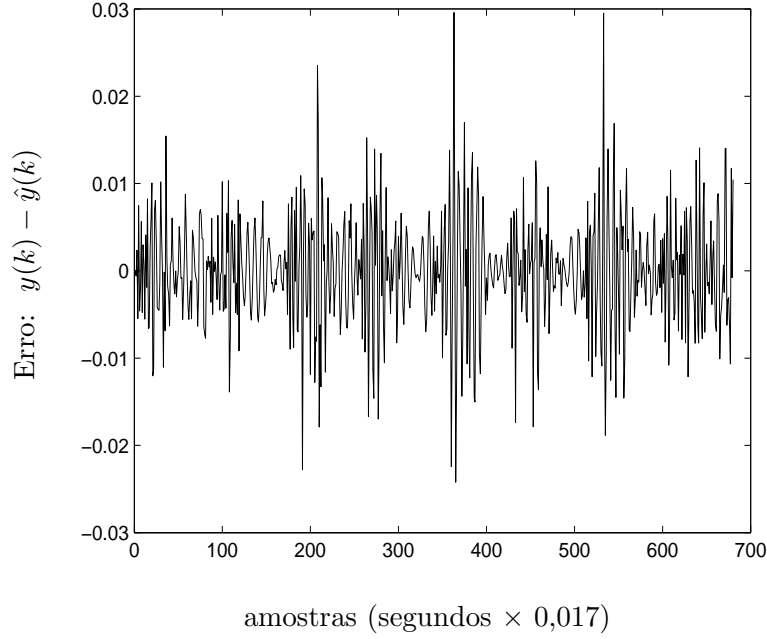


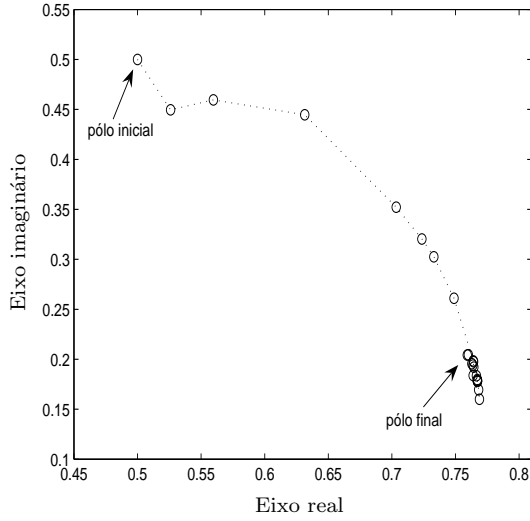
Figura 4.17: Erro de estimação entre a saída real e a saída do modelo do levitador magnético.

4.6.2 Modelo com funções GOBF

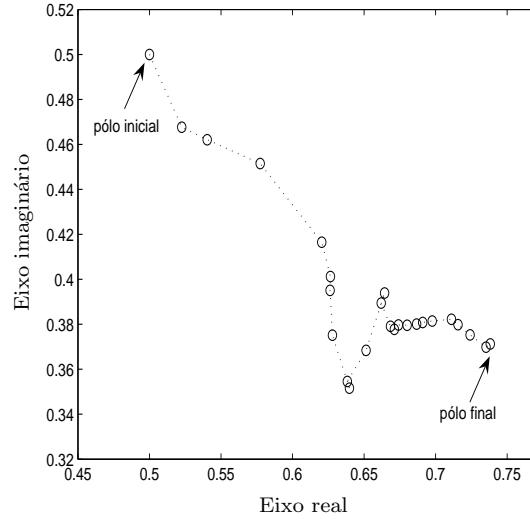
O modelo adotado nesta seção é uma representação de Volterra de segunda ordem, como aquele na equação (4.96), porém agora baseado em funções ortonormais generalizadas (GOBF) com dois pares de pólos complexos.

Os dados entrada/saída para estimação do modelo são os mesmos apresentados na Figura 3.8. O modelo utilizado possui 4 funções GOBF independentes para cada ordem, contemplando dois pares de pólos complexos, baseado na abordagem apresentada na seção 4.3.3. Nesta simulação, adotou-se $\gamma' = \gamma'' = \mu' = \mu'' = 0$. Em Ziaei e Wang (2006), discute-se que a escolha destes parâmetros não tem efeito significativo na modelagem do processo. Este exemplo também foi simulado utilizando-se $\lambda' = \lambda'' = \rho' = \rho'' = 1$, e obteve-se resultados similares.

Aplicando a metodologia proposta para a obtenção do modelo do levitador magnético usando os pólos iniciais $\beta_{1,1}^0 = \beta_{1,2}^0 = \beta_{2,1}^0 = \beta_{2,1}^0 = 0,5 + i0,5$, e considerando o critério de parada tal que a diferença entre o valor da função de custo J em (4.9) em iterações sucessivas seja menor do que 10^{-5} , o método convergiu após 26 iterações. A evolução dos pólos ao longo das iterações é ilustrada nas Figuras 4.18(a) e 4.18(b) para os pólos de primeira ordem, e nas Figuras 4.19(a) e 4.19(b) para os pólos de segunda ordem. A Tabela 4.6 mostra os valores finais obtidos para os pólos das funções GOBF.

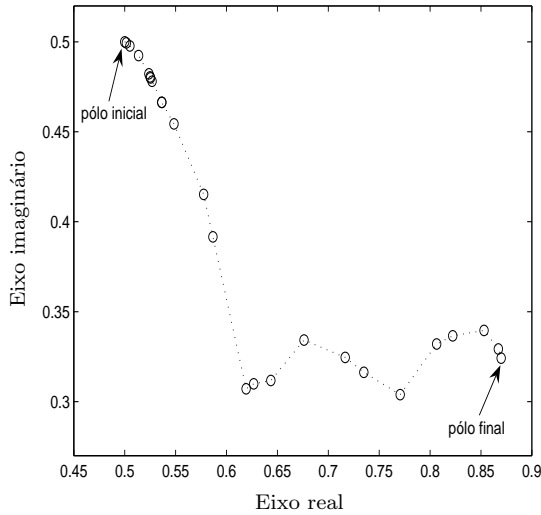


(a) Evolução de $\beta_{1,1}$

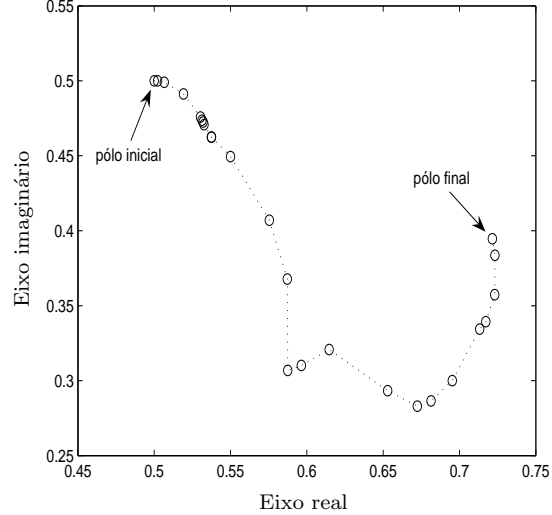


(b) Evolução de $\beta_{1,2}$

Figura 4.18: Evolução, ao longo das iterações, dos pólos do termo de primeira ordem para o modelo do levitador magnético usando funções GOBF.



(a) Evolução de $\beta_{2,1}$



(b) Evolução de $\beta_{2,2}$

Figura 4.19: Evolução, ao longo das iterações, dos pólos do termo de segunda ordem para o modelo do levitador magnético usando funções GOBF.

Por fim, a Tabela 4.7 apresenta uma comparação entre diferentes abordagens encontradas na literatura com relação ao erro de estimação obtido para o modelo do levitador magnético.

Tabela 4.6: Pólos obtidos para o modelo do sistema de levitação magnética com GOBF.

Ordem do kernel (η)	Pólos obtidos
1	$\beta_{1,1} = 0,7592 \pm i0,2042$
	$\beta_{1,2} = 0,7382 \pm i0,3713$
2	$\beta_{2,1} = 0,8698 \pm i0,3242$
	$\beta_{2,2} = 0,7215 \pm i0,3947$

Tabela 4.7: Comparação dos erros obtidos usando diferentes abordagens para otimizar os parâmetros do modelo do sistema de levitação magnética.

Método	EQS (em dB)
Em (da Rosa 2005, da Rosa <i>et al.</i> 2006) (base única de Kautz)	-22,9
No Capítulo 3 e em (da Rosa <i>et al.</i> 2008a) (bases múltiplas de Kautz)	-30,1
Na seção 4.6.1 e em (da Rosa <i>et al.</i> aceito) (base de Kautz)	-35,6
Nesta seção (base GOBF)	-33,2

O erro obtido com o modelo GOBF é ligeiramente maior quando comparado com o erro obtido pelo modelo de Kautz (-33,2dB contra -35,6dB), contudo utiliza-se somente a metade do número de funções (4 contra 8). Como um modelo de Volterra simétrico com termos de primeira e segunda ordem, usando M funções, possui um número total de coeficientes da expansão ortonormal dado por $M + M \cdot (M + 1)/2$, e considerando também o número de pólos, o modelo de Kautz tem 48 parâmetros, enquanto que o modelo GOBF tem somente 22 parâmetros. Assim, o modelo GOBF utiliza menos da metade do número de parâmetros do que modelo de Kautz para representar o levitador magnético, e apresenta erro de modelagem similar.

4.7 Resumo e contribuições

Este capítulo apresentou uma nova estratégia para otimização numérica dos pólos das bases ortonormais em modelos OBF-Volterra de qualquer ordem. Esta estratégia é baseada no cálculo de expressões analíticas para os gradientes dos filtros ortonormais em relação aos pólos da base. As expressões obtidas fornecem direções de busca exatas a serem utilizadas por um método numérico para otimizar os pólos em modelos lineares e não-lineares. O problema central foi abordado por meio da minimização de uma função de custo quadrática que leva em consideração o erro de estimação da saída do sistema. Apesar da ênfase dada para modelos baseados em funções de Kautz, um estudo equivalente baseado em funções de Laguerre pode ser facilmente obtido como um caso particular. A principal vantagem desta metodologia é que os gradientes são calculados analiticamente usando somente da-

dos entrada/saída medidos a partir do sistema a ser modelado, sem a necessidade de se ter informação prévia sobre os kernels do sistema. Estudou-se também modelos baseados em funções ortonormais generalizadas (GOBF), através da obtenção dos gradientes destas funções com relação aos seus pólos.

O método proposto foi implementado para obter a seleção ótima dos pólos de Kautz e de pólos GOBF em exemplos de simulação de sistemas dinâmicos lineares e não-lineares. Uma discussão dos resultados numéricos obtidos permitiu verificar a qualidade da aproximação dos sistemas estudados, comparando-se também estes resultados com outros métodos encontrados na literatura.

Aplicou-se ainda esta metodologia na modelagem de um sistema não-linear de levitação magnética com comportamento oscilatório. A partir da medição de dados de entrada/saída, obteve-se modelos de Volterra de segunda ordem baseados em funções de Kautz e GOBF. Simulações computacionais ilustraram que a metodologia proposta permitiu obter um modelo que representa adequadamente as dinâmicas do sistema.

Capítulo 5

Modelagem de Sistemas Não-Lineares Incertos Usando Funções Ortonormais

O desempenho de um sistema em malha fechada depende da confiabilidade do modelo utilizado para o cálculo das previsões das saídas. Se o modelo disponível representa a dinâmica satisfatoriamente na faixa de operação de interesse, uma previsão correta da saída do sistema pode ser calculada no instante atual. Por outro lado, diversos processos reais apresentam comportamento incerto ou variante no tempo. Quando as incertezas sobre os parâmetros do modelo são definidas por intervalos de pertinência e se tem certeza quanto à ordem do modelo, diz-se que a incerteza é do tipo estruturada (Oliveira 1997, Moreira 2006).

Este capítulo apresenta resultados referentes a seleção de parâmetros de modelos OBF-Volterra quando o modelo do processo possui incerteza estruturada. No contexto dos sistemas lineares, os intervalos de pertinência dos coeficientes do modelo OBF, que representam os limitantes das incertezas, já foram estimados na literatura usando um conjunto de respostas ao impulso (Oliveira 1997, Oliveira *et al.* 2000, Moreira 2006). Este conjunto pode ser obtido, por exemplo, a partir da planta do processo para diversos valores de incertezas. Nos trabalhos anteriores, também foram propostas condições para assegurar que o modelo estimado represente todas as respostas ao impulso (kernel de Volterra de primeira ordem) do processo, o que é exigido no projeto de um controlador robusto. O objetivo aqui é estender esses resultados para o contexto dos modelos não-lineares de Volterra através do cálculo de limitantes das incertezas baseado em um conjunto de kernels de Volterra incertos. Um exemplo ilustrativo também é apresentado, onde se compara a abordagem proposta com outra existente na literatura.

5.1 Formulação de modelos com incertezas

A seguir, apresenta-se a formulação de modelos lineares e não-lineares (Volterra) quando este possui incerteza sobre os kernels do modelo.

5.1.1 Modelos OBF lineares incertos

Considere um sistema linear representado por um modelo de Volterra de primeira ordem, como aquele na equação (2.1). Suponha que este modelo possua uma incerteza sobre seus parâmetros, representados aqui pela resposta ao impulso do processo. Desta forma, estas incertezas podem ser descritas de acordo com o seguinte conjunto de respostas ao impulso (Oliveira 1997, Oliveira *et al.* 2000):

$$\{h_1(k)\} = \left\{ \bar{h}_1(k) + \rho_1(k)\Delta h_1(k) \right\} \quad (5.1)$$

onde:

- $\bar{h}_1(k)$ é o valor central do conjunto de respostas ao impulso;
- $\Delta h_1(k)$ representa o módulo da máxima variação das respostas ao impulso em torno de $\bar{h}_1(k)$;
- $\rho_1(k)$ representa uma incerteza sobre $h_1(k)$ satisfazendo $-1 \leq \rho_1(k) \leq 1$.

Os coeficientes do desenvolvimento ortonormal deste modelo incerto podem ser calculados por meio de uma expansão em uma base de funções ortonormais $\{\psi_m\}$. De acordo com a equação (2.16), estes coeficientes são dados por:

$$\alpha_m = \sum_{k=0}^{\infty} h_1(k)\psi_m(k)$$

Usando o conjunto dado em (5.1) para reescrever a equação acima, tem-se o seguinte conjunto de coeficientes incertos:

$$\begin{aligned} \{\alpha_m\} &= \sum_{k=0}^{\infty} \{h_1(k)\}\psi_m(k) \\ &= \sum_{k=0}^{\infty} \left\{ \bar{h}_1(k) + \rho_1(k)\Delta h_1(k) \right\} \psi_m(k) \end{aligned} \quad (5.2)$$

Embora não seja possível determinar analiticamente a máxima incerteza dos coeficientes do desenvolvimento ortonormal $\alpha_{(\cdot)}$ do modelo incerto, pode-se obter uma estimativa desta incerteza, conforme apresentado em trabalhos anteriores

(Oliveira 1997, Oliveira *et al.* 2000). Dado que $\rho_1(k)$ está contido no intervalo $[-1, 1]$, e ainda considerando que seja válida a seguinte condição:

$$\psi_m(k) > 0, \quad \forall k \quad m = 1, \dots, M \quad (5.3)$$

pode-se assim escrever a desigualdade a seguir (Oliveira 1997, Oliveira *et al.* 2000):

$$-\sum_{k=0}^{\infty} \Delta h_1(k) \psi_m(k) \leq \sum_{k=0}^{\infty} \rho_1(k) \Delta h_1(k) \psi_m(k) \leq \sum_{k=0}^{\infty} \Delta h_1(k) \psi_m(k) \quad (5.4)$$

$m = 1, \dots, M$

Os extremos da desigualdade dada em (5.4) são então tomados como a máxima incerteza (inferior e superior) sobre os coeficientes da expansão em base ortonormal do modelo linear, como feito em outras trabalhos (Oliveira 1997, Oliveira *et al.* 2000, Moreira *et al.* 2004, Moreira 2006). Desta forma, o valor central dos coeficientes da expansão ortonormal e a máxima incerteza sobre este valor central são dados, respectivamente, por:

$$\begin{aligned} \bar{\alpha}_m &= \sum_{k=0}^{\infty} \bar{h}_1(k) \psi_m(k) \\ \Delta \alpha_m &= \sum_{k=0}^{\infty} \Delta h_1(k) \psi_m(k) \end{aligned}$$

Logo, os coeficientes da expansão ortonormal do modelo também podem ser representados por meio de intervalos de pertinência, como a seguir:

$$\{\alpha_m\} = \{\bar{\alpha}_m + \sigma_m \Delta \alpha_m\} \quad (5.5)$$

onde:

- $\bar{\alpha}_m$ são os valores centrais dos coeficientes da expansão em base ortonormal;
- $\Delta \alpha_m$ representa um desvio dos coeficientes em torno dos respectivos valores centrais;
- σ_m representa um fator de escalamento sobre $\Delta \alpha_m$ satisfazendo $-1 \leq \sigma_m \leq 1$.

5.1.2 Modelos OBF não-lineares incertos

Esta seção objetiva estender os conceitos apresentados na seção anterior para o contexto dos modelos OBF-Volterra. Considere então um modelo de Volterra de ordem η , como aquele em (2.1), que possui incerteza sobre seus parâmetros, aqui representados pelos kernels do modelo. Seguindo as linhas de raciocínio da seção

anterior, estas incertezas podem ser descritas matematicamente por um conjunto de kernels do tipo:

$$\{h_\eta(k_1, \dots, k_\eta)\} = \left\{ \bar{h}_\eta(k_1, \dots, k_\eta) + \rho_\eta(k_1, \dots, k_\eta) \Delta h_\eta(k_1, \dots, k_\eta) \right\} \quad (5.6)$$

onde:

- $\bar{h}_\eta(k_1, \dots, k_\eta)$ é o valor central do conjunto de kernels de Volterra;
- $\Delta h_\eta(k_1, \dots, k_\eta)$ representa o módulo da máxima variação dos kernels em torno de $\bar{h}_\eta(k_1, \dots, k_\eta)$;
- $\rho_\eta(k_1, \dots, k_\eta)$ representa uma incerteza sobre $h_\eta(k_1, \dots, k_\eta)$ satisfazendo $-1 \leq \rho_\eta(k_1, \dots, k_\eta) \leq 1$.

Os coeficientes do desenvolvimento ortonormal do modelo com incertezas são então calculados por meio da equação (2.16), i.e.:

$$\alpha_{i_1, \dots, i_\eta} = \sum_{k_1=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} h_\eta(k_1, \dots, k_\eta) \prod_{l=1}^{\eta} \psi_{i_l}(k_l)$$

Usando o conjunto dado em (5.6) para reescrever a equação acima, tem-se o seguinte conjunto de coeficientes incertos:

$$\begin{aligned} \{\alpha_{i_1, \dots, i_\eta}\} &= \sum_{k_1=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} \{h_\eta(k_1, \dots, k_\eta)\} \prod_{l=1}^{\eta} \psi_{i_l}(k_l) \\ &= \sum_{k_1=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} \left\{ \bar{h}_\eta(k_1, \dots, k_\eta) + \rho_\eta(k_1, \dots, k_\eta) \Delta h_\eta(k_1, \dots, k_\eta) \right\} \prod_{l=1}^{\eta} \psi_{i_l}(k_l) \end{aligned} \quad (5.7)$$

Uma estimativa das incertezas sobre os coeficientes da expansão ortonormal no caso de modelos não-lineares pode ser obtida através de uma desigualdade similar àquela em (5.4). Dado que $\rho_\eta(k_1, \dots, k_\eta)$ está contido no intervalo $[-1, 1]$ e considerando que seja válida uma extensão da condição dada pela equação (5.3), i.e. $\prod_{l=1}^{\eta} \psi_{i_l}(k_l) > 0 \quad \forall k_l, \quad i_l = 1, \dots, M$, pode-se assim escrever a desigualdade a seguir:

$$\begin{aligned} - \sum_{k_1=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} \Delta h_\eta(k_1, \dots, k_\eta) \prod_{l=1}^{\eta} \psi_{i_l}(k_l) &\leq \\ \sum_{k_1=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} \rho_\eta(k_1, \dots, k_\eta) \Delta h_\eta(k_1, \dots, k_\eta) \prod_{l=1}^{\eta} \psi_{i_l}(k_l) &\leq \end{aligned} \quad (5.8)$$

$$\sum_{k_1=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} \Delta h_\eta(k_1, \dots, k_\eta) \prod_{l=1}^{\eta} \psi_{i_l}(k_l)$$

Os extremos da desigualdade em (5.8) são então tomados como a máxima incerteza (inferior e superior) sobre os coeficientes da expansão em base ortonormal do modelo não-linear. Desta forma, estes coeficientes também podem ser representados por meio de intervalos de pertinência. A parcela nominal destes intervalos é dada em termos dos kernels de Volterra centrais $\bar{h}_\eta(k_1, \dots, k_\eta)$, enquanto que a parcela incerta é obtida em termos de Δh_η , como a seguir:

$$\bar{\alpha}_{i_1, \dots, i_\eta} = \sum_{k_1=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} \bar{h}_\eta(k_1, \dots, k_\eta) \prod_{l=1}^{\eta} \psi_{i_l}(k_l) \quad (5.9)$$

$$\Delta \alpha_{i_1, \dots, i_\eta} = \sum_{k_1=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} \Delta h_\eta(k_1, \dots, k_\eta) \prod_{l=1}^{\eta} \psi_{i_l}(k_l) \quad (5.10)$$

Assim, pode-se escrever o seguinte conjunto:

$$\{\alpha_{i_1, \dots, i_\eta}\} = \{\bar{\alpha}_{i_1, \dots, i_\eta} + \sigma_{i_1, \dots, i_\eta} \cdot \Delta \alpha_{i_1, \dots, i_\eta}\} \quad (5.11)$$

onde:

- $\bar{\alpha}_{i_1, \dots, i_\eta}$ são os valores centrais dos coeficientes da expansão em base ortonormal;
- $\Delta \alpha_{i_1, \dots, i_\eta}$ representa um desvio dos coeficientes em torno dos respectivos valores centrais;
- $\sigma_{i_1, \dots, i_\eta}$ representa um fator de escalamento sobre $\Delta \alpha_{i_1, \dots, i_\eta}$ satisfazendo $-1 \leq \sigma_{i_1, \dots, i_\eta} \leq 1$.

O desenvolvimento apresentado acima permite que a incerteza de um conjunto de kernels do sistema seja mapeada em intervalos de pertinência dos coeficientes da expansão ortonormal. Entretanto, dado que $\Delta h_\eta(k_1, \dots, k_\eta) \geq 0 \quad \forall k_l, \quad l \in \{1, \dots, \eta\}$, as desigualdades dadas por (5.4) e (5.8) são sempre verdadeiras somente se a condição (5.3) for satisfeita. Quando esta não é satisfeita, não se pode assegurar que o modelo com incertezas represente todos os kernels do processo incerto. Esta é uma condição necessária do problema de controle preditivo robusto, uma vez que o sinal de entrada é calculado considerando-se o pior desempenho possível do sinal de referência, que é determinado utilizando-se o modelo. Para contornar estes problemas, na seção a seguir propõe-se condições adicionais para assegurar que o modelo ortonormal com incertezas estruturadas seguramente represente todos os comportamentos possíveis do processo.

5.2 Cálculo de limitantes de incertezas em modelos OBF

Uma série de problemas de otimização são formulados nesta seção para o cálculo dos limitantes de incertezas em modelos dinâmicos baseados em funções ortonormais. Este cálculo é feito através de uma sobre-estimação dos kernels de Volterra. Apresenta-se inicialmente a formulação do problema no contexto dos sistemas lineares, que foi abordado em um trabalho anterior (Moreira 2006). Posteriormente, esta formulação é estendida para o contexto dos modelos não-lineares de Volterra.

5.2.1 Modelos lineares

Em Moreira (2006), propôs-se resolver um problema de otimização por meio de uma sobre-estimação do conjunto de respostas ao impulso do processo, como apresentado a seguir.

$$\begin{aligned} \min_{\sigma_m} \quad & 2^M \prod_{m=1}^M \sigma_m \Delta \alpha_m \\ \text{sujeito a} \quad & \left\{ \begin{array}{l} \{h_1(k)\} \subseteq \{\check{h}_1(k)\} \\ \check{h}_1(k) = \sum_{m=1}^M [\bar{\alpha}_m + \tilde{\sigma}_m \Delta \alpha_m] \psi_m(k), \\ |\tilde{\sigma}_m| \leq \sigma_m, \quad m = 1, \dots, M \\ \bar{\alpha}_m = \sum_{k=0}^{\infty} \bar{h}_1(k) \psi_m(k) \\ \Delta \alpha_m = \sum_{k=0}^{\infty} \rho_1(k) \Delta h_1(k) \psi_m(k) \end{array} \right. \end{aligned} \quad (5.12)$$

O critério de custo do problema (5.12) é o volume do hiper-retângulo formado pelo prolongamento da incerteza sobre os coeficientes da expansão ortonormal, i.e., $\sigma_m \Delta \alpha_m$. Neste problema, o modelo ortonormal é inicialmente determinado conforme proposto em trabalhos anteriores (Oliveira 1997, Oliveira *et al.* 2000). As parcelas incertas dos coeficientes do modelo, ou seja $\Delta \alpha_m$ ($m = 1, 2, \dots, M$), são então escalonadas pelos escalares σ_m minimizando-se o volume do hiper-retângulo que delimita as incertezas até que todas as respostas ao impulso do processo, representadas pelo conjunto $\{h_1(k)\}$, ainda estejam contidas no conjunto de respostas ao impulso do modelo $\{\check{h}_1(k)\}$. As restrições do problema asseguram que o modelo obtido contém todas as respostas ao impulso do processo.

As variáveis do problema de otimização (5.12) são os coeficientes σ_m ($m = 1, \dots, M$). Para o critério minimizado, estas variáveis correspondem ao

escalamento de menor volume do hiper-retângulo gerado pelas incertezas, tal que qualquer elemento do conjunto de respostas ao impulso do processo esteja representado no modelo. Tem-se assim que:

$$\alpha_{m,\min} \leq \alpha_m \leq \alpha_{m,\max} \quad (5.13)$$

com:

$$\begin{cases} \alpha_{m,\min} &= \bar{\alpha}_m - \sigma_{m,\text{otimo}} \cdot \Delta\alpha_m \\ \alpha_{m,\max} &= \bar{\alpha}_m + \sigma_{m,\text{otimo}} \cdot \Delta\alpha_m \end{cases} \quad (5.14)$$

onde $\sigma_{m,\text{otimo}}$ são as soluções do problema (5.12).

5.2.2 Modelos não-lineares

Uma extensão do problema formulado na seção anterior para sistemas lineares é proposto a seguir no contexto dos modelos não-lineares.

Considere portanto um modelo OBF-Volterra de ordem η , como aquele em (2.22). As incertezas sobre os coeficientes do modelo são representadas pelos termos $\sigma_{i_1} \cdot \Delta\alpha_{i_1}$, $\sigma_{i_1,i_2} \cdot \Delta\alpha_{i_1,i_2}$ e $\sigma_{i_1,\dots,i_\eta} \cdot \Delta\alpha_{i_1,\dots,i_\eta}$ para a primeira, segunda e η -ésima ordens do modelo, respectivamente. Supondo que todos os η kernels do modelo são independentes entre si, ou seja, que não existe uma interdependência entre h_1, h_2, \dots, h_η , deve-se resolver η problemas separadamente, como a seguir.

Kernel de primeira ordem:

$$\begin{aligned} \min_{\sigma_{i_1}} \quad & \prod_{i_1=1}^M \sigma_{i_1} \cdot \Delta\alpha_{i_1} \\ \text{sujeito a} \quad & \left\{ \begin{aligned} \{h_1(k_1)\} &\subseteq \{\check{h}_1(k_1)\} \\ \check{h}_1(k_1) &= \sum_{i_1=1}^M [\bar{\alpha}_{i_1} + \tilde{\sigma}_{i_1} \Delta\alpha_{i_1}] \psi_{i_1}(k_1), \\ &|\tilde{\sigma}_{i_1}| \leq \sigma_{i_1}, \quad i_1 = 1, \dots, M \\ \bar{\alpha}_{i_1} &= \sum_{k_1=0}^{\infty} \bar{h}_1(k_1) \psi_{i_1}(k_1) \\ \Delta\alpha_{i_1} &= \sum_{k_1=0}^{\infty} \Delta h_1(k_1) \psi_{i_1}(k_1) \end{aligned} \right. \end{aligned} \quad (5.15)$$

Kernel de segunda ordem:

$$\begin{aligned} \min_{\sigma_{i_1, i_2}} \quad & \prod_{i_1=1}^M \prod_{i_2=1}^M \sigma_{i_1, i_2} \cdot \Delta \alpha_{i_1, i_2} \\ \text{sujeito a} \quad & \left\{ \begin{aligned} \{h_2(k_1, k_2)\} &\subseteq \{\check{h}_2(k_1, k_2)\} \\ \check{h}_2(k_1, k_2) &= \sum_{i_1=1}^M \sum_{i_2=1}^M [\bar{\alpha}_{i_1, i_2} + \tilde{\sigma}_{i_1, i_2} \Delta \alpha_{i_1, i_2}] \psi_{i_1}(k_1) \psi_{i_2}(k_2), \\ &|\tilde{\sigma}_{i_1, i_2}| \leq \sigma_{i_1, i_2}, \quad i_1, i_2 = 1, \dots, M \\ \bar{\alpha}_{i_1, i_2} &= \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \bar{h}_2(k_1, k_2) \psi_{i_1}(k_1) \psi_{i_2}(k_2) \\ \Delta \alpha_{i_1, i_2} &= \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \Delta h_2(k_1, k_2) \psi_{i_1}(k_1) \psi_{i_2}(k_2) \end{aligned} \right. \quad (5.16) \end{aligned}$$

Kernel de η -ésima ordem:

$$\begin{aligned} \min_{\sigma_{i_1, \dots, i_\eta}} \quad & \prod_{i_1=1}^M \cdots \prod_{i_\eta=1}^M \sigma_{i_1, \dots, i_\eta} \cdot \Delta \alpha_{i_1, \dots, i_\eta} \\ \text{sujeito a} \quad & \left\{ \begin{aligned} \{h_\eta(k_1, \dots, k_\eta)\} &\subseteq \{\check{h}_\eta(k_1, \dots, k_\eta)\} \\ \check{h}_\eta(k_1, \dots, k_\eta) &= \sum_{i_1=1}^M \cdots \sum_{i_\eta=1}^M [\bar{\alpha}_{i_1, \dots, i_\eta} + \tilde{\sigma}_{i_1, \dots, i_\eta} \Delta \alpha_{i_1, \dots, i_\eta}] \prod_{l=1}^{\eta} \psi_{i_l}(k_l), \\ &|\tilde{\sigma}_{i_1, \dots, i_\eta}| \leq \sigma_{i_1, \dots, i_\eta}, \quad i_l = 1, \dots, M \\ \bar{\alpha}_{i_1, \dots, i_\eta} &= \sum_{k_1=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} \bar{h}_\eta(k_1, \dots, k_\eta) \prod_{l=1}^{\eta} \psi_{i_l}(k_l) \\ \Delta \alpha_{i_1, \dots, i_\eta} &= \sum_{k_1=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} \Delta h_\eta(k_1, \dots, k_\eta) \prod_{l=1}^{\eta} \psi_{i_l}(k_l) \end{aligned} \right. \quad (5.17) \end{aligned}$$

A interpretação para a solução dos problemas acima é minimizar os volumes gerados pelo escalamento das incertezas sobre os coeficientes $\alpha_{(\cdot)}$ do modelo ortonormal. De acordo com a metodologia proposta, estas incertezas são independentes entre si. Em um modelo de Volterra de ordem η , a solução dos η problemas minimiza os volumes gerados pelas incertezas, até que todos os kernels do processo, representados pelos conjuntos $\{h_l(k_1, \dots, k_l)\}$, $l = 1, 2, \dots, \eta$, estejam contidos no conjunto de kernels de Volterra do modelo $\{h_l(k_1, \dots, k_l)\}$.

As variáveis dos problemas de otimização (5.15), (5.16) e (5.17) são os coeficientes σ_{i_1} , σ_{i_1, i_2} e $\sigma_{i_1, \dots, i_\eta}$, respectivamente. Para o critério minimizado, estas variáveis correspondem ao escalamento dos lados do hiper-retângulo das incertezas tal que qualquer kernel do processo esteja contido no conjunto dos kernels factíveis do modelo. A solução numérica fornece o escalamento necessário nos lados do hiper-retângulo para atender as especificações impostas pelas restrições do problema.

5.3 Exemplo ilustrativo

O objetivo deste exemplo é obter os limites das incertezas em modelos OBF-Volterra por meio da formulação proposta na seção 5.2. Suponha um sistema não-linear que possui os seguintes kernels de Volterra incertos de primeira e segunda ordem:

$$h_1(k_1) = \mathcal{Z}^{-1} \left[\frac{z + 0,5}{z^2 - 1, 2z + r_1} \right] \quad (5.18)$$

$$h_2(k_1, k_2) = g_{i_1}(k_1)g_{i_2}(k_2) \quad \text{onde} \quad g_{i_l}(k) = \mathcal{Z}^{-1} \left[\frac{z + 1}{z^2 - r_{i_l}z + 0,5} \right] \quad (5.19)$$

Note que as incertezas sobre os kernels h_1 e h_2 são dadas em função de parâmetros incertos do processo, aqui representados por r_1 e (r_{i_1}, r_{i_2}) , respectivamente. Considere que estes parâmetros estejam restritos aos seguintes intervalos: $r_1 \in [0, 3; 0, 5]$ e $(r_{i_1}, r_{i_2}) \in [0, 4; 0, 8] \times [0, 4; 0, 8]$. Para efeitos de simulação computacional, estes intervalos foram discretizados com uma taxa de amostragem igual a 0,01. Tem-se assim:

$$\begin{cases} r_1(j) = 0,3 + j \times 0,01 & j = 0, 1, \dots, 20 \\ r_{i_1}(j') = 0,4 + j' \times 0,01 & j' = 0, 1, \dots, 40 \\ r_{i_2}(j'') = 0,4 + j'' \times 0,01 & j'' = 0, 1, \dots, 40 \end{cases}$$

Os valores centrais dos kernels de primeira e segunda ordem, i.e. $\bar{h}_1(k_1)$ e $\bar{h}_2(k_1, k_2)$, são computados como uma média de todos os kernels factíveis de $h_1(k_1)$ e $h_2(k_1, k_2)$, respectivamente. Em outras palavras, \bar{h}_1 é a média temporal de todas as possíveis funções dadas em (5.18), ao se variar o parâmetro r_1 dentro de seu intervalo de factibilidade (Moreira 2006). Este raciocínio também vale para \bar{h}_2 e (r_{i_1}, r_{i_2}) em (5.19). Assim, os kernels centrais são obtidos como segue:

$$\bar{h}_1(k_1) = \frac{1}{21} \sum_{j=1}^{21} h_1(k_1, r_1(j)) \quad (5.20)$$

$$\bar{h}_2(k_1, k_2) = \frac{1}{41 \times 41} \sum_{j'=1}^{41} \sum_{j''=1}^{41} h_2(k_1, k_2, r_{i_1}(j'), r_{i_2}(j'')) \quad (5.21)$$

A Figura 5.1 apresenta o conjunto de kernels de Volterra de primeira ordem $\{h_1(k_1)\}$, obtido para cada valor do parâmetro incerto r_1 em (5.18). Na mesma figura, mostra-se também os limites inferior e superior deste conjunto. A Figura 5.2 ilustra os limites inferior e superior do conjunto de kernels de segunda ordem $\{h_2(k_1, k_2)\}$ gerados por (5.19).

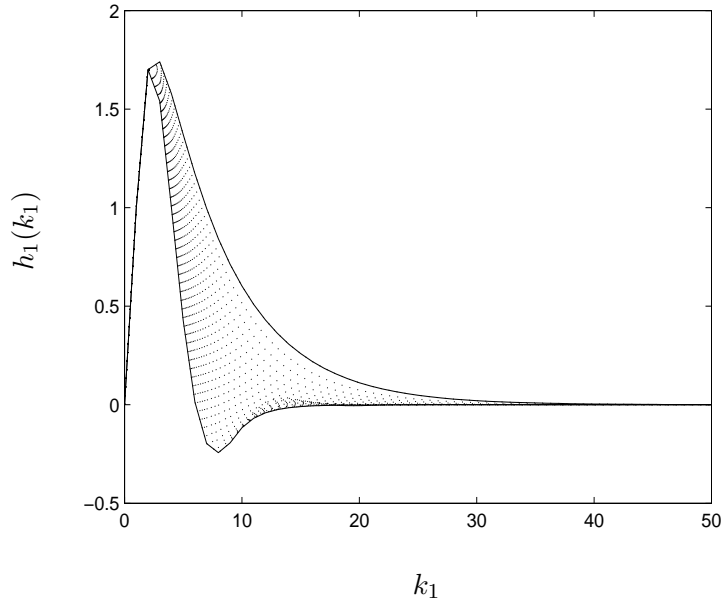


Figura 5.1: Conjunto de kernels de primeira ordem gerado por (5.18) (linhas pontilhadas), mostrado juntamente com os limites inferior e superior deste conjunto (linhas contínuas).

Dado que a base de Kautz representa o enfoque maior deste trabalho, ela será considerada como o conjunto de funções ortonormais a ser utilizado no modelo OBF incerto. Além disso, como os kernels em (5.18) e (5.19) possuem comportamento oscilatório, por serem regidos por dinâmicas sub-amortecidas, é mais conveniente a utilização de uma base de funções com pólos complexos, como a base de Kautz. Neste exemplo, são utilizadas $M = 8$ funções de Kautz. O valor central dos coeficientes da expansão ortonormal do modelo incerto e a máxima variação sobre este valor central são calculados usando as equações (5.9) e (5.10), respectivamente. Como já comentado anteriormente, estas equações representam uma extensão de uma abordagem original (Oliveira 1997, Oliveira *et al.* 2000) para o contexto dos modelos de Volterra. Usando os pólos ótimos dos kernels centrais $\bar{h}_1(k)$ e $\bar{h}_2(k_1, k_2)$,

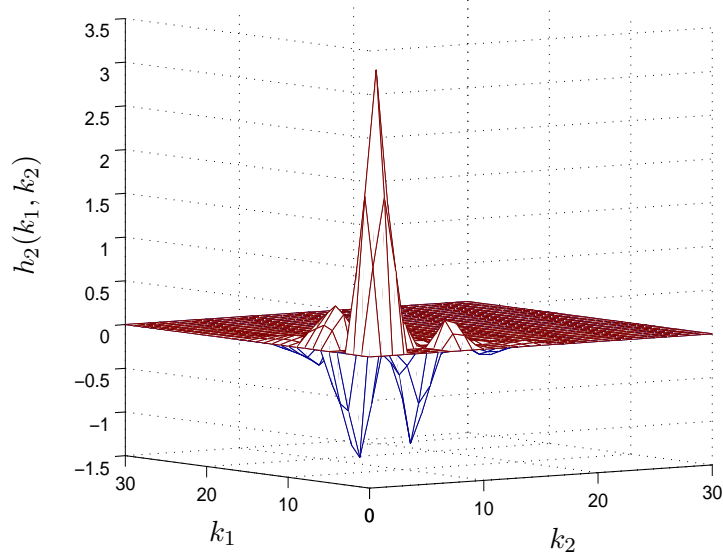


Figura 5.2: Limites inferior (azul) e superior (vermelho) do conjunto de kernels de segunda ordem $\{h_2(k_1, k_2)\}$ gerado por (5.19).

os coeficientes de suas expansões ortonormais são dados por:

$$\bar{\alpha}_{i_1} = \begin{bmatrix} -0,2078 \\ 2,9920 \\ 0,7334 \\ -0,0084 \\ 0,1158 \\ 0,1251 \\ 0,0299 \\ -0,0053 \end{bmatrix}$$

$$\bar{\alpha}_{i_1, i_2} =$$

$$\begin{bmatrix} 0,2141 & -0,5882 & -0,7033 & -0,1818 & -0,0676 & 0,0768 & 0,0041 & 0,0029 \\ -0,5882 & 1,6163 & 1,9326 & 0,4997 & 0,1857 & -0,2111 & -0,0113 & -0,0079 \\ -0,7033 & 1,9326 & 2,3107 & 0,5974 & 0,2220 & -0,2524 & -0,0136 & -0,0095 \\ -0,1818 & 0,4997 & 0,5974 & 0,1545 & 0,0574 & -0,0653 & -0,0035 & -0,0024 \\ -0,0676 & 0,1857 & 0,2220 & 0,0574 & 0,0213 & -0,0242 & -0,0013 & -0,0009 \\ 0,0768 & -0,2111 & -0,2524 & -0,0653 & -0,0242 & 0,0276 & 0,0015 & 0,0010 \\ 0,0041 & -0,0113 & -0,0136 & -0,0035 & -0,0013 & 0,0015 & 0,0001 & 0,0001 \\ 0,0029 & -0,0079 & -0,0095 & -0,0024 & -0,0009 & 0,0010 & 0,0001 & 0,0000 \end{bmatrix}$$

O valor de $\Delta h_1(k)$ correspondente ao kernel de primeira ordem em (5.18) é calculado como o módulo da máxima variação sobre o kernel central $\bar{h}_1(k)$, ao se variar o valor do parâmetro incerto r_1 dentro de seu intervalo de factibilidade. Esta mesma ideia vale também para o kernel de segunda ordem em (5.19). Tem-se portanto que:

$$\Delta h_1(k_1) = |\max\{h_1(k_1, r_1) - \bar{h}_1(k_1)\}| \quad (5.22)$$

$$\Delta h_2(k_1, k_2) = |\max\{h_2(k_1, k_2, r_{i_1}, r_{i_2}) - \bar{h}_2(k_1, k_2)\}| \quad (5.23)$$

As Figuras 5.3 e 5.4 ilustram os desvios $\Delta h_1(k_1)$ e $\Delta h_2(k_1, k_2)$, respectivamente.

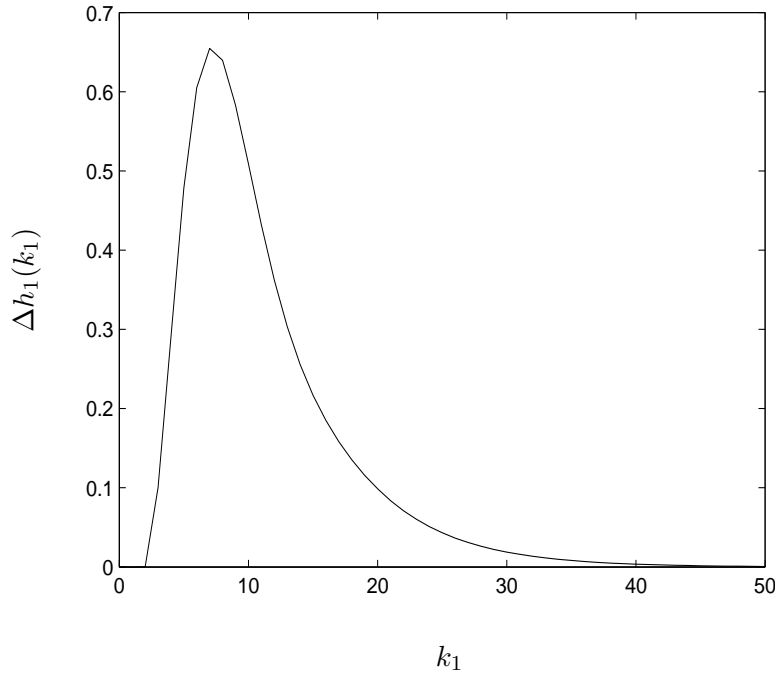


Figura 5.3: Máxima variação sobre o kernel de primeira ordem central $\bar{h}_1(k_1)$.

Obtém-se agora os valores ótimos das incertezas σ_{i_1} e σ_{i_1, i_2} usando a metodologia proposta na seção 5.2, através das soluções dos problemas de otimização (5.15) e (5.16), respectivamente. As soluções obtidas são dadas por:

$$\sigma_{i_1, \text{otimo}} = \begin{bmatrix} 0,7714 \\ 0,3474 \\ 0,5102 \\ 0,1147 \\ 0,0404 \\ 0,0288 \\ 0,3865 \\ 0,5512 \end{bmatrix} \quad (5.24)$$

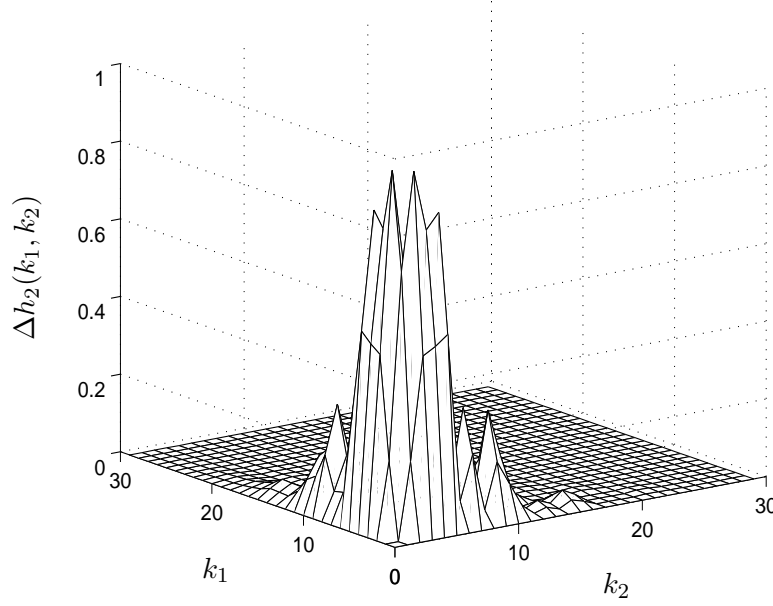


Figura 5.4: Máxima variação sobre o kernel de segunda ordem central $\bar{h}_2(k_1, k_2)$.

$$\sigma_{i_1, i_2, \text{otimo}} = \begin{bmatrix} 0.8180 & 0.8385 & 0.7948 & 0.8757 & 0.2844 & 0.4329 & 0.7833 & 0.4154 \\ 0.6602 & 0.5681 & 0.9568 & 0.7373 & 0.4692 & 0.2259 & 0.6808 & 0.3050 \\ 0.3420 & 0.3704 & 0.5226 & 0.1365 & 0.0648 & 0.5798 & 0.4611 & 0.8744 \\ 0.2897 & 0.7027 & 0.8801 & 0.0118 & 0.9883 & 0.7604 & 0.5678 & 0.0150 \\ 0.3412 & 0.5466 & 0.1730 & 0.8939 & 0.5828 & 0.5298 & 0.7942 & 0.7680 \\ 0.5341 & 0.4449 & 0.9797 & 0.1991 & 0.4235 & 0.6405 & 0.0592 & 0.9708 \\ 0.7271 & 0.6946 & 0.2714 & 0.2987 & 0.5155 & 0.2091 & 0.6029 & 0.9901 \\ 0.3093 & 0.6213 & 0.2523 & 0.6614 & 0.3340 & 0.3798 & 0.0503 & 0.7889 \end{bmatrix} \quad (5.25)$$

A validade destas soluções pode ser verificada observando se os limites dos kernels obtidos usando a metodologia proposta abrangem os kernels do processo, que são dados por (5.18) e (5.19). A Figura 5.5 mostra os limites inferior e superior do kernel de primeira ordem usando duas diferentes abordagens. Para obter uma melhor visualização gráfica, as Figuras 5.6 e 5.7 mostram um *zoom* em diferentes regiões do intervalo de tempo observado.

Conforme pode ser visto na Figura 5.7, a extensão de um método anterior da

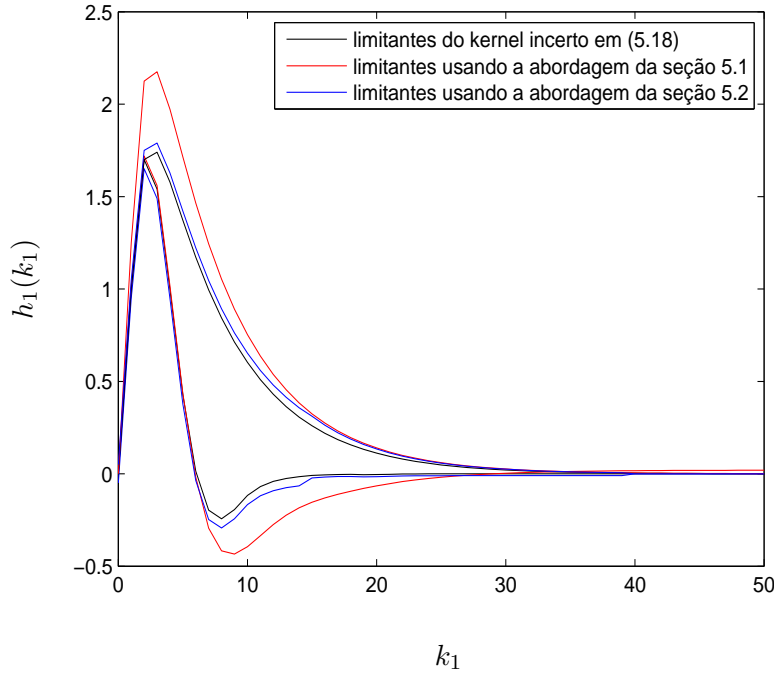


Figura 5.5: Comparação entre os limitantes inferior e superior do kernel incerto de primeira ordem calculados utilizando a abordagem da seção 5.1 e a abordagem da seção 5.2.

literatura (Oliveira 1997, Oliveira *et al.* 2000) (seção 5.1) não foi capaz de representar os limites de h_1 em todo o intervalo de tempo observado. A partir aproximadamente do instante de tempo $k_1 = 29$, observa-se claramente que o limitante inferior do modelo do kernel obtido não contém o limitante inferior do processo. A explicação para este fato é porque a condição (5.3) não foi satisfeita, como pode ser verificado numericamente. A formulação proposta (seção 5.2) mostrou ser mais eficiente para o cálculo dos limites das incertezas, uma vez que os limitantes de h_1 calculados usando a solução obtida para $\sigma_{i_1, \text{otimo}}$ contêm os limitantes inferior e superior do processo. Matematicamente, tem-se que $\{h_1(k_1)\} \subseteq \{\check{h}_1(k_1)\}$, atendendo portanto as especificações impostas pelas restrições do problema de otimização (5.15).

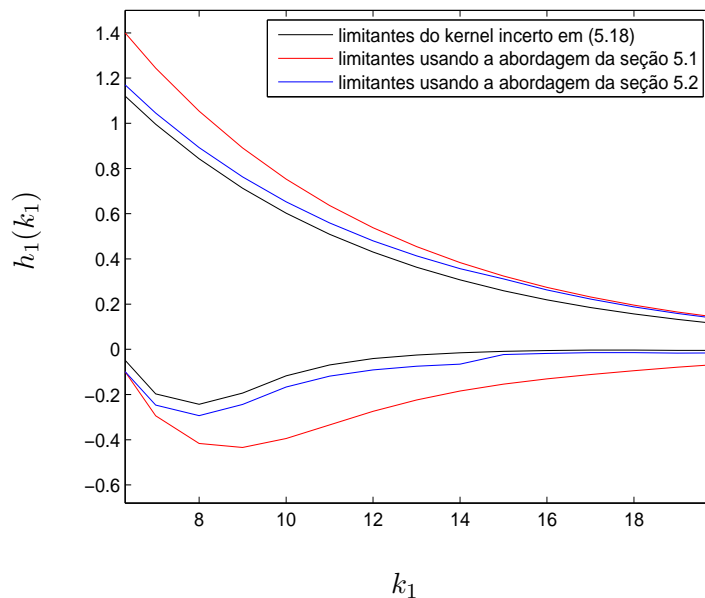


Figura 5.6: *Zoom* na região $6 \leq k_1 \leq 20$.

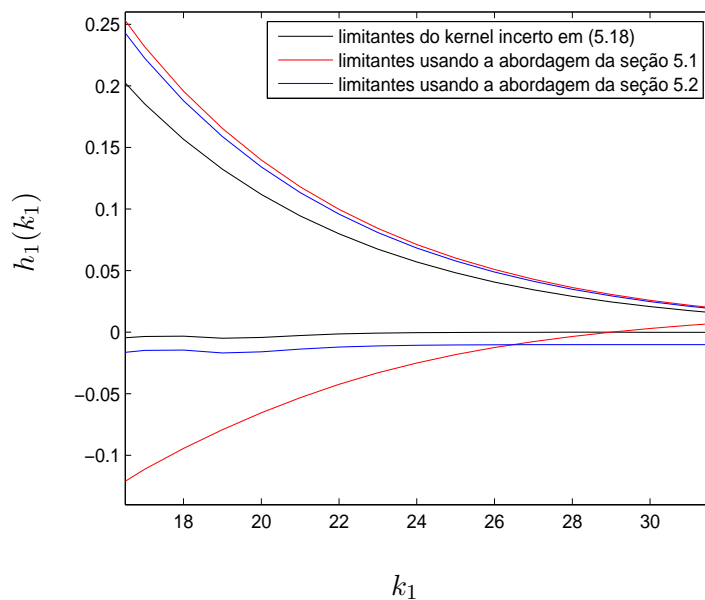


Figura 5.7: *Zoom* na região $16 \leq k_1 \leq 32$.

A Figura 5.8 apresenta uma comparação correspondente aos limites obtidos para o kernel de segunda ordem, onde foi feita uma projeção no plano $k_1 = 10$. Observa-

se que a extensão do método anterior proposto na literatura (Oliveira 1997, Oliveira *et al.* 2000) (seção 5.1) não foi capaz de representar os limites de h_2 em todo o intervalo de tempo observado. A explicação para este fato é porque a extensão da condição (5.3), i.e. $\psi_{i_1}(k_1)\psi_{i_2}(k_2) > 0 \quad \forall k_1, k_2, \quad i_1, i_2 = 1, \dots, M$ não foi satisfeita. Para obter uma melhor visualização gráfica, a Figura 5.9 mostra um *zoom* em um intervalo de tempo observado.

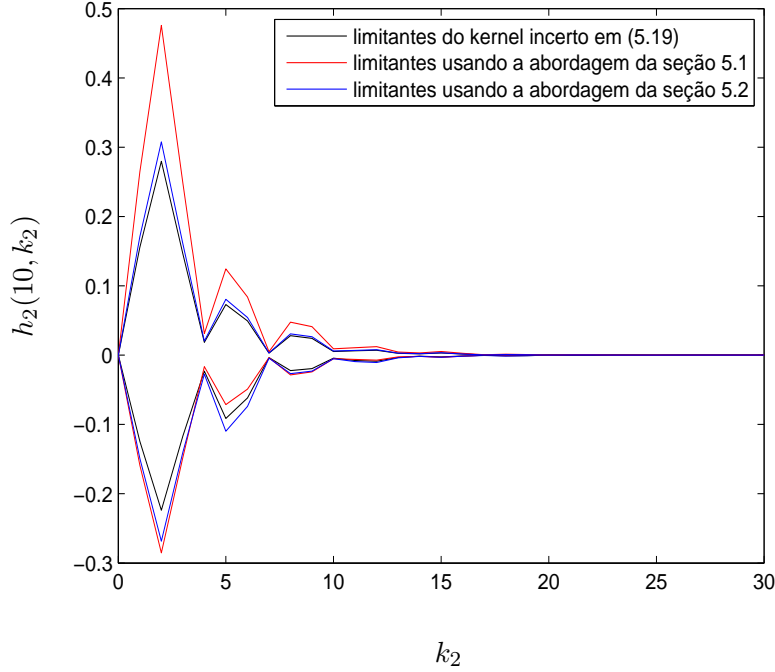


Figura 5.8: Comparação entre os limitantes inferior e superior do kernel incerto de segunda ordem calculados utilizando a abordagem da seção 5.1 e a abordagem da seção 5.2.

Além da comparação visual apresentada nas Figuras 5.8 e 5.9, também verificou-se numericamente que o modelo obtido a partir da solução $\sigma_{i_1, i_2, \text{otimo}}$ em (5.25) consegue representar os kernels do processo dados por (5.19). Matematicamente, tem-se que $\{h_2(k_1, k_2)\} \subseteq \{\hat{h}_2(k_1, k_2)\}$, atendendo portanto as especificações impostas pelas restrições do problema de otimização (5.16).

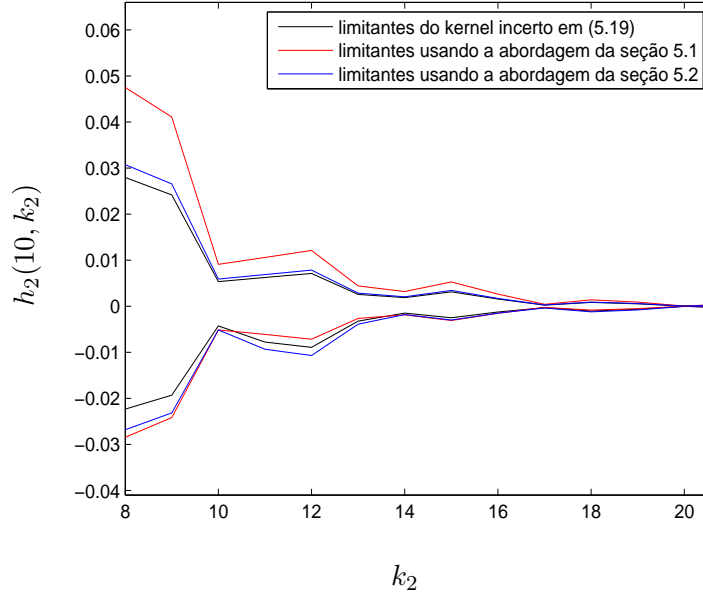


Figura 5.9: *Zoom* na região $8 \leq k_2 \leq 20$.

5.4 Resumo e contribuições

Este capítulo discutiu a modelagem de sistemas dinâmicos incertos baseada em funções ortonormais. Abordou-se este problema no contexto dos modelos não-lineares de Volterra em que os intervalos de pertinência dos coeficientes do modelo OBF, que representam os limites da incerteza, foram estimados utilizando um conjunto de kernels de Volterra.

O problema central, porém, apresenta a desvantagem de não assegurar que o modelo com incertezas represente todos os kernels do processo incerto. Esta desvantagem foi posteriormente contornada propondo-se condições adicionais para assegurar que o modelo ortonormal com incertezas estruturadas represente todos os comportamentos possíveis do processo. As condições adicionais foram formuladas por meio de uma série de problemas de otimização, que permitem obter o cálculo das incertezas através de uma sobre-estimação dos kernels de Volterra. Estas condições são úteis em problemas de controle preditivo robusto, em que o sinal de entrada é calculado considerando-se o pior desempenho possível do sinal de referência. Um exemplo simulado não-linear foi apresentado para ilustrar o método proposto, sendo aplicado na modelagem de kernels de Volterra de primeira e de segunda ordem.

Para utilizar a estratégia apresentada neste capítulo como base para uma abordagem em termos de incertezas sobre o sinal de saída, a formulação envolve considerar as restrições do problema de otimização correspondente à uma determinada ordem do modelo de Volterra como uma restrição para os problemas correspondentes

às ordens superiores. Esta formulação permite sempre garantir que todas as ordens do modelo são consideradas simultaneamente na solução do problema.

Capítulo 6

Conclusões, Contribuições e Perspectivas Futuras

6.1 Conclusões e contribuições

Há algumas décadas, o campo de identificação e modelagem de sistemas não-lineares era uma área com abordagens aplicáveis a uma classe restrita de sistemas. Com o surgimento de técnicas modernas de otimização, uma classe maior de sistemas tem sido estudada. Embora uma importante característica dos sistemas não-lineares seja o fato de que quase todo sistema é único, inúmeras ferramentas têm sido desenvolvidas permitindo o uso da mesma abordagem para uma grande variedade de sistemas. Nos anos mais recentes, as áreas de identificação e controle de sistemas dinâmicos não-lineares têm recebido um maior interesse em engenharia. O estudo destas áreas tem crescido em importância, levando a necessidade de modelos mais detalhados para descrever a dinâmica dos processos físicos. Por estas razões, este trabalho abordou a representação de sistemas não-lineares por meio de modelos de Volterra, que são comumente utilizados para a representação de processos não-lineares. A parametrização destes modelos é atrativa porque eles possuem uma estrutura linear nos parâmetros, permitindo assim utilizar técnicas de regressão linear para estimar os parâmetros do modelo.

Os modelos de Volterra representam uma generalização dos modelos lineares de resposta ao impulso para a descrição de sistemas dinâmicos não-lineares. Estes modelos são capazes de aproximar arbitrariamente bem os sistemas com memória finita. Contudo, eles geralmente necessitam de um número elevado de parâmetros para representar os kernels de Volterra. Esta desvantagem pode ser reduzida adotando-se uma representação do tipo Wiener, que possui uma dinâmica linear constituída por um conjunto de filtros de funções ortonormais, seguida de um mapeamento estático não-linear. No modelo resultante, conhecido como modelo OBF-Volterra, os kernels são desenvolvidos utilizando-se uma expansão em bases de funções ortonormais. A utilização de funções ortonormais permite uma redução na ordem da representação do modelo, simplificando os problemas de identificação e controle associados. Uma

outra vantagem da abordagem com bases ortonormais é que ela possibilita incorporar conhecimento prévio sobre a dinâmica do sistema no processo de identificação. Desta forma, o problema de estimação dos coeficientes dos modelos se beneficia de uma rápida taxa de convergência, devido ao número reduzido de coeficientes a serem determinados para modelar o sistema. Por sua vez, os modelos de OBF-Volterra podem ter sua complexidade reduzida ao se fazer uma seleção ótima da base de funções ortonormais, o que permite uma redução no grau de sua representação. O estudo da escolha de pólos das bases ortonormais representa a principal motivação para sua aplicação na modelagem e controle de sistemas dinâmicos. O cálculo de pólos ótimos das funções ortonormais é o tema central abordado no presente trabalho.

Uma das contribuições deste trabalho foi a obtenção de uma expressão analítica para o limitante superior do erro de aproximação truncada dos kernels de Volterra quando estes são representados em bases de funções ortonormais de Kautz. Na abordagem apresentada, utiliza-se um conjunto de funções de Kautz independentes para direção axial do kernel de Volterra. A minimização deste limitante permitiu obter uma solução analítica para a seleção ótima de um dos parâmetros de Kautz em modelos de OBF-Volterra. Esta solução, por sua vez, pode ser interpretada como uma expressão que fornece o pólo sub-ótimo de Kautz, e representa uma extensão de um resultado prévio cuja solução utiliza pólos iguais para todas as direções do kernel.

Por meio de um exemplo ilustrativo, verificou-se que o uso de bases independentes para representar cada direção dos kernels reduz o erro de truncamento quando o kernel apresenta diferentes dinâmicas em suas direções. Resultados de simulação mostraram que o método proposto fornece uma boa aproximação de sistemas não-lineares com comportamento oscilatório. A seleção ótima dos pólos permite obter uma boa qualidade de aproximação utilizando um número reduzido de funções da base de Kautz, o que significa uma redução na ordem de representação do modelo de Volterra resultante. A metodologia proposta também foi aplicada para otimização dos pólos de Kautz na modelagem de um sistema não-linear de levitação magnética com forte comportamento oscilatório. O modelo OBF-Volterra de segunda ordem obtido conseguiu representar adequadamente as dinâmicas do sistema.

Outra contribuição foi a proposta de uma nova estratégia de otimização numérica de bases ortonormais usadas na aproximação de sistemas dinâmicos. Esta estratégia, que é válida para modelos lineares e também para não-lineares de Volterra de qualquer ordem, é baseada no cálculo das expressões analíticas para as derivadas da saída dos filtros ortonormais com relação aos pólos da base. Estas expressões fornecem direções de busca exatas a serem utilizadas por um algoritmo de otimização para obter os pólos ótimos das bases ortonormais. A solução para o problema principal envolve a minimização de uma função de custo quadrática que leva em consideração o erro de estimação da saída do sistema. Particular ênfase foi dada para a base de Kautz, embora as expressões relativas à base de Laguerre podem ser obtidas de maneira similar. Um estudo equivalente com modelos baseados em funções ortonormais generalizadas (GOBF) também foi apresentado.

A principal vantagem dessa metodologia é que os gradientes são calculados analiticamente usando somente dados entrada/saída medidos a partir do sistema a ser modelado. Não é necessário ter informação prévia sobre os kernels de Volterra do sistema, uma vez que os kernels do sistema podem ser recuperados diretamente a partir de expansões na base ortonormal adotada. Por meio de simulações computacionais, verificou-se que a aplicação da metodologia proposta forneceu resultados satisfatórios na modelagem de sistemas dinâmicos lineares e não-lineares. Este método surge assim como uma ferramenta promissora nas áreas de identificação e controle de sistemas baseados em funções ortonormais.

Este trabalho abordou ainda a modelagem de sistemas não-lineares incertos usando bases de funções ortonormais. Os intervalos de pertinência dos coeficientes do modelo OBF-Volterra, que representam os limites da incerteza, foram estimados utilizando-se um conjunto de kernels de Volterra. A contribuição central neste sentido foi a proposta de condições adicionais em relação a formulação original para assegurar que o modelo ortonormal com incertezas estruturadas represente todos os comportamentos possíveis do processo incerto. Estas condições foram formuladas por meio de uma série de problemas de otimização, que permitem obter o cálculo das incertezas através de uma sobre-estimação dos kernels de Volterra.

6.2 Perspectivas futuras

Tendo em vista as contribuições apresentadas neste trabalho, as perspectivas para estudos futuros estão baseadas na seleção de pólos das funções ortonormais, em especial das bases de Kautz e GOBF. Neste contexto, está o cálculo de pólos ótimos para representação de sistemas não-lineares por meio de modelos OBF-Volterra, através de abordagens analíticas e numéricas.

De forma mais detalhada, uma perspectiva para trabalho futuro está relacionada com a otimização analítica dos pólos de bases GOBF, que é um problema ainda em aberto na literatura. No contexto da base de Kautz, por exemplo, ao invés da otimização de um limitante superior do erro de aproximação dos kernels de Volterra usando-se bases de Kautz, uma linha de pesquisa envolveria a otimização do próprio erro de aproximação dos kernels. Para isso, sugere-se o estudo de outras normas para o cálculo do erro. Neste trabalho, foi adotada a norma-2, que representa uma soma quadrática em cada componente do erro de aproximação. A utilização de outras normas, como a norma- ∞ e a norma-1, pode fornecer resultados mais interessantes, por serem calculados como funções de máximo e mínimo.

Uma outra linha de pesquisa a ser abordada no futuro envolve o estudo de diferentes níveis de recursão temporal utilizados no algoritmo *back-propagation-through-time*. Neste trabalho, estas recursões foram tomadas até que as condições do modelo no instante de tempo inicial $k = 0$ fossem atingidas. Uma abordagem consiste em adotar diferentes truncamentos dessas recursões, até instantes de tempo em que os gradientes calculados não mais contribuem de forma significativa para a re-

apresentação do modelo. Espera-se com isso reduzir o custo computacional associado com o problema.

Outra perspectiva para pesquisas futuras consiste também na extensão dos resultados obtidos no Capítulo 4 com relação a utilização de pólos distintos para cada direção dos kernels de Volterra, similarmente ao estudo apresentado no Capítulo 3. As expressões analíticas para os gradientes da saída dos filtros ortonormais foram obtidas adotando-se uma única base para todas as direções de cada kernel. Portanto, uma abordagem a ser investigada envolve a obtenção de gradientes para a representação de modelos OBF-Volterra com bases independentes (múltiplas) em cada direção.

A metodologia proposta no Capítulo 5 permite calcular os limitantes de incertezas em modelos de Volterra baseados em funções ortonormais quando se dispõe de informações sobre as incertezas nos kernels de Volterra. Quando se dispõe de informações sobre as incertezas no sinal de saída, deve-se utilizar outra abordagem. Assim, uma perspectiva para ser estudada no futuro envolve abordar um problema em que as incertezas estão sobre o sinal de saída, usando métodos de identificação robusta baseados na abordagem *Unknown-But-Bounded-Error* (UBBE) (Milanese e Tempo 1985, Vicino e Milanese 1991, Akçay e Ninness 1998). A abordagem UBBE consiste em determinar um região no espaço de parâmetros que seja consistente com as medidas, com a estrutura do modelo e com os limites do erro. As incertezas sobre os parâmetros são caracterizadas por intervalos de possíveis valores destes parâmetros, que são estimados a partir dos valores mínimo e máximo do intervalo de incerteza.

Perspectivas para trabalhos futuros envolvem ainda o estudo de modelos de Volterra com ruído nos sinais de entrada e saída, e também o estudo de modelos incluindo kernels de terceira ordem.

Apêndice A

Provas e Demonstrações

Este apêndice destina-se a apresentar as demonstrações de teoremas e de algumas equações, que foram omitidas no corpo principal do trabalho, por conveniência.

A.1 Demonstração da equação (3.2)

Inicialmente, usando a equação (2.14), pode-se escrever o seguinte:

$$\begin{aligned} & \sum_{k_1=0}^{\infty} \cdots \sum_{k_{\eta}=0}^{\infty} h_{\eta}^2(k_1, \dots, k_{\eta}) = \\ &= \sum_{k_1=0}^{\infty} \cdots \sum_{k_{\eta}=0}^{\infty} \left(\sum_{i_1=1}^{\infty} \cdots \sum_{i_{\eta}=1}^{\infty} \alpha_{i_1, \dots, i_{\eta}} \prod_{l=1}^{\eta} \psi_{l, i_l}(k_l) \right)^2 \\ &= \sum_{k_1=0}^{\infty} \cdots \sum_{k_{\eta}=0}^{\infty} \left[\left(\sum_{i_1=1}^{\infty} \cdots \sum_{i_{\eta}=1}^{\infty} \alpha_{i_1, \dots, i_{\eta}} \prod_{l=1}^{\eta} \psi_{l, i_l}(k_l) \right) \cdot \left(\sum_{i'_1=1}^{\infty} \cdots \sum_{i'_{\eta}=1}^{\infty} \alpha_{i'_1, \dots, i'_{\eta}} \prod_{l'=1}^{\eta} \psi_{l', i_{l'}}(k_{l'}) \right) \right] \\ &= \sum_{i_1=1}^{\infty} \cdots \sum_{i_{\eta}=1}^{\infty} \sum_{i'_1=1}^{\infty} \cdots \sum_{i'_{\eta}=1}^{\infty} (\alpha_{i_1, \dots, i_{\eta}} \alpha_{i'_1, \dots, i'_{\eta}}) \left(\sum_{k_1=0}^{\infty} \cdots \sum_{k_{\eta}=0}^{\infty} \prod_{l=1}^{\eta} \psi_{l, i_l}(k_l) \prod_{l'=1}^{\eta} \psi_{l', i_{l'}}(k_{l'}) \right) \quad (\text{A.1}) \end{aligned}$$

Utilizando a propriedade de ortonormalidade das funções $\psi_{l, i_l}(k_l)$, isto é:

$$\sum_{k=0}^{\infty} \psi_{l, q}(k) \psi_{l, r}(k) = \begin{cases} 0 & \text{se } q \neq r \\ 1 & \text{se } q = r \end{cases}$$

a equação (A.1) é então reescrita como:

$$\sum_{k_1=0}^{\infty} \cdots \sum_{k_{\eta}=0}^{\infty} h_{\eta}^2(k_1, \dots, k_{\eta}) = \sum_{i_1=1}^{\infty} \cdots \sum_{i_{\eta}=1}^{\infty} \alpha_{i_1, \dots, i_{\eta}}^2$$

□

A.2 Demonstração do Teorema 3.1.1

Considere uma função $x(k_1, \dots, k_\eta)$ que é nula para todo $k_l < 0$ ($l = 1, 2, \dots, \eta$), e absolutamente somável em $[0, \infty)$. Sejam $\alpha_{i_1, \dots, i_\eta}$ os coeficientes da expansão do η -ésimo kernel h_η usando base de Kautz, e as constantes reais c_l representando um dos parâmetros de Kautz. É possível mostrar então que a seguinte relação é válida (da Rosa *et al.* 2008a):

$$\begin{aligned}
 & -2c_l \sum_{k_1=0}^{\infty} \cdots \sum_{k_l=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} k_l x(k_1, \dots, k_l - 1, \dots, k_\eta) x(k_1, \dots, k_l, \dots, k_\eta) \\
 & + (1 + c_l^2) \sum_{k_1=0}^{\infty} \cdots \sum_{k_l=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} k_l x^2(k_1, \dots, k_l, \dots, k_\eta) \\
 & + \sum_{k_1=0}^{\infty} \cdots \sum_{k_l=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} x^2(k_1, \dots, k_l, \dots, k_\eta) = \\
 & = (1 - c_l^2) \sum_{i_1=1}^{\infty} \cdots \sum_{i_l=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} i_l \alpha_{i_1, \dots, i_l, \dots, i_\eta}^2
 \end{aligned} \tag{A.2}$$

Usando as expressões para $\mu_{1,l}(x)$, $\mu_{2,l}(x)$, $\mu_{3,l}(x)$ dadas nas equações (3.10), (3.11) e (3.12), respectivamente, pode-se escrever ainda que:

$$\sum_{i_1=1}^{\infty} \cdots \sum_{i_l=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} i_l \alpha_{i_1, \dots, i_l, \dots, i_\eta}^2 = \frac{-2c_l \mu_{1,l}(x) + (1 + c_l^2) \mu_{2,l}(x) + \eta \mu_{3,l}(x)}{1 - c_l^2} \tag{A.3}$$

$l = 1, \dots, \eta$

Somando em l cada a equação em (A.3), chega-se a:

$$\sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} (i_1 + \cdots + i_\eta) \alpha_{i_1, \dots, i_\eta}^2 = \sum_{l=1}^{\eta} \left[\frac{-2c_l \mu_{1,l}(x) + (1 + c_l^2) \mu_{2,l}(x) + \eta \mu_{3,l}(x)}{1 - c_l^2} \right] \tag{A.4}$$

Considere a expansão em base de Laguerre das funções $g_{\text{par},l}$ e $g_{\text{impar},l}$, definidas em (3.4) e (3.5), com coeficientes $\gamma_{i_1, \dots, i_\eta}$ e $\sigma_{i_1, \dots, i_\eta}$, respectivamente, tais que:

$$g_{\text{par},l}(k_1, \dots, k_\eta) = \sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} \gamma_{i_1, \dots, i_\eta} \prod_{l=1}^{\eta} \phi_{l,i_l}(k_l) \tag{A.5}$$

$$g_{\text{impar},l}(k_1, \dots, k_\eta) = \sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} \sigma_{i_1, \dots, i_\eta} \prod_{l=1}^{\eta} \phi_{l,i_l}(k_l) \tag{A.6}$$

Comparando as equações (3.4) e (3.5) com (A.5) e (A.6), observa-se que os coeficientes $\alpha_{(\cdot)}$ da expansão do kernel $h_\eta(k_1, \dots, k_\eta)$ usando funções de Kautz estão relacionados com os coeficientes das expansões acima como $\gamma_{i_1, \dots, i_\eta} = \alpha_{2i_1, \dots, 2i_\eta}$ e $\sigma_{i_1, \dots, i_\eta} = \alpha_{2i_1-1, \dots, 2i_\eta-1}$, $\forall i_1, \dots, i_\eta$.

Denotando M como o número de termos da base de Kautz utilizados na aproximação do kernel h_η , de acordo com a equação (2.21), o limitante superior dado por (3.6) pode ser deduzido a partir da seguinte desigualdade (Campello *et al.* 2004):

$$\begin{aligned}
 \eta(M+1) \sum_{i_1=M+1}^{\infty} \cdots \sum_{i_\eta=M+1}^{\infty} \alpha_{i_1, \dots, i_\eta}^2 &\leq \sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} (i_1 + \cdots + i_\eta) \alpha_{i_1, \dots, i_\eta}^2 \\
 &= \sum_{i_1 \text{ par}}^{\infty} \cdots \sum_{i_\eta \text{ par}}^{\infty} (i_1 + \cdots + i_\eta) \alpha_{i_1, \dots, i_\eta}^2 + \\
 &+ \sum_{i_1 \text{ impar}}^{\infty} \cdots \sum_{i_\eta \text{ impar}}^{\infty} (i_1 + \cdots + i_\eta) \alpha_{i_1, \dots, i_\eta}^2 \\
 &= \sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} (2i_1 + \cdots + 2i_\eta) \alpha_{2i_1, \dots, 2i_\eta}^2 + \\
 &+ \sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} [(2i_1 - 1) + \cdots + (2i_\eta - 1)] \alpha_{2i_1-1, \dots, 2i_\eta-1}^2 \\
 &\leq 2 \sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} (i_1 + \cdots + i_\eta) \gamma_{i_1, \dots, i_\eta}^2 + \\
 &+ 2 \sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} (i_1 + \cdots + i_\eta) \sigma_{i_1, \dots, i_\eta}^2 \tag{A.7}
 \end{aligned}$$

O erro (quadrático) de aproximação do kernel $h_\eta(k_1, \dots, k_\eta)$ aparece no lado esquerdo da equação (A.7) multiplicado por $\eta(M+1)\|h_\eta\|^2$. Então, dividindo (A.7) por $\eta(M+1)\|h_\eta\|^2$ e usando (A.4) resulta em:

$$\begin{aligned}
 \text{EQN} &\leq 2 \sum_{l=1}^{\eta} \left[\frac{-2c_l \mu_{1,l}(g_{\text{par},l}) + (1+c_l^2) \mu_{2,l}(g_{\text{par},l}) + \mu_{3,l}(g_{\text{par},l})}{\eta(M+1)\|h_\eta\|^2(1-c_l^2)} \right] \\
 &+ 2 \sum_{l=1}^{\eta} \left[\frac{-2c_l \mu_{1,l}(g_{\text{impar},l}) + (1+c_l^2) \mu_{2,l}(g_{\text{impar},l}) + \mu_{3,l}(g_{\text{impar},l})}{\eta(M+1)\|h_\eta\|^2(1-c_l^2)} \right] \tag{A.8}
 \end{aligned}$$

Finalmente, com o uso das equações (3.7)–(3.9) na desigualdade (A.8), chega-se a:

$$\text{EQN} \leq \frac{2}{\eta(M+1)\|h_\eta\|^2} \sum_{l=1}^{\eta} \left[\frac{m_{2,l}c_l^2 - 2m_{1,l}c_l + m_{3,l}}{1-c_l^2} \right]$$

□

A.3 Demonstração da factibilidade da solução (3.17)

A prova da factibilidade da solução (3.17) para o problema de otimização em (3.13) será feita para o caso em que $\xi_l > 1$. A demonstração para $\xi_l < -1$ pode ser obtida de maneira análoga. Portanto se $\xi_l > 1$, tem-se que:

$$\begin{aligned}\xi_l &> 1 \\ -2\xi_l &< -2 \\ \xi_l^2 + 1 - 2\xi_l &< \xi_l^2 + 1 - 2 \\ (\xi_l - 1)^2 &< \xi_l^2 - 1 \\ |\xi_l - 1| &< \sqrt{\xi_l^2 - 1}\end{aligned}\tag{A.9}$$

Como $\xi_l > 1$ implica que $|\xi_l - 1| = \xi_l - 1$, pode-se escrever ainda:

$$\begin{aligned}|\xi_l - 1| &< \sqrt{\xi_l^2 - 1} \\ \xi_l - 1 &< \sqrt{\xi_l^2 - 1} \\ \xi_l - \sqrt{\xi_l^2 - 1} &< 1\end{aligned}\tag{A.10}$$

É possível escrever também $\xi_l^2 > \xi_l^2 - 1$, implicando que $|\xi_l| > \sqrt{\xi_l^2 - 1}$. Logo:

$$\begin{aligned}|\xi_l| &> \sqrt{\xi_l^2 - 1} \\ \xi_l &> \sqrt{\xi_l^2 - 1} \\ \xi_l - \sqrt{\xi_l^2 - 1} &> 0\end{aligned}\tag{A.11}$$

Agrupando as condições dadas em (A.10) e (A.11), conclui-se finalmente que $|c_{\text{otimo},l}| < 1$, $\forall l \in \{1, \dots, \eta\}$. Em outras palavras, a equação (3.17) é solução factível para o problema (3.13). \square

A.4 Demonstração do Teorema 3.1.2

Para demonstrar que a condição $-1 \leq \xi_l \leq 1$ é infactível, parte-se da seguinte relação, que é válida para $\forall l = \{1, \dots, \eta\}$:

$$\sum_{k_1=0}^{\infty} \cdots \sum_{k_l=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} k_l \left[x(k_1, \dots, k_l, \dots, k_\eta) \pm x(k_1, \dots, k_l - 1, \dots, k_\eta) \right]^2 > 0 \tag{A.12}$$

$l = 1, \dots, \eta$

A partir da equação (A.12), pode-se inferir o seguinte:

$$\begin{aligned}
& \sum_{k_1=0}^{\infty} \cdots \sum_{k_l=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} k_l x^2(k_1, \dots, k_l, \dots, k_\eta) \\
& \pm 2 \sum_{k_1=0}^{\infty} \cdots \sum_{k_l=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} k_l x(k_1, \dots, k_l, \dots, k_\eta) x(k_1, \dots, k_l - 1, \dots, k_\eta) \quad (\text{A.13}) \\
& + \sum_{k_1=0}^{\infty} \cdots \sum_{k_l=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} k_l x^2(k_1, \dots, k_l - 1, \dots, k_\eta) > 0
\end{aligned}$$

Equivalentemente, pode-se escrever também:

$$\begin{aligned}
& \sum_{k_1=0}^{\infty} \cdots \sum_{k_l=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} k_l x^2(k_1, \dots, k_l, \dots, k_\eta) \\
& \pm 2 \sum_{k_1=0}^{\infty} \cdots \sum_{k_l=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} k_l x(k_1, \dots, k_l, \dots, k_\eta) x(k_1, \dots, k_l - 1, \dots, k_\eta) \\
& + \sum_{k_1=0}^{\infty} \cdots \sum_{k_l=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} k_l x^2(k_1, \dots, k_l, \dots, k_\eta) \\
& - \sum_{k_1=0}^{\infty} \cdots \sum_{k_l=1}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} (k_l - 1) x^2(k_1, \dots, k_l - 1, \dots, k_\eta) \quad \left. \vphantom{\sum_{k_1=0}^{\infty}} \right\} \text{termos iguais} \\
& + \sum_{k_1=0}^{\infty} \cdots \sum_{k_l=1}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} k_l x^2(k_1, \dots, k_l - 1, \dots, k_\eta) > 0
\end{aligned}$$

Ou ainda:

$$\begin{aligned}
& 2 \sum_{k_1=0}^{\infty} \cdots \sum_{k_l=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} k_l x^2(k_1, \dots, k_l, \dots, k_\eta) \\
& \pm 2 \sum_{k_1=0}^{\infty} \cdots \sum_{k_l=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} k_l x(k_1, \dots, k_l, \dots, k_\eta) x(k_1, \dots, k_l - 1, \dots, k_\eta) \quad (\text{A.14}) \\
& + \sum_{k_1=0}^{\infty} \cdots \sum_{k_l=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} x^2(k_1, \dots, k_l, \dots, k_\eta) > 0
\end{aligned}$$

Usando as expressões para os momentos definidos no Teorema 3.1.1, a equação acima pode ser escrita como segue:

$$2\mu_{2,l}(x) \pm 2\mu_{1,l}(x) + \eta\mu_{3,l}(x) > 0 \quad (\text{A.15})$$

Na equação (A.15), faça primeiramente $x = g_{\text{par},l}$ e depois $x = g_{\text{impar},l}$ para obter as duas relações a seguir:

$$2\mu_{2,l}(g_{\text{par},l}) \pm 2\mu_{1,l}(g_{\text{par},l}) + \eta\mu_{3,l}(g_{\text{par},l}) > 0 \quad (\text{A.16})$$

$$2\mu_{2,l}(g_{\text{impar},l}) \pm 2\mu_{1,l}(g_{\text{impar},l}) + \eta\mu_{3,l}(g_{\text{impar},l}) > 0 \quad (\text{A.17})$$

Somando as desigualdades (A.16) e (A.17), tem-se que:

$$2\mu_{2,l}(g_{\text{par},l}) + 2\mu_{2,l}(g_{\text{impar},l}) + \eta\mu_{3,l}(g_{\text{par},l}) + \eta\mu_{3,l}(g_{\text{impar},l}) > \\ \pm [2\mu_{1,l}(g_{\text{par},l}) + 2\mu_{1,l}(g_{\text{impar},l})]$$

o que permite concluir o seguinte:

$$\left| \frac{2\mu_{2,l}(g_{\text{par},l}) + 2\mu_{2,l}(g_{\text{impar},l}) + \eta\mu_{3,l}(g_{\text{par},l}) + \eta\mu_{3,l}(g_{\text{impar},l})}{2\mu_{1,l}(g_{\text{par},l}) + 2\mu_{1,l}(g_{\text{impar},l})} \right| > 1 \quad (\text{A.18})$$

Finalmente, substituindo as relações dadas em (3.7), (3.8) e (3.9) em (A.18), obtém-se:

$$\left| \frac{m_{2,l} + m_{3,l}}{2m_{1,l}} \right| = |\xi_l| > 1 \quad (\text{A.19})$$

□

A.5 Demonstração do Teorema 3.1.4

A demonstração apresentada aqui contempla uma extensão, para casos mais gerais, de teoremas semelhantes já existentes na literatura (Tanguy *et al.* 2002, da Rosa *et al.* 2005, da Rosa *et al.* 2007, da Rosa *et al.* 2008a). Muitos dos argumentos e fundamentos matemáticos usados para sua demonstração podem ser encontrados nesses trabalhos. Apresenta-se somente a demonstração para a função $g_{\text{par},l}(k_1, \dots, k_\eta)$, uma vez que a prova para $g_{\text{impar},l}(k_1, \dots, k_\eta)$ pode ser obtida de forma análoga.

Inicialmente, utilizando (2.16) para reescrever a equação (3.4), leva a:

$$\begin{aligned} g_{\text{par},l}(k_1, \dots, k_\eta) &= \sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} \alpha_{2i_1, \dots, 2i_\eta} \prod_{l=1}^{\eta} \phi_{l,i_l}(k_l) \\ &= \sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} \left(\sum_{\tau_1=0}^{\infty} \cdots \sum_{\tau_\eta=0}^{\infty} h_\eta(\tau_1, \dots, \tau_\eta) \prod_{l=1}^{\eta} \psi_{l,2i_l}(\tau_l) \right) \prod_{l=1}^{\eta} \phi_{l,i_l}(k_l) \\ &= \sum_{\tau_1=0}^{\infty} \cdots \sum_{\tau_\eta=0}^{\infty} h_\eta(\tau_1, \dots, \tau_\eta) \left(\sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} \prod_{l=1}^{\eta} \psi_{l,2i_l}(\tau_l) \phi_{l,i_l}(k_l) \right) \end{aligned} \quad (\text{A.20})$$

Então, tomando a transformada Z com relação a τ_l (para $l = 1, 2, \dots, \eta$) do termo entre parênteses acima resulta em:

$$\begin{aligned}
 & \mathcal{Z} \left[\sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} \prod_{l=1}^{\eta} \psi_{l,2i_l}(\tau_l) \phi_{l,i_l}(k_l) \right] = \\
 &= \sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} \left\{ \prod_{l=1}^{\eta} \left[\frac{\sqrt{(1-c_l^2)(1-b_l^2)}}{z_l^2 + b_l(c_l-1)z_l - c_l} z_l \cdot \left(\frac{-c_l z_l^2 + b_l(c_l-1)z_l + 1}{z_l^2 + b_l(c_l-1)z_l - c_l} \right)^{i_l-1} \phi_{l,i_l}(k_l) \right] \right\} \\
 &= \left[\prod_{l=1}^{\eta} \frac{\sqrt{(1-c_l^2)(1-b_l^2)}}{z_l^2 + b_l(c_l-1)z_l - c_l} z_l \right] \cdot \left[\sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} \prod_{l=1}^{\eta} \phi_{l,i_l}(k_l) w_l^{1-i_l} \right] \quad (\text{A.21})
 \end{aligned}$$

onde a seguinte simplificação foi feita:

$$w_l \triangleq \frac{z_l^2 + b_l(c_l-1)z_l - c_l}{-c_l z_l^2 + b_l(c_l-1)z_l + 1} \quad l = 1, \dots, \eta \quad (\text{A.22})$$

O último termo em (A.21) pode então ser reescrito como segue:

$$\begin{aligned}
 \sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} \prod_{l=1}^{\eta} \phi_{l,i_l}(k_l) w_l^{1-i_l} &= \prod_{l=1}^{\eta} \left(w_l \sum_{i_l=1}^{\infty} \phi_{l,i_l}(k_l) w_l^{-i_l} \right) \\
 &= \prod_{l=1}^{\eta} \left[\frac{w_l \sqrt{1-c_l^2}}{w_l + c_l} \left(\frac{1 + c_l w_l}{w_l + c_l} \right)^{k_l} \right] \quad (\text{A.23})
 \end{aligned}$$

após ser usada a seguinte relação válida para as funções de Laguerre $\phi_{l,n}(k)$ (Tanguy *et al.* 2002):

$$\sum_{n=1}^{\infty} \phi_{l,n}(k) w_l^{-n} = \frac{\sqrt{1-c_l^2}}{w_l + c_l} \left(\frac{1 + c_l w_l}{w_l + c_l} \right)^k$$

Substituindo a equação (A.22) em (A.23), e usando a equação resultante para reescrever (A.21), chega-se após algumas manipulações algébricas à seguinte relação:

$$\begin{aligned}
 \mathcal{Z} \left[\sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} \prod_{l=1}^{\eta} \psi_{l,2i_l}(\tau_l) \phi_{l,i_l}(k_l) \right] &= \prod_{l=1}^{\eta} \left[\frac{\sqrt{1-b_l^2}}{z_l - b_l} \left(\frac{1 - b_l z_l}{z_l - b_l} \right)^{k_l} z_l^{-k_l} \right] \\
 &= \prod_{l=1}^{\eta} \mathcal{Z} \left\{ [\psi_{l,2(k_l+1)}(\tau_l)]_{c_l=0} \right\} \\
 &= \mathcal{Z} \left\{ \prod_{l=1}^{\eta} [\psi_{l,2(k_l+1)}(\tau_l)]_{c_l=0} \right\} \quad (\text{A.24})
 \end{aligned}$$

A equação (A.24) pode ser reescrita como:

$$\sum_{i_1=1}^{\infty} \cdots \sum_{i_{\eta}=1}^{\infty} \prod_{l=1}^{\eta} \psi_{l,2i_l}(\tau_l) \phi_{l,i_l}(k_l) = \prod_{l=1}^{\eta} \hat{\psi}_{l,2(k_l+1)}(\tau_l) \quad (\text{A.25})$$

onde $\hat{\psi}_{l,n}(\tau_l) \triangleq [\psi_{l,n}(\tau_l)]_{c_l=0}$. Finalmente, substituindo (A.25) em (A.20), obtém-se o resultado desejado, completando a prova.

Apêndice B

Análise do Custo Computacional do Método Proposto no Capítulo 4

Este apêndice apresenta uma análise de complexidade computacional do método proposto em termos de tempo de execução assintótico. Para o leitor não familiar com análise assintótica, recomenda-se (Cormen *et al.* 2002). É importante mencionar que as direções de busca geradas pelo método proposto podem ser utilizadas conjuntamente com qualquer algoritmo de otimização baseado em gradiente, não apenas aquele sugerido e adotado nos experimentos do Capítulo 4 (Levenberg-Marquardt). Por esta razão, as discussões a seguir restringem-se à complexidade computacional dos procedimentos para cálculo dessas direções de busca (gradientes) para os diferentes modelos investigados, a saber, modelos OBF lineares e de Volterra com funções de Kautz e GOBF.

B.1 Modelo de Kautz linear

No caso do modelo OBF linear com funções de Kautz, as variáveis possivelmente críticas em termos de tempo de execução são o número de amostras de entrada e saída, N_d , e o número de funções ortonormais na base, M . A análise de complexidade do método proposto consiste em avaliar o custo computacional assintótico dos seus passos com relação a essas variáveis. Os passos são aqueles necessários para calcular as equações (4.10), (4.11), (4.22), (4.33), (4.35) e (4.39) (vide algoritmo na Seção 4.2.3 do Capítulo 4).

(I) De acordo com as respectivas definições das matrizes \mathcal{A} , \mathcal{B} , \mathcal{C} , $\partial\mathcal{A}/\partial b$, $\partial\mathcal{A}/\partial c$, $\partial\mathcal{B}/\partial b$, $\partial\mathcal{B}/\partial c$, \mathcal{B}_1 , \mathcal{B}_2 , \mathcal{B}_3 , \mathcal{B}_4 , \mathcal{B}_5 , $\partial\mathcal{B}_1/\partial c$, $\partial\mathcal{B}_2/\partial b$, $\partial\mathcal{B}_2/\partial c$, $\partial\mathcal{B}_4/\partial c$, $\partial\mathcal{B}_5/\partial b$, $\partial\mathcal{B}_5/\partial c$ nas Seções 4.2.1 e 4.2.2, é evidente que nenhuma dessas matrizes depende de qualquer das variáveis críticas mencionadas anteriormente. Logo, o cálculo de todas essas matrizes é feito em tempo constante com relação às variáveis críticas da análise, ou seja, em tempo $O(1)$.

(II) O custo de calcular \mathcal{A}^k em (4.25) ou (4.26) é determinado pelo custo de calcular os termos β^k . Como esse cálculo pode ser feito incrementalmente para

$k = 1, \dots, N_d$, tem-se um custo total de ordem $O(N_d)$.

(III) Cada multiplicação matricial em (4.23) e (4.24) é feita em tempo constante, mas a somatória faz com que o cálculo de $\partial \mathcal{A}^k / \partial b$ e $\partial \mathcal{A}^k / \partial c$ para $k = 1, \dots, N_d$ demande tempo $O(1 + 2 + \dots + N_d)$, ou seja, $O(N_d \cdot (N_d + 1)/2)$, que é $O(N_d^2)$.

(IV) É trivial mostrar que, utilizando a representação da dinâmica de Kautz na forma de espaço de estados, o cálculo das saídas dos M filtros de Kautz em um dado instante de tempo k , $\tilde{\psi}_i(k)$ ($i = 1, \dots, M$), pode ser feito em tempo $O(M^2)$. Como é preciso calcular as saídas dos filtros para todo $k = 1, \dots, N_d$, tem-se $O(M^2 N_d)$. Uma vez que se dispõe das saídas dos filtros, pode-se calcular a saída do modelo, $\hat{y}(k)$, para todo k , utilizando a equação (4.8), em tempo adicional $O(M N_d)$. Logo, o cálculo das saídas dos filtros e da saída do modelo para todos os instantes de amostragem demanda custo $O(M^2 N_d + M N_d)$, que é $O(M^2 N_d)$.

(V) Uma vez que todos os termos mencionados nos itens anteriores já estejam calculados, as equações (4.22), (4.33), (4.35) e (4.39) podem ser computadas. Cada uma demanda, para m e k dados (fixos), um número proporcional a k conjuntos de operações algébricas, cada um dos quais passíveis de realização em tempo constante, o que implica tempo $O(k)$. Mas, para um dado m , é preciso calcular tais equações para $k = 1, \dots, N_d$, o que resulta $O(1 + 2 + \dots + N_d)$, ou seja, $O(N_d^2)$. Por fim, é preciso realizar esses cálculos para cada um dos filtros, ou seja, $m = 1, \dots, M$, o que implica um tempo total da ordem de $O(M N_d^2)$.

(VI) Seja n_p a dimensão do vetor de parâmetros \mathbf{p} em (4.10). O tempo computacional para cálculo de todas as n_p componentes escalares de $\nabla_{\mathbf{p}} J$ em (4.10) é da ordem de $O(n_p M N_d)$. Porém, para a base de Kautz tem-se $n_p = 2$ ($\mathbf{p} = [b \quad c]^T$), o que implica $O(M N_d)$.

(VII) O tempo requerido para cálculo de cada uma das M componentes escalares do vetor gradiente $\nabla_{\alpha} J$ em (4.11) é de ordem $O(N_d)$, o que resulta em tempo $O(M N_d)$ para cálculo de todo o gradiente.

Dos itens (I) a (VII) acima tem-se que o limitante superior assintótico para o tempo computacional de cálculo das direções de busca para otimização de modelos lineares de Kautz segundo o método proposto é $O(M^2 N_d + N_d^2 M)$. No entanto, na prática tem-se usualmente $N_d \gg M$, o que permite afirmar que a demanda computacional para o cálculo de tais direções é $O(N_d^2)$.

B.2 Modelo GOBF linear

A única diferença quando se usa GOBF ao invés de Kautz é que a dimensão do vetor de parâmetros \mathbf{p} em (4.10) não é mais igual a dois, mas é tal que $n_p \propto M$. Isso afeta apenas o item (VI) da análise anterior para o caso Kautz, que passa a ser $O(M^2 N_d)$ quando se utiliza GOBF. Note que isso, no entanto, não afeta a complexidade final do método, que continua sendo $O(M^2 N_d + N_d^2 M)$, ou simplesmente $O(N_d^2)$ dada a hipótese prática de que o número de amostras dos sinais de entrada e saída do sistema usados para a otimização do modelo é muito maior que o número de funções

ortonormais ($N_d \gg M$).

B.3 Modelos de Volterra

No caso mais geral onde uma base de funções ortonormais independente é usada para a expansão ortonormal de cada um dos N kernels de Volterra, tem-se que os cálculos descritos nos itens (I), (II), (III) e (V) da análise anterior (caso linear) precisarão ser realizados N vezes, uma vez para cada kernel. Isso implica um tempo computacional de ordem $O(NMN_d^2)$.

No que se refere ao item (IV) da análise anterior, tem-se que o cálculo das saídas dos M filtros de cada uma das bases ortonormais, para $k = 1, \dots, N_d$, continua sendo feito em tempo $O(M^2 N_d)$, mas, como são N bases independentes, tem-se no caso de Volterra um tempo total de $O(NM^2 N_d)$. O tempo requerido para cálculo da saída do modelo também não permanece o mesmo. De fato, para um modelo de Volterra de ordem N descrito pela equação (2.22), tem-se que cada um dos termos de ordem η – i.e. $\hat{y}_\eta(k)$ em (4.81) – demanda tempo $O(M^\eta)$ para ser calculado. Logo, calcular o conjunto de $\eta = 1, \dots, N$ termos que compõem a saída $\hat{y}(k)$, para $k = 1, \dots, N_d$, demanda tempo $O(N_d(M + M^2 + \dots + M^N))$, ou seja, $O(NM^N N_d)$. Por outro lado, o cálculo de cada uma das n_p componentes do gradiente $\nabla_{\mathbf{p}} \hat{y}_\eta(k)$ em (4.82), para η e k fixos, demanda tempo $O(\eta M^\eta)$. Logo, as n_p componentes desse gradiente, para $k = 1, \dots, N_d$ e $\eta = 1, \dots, N$, demandam tempo $O(n_p(M + 2M^2 + \dots + NM^N) N_d)$, que pode ser superestimado como $O(n_p N^2 M^N N_d)$. Uma vez que esse gradiente esteja calculado, o cálculo do gradiente $\nabla_{\mathbf{p}} J$ em (4.83) demanda apenas um tempo adicional de ordem $O(n_p N N_d)$. Finalmente, cada uma das n_α componentes do termo $\nabla_{\alpha} \hat{y}_\eta(k)$ em (4.84), para k e η fixos, pode ser facilmente calculada a partir de (4.81) em tempo $O(\eta)$, o que implica tempo $O(n_\alpha(1 + 2 + \dots + N) N_d)$ para todas as n_α componentes, para todo $k = 1, \dots, N_d$ e para todo $\eta = 1, \dots, N$, ou seja, $O(n_\alpha N^2 N_d)$. Com esses termos calculados, o cálculo do gradiente $\nabla_{\alpha} J$ em (4.84) demanda apenas um tempo adicional de ordem $O(n_\alpha N N_d)$. Em resumo, dadas todas as considerações acima, pode-se concluir, a partir da eliminação dos termos menos significativos, que a ordem de complexidade do tempo total para o cálculo das direções de busca no caso de Volterra é $O(NMN_d^2 + n_p N^2 M^N N_d + n_\alpha N^2 N_d)$.

Note, no entanto, que N é a ordem do polinômio de Volterra, que é, na prática, usualmente igual a 2 (raríssimamente maior que 3). Ou seja, N é uma constante, não uma variável crítica do problema. Além disso, independente da hipótese de simetria dos kernels, tem-se que a quantidade de coeficientes da expansão polinomial (dimensão do vetor α), n_α , é tal que $n_\alpha \propto M + M^2 + \dots + M^N$. A partir dessas duas observações, pode-se rescrever a ordem de magnitude do tempo total para o cálculo dos gradientes apresentada acima como $O(MN_d^2 + n_p M^N N_d)$. Para Kautz, tem-se $n_p = 2$, o que implica $O(MN_d^2 + M^N N_d)$. Já para GOBF tem-se $n_p \propto M$, o que implica $O(MN_d^2 + M^{(N+1)} N_d)$. Em qualquer dos casos, como na prática tem-se $N_d \gg M$, pode-se estimar a ordem de complexidade para o cálculo das direções

de busca no caso de modelos de Volterra como $O(N_d^2)$. Ou seja, em termos da variável crítica para o problema, N_d , a ordem de complexidade para o caso Volterra é a mesma daquela para o caso linear, não importa a base de funções (Kautz ou GOBF).

Apêndice C

Trabalhos Publicados

Os trabalhos publicados a partir desta tese estão listados a seguir.

Artigos em congressos

- da Rosa, A., Amaral, W. C. e Campello, R. J. G. B. (2006). Desenvolvimento de modelos de Volterra usando funções de Kautz e sua aplicação à modelagem de um sistema de levitação magnética. Em: *Proc. XVI Congresso Brasileiro de Automática*. Salvador-BA, Brasil.
- da Rosa, A., Campello, R. J. G. B. e Amaral, W. C. (2008). Cálculo de direções de busca exatas para otimização de filtros de Laguerre e de Kautz. Em: *Proc. XVII Congresso Brasileiro de Automática*. Juiz de Fora-MG, Brasil.

Artigos em periódicos

- 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. (2008). 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. (aceito). Exact search directions for optimization of linear and nonlinear models based on generalized orthonormal functions. *IEEE Transactions on Automatic Control*.

Referências Bibliográficas

- Aguirre, L. A. (2007). *Introdução à Identificação de Sistemas – Técnicas Lineares e Não-Lineares Aplicadas a Sistemas Reais*. 3a. edição, Editora UFMG, Belo Horizonte.
- Akçay, H. e Ninness, B. (1998). Rational basis functions for robust identification from frequency and time-domain measurements. *Automatica* **34**(9), 1101–1117.
- Bazaraa, M. S., Sherali, H. D. e Shetty, C. M. (1993). *Nonlinear Programming: Theory and Algorithms*. John Wiley & Sons.
- Billings, S. A. (1980). Identification of nonlinear systems – a survey. *IEE Proc. Pt D* **127**(6), 272–285.
- Bodin, P., Villemoes, L. F. e Wahlberg, B. (1997). An algorithm for selection of best orthonormal rational basis. Em: *Proc. 36th Conference on Decision and Control*. San Diego, California, USA.
- 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 approximation nonlinear operators with Volterra series. *IEEE Transactions 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.
- Campello, R. J. G. B., Amaral, W. C. e Favier, G. (2001). Optimal Laguerre series expansion of discrete Volterra models. Em: *Proc. European Control Conference*. Porto, Portugal.
- Campello, R. J. G. B. (2002). *Arquiteturas e Metodologias para Modelagem e Controle de Sistemas Complexos Utilizando Ferramentas Clássicas e Modernas*. Tese de Doutorado. Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas (UNICAMP). Campinas-SP, Brasil.

- Campello, R. J. G. B., Favier, G e Amaral, W. C. (2003). Optimal expansions of discrete-time Volterra models using Laguerre functions. Em: *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**(5), 815–822.
- 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., Oliveira, G. H. C. e Amaral, W. C. (2007a). Identificação e controle de processos via desenvolvimentos em séries ortonormais: Parte A: Identificação. *Controle & Automação* **18**(3), 301–321.
- Campello, R. J. G. B., Oliveira, G. H. C. e Amaral, W. C. (2007b). Identificação e controle de processos via desenvolvimentos em séries ortonormais. Parte B: Controle. *Controle & Automação* **18**(3), 321–336.
- Campello, R. J. G. B. e Oliveira, G. H. C. (2007c). Modelos não-lineares. Em: *Aguirre, L. A., Silva, A. P. A., Campos, M. F. M. e Amaral, W. C. (editores). Enciclopédia de Automática*, Vol. 3, Edgard Blücher.
- 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.
- Cormen, T. H., Leiserson, C. E., Rivest, R. L. e Stein, C. (2002). *Introduction to Algorithms*. MIT Press, 2a. edição.
- da Rosa, A. (2005). *Desenvolvimento de Modelos Discretos de Volterra Usando Funções de Kautz*. Dissertação de Mestrado. Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas (UNICAMP). Campinas-SP, Brasil.
- 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. Em: *Proc. 16th IFAC World Congress*. Praga, República Tcheca.
- da Rosa, A., Amaral, W. C. e Campello, R. J. G. B. (2006). Desenvolvimento de modelos de Volterra usando funções de Kautz e sua aplicação à modelagem de um sistema de levitação magnética. Em: *Proc. XVI Congresso Brasileiro de Automática*. Salvador-BA, Brasil.

- 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). 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. (2008b). Cálculo de direções de busca exatas para otimização de filtros de Laguerre e de Kautz. Em: *Proc. XVII Congresso Brasileiro de Automática*. Juiz de Fora-MG, Brasil.
- da Rosa, A., Campello, R. J. G. B. e Amaral, W. C. (aceito). Exact search directions for optimization of linear and nonlinear models based on generalized orthonormal functions. *IEEE Transactions on Automatic Control*.
- den Brinker, A. C., Benders, F. P. A. e Oliveira e Silva, T. A. M. (1996). Optimality conditions for truncated Kautz series. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing* **43**(2), 117–122.
- Doyle III, F. J., Pearson, R. K. e Ogunnaike, B. A. (2002). *Identification and Control Using Volterra Models*. Springer-Verlag, Berlin.
- Dumont, G. A. e Fu, Y. (1993). Non-linear adaptive control via Laguerre expansion of Volterra kernels. *Int. J. Adaptive Control and Signal Processing* **7**(5), 367–382.
- Educational Control Products (1999). *Manual for Model 730 – Magnetic Levitation System*. California, USA.
- Eykhoff, P. (1974). *System Identification: Parameter and State Estimation*. John Wiley & Sons, New York.
- Fu, Y. e Dumont, G. A. (1993). An optimum time scale for discrete Laguerre network. *IEEE Transactions on Automatic Control* **38**(6), 934–938.
- Gabel, R. A. e Roberts, R. A. (1987). *Signals and Linear Systems*. John Wiley & Sons.
- Hacıoğlu, R. e Williamson, G. A. (2001). Reduced complexity Volterra models for nonlinear system identification. *EURASIP Journal on Applied Signal Processing* **2001**(4), 257–265.
- Heuberger, P. S. C., Van den Hof, P. M. J. e Bosgra, 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, London.
- Kautz, W. H. (1954). Transient synthesis in the time domain. *IRE Transactions on Circuit Theory* **1**(3), 29–39.
- Kibangou, A. Y., Favier, G. e Hassani, M. M. (2003). Generalized orthonormal basis selection for expanding quadratic Volterra filters. Em: *Proc. 13th IFAC Symposium on System Identification*. Rotterdam, Holanda.
- Kibangou, A. Y., Favier, G. e Hassani, M. M. (2005a). Iterative optimization method of GOB-Volterra filters. Em: *Proc. 16th IFAC World Congress*. Praga, República Tcheca.
- Kibangou, A. Y., Favier, G. e Hassani, M. M. (2005b). Laguerre-Volterra filters optimization based on Laguerre spectra. *EURASIP Journal on Applied Signal Processing* **2005**(17), 2874–2887.
- Kibangou, A. Y., Favier, G. e Hassani, M. M. (2005c). Selection of generalized orthonormal bases for second-order Volterra filters. *Signal Processing* **85**(12), 2371–2385.
- Levenberg, K. (1944). A method for the solution of certain nonlinear problems in least squares. *Quarterly of Applied Mathematics* **2**(2), 164–168.
- Marquardt, D. W. (1963). An algorithm for the least-squares 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.
- Milanese, M. e Tempo, R. (1985). Optimal algorithms theory for robust estimation and prediction. *IEEE Transactions on Automatic Control* **30**(8), 730–738.
- Moreira, V. D. (2006). *Controle Preditivo Robusto de Sistemas Híbridos Incertos Integrando Restrições, Lógica e Dinâmica Baseada em Séries de Funções Ortonormais*. Tese de Doutorado. Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas (UNICAMP). Campinas-SP, Brasil.
- Moreira, V. D., Amaral, W. C. e Ferreira, P. A. V. (2004). A new approach for robust model predictive control with an application to an air-conditioning system. Em: *Proc. IEEE International Conference on Control Applications*. Taipei, Taiwan.

- Nalbantoğlu, V., Bokor, J., Balas, G. e Gaspar, P. (2003). System identification with generalized orthonormal basis functions: an application to flexible structures. *Control Engineering Practice* **11**(3), 245–259.
- Nelles, O. (2001). *Nonlinear System Identification*. Springer-Verlag, Berlin.
- Ngia, L. S. H. (2001). Separable nonlinear least-squares methods for efficient off-line and on-line modeling of systems using Kautz and Laguerre filters. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing* **48**(6), 562–579.
- Ninness, B. e Gustafsson, F. (1997). A unifying construction of orthonormal bases for system identification. *IEEE Transactions on Automatic Control* **42**(4), 515–521.
- Ninness, B., Hjalmarsson, H. e Gustafsson, F. (1999). The fundamental role of general orthonormal bases in system identification. *IEEE Transactions on Automatic Control* **44**(7), 1384–1406.
- Nocedal, J. e Wright, S. J. (1999). *Numerical Optimization*. Springer-Verlag, New York.
- Oliveira, G. H. C. (1997). *Controle Preditivo para Processos com Incertezas Estruturadas Baseado em Série de Funções Ortonormais*. Tese de Doutorado. Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas (UNICAMP). Campinas-SP, Brasil.
- Oliveira, G. H. C., Amaral, W. C., Favier, G. e Dumont, G. A. (2000). Constrained robust predictive controller for uncertain processes modeled by orthonormal series functions. *Automatica* **36**(4), 563–571.
- Oliveira e Silva, T. A. M. (1994). Optimality conditions for truncated Laguerre networks. *IEEE Transactions on Signal Processing* **42**(9), 2528–2530.
- Oliveira e Silva, T. A. M. (1995a). On the determination of the optimal pole position of Laguerre filters. *IEEE Transactions on Signal Processing* **43**(9), 2079–2087.
- Oliveira e Silva, T. A. M. (1995b). Optimality conditions for truncated Kautz networks with two periodically repeating complex conjugate poles. *IEEE Transactions on Automatic Control* **40**(2), 342–346.
- Oliveira e Silva, T. A. M. (1997). Stationary conditions for the L^2 error surface of the generalized orthonormal basis functions lattice filter. *Signal Processing* **56**(3), 233–253.
- Patwardhan, S. C. e Shah, S. L. (2005). From data to diagnosis and control using generalized orthonormal basis filters. Part I: Development of state observers. *Journal of Process Control* **15**(7), 819–835.

- Patwardhan, S. C., Manuja, S., Narasimhan, S. e Shah, S. L. (2006). From data to diagnosis and control using generalized orthonormal basis filters. Part II: Model predictive and fault tolerant control. *Journal of Process Control* **16**(2), 157–175.
- Rugh, W. J. (1981). *Nonlinear System Theory – The Volterra/Wiener Approach*. The Johns Hopkins University Press, Baltimore.
- Sarroukh, B. E., van Eijndhoven, S. J. L. e den Brinker, A. C. (2001). An iterative solution for the optimal poles in a Kautz series. Em: *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*. Salt Lake City, Utah, USA.
- Schetzen, M. (1980). *The Volterra and Wiener Theories of Nonlinear Systems*. Robert Krieger Publishing Company, New York.
- Takenaka, S. (1925). On the orthogonal functions and a new formula of interpolation. *Japanese Journal of Mathematics* (II), 129–145.
- Tanguy, N., Morvan, R., Vilbé, P. e Calvez L. C. (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 Calvez L. C. (2000). Pertinent parameters for Kautz approximation. *Electronic Letters* **36**(8), 769–771.
- Tanguy, N., Morvan, R., Vilbé, P. e Calvez L. C. (2002). Pertinent choice of parameters for discrete Kautz approximation. *IEEE Transactions on Automatic Control* **47**(5), 783–787.
- Van den Hof, P. M. J., Heuberger, P. S. C. e Bokor, J. (1995). System identification with generalized orthonormal basis functions. *Automatica* **31**(12), 1821–1834.
- Vicino, A. e Milanese, M. (1991). Optimal inner bounds of feasible parameter set in linear estimation with bounded noise. *IEEE Transactions on Automatic Control* **36**(6), 759–763.
- 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 Mäkilä, P. M. (1996). Approximation of stable linear dynamical systems using Laguerre e Kautz functions. *Automatica* **32**(5), 693–708.
- Wiener, N. (1958). *Nonlinear Problems in Random Theory*. The MIT Press, New York.

- Ziaei, K. e Wang, D. W. L. (2006). Application of orthonormal basis functions for identification of flexible-link manipulators. *Control Engineering Practice* **14**(2), 99–106.