

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

Modelos de Volterra: Identificação Não Paramétrica e Robusta Utilizando Funções Ortonormais de Kautz e Generalizadas

Autor: Márcio Feliciano Braga

Orientador: Prof. Dr. Wagner Caradori do Amaral

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

Dissertação de Mestrado apresentada à Faculdade de Engenharia Elétrica e de Computação como parte dos requisitos para obtenção do título de Mestre em Engenharia Elétrica. Área de concentração: **Automação**.

BANCA EXAMINADORA:

Prof. Dr. Antonio Augusto Rodrigues Coelho – UFSC
Prof. Dr. Paulo Augusto Valente Ferreira – DT/FEEC/UNICAMP

Campinas, SP
Junho/2011

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

B73m Braga, Márcio Feliciano
Modelos de Volterra: identificação não paramétrica e robusta utilizando funções ortonormais de Kautz e generalizadas / Márcio Feliciano Braga. – Campinas, SP: [s.n.], 2011.

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

Dissertação de Mestrado - 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 matemática. I. Amaral, Wagner Caradori do. II. Campello, Ricardo José Gabrielli Barreto. III. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. IV. Título

Título em Inglês:	Volterra models: nonparametric and robust identification using Kautz and generalized orthonormal functions
Palavras-chave em Inglês:	Systems identification, Nonlinear systems, Volterra series, Mathematical optimization
Área de concentração:	Automação
Titulação:	Mestre em Engenharia Elétrica
Banca Examinadora:	Antonio Augusto Rodrigues Coelho, Paulo Augusto Valente Ferreira
Data da defesa:	30-06-2011
Programa de Pós Graduação:	Engenharia Elétrica

COMISSÃO JULGADORA - TESE DE MESTRADO

Candidato: Márcio Feliciano Braga

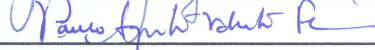
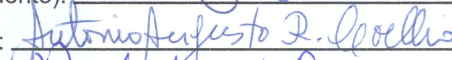
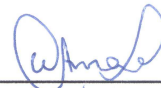
Data da Defesa: 30 de junho de 2011

Título da Tese: "Modelos de Volterra: Identificação Não-Paramétrica e Robusta Utilizando Funções Ortonormais de Kautz e Generalizadas"

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

Prof. Dr. Antônio Augusto Rodrigues Coelho:

Prof. Dr. Paulo Augusto Valente Ferreira:



*Dedico este trabalho ao meu pai,
Edson Braga, e à minha mãe,
Maria das Dores Braga.*

“What goes on in the modeller’s head is not purely formalizable, either in abstract terms or in taxonomic views. It has structure, it has technique that can be taught and learned, but involves also a personal touch, not only in trivialities but in deeper considerations of skill and suitability.”

- R. ARIS e M. PENN (1980)

ARIS, R., PENN, M. The Mere Notion of a Mathematical Model. Int. J. Math. Modelling, 1,1, 1980.

Agradecimentos

Gostaria de expressar os meus sinceros agradecimentos aos meus orientadores, Professor Doutor Wagner Caradori do Amaral e Professor Doutor Ricardo Campello, pela supervisão deste mestrado, especialmente pela dedicação na orientação, pela paciência dispensada na leitura das versões preliminares deste texto. Seus comentários perspicazes foram de suma importância para o aprimoramento da qualidade desta dissertação.

Agradeço imensamente os meus pais, Edson e Dora, e meus irmãos, Patrícia e Marcelo, pelo amor, incentivo e confiança demonstrados que sempre foram constantes em todas as etapas de minha vida.

Gostaria de agradecer aos amigos e professores da Unicamp, em especial ao Professor Doutor Paulo Valente, por sugestões e revisão de partes do texto desta dissertação. Estendo este agradecimento a todos os professores que fizeram parte de minha formação acadêmica, desde a minha graduação na Universidade Federal de Ouro Preto, e também aos colegas Danny e Adrielle que cursavam o mestrado em outras instituições com que tive a oportunidade de discutir as dificuldades e novidades desta nova etapa de minha vida.

Ao CNPq pelo apoio financeiro durante o desenvolvimento deste trabalho.

Resumo

Enfoca-se a modelagem de sistemas não-lineares usando modelos de Volterra com bases de funções ortonormais (*Orthonormal Basis Functions* – OBF) distintas para cada direção do kernel. Os modelos de Volterra constituem uma classe de modelos polinomiais não-recursivos, modelos sem realimentação da saída. Tais modelos são parametrizados por funções multidimensionais, chamadas *kernels* de Volterra, e representam uma generalização do bem conhecido modelo de resposta ao impulso (FIR) para a descrição de sistemas não-lineares. Como os modelos de Volterra não possuem realimentação do sinal de saída, um número elevado de parâmetros é necessário para representar os kernels de Volterra, especialmente quando o comportamento não-linear do sistema depende fortemente do sinal de saída. No entanto, é possível contornar esta desvantagem por descrever cada kernel por meio de uma expansão em bases de funções ortonormais (OBF). Resultando num modelo que, em geral, possui um número menor de termos para representar o sistema. 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 reside na questão de como selecionar os polos livres que completamente parametrizam estas funções de forma a reduzir o número de termos a serem utilizados em cada base. Uma abordagem já utilizada 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 polos da base. Estes gradientes fornecem direções de busca exatas para otimizar 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. Esta abordagem considerou apenas os modelos lineares e não-lineares cujas direções dos kernels foram todas parametrizadas por um mesmo conjunto de polos. Neste trabalho, estes resultados foram estendidos de forma a permitir o uso de uma base independente para cada direção dos kernels. Isto permite reduzir ainda mais o erro de truncamento quando as dinâmicas dominantes do kernel ao longo das múltiplas direções são diferentes entre si. As expressões dos gradientes relativas à base de Kautz e à base GOBF são obtidas recursivamente o que permite uma redução no tempo de processamento. Esta metodologia utiliza somente dados de entrada-saída medidos do sistema a ser modelado, isto é, não exige nenhuma informação prévia sobre os kernels de Volterra. Exemplos de simulação ilustram a aplicação dessas abordagens para a modelagem de sistemas não-lineares. Por último, apresentam-se resultados referentes à identificação robusta de modelos não-lineares sob a hipótese de erro desconhecido mas limitado, cujo objetivo é definir os limites superior e inferior dos parâmetros de modelos (intervalos de pertinência paramétrica). É analisado o caso em que se tem informação somente sobre a incerteza na saída do sistema, fornecendo-se o cálculo dos limitantes das incertezas para modelos OBF-Volterra. Estuda-se também os processos que possuem incerteza estruturada, i.e., os parâmetros do modelo, ou os *kernels* de Volterra, são definidos por meio de intervalos de pertinência e a ordem do modelo é conhecida. Apresenta-se uma solução exata para este problema, eliminando restrições impostas por metodologias anteriores.

Palavras-chave: Identificação de sistemas, Sistemas não-lineares, Séries de Volterra, Base de Funções ortonormais, Otimização, Identificação Robusta, Método UBBE, Incerteza estruturada.

Abstract

It focuses in the modeling of nonlinear systems using Volterra models with distinct orthonormal basis functions (OBF) to each kernel direction. The Volterra models are a class of nonrecursive polynomial models, models without output feedback. Such models are parameterized by multidimensional functions, called Volterra kernels, they represent a generalization of the well-known impulse response model and are used to describe nonlinear systems. As the Volterra models do not have output feedback, it is required a large number of parameters to represent the Volterra kernels, especially when the nonlinear behavior strongly depends of the output signal. However, such drawback can be overwhelmed by describing each kernel by an expansion in orthonormal basis functions (OBF). Resulting in a model that, in general, requires fewer parameters to represent the system. The resulting model, so-called OBF-Volterra, 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 in order to reduce the number of terms to be used in each bases. An approach, already used, involves the numeric optimization of orthonormal bases of function used for approximation of dynamic systems. This strategy is based on the computation of analytical expressions for the gradient of the orthonormal filters output with respect to the basis poles. Such gradient provides exact search directions for optimizing the poles of a given orthonormal basis. The search direction can, in turn, be used as part of an optimization procedure to locate the minimum of a cost-function that takes into consideration the estimation error of the system output. Although, that approach took in count only the linear models and nonlinear models which kernels directions were parameterized by a single set of poles. In this work, these results are extended in such a way to allow a use of an independent basis to each kernel direction. It can reduce even more the truncation error when dominant dynamics of the kernel are different along its directions. The gradient expressions to Kautz and GOBF bases are obtained in a recursive way which allows reducing the time processing. This 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 nonlinear systems. At last, it is presented some results about robust identification of nonlinear models under the hypothesis of unknown but bounded error, whose aim is to define the upper and lower bounds of the model parameters (parameter uncertainty interval). It is analyzed the case where the information available is about the uncertainty in the system output signal, providing the calculation for the uncertainty intervals to OBF-Volterra models. The process having structured uncertainty, i.e., the models parameters, or the Volterra kernels, are defined by intervals and the model order is known, is also studied. An exact solution to this problem is developed, eliminating restrictions imposed by previous approach.

Keywords: Systems Identification, Nonlinear Systems, Volterra Series, Orthonormal Basis Function, Optimization, Robust Identification, UBBE Method, Structured Uncertainty.

Sumário

Lista de Figuras	xvii
Lista de Tabelas	xix
Lista de Abreviaturas	xxi
1 Introdução	1
1.1 Organização do Trabalho	4
2 Modelos de Volterra Baseados em Funções Ortonormais (FBO)	7
2.1 Modelos de Volterra	8
2.1.1 Exemplo	11
2.1.2 Simetrização dos <i>kernels</i> de Volterra	12
2.2 Modelos de Volterra Baseados em Funções Ortonormais	13
2.2.1 Motivação Inicial	13
2.2.2 Representação matemática	14
2.3 Base de Funções Ortonormais	17
2.3.1 Base de Funções Ortonormais Generalizadas – GOBF	17
2.3.2 Funções de Laguerre	19
2.3.3 Funções de Kautz	19
2.3.4 Considerações finais sobre as Funções Ortonormais	23
2.4 Resumo	23
3 Direções de Busca Exatas Para Otimização dos Polos de Bases Ortonormais Distintas	25
3.1 Formulação do Problema	26
3.1.1 Otimização dos polos	26
3.2 Gradientes das Saídas dos Filtros	30
3.2.1 Funções de Kautz	31
3.2.2 Funções Ortonormais Generalizadas (GOBF)	35
3.2.3 Algoritmo	44
3.3 Exemplos Ilustrativos: Modelos não lineares	45
3.3.1 Exemplo ilustrativo 1	45
3.3.2 Exemplo ilustrativo 2	52
3.4 Comparação: Base Única <i>versus</i> Bases Múltiplas	58

3.4.1	Análise com funções de Kautz	59
3.4.2	Análise com GOBF	62
3.4.3	Considerações Finais sobre a Utilização de uma Base Distinta para Cada Di- reção do <i>Kernel</i>	66
3.5	Comparação: Tempo de Processamento	68
3.6	Resumo e Contribuições	70
4	Estimação Robusta de Regiões Paramétricas com Modelos OBF-Volterra	73
4.1	Estimação Robusta de Processos com Incertezas sobre a Saída	74
4.1.1	Formulação do Problema	75
4.1.2	Método UBBE para Estimação de Modelos Lineares	76
4.1.3	Solução por Hipercubo Externo	79
4.1.4	Descrição Exata da Região Politópica	80
4.1.5	Estimação Robusta via Modelos de Volterra	84
4.2	Exemplos Ilustrativos - Modelos com Incerteza na Saída	86
4.2.1	Modelo de Volterra simétrico de segunda ordem	86
4.2.2	Modelo de Volterra assimétrico de segunda ordem	88
4.3	Estimação Robusta de Processos com Incertezas sobre o <i>Kernel</i>	92
4.3.1	Cálculo dos limitantes dos coeficientes de modelos com incertezas	93
4.3.2	Relação com a Solução de Sobre-estimação do Conjunto de Respostas ao Impulso do Processo	98
4.4	Exemplo Ilustrativo - Modelos com Incertezas no <i>Kernel</i>	101
4.5	Resumo e Conclusão	105
5	Conclusões	109
5.1	Conclusões e Contribuições	109
5.2	Perspectivas e Trabalhos Futuros	110
	Referências Bibliográficas	111
A	Equações dos gradientes das saídas dos filtros ortonormais	117
A.1	Funções de Kautz	117
A.2	Funções Ortonormais Generalizadas (GOBF)	118
A.2.1	Caso GOBF com polos reais	118
A.2.2	Caso GOBF com um par de polos complexos	119
A.2.3	Caso GOBF com dois pares de polos complexos	120
B	Cálculo dos Limitantes de Incertezas em Modelos OBF-Volterra	123
B.1	Modelos Lineares	123
B.2	Modelos Não-Lineares	124
C	Trabalhos Publicados	127

Lista de Figuras

2.1	Sistema não linear com <i>kernels</i> de Volterra de primeira e segunda ordens.	11
2.2	Diagrama de blocos de um sistema linear que gera as funções ortonormais $\psi_{l,n}(k)$. .	18
2.3	Diagrama de blocos de um modelo OBF com Funções de Kautz	20
3.1	Kernel de segunda ordem com dinâmicas dominantes diferentes em cada direção – Exemplo 1.	47
3.2	Planos transversais do <i>kernel</i> de Volterra de segunda ordem mostrando as dinâmicas em cada direção.	47
3.3	Validação do modelo com Funções de Kautz. Acima: saída real do sistema. Abaixo: modelo correspondente com os polos ótimos obtidos: $[0, 6868 \quad 0, 2990 \pm 0, 7987i]$. .	49
3.4	Erro de estimação da saída do modelo com Funções de Kautz.	50
3.5	Validação do modelo com GOBF. Acima: Saída Real do sistema. Meio: Modelo correspondente utilizando os polos ótimos obtidos. Abaixo: Erro de estimação. . . .	52
3.6	Kernel de segunda ordem com dinâmicas dominantes diferentes em cada direção – Exemplo 2.	53
3.7	Planos transversais do <i>kernel</i> de Volterra de segunda ordem onde as dinâmicas em cada direção são realçadas.	54
3.8	Validação do modelo com Funções de Kautz. Acima: saída real do sistema. Abaixo: modelo correspondente com os polos ótimos: $[0, 2768 \pm 0, 4092i \quad 0, 4743 \pm 0, 7494i]$. .	56
3.9	Erro de estimação da saída do modelo com Funções de Kautz.	56
3.10	Validação do modelo com GOBF. Acima: Saída Real do sistema. Meio: Modelo correspondente utilizando os polos ótimos obtidos. Abaixo: Erro de estimação. . . .	57
3.11	Validação do modelo com funções de Kautz. Acima: saída real do sistema. Meio: modelo correspondente com o polo ótimo obtido: $[0, 3611 \pm 0, 5327i]$. Abaixo: Erro de estimação.	60
3.12	Validação do modelo com funções de Kautz. Acima: saída real do sistema. Meio: modelo correspondente com o polo ótimo obtido: $[0, 4529 \pm 0, 6570i]$. Abaixo: Erro de estimação.	62
3.13	Validação do modelo com GOBF. Acima: saída real do sistema. Abaixo: modelo correspondente com os polos ótimos obtidos: $[0, 7183 \quad 0, 3005 \pm 0, 7982i]$. Abaixo: Erro de estimação.	64
3.14	Estrutura do <i>kernel</i> simétrico em (3.86).	67
4.1	Região de pertinência paramétrica para modelos lineares	77

4.2	Região de pertinência paramétrica para modelos lineares com $ e(k) = e^M(k)$	80
4.3	Interpolação de um novo vértice com a introdução de uma nova observação.	83
4.4	Região de pertinência paramétrica para modelos não lineares.	85
4.5	Conjunto de <i>kernels</i> de primeira ordem.	102
4.6	Limites inferior (preto) e superior (azul) do conjunto de <i>kernels</i> de segunda ordem. .	103
4.7	Limites superior e inferior do modelo (linhas pontilhadas) e limitantes superior e inferior (linhas contínuas) do conjunto de repostas ao impulso dos <i>kernels</i> incertos de primeira ordem do processo.	104
4.8	Limites superior e inferior do modelo (linhas pontilhadas) e limitantes superior e inferior (linhas contínuas) do conjunto de repostas ao impulso dos <i>kernels</i> incertos de primeira ordem do processo utilizando a metodologia proposta em Oliveira et al. (2000) e da Rosa (2009).	105
4.9	Comparação entre os limites inferior e superior do <i>kernel</i> de primeira ordem obtidos por meio da abordagem proposta neste trabalho (linhas tracejadas), da abordagem em Oliveira et al. (2000) (linhas pontilhadas) e os limitantes do <i>kernel</i> incerto de primeira ordem (linhas contínuas).	106
4.10	Comparação entre os limites inferior e superior do <i>kernel</i> de segunda ordem obtidos por meio da abordagem proposta neste trabalho (linhas tracejadas), da abordagem em da Rosa (2009) (linhas pontilhadas) e os limitantes do <i>kernel</i> incerto de primeira ordem (linhas contínuas).	107

Lista de Tabelas

3.1	Valores ótimos dos polos de Kautz relativos ao número de funções utilizadas – Exemplo 1.	48
3.2	Erro de aproximação da saída de acordo com o número de funções de Kautz utilizadas – Exemplo 1.	49
3.3	Valores ótimos dos polos relativos ao número de funções generalizadas (GOBF) utilizadas – Exemplo 1.	51
3.4	Erro de aproximação da saída de acordo com o número de funções GOBF utilizadas – Exemplo 1.	51
3.5	Valores ótimos dos polos de Kautz relativos ao número de funções utilizadas – Exemplo 2.	54
3.6	Erro de aproximação da saída de acordo com o número de funções de Kautz utilizadas – Exemplo 2.	55
3.7	Valores ótimos dos polos relativos ao número de funções generalizadas (GOBF) utilizadas – Exemplo 2.	57
3.8	Erro de aproximação da saída de acordo com o número de funções (GOBF) utilizadas – Exemplo 2.	57
3.9	Valores ótimos dos polos de Kautz relativos a M funções utilizadas – Exemplo 1.	59
3.10	Erro de aproximação da saída de acordo com o número de funções de Kautz utilizadas – Exemplo 1.	59
3.11	Relação: número de parâmetros (NP) e erro de estimação (EQS) do modelo com funções de Kautz – Exemplo 1.	61
3.12	Valores ótimos dos polos de Kautz relativos a M funções utilizadas – Exemplo 2.	61
3.13	Erro de aproximação da saída de acordo com o número de funções de Kautz utilizadas – Exemplo 2.	61
3.14	Relação: número de parâmetros (NP) e erro de estimação (EQS) do modelo com funções de Kautz – Exemplo 2.	63
3.15	Valores ótimos dos polos relativos ao número de funções generalizadas (GOBF) utilizadas – Exemplo 1.	63
3.16	Erro de aproximação da saída de acordo com o número de funções (GOBF) utilizadas – Exemplo 1.	63
3.17	Relação: número de parâmetros (NP) e erro de estimação (EQS) do modelo com GOBF – Exemplo 1.	65
3.18	Valores ótimos dos polos relativos ao número de funções generalizadas (GOBF) utilizadas – Exemplo 2.	65

3.19	Erro de aproximação da saída de acordo com o número de funções (GOBF) utilizadas – Exemplo 2.	65
3.20	Relação: número de parâmetros (NP) e erro de estimação (EQS) do modelo com GOBF – Exemplo 2.	66
3.21	Comparativo: Modelagens Assimétrica e Simétrica para o modelo em (3.86) – Funções de Kautz.	67
3.22	Comparativo: Modelagens Assimétrica e Simétrica para o modelo em (3.86) – GOBF.	68
3.23	Tempo de processamento para os Exemplos Ilustrativos quando utilizadas as equações recursivas (Funções de Kautz).	69
3.24	Tempo de processamento para os Exemplos Ilustrativos utilizando as equações não recursivas (Funções de Kautz).	69
3.25	Valores ótimos dos polos de Kautz utilizando as equações não recursivas.	69
3.26	Tempo de processamento para os Exemplos Ilustrativos quando utilizadas as equações recursivas (GOBF).	70
3.27	Tempo de processamento para os Exemplos Ilustrativos utilizando as equações não recursivas (GOBF).	70
4.1	Valores ótimos dos polos encontrados na etapa de identificação - Exemplo 1.	87
4.2	Limites de Incerteza Paramétrica dos parâmetros α do modelo $y(k)$ com 100 medidas e 2 funções na base - Exemplo 1.	88
4.3	Limites de Incerteza Paramétrica dos parâmetros α do modelo $y(k)$ com 100 medidas e 4 funções na base - Exemplo 1.	89
4.4	Limites de Incerteza Paramétrica <i>PUI</i> para os parâmetros α do modelo simétrico $y(k)$ com 50 dados - Exemplo 1.	90
4.5	Valores ótimos dos polos encontrados na etapa de identificação – Base de Kautz - Exemplo 2.	90
4.6	Valores ótimos dos polos encontrados na etapa de identificação – GOBF - Exemplo 2.	91
4.7	Limites de Incertezas Paramétricas <i>PUI</i> para os parâmetros α do modelo $y(k)$ com 100 dados e $e^M = 0,57$ – Bases de Kautz - Exemplo 2.	91
4.8	Limites de Incertezas Paramétricas <i>PUI</i> para os parâmetros α do modelo $y(k)$ com 100 dados e $e^M = 0,57$ – GOBF - Exemplo 2.	92
4.9	Limites dos coeficientes α dos <i>kernels</i> incertos $\{h_1(k_1)\}$ com 4 funções (GOBF) na base.	102
4.10	Limites dos coeficientes α dos <i>kernels</i> incertos $\{h_2(k_1, k_2)\}$ com 4 funções (GOBF) na base.	104

Lista de Abreviaturas

DE	-	Método de descrição exata
EIB	-	Algoritmos de aproximação por elipsoide interno (<i>Algorithms Ellipsoidal-Inner-Bounding</i>)
EOB	-	Algoritmos de aproximação por elipsoide externo (<i>Algorithms Ellipsoidal-Outer-Bounding</i>)
FIR	-	Resposta ao Impulso Finita (<i>Finite Impulse Response</i>)
GOBF	-	Base de Funções Ortonormais Generalizadas (<i>Generalized Orthonormal Basis Functions</i>)
HE	-	Método do hipercubo externo
OBF	-	Base de Funções Ortonormais (<i>Orthonormal Basis Functions</i>)
<i>PUI</i>	-	Intervalo de Incerteza Paramétrica (<i>Parameter Uncertainty Interval</i>)
UBBE	-	Erro Desconhecido Mas Limitado (<i>Unknown But Bounded Error</i>)

Capítulo 1

Introdução

No cerne das ciências naturais e da engenharia, encontra-se a noção geral de observação e medida. Baseando-se em suas observações do mundo, o cientista constrói uma ideia inicial do problema e, a partir desta ideia primordial, formula, por tentativa e erro, a teoria que tem por objetivo descrever o fenômeno estudado. A seguir, esta teoria deve ser testada por meio de experimentos. A observação dos resultados, provenientes dos experimentos, pode confirmar a sua teoria, exigir mudanças ou mesmo rejeitá-la completamente. Este esforço persistente permite ao cientista encontrar algumas respostas a problemas que atormentam as mentes humanas e, durante este processo, deriva o prazer de pensar e, quem sabe, ver suas ideias difundidas na sociedade, chegando a ponto de se fundir ao tecido da vida cotidiana e tornar-se nele indistinguível.

De igual importância é a ideia da construção de modelos. Isto está diretamente relacionado com a formulação da teoria, uma vez que a teoria pode ser vista como um possível modelo matemático da realidade. Em Teoria de Sistemas, um modelo pode ser definido como uma representação dos aspectos essenciais de um sistema que o represente de forma útil para um determinado propósito (Eykhoff, 1974).

A obtenção de um modelo que descreva a realidade é de fundamental importância em engenharia, pois permite a análise e projeto de sistemas. O procedimento da modelagem de sistemas baseia-se, portanto, na noção geral, descrita anteriormente, de observação e medida. O problema central nesta etapa é: dado um histórico temporal da entrada e saída determinar as equações que descrevem o comportamento do sistema em estudo. Este processo é denominado de identificação. Identificação é definido por Zadeh (1962) como a determinação, com base nas observações das entradas e saídas, de um “modelo” dentro de uma classe específica de “modelos”, que descreva equivalentemente o comportamento do sistema em teste. Percebe-se assim que a qualidade do modelo está diretamente relacionada à qualidade da solução final do problema, ou seja, quanto mais precisa for a descrição do comportamento do sistema pelo modelo, melhores são os resultados no que tange, para citar algumas, simulação, controle e otimização do processo. Isto torna a modelagem parte essencial da Teoria de Sistemas, como sendo uma pré-etapa no desenvolvimento de projetos de controladores, etc.

Para a determinação do modelo no problema de identificação, duas linhas de estudo podem ser adotadas (Aguirre, 2007):

- (i) Análise física matemática (*model building*), também conhecida como modelagem conceitual, em que obtém-se o modelo por meio da descrição do comportamento físico do sistema, o que exige conhecer bem o sistema com o qual se está lidando.

- (ii) Identificação experimental, ou ainda modelagem caixa preta ou modelagem empírica, em que utilizam-se informações coletadas diretamente do processo em funcionamento para construir um modelo paramétrico do sistema, o que exige pouco ou nenhum conhecimento prévio do sistema a ser modelado, sendo esta última a utilizada neste trabalho.

Com o passar dos anos e com a melhoria nas áreas de computação e eletrônica, o estudo envolvendo identificação, modelagem e controle de sistemas dinâmicos tem crescido em importância e permitido que especificações mais rígidas de desempenho sejam impostas aos projetos. Tais especificações mais rígidas permitem atender as necessidades, cada vez maiores, dos meios de produção na busca de melhor qualidade e maior eficiência dos meios produtivos. E com isto, exigem-se modelos mais detalhados do processo, por exemplo modelos que descrevam de forma mais precisa o comportamento transitório do sistema, algo que há alguns anos era proibitivo, devido ao alto tempo de processamento e a erros provenientes das simulações.

Em especial, a identificação de sistemas não lineares era um campo de abordagens aplicáveis a classes restritas de sistemas. Com as modernas técnicas de otimização, uma maior classe de problemas podem agora ser investigadas, permitindo a utilização de uma mesma abordagem para uma variedade de sistemas não lineares com diferentes propriedades. Neste contexto, apresentam-se os modelos de Volterra que são utilizados para descrever ou representar processos não lineares.

Este trabalho lida com os problemas de identificação e modelagem de sistemas não lineares descritos por modelos de Volterra (Eykhoff, 1974; Schetzen, 1980; Rugh, 1981; Ljung, 1987; Nelles, 2001; Doyle III et al., 2002; Aguirre, 2007; Ogunfunmi, 2007). Os modelos de Volterra são modelos polinomiais sem realimentação do sinal de saída. Devido à sua estrutura geral, não linear, são de grande importância na representação de sistemas não lineares. Podem ser vistos como uma generalização do modelo de resposta ao impulso, que relaciona a entrada e saída dos sistemas lineares (Eykhoff, 1974; Schetzen, 1980; Aguirre, 2007; Campello and Oliveira, 2007).

Os modelos de Volterra são parametrizados por funções multidimensionais conhecidas como *kernels* de Volterra. Devido a não realimentação do sinal de saída, um grande número de termos pode ser necessário para representar os *kernels* de Volterra, em especial quando o comportamento não linear do sistema está fortemente relacionado ao sinal de saída do sistema (Nelles, 2001). Entretanto, esta desvantagem pode ser contornada por descrever cada *kernel* por meio de uma expansão em uma base de funções ortonormais (*Orthonormal Basis Functions* - OBF). Neste caso, o modelo é descrito por uma dinâmica linear (conjunto de filtros de funções ortonormais que relaciona a entrada do sistema aos estados ortonormais) seguida de um mapeamento estático não linear (relaciona os estados ortonormais à saída do sistema), ou seja, um modelo do tipo Wiener (Wiener, 1958; Campello and Oliveira, 2007). Gerando o modelo resultante conhecido como OBF-Volterra, o qual permite aproximar um dado sistema utilizando um número reduzido de termos se a base de funções ortonormais for adequadamente projetada. Como as bases de funções são caracterizadas por parâmetros de projeto livres, polos do modelo, a escolha adequada, i.e., quanto mais próximos estiverem os polos escolhidos dos polos reais do sistema em estudo, menor será o número de parâmetros a serem estimados, fixando-se uma mesma precisão (Campello and Oliveira, 2007). A estrutura OBF-Volterra possui, entre outros benefícios, a vantagem de ser linear nos parâmetros da expansão linear (Eykhoff, 1974; Schetzen, 1980; Rugh, 1981). Esta característica permite utilizar técnicas de regressão linear para identificar os parâmetros do modelo com base nas medidas realizadas.

Além da característica já citada, as bases de funções ortonormais possuem, entre outras, as se-

guintes características:

- permitem incorporar ao procedimento de identificação um conhecimento prévio (*a priori*) sobre a dinâmica do sistema (Heuberger et al., 1995; Ninness and Gustafsson, 1997);
- ausência de recursão da saída o que impede a realimentação de erros de predição levando a modelos mais precisos (Nelles, 2001);
- não necessidade de se conhecer previamente a estrutura do vetor de regressão, como a ordem do modelo, o que torna a sua identificação não-paramétrica, e atrasos de transportes, que podem ou não ser incorporados ao modelo;
- a possibilidade de aumentar a capacidade de representação do modelo por incrementar o número de funções utilizadas na base ortonormal;
- tolerância a dinâmicas não-modeladas e sensibilidade reduzida aos parâmetros incertos (Dumont and Fu, 1993).

As bases de funções ortonormais vem sendo aplicadas em diversas áreas como identificação de sistemas, processamento de sinais, controle robusto, entre outras (Oliveira e Silva, 1995a; Heuberger et al., 1995; Campello and Oliveira, 2007).

A utilização de modelos baseados em funções ortonormais em problemas de identificação e controle de processos dinâmicos lineares e não lineares tem sido objeto de amplo estudo na literatura (Dumont and Fu, 1993; Ninness and Gustafsson, 1997; Bokor and Schipp, 1998; Oliveira et al., 2000; Doyle III et al., 2002; Nalbantoglu et al., 2003; Campello et al., 2007a,b; da Rosa et al., 2007, 2008, 2009). Conforme comentado se for selecionado adequadamente os polos que parametrizam as bases, uma boa aproximação pode ser conseguida com um número reduzido de parâmetros. Muitos trabalhos recentes tem abordado a seleção ótima dos polos que caracterizam as bases de funções ortonormais discretas mais utilizadas na representação de sistemas lineares e não lineares, as quais são as bases de Laguerre, de Kautz e as bases de funções ortonormais generalizadas (GOBF - *Generalized Orthonormal Basis Functions*) (Broome, 1965; Wahlberg and Mäkilä, 1996; Bokor and Schipp, 1998; Heuberger et al., 1995).

Este trabalho tem por objetivo tratar da seleção ótima dos polos (parâmetros livres das bases de funções ortonormais) dos modelos OBF-Volterra. Dá-se uma ênfase especial aos modelos cujos *kernels* são assimétricos, ou seja, apresentam dinâmicas distintas ao longo de suas múltiplas direções, conforme será melhor detalhado no Capítulo 3. Além disso, aborda-se a questão da identificação robusta, quando existem incertezas ou perturbações externas associadas ao sistema. Tais perturbações impedem a obtenção de um conjunto único de parâmetros que garanta a um único modelo uma representação fiel do sistema em estudo. Portanto, torna-se imperativo a inclusão dessas incertezas aos parâmetros do modelo, em especial duas abordagens são estudadas: quando a incerteza está associada ao sinal de saída, conhecida na literatura como abordagem UBBE (Milanese and Belforte, 1982; Vicino and Milanese, 1989; Mo and Norton, 1990), e; ainda quando a incerteza está associada aos parâmetros do modelo, ou aos *kernels* de Volterra (Oliveira, 1997; Moreira, 2006).

1.1 Organização do Trabalho

Este trabalho, após esta introdução, encontra-se organizado na seguinte estrutura:

- **Capítulo 2 – Modelos de Volterra Baseados em Funções Ortonormais (FBO):** Estudam-se os modelos discretos de Volterra e como podem ser aplicados à modelagem de sistemas dinâmicos. Apresentam-se a terminologia, a notação e as manipulações matemáticas básicas relativas às representações de sistemas. Mostra-se que estes modelos podem ser vistos como uma generalização do conhecido modelo de resposta ao impulso dos sistemas lineares. Como tais modelos são parametrizados por funções multidimensionais, os chamados *kernels* de Volterra, também apresenta-se a ideia de simetrização dos *kernels* e faz-se um levantamento sobre as vantagens e desvantagens de proceder a simetrização de *kernels* assimétricos. Além disso, por não possuir realimentação de saída os modelos de Volterra geralmente necessitam de um número elevado de parâmetros, porém mostra-se que esta desvantagem pode ser contornada por descrever os *kernels* de Volterra por meio de uma base de funções ortonormais. Este procedimento gera o modelo conhecido como OBF-Volterra, o qual pode ser visto como um modelo tipo Wiener, uma dinâmica linear descrita pelo conjunto de filtros ortonormais, seguida pelo mapeamento estático, descrito neste trabalho pela série de Volterra. Descrevem-se também as principais bases de funções utilizadas na literatura: as bases de Laguerre e de Kautz, neste trabalho são utilizadas uma representação em espaço de estados para a base de Kautz, conforme desenvolvida em Machado (2011). Estas duas bases de funções podem ser vistas como um caso especial das funções ortonormais generalizadas (GOBF). Estas bases são caracterizadas por polos livres e, como já discutido, quanto mais próximos os polos do modelo estiverem dos polos reais do sistema, menor será o número de parâmetros a serem estimados, daí o interesse na seleção ótima dos polos, problema abordado no Capítulo 3.
- **Capítulo 3 – Direções de Busca Exatas Para Otimização dos Polos de Bases Ortonormais Distintas:** Estuda-se a seleção numérica dos polos livres que parametrizam as bases de funções ortonormais. Em da Rosa et al. (2009), apresentou-se uma abordagem para a seleção ótima dos polos onde minimizava-se uma função de custo quadrática que leva em conta o erro de estimação do sistema, ou seja, a diferença entre a saída do modelo e a saída real do sistema, no entanto, exigia-se que os *kernels* de Volterra fossem simétricos. Neste capítulo, estendem-se os resultados apresentados naquele trabalho, de forma a permitir que cada direção do *kernel* possa ser parametrizada por uma base independente de funções. Apresenta-se também um novo método, computacionalmente mais eficiente, para calcular, analítica e recursivamente, as direções de busca (gradientes) dos polos das bases. Esta nova abordagem permite reduzir o tempo de processamento envolvido nos cálculos dos polos ótimos da base. Utiliza-se o algoritmo de Levenberg-Marquardt (Levenberg, 1944; Marquardt, 1963; Nelles, 2001), o qual é uma escolha usual em problemas de otimização não linear (Nocedal and Wright, 1999; Nelles, 2001). Um conjunto de exemplos ilustrativos são apresentados para ilustrar o desempenho da metodologia desenvolvida. Apresenta-se, também, uma comparação com relação à redução do número de parâmetros, entre o método proposto, o qual utiliza *kernels* assimétricos com múltiplas bases ortonormais, e a metodologia original, onde exige-se que os *kernels* sejam simétricos parametrizados por uma única base de funções. Compara-se também o tempo de processamento

quando utilizam-se as equações desenvolvidas neste trabalho, equações recursivas, para calcular os gradientes das saídas dos filtros ortonormais e as equações apresentadas na metodologia anterior.

- **Capítulo 4 – Estimação Robusta de Regiões Paramétricas com Modelos OBF-Volterra:** O problema da estimação robusta sob a hipótese de erro desconhecido mas limitado (UBBE) é formulado nesta seção. São apresentados dois métodos de identificação robusta para a determinação dos intervalos de incerteza paramétricos: solução por Hipercubo Externo (Milanese and Belforte, 1982) e o algoritmo de descrição exata da região politópica (Mo and Norton, 1990). Os intervalos de pertinência, citados anteriormente, fornecem os limites superior e inferior de cada parâmetro do modelo. Entretanto, tanto a abordagem UBBE, como a descrita neste capítulo, quanto os métodos de identificação apresentados estão limitados à aplicação em sistemas lineares nos parâmetros, o que torna impeditiva sua aplicação em sistemas não lineares. Diante desta constatação, os modelos OBF-Volterra, modelos de Volterra baseados em funções ortonormais, apresentam-se como uma escolha viável para o tratamento de sistemas não lineares incertos, pois descrevem sistemas não lineares e são lineares nos parâmetros da expansão ortonormal, podendo a técnica UBBE ser diretamente aplicada. Aborda-se também a questão da estimação robusta de processos com incerteza estruturada, nestes processos as incertezas estão associadas aos parâmetros do modelo, ou aos *kernels* do modelo não linear, são definidas por intervalos de pertinência e não há dúvida relativas à ordem do modelo. Fornece-se o cálculo dos limitantes de incertezas para sistemas lineares e não lineares incertos usando modelos OBF-Volterra. Embora trabalhos anteriores tenham tratado da questão (Oliveira, 1997; Oliveira et al., 2000; da Rosa, 2009), a metodologia proposta não necessita supor a positividade das funções ortonormais. Um conjunto de exemplos ilustram as metodologias apresentadas, tanto para o caso UBBE quanto para o caso de incerteza estruturada.
- **Capítulo 5 – Conclusões:** As conclusões pertinentes a este trabalho são apresentadas neste capítulo, juntamente com as contribuições na área de modelagem e as perspectivas de trabalhos futuros.

Capítulo 2

Modelos de Volterra Baseados em Funções Ortonormais (FBO)

O modelo matemático de um sistema dinâmico é constituído por um conjunto de equações que representa as dinâmicas deste sistema de forma precisa ou o mais próximo disto. Tal modelo matemático não é único para um dado sistema, i.e., o mesmo sistema pode ser representado de diversas maneiras e, portanto, existem diferentes modelos matemáticos que atingem este objetivo. Dependendo do sistema e das circunstâncias relativas àquele sistema, um dado modelo matemático é mais adequado do que outro.

Particularmente, neste capítulo, faz-se uma análise da modelagem de sistemas dinâmicos baseada na representação em modelos de Volterra. Estes modelos constituem uma classe de modelos polinomiais não-recursivos, ou seja, modelos sem realimentação da saída. Serão apresentadas as principais vantagens e desvantagens dos modelos de Volterra na modelagem de sistemas não lineares. Apresenta-se também o conceito de simetrização dos *kernels* de Volterra, ferramenta matemática para se obter uma representação de sistemas via modelo de Volterra com menor complexidade.

Este capítulo apresenta a terminologia, a notação e as manipulações matemáticas básicas concernentes às representações de sistemas não lineares por modelos de Volterra. Estes modelos são compostos por funções multidimensionais, denominadas *kernels*, ou núcleos, de Volterra, e representam uma generalização do modelo de resposta ao impulso (FIR) para a descrição de sistemas lineares. Como os modelos de Volterra não possuem realimentação do sinal de saída, um número elevado de parâmetros é necessário para representar os *kernels* de Volterra (Schetzen, 1980; Doyle III et al., 2002; Aguirre, 2007; Campello and Oliveira, 2007), especialmente quando o comportamento não linear do sistema depende fortemente do sinal de saída (Nelles, 2001). No entanto, é possível contornar esta desvantagem descrevendo cada *kernel* por meio de uma expansão em bases de funções ortonormais (FBO), resultando num modelo que, em geral, possui um número menor de termos para descrever o sistema.

Em seguida, os conjuntos de funções ortonormais mais comumente utilizados na aproximação de sinais e sistemas são apresentados, sendo estes as bases de Laguerre, de Kautz e GOBF (Broome, 1965; Wahlberg and Mäkilä, 1996; Bokor and Schipp, 1998).

2.1 Modelos de Volterra

Para os sistemas lineares, o conhecimento de uma função temporal que caracterize o processo real, por exemplo a resposta ao impulso, é suficiente para determinar a saída do processo a um sinal de entrada arbitrário (Eykhoff, 1974; Schetzen, 1980). Para processos não lineares, também é desejável obter uma função temporal que os caracterizem e, por isso, o interesse na série de Volterra que cumpre este requisito.

Os modelos de Volterra são uma expansão funcional entrada-saída de um sistema não linear, sendo sua estrutura uma generalização direta do modelo de resposta ao impulso (Eykhoff, 1974; Nelles, 2001; Aguirre, 2007; Ogunfunmi, 2007). A conexão destes modelos com a teoria dos sistemas lineares, por tratá-los como uma subclasse do caso não linear (Zadeh, 1953), e a sua capacidade de representar uma grande classe de sistemas não lineares os tornam bastante atrativos.

Dentro dos modelos matemáticos, a série de Volterra encontra-se na categoria de modelos não-paramétricos. Para estes modelos não é necessário especificar explícita e previamente a ordem do processo. Isto é uma grande vantagem, pois as representações paramétricas podem resultar em grandes erros quando a ordem do modelo não concorda com a ordem do processo. Além disso, por ser uma representação intrinsecamente de dimensão infinita é possível obter um modelo cuja saída concorda exatamente com a saída do processo (Eykhoff, 1974).

A representação de sistemas por meio de modelos de Volterra foi introduzida pelo italiano Vito Volterra ¹, no seu livro “*Theory of Functional*” (Volterra, 1959) em 1887. Pode ser vista como uma generalização da série de Taylor (Anton, 2007; Eykhoff, 1974) para sistemas com memória. A primeira utilização da série de Volterra em sistemas não lineares é creditada a Norbert Wiener, já no século XX, que a utilizou num grande número de problemas (Wiener, 1958). Por exemplo em Wiener (1942), é analisada a resposta de um circuito RLC em série com um resistor não linear quando a entrada do sistema é um ruído branco gaussiano. Desde então, a série de Volterra tem sido aplicada em diversas áreas do conhecimento. A utilização destes modelos permite obter uma relação explícita entre os sinais de entrada e saída, analisar sistemas excitados por sinais aleatórios, descrever um sistema linear como uma subclasse dos sistemas não lineares, e estender aos sistemas não lineares conceitos definidos para sistemas lineares (Eykhoff, 1974).

A descrição matemática dos modelos de Volterra, no domínio do tempo discreto, que relaciona a saída $y(k)$ de um processo físico causal com sua entrada $u(k)$ é dada por (Schetzen, 1980; Rugb, 1981; Doyle III et al., 2002; Ogunfunmi, 2007):

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

¹Vito Volterra (1860-1940) – Nascido em 03 de maio de 1860 na cidade de Ancona, Itália. Foi um físico e matemático italiano, conhecido por suas contribuições à biologia matemática e equações integrais, sendo este último seu trabalho mais relevante. Publicou em 1896 artigos sobre o tema, conhecido como Equações Integrais de Volterra. E em 1887 publica o livro “*Theory of Functional*” onde apresenta a Série de Volterra e o Teorema de Volterra. Faleceu em 11 de outubro de 1940 em Roma, Itália.

onde as funções multivariáveis $h_\eta(\tau_1, \dots, \tau_\eta)$ são os *kernels* da série de Volterra, ou como definido em Eykhoff (1974) a resposta ao impulso de grau η , e h_0 , um termo adicional inserido para representar um eventual nível constante não-nulo (nível DC) na saída do sistema. Portanto, o modelo de Volterra descreve a saída de um sistema não linear como a soma das respostas de operadores de primeira ordem, de segunda ordem, até a ordem η desejada.

Como observado na equação (2.1), a série de Volterra é uma generalização do modelo de resposta ao impulso, tradicionalmente utilizado para a análise de sistemas lineares. Esse modelo descreve a saída de um sistema como a convolução de sua resposta ao impulso $h(k)$ com a entrada do sistema, ou seja:

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

A equação (2.2) descreve a saída dos sistemas dinâmicos lineares invariantes no tempo, cuja dinâmica do sistema fica completamente determinada por sua resposta ao impulso (Eykhoff, 1974; Doyle III et al., 2002; Ogunfunmi, 2007; Aguirre, 2007).

A equação (2.1) pode ser reescrita, desprezando-se o termo h_0 , na seguinte forma compacta:

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

Os modelos de Volterra possuem somas de infinitos termos sendo possível, em alguns casos, investigar a convergência da série de Volterra usando uma série de potências (Eykhoff, 1974). Todavia, raramente a série obtida possui uma forma fechada (expressão analítica). Na maioria dos casos somente uma série truncada pode ser usada, desde que esta forneça uma aproximação satisfatória do sistema quando são desprezados os termos de ordem superior (Boyd and Chua, 1985).

Os *kernels* da série de Volterra de sistemas dinâmicos estáveis possuem as seguintes propriedades (Eykhoff, 1974):

- $h_\eta(\tau_1, \tau_2, \dots, \tau_\eta) = 0 \quad \forall \tau_i < 0 \quad i = 1, 2, 3, \dots, \eta;$
- $\lim_{\tau_i \rightarrow \infty} h_\eta(\tau_1, \tau_2, \dots, \tau_\eta) = 0 \quad i = 1, 2, 3, \dots, \eta;$
- $h_\eta(\tau_1, \tau_2, \dots, \tau_\eta)$ são ou podem ser construídos de forma simétrica.

A primeira propriedade é devido ao fato que o sistema é considerado causal e com condições iniciais nulas, i.e., a resposta de um sistema físico realizável não depende de valores futuros do sinal de entrada, apenas de valores passados da entrada (Schetzen, 1980). A segunda propriedade é consequência da modelagem de sistemas estáveis (Eykhoff, 1974; Boyd and Chua, 1985), ou seja, a resposta a uma dada excitação na entrada desaparece com o tempo até não contribuir mais significativamente com a saída após o período finito de tempo τ_i . Sistemas que possuem esta característica são também conhecidos como sistemas com *fading memory* (Boyd and Chua, 1985; Campello et al., 2007a). A importância dessa propriedade, no contexto da série de Volterra, é permitir o truncamento

da série. Efetua-se o truncamento do termo de ordem η , por exemplo, da seguinte maneira, primeiro defini-se o escalar δ_η que representa o período de tempo a partir do qual a contribuição do respectivo *kernel* é desprezível, em seguida considera-se $h_\eta(\tau_1, \dots, \tau_\eta) = 0$ para $\tau_i > \delta_\eta$ ($\forall i \in \{1, \dots, \eta\}$). Quanto à terceira e última propriedade, tem-se que os termos $h_\eta(\tau_1, \tau_2, \dots, \tau_\eta)$ referentes a qualquer permutação dos índices $\tau_1, \tau_2, \dots, \tau_\eta$ sempre multiplicam o mesmo valor $u(k - \tau_1) \cdots u(k - \tau_\eta)$ na equação (2.3). Por conseguinte, pode-se proceder a simetrização do *kernel* sem nenhuma consequência para a capacidade de representação do modelo, o que permite uma redução no número de termos a ser estimado, sendo necessário estimar apenas metade do domínio (Eykhoff, 1974).

Embora a série de Volterra seja capaz de modelar sistemas dinâmicos não lineares, ela possui algumas desvantagens. A principal está associada à dificuldade de obter um número necessário de *kernels* para aproximar com precisão um sistema. Como já discutido, a ordem do *kernel* dá uma ideia do grau de não-linearidade do sistema e está associado ao número de *kernels* necessários para descrevê-lo. Devido a isto, um sistema que é levemente não linear pode requerer apenas um modelo de Volterra de segunda ordem, enquanto que um sistema altamente não linear pode necessitar de um número muito maior de *kernels* para fornecer resultados satisfatórios.

O problema de identificação de sistemas não lineares utilizando modelos de Volterra, como em (2.1), remete ao problema da estimação dos *kernels* $h_\eta(\tau_1, \dots, \tau_\eta)$ do modelo. A principal dificuldade neste caso é que os *kernels* são funções multidimensionais não-parametrizadas (Schetzen, 1980; Rugh, 1981) que só podem ser mensuradas se suas contribuições individuais puderem ser separadas da resposta total do sistema. Uma abordagem intuitiva e direta seria a de generalizar a ideia dos modelos FIR – *Finite Impulse Response* – (Eykhoff, 1974; Aguirre, 2007), tratando o valor de cada componente do *kernel* como um parâmetro individual a ser estimado (Doyle III et al., 2002). Nesse caso, o modelo de Volterra torna-se linear nos parâmetros e algoritmos de estimação clássicos podem ser utilizados. No entanto, esta abordagem seria possível apenas para a modelagem de sistemas com *fading memory* onde o conjunto de parâmetros pode ser feito finito por meio de um procedimento de truncamento dos termos da série, conforme definido na segunda propriedade dos *kernels* de Volterra. Entretanto, mesmo no caso dos modelos FIR que utilizam apenas o *kernel* de primeira ordem, a quantidade de parâmetros necessários para uma aproximação satisfatória do modelo pode ser elevada, em especial quando o sistema possui dinâmicas lentas, o que seria mais uma desvantagem da modelagem utilizando a série de Volterra. A inclusão de *kernels* de ordens superiores agrava ainda mais esta desvantagem pois provoca um aumento exponencial da quantidade de termos a serem estimados, tornando esta abordagem infactível para grande parte das aplicações práticas (Billings, 1980).

Assim, os modelos de Volterra são úteis na representação de sinais e sistemas devido à sua estrutura geral não linear. Entretanto, necessitam de um número considerável de termos para a representação dos *kernels* de Volterra, sendo esta considerada a sua principal desvantagem (Aguirre, 2007; Campello and Oliveira, 2007; Doyle III et al., 2002; Eykhoff, 1974; Nelles, 2001; Rugh, 1981). Nenhum método exato de isolar um operador de Volterra individual existe para modelos que não são de ordem finita.

Uma estratégia de particular interesse que visa contornar os problemas discutidos anteriormente, originalmente apresentada por Wiener (Wiener, 1958), consiste em expandir os *kernels* de Volterra usando conjuntos de funções ortonormais. Desta forma, o número de termos necessários para a representação dos modelos de Volterra pode ser drasticamente reduzido (Schetzen, 1980; Rugh, 1981). Estes modelos serão tratados posteriormente na Seção 2.2. A seguir é apresentado um exemplo que ilustra a descrição de um sistema não linear usando um modelo de Volterra.

2.1.1 Exemplo

Identificar um sistema utilizando a série de Volterra consiste em estimar os *kernels* $h_\eta(\tau_1, \dots, \tau_\eta)$ do modelo (2.1). Em princípio, como já discutido, trata-se de um problema complexo, pois os *kernels* são funções multidimensionais não parametrizadas. Este exemplo ilustra a descrição de um sistema não linear usando um modelo de Volterra com *kernels* de primeira e segunda ordens (adaptado de Rugh (1981)).

Considere a conexão não linear de três subsistemas, conforme apresentado na Figura 2.1.

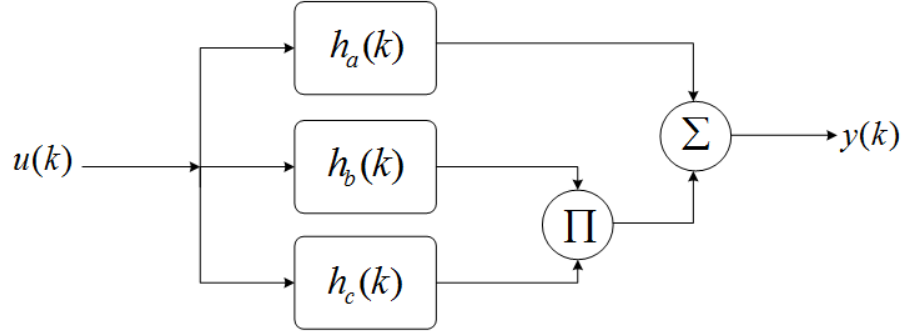


Fig. 2.1: Sistema não linear com *kernels* de Volterra de primeira e segunda ordens.

Os três subsistemas desta figura são lineares, caracterizados por suas respectivas respostas ao impulso, $h_a(k)$, $h_b(k)$ e $h_c(k)$. Esses subsistemas são descritos pela operação de convolução:

$$y_j(k) = \sum_{\tau=0}^k h_j(\tau)u(k-\tau) \quad j = a, b, c. \quad (2.4)$$

De acordo com o diagrama ilustrado na Figura 2.1, o sistema completo é descrito pelo produto das saídas dos dois subsistemas lineares $y_b(k)$ e $y_c(k)$ somado à saída do subsistema $y_a(k)$, i.e.:

$$\begin{aligned} y(k) &= y_a(k) + y_b(k)y_c(k) \\ &= \sum_{\tau_1=0}^k h_a(\tau_1)u(k-\tau_1) + \sum_{\tau_1=0}^k h_b(\tau_1)u(k-\tau_1) \sum_{\tau_2=0}^k h_c(\tau_2)u(k-\tau_2) \\ &= \sum_{\tau_1=0}^k h_a(\tau_1)u(k-\tau_1) + \sum_{\tau_1=0}^k \sum_{\tau_2=0}^k h_b(\tau_1)u(k-\tau_1)h_c(\tau_2)u(k-\tau_2). \end{aligned} \quad (2.5)$$

Observando a equação (2.5), nota-se que os *kernels* de primeira e segunda ordens do sistema são dados, respectivamente, por:

$$h_1(\tau_1) = h_a(\tau_1) \quad \text{e} \quad h_2(\tau_1, \tau_2) = h_b(\tau_1)h_c(\tau_2). \quad (2.6)$$

Portanto, o sistema da Figura 2.1 pode ser representado por um modelo de Volterra de segunda ordem, composto por um *kernel* de primeira ordem e um *kernel* de segunda ordem, como apresentado em (2.6). Nesse caso, o *kernel* de segunda ordem do modelo é composto pelo produto das respostas ao impulso individuais de cada um dos subsistemas, $h_b(k)$ e $h_c(k)$, que compõem o sistema.

2.1.2 Simetrização dos *kernels* de Volterra

Como discutido na terceira propriedade dos *kernels* da série de Volterra, todo *kernel* é, ou pode ser feito, simétrico. Isto significa que qualquer *kernel* assimétrico pode ser simetrizado, de maneira que a análise pode ser realizada considerando-se apenas sua versão simétrica sem qualquer perda de generalidade (Schetzen, 1980). O principal objetivo da simetrização é reduzir o número de parâmetros a serem obtidos e com isto simplificar os cálculos computacionais associados à etapa de estimação.

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

$$h_{\eta}^{sim}(\tau_1, \dots, \tau_{\eta}) = \frac{1}{\eta!} \sum_{\pi(\cdot)} h_{\eta}(\tau_{\pi(1)}, \dots, \tau_{\pi(\eta)}) \quad (2.7)$$

onde $\pi(\cdot)$ denota qualquer permutação dos inteiros $1, 2, \dots, \eta$, o que significa tomar todas as permutações possíveis dos elementos τ_1 a τ_{η} . A soma indicada na equação (2.7) é sobre todas as $\eta!$ permutações dos inteiros de 1 a η . Em Eykhoff (1974), essa definição é reescrita como:

$$h_{\eta}^{sim}(\tau_1, \dots, \tau_{\eta}) = \frac{1}{\eta!} \{h_{\eta}^{assim}(\tau_1, \dots, \tau_{\eta}) + \dots + h_{\eta}^{assim}(\tau_{\eta}, \dots, \tau_1)\} \quad (2.8)$$

assim, fica claro que um *kernel* de segunda ordem, por exemplo, é dito simétrico se $h_2(\tau_1, \tau_2) = h_2(\tau_2, \tau_1)$.

Para efeito de ilustração do processo de simetrização de um *kernel* de Volterra, considere que o *kernel* de terceira ordem de um determinado sistema seja dado por (Rugh, 1981):

$$h_3(\tau_1, \tau_2, \tau_3) = g(\tau_1)g(\tau_2)g(\tau_3)f(\tau_1 + \tau_2). \quad (2.9)$$

Aplicando a definição dada em (2.7) ou (2.8), tem-se que:

$$\begin{aligned} h_3^{sim}(\tau_1, \tau_2, \tau_3) &= \frac{1}{3!} \left[h_3(\tau_1, \tau_2, \tau_3) + h_3(\tau_1, \tau_3, \tau_2) + h_3(\tau_2, \tau_1, \tau_3) + h_3(\tau_2, \tau_3, \tau_1) + \right. \\ &\quad \left. h_3(\tau_3, \tau_1, \tau_2) + h_3(\tau_3, \tau_2, \tau_1) \right] \\ &= \frac{1}{6} \left[2g(\tau_1)g(\tau_2)g(\tau_3)f(\tau_1 + \tau_2) + 2g(\tau_1)g(\tau_2)g(\tau_3)f(\tau_1 + \tau_3) + \right. \\ &\quad \left. 2g(\tau_1)g(\tau_2)g(\tau_3)f(\tau_2 + \tau_3) \right] \\ &= \frac{1}{3} g(\tau_1)g(\tau_2)g(\tau_3) \left[f(\tau_1 + \tau_2) + f(\tau_1 + \tau_3) + f(\tau_2 + \tau_3) \right]. \end{aligned} \quad (2.10)$$

Ao comparar (2.10) e (2.9), nota-se que o *kernel* simétrico possui uma expressão matemática mais complexa do que sua versão assimétrica e, usualmente, é o que ocorre. No entanto, mesmo com o

aumento da complexidade, este pode ser um procedimento importante. Tal constatação é justificada porque, 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, produz-se uma significativa redução no número de parâmetros a serem estimados, pois neste caso necessita-se determinar apenas umas das partes simétricas para a caracterização completa do *kernel*. Outro fator importante é que a forma simétrica do *kernel* é única enquanto que a forma assimétrica não o é (Rugh, 1981; Schetzen, 1980).

Contudo, em algumas circunstâncias, é mais conveniente manter a assimetria do *kernel* a proceder a sua simetrização, mais especificamente quando uma base de funções ortonormais é utilizada para desenvolver os *kernels* da série de Volterra. Se o *kernel* apresentar dinâmicas dominantes diferenciadas ao longo de suas múltiplas direções, a manutenção da assimetria do *kernel* permite reduzir o erro de estimação associado ao modelo. No Capítulo 3, estas questões serão melhor detalhadas.

2.2 Modelos de Volterra Baseados em Funções Ortonormais

Devido às dificuldades apresentadas neste capítulo para a representação de sinais e sistemas por meio da série de Volterra, uma abordagem usando desenvolvimento ortonormal se apresenta como uma importante alternativa. Esta estratégia consiste em representar um determinado sistema usando uma base ortonormal para o espaço de interesse.

2.2.1 Motivação Inicial

Uma desvantagem associada aos modelos de Volterra na representação de sistemas dinâmicos é que, para descrever os *kernels* do modelo, é necessário uma grande quantidade de termos. No entanto, se uma base de funções ortonormais for utilizada para descrever os *kernels* do modelo, a quantidade de termos necessários pode ser reduzida. Isto ocorre porque as bases de funções ortonormais permitem incorporar conhecimento prévio da dinâmica do sistema no procedimento de identificação (Van den Hof et al., 1995; Oliveira, 1997; Ninness and Gustafsson, 1997; Nelles, 2001; Campello et al., 2007a; da Rosa, 2009). Algumas formas de obter esse conhecimento são intuições físicas sobre o sistema ou a partir de medições experimentais realizadas.

O uso de funções ortonormais na modelagem de sistemas dinâmicos para controle e processamento de sinais remonta a primeira utilização por Wiener em (Wiener, 1958). Como vantagem pode-se citar a simplificação da solução do problema de aproximação devido às propriedades de ortonormalidade das funções (Schetzen, 1980; Doyle III et al., 2002; Ogunfunmi, 2007). Se a dinâmica das funções do modelo aproxima-se da dinâmica dominante do processo, a representação do sistema estável torna-se simples. Outra vantagem do uso das funções ortonormais está relacionada à rápida taxa de convergência no problema de estimação dos coeficientes da expansão ortonormal, o que implica em um conjunto menor de coeficientes a serem determinados na modelagem (Van den Hof et al., 1995). Por serem geradas por filtros passa-todas, as funções ortonormais são muito robustas a implementações numéricas (Wahlberg, 1994).

A utilização de modelos baseados em funções ortonormais em problemas de identificação e controle de processos dinâmicos tem sido amplamente divulgada na literatura (Dumont and Fu, 1993; Ninness and Gustafsson, 1997; Bokor and Schipp, 1998; Oliveira et al., 2000; Doyle III et al., 2002;

Nalbantoglu et al., 2003; Campello et al., 2007a,b; da Rosa et al., 2007, 2008, 2009). Modelos que utilizam bases de funções ortonormais possuem propriedades atrativas também em termos da variância da sua estimação. Sabe-se que quanto maior o conjunto de parâmetros a serem estimados, maior é a variância da estimação dos parâmetros. Se houver uma seleção adequada dos parâmetros/polos da base ortonormal, um menor número de termos será utilizado e com isso obtém-se uma menor variância dos parâmetros.

Como os polos das funções ortonormais são parâmetros livres de projeto, um problema diretamente relacionado é como realizar uma boa seleção destes parâmetros (Tanguy et al., 2002; Oliveira e Silva, 1995a). Parece razoável escolher os polos das funções ortonormais o mais próximo possível dos polos reais do sistema estudado. Esse assunto será abordado no Capítulo 3.

2.2.2 Representação matemática

Apresenta-se agora a descrição matemática dos modelos de Volterra baseados em funções ortonormais (OBF), levando à representação conhecida como modelos OBF-Volterra. A ideia por trás desta estrutura é fundamentada na propriedade de completude das bases ortonormais. Esta representação é possível para sistemas dinâmicos em que os *kernels* do sistema são absolutamente somáveis no espaço $[0, \infty)$, i.e.:

$$\sum_{k_1=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} |h_\eta(k_1, \dots, k_\eta)| < \infty, \quad (2.11)$$

esta condição implica em estabilidade do sistema e permite que os *kernels* de Volterra sejam truncados ao serem representados em uma base ortonormal, conforme discutido nas propriedades dos *kernels* apresentados na Seção 2.1.

Nos modelos OBF-Volterra, os *kernels* h_η são descritos matematicamente como uma expansão numa base de funções ortonormais $\{\psi_{l,n}\}$, como (Schetzen, 1980; Doyle III et al., 2002; Ogunfunmi, 2007):

$$h_\eta(k_1, \dots, k_\eta) = \sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} \alpha_{i_1, \dots, i_\eta} \prod_{l=1}^{\eta} \psi_{l, i_l}(k_l). \quad (2.12)$$

A expansão dada em (2.12) também pode ser expressa no domínio da frequência complexa z , sendo representada em termos da transformada \mathcal{Z} das funções ortonormais como:

$$H_\eta(z_1, \dots, z_\eta) = \sum_{i_1=1}^{\infty} \cdots \sum_{i_\eta=1}^{\infty} \alpha_{i_1, \dots, i_\eta} \prod_{l=1}^{\eta} \Psi_{l, i_l}(z_l). \quad (2.13)$$

onde $\Psi_{l, i_l}(z) = \mathcal{Z}[\psi_{l, i_l}(k)]$.

O conjunto de funções $\{\psi_{l, i}(k)\}$ satisfaz a propriedade de ortonormalidade

$$\sum_{k=0}^{\infty} \psi_{l, i}(k) \psi_{l, j}(k) = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases} \quad (2.14)$$

descrita em (2.14) no domínio do tempo. Ou ainda, no domínio da frequência complexa z , como:

$$\frac{1}{2\pi i} \oint_{\mathbb{T}} \Psi_{l,i}(z) \bar{\Psi}_{l,i}(\bar{z}^{-1}) \frac{dz}{z} = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j, \end{cases} \quad (2.15)$$

onde a curva \mathbb{T} é o círculo unitário no plano complexo z : $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$.

Logo, os coeficientes $\alpha_{(\cdot)}$ podem ser calculados a partir de (2.12) 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 efeito de ilustração e comparação com os sistemas lineares, considere um *kernel* de primeira ordem. Isso significa fazer $\eta = 1$ na equação (2.12) e, consequentemente, escreve-se:

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

Perceba que (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 $\{\psi_{1,i_1}\}$.

Assim, para qualquer inteiro $j_1 > 0$ tem-se de (2.17) que

$$\sum_{k_1=0}^{\infty} h_1(k_1) \psi_{1,j_1}(k_1) = \sum_{k_1=0}^{\infty} \psi_{1,j_1}(k_1) \sum_{i_1=1}^{\infty} \alpha_{i_1} \psi_{1,i_1}(k_1) = \sum_{i_1=1}^{\infty} \alpha_{i_1} \sum_{k_1=0}^{\infty} \psi_{1,j_1}(k_1) \psi_{1,i_1}(k_1) \quad (2.18)$$

e utilizando a propriedade de ortonormalidade das funções, equação (2.14), os coeficientes da série são dados por:

$$\alpha_{j_1} = \sum_{k_1=0}^{\infty} h_1(k_1) \psi_{1,j_1}(k_1). \quad (2.19)$$

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

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

$$\alpha_{i_1, i_2} = \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} h_2(k_1, k_2) \psi_{1,i_1}(k_1) \psi_{2,i_2}(k_2). \quad (2.21)$$

Em sistemas com *fading memory*, na equação (2.12) limita-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. Isto significa dizer que, se a estabilidade do *kernel* for garantida, então as somas infinitas em (2.16) podem ser truncadas. Isto representa um truncamento prático para os *kernels* de Volterra, isto é, define-se uma constante $\delta < \infty$ tal que $h_\eta(\tau_1, \dots, \tau_\eta)$ seja assumido nulo para $k_l > \delta$, ($\forall l \in \{1, \dots, \eta\}$). Tal valor pode ser selecionado de acordo com a dinâmica do sistema a ser modelado, observando, por exemplo, o tempo de acomodação do sistema. Assim, a quantidade de funções utilizadas é limitada a um número finito M_{i_κ} , com $\kappa = 1, \dots, \eta$, de funções que seja o suficiente para garantir uma boa aproximação do modelo. Isto permite escrever (2.12) como:

$$\tilde{h}_\eta(k_1, \dots, k_\eta) = \sum_{i_1=1}^{M_{i_1}} \cdots \sum_{i_\eta=1}^{M_{i_\eta}} \alpha_{i_1, \dots, i_\eta} \prod_{l=1}^{\eta} \psi_{l, i_l}(k_l). \quad (2.22)$$

O truncamento apresentado anteriormente refere-se ao número de funções utilizadas para a expansão do *kernel*. No entanto, também é necessário um truncamento na ordem do *kernel*. Se certas condições forem impostas ao sinal de entrada $u(k)$ do modelo de Volterra (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, exige-se que a entrada seja limitada tal que $|u(k)| < 1$ (Dumont and Fu, 1993; Eykhoff, 1974), independentemente do tipo de não-linearidade. Se, além disso, for assumido que $u(k) = 0$ para $k < 0$, então a equação (2.3) pode ser reescrita como:

$$\tilde{y}(k) = \sum_{\eta=1}^N \left[\sum_{i_1=1}^{M_{i_1}} \cdots \sum_{i_\eta=1}^{M_{i_\eta}} \alpha_{i_1, \dots, i_\eta} \prod_{l=1}^{\eta} \left(\sum_{\tau_l=0}^k \psi_{l, i_l}(\tau_l) u(k - \tau_l) \right) \right]. \quad (2.23)$$

Devido à completude da bases ortonormais, o número $M_{(\cdot)}$ de funções a serem utilizadas no modelo (2.23) representa apenas um compromisso entre a capacidade de aproximação e a quantidade de parâmetros do modelo (Campello, 2002). O número de parâmetros, por sua vez, depende da quantidade de funções ortonormais utilizadas nos desenvolvimentos dos *kernels*. Modelos truncados de Volterra (2.23) podem aproximar qualquer sistema não linear invariante no tempo, com memória finita e entrada limitada (Boyd and Chua, 1985).

Deve-se observar que o modelo em (2.23) é linear nos parâmetros $\alpha_{i_1, \dots, i_\eta}$ e o problema de estimação dos *kernels* de Volterra resume-se à estimação linear desses parâmetros, o que pode ser feito por meio de algoritmos do tipo mínimos quadrados.

Outra importante característica dos filtros ortonormais é permitirem a incorporação de conhecimento *a priori* sobre a dinâmica do sistema no processo de identificação (Heuberger et al., 1995; Nelles, 2001; Ninness and Gustafsson, 1997; Van den Hof et al., 1995). Assim, é possível obter uma aproximação satisfatória utilizando um número reduzido de funções. No entanto, o número de funções necessárias para proporcionar uma dada representação depende da complexidade do problema. Por exemplo, sistemas com múltiplos modos dominantes requerem uma quantidade maior de funções. A quantidade de funções necessárias para atingir uma precisão pré-determinada também depende da base utilizada.

A principal limitação dos modelos OBF-Volterra decorre do truncamento na ordem do modelo, ou ainda, o número de *kernels* a ser utilizado. Por razões práticas, os modelos de segunda ordem ($N = 2$) são aqueles usualmente adotados tanto em problemas de natureza acadêmica quanto em problemas

reais (Billings, 1980; Dumont and Fu, 1993; Campello et al., 2007a). Embora o truncamento permita reduzir o número de parâmetros a ser estimado, ele também limita a capacidade de representação dos modelos. Novamente, verifica-se o compromisso (*trade-off*) entre a capacidade de representação do modelo e a redução do número de termos a serem estimados.

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

2.3 Base de Funções Ortonormais

A utilização de conjuntos de funções ortonormais na identificação de sistemas dinâmicos remonta há várias décadas (Takenaka, 1925; Wiener, 1942, 1958). Dentro deste conjunto de funções, apresentam-se as funções de Laguerre e Kautz (Broome, 1965; Wahlberg and Mäkilä, 1996; Bokor and Schipp, 1998) como as bases de funções ortonormais mais utilizadas em identificação e controle de sistemas. Essas funções são mais apropriadas para modelar sistemas com dinâmica dominante de primeira e segunda ordem, respectivamente.

Sistemas que apresentam dinâmicas complexas são melhor aproximados por uma base de funções que permita a incorporação de múltiplos modos, ou seja, um conjunto variado de polos (reais e complexos). Este contexto remete à 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 Kautz, é a capacidade de incorporar uma variedade de polos reais e complexos no modelo do sistema. No entanto, como desvantagem, as funções GOBF envolvem uma parametrização mais complexa da base ortonormal.

2.3.1 Base de Funções Ortonormais Generalizadas – GOBF

A base de funções ortonormais generalizadas (GOBF) pode ser gerada pela conexão, ou cascadeamento, de diferentes filtros passa-todas de ordem um ou dois, como a seguir (Ninness and Gustafsson, 1997; Heuberger et al., 1995):

$$\Psi_{l,n}(z) = \frac{z\sqrt{1-|\beta_{l,n}|^2}}{z-\beta_{l,n}} \prod_{j=1}^{n-1} \left(\frac{1-\bar{\beta}_{l,j}z}{z-\beta_{l,j}} \right) \quad n = 1, 2, \dots, m \quad (2.24)$$

onde $\bar{\beta}_{l,n}$ é o complexo conjugado de $\beta_{l,n}$ que são os polos estáveis da base ortonormal ($|\beta_{l,n}| < 1$). Esta base de funções, equação (2.24), é também conhecida como funções de Takenaka-Malmquist (Heuberger et al., 1995).

A partir da equação (2.24), pode-se obter uma estrutura em blocos equivalente, considerando cada filtro ortonormal como um sistema físico, conforme ilustrado na Figura 2.2.

Geralmente, as funções $\psi_{l,n}(k)$ serão complexas. Tal característica é fisicamente indesejável em problemas reais, com exceção de alguns problemas na área de telecomunicações. Em Ninness and

²Embora o termo Base de Funções Ortonormais Generalizadas é originalmente associado à formula em Heuberger et al. (1995) e Van den Hof et al. (1995), este termo será utilizado neste trabalho para se referir a bases de funções ortonormais com múltiplos modos.

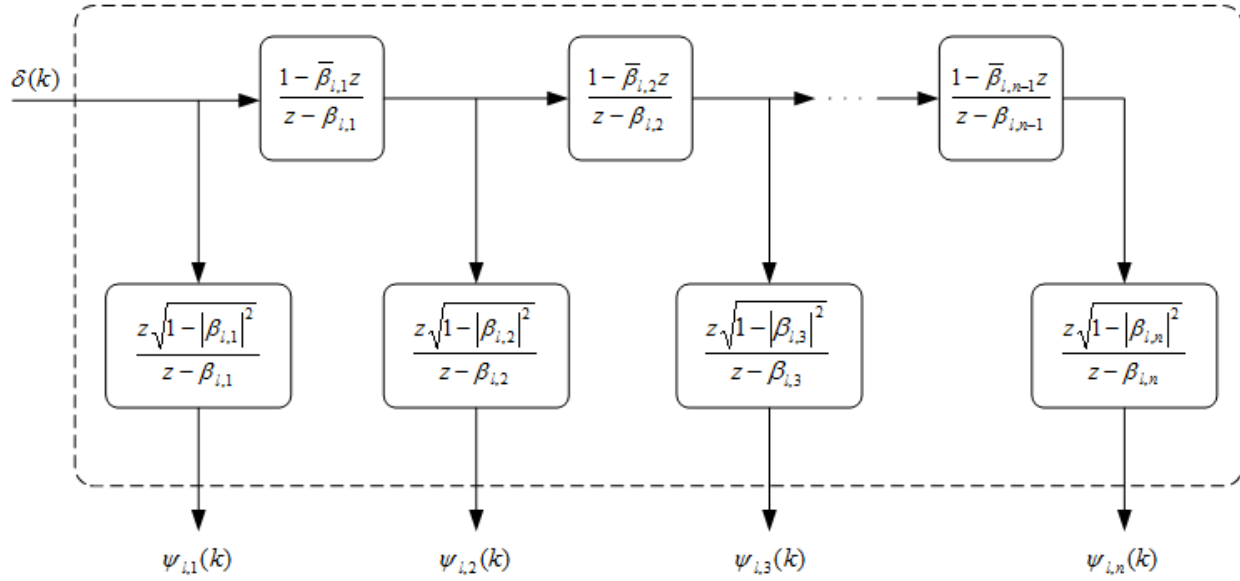


Fig. 2.2: Diagrama de blocos de um sistema linear que gera as funções ortonormais $\psi_{l,n}(k)$

Gustafsson (1997), apresenta-se um método de construção de GOBF que contorna esta desvantagem obtendo uma nova base de funções ortonormais com respostas ao impulso reais. Nesta 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 polos reais, o que significa fazer β_l em (2.24) igual a $p_l \in \mathbb{R}$. Assim, o conjunto de polos pode ser expresso como $\{p_1, p_2, \dots, p_{m-1}\}$. A introdução de um polo complexo $\beta_m \in \mathbb{C}$ exige que duas novas funções com respostas ao impulso reais sejam formadas como uma combinação linear de $\Psi_{l,m}(z)$ e $\Psi_{l,m+1}$ em (2.24). A nova base formada contém apenas funções com respostas ao impulso reais e preserva a propriedade de ortonormalidade. Esta nova representação das GOBF é apropriada para utilização em problemas de identificação e controle de sistemas, pois supera a desvantagem já comentada, e será melhor analisada na Seção 3.2.2 do Capítulo 3.

As funções dadas em (2.24) formam um conjunto completo no espaço $\ell^2 [0, \infty)$ se e somente se (Ninness and Gustafsson, 1997; Heuberger et al., 1995):

$$\sum_{n=1}^{\infty} (1 - |\beta_{l,n}|) = \infty. \quad (2.25)$$

Satisfazendo esta condição, qualquer sinal de energia finita pode ser aproximado com uma dada precisão procedendo-se o truncamento de uma expansão infinita em bases de funções ortonormais. Uma outra implicação da completude das funções ortonormais é que a parametrização do conjunto de funções não é crítica. Ou seja, caso a escolha dos parâmetros $\beta_{l,n}$, durante o processo de identificação, seja destoante dos valores reais do sistema, a capacidade de representação da série não é afetada. Exige-se apenas que ocorra o aumento do número de funções utilizadas para que a precisão estabelecida seja atingida.

Casos particulares dessas funções são as funções de Laguerre e Kautz, objeto de estudo a seguir.

2.3.2 Funções de Laguerre

As funções de Laguerre ³ podem ser vistas como um caso particular das funções ortonormais generalizadas, bastando fazer todos os polos da base, na equação (2.24), reais e iguais, isto é, $\beta_{l,n} = \bar{\beta}_{l,n} = p_l \quad \forall n$. As funções de Laguerre são expressas em termos de sua transformada \mathcal{Z} como (Wahlberg, 1991; Fu and Dumont, 1993; Oliveira e Silva, 1994; Heuberger et al., 1995):

$$\Psi_{l,n}(z) = \frac{z\sqrt{1-p_l^2}}{z-p_l} \left(\frac{1-p_l z}{z-p_l} \right)^{n-1} \quad (2.26)$$

onde o parâmetro p_l é conhecido como polo de Laguerre. Esse parâmetro é arbitrário, exceto pelo fato de que deve estar dentro do círculo unitário, ou seja, $p_l \in \{\mathbb{R} : |p_l| < 1\}$.

Quando $p_l = 0$, as funções de transferência em (2.26) são dadas por $\Psi_{l,n}(z) = z^{-(n-1)}$ e, portanto, representam o modelo FIR - *Finite Impulse Response* (Oliveira e Silva, 1995a).

Uma das vantagens das funções de Laguerre é que suas transformadas são funções racionais com uma forma repetitiva simples (Oliveira e Silva, 1994), por serem parametrizadas por um único polo real. É possível escrevê-las na forma de uma equação de estados, que não será apresentada aqui, mas pode ser encontrada em Campello et al. (2007a) ou Oliveira (1997). Estas funções são adequadas pra modelar sistemas com dinâmicas bem amortecidas com uma constante de tempo dominante, ou seja, sistemas com polos puramente reais ou com parte imaginária com valor reduzido (Nelles, 2001). Alocando os polos p_l próximos aos polos dominantes do sistema, o número de termos necessários para representá-lo adequadamente pode ser reduzido consideravelmente. Todavia, as funções de Laguerre possuem algumas desvantagens (Wahlberg, 1994):

- Sistemas com polos dominantes situados dispersamente no plano complexo não são bem descritos usando-se um número reduzido de funções;
- A presença de polos que produzam oscilação, pares complexos conjugados, pode causar problemas computacionais em termos da baixa velocidade de convergência da série.

Funções que possuem um ou mais pares de polos complexos dominantes se forem representadas por meio de uma série de Laguerre requerem um número elevado de termos (Oliveira e Silva, 1995b; Nelles, 2001). Isto ocorre porque os polos reais que parametrizam a base de Laguerre em (2.26) estarão longe dos polos complexos conjugados que descrevem o sistema. Uma alternativa para superar esta desvantagem é utilizar uma base de funções ortonormais que seja caracterizada por polos complexos conjugados. Esse é o caso das funções de Kautz, apresentadas na seção seguinte.

2.3.3 Funções de Kautz

A base de funções ortonormais de Kautz também pode ser vista como um caso particular das funções geradas pela equação (2.24) quando o conjunto de polos $\{\beta_{l,n}\}$ é $\{\beta_l, \bar{\beta}_l, \beta_l, \bar{\beta}_l, \dots\}$, com

³Edmond Nicolas Laguerre – Nascido em 9 de abril de 1834 em Bar-le-Duc, França. Foi um matemático francês que graduou-se em 1854 pela École Polytechnique em Paris. Publicou uma série de trabalhos a respeito das funções polinomiais de Laguerre e polinômios ortogonais, em 1879, e mostrou que uma função arbitrária pode ser expandida em série de funções polinomiais de Laguerre. Foi eleito membro da Academia Francesa em 1885. Suas principais áreas de investigação foram geometria e análise complexa. Faleceu em 14 de agosto de 1886 em sua cidade natal.

$\beta_l, \bar{\beta}_l \in \mathbb{C}$. Neste caso, as funções discretas de Kautz são definidas como (Wahlberg, 1994; Heuberger et al., 1995; Tanguy et al., 2002):

$$\Psi_{l,2n-1}(z) = \frac{z(z-b_l)\sqrt{1-c_l^2}}{z^2+b_l(c_l-1)z-c_l} \left[\frac{-c_l z^2 + b_l(c_l-1)z+1}{z^2+b_l(c_l-1)z-c_l} \right]^{n-1} \quad (2.27)$$

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

onde os parâmetros b_l e c_l são reais e estão relacionados com o par de polos de Kautz $\beta_l, \bar{\beta}_l$ da seguinte forma:

$$b_l = \frac{\beta_l + \bar{\beta}_l}{1 + \beta_l \bar{\beta}_l} \quad (2.28)$$

$$c_l = -\beta_l \bar{\beta}_l \quad (2.29)$$

sendo $|b_l| < 1$ e $|c_l| < 1$. As equações em (2.27) estão relacionadas por:

$$\Psi_{l,2n-1}(z) = \frac{z-b_l}{\sqrt{1-b_l^2}} \Psi_{l,2n}(z). \quad (2.30)$$

A partir da definição das funções de Kautz em (2.27) e (2.30) também é possível construir um diagrama de blocos equivalente, como aquele apresentado na Figura 2.2. Porém, o diagrama ilustrado na Figura 2.3 é mais completo por mostrar a relação entrada/saída de um modelo OBF.

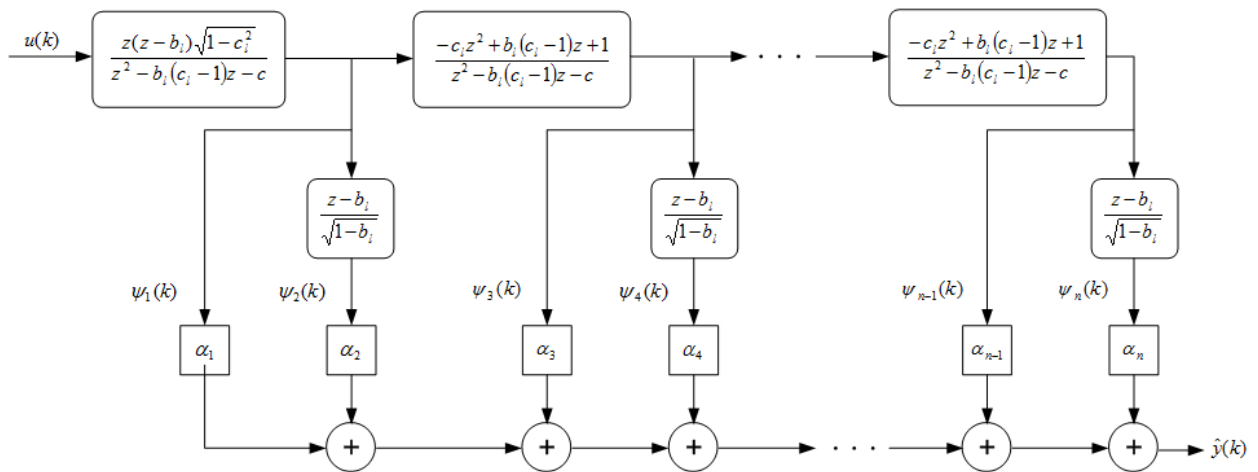


Fig. 2.3: Diagrama de blocos de um modelo OBF com Funções de Kautz

Em trabalhos recentes, as funções de Kautz tem recebido maior atenção porque permitem aproximar eficientemente sinais e sistemas com comportamento oscilatório (Wahlberg, 1994; Oliveira e

Silva, 1995b; Tanguy et al., 2002; da Rosa et al., 2005, 2007, 2008, 2009). A utilização destas funções na representação de sistemas com esta característica permite ajustar não somente o fator de amortecimento do modelo, mas também a sua frequência de oscilação.

Assim como as funções de Laguerre, as funções de Kautz também permitem uma descrição por meio de equações de estado que é apresentada a seguir.

Representação das Funções de Kautz em Espaço de Estados

A estrutura repetitiva das funções de Kautz permite uma representação em espaço de estados. Em Mbarek et al. (2003) apresentou-se uma representação que utiliza duas matrizes \mathcal{A}_1 e \mathcal{A}_2 , no entanto, em Machado (2011) foi desenvolvida uma representação mais concisa que permite a substituição daquelas duas matrizes por apenas uma matriz \mathcal{A}_K e será esta a representação utilizada neste trabalho. Portanto, a representação em espaço de estados das funções de Kautz é dada por:

$$\begin{aligned}\mathcal{X}(k+1) &= \mathcal{A}_K \mathcal{X}(k) + \mathcal{B}_K \mathcal{U}(k) \\ \hat{y}(k) &= \mathcal{C}_K \mathcal{X}(k)\end{aligned}\tag{2.31}$$

onde,

$$\mathcal{X}(k) = [\psi_{l,1}(k-1) \quad \psi_{l,1}(k) \quad \cdots \quad \psi_{l,n-1}(k-1) \quad \psi_{l,n-1}(k) \quad \psi_{l,n}(k-1) \quad \psi_{l,n}(k)]^T$$

$$\mathcal{U}(k) = [u(k-1) \quad u(k)]^T$$

$$\mathcal{A}_K = \begin{bmatrix} 0 & a_{1,2} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ a_{2,1} & a_{2,2} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a_{3,4} & \cdots & 0 & 0 & 0 & 0 \\ 0 & a_{4,2} & 0 & a_{4,4} & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & a_{2(n-1)-1,2(n-1)} & 0 & 0 \\ a_{2(n-1),1} & a_{2(n-1),1} & 0 & 0 & \cdots & a_{2(n-1)-1,2n-3} & a_{2(n-1)-1,2n-2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & a_{2n-1,2n} \\ 0 & 0 & 0 & 0 & \cdots & 0 & a_{2n,2n-2} & 0 & a_{2n,2n} \end{bmatrix}$$

$$\mathcal{B}_K = \sqrt{1 - c_l^2} \begin{bmatrix} 0 & 0 \\ b_{2,1} & b_{2,2} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ b_{6,1} & b_{6,2} \\ \vdots & \vdots \\ 0 & 0 \\ b_{2(n-1),1} & b_{2(n-1),2} \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathcal{C}_K = \begin{bmatrix} 0 & \alpha_1 & 0 & \alpha_2 & \cdots & 0 & \alpha_{n-1} & 0 & \alpha_n \end{bmatrix}$$

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

- Para i ímpar:
 - $a_{i,2i} = 1$;
 - demais termos iguais a 0.
- Para $i = 2$:
 - $a_{i,i} = -b_l(c_l - 1)$;
 - $a_{i,i-1} = c_l$;
 - demais termos iguais a 0.
- Para i múltiplo de 4:
 - $a_{i,i} = b_l$;
 - $a_{i,i-2} = \sqrt{1 - b_l^2}$;
 - demais termos iguais a 0.
- Para outras linhas i pares os termos são:
 - $a_{i,i} = -b_l(c_l - 1)$;
 - $a_{i,i-1} = c_l$;
 - $a_{i,i-4j} = (-c_l)^{j-1} [b_l(c_l - 1) + (-c_l)(-b_l(c_l - 1))]$ e
 $a_{i,i-(4j+1)} = (-c_l)^{j-1} ((-c_l)(c_l) + 1)$, para $j = 0, 1, 2, \dots$ tal que $4j < i$;
 - demais termos iguais a 0.

Já os elementos da matriz \mathcal{B}_K serão definidos por:

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

- $b_{i,1} = (-c_l)^j$;
- $b_{i,2} = (-c_l)^j (-b_l)$.

- Para os demais i 's:

- $b_{i,1} = b_{i,2} = 0$.

E os elementos da matriz \mathcal{C}_K são os coeficientes α da expansão ortonormal.

2.3.4 Considerações finais sobre as Funções Ortonormais

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 and Gustafsson, 1997; Nelles, 2001; Heuberger et al., 1995; Campello et al., 2007a), sendo cada uma (Laguerre, Kautz e GOBF) indicada para uma dinâmica específica. As GOBF permitem aproximar com melhor precisão sistemas que possuem dinâmica dominante superior a dois, pois permitem associar polos complexos e polos reais a uma mesma função. No processo de identificação, 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). Neste caso, o número de parâmetros a serem determinados para modelar o sistema é reduzido, e como consequência, tem-se o aumento da precisão.

É importante salientar ainda que, embora nenhum atraso de transporte esteja explicitamente representado nas descrições dos modelos OBF-Volterra, as funções ortonormais são capazes de representar dinâmicas com essa característica (Fu and Dumont, 1993). Também, pode-se incorporar explicitamente ao modelo qualquer informação aproximada disponível sobre o atraso real do sistema, bastando para isto proceder a substituição de $u(k)$ por $u(k - \tau_d)$ nas equações, onde τ_d é o atraso estimado. A incorporação deste tipo de informação permite reduzir ainda mais o número de funções necessárias para modelar o sistema com uma precisão.

O problema central do desenvolvimento dos modelos OBF-Volterra reside na seleção dos polos das funções ortonormais e do número de funções necessário para se obter uma boa aproximação da dinâmica do sistema. Como já comentado, uma base ortonormal pode ser construída selecionando-se os polos de maneira arbitrária, porém, ao utilizar este procedimento, exige-se um número elevado de termos na expansão ortonormal (equação 2.23) e, conseqüentemente, o modelo resultante pode conter muitos parâmetros. Assim, se a escolha da base de funções for adequada para o sistema em questão, uma boa aproximação pode ser gerada com um conjunto reduzido de coeficientes.

2.4 Resumo

Este capítulo apresentou conceitos matemáticos referentes à aproximação de sistemas dinâmicos não lineares por meio de modelos de Volterra, uma generalização do modelo linear de resposta ao

impulso. Trata-se de um modelo sem realimentação da saída que, embora evite a realimentação de erros de predição, exige que um grande número de parâmetros seja necessário para representar os seus *kernels*.

Esse número elevado de parâmetros pode ser reduzido com a utilização de uma base de funções ortonormais para descrever os *kernels* de Volterra, representação conhecida como OBF-Volterra. Esta modelagem permite incorporar conhecimento prévio, na base de funções, sobre a dinâmica do processo. Outro fator marcante deste modelo é que ele é linear nos parâmetros, o que permite a utilização de algoritmos de estimação clássicos como mínimos quadrados.

Apresentou-se, também, as principais bases de funções ortonormais utilizadas na modelagem e controle de sistemas. Deu-se ênfase às Bases de Funções Ortonormais Generalizadas (GOBF), de Laguerre e Kautz, sendo que essas últimas podem ser vistas como um caso particular da primeira. Apresentaram-se ainda as principais vantagens de utilização de cada uma destas bases.

Capítulo 3

Direções de Busca Exatas Para Otimização dos Polos de Bases Ortonormais Distintas

A principal limitação da utilização de funções ortonormais para a identificação de sistemas não-lineares utilizando modelos de Volterra é a necessidade do conhecimento prévio sobre os *kernels* do sistema a ser modelado. Trata-se de uma tarefa de difícil imposição prática, uma vez que os kernels são, em princípio, funções não-parametrizadas. Logo sua medição só é possível se, da resposta total do sistema, cada contribuição individual puder ser separada.

Para eliminar esta desvantagem, uma estratégia de otimização global foi apresentada em da Rosa (2009). Trata-se da aproximação de modelos OBF-Volterra por meio de seleção numérica dos polos das bases ortonormais. Os gradientes das saídas dos filtros ortonormais em relação aos polos da base foram obtidos e utilizados juntamente com um algoritmo de otimização para obter uma direção a ser seguida no espaço de busca dos polos ótimos da base ortonormal. Desta maneira, deseja-se que a partir de um conjunto de polos iniciais qualquer, o algoritmo convirja para os polos ótimos do modelo.

Naquele trabalho, foram apresentados resultados para modelos lineares e, posteriormente, estendidos para modelos de Volterra de qualquer ordem. As expressões analíticas dos gradientes para os modelos baseados em funções de Kautz e funções ortonormais generalizadas (GOBF) também foram apresentadas. Contudo, considerou-se somente o caso onde os *kernels* eram simétricos, ou seja utiliza-se apenas uma base de funções ortonormais para a expansão dos *kernels*.

Trabalhos anteriores abordaram a questão da utilização de uma base de funções ortonormais independente para cada direção de cada *kernel* do modelo OBF-Volterra utilizando uma solução analítica para a obtenção dos polos ótimos que parametrizam as bases ortonormais. Em Campello et al. (2006), é derivada uma solução analítica global onde as funções de Laguerre são utilizadas para a expansão dos *kernels* de Volterra, enquanto que em da Rosa et al. (2007), obtém-se uma solução analítica global para a expansão ótima dos *kernels* usando funções de Kautz.

Este capítulo utiliza uma conexão das duas ideias descritas nos parágrafos anteriores e apresenta resultados referentes a modelos de Volterra com o uso de uma base de funções ortonormais independente para cada direção de cada *kernel*, ou seja, cada *kernel* $h_\eta(k_1, \dots, k_\eta)$ é decomposto em um conjunto finito de funções ortonormais independentes para cada direção $l = 1, \dots, \eta$, conjuntamente com a estratégia de seleção numérica dos polos (gradientes analíticos). Espera-se, assim, reduzir o erro de truncamento quando as dinâmicas dominantes do *kernel* ao longo das múltiplas direções são distintas. Além disso, apresenta-se um novo método, computacionalmente mais eficiente, para calcu-

lar, analítica e recursivamente, as direções de busca (gradientes) dos polos das bases. Tal abordagem permite reduzir o tempo de processamento envolvido nos cálculos dos polos ótimos.

O problema consiste na minimização de uma função de custo quadrática que considere o erro de estimação da saída do sistema, definido como a diferença entre a saída real do processo e a saída do modelo. As derivadas obtidas são utilizadas no método de otimização não linear, para a obtenção da direção de busca para os polos da base. O algoritmo adotado é o de Levenberg-Marquardt (Levenberg, 1944; Marquardt, 1963; Nocedal and Wright, 1999; Nelles, 2001), que é uma escolha usual em problemas de otimização não linear, devido a sua característica dual de representar o método de descida do gradiente e o método do gradiente conjugado em problemas de médio porte. O método de Levenberg-Marquardt é um algoritmo de pseudo-segunda ordem, já que trabalha apenas com avaliações da função de custo e informações do gradiente. Sendo a matriz Hessiana da função de custo estimada por meio destas informações.

3.1 Formulação do Problema

O problema de se ajustar a relação dinâmica entre os parâmetros dos modelos OBF e a saída do modelo é formulado nesta seção. Primeiro, as equações do gradiente do critério quadrático são desenvolvidas. A seguir, as equações do gradiente das saídas dos filtros ortonormais de forma recursiva são desenvolvidas. Por fim, o desenvolvimento da técnica para modelos de Volterra utilizando bases independentes para cada direção.

3.1.1 Otimização dos polos

Para facilitar o entendimento, primeiro será apresentada a metodologia de otimização dos polos para o caso linear e, em seguida, sua extensão para o caso geral destes modelos.

Modelo linear

Em um sistema linear, a resposta do filtro $\psi_m(k)$ à entrada $u(k)$ é dada por:

$$\check{\psi}_m(k) = \sum_{\tau=0}^k \psi_m(\tau) u(k - \tau) \quad (3.1)$$

$$\check{\Psi}_m(z) = \Psi_m(z) U(z). \quad (3.2)$$

Portanto, um modelo linear (modelo OBF-Volterra de primeira ordem) composto por M funções de uma determinada base $\{\psi_m\}$ pode ser expresso como:

$$\hat{y}(k) = \sum_{m=0}^M \alpha_m \check{\psi}_m(k) \quad (3.3)$$

onde $\check{\psi}_m(k)$ foi definido na equação (3.1).

Como apresentado em da Rosa (2009), a estratégia adotada foi a otimização simultânea do vetor θ , composto pelo vetor de polos \mathbf{p} que parametriza a base ortonormal $\{\psi_m\}$ e do vetor de coeficientes correspondentes $\alpha := [\alpha_1 \ \alpha_2 \ \cdots \ \alpha_M]^T$. Para isto, defini-se o problema de otimização (3.4) que leva em conta o erro de estimação da saída do sistema:

$$\min_{\theta} J \triangleq \frac{1}{2} \sum_{k=1}^{N_d} [y(k) - \hat{y}(k)]^2 \quad (3.4)$$

onde $y(k)$ é a saída medida do sistema, $\hat{y}(k)$ é a saída do modelo, ambas no instante k , e N_d é o número de amostras entrada-saída disponíveis. O gradiente do critério quadrático J com relação ao vetor de polos \mathbf{p} e α são calculados como:

$$\begin{aligned} \nabla_{\mathbf{p}} J &= \sum_{k=1}^{N_d} [\hat{y}(k) - y(k)] \nabla_{\mathbf{p}} \hat{y}(k) \\ &= \sum_{k=1}^{N_d} [\hat{y}(k) - y(k)] \sum_{m=1}^M \alpha_m \nabla_{\mathbf{p}} \check{\psi}_m(k) \\ &= \sum_{k=1}^{N_d} \sum_{m=1}^M [\hat{y}(k) - y(k)] \alpha_m \nabla_{\mathbf{p}} \check{\psi}_m(k) \end{aligned} \quad (3.5)$$

$$\begin{aligned} \nabla_{\alpha} J &= \sum_{k=1}^{N_d} [\hat{y}(k) - y(k)] \nabla_{\alpha} \hat{y}(k) \\ &= \sum_{k=1}^{N_d} [\hat{y}(k) - y(k)] \check{\psi}(k) \end{aligned} \quad (3.6)$$

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

Modelos de Volterra de ordem superior

A seguir são desenvolvidos os gradientes do critério quadrático J para os modelos de Volterra de ordem η , onde se utiliza uma base de funções ortonormais distinta para cada direção do *kernel*. Por questão de simplicidade, considere em (2.23) apenas o η -ésimo termo do modelo OBF-Volterra:

$$\begin{aligned} \hat{y}_{\eta}(k) &= \sum_{i_1=1}^{M_{i_1}} \cdots \sum_{i_{\eta}=1}^{M_{i_{\eta}}} \alpha_{i_1, \dots, i_{\eta}} \prod_{l=1}^{\eta} \left(\sum_{\tau_l=0}^k \psi_{l, i_l}(\tau_l) u(k - \tau_l) \right) \\ &= \sum_{i_1=1}^{M_{i_1}} \cdots \sum_{i_{\eta}=1}^{M_{i_{\eta}}} \alpha_{i_1, \dots, i_{\eta}} \prod_{l=1}^{\eta} \check{\psi}_{l, i_l}(k) \end{aligned} \quad (3.7)$$

tal que a saída total do modelo (2.23) pode ser reescrita como a soma de N componentes, ou seja, $\hat{y}(k) = \sum_{\eta=1}^N \hat{y}_{\eta}(k)$.

A estratégia permanece igual àquela adotada no caso linear, a otimização simultânea do vetor de polos \mathbf{p} que parametriza as bases ortonormais $\{\psi_{1,i_l}\}, \dots, \{\psi_{\eta,i_l}\}$ e o vetor de coeficientes correspondentes $\boldsymbol{\alpha} := [\alpha_{1,\dots,1} \ \cdots \ \alpha_{1,\dots,M} \ \cdots \ \alpha_{M,\dots,M}]^T$ para cada componente do modelo dado em (3.7). Com base nestas observações define-se o seguinte problema de otimização:

$$\min_{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_\eta} J \triangleq \frac{1}{2} \sum_{k=1}^{N_d} [y(k) - \hat{y}(k)]^2 \quad (3.8)$$

onde $\boldsymbol{\theta}_\eta (\eta \in \{1, \dots, N\})$ é o vetor de parâmetros associado com a η -ésima componente do modelo e $y(k)$, $\hat{y}(k)$ e N_d são os mesmos definidos anteriormente.

Como o modelo apresentado em (3.7) utiliza uma base ortonormal distinta para cada direção l do correspondente *kernel* de ordem η , cada base ortonormal possui um conjunto específico e distinto de polos $\mathbf{p}_l = [p_{l,1}^T \ p_{l,2}^T \ \cdots \ p_{l,M_{i_l}}^T]$ para sua parametrização. Todos estes podem ser dispostos em um único vetor de polos \mathbf{p} , i.e., $\mathbf{p} = [\mathbf{p}_1^T \ \mathbf{p}_2^T \ \cdots \ \mathbf{p}_\eta^T]$. Este vetor pode, então, ser concatenado com o correspondente vetor de coeficientes $\boldsymbol{\alpha}$ para formar o vetor de parâmetros $\boldsymbol{\theta}_\eta$ para a expansão do η -ésimo *kernel*, i.e., $\boldsymbol{\theta}_\eta = [\mathbf{p}^T \ \boldsymbol{\alpha}^T]^T = [\mathbf{p}_1^T \ \mathbf{p}_2^T \ \cdots \ \mathbf{p}_\eta^T \ \boldsymbol{\alpha}^T]$.

Dessa forma, o gradiente do funcional J em (3.8) com relação aos vetores \mathbf{p} e $\boldsymbol{\alpha}$ de $\boldsymbol{\theta}_\eta$ são obtidos como:

$$\begin{aligned} \nabla_{\mathbf{p}} J &= \sum_{k=1}^{N_d} [\hat{y}(k) - y(k)] \nabla_{\mathbf{p}} \hat{y}(k) \\ &= \sum_{k=1}^{N_d} [\hat{y}(k) - y(k)] \sum_{\eta=1}^N \nabla_{\mathbf{p}} \hat{y}_\eta(k) \\ &= \sum_{k=1}^{N_d} \sum_{\eta=1}^N [\hat{y}(k) - y(k)] \nabla_{\mathbf{p}} \hat{y}_\eta(k) \end{aligned} \quad (3.9)$$

e

$$\begin{aligned} \nabla_{\boldsymbol{\alpha}} J &= \sum_{k=1}^{N_d} [\hat{y}(k) - y(k)] \nabla_{\boldsymbol{\alpha}} \hat{y}(k) \\ &= \sum_{k=1}^{N_d} [\hat{y}(k) - y(k)] \sum_{\eta=1}^N \nabla_{\boldsymbol{\alpha}} \hat{y}_\eta(k) \\ &= \sum_{k=1}^{N_d} \sum_{\eta=1}^N [\hat{y}(k) - y(k)] \nabla_{\boldsymbol{\alpha}} \hat{y}_\eta(k). \end{aligned} \quad (3.10)$$

onde $\nabla_{\mathbf{p}} \hat{y}_\eta(k)$ e $\nabla_{\boldsymbol{\alpha}} \hat{y}_\eta(k)$ são os gradientes de (3.7) com relação ao vetor de polos e coeficientes do respectivo *kernel*. Quando \mathbf{p} e $\boldsymbol{\alpha}$ referem-se a um outro *kernel*, senão o η -ésimo, estes gradientes são nulos, uma vez que as componentes do modelo em (3.7) são independentes uma das outras. Portanto, $\nabla_{\mathbf{p}} \hat{y}_\eta(k)$ e $\nabla_{\boldsymbol{\alpha}} \hat{y}_\eta(k)$ são, respectivamente, calculados como:

$$\nabla_{\mathbf{p}} \hat{y}_\eta(k) = \left[\nabla_{\mathbf{p}_1} \hat{y}_\eta(k) \cdots \nabla_{\mathbf{p}_l} \hat{y}_\eta(k) \cdots \nabla_{\mathbf{p}_\eta} \hat{y}_\eta(k) \right]^T \quad (3.11)$$

$$\nabla_{\mathbf{p}_l} \hat{y}_\eta(k) = \sum_{i_1=1}^{M_{i_1}} \cdots \sum_{i_\eta=1}^{M_{i_\eta}} \alpha_{i_1, \dots, i_\eta} \left[\nabla_{\mathbf{p}_l} \check{\psi}_{l, i_l}(k) \prod_{\substack{j=1 \\ j \neq l}}^{\eta} \check{\psi}_{j, i_j}(k) \right], \quad (3.12)$$

e

$$\nabla_{\boldsymbol{\alpha}} \hat{y}_\eta(k) = \left[\frac{\partial \hat{y}_\eta(k)}{\partial \alpha_{i_1, \dots, i_1}} \cdots \frac{\partial \hat{y}_\eta(k)}{\partial \alpha_{i_n, \dots, i_l}} \cdots \frac{\partial \hat{y}_\eta(k)}{\partial \alpha_{i_\eta, \dots, i_\eta}} \right]^T \quad (3.13)$$

$$\frac{\partial \hat{y}_\eta(k)}{\partial \alpha_{i_1, \dots, i_\eta}} = \check{\psi}_{\eta, i_1}(k) \cdot \dots \cdot \check{\psi}_{\eta, i_\eta}(k).$$

O cálculo de (3.12) pode ser verificado expandindo o segundo termo da equação (3.7) e calculando o gradiente com relação ao vetor de polos \mathbf{p}_l da base $\check{\psi}_{l, i_l}(k)$, como a seguir:

$$\begin{aligned} \hat{y}_\eta(k) &= \sum_{i_1=1}^{M_{i_1}} \cdots \sum_{i_\eta=1}^{M_{i_\eta}} \alpha_{i_1, \dots, i_\eta} \left(\check{\psi}_{1, i_1}(k) \cdot \dots \cdot \check{\psi}_{l, i_l}(k) \cdot \dots \cdot \check{\psi}_{\eta, i_\eta}(k) \right) \\ \nabla_{\mathbf{p}_l} \hat{y}_\eta(k) &= \sum_{i_1=1}^{M_{i_1}} \cdots \sum_{i_\eta=1}^{M_{i_\eta}} \alpha_{i_1, \dots, i_\eta} \left(\check{\psi}_{1, i_1}(k) \cdot \dots \cdot \nabla_{\mathbf{p}_l} \check{\psi}_{l, i_l}(k) \cdot \dots \cdot \check{\psi}_{\eta, i_\eta}(k) \right). \end{aligned} \quad (3.14)$$

levando à generalização apresentada em (3.12).

Apenas para efeito de melhor entendimento a respeito do cálculo dos gradientes, considere um modelo de Volterra de segunda ordem. Isto significa fazer $\eta = 2$ na equação (3.7). Além disso, considere que o número de funções utilizadas em cada direção é igual a M . Assim o termo de segunda ordem é dado por:

$$\begin{aligned} \hat{y}_2(k) &= \sum_{i_1=1}^M \sum_{i_2=1}^M \alpha_{i_1, i_2} \sum_{\tau_1=0}^k \sum_{\tau_2=0}^k \psi_{1, i_1}(\tau_1) u(k - \tau_1) \psi_{2, i_2}(\tau_2) u(k - \tau_2) \\ &= \sum_{i_1=1}^M \sum_{i_2=1}^M \alpha_{i_1, i_2} \check{\psi}_{1, i_1}(k) \check{\psi}_{2, i_2}(k). \end{aligned} \quad (3.15)$$

O gradiente de $\hat{y}_2(k)$ com relação ao vetor \mathbf{p} composto pelos vetores de polos \mathbf{p}_1 e \mathbf{p}_2 que parametrizam as bases $\{\check{\psi}_{1, i_1}(k)\}$ e $\{\check{\psi}_{2, i_2}(k)\}$, respectivamente, resulta em:

$$\begin{aligned}
\nabla_{\mathbf{p}_1} \hat{y}_2(k) &= \sum_{i_1=1}^M \sum_{i_2=1}^M \alpha_{i_1, i_2} (\nabla_{\mathbf{p}_1} \check{\psi}_{1, i_1}(k) \cdot \check{\psi}_{2, i_2}(k)) \\
\nabla_{\mathbf{p}_2} \hat{y}_2(k) &= \sum_{i_1=1}^M \sum_{i_2=1}^M \alpha_{i_1, i_2} (\check{\psi}_{1, i_1}(k) \cdot \nabla_{\mathbf{p}_2} \check{\psi}_{2, i_2}(k)) \\
\nabla_{\mathbf{p}} \hat{y}_2(k) &= [\nabla_{\mathbf{p}_1} \hat{y}_2(k) \quad \nabla_{\mathbf{p}_2} \hat{y}_2(k)]^T
\end{aligned} \tag{3.16}$$

onde $\mathbf{p}_1 = [b_1 \ c_1]^T$ e $\mathbf{p}_2 = [b_2 \ c_2]^T$ quando os conjuntos de funções ortonormais adotados são a base de Kautz, ou $\mathbf{p}_1 = [p_{1,1} \ \cdots \ p_{1,M}]^T$ e $\mathbf{p}_2 = [p_{2,1} \ \cdots \ p_{2,M}]^T$ quando os conjuntos adotados são as bases GOBF com polos reais, ou $\mathbf{p}_1 = [\beta_{1,1} \ \cdots \ \beta_{1,M}]^T$ e $\mathbf{p}_2 = [\beta_{2,1} \ \cdots \ \beta_{2,M}]^T$ quando os conjuntos adotados são as bases GOBF com polos complexos.

A ideia é utilizar as equações de $\nabla_{\mathbf{p}_l} \check{\psi}_{l, i_l}(k)$ junto ao algoritmo de otimização, no caso o algoritmo de Levenberg-Marquardt (Levenberg, 1944; Marquardt, 1963; Nocedal and Wright, 1999; Nelles, 2001). Esta escolha decorre do fato que este algoritmo trabalha com uma aproximação quadrática de J na vizinhança dos argumentos (parâmetros do modelo – \mathbf{p} e α). Usando somente informação das derivadas de primeira ordem obtidas por meio do cálculo dos gradientes, o algoritmo aproxima a matriz Hessiana de J . Esta estimativa da matriz Hessiana é gerada usando a soma dos produtos dos gradientes em (3.5) e (3.6). A atualização da variável de otimização θ ocorre de acordo com a seguinte equação (Nelles, 2001):

$$\theta_n = \theta_{n-1} - [\nabla_{\theta} J_{n-1}(\theta) \nabla_{\theta} J_{n-1}^T(\theta) + \epsilon_{n-1} I]^{-1} \nabla_{\theta} J_{n-1}(\theta) J_{n-1}(\theta) \tag{3.17}$$

onde ϵ_{n-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, como dito anteriormente, utiliza somente avaliações da função de custo e seu gradiente. É um algoritmo globalmente convergente, o que significa dizer que dado um ponto inicial qualquer θ^0 , este converge para um ponto θ^* satisfazendo as condições de otimalidade necessárias para um minimizador local de $J(\theta)$, ou ainda, $\nabla_{\theta} J = 0$. No entanto, não é possível garantir que o ponto θ^* seja o minimizador global de $J(\theta)$, a menos que $J(\theta)$ seja uma função convexa (Nocedal and Wright, 1999; Bazaraa et al., 2006). O algoritmo de Levenberg-Marquardt pode ser interpretado como uma combinação dos métodos do gradiente de maior descida (*Steepest Descent*) e de Gauss-Newton (Nelles, 2001): longe da solução ótima, o algoritmo se comporta como o método do gradiente, lento mas com convergência garantida, implicando num alto valor para ϵ na equação (3.17); próximo da solução ótima, o algoritmo se comporta como o método de Gauss-Newton (ϵ pequeno) que o faz convergir mais rapidamente.

3.2 Gradientes das Saídas dos Filtros

Nesta seção, são desenvolvidas as equações dos gradientes das saídas dos filtros com relação aos polos da base ortonormal, ou seja, $\nabla_{\mathbf{p}_l} \check{\psi}_{l, i_l}(k)$ em (3.5) ou (3.12) que, posteriormente, serão utilizados junto com o algoritmo de otimização nos modelos de Volterra com uma base independente para cada direção.

3.2.1 Funções de Kautz

Dois trabalhos anteriores apresentam duas formulações distintas para o cálculo das derivadas das funções de Kautz. Em da Rosa et al. (2009), propõe-se uma abordagem para o cômputo das derivadas onde os cálculos são feitos individualmente para cada filtro com relação aos parâmetros b e c . Já em Machado (2011), o cálculo é feito em batelada (*batch*), isto é, todo o cálculo dos gradientes é realizado de maneira simultânea para todos os filtros (Veja Apêndice A). Embora essa última abordagem seja muito eficaz em relação à primeira, ela ainda conserva algumas desvantagens, como, por exemplo, o elevado tempo de processamento para proceder aos cálculos das derivadas. Isto ocorre porque, em ambas as abordagens, para cada nova amostra k , todos os cálculos devem ser repetidos integralmente. Ou seja, para cada nova amostra, as equações (A.3) e (A.5) devem ser recalculadas partindo da medida do instante atual até as condições iniciais, $k = 0$, onerando o tempo de processamento quando o número de medidas k aumenta. A abordagem desenvolvida a seguir permite calcular o gradiente analiticamente de forma recursiva e em batelada. Utilizando as informações do instante $k - 1$ para o cálculo no instante k , reduz-se o tempo de processamento e o esforço computacional envolvido. Outra vantagem é a possibilidade de implementação da técnica em tempo real (*on-line*).

Para a base de Kautz, convenientemente, adota-se o vetor \mathbf{p}_l contendo os parâmetros (reais) que descrevem os polos de Kautz. E, portanto,

$$\mathbf{p}_l \triangleq \begin{bmatrix} b_l \\ c_l \end{bmatrix} \Rightarrow \nabla_{\mathbf{p}_l} \check{\psi}_{l,i_l}(k) = \begin{bmatrix} \frac{\partial \check{\psi}_{l,i_l}(k)}{\partial b_l} \\ \frac{\partial \check{\psi}_{l,i_l}(k)}{\partial c_l} \end{bmatrix}. \quad (3.18)$$

O cálculo do gradiente utiliza o modelo das funções de Kautz em espaço de estado, como descrito na equação (2.31), sendo as saídas dos filtros $\check{\psi}_{l,i_l}(k)$ descritas como:

$$\begin{aligned} \mathcal{X}(k+1) &= \mathcal{A}_K \mathcal{X}(k) + \mathcal{B}_K \mathcal{U}(k) \\ \check{\Psi}_l(k) &= \mathcal{C}_{K_m} \mathcal{X}(k) \end{aligned} \quad (3.19)$$

onde $\mathcal{X}(k)$, $\mathcal{U}(k)$, \mathcal{A}_K e \mathcal{B}_K são as mesmas matrizes definidas na Seção 2.3.3 para as funções de Kautz, \mathcal{C}_{K_m} e $\check{\Psi}_l(k)$ são definidas, respectivamente, como:

$$\mathcal{C}_{K_m} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

e

$$\check{\Psi}_l(k) = [\check{\psi}_{l,1}(k) \quad \check{\psi}_{l,2}(k) \quad \cdots \quad \check{\psi}_{l,n-1}(k) \quad \check{\psi}_{l,n}(k)]^T.$$

O modelo em espaço de estados dos filtros ortonormais de Kautz (3.19) apresenta uma recursão no estado, ou ainda, o estado $\mathcal{X}(k+1)$ depende do estado no instante k , ou seja, $\mathcal{X}(k)$. Assim, mantendo a estrutura em espaço de estado e tomando a derivada com relação ao parâmetro b_l chega-se a:

$$\frac{\partial \mathcal{X}(k+1)}{\partial b_l} = \frac{\partial \mathcal{A}_K}{\partial b_l} \mathcal{X}(k) + \mathcal{A}_K \frac{\partial \mathcal{X}(k)}{\partial b_l} + \frac{\partial \mathcal{B}_K}{\partial b_l} \mathcal{U}(k) \quad (3.20)$$

$$\frac{\partial \check{\Psi}_l(k+1)}{\partial b_l} = \mathcal{C}_{K_m} \frac{\partial \mathcal{X}(k+1)}{\partial b_l}. \quad (3.21)$$

As derivadas das matrizes \mathcal{A}_K e \mathcal{B}_K com relação ao parâmetro b_l são dadas por (Machado, 2011):

$$\frac{\partial \mathcal{A}_K}{\partial b_l} = \begin{bmatrix} 0 & dab_{1,2} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ dab_{2,1} & dab_{2,2} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & dab_{3,4} & \cdots & 0 & 0 & 0 & 0 \\ 0 & dab_{4,2} & 0 & dab_{4,4} & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & dab_{2n-1,2n} \\ 0 & 0 & 0 & 0 & \cdots & 0 & dab_{2n,2n-2} & 0 & dab_{2n,2n} \end{bmatrix} \quad (3.22)$$

onde os elementos $dab_{i,j}$ pertencentes à matriz $\partial \mathcal{A}_K / \partial b_l$, com i indicando a linha e j a coluna do respectivo elemento, são definidos por:

- Para i ímpar:
 - todos os termos iguais a 0.
- Para $i = 2$:
 - $dab_{i,i} = -(c_l - 1)$;
 - demais termos iguais a 0.
- Para i múltiplo de 4:
 - $dab_{i,i} = 1$;
 - $dab_{i,i-2} = \frac{-b_l}{\sqrt{1-b_l^2}}$;
 - demais termos iguais a 0.
- Para outras linhas i pares, os termos são:
 - $dab_{i,i} = -(c_l - 1)$;
 - $dab_{i,i-4j} = (-c_l)^{j-1} [(c_l - 1) + (-c_l) (- (c_l - 1))]$ para $j = 0, 1, 2, \dots$ tal que $4j < i$;
 - demais termos iguais a 0.

Para a matriz \mathcal{B}_K , obtém-se a derivada com relação ao parâmetro b_l como (Machado, 2011):

$$\frac{\partial \mathcal{B}_K}{\partial b_l} = \sqrt{1 - c_l^2} \begin{bmatrix} 0 & 0 \\ dbb_{2,1} & dbb_{2,2} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ dbb_{6,1} & dbb_{6,2} \\ \vdots & \vdots \\ 0 & 0 \\ dbb_{2(n-1),1} & dbb_{2(n-1),2} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad (3.23)$$

cujos elementos são definidos como:

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

Assim, resta apenas o cálculo recursivo da derivada dos filtros ortonormais com respeito ao parâmetro c_l , o qual se procede de forma similar, bastando proceder a substituição de b_l por c_l nas equações (3.20) e (3.21), ou seja:

$$\frac{\partial \mathcal{X}(k+1)}{\partial c_l} = \frac{\partial \mathcal{A}_K}{\partial c_l} \mathcal{X}(k) + \mathcal{A}_K \frac{\partial \mathcal{X}(k)}{\partial c_l} + \frac{\partial \mathcal{B}_K}{\partial c_l} \mathcal{U}(k) \quad (3.24)$$

$$\frac{\partial \check{\Psi}_l(k+1)}{\partial c_l} = \mathcal{C}_{K_m} \frac{\partial \mathcal{X}(k+1)}{\partial c_l} \quad (3.25)$$

onde as matrizes \mathcal{A}_K , \mathcal{B}_K e \mathcal{C}_{K_m} são as citadas anteriormente. Os cálculos das derivadas das duas primeiras matrizes são apresentados a seguir.

A derivada da matriz \mathcal{A}_K com relação ao parâmetro c_l é dada por (Machado, 2011):

$$\frac{\partial \mathcal{A}_K}{\partial c_l} = \begin{bmatrix} 0 & dac_{1,2} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ dac_{2,1} & dac_{2,2} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & dac_{3,4} & \cdots & 0 & 0 & 0 & 0 \\ 0 & dac_{4,2} & 0 & dac_{4,4} & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & dac_{2n-1,2n} \\ 0 & 0 & 0 & 0 & \cdots & 0 & dac_{2n,2n-2} & 0 & dac_{2n,2n} \end{bmatrix} \quad (3.26)$$

onde os elementos $dac_{i,j}$ pertencentes à matriz $\partial\mathcal{A}_K/\partial c_l$, com i indicando a linha e j a coluna do respectivo elemento, são definidos como:

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

Já a matriz \mathcal{B}_K possui sua derivada com relação ao parâmetro c_l definida por (Machado, 2011):

$$\frac{\partial \mathcal{B}_K}{\partial c_l} = \sqrt{1 - c_l^2} \begin{bmatrix} 0 & 0 \\ dbc_{2,1} & dbc_{2,2} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ dbc_{6,1} & dbc_{6,2} \\ \vdots & \vdots \\ 0 & 0 \\ dbc_{2(n-1),1} & dbc_{2(n-1),2} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad (3.27)$$

cujos elementos são:

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

$$\begin{aligned}
- \text{dbc}_{i,1} &= (-c_l)^j + \left(\frac{1-c_l^2}{c_l} \right) (j)(-c_l)^{j-1}; \\
- \text{dbc}_{i,2} &= (-b_l)(-c_l)^j + \left(\frac{1-c_l^2}{c_l} \right) (j)(-c_l)^{j-1}.
\end{aligned}$$

- Para os demais i 's:

$$- \text{dbc}_{i,1} = \text{dbc}_{i,2} = 0.$$

É válido destacar que tanto o vetor $\partial \mathcal{X}(0)/\partial b_l$ quanto $\partial \mathcal{X}(0)/\partial c_l$ são nulos devido às condições iniciais.

3.2.2 Funções Ortonormais Generalizadas (GOBF)

Sistemas cujos *kernels* apresentam dinâmicas mais complexas são melhor representados por modelos baseados em funções ortonormais generalizadas (GOBF), já que a descrição matemática destas funções envolve parâmetros associados com múltiplas dinâmicas por incorporarem uma variedade de polos, reais (dinâmicas amortecidas) e complexos (dinâmicas sub-amortecidas), como já discutido no Capítulo 2.

Em da Rosa et al. (2009), o cálculo da derivada da saída de cada filtro GOBF foi obtido individualmente, assim como neste trabalho. Entretanto, as equações apresentadas nessa seção possuem a vantagem de poderem ser implementadas recursivamente, o que, como já comentado, permite reduzir o tempo de processamento. Na apresentação dos exemplos, um estudo comparativo será feito mostrando a superioridade, no que tange ao tempo de processamento, quando são utilizadas as equações recursivas apresentadas aqui e não aquelas apresentadas em Machado (2011) e da Rosa et al. (2009) (Apêndice A).

Caso GOBF com polos reais

Quando a base de funções ortonormais generalizadas é parametrizada por um conjunto de polos reais e distintos, i.e., fazendo $\beta_{l,j}$ na equação (2.24) igual a $p_{l,j}$, o vetor de polos pode ser escrito como $\mathbf{p}_l = [p_{l,1} \ p_{l,2} \ \dots \ p_{l,M}]^T$ isto significa dizer que a base é composta por M funções e aquela equação pode ser reescrita como:

$$\Psi_{l,m}(z) = \frac{z \sqrt{1 - |p_{l,m}|^2}}{z - p_{l,m}} \prod_{j=1}^{m-1} \left(\frac{1 - p_{l,j}z}{z - p_{l,j}} \right) \quad m = 1, 2, \dots \quad (3.28)$$

É válido destacar que a m -ésima função da base l depende apenas dos m primeiros polos $\{p_{l,1}, \dots, p_{l,m}\}$ e não depende dos demais polos $\{p_{l,m+1}, \dots, p_{l,M}\}$. Portanto, $\partial \check{\psi}_{l,m}(k)/\partial p_{l,k} = 0$ para $k > m$ e o gradiente $\nabla_{\mathbf{p}_l} \check{\psi}_{l,m}(k)$ é dado por:

$$\begin{aligned}
\nabla_{\mathbf{p}_l} \check{\psi}_{l,m}(k) &= \left[\frac{\partial \check{\psi}_{l,m}(k)}{\partial p_{l,1}} \quad \dots \quad \frac{\partial \check{\psi}_{l,m}(k)}{\partial p_{l,m}} \quad \dots \quad \frac{\partial \check{\psi}_{l,m}(k)}{\partial p_{l,M}} \right]^T \\
&= \left[\frac{\partial \check{\psi}_{l,m}(k)}{\partial p_{l,1}} \quad \dots \quad \frac{\partial \check{\psi}_{l,m}(k)}{\partial p_{l,m}} \quad \underbrace{0 \quad \dots \quad 0}_{(M-m) \text{ vezes}} \right]^T
\end{aligned} \tag{3.29}$$

A saída do primeiro filtro ($m = 1$) pode ser reescrita de acordo com a equação (3.28) da seguinte maneira:

$$\check{\Psi}_{l,1}(z) = \frac{z\sqrt{1-p_{l,1}^2}}{z-p_{l,1}}U(z) \tag{3.30}$$

que no domínio do tempo pode ser reescrita como:

$$\check{\psi}_{l,1}(k+1) = p_{l,1}\check{\psi}_{l,1}(k) + \sqrt{1-p_{l,1}^2}u(k+1). \tag{3.31}$$

Logo, a derivada do primeiro filtro com relação ao polo $p_{l,1}$ é dada por:

$$\frac{\partial \check{\psi}_{l,1}(k+1)}{\partial p_{l,1}} = \check{\psi}_{l,1}(k) + p_{l,1} \frac{\partial \check{\psi}_{l,1}(k)}{\partial p_{l,1}} - \frac{p_{l,1}}{\sqrt{1-p_{l,1}^2}}u(k+1) \tag{3.32}$$

Como o polo $p_{l,1}$ parametriza apenas a base $\check{\psi}_{l,1}(k)$, logo $\partial \check{\psi}_{l,1}(k)/\partial p_{l,k} = 0$ para $k = 2, \dots, M$.

Para os demais filtros, ou ainda, para $m \geq 2$, utiliza-se as equações (3.28) e (3.2), com as quais pode-se escrever:

$$\check{\Psi}_{l,m}(z) = \sqrt{\frac{1-p_{l,m}^2}{1-p_{l,m-1}^2}} \left(\frac{1-p_{l,m-1}z}{z-p_{l,m}} \right) \check{\Psi}_{l,m-1}(z), \tag{3.33}$$

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

$$\check{\psi}_{l,m}(k+1) = p_{l,m}\check{\psi}_{l,m}(k) + \sqrt{\frac{1-p_{l,m}^2}{1-p_{l,m-1}^2}} (\check{\psi}_{l,m-1}(k) - p_{l,m-1}\check{\psi}_{l,m-1}(k+1)). \tag{3.34}$$

Para obter as equações do gradiente dos filtros é necessário calcular as derivadas de $\check{\psi}_{l,m}(k+1)$ em (3.34) com relação a $p_{l,k}$ para $k = 1, \dots, m$. A derivada de (3.34) com relação ao polo $p_{l,m}$ é dado por:

$$\begin{aligned} \frac{\partial \check{\psi}_{l,m}(k+1)}{\partial p_{l,m}} &= \check{\psi}_{l,m}(k) + p_{l,m} \frac{\partial \check{\psi}_{l,m}(k)}{\partial p_{l,m}} + \\ &+ \frac{1}{\sqrt{(1-p_{l,m}^2)(1-p_{l,m-1}^2)}} \left[\check{\psi}_{l,m-1}(k) - p_{l,m-1} \check{\psi}_{l,m-1}(k+1) \right] \end{aligned} \quad (3.35)$$

onde utiliza-se o fato de que $\partial \check{\psi}_{l,m-1}(k)/\partial p_{l,m} = 0$, pois a função $\check{\psi}_{l,m-1}(k)$ depende somente do conjunto de polos $\{p_{l,1}, p_{l,2}, \dots, p_{l,m-1}\}$.

A derivada de (3.34) com relação ao polo $p_{l,m-1}$ é calculada como:

$$\begin{aligned} \frac{\partial \check{\psi}_{l,m}(k+1)}{\partial p_{l,m-1}} &= p_{l,m} \frac{\partial \check{\psi}_{l,m}(k)}{\partial p_{l,m-1}} + \sqrt{\frac{1-p_{l,m}^2}{(1-p_{l,m-1}^2)^3}} \left(\check{\psi}_{l,m-1}(k) - p_{l,m-1} \check{\psi}_{l,m-1}(k+1) \right) \\ &+ \sqrt{\frac{1-p_{l,m}^2}{1-p_{l,m-1}^2}} \left(\frac{\partial \check{\psi}_{l,m-1}(k)}{\partial p_{l,m-1}} - \check{\psi}_{l,m-1}(k+1) - p_{l,m-1} \frac{\partial \check{\psi}_{l,m-1}(k+1)}{\partial p_{l,m-1}} \right). \end{aligned} \quad (3.36)$$

Por fim, as derivadas de (3.34) com relação a $p_{l,k}$, $k = 1, \dots, m-2$, são calculadas usando a equação:

$$\frac{\partial \check{\psi}_{l,m}(k+1)}{\partial p_{l,k}} = p_{l,m} \frac{\partial \check{\psi}_{l,m}(k)}{\partial p_{l,k}} + \sqrt{\frac{1-p_{l,m}^2}{1-p_{l,m-1}^2}} \left(\frac{\partial \check{\psi}_{l,m-1}(k)}{\partial p_{l,k}} - p_{l,m-1} \frac{\partial \check{\psi}_{l,m-1}(k+1)}{\partial p_{l,k}} \right) \quad (3.37)$$

Lembrando que $\partial \check{\psi}_{l,M}(0)/\partial p_{l,M} = 0$, devido às condições iniciais, para $M = 1, 2, \dots, m$.

Caso GOBF com um par de polos complexos

Como pode ser verificado pela equação (2.24), em geral, as funções ortonormais generalizadas possuem transformada Z inversa com coeficientes complexos. Esta desvantagem pode ser superada (Ninness and Gustafsson, 1997), construindo-se uma nova base de funções ortonormais com respostas ao impulso reais. Parte deste desenvolvimento será apresentado a seguir.

Suponha um modelo baseado em funções ortonormais, tal que as $M-1$ primeiras funções da base $\{\Psi_{l,1}(z), \Psi_{l,2}(z), \dots, \Psi_{l,M-1}(z)\}$ sejam parametrizadas pelos seus $M-1$ respectivos polos reais, dados no conjunto $\{p_{l,1}, p_{l,2}, \dots, p_{l,M-1}\}$ e que se deseja adicionar um polo complexo $\beta_{l,M}$ nesse conjunto de polos. De acordo com Ninness and Gustafsson (1997), para cada inclusão de um polo complexo $\beta_{l,M}$ duas novas funções $\Psi'_{l,M}(z)$ e $\Psi''_{l,M}(z)$ com respostas real ao impulso são construídas a partir de uma combinação linear de $\Psi_{l,M}(z)$ e $\Psi_{l,M+1}(z)$ geradas por (2.24). Portanto, o conjunto de funções será dado por $\{\Psi_{l,1}(z), \Psi_{l,2}(z), \dots, \Psi_{l,M-1}(z), \Psi'_{l,M}(z), \Psi''_{l,M}(z)\}$, tendo como polos o conjunto $\{p_{l,1}, p_{l,2}, \dots, p_{l,M-1}, \beta_{l,m}, \bar{\beta}_{l,m}\}$.

A combinação linear que gera as funções é definida como:

$$\begin{bmatrix} \Psi'_{l,M}(z) \\ \Psi''_{l,M}(z) \end{bmatrix} = \begin{bmatrix} c_0 & c_1 \\ c'_0 & c'_1 \end{bmatrix} \cdot \begin{bmatrix} \Psi_{l,M}(z) \\ \Psi_{l,M+1}(z) \end{bmatrix} \quad (3.38)$$

onde c_0, c_1, c'_0, c'_1 são escalares complexos arbitrários e possivelmente distintos. Para garantir a propriedade de ortonormalidade das funções geradas é necessário que as seguintes restrições sejam atendidas (Ziaei and Wang, 2006):

$$|c_0| + |c_1| = 1, \quad (3.39)$$

$$|c'_0| + |c'_1| = 1, \quad (3.40)$$

$$c_0 \bar{c}'_0 + c_1 \bar{c}'_1 = 0. \quad (3.41)$$

A equação (2.24) também pode ser reescrita como:

$$\Psi_{l,M}(z) = \frac{\sqrt{1 - |\beta_{l,M}|^2}}{1 - |\beta_{l,M-1}|^2} \left(\frac{1 - \bar{\beta}_{l,M-1}z}{z - \beta_{l,M}} \right) \Psi_{l,M-1}(z) \quad (3.42)$$

Substituindo a equação (3.42) em (3.38), chega-se à descrição matemática da base GOBF quando um polo complexo $\beta_{l,M}$ é incorporado ao modelo (Ninness and Gustafsson, 1997):

$$\begin{aligned} \Psi'_{l,M}(z) &= \frac{z\sqrt{1 - |\beta_{l,M}|^2}(\lambda'z + \gamma')}{z^2 - (\beta_{l,M} + \bar{\beta}_{l,M})z + |\beta_{l,M}|^2} \prod_{j=1}^{M-1} \left(\frac{1 - p_{l,j}z}{z - p_{l,j}} \right) \\ \Psi''_{l,M}(z) &= \frac{z\sqrt{1 - |\beta_{l,M}|^2}(\lambda''z + \gamma'')}{z^2 - (\beta_{l,M} + \bar{\beta}_{l,M})z + |\beta_{l,M}|^2} \prod_{j=1}^{M-1} \left(\frac{1 - p_{l,j}z}{z - p_{l,j}} \right) \end{aligned} \quad (3.43)$$

onde $\beta_{l,M}, \bar{\beta}_{l,M} \in \mathcal{C}$ é um par de polos complexos conjugados incorporados na base ortonormal GOBF, e os parâmetros reais $\lambda', \gamma', \lambda''$ e γ'' estão relacionados com o polo complexo $\beta_{l,M}$ da seguinte forma (da Rosa, 2009):

$$\begin{bmatrix} \lambda' & \gamma' \end{bmatrix} \begin{bmatrix} 1 + |\beta_{l,M}|^2 & \beta_{l,M} + \bar{\beta}_{l,M} \\ \beta_{l,M} + \bar{\beta}_{l,M} & 1 + |\beta_{l,M}|^2 \end{bmatrix} \begin{bmatrix} \lambda' \\ \gamma' \end{bmatrix} = |1 - \beta_{l,M}^2|^2 \quad (3.44)$$

$$\begin{bmatrix} \lambda'' \\ \gamma'' \end{bmatrix} = \frac{1}{\sqrt{1 - \xi^2}} \begin{bmatrix} \xi & 1 \\ -1 & -\xi \end{bmatrix} \begin{bmatrix} \lambda' \\ \gamma' \end{bmatrix} \quad (3.45)$$

$$\xi = \frac{\beta_{l,M} + \bar{\beta}_{l,M}}{1 + |\beta_{l,M}|^2} \quad (3.46)$$

Dado que a matriz apresentada em (3.44) é definida positiva, os vetores que pré e pós-multiplicam esta matriz encontram-se numa elipse e, portanto, admitem infinitas soluções. A princípio, qualquer

escolha destes parâmetros pode ser adotada, pois, satisfazendo as equações (3.44) - (3.46), a ortonormalidade da base está garantida. Para uma escolha particular de um dos parâmetros, por exemplo $\gamma' = 0$, teríamos a seguinte solução para os parâmetros restantes:

$$\begin{aligned}\lambda' &= \frac{|1 - \beta_{l,M}^2|}{\sqrt{1 + |\beta_{l,M}|^2}} \\ \lambda'' &= \frac{|1 - \beta_{l,M}^2|}{\sqrt{1 + |\beta_{l,M}|^2}} \\ \gamma'' &= -\sqrt{1 + |\beta_{l,M}|^2}.\end{aligned}\tag{3.47}$$

De posse das equações (3.47) e (3.2), a seguinte relação pode ser deduzida:

$$\check{\Psi}'_{l,M}(z) = \sqrt{\frac{1 - |\beta_{l,M}|^2}{1 - p_{l,M-1}^2}} \cdot \frac{(\lambda'z + \gamma')(1 - p_{l,M-1}z)}{z^2 - (\beta_{l,M} + \bar{\beta}_{l,M})z + |\beta_{l,M}|^2} \cdot \check{\Psi}_{l,M-1}(z).\tag{3.48}$$

Tomando a equação (3.48) no domínio do tempo, como feito para o caso onde apenas polos reais parametrizavam as bases, e usando uma descrição em espaço de estados, a equação torna-se:

$$\begin{cases} x'_{l,M}(k+1) = Ax'_{l,M}(k) + B_1\check{\psi}_{l,M-1}(k+2) + B_2\check{\psi}_{l,M-1}(k+1) + B_3\check{\psi}_{l,M-1}(k) \\ \check{\psi}'_{l,M}(k) = Cx'_{l,M}(k) \end{cases}\tag{3.49}$$

onde as matrizes A , B_1 , B_2 , B_3 , C são definidas como (da Rosa et al., 2009):

$$\begin{aligned}A &= \begin{bmatrix} 0 & 1 \\ -|\beta_{l,M}|^2 & \beta_{l,M} + \bar{\beta}_{l,M} \end{bmatrix} & B_1 &= K_M \begin{bmatrix} 0 \\ -\lambda'p_{l,M-1} \end{bmatrix} \\ B_2 &= K_M \begin{bmatrix} 0 \\ \lambda' - \gamma'p_{l,M-1} \end{bmatrix} & B_3 &= K_M \begin{bmatrix} 0 \\ \gamma' \end{bmatrix} \\ C &= [1 \quad 0] & K_M &= \sqrt{\frac{1 - |\beta_{l,M}|^2}{1 - p_{l,M-1}^2}}\end{aligned}$$

Os gradientes das saídas dos filtros ortonormais utilizando GOBF com um par de polos complexos são definidos tratando o polo $\beta_{l,M}$ como um vetor bidimensional real, ou seja, define-se $\beta_{l,M} \triangleq [\mathbf{Re}(\beta_{l,M}) \quad \mathbf{Im}(\beta_{l,M})]^T$. O gradiente da função ortonormal passa a ser definido como $\nabla_{\beta_{l,M}}\check{\psi}'_{l,M}(k) \triangleq [\partial\check{\psi}'_{l,M}(k)/\partial\mathbf{Re}(\beta_{l,M}) \quad \partial\check{\psi}'_{l,M}(k)/\partial\mathbf{Im}(\beta_{l,M})]^T$, cujo primeiro elemento (derivada com relação à parte real) é dado por:

$$\begin{aligned} \frac{\partial x'_{l,M}(k+1)}{\partial \text{Re}(\beta_{l,M})} &= \frac{\partial A}{\partial \text{Re}(\beta_{l,M})} x'_{l,M}(k) + A \frac{\partial x'_{l,M}(k)}{\partial \text{Re}(\beta_{l,M})} + \frac{\partial B_1}{\partial \text{Re}(\beta_{l,M})} \check{\psi}_{l,M-1}(k+2) \\ &\quad + \frac{\partial B_2}{\partial \text{Re}(\beta_{l,M})} \check{\psi}_{l,M-1}(k+1) + \frac{\partial B_3}{\partial \text{Re}(\beta_{l,M})} \check{\psi}_{l,M-1}(k) \end{aligned} \quad (3.50)$$

$$\frac{\partial \check{\psi}'_{l,M}(k+1)}{\partial \text{Re}(\beta_{l,M})} = C \cdot \frac{\partial x'_{l,M}(k+1)}{\partial \text{Re}(\beta_{l,M})} \quad (3.51)$$

onde utilizou-se o fato de $\check{\psi}'_{l,M-1}$ não depender do polo $\beta_{l,M}$ e, portanto, sua derivada com relação a ele é nula. Os cálculos das derivadas das matrizes A , B_1 , B_2 e B_3 com relação à parte real são desenvolvidos a seguir.

Reescrevendo a matriz A em função das componentes do polo $\beta_{l,M}$ tem-se:

$$A = \begin{bmatrix} 0 & 1 \\ -\text{Re}(\beta_{l,M})^2 - \text{Im}(\beta_{l,M})^2 & 2\text{Re}(\beta_{l,M}) \end{bmatrix}, \quad (3.52)$$

e, portanto, $\partial A / \partial \text{Re}(\beta_{l,M})$ é dada por:

$$\frac{\partial A}{\partial \text{Re}(\beta_{l,M})} = \begin{bmatrix} 0 & 0 \\ -2\text{Re}(\beta_{l,M}) & 2 \end{bmatrix}.$$

De forma similar, as derivadas das demais matrizes com relação à parte real do polo $\beta_{l,M}$ são calculadas da seguinte maneira:

$$\frac{\partial B_1}{\partial \text{Re}(\beta_{l,M})} = \frac{\partial K_M}{\partial \text{Re}(\beta_{l,M})} \begin{bmatrix} 0 \\ -\lambda' p_{l,M-1} \end{bmatrix} \quad \frac{\partial B_2}{\partial \text{Re}(\beta_{l,M})} = \frac{\partial K_M}{\partial \text{Re}(\beta_{l,M})} \begin{bmatrix} 0 \\ \lambda' - \gamma' p_{l,M-1} \end{bmatrix}$$

$$\frac{\partial B_3}{\partial \text{Re}(\beta_{l,M})} = \frac{\partial K_M}{\partial \text{Re}(\beta_{l,M})} \begin{bmatrix} 0 \\ \gamma' \end{bmatrix} \quad \frac{\partial K_{l,M}}{\partial \text{Re}(\beta_{l,M})} = \frac{-\text{Re}(\beta_{l,M})}{\sqrt{(1 - |\beta_{l,M}|^2)(1 - p_{l,M-1}^2)}}$$

Já a derivada com relação a parte imaginária de $\beta_{l,M}$ é idêntica às equações (3.50) e (3.51), bastando proceder a respectiva substituição de $\text{Re}(\beta_{l,M})$ por $\text{Im}(\beta_{l,M})$ naquelas equações, inclusive nas derivadas das matrizes B_1 , B_2 e B_3 . Somente a derivada de A com relação à parte imaginária do polo $\beta_{l,M}$ apresenta uma mudança significativa e pode ser obtida a partir de (3.52), sendo calculada como:

$$\frac{\partial A}{\partial \text{Im}(\beta_{l,M})} = \begin{bmatrix} 0 & 0 \\ -2\text{Im}(\beta_{l,M}) & 0 \end{bmatrix}.$$

As derivadas com relação a $p_{l,k}$ ($k = 1, \dots, M-1$) são calculadas como:

$$\begin{aligned} \frac{\partial x'_{l,M}(k+1)}{\partial p_{l,k}} &= A \frac{\partial x'_{l,M}(k)}{\partial p_{l,k}} + \frac{\partial B_1}{\partial p_{l,k}} \check{\psi}_{l,M-1}(k+2) + B_1 \frac{\partial \check{\psi}_{l,M-1}(k+2)}{\partial p_{l,k}} + \frac{\partial B_2}{\partial p_{l,k}} \check{\psi}_{l,M-1}(k+1) \\ &\quad + B_2 \frac{\partial \check{\psi}_{l,M-1}(k+1)}{\partial p_{l,k}} + \frac{\partial B_3}{\partial p_{l,k}} \check{\psi}_{l,M-1}(k) + B_3 \frac{\partial \check{\psi}_{l,M-1}(k)}{\partial p_{l,k}} \end{aligned} \quad (3.53)$$

$$\frac{\partial \check{\psi}'_{l,M}(k+1)}{\partial p_{l,k}} = C \cdot \frac{\partial x'_{l,M}(k+1)}{\partial p_{l,k}} \quad (3.54)$$

onde as matrizes A_1 , B_1 , B_2 e B_3 são aquelas definidas anteriormente. É digno de nota que as derivadas de $\partial B_1/\partial p_{l,k}$, $\partial B_2/\partial p_{l,k}$ e $\partial B_3/\partial p_{l,k}$ são nulas se $k \neq M-1$, pois tais matrizes não são dependentes de $p_{l,k}$. Além disso, as expressões $\partial x'_{l,M}(0)/\partial \text{Re}(\beta_{l,M})$, $\partial \check{\psi}_{l,k}(0)/\partial \text{Re}(\beta_{l,M})$, $\partial x'_{l,M}(0)/\partial p_{l,k}$ e $\partial \check{\psi}_{l,k}(0)/\partial p_{l,k}$ são nulas, $\forall k$.

E as derivadas com relação ao polo $p_{l,M-1}$ são calculadas da seguinte forma:

$$\begin{aligned} \frac{\partial B_1}{\partial p_{l,M-1}} &= \begin{bmatrix} 0 \\ -\lambda' K_M - \lambda' p_{l,M-1} \frac{\partial K_M}{\partial p_{l,M-1}} \end{bmatrix} & \frac{\partial B_2}{\partial p_{l,M-1}} &= \begin{bmatrix} 0 \\ (\lambda' - \gamma' p_{l,M-1}) \frac{\partial K_M}{\partial p_{l,M-1}} - \gamma' K_M \end{bmatrix} \\ \frac{\partial B_3}{\partial p_{l,M-1}} &= \frac{\partial K_M}{\partial p_{l,M-1}} \begin{bmatrix} 0 \\ \gamma' \end{bmatrix} & \frac{\partial K_M}{\partial p_{l,M-1}} &= p_{l,M-1} \sqrt{\frac{1 - |\beta_{l,M}|^2}{(1 - p_{l,M-1})^3}} \end{aligned}$$

O cálculo das derivadas das funções complementares $\check{\Psi}''_{l,M}(z)$ procede-se de forma semelhante, necessitando apenas substituir λ' e γ' por λ'' e γ'' , respectivamente, nas equações.

Caso GOBF com dois pares de polos complexos

Como discutido anteriormente, para cada inclusão de um par de polos complexos, duas novas funções são acrescentadas à base ortonormal. Portanto, uma base ortonormal, parametrizada por $M-1$ polos reais e uma par de polos complexos $\beta_{l,M}$, $\bar{\beta}_{l,M}$, é composta pelo conjunto de funções $\{\Psi_{l,1}(z), \Psi_{l,2}(z), \dots, \Psi_{l,M-1}(z), \Psi'_{l,M}(z), \Psi''_{l,M}(z)\}$.

Se for desejado incluir mais um par de polos complexos, $\beta_{l,M+1}$ e $\bar{\beta}_{l,M+1}$ na base GOBF, mais duas novas funções $\Psi'_{l,M+1}(z)$ e $\Psi''_{l,M+1}(z)$ com respostas reais ao impulso são construídas por meio da combinação linear de $\Psi_{l,M+2}(z)$ e $\Psi_{l,M+3}(z)$ conforme definido em (2.24). Assim, o novo conjunto de funções passa a ser $\{\Psi_{l,1}(z), \Psi_{l,2}(z), \dots, \Psi_{l,M-1}(z), \Psi'_{l,M}(z), \Psi''_{l,M}(z), \Psi'_{l,M+1}(z), \Psi''_{l,M+1}(z)\}$, com o correspondente vetor de polos $\{p_{l,1}, p_{l,2}, \dots, p_{l,M-1}, \beta_{l,M}, \bar{\beta}_{l,M}, \beta_{l,M+1}, \bar{\beta}_{l,M+1}\}$. Após algumas manipulações algébricas, as duas novas funções geradas são definidas por:

$$\Psi'_{l,M+1}(z) = \frac{z \sqrt{1 - |\beta_{l,M+1}|^2} (\rho' z + \mu')}{z^2 - (\beta_{l,M+1} + \bar{\beta}_{l,M+1})z + |\beta_{l,M+1}|^2} \left(\frac{1 - \bar{\beta}_{l,M} z}{z - \beta_{l,M}} \right) \left(\frac{1 - \beta_{l,M} z}{z - \bar{\beta}_{l,M}} \right) \prod_{j=1}^{M-1} \left(\frac{1 - p_{l,j} z}{z - p_{l,j}} \right) \quad (3.55)$$

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

onde ρ' , μ' , ρ'' e μ'' são parâmetros reais relacionados com o polo $\beta_{l,M+1}$ como segue:

$$\begin{bmatrix} \rho' & \mu' \end{bmatrix} \begin{bmatrix} 1 + |\beta_{l,M+1}|^2 & \beta_{l,M+1} + \bar{\beta}_{l,M+1} \\ \beta_{l,M+1} + \bar{\beta}_{l,M+1} & 1 + |\beta_{l,M+1}|^2 \end{bmatrix} \begin{bmatrix} \rho' \\ \mu' \end{bmatrix} = |1 - \beta_{l,M+1}^2|^2 \quad (3.56)$$

$$\begin{bmatrix} \rho'' \\ \mu'' \end{bmatrix} = \frac{1}{\sqrt{1 - \xi^2}} \begin{bmatrix} \xi & 1 \\ -1 & -\xi \end{bmatrix} \begin{bmatrix} \rho' \\ \mu' \end{bmatrix} \quad (3.57)$$

$$\xi = \frac{\beta_{l,M+1} + \bar{\beta}_{l,M+1}}{1 + |\beta_{l,M+1}|^2}. \quad (3.58)$$

Assim, como no caso para um par de polos reais, existem infinitas soluções para ρ' , μ' , ρ'' e μ'' que satisfazem as equações (3.56) - (3.58), garantindo a ortonormalidade das novas funções da base.

Para o cálculo das derivadas das saídas dos filtros ortonormais um procedimento similar ao adotado anteriormente é utilizado. Primeiro, utilizando as equações (3.56) e (3.2), pode-se escrever a seguinte relação:

$$\check{\Psi}'_{l,M+1}(z) = \sqrt{\frac{1 - |\beta_{l,M+1}|^2}{1 - |\beta_{l,M+1}|^2}} \left(\frac{\rho' z + \mu'}{\lambda' z + \gamma'} \right) \left(\frac{1 - (\beta_{l,M} + \bar{\beta}_{l,M})z + |\beta_{l,M}|^2 z^2}{z^2 - (\beta_{l,M+1} + \bar{\beta}_{l,M+1})z + |\beta_{l,M+1}|^2} \right) \check{\Psi}'_{l,M}(z). \quad (3.59)$$

Em seguida, reescrevendo a equação (3.59) no domínio do tempo e utilizando uma descrição em espaço de estados, chega-se a:

$$\begin{cases} x'_{l,M+1}(k+1) = A_1 x'_{l,M+1}(k) + B_4 \check{\psi}'_{l,M}(k+3) + B_5 \check{\psi}'_{l,M}(k+2) \\ \quad \quad \quad + B_6 \check{\psi}'_{l,M}(k+1) + B_7 \check{\psi}'_{l,M}(k) \\ \check{\psi}'_{l,M+1}(k) = C_1 x'_{l,M+1}(k) \end{cases} \quad (3.60)$$

onde as matrizes A_1 , B_4 , B_5 , B_6 , B_7 e C_1 são definidas como (da Rosa, 2009; da Rosa et al., 2009):

$$A_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -\frac{\gamma' |\beta_{l,M+1}|^2}{\lambda'} & \frac{\gamma' (\beta_{l,M+1} + \bar{\beta}_{l,M+1})}{\lambda'} - |\beta_{l,M+1}|^2 & \beta_{l,M+1} + \bar{\beta}_{l,M+1} - \frac{\gamma'}{\lambda'} \end{bmatrix}$$

$$B_4 = Q_M \begin{bmatrix} 0 \\ 0 \\ \frac{\rho' |\beta_{l,M}|^2}{\lambda'} \end{bmatrix} \quad B_5 = Q_{l,M} \begin{bmatrix} 0 \\ 0 \\ \frac{\mu' |\beta_{l,M}|^2 - \rho' (\beta_{l,M} + \bar{\beta}_{l,M})}{\lambda'} \end{bmatrix} \quad B_6 = Q_M \begin{bmatrix} 0 \\ 0 \\ \frac{\rho' - \mu' (\beta_{l,M} + \bar{\beta}_{l,M})}{\lambda'} \end{bmatrix}$$

$$B_7 = Q_M \begin{bmatrix} 0 \\ 0 \\ \frac{\mu'}{\lambda'} \end{bmatrix} \quad C_1 = Q_M \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad Q_M = \sqrt{\frac{1 - |\beta_{l,M+1}|^2}{1 - |\beta_{l,M}|^2}}.$$

Partindo de (3.60), os elementos do gradiente $\nabla_{\mathbf{p}_l} \check{\psi}'_{l,M+1}(k)$ são como segue. Com relação a $p_{l,k}$ para $k = 1, \dots, M-1$, tem-se:

$$\begin{aligned} \frac{\partial x'_{l,M+1}(k+1)}{\partial p_{l,k}} &= A_1 \frac{\partial x'_{l,M+1}(k)}{\partial p_{l,k}} + B_4 \frac{\partial \check{\psi}_{l,M}(k+3)}{\partial p_{l,k}} + B_5 \frac{\partial \check{\psi}_{l,M}(k+2)}{\partial p_{l,k}} \\ &\quad + B_6 \frac{\partial \check{\psi}_{l,M}(k+1)}{\partial p_{l,k}} + B_7 \frac{\partial \check{\psi}_{l,M}(k)}{\partial p_{l,k}} \end{aligned} \quad (3.61)$$

$$\frac{\partial \check{\psi}'_{l,M+1}(k+1)}{\partial p_{l,k}} = C_1 \cdot \frac{\partial x'_{l,M+1}(k+1)}{\partial p_{l,k}}. \quad (3.62)$$

Já a derivada com relação à parte real do polo $\beta_{l,M}$, o cálculo pode ser efetuado por:

$$\begin{aligned} \frac{\partial x'_{l,M+1}(k+1)}{\partial \text{Re}(\beta_{l,M})} &= A_1 \frac{\partial x'_{l,M+1}(k)}{\partial \text{Re}(\beta_{l,M})} + \frac{\partial B_4}{\partial \text{Re}(\beta_{l,M})} \check{\psi}_{l,M}(k+3) + B_4 \frac{\partial \check{\psi}_{l,M}(k+3)}{\partial \text{Re}(\beta_{l,M})} \\ &\quad + \frac{\partial B_5}{\partial \text{Re}(\beta_{l,M})} \check{\psi}_{l,M}(k+2) + B_5 \frac{\partial \check{\psi}_{l,M}(k+2)}{\partial \text{Re}(\beta_{l,M})} + \frac{\partial B_6}{\partial \text{Re}(\beta_{l,M})} \check{\psi}_{l,M}(k+1) \\ &\quad + B_6 \frac{\partial \check{\psi}_{l,M}(k+1)}{\partial \text{Re}(\beta_{l,M})} + \frac{\partial B_7}{\partial \text{Re}(\beta_{l,M})} \check{\psi}_{l,M}(k) + B_7 \frac{\partial \check{\psi}_{l,M}(k)}{\partial \text{Re}(\beta_{l,M})} \end{aligned} \quad (3.63)$$

$$\frac{\partial \check{\psi}'_{l,M+1}(k+1)}{\partial \text{Re}(\beta_{l,M})} = C_1 \cdot \frac{\partial x'_{l,M+1}(k+1)}{\partial \text{Re}(\beta_{l,M})}. \quad (3.64)$$

Por fim, a derivada com relação ao polo $\beta_{l,M+1}$ é obtida por meio das equações (3.65) e (3.66):

$$\begin{aligned} \frac{\partial x'_{l,M+1}(k+1)}{\partial \text{Re}(\beta_{l,M+1})} &= \frac{\partial A_1}{\partial \text{Re}(\beta_{l,M+1})} + A_1 \frac{\partial x'_{l,M+1}(k)}{\partial \text{Re}(\beta_{l,M+1})} + \frac{\partial B_4}{\partial \text{Re}(\beta_{l,M+1})} \check{\psi}_{l,M}(k+3) \\ &\quad + \frac{\partial B_5}{\partial \text{Re}(\beta_{l,M+1})} \check{\psi}_{l,M}(k+2) + \frac{\partial B_6}{\partial \text{Re}(\beta_{l,M+1})} \check{\psi}_{l,M}(k+1) \\ &\quad + \frac{\partial B_7}{\partial \text{Re}(\beta_{l,M+1})} \check{\psi}_{l,M}(k) \end{aligned} \quad (3.65)$$

$$\frac{\partial \check{\psi}'_{l,M+1}(k+1)}{\partial \text{Re}(\beta_{l,M+1})} = C_1 \cdot \frac{\partial x'_{l,M+1}(k+1)}{\partial \text{Re}(\beta_{l,M+1})}. \quad (3.66)$$

As derivadas com relação às partes imaginárias de $\beta_{l,M}$ e $\beta_{l,M+1}$ são obtidas pela substituição de $\text{Re}(\beta_{l,M})$ e $\text{Re}(\beta_{l,M+1})$ por $\text{Im}(\beta_{l,M})$ e $\text{Im}(\beta_{l,M+1})$, respectivamente, nas equações (3.63) - (3.66). As derivadas das funções complementares $\check{\Psi}''_{l,M+1}(k)$ são obtidas procedendo a respectiva substituição das variáveis ρ' e μ' por ρ'' e μ'' .

Caso GOBF com qualquer número de pares de polos complexos

Para incluir um número qualquer de polos complexos, repete-se o procedimento apresentado nas seções anteriores. Suponha uma base de funções ortonormais composta por $r - 1$ polos reais e c pares de polos complexos. A generalização do método de construção apresentado em Ninness and Gustafsson (1997) permite escrever o c -ésimo par de funções da base ortonormal como:

$$\begin{aligned}\Psi'_{l,c}(z) &= \frac{z\sqrt{1-|\beta_{l,c}|^2}(\rho'_c z + \mu'_c)}{z^2 - (\beta_{l,c} + \bar{\beta}_{l,c})z + |\beta_{l,c}|^2} \prod_{j=1}^{c-1} \left(\frac{1 - \bar{\beta}_{l,j}z}{z - \beta_{l,j}} \right) \left(\frac{1 - \beta_{l,j}z}{z - \bar{\beta}_{l,j}} \right) \prod_{j=1}^{r-1} \left(\frac{1 - p_{l,j}z}{z - p_{l,j}} \right) \\ \Psi''_{l,c}(z) &= \frac{z\sqrt{1-|\beta_{l,c}|^2}(\rho''_c z + \mu''_c)}{z^2 - (\beta_{l,c} + \bar{\beta}_{l,c})z + |\beta_{l,c}|^2} \prod_{j=1}^{c-1} \left(\frac{1 - \bar{\beta}_{l,j}z}{z - \beta_{l,j}} \right) \left(\frac{1 - \beta_{l,j}z}{z - \bar{\beta}_{l,j}} \right) \prod_{j=1}^{r-1} \left(\frac{1 - p_{l,j}z}{z - p_{l,j}} \right)\end{aligned}\quad (3.67)$$

onde ρ'_c , μ'_c , ρ''_c e μ''_c são parâmetros reais relacionados com o polo $\beta_{l,c}$ da seguinte maneira:

$$\begin{bmatrix} \rho'_c & \mu'_c \end{bmatrix} \begin{bmatrix} 1 + |\beta_{l,c}|^2 & \beta_{l,c} + \bar{\beta}_{l,c} \\ \beta_{l,c} + \bar{\beta}_{l,c} & 1 + |\beta_{l,c}|^2 \end{bmatrix} \begin{bmatrix} \rho'_c \\ \mu'_c \end{bmatrix} = |1 - \beta_{l,c}^2|^2 \quad (3.68)$$

$$\begin{bmatrix} \rho''_c \\ \mu''_c \end{bmatrix} = \frac{1}{\sqrt{1 - \xi_c^2}} \begin{bmatrix} \xi_c & 1 \\ -1 & -\xi_c \end{bmatrix} \begin{bmatrix} \rho'_c \\ \mu'_c \end{bmatrix} \quad (3.69)$$

$$\xi_c = \frac{\beta_{l,c} + \bar{\beta}_{l,c}}{1 + |\beta_{l,c}|^2} \quad (3.70)$$

3.2.3 Algoritmo

O método proposto pode ser resumido nos seguintes passos. Partindo de um ponto inicial θ^0 , faça:

1. Calcule o gradiente de $\nabla_{\theta}^T J = [\nabla_{\mathbf{p}}^T J \quad \nabla_{\alpha}^T J]$ usando as equações (3.9) e (3.10). Para o cálculo do termo $\nabla_{\mathbf{p}} \tilde{\psi}_m(k)$, em (3.11), utilize:
 - As equações (3.20), (3.21) e (3.24), (3.25), se a base utilizada for a de Kautz;
 - As equações (3.32), (3.35), (3.36) e (3.37), para o caso GOBF com polos reais;
 - As equações (3.32), (3.50) - (3.54) para o caso GOBF com um par de polos complexos;
 - As equações (3.32), (3.61) - (3.66) para o caso GOBF com dois pares de polos complexos.
2. Use $\nabla_{\theta}^T J$ 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 polos da base adotada no modelo como restrições do problema de otimização (Veja **Observação 1**).
3. Volte ao passo 1 até que o critério de parada seja atingido.

Observação 1. As restrições mencionadas no passo 2 são necessárias para impedir que o procedimento de otimização retorne polos instáveis, fora do círculo unitário. Mais especificamente, para a base de Kautz, além de garantir a estabilidade dos polos, é preciso garantir que, se o algoritmo retornar polos com parte imaginária nula (polos puramente reais), então os dois polos devem ser necessariamente iguais. Esta imposição vem da propriedade da base de Kautz que, por construção, possui sempre polos complexos conjugados.

A estabilidade dos polos de Kautz é garantida quando $|b| < 1$ e $|c| < 1$. No entanto, uma única condição para garantir a estabilidade dos polos precisa ser imposta, sendo (da Rosa, 2009):

$$-1 < c < 0. \quad (3.71)$$

Além disso, é necessário garantir que os polos de Kautz sejam complexos conjugados e, portanto, outra restrição deve ser imposta. O par de polos de Kautz $(\beta, \bar{\beta})$ pode ser reescrito em termos dos parâmetros reais b e c :

$$(\beta, \bar{\beta}) = \frac{b(1-c)}{2} \pm i\sqrt{\frac{-b^2(1-c)^2}{4} - c}. \quad (3.72)$$

Logo, os polos na equação (3.72) serão complexos conjugados se:

$$b^2(1-c)^2 + 4c \leq 0. \quad (3.73)$$

Caso a restrição (3.73) não seja atendida, os dois polos de Kautz em (3.72) serão reais e distintos o que, por hipótese, não é possível.

3.3 Exemplos Ilustrativos: Modelos não lineares

Resultados de simulação são apresentados para ilustrar o desempenho da metodologia desenvolvida. São utilizados dois modelos não lineares de Volterra, cujos *kernels* de segunda ordem apresentem dinâmicas distintas (não-simétrico) ao longo de suas direções. Inicialmente, modela-se utilizando as bases de Kautz e, em seguida, as bases de funções generalizadas. Apresenta-se um estudo comparativo entre a técnica simétrica (da Rosa et al., 2009; da Rosa, 2009), quando uma única base de funções ortonormais é utilizada para a expansão de todas as direções do *kernel*, e não simétrica, quando utiliza-se uma base independente e distinta de funções ortonormais para a expansão de cada direção do *kernel*. Ilustra-se, também, a superioridade da utilização das derivadas das saídas dos filtros de forma recursiva no que tange à redução do tempo de processamento.

3.3.1 Exemplo ilustrativo 1

Seja um modelo OBF-Volterra de segunda ordem definido por:

$$h(k_1, k_2) = 0,25 \cdot f(k_1) \cdot g(k_2) \quad (3.74)$$

onde

$$f(k_1) = \mathcal{Z}^{-1} \left[\frac{z}{(z-0,1)(z-0,7)} \right] \quad \text{e} \quad g(k_2) = \mathcal{Z}^{-1} \left[\frac{0,6z-1,46}{z^2-0,6z+0,73} \right] \quad (3.75)$$

em que \mathcal{Z}^{-1} denota a transformada Z inversa unilateral. O cálculo das sequências discretas de (3.75) resulta em:

$$\begin{aligned} f(k_1) &= 5/3 \cdot [(0,7)^{k_1} - (0,1)^{k_1}] \\ g(k_2) &= (0,3+0,8i)^{k_2} + (0,3-0,8i)^{k_2} \end{aligned}$$

e, portanto,

$$h(k_1, k_2) = 0,25 \cdot \left[\frac{5}{3} \cdot ((0,7)^{k_1} - (0,1)^{k_1}) \right] \cdot [(0,3+0,8i)^{k_2} + (0,3-0,8i)^{k_2}] \quad (3.76)$$

para $k_1, k_2 \geq 1$.

O modelo apresentado é um modelo OBF-Volterra de segunda ordem. Sob a perspectiva da saída do modelo de Volterra, é possível substituir qualquer *kernel* assimétrico por um *kernel* equivalente simétrico por meio de um procedimento de simetrização, conforme discutido na Seção 2.1.2. No entanto, *kernels* simétricos são equivalentes a suas partes não-simétricas somente em termos da saída do modelo. Isto significa que a utilização de uma única base de funções ortonormais para a expansão de todas as direções do *kernel* e bases de funções ortonormais distintas para a expansão de cada direção do *kernel* não produzirão a mesma estrutura do *kernel*, porém apresentarão a mesma saída. A vantagem de se usar uma base de funções ortonormais independente para cada direção é que oferece maior flexibilidade nas escolhas dos parâmetros e, com isto, redução no erro de truncamento quando o *kernel* possui dinâmicas dominantes diferentes ao longo de múltiplas direções, i.e., $h(k_1, k_2) \neq h(k_2, k_1)$. Esta condição apresenta-se no modelo (3.74) que tem uma dinâmica amortecida em uma das direções e sub-amortecida na outra (Figura 3.1). Neste caso, deve ser apropriado utilizar bases de funções ortonormais parametrizadas por parâmetros distintos em cada uma das direções do *kernel*¹, como sugerido pelas Figuras 3.2(a) e 3.2(b).

São utilizadas bases $\{\psi_{21,m}\}$ e $\{\psi_{22,m}\}$ para a expansão do *kernel* de segunda ordem. Como o *kernel* não é simétrico, o modelo resultante é definido como:

$$\hat{y}(k) = \sum_{i_1=1}^M \sum_{i_2=1}^M \alpha_{i_1, i_2} \check{\psi}_{21, i_1}(k) \check{\psi}_{22, i_2}(k) \quad (3.77)$$

onde $\{\check{\psi}_{21,m}(k)\}$ e $\{\check{\psi}_{22,m}(k)\}$ denotam o resultado da filtragem do sinal $u(k)$ pelas funções ortonormais $\{\psi_{21,m}\}$ e $\{\psi_{22,m}\}$, respectivamente. O sinal de entrada $u(k)$ é aleatório com distribuição gaussiana, média nula e variância unitária.

¹Embora os *kernels* sejam ilustrados nos exemplos apresentados, eles são considerados desconhecidos sob a perspectiva do procedimento de identificação onde utilizam-se apenas dados de entrada e saída dos sistemas para estudá-los.

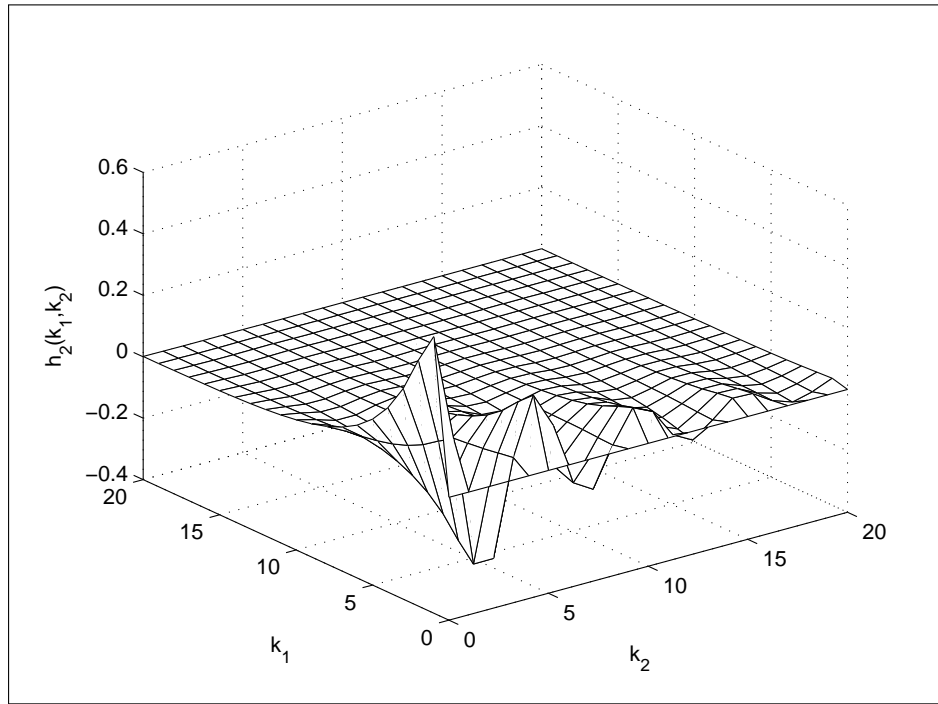
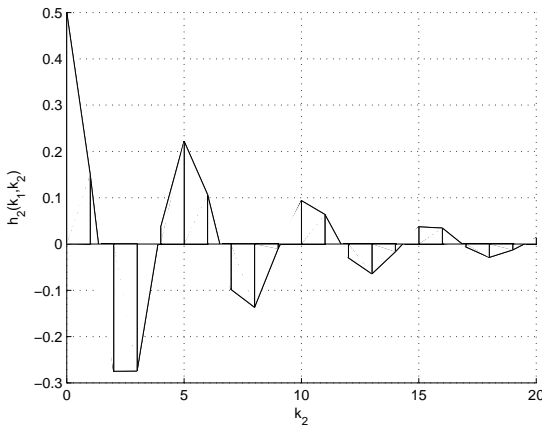
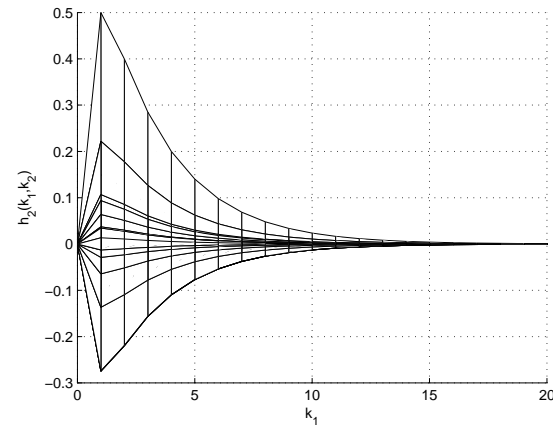


Fig. 3.1: Kernel de segunda ordem com dinâmicas dominantes diferentes em cada direção – **Exemplo 1.**



(a) Dinâmica sub-amortecida



(b) Dinâmica mais amortecida

Fig. 3.2: Planos transversais do *kernel* de Volterra de segunda ordem mostrando as dinâmicas em cada direção.

Cada base ortonormal em (3.77) é parametrizada por um vetor individual de polos, como discutido no exemplo (3.16). Logo, o gradiente de (3.77) com relação ao vetor de polos $\mathbf{p} \triangleq [\mathbf{p}_{21}^T \quad \mathbf{p}_{22}^T]^T$ é dado por:

$$\nabla_{\mathbf{p}} \hat{y}(k) = \left[\nabla_{\mathbf{p}_{21}} \hat{y}(k) \quad \nabla_{\mathbf{p}_{22}} \hat{y}(k) \right]^T \quad (3.78)$$

onde

$$\begin{aligned}\nabla_{\mathbf{p}_{21}} \hat{y}(k) &= \sum_{i_1=1}^M \sum_{i_2=1}^M \alpha_{i_1, i_2} (\nabla_{\mathbf{p}_{21}} \check{\psi}_{21, i_1}(k) \cdot \check{\psi}_{22, i_2}(k)) \\ \nabla_{\mathbf{p}_{22}} \hat{y}(k) &= \sum_{i_1=1}^M \sum_{i_2=1}^M \alpha_{i_1, i_2} (\check{\psi}_{21, i_1}(k) \cdot \nabla_{\mathbf{p}_{22}} \check{\psi}_{22, i_2}(k)).\end{aligned}$$

Modelo não linear com funções de Kautz

Nesta seção, as bases de funções $\{\psi_{21, i_1}\}$ e $\{\psi_{22, i_2}\}$ escolhidas para o modelo (3.77) são as de Kautz. Para o cálculo dos gradientes $\nabla_{\mathbf{p}}^T J$ e $\nabla_{\boldsymbol{\alpha}}^T J$ utiliza-se as equações (3.9) e (3.10), respectivamente. A metodologia proposta foi implementada utilizando polos iniciais gerados aleatoriamente.

Foram feitos diversos testes, utilizando um número diferentes de funções. Assumindo que o sistema seja desconhecido e apenas observações entrada/saída estão disponíveis, pode-se testar modelos com diferentes números de funções até que a precisão desejada seja atingida. Neste trabalho, testes foram efetuados variando o número de funções de Kautz de 2 a 8.

Para efeito de comparação, a qualidade do modelo obtido é avaliada calculando-se o erro entre a saída real do sistema e a saída do modelo. Para isto, define-se o Erro Quadrático da Saída (EQS), em dB, como:

$$\text{EQS} \triangleq 10 \log \frac{\sum_{k=1}^{N_d} [y(k) - \hat{y}(k)]^2}{\sum_{k=1}^{N_d} [y(k)]^2} \quad (3.79)$$

onde N_d é o número de dados entrada/saída disponíveis (dados de validação). Neste exemplo, utilizou-se $N_d = 200$.

Na Tabela 3.1, apresenta-se um resumo contendo os polos ótimos de Kautz para cada direção associada.

Tab. 3.1: Valores ótimos dos polos de Kautz relativos ao número de funções utilizadas – **Exemplo 1**.

Número de funções (M)	Direção de h_2	Polos ótimos
8	$l = 1$	0,5847
	$l = 2$	$0,2848 \pm 0,7484i$
6	$l = 1$	0,6301
	$l = 2$	$0,3143 \pm 0,7821i$
4	$l = 1$	0,6868
	$l = 2$	$0,2990 \pm 0,7987i$
2	$l = 1$	0,7002
	$l = 2$	$0,3000 \pm 0,8000i$

Na Tabela 3.2, apresentam-se resultados quantitativos a respeito da aproximação do modelo ao sistema original, em que cada direção é representada por um polo individual de Kautz. Esses resulta-

dos indicam que a seleção ótima dos polos permite obter uma qualidade satisfatória de aproximação utilizando um menor número de funções.

Tab. 3.2: Erro de aproximação da saída de acordo com o número de funções de Kautz utilizadas – **Exemplo 1.**

Número de funções (M)	EQS (dB)
8	−74,9023
6	−74,9581
4	−75,0588
2	−68,5840

De posse dos polos de Kautz obtidos, a saída real do sistema com *kernel* dado por (3.76) pode ser comparada com a saída do modelo obtido com relação a um outro conjunto de dados (dados de validação). Na Figura 3.3, ilustra-se a saída medida $y(k)$ do sistema, juntamente com a saída $\hat{y}(k)$ do modelo. As funções $\{\check{\psi}_{21,i_1}(k)\}$ e $\{\check{\psi}_{22,i_2}(k)\}$ do modelo são parametrizadas pelos polos ótimos de Kautz apresentados na Tabela 3.1, com os dados correspondentes a dados de validação. Foram utilizadas 4 funções em cada base, uma vez que a partir deste número não houve melhorias significativas nos resultados.

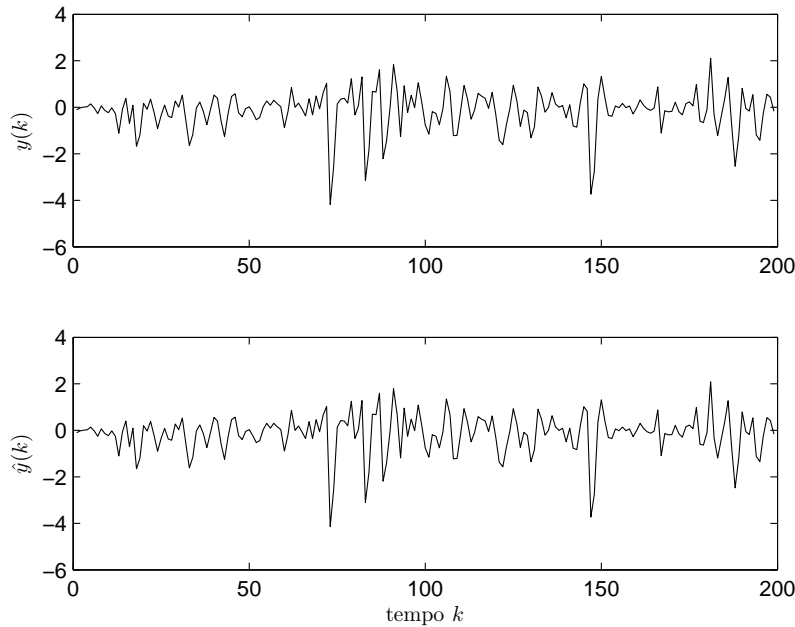


Fig. 3.3: Validação do modelo com Funções de Kautz. Acima: saída real do sistema. Abaixo: modelo correspondente com os polos ótimos obtidos: $[0,6868 \quad 0,2990 \pm 0,7987i]$.

Uma outra abordagem para verificar a capacidade de representação do modelo é observando o seu erro de predição (Figura 3.4), o qual permite avaliar quanto a saída do modelo desviou-se da saída

real do sistema em cada instante k . A análise do gráfico indica que o maior desvio ocorrido foi de, aproximadamente, $|0.08|$.

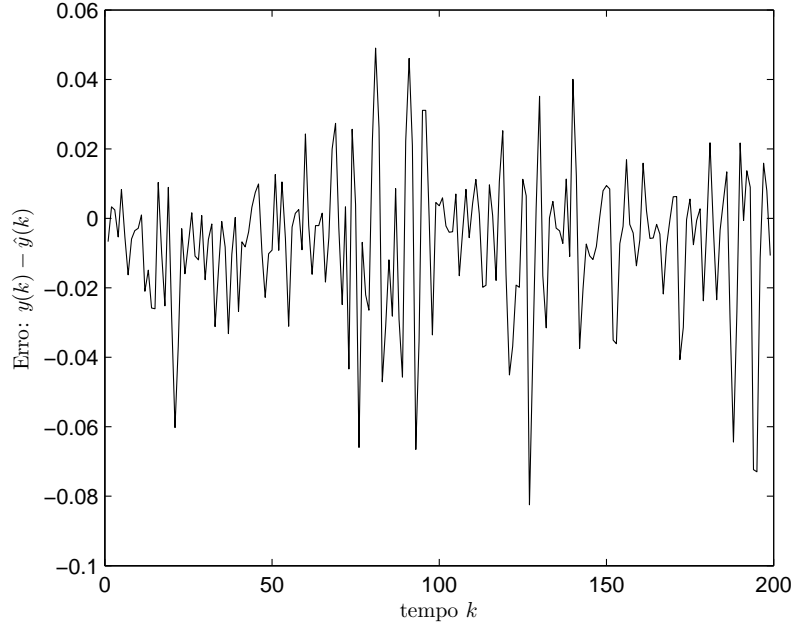


Fig. 3.4: Erro de estimação da saída do modelo com Funções de Kautz.

Modelo não linear com GOBF

O sistema de segunda ordem apresentado em (3.77) foi também modelado utilizando as bases GOBF $\{\psi_{21,i_1}\}$ e $\{\psi_{22,i_2}\}$ para a expansão do *kernel* de segunda ordem em suas duas direções. O cálculo dos gradientes $\nabla_{\mathbf{p}}^T J$ e $\nabla_{\mathbf{\alpha}}^T J$ foi efetuado por meio as equações (3.9) e (3.10). Os polos iniciais foram gerados aleatoriamente e aplicados à metodologia proposta.

Diversos testes foram efetuados com um número diferentes de funções. Assim como no caso em que a base de funções utilizada foi a de Kautz, assume-se que o sistema seja desconhecido e apenas observações entrada/saída estão disponíveis, portanto testar modelos com diferentes números de funções até que a precisão desejada seja atingida é uma boa estratégia.

Os testes utilizaram de 2 a 4 funções, variando o número e o tipo de polos que parametrizam as bases. Para o exemplo em questão, fixou-se $\gamma' = 0$ e os outros parâmetros foram calculados segundo as equações (3.47). Como discutido na Seção 3.2.2, a escolha dos parâmetros γ' , γ'' , μ' , μ'' , λ' , λ'' , ρ' e ρ'' não tem efeito significativo na modelagem do processo (Ziaei and Wang, 2006). No entanto, é fundamental que sejam atendidas as equações (3.68)-(3.70), que garantem a ortonormalidade das GOBF, e assim estes parâmetros devem ser recalculados a cada iteração do algoritmo. Tal imposição torna-se necessária pois estes parâmetros são dependentes do polo complexo em questão, o qual altera-se a cada nova iteração. Testes com valores diferentes para os parâmetros foram efetuados e os resultados obtidos foram similares.

A qualidade do modelo é avaliada pelo seu Erro Quadrático de Saída (3.79), como no caso das bases de Kautz.

O teste com duas funções na base parametrizadas por um par de polos complexos se comportou como no caso em que duas funções de Kautz foram utilizadas. Isto já era de se esperar uma vez que a parcela bem amortecida do modelo (polos reais) não seria bem descrita. Na Tabela 3.3, pode-se observar um resumo com os polos ótimos encontrados. Como o modelo é completamente descrito por um conjunto de funções que possui 2 polos reais e um par de polos complexos, os testes com mais de quatro funções na base não apresentaram melhorias significativas nos resultados, desde que contivessem 1 par de polos complexos.

Tab. 3.3: Valores ótimos dos polos relativos ao número de funções generalizadas (GOBF) utilizadas – **Exemplo 1.**

Número de funções (M)	Direção de h_2	Polos ótimos
2 (1 par complexo)	$l = 1$	0,6046
	$l = 2$	$0,2996 \pm 0,8001i$
3 (1 real e 1 par complexo)	$l = 1$	$[0,7000 \ 0,3000 \pm 0,8000i]$
	$l = 2$	$[0,1000 \ 0,3000 \pm 0,8000i]$
4 (2 reais e 1 par complexo)	$l = 1$	$[0,1000 \ 0,7000 \ 0,7000 \pm 0,0006i]$
	$l = 2$	$[0,2699 \ -0,6194 \ 0,3000 \pm 0,8000i]$

Na Tabela 3.4, apresenta-se um resumo quantitativo do Erro Quadrático de Saída, referente ao número de funções e tipos de polos (Tabela 3.3).

Tab. 3.4: Erro de aproximação da saída de acordo com o número de funções GOBF utilizadas – **Exemplo 1.**

Número de funções (M)	EQS (dB)
2	-33,4973
3	-120,7798
4	-123,9039

A saída real do sistema com *kernel* dado por (3.76) pode ser comparada com a saída do modelo obtido. Na Figura 3.5, ilustra-se a saída medida $y(k)$ do sistema, juntamente com a saída $\hat{y}(k)$ do modelo. As funções $\{\check{\psi}_{21,i_1}(k)\}$ e $\{\check{\psi}_{22,i_2}(k)\}$ do modelo são parametrizadas pelos polos ótimos da base GOBF apresentados na Tabela 3.3, utilizando 3 funções na base com 1 polo real e 1 par de polos complexos, com os dados correspondentes a dados de validação. A análise desta figura permite concluir que foi obtido um modelo cuja resposta é acurada quando comparada com a saída real do sistema.

A qualidade da representação do modelo também é evidenciada pelo erro de estimação, apresentado conjuntamente na Figura 3.5. Este erro permite analisar a cada instante k qual o desvio cometido pelo modelo em relação à saída real do sistema, cujos valores ficaram compreendidos no intervalo $|e(k)| < 2 \times 10^{-5}$, resultado este muito superior àquele encontrado para a base de Kautz.

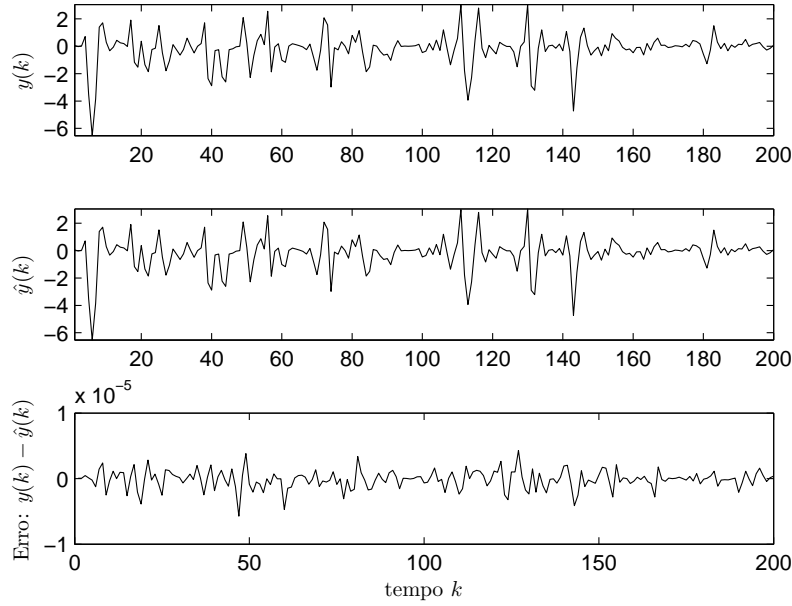


Fig. 3.5: Validação do modelo com GOBF. Acima: Saída Real do sistema. Meio: Modelo correspondente utilizando os polos ótimos obtidos. Abaixo: Erro de estimação.

A simples análise das Tabelas 3.2 e 3.4 mostra a superioridade da base GOBF, quando comparada com a base de Kautz, em descrever sistemas com múltiplas dinâmicas devido à sua principal característica de poder incorporar um número irrestrito de polos com características diferentes.

Mesmo utilizando 8 funções de Kautz, o que significa estimar 68 parâmetros, para modelar este sistema, não conseguiu-se atingir o mesmo EQS que foi atingido por apenas 3 funções GOBF em cada base, sendo que neste caso estimou-se 15 parâmetros.

3.3.2 Exemplo ilustrativo 2

Considere, agora, outro modelo não linear, cujo *kernel* apresenta dinâmicas oscilatórias diferenciadas ao longo de suas direções. Definido por:

$$h(k_1, k_2) = 0,25 \cdot f(k_1) \cdot g(k_2) \quad (3.80)$$

onde

$$f(k) = \mathcal{Z}^{-1} \left[\frac{1,3000z^2 - 2,0000z + 1,2180}{(z - 0,7000)(z^2 - 0,6000z + 0,5800)} \right] \quad (3.81)$$

e

$$g(k) = \mathcal{Z}^{-1} \left[\frac{1,1000z^2 - 1,9800z + 0,2670}{(z - 0,1000)(z^2 - z + 0,8900)} \right] \quad (3.82)$$

em que \mathcal{Z}^{-1} denota a transformada Z inversa unilateral. O cálculo das sequências discretas de (3.81) e (3.82) resulta em:

$$f(k_1) = (0,7)^{k_1} + (0,3 + 0,7i)^{k_1} + (0,3 - 0,7i)^{k_1}$$

$$g(k_2) = (0,1)^{k_2} + (0,5 + 0,8i)^{k_2} + (0,5 - 0,8i)^{k_2}$$

para $k_1, k_2 \geq 1$.

O *kernel* do modelo é ilustrado na Figura 3.6. A análise das Figuras 3.7(a) e 3.7(b) indica que o *kernel* apresenta dinâmicas diferenciadas em suas direções. Como já comentado, a vantagem da utilização de uma base independente para cada direção é o aumento da flexibilidade na escolha dos parâmetros, além da redução do erro de truncamento. Entretanto, as dinâmicas deste exemplo não são completamente diferentes, como ocorria com o exemplo 1, pois ambas possuem característica oscilatória, embora uma direção a apresente em menor intensidade que a outra.

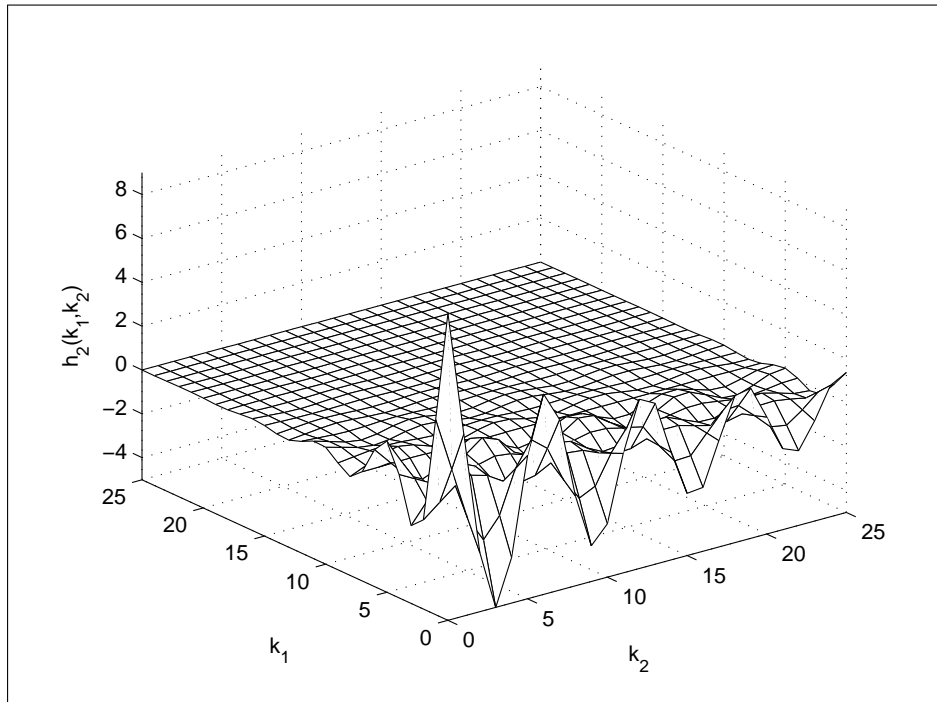


Fig. 3.6: Kernel de segunda ordem com dinâmicas dominantes diferentes em cada direção – **Exemplo 2.**

Assim como no exemplo anterior, foram utilizadas bases distintas, ou seja, $\{\psi_{21,m}\}$ e $\{\psi_{22,m}\}$ para a expansão do *kernel* de segunda ordem. O modelo resultante é idêntico àquele definido pela

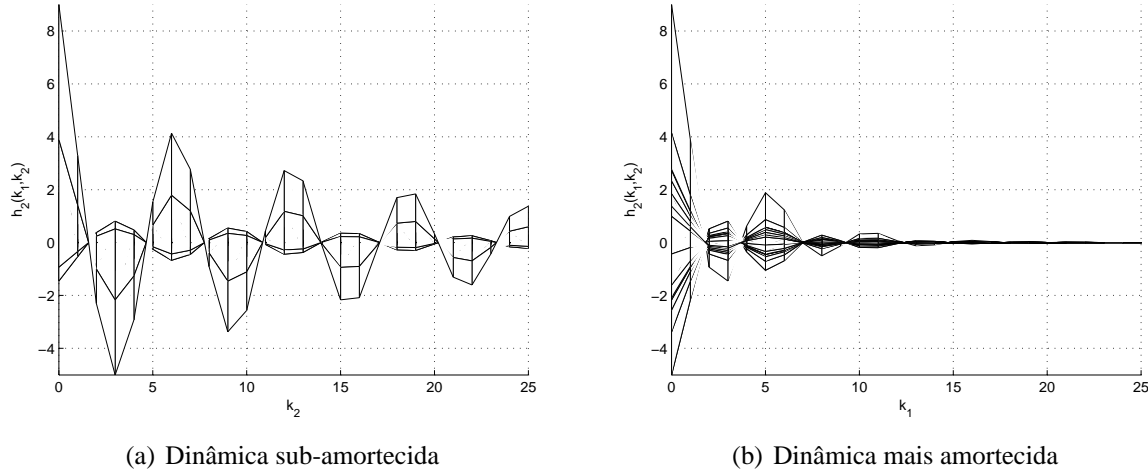


Fig. 3.7: Planos transversais do *kernel* de Volterra de segunda ordem onde as dinâmicas em cada direção são realçadas.

equação (3.77), sendo $\{\check{\psi}_{21,m}(k)\}$ e $\{\check{\psi}_{22,m}(k)\}$, o resultado da filtragem do sinal $u(k)$ pelas funções ortonormais $\{\psi_{21,m}\}$ e $\{\psi_{22,m}\}$, respectivamente. Novamente, o sinal de entrada $u(k)$ é aleatório com média nula e variância unitária.

O cálculo do gradiente do modelo é o mesmo apresentado em (3.78). Foram utilizadas as bases de Kautz e GOBF para modelar o *kernel*.

Modelo não linear com funções de Kautz

O cálculo dos gradientes $\nabla_{\mathbf{p}}^T J$ e $\nabla_{\boldsymbol{\alpha}}^T J$ foi efetuado utilizando as equações (3.9) e (3.10), respectivamente. A metodologia proposta foi implementada utilizando polos iniciais gerados aleatoriamente. O número de dados entrada/saída disponível foi $N_d = 150$.

Os testes efetuados utilizaram 2 a 8 funções de Kautz e o Erro Quadrático de Saída (3.79) como forma de avaliar a qualidade do modelo. Os valores ótimos dos polos são apresentados na Tabela 3.5.

Tab. 3.5: Valores ótimos dos polos de Kautz relativos ao número de funções utilizadas – **Exemplo 2**.

Número de funções (M)	Direção de h_2	Polos ótimos
8	$l = 1$	$0,2768 \pm 0,4092i$
	$l = 2$	$0,4743 \pm 0,7494i$
6	$l = 1$	$0,4412 \pm 0,4685i$
	$l = 2$	$0,4842 \pm 0,7808i$
4	$l = 1$	$0,3179 \pm 0,4324i$
	$l = 2$	$0,5107 \pm 0,7816i$
2	$l = 1$	$0,4068 \pm 0,7558i$
	$l = 2$	$0,4067 \pm 0,7557i$

Na Tabela 3.6, apresentam-se resultados quantitativos a respeito da aproximação do modelo ao sistema original, em que cada direção é representada por um polo individual de Kautz.

Tab. 3.6: Erro de aproximação da saída de acordo com o número de funções de Kautz utilizadas – Exemplo 2.

Número de funções (M)	EQS (dB)
8	−29, 6204
6	−25, 2360
4	−20, 5692
2	−5, 8608

Com os polos obtidos pelo procedimento de otimização, Tabela 3.5, a saída do modelo pode ser comparada com a saída real do processo, com dados correspondentes a dados de validação, conforme ilustrado pela Figura 3.8. Foram empregadas 8 funções em cada base ortonormal, uma vez que este número de funções foi o que forneceu melhor resultado.

Com a utilização das bases de Kautz, os resultados obtidos para o Exemplo 1 foram superior ao encontrado para o Exemplo 2. Este último exemplo apresenta múltiplas dinâmicas (1 polo real e 1 par de polos complexos conjugados) ao longo de suas direções, ao passo que o Exemplo 1, embora assimétrico quando comparadas as direções k_1 e k_2 , possui apenas uma dinâmica em cada uma delas (polos reais na direção k_1 e 1 par de polos complexos conjugados na direção k_2). Como a base de Kautz é parametrizada por apenas um único modo dominante, vê-se a dificuldade desta base em descrever modelos com múltiplas dinâmicas, como é o caso do Exemplo 2.

Além disso, na Figura 3.9, pode-se verificar o erro de estimação do modelo. Como já comentado, este erro permite verificar a diferença instantânea entre a saída real e a saída do modelo e, neste caso, apresentou um valor máximo em módulo de, aproximadamente, 0, 15.

Modelo não linear com GOBF

Nesta seção, utiliza-se as bases de funções generalizadas (GOBF) para a expansão do *kernel* do modelo (3.80). Os gradientes $\nabla_p^T J$ e $\nabla_\alpha^T J$ foram calculados segundo as equações (3.9) e (3.10) e polos iniciais aleatórios foram utilizados para o procedimento de otimização de acordo com a metodologia proposta.

Os testes utilizaram de 2 a 4 funções na base. Como utilizado no exemplo 1, fixou-se $\gamma' = 0$ e os outros parâmetros foram calculados conforme as equações (3.47) e a qualidade do modelo avaliada pelo Erro Quadrático de Saída (3.79).

Na Tabela 3.7, encontra-se um resumo com os polos ótimos encontrados e um resumo quantitativo do Erro Quadrático de Saída, referente ao número de funções e tipos de polos (Tabela 3.7) é encontrado na Tabela 3.8.

Na Figura 3.10, ilustra-se, de forma condensada, a saída medida $y(k)$ do sistema, juntamente com a saída $\hat{y}(k)$ do modelo, com dados correspondentes a dados de validação. As funções ortonormais foram parametrizadas pelos polos ótimos da base GOBF conforme apresentado na Tabela 3.7, onde

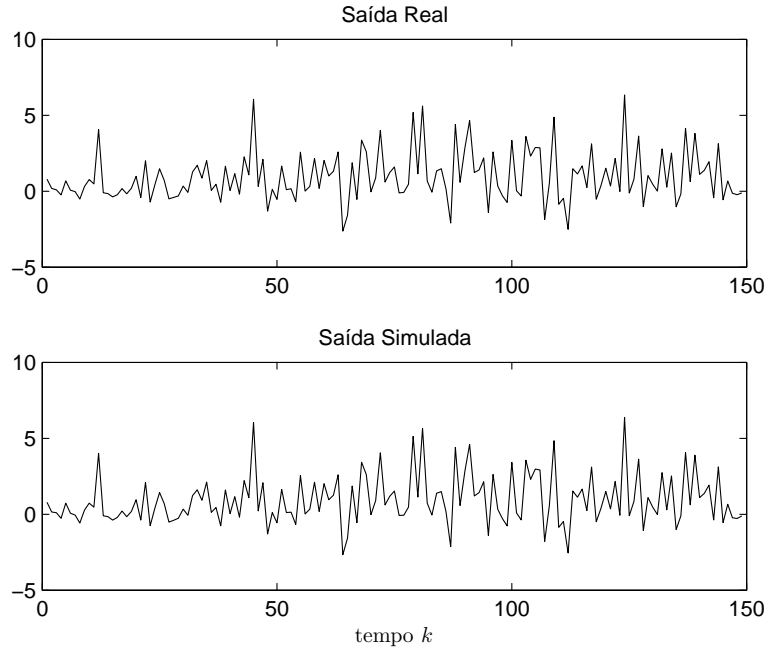


Fig. 3.8: Validação do modelo com Funções de Kautz. Acima: saída real do sistema. Abaixo: modelo correspondente com os polos ótimos: $[0, 2768 \pm 0, 4092i \ 0, 4743 \pm 0, 7494i]$.

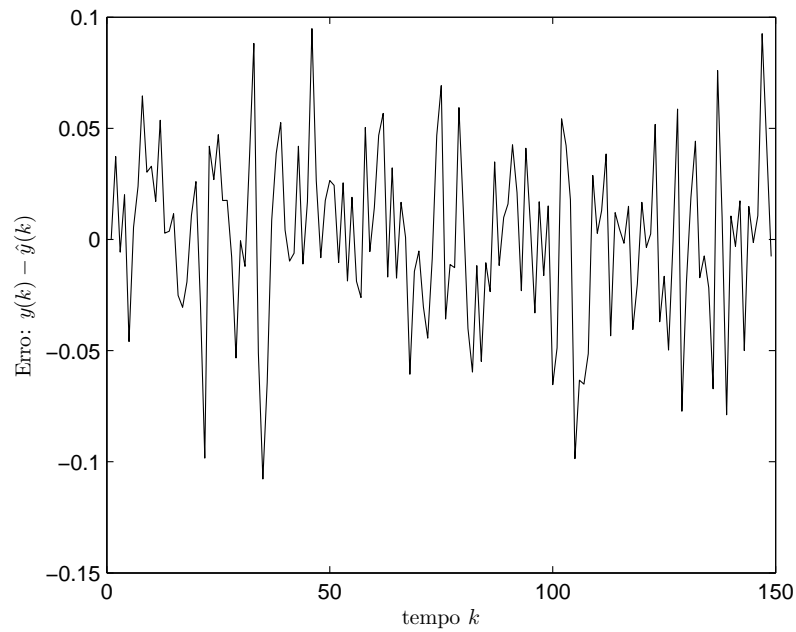


Fig. 3.9: Erro de estimação da saída do modelo com Funções de Kautz.

cada base utilizou um conjunto de 3 funções, o que significa ser parametrizada por um polo real e um

Tab. 3.7: Valores ótimos dos polos relativos ao número de funções generalizadas (GOBF) utilizadas – **Exemplo 2**.

Número de funções (M)	Direção de h_2	Polos ótimos
2 (1 par complexo)	$l = 1$	$0,1765 \pm 0,6987i$
	$l = 2$	$0,4964 \pm 0,8020i$
3 (1 real e 1 par complexo)	$l = 1$	$[0,6996 \ 0,3001 \pm 0,7004i]$
	$l = 2$	$[-0,0412 \ 0,5000 \pm 0,8000i]$
4 (2 reais e 1 par complexo)	$l = 1$	$[0,0234 \ 0,7000 \ 0,3000 \pm 0,7000i]$
	$l = 2$	$[0,1001 \ 0,7710 \ 0,5000 \pm 0,8000i]$

Tab. 3.8: Erro de aproximação da saída de acordo com o número de funções (GOBF) utilizadas – **Exemplo 2**.

Número de funções (M)	EQS (dB)
2	-12,1865
3	-55,2071
4	-95,1547

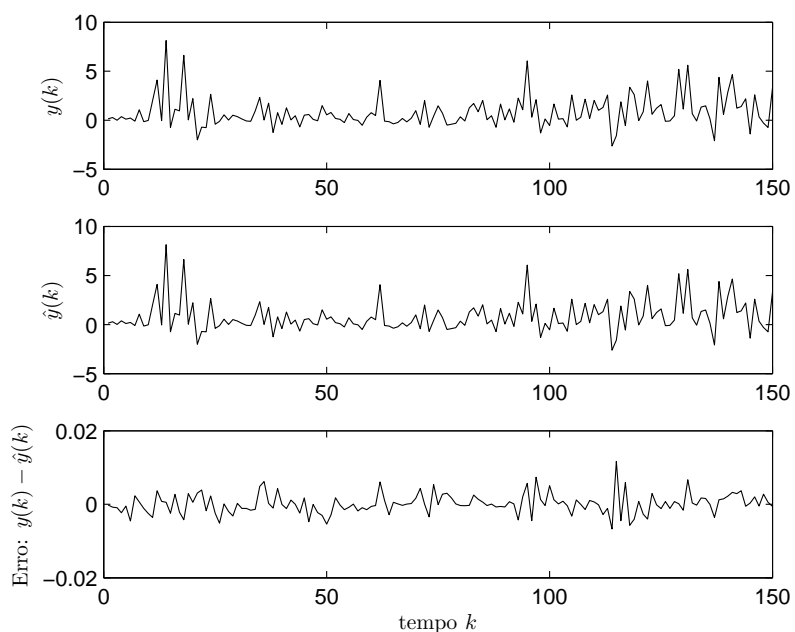


Fig. 3.10: Validação do modelo com GOBF. Acima: Saída Real do sistema. Meio: Modelo correspondente utilizando os polos ótimos obtidos. Abaixo: Erro de estimação.

par de polos complexos, e os dados utilizados correspondem a dados de validação do modelo.

A comparação direta entre os erros EQS da modelagem por bases de Kautz e por bases GOBF (comparativo entre as Tabelas 3.6 e 3.8) deixa claro novamente a superioridade das bases GOBF em descrever sistemas com múltiplas dinâmicas, pois possuem a capacidade de incorporar um conjunto irrestrito de polos com características diferentes. Esta superioridade apresenta-se não só na redução do erro de estimação, mas também no número de parâmetros a serem estimados. Apenas para efeito de comparação, utilizando os melhores resultados encontrados, foram estimados 68 parâmetros com a metodologia empregando a base de Kautz enquanto que com GOBF foram apenas 15 parâmetros.

3.4 Comparação: Base Única *versus* Bases Múltiplas

Para efeito de comparação da técnica apresentada, os mesmos sistemas foram modelados utilizando uma única base (Kautz e GOBF) para todas as direções. Ou seja, o modelo apresentado em (3.77) é descrito como um *kernel* simétrico e os coeficientes da expansão ortonormal satisfazem $\alpha_{i_1, i_2} = \alpha_{i_2, i_1}$, gerando uma matriz simétrica para os coeficientes. Portanto, aquele modelo pode ser descrito como:

$$\hat{y}(k) = \sum_{i_1=1}^M \sum_{i_2=1}^{i_1} \alpha_{i_1, i_2} \check{\psi}_{2, i_1}(k) \check{\psi}_{2, i_2}(k). \quad (3.83)$$

Assim como no caso onde uma base distinta foi utilizada para cada direção, a estratégia utilizada foi a mesma com a devida alteração nos cálculos dos gradientes. No cálculo de $\nabla_{\mathbf{p}}^T J$, o termo $\nabla_{\mathbf{p}} \hat{y}_\eta$ que aparece na equação (3.9) não pode ser calculado utilizando a equação (3.11). Isto ocorre porque essa equação foi desenvolvida para o caso em que bases distintas são utilizadas. Quando uma mesma base é utilizada, o cálculo do gradiente é dado por (da Rosa, 2009):

$$\nabla_{\mathbf{p}} \hat{y}_\eta(k) = \sum_{l=1}^{\eta} \left[\sum_{i_1=1}^M \cdots \sum_{i_{\eta}=1}^{i_{\eta-1}} \alpha_{i_1, \dots, i_{\eta}} \nabla_{\mathbf{p}_l} \check{\psi}_{l, i_l}(k) \prod_{\substack{j=1 \\ j \neq l}}^{\eta} \check{\psi}_{j, i_j}(k) \right]. \quad (3.84)$$

Logo, o gradiente do modelo em (3.83), utilizando as equações (3.9) e (3.84) é:

$$\nabla_{\mathbf{p}} \hat{y}(k) = \sum_{i_1=1}^M \sum_{i_2=1}^{i_1} \alpha_{i_1, i_2} \left[\nabla_{\mathbf{p}} \check{\psi}_{2, i_1}(k) \check{\psi}_{2, i_2}(k) + \check{\psi}_{2, i_1}(k) \nabla_{\mathbf{p}} \check{\psi}_{2, i_2}(k) \right]. \quad (3.85)$$

Para efeito de comparação, os testes seguiram a abordagem do caso não-simétrico e foram efetuados variando de 2 a 8 funções de Kautz e de 2 a 4 funções generalizadas (GOBF), tanto para o exemplo ilustrativo 1 quanto para o exemplo ilustrativo 2.

3.4.1 Análise com funções de Kautz

Nesta seção, será desenvolvido um estudo comparativo onde as bases utilizadas foram as de Kautz. Num primeiro momento, o exemplo ilustrativo 1 será o ponto focal, seguido de uma análise do exemplo ilustrativo 2. O objetivo é mostrar o desempenho da abordagem assimétrica ante a metodologia simétrica quando o *kernel* é assimétrico.

Exemplo ilustrativo 1

Com a utilização das funções de Kautz, o melhor resultado ocorreu com $M = 8$ funções de Kautz, cujo polo ótimo foi $0,3611 \pm 0,5327i$. Um resumo dos resultados são apresentados nas Tabelas 3.9 e 3.10.

Tab. 3.9: Valores ótimos dos polos de Kautz relativos a M funções utilizadas – **Exemplo 1.**

Número de funções (M)	Polos ótimos
8	$[0,3611 \pm 0,5327i]$
6	$[0,3712 \pm 0,5116i]$
4	$[0,2482 \pm 0,4243i]$
2	$[0,3324 \pm 0,4707i]$

Tab. 3.10: Erro de aproximação da saída de acordo com o número de funções de Kautz utilizadas – **Exemplo 1.**

Número de funções (M)	EQS (dB)
8	-20,6285
6	-13,5508
4	-8,5573
2	-4,4835

De posse dos polos de Kautz obtidos, a saída real do sistema com *kernel* dado por (3.74) pode ser comparada com a saída do modelo obtido. Na Figura 3.11, ilustra-se a saída medida $y(k)$ do sistema, juntamente com a saída $\hat{y}(k)$ do modelo (dados de validação) e o erro de estimação do modelo. As funções $\{\psi_{2,i_1}(k)\}$ e $\{\psi_{2,i_2}(k)\}$ do modelo são parametrizadas pelos polos ótimos de Kautz apresentados na Tabela (3.9) utilizando 8 funções na base, com os dados correspondentes a dados de validação.

Uma análise direta das Tabelas 3.2 e 3.10 mostra a superioridade dos resultados quando uma base distinta é utilizada para cada direção em detrimento de apenas uma. O erro EQS dos melhores resultados variou de -75,0588 dB, no caso assimétrico, para -20,6285 dB, no caso simétrico. Isto significou uma melhoria de 368,86% no erro de estimação do modelo.

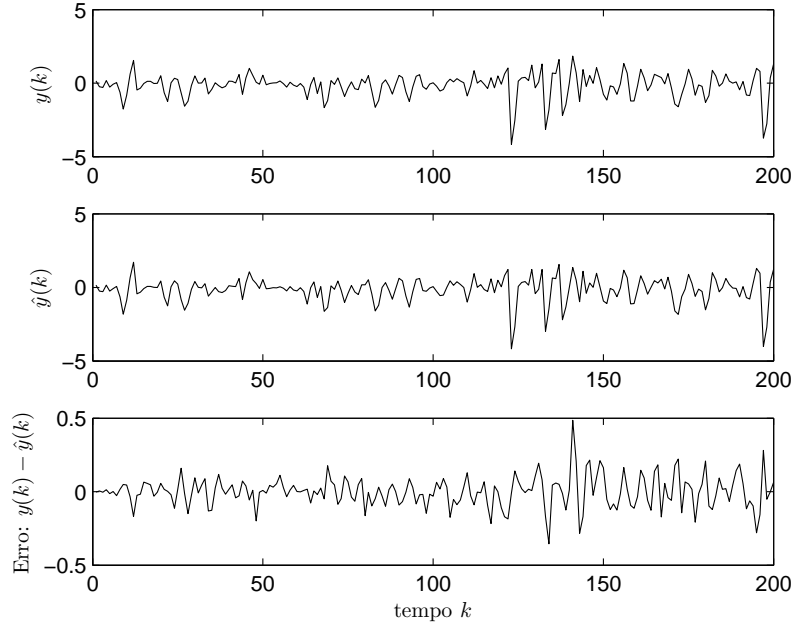


Fig. 3.11: Validação do modelo com funções de Kautz. Acima: saída real do sistema. Meio: modelo correspondente com o polo ótimo obtido: $[0, 3611 \pm 0, 5327i]$. Abaixo: Erro de estimação.

Outro fato marcante é o número de funções utilizadas. Utilizaram-se apenas 4 funções na base, no caso assimétrico, enquanto que no caso simétrico foram necessárias 8 funções. Ou seja, mesmo com o dobro de funções na base, o erro de estimação ainda foi pelo menos 3 vezes maior que o anterior.

Sabe-se que para modelos assimétricos, o número de parâmetros (NP) a ser estimado é igual a 2η , onde η é a ordem do modelo, mais M^2 coeficientes α_{i_1, i_2} , sendo M o número de funções utilizadas, i.e., $NP = 2\eta + M^2$. Para o caso simétrico, o número de parâmetros é dado por $NP = 2 + M(M + 1)/2$, o que aparentemente significa uma vantagem para a modelagem simétrica do *kernel*. No entanto, mesmo quando o número de parâmetros é levado em consideração, ainda assim a abordagem assimétrica mostrou-se vantajosa quando o *kernel* apresenta dinâmicas diferenciadas em suas direções. Mais precisamente, no caso assimétrico foram estimados $2 \cdot 2 + 4^2 = 20$ parâmetros, enquanto que no caso simétrico foram $2 + (8 \cdot 9)/2 = 38$ parâmetros. Uma redução de 66,67% no número de parâmetros a serem estimados, para os melhores resultados apresentados na Tabela 3.11.

Exemplo ilustrativo 2

Neste exemplo, o melhor resultado ocorreu com a utilização de 8 funções na base ortonormal, cujo polo ótimo foi $0,4529 \pm 0,6570i$. Resumem-se os resultados nas Tabelas 3.12 e 3.13, onde apresentam-se os polos ótimos encontrados e o erro de estimação de aproximação relativos ao número M de funções utilizadas.

A saída real $y(k)$ do sistema com *kernel* dado por (3.80), a saída do modelo $\hat{y}(k)$ e o erro de estimação são apresentadas, compactamente, na Figura 3.12, onde as funções $\{\check{\psi}_{2, i_1}(k)\}$ e $\{\check{\psi}_{2, i_2}(k)\}$ do modelo foram parametrizadas pelo polo ótimo de Kautz encontrado para 8 funções na base e os

Tab. 3.11: Relação: número de parâmetros (NP) e erro de estimação (EQS) do modelo com funções de Kautz – **Exemplo 1**.

Metodologia	Número de Funções (M)	EQS (dB)	Número de Parâmetros (NP)
Simétrica	8	-20,6285	38
	6	-13,5508	23
	4	-8,5573	12
	2	-4,4835	5
Assimétrica	8	-74,9023	68
	6	-74,9581	40
	4	-75,0588	20
	2	-68,5840	8

Tab. 3.12: Valores ótimos dos polos de Kautz relativos a M funções utilizadas – **Exemplo 2**.

Número de funções (M)	Polos ótimos
8	$[0,4529 \pm 0,6570i]$
6	$[0,3801 \pm 0,6677i]$
4	$[0,4011 \pm 0,8170i]$
2	$[0,4258 \pm 0,7498i]$

Tab. 3.13: Erro de aproximação da saída de acordo com o número de funções de Kautz utilizadas – **Exemplo 2**.

Número de funções (M)	EQS (dB)
8	-16,0618
6	-11,7735
4	-6,5711
2	-5,3744

dados entrada/saída correspondentes a dados de validação.

A simples análise das Tabelas 3.6 e 3.13 ressalta a superioridade dos resultados quando uma base distinta é utilizada para cada direção do *kernel*. Perceba que o EQS variou de -29,6204 dB, no caso assimétrico, para -16,0618 dB, no caso simétrico. Isto significou uma melhoria de 184,41% no erro de estimação do modelo.

Este exemplo permite fazer uma consideração mais detalhada do número de parâmetros a serem estimados. Como comentado anteriormente, o número de parâmetros a ser estimado, no caso assimétrico, é igual a $2\eta + M^2$, sendo η a ordem do modelo e M o número de funções utilizadas e, no caso simétrico, o número de parâmetros é dado por $2 + M(M+1)/2$. Com base nessas equações, percebe-se que significou estimar $2 \cdot 2 + 8^2 = 68$ parâmetros no caso assimétrico, enquanto no caso simétrico

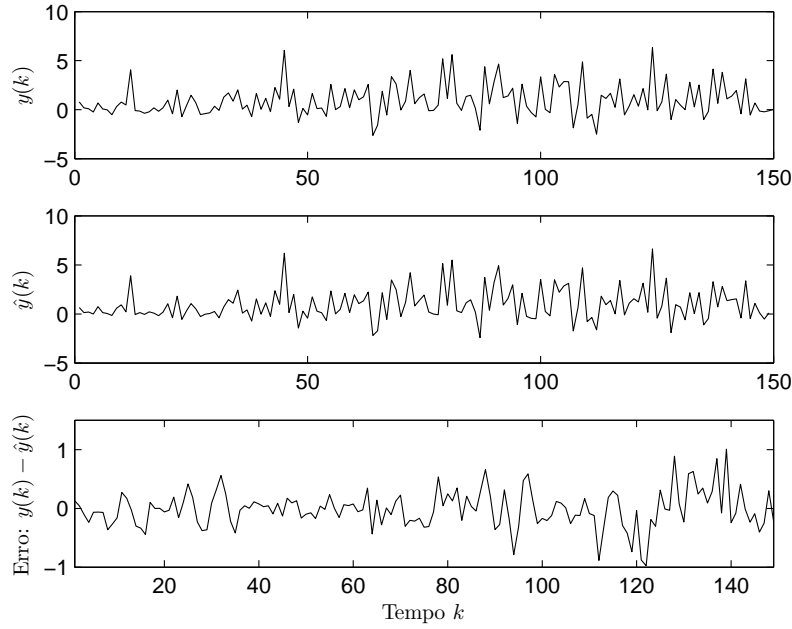


Fig. 3.12: Validação do modelo com funções de Kautz. Acima: saída real do sistema. Meio: modelo correspondente com o polo ótimo obtido: $[0, 4529 \pm 0, 6570i]$. Abaixo: Erro de estimação.

foram $2 + (8 \cdot 9)/2 = 38$ parâmetros. No entanto, a melhoria na qualidade do modelo assimétrico foi substancial quando compara-se as duas abordagens com relação à precisão do modelo. Frente a este cenário, uma avaliação adicional foi fixar a precisão do modelo, EQS, em aproximadamente -29 dB, e aumentar o número de funções na modelagem simétrica até que a mesma fosse atingida. Feito isso, seriam necessárias 18 funções para obter esta precisão. A utilização desse número de funções significa estimar 173 parâmetros (2 parâmetros de Kautz e 171 coeficientes da expansão ortonormal), enquanto na modelagem assimétrica foram 68 parâmetros, como já comentado. Isto significa uma redução de 60,69% na quantidade de parâmetros a estimar. Na Tabela 3.14, apresenta-se um resumo que permite comparar o erro de modelagem e o número de parâmetros estimados em cada abordagem (Simétrica e Assimétrica).

3.4.2 Análise com GOBF

Assim como procedido na seção anterior, agora o estudo comparativo envolve a utilização da base de funções ortonormais generalizadas associada à metodologia assimétrica e simétrica.

Exemplo ilustrativo 1

O *kernel* apresentado em (3.76) é completamente descrito por uma base que contenha 2 polos reais e 1 par de polos complexos. Assim, os testes efetuados em que as bases possuíam estas características tiveram resultados excelentes, como pode ser observado pela análise da Tabela 3.16. No entanto, até que estes parâmetros fossem atingidos o melhor resultado ocorreu com a utilização de 3

Tab. 3.14: Relação: número de parâmetros (NP) e erro de estimação (EQS) do modelo com funções de Kautz – **Exemplo 2**.

Metodologia	Número de Funções (M)	EQS (dB)	Número de Parâmetros (NP)
Simétrica	18	-29,8888	173
	8	-16,0618	38
	6	-11,7735	23
	4	-6,5711	12
	2	-5,3744	5
Assimétrica	8	-29,6204	68
	6	-25,2360	40
	4	-20,5692	20
	2	-5,8608	8

funções na base parametrizada por 1 polo real e 1 par de polos complexos, cujos polos ótimos foram $[0,7183 \ 0,3005 \pm 0,7982i]$. O resumo dos resultados são apresentados nas Tabelas 3.15 e 3.16.

Para o cálculo dos parâmetros que garantem a ortogonalidade da base, uma estratégia similar àquela do caso não-simétrico foi adotada. Ou seja, fixou-se $\gamma' = 0$ e calculou-se os outros valores utilizando as equações em (3.47). Testes com valores diferentes para os parâmetros também foram efetuados e, novamente, obtiveram-se resultados similares.

Tab. 3.15: Valores ótimos dos polos relativos ao número de funções generalizadas (GOBF) utilizadas – **Exemplo 1**.

Número de funções (M)	Polos ótimos
2 (1 par complexo)	$0,3271 \pm 0,4938i$
3 (1 real e 1 par complexo)	$[0,7183 \ 0,3005 \pm 0,7982i]$
4 (2 reais e 1 par complexo)	$[0,1000 \ 0,7000 \ 0,3000 \pm 0,8000i]$

Tab. 3.16: Erro de aproximação da saída de acordo com o número de funções (GOBF) utilizadas – **Exemplo 1**.

Número de funções (M)	EQS (dB)
2	-4,4835
3	-31,3226
4	-123,0385

Com os polos ótimos obtidos, a saída real do sistema, com *kernel* dado por (3.76), pode ser comparada com a saída do modelo obtido. Na Figura 3.13, ilustra-se a saída medida $y(k)$ do sistema, juntamente com a saída $\hat{y}(k)$ do modelo, com dados correspondentes a dados de validação, e o erro

de estimação do modelo. As funções $\{\check{\psi}_{2,i_1}(k)\}$ e $\{\check{\psi}_{2,i_2}(k)\}$ do modelo são parametrizadas pelos polos ótimos apresentados na Tabela (3.15) utilizando 3 funções generalizadas, com os dados correspondentes a dados de validação. A análise das Figuras 3.13 e 3.5 permite fazer uma consideração sobre a superioridade da modelagem assimétrica quando compara-se o erro de estimação. Note que enquanto o erro na modelagem assimétrica ficou compreendido no intervalo $e(k) \ll 2 \times 10^{-5}$, neste caso o intervalo aumentou para $e(k) < 0,2$, um aumento significativo na imprecisão do modelo.

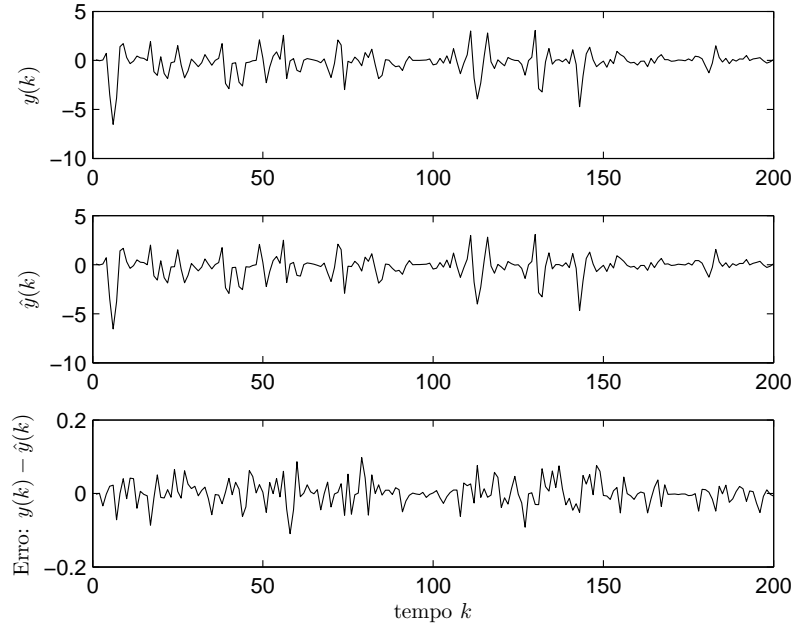


Fig. 3.13: Validação do modelo com GOBF. Acima: saída real do sistema. Abaixo: modelo correspondente com os polos ótimos obtidos: $[0, 7183 \ 0, 3005 \pm 0, 7982i]$. Abaixo: Erro de estimação.

Comparando-se as Tabelas 3.4 e 3.16, percebe-se a vantagem da utilização de bases distintas para descrever o *kernel* assimétrico quando as bases estão subparametrizadas. No entanto, isto significou aumento no número de parâmetros a ser estimados. O número de parâmetros a serem estimados em cada caso, quando a ordem do *kernel*, η , é igual a 2, como a utilizada aqui, é dado por $NP = 2M + M^2$, no caso assimétrico, e $NP = M + M(M + 1)/2$, no caso simétrico, onde M é o número de funções utilizadas. Conforme resumo apresentado na Tabela 3.17, isto significou estimar, na abordagem assimétrica, 15 parâmetros (6 relativos aos polos das funções generalizadas e 9 coeficientes), enquanto no caso simétrico foram 9 (3 relativos aos polos das funções e 6 para os coeficientes). Entretanto, mesmo com o aumento do número de parâmetros a serem estimados, a abordagem assimétrica apresenta-se como uma metodologia viável, já que implicou em uma melhoria significativa na qualidade do modelo, mais precisamente, em 385,60%.

Se, no entanto, desejar fixar a precisão e calcular o número necessário de funções para atingi-la, pode-se comparar a quantidade de parâmetros estimados em cada caso. Por exemplo, fixando-se a precisão em, aproximadamente, -120 dB, seriam necessárias 3 funções em cada base, para o caso assimétrico, o que implica em estimar 15 parâmetros, enquanto que, no caso simétrico, seriam necessárias 4 funções e, conseqüentemente, 14 parâmetros seriam estimados. O que ainda manteria a

Tab. 3.17: Relação: número de parâmetros (NP) e erro de estimação (EQS) do modelo com GOBF – Exemplo 1.

Metodologia	Número de Funções (M)	EQS (dB)	Número de Parâmetros (NP)
Assimétrica	2	−33,4973	8
	3	−120,7798	15
	4	−123,9039	24
Simétrica	2	−4,4835	5
	3	−31,3226	9
	4	−123,0385	14

modelagem simétrica vantajosa em relação ao número de parâmetros a serem estimados. O exemplo seguinte, por possuir uma dinâmica complexa, ilustrará a vantagem da utilização da metodologia assimétrica não só quando as bases estão subparametrizadas.

Exemplo ilustrativo 2

Utilizando de 2 a 4 funções na base ortonormal, o *kernel* apresentado em (3.80) foi identificado. Os resultados com os polos ótimos obtidos e o erro de aproximação do modelo são apresentados nas Tabelas 3.18 e 3.19.

Tab. 3.18: Valores ótimos dos polos relativos ao número de funções generalizadas (GOBF) utilizadas – Exemplo 2.

Número de funções (M)	Polos ótimos
2 (1 par complexo)	$0,4264 \pm 0,4938i$
3 (1 real e 1 par complexo)	$[0,1211 \ 0,4875 \pm 0,7718i]$
4 (2 reais e 1 par complexo)	$[0,8136 \ 0,0463 \ 0,4829 \pm 0,7687i]$

Tab. 3.19: Erro de aproximação da saída de acordo com o número de funções (GOBF) utilizadas – Exemplo 2.

Número de funções (M)	EQS (dB)
2	−5,3798
3	−11,5155
4	−12,3424

A diferença, neste caso, ficou ainda mais marcante. Ao comparar as Tabelas 3.8 e 3.19, percebe-se a possível vantagem da utilização de bases distintas para descrever *kernels* assimétricos.

O uso de uma base independente para cada direção aumenta a flexibilidade na escolha dos polos, como pode ser observado pelo exemplo analisado. Na abordagem simétrica, analisada nesta seção, utilizando 3 funções na base obteve-se um EQS de $-11,5155$ dB, enquanto, na metodologia assimétrica, o EQS encontrado foi de $-55,2071$ dB com o mesmo número de funções em cada base ortonormal, o que significou uma grande melhoria no resultado. Infelizmente, assim como no Exemplo 1, o número de parâmetros estimados aumenta de 9, no caso simétrico, para 15 parâmetros, na modelagem assimétrica. No entanto, isto significa uma melhoria de 479,42% na qualidade do modelo.

Por outro lado, se for fixada uma mesma precisão para ambas as abordagens, EQS de aproximadamente -55 dB, e aumentar o número de funções na modelagem simétrica até que a mesma seja atingida, são necessárias 6 funções. Isto implica na necessidade de estimar 27 parâmetros (6 parâmetros para os polos da base e 21 coeficientes da expansão ortonormal). Isto ilustra que a abordagem assimétrica pode reduzir o erro de truncamento e, fixando uma mesma precisão, também pode reduzir o número de parâmetros a serem estimados. Na Tabela 3.20, pode-se observar um resumo comparativo entre cada abordagem, seu respectivo EQS e número de parâmetros estimados.

Tab. 3.20: Relação: número de parâmetros (NP) e erro de estimação (EQS) do modelo com GOBF – Exemplo 2.

Metodologia	Número de Funções (M)	EQS (dB)	Número de Parâmetros (NP)
Assimétrica	2	$-12,1865$	8
	3	$-55,2071$	15
	4	$-95,1547$	24
Simétrica	2	$-5,3798$	5
	3	$-11,5155$	9
	4	$-12,3424$	14
	6	$-55,2474$	27

Este é, portanto, um exemplo em que o uso de um modelo assimétrico permite obter um desempenho semelhante (mesma precisão) de um modelo simétrico com menor número de parâmetros a estimar ou um melhor desempenho com um número similar de parâmetros.

3.4.3 Considerações Finais sobre a Utilização de uma Base Distinta para Cada Direção do *Kernel*

A metodologia apresentada anteriormente foi desenvolvida para o caso onde o *kernel* possui dinâmicas dominantes diferenciadas ao longo de suas múltiplas direções. Entretanto quando o *kernel* é simétrico obtém-se resultados similares tanto para a técnica assimétrica, desenvolvida neste trabalho, quanto para a abordagem simétrica, apresentada em da Rosa et al. (2009).

Para isto, analise o seguinte modelo OBF-Volterra de segunda ordem:

$$h_2(k_1, k_2) = 0,25 \exp(-\rho_1 k_1 - \rho_2 k_2) \cos(\omega_1 k_1 + \omega_2 k_2) \quad (3.86)$$

para $k_1, k_2 \geq 0$ e fazendo $\rho_1 = \rho_2 = 0,2$ e $\omega_1 = \omega_2 = 0,5$, obtém-se um *kernel* simétrico, como pode ser visualizado na Figura 3.14.

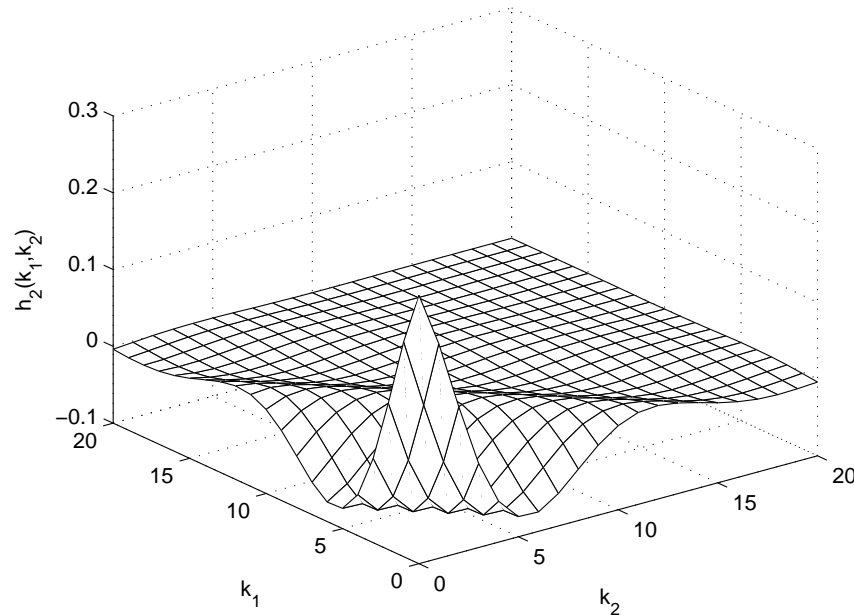


Fig. 3.14: Estrutura do *kernel* simétrico em (3.86).

Aplicando ambas as metodologias para modelar o sistema descrito em 3.86, obtêm-se os resultados apresentados nas Tabelas 3.21 e 3.22 onde foram utilizadas as funções de Kautz e GOBF, respectivamente.

Tab. 3.21: Comparativo: Modelagens Assimétrica e Simétrica para o modelo em (3.86) – Funções de Kautz.

Metodologia	Nº de Funções	Polos ótimos	EQS (dB)	Nº de Parâmetros
Assimétrica	2	$0,7185 \pm 0,3925i$	-90,3210	8
	4	$0,7125 \pm 0,3932i$	-69,9381	20
	6	$0,6426 \pm 0,4055i$	-65,9699	40
Simétrica	2	$0,7185 \pm 0,3925i$	-91,7789	5
	4	$0,7141 \pm 0,3945i$	-69,5516	12
	6	$0,6894 \pm 0,4000i$	-65,4706	23

Como pode ser observado nos resultados obtidos, os conjuntos de polos encontrados na modelagem assimétrica foram os mesmos para as duas direções do *kernel*, resultado já esperado pois trata-se de *kernel* simétrico. Além disso, a qualidade do modelo obtido foi similar em ambas as abordagens como pode ser observado pela comparação do EQS obtido em cada abordagem. Portanto, conclui-se

Tab. 3.22: Comparativo: Modelagens Assimétrica e Simétrica para o modelo em (3.86) – GOBF.

Metodologia	Nº de Funções	Polos ótimos	EQS (dB)	Nº de Parâmetros
Assimétrica	2	$0,7185 \pm 0,3925i$	-131,6605	8
	3	$0,1514 \quad 0,7185 \pm 0,3925i$	-114,9969	15
Simétrica	2	$0,7185 \pm 0,3925i$	-132,1116	5
	3	$0,2211 \quad 0,7185 \pm 0,3925$	-115,8644	9

que em modelos simétricos a utilização das abordagens simétricas ou assimétricas produzem resultados similares, com desvantagem para a modelagem assimétrica devido ao aumento do número de parâmetros a estimar, como pode ser observado nas Tabelas 3.21 e 3.22.

Em suma, os exemplos mostrados, tanto nesta seção quanto na Seção 3.3, sugerem que: (i) se deseja-se identificar um modelo de Volterra e este for simétrico, como usualmente adota-se na prática, então não há razão para utilizar múltiplas bases de funções ortonormais para a aproximação de cada *kernel*; (ii) por outro lado, se decidir adotar uma base distinta para cada direção do *kernel*, e este for um *kernel* assimétrico, então o uso da metodologia assimétrica pode produzir melhores resultados.

3.5 Comparação: Tempo de Processamento

Uma contribuição deste trabalho é o desenvolvimento dos cálculos das derivadas das saídas dos filtros ortonormais de forma recursiva (Seção 3.2). Como comentado naquela seção, equações recursivas permitem utilizar informações de instantes de tempo anteriores para efetuar os cálculos atuais. Com isso, tem-se uma significativa redução no tempo de processamento e no esforço computacional requeridos nos cálculos dos gradientes.

O objetivo desta seção é mostrar que a utilização das equações recursivas desenvolvidas neste trabalho permitiram a redução no tempo de processamento para os exemplos apresentados na Seção 3.3. Todos os cálculos efetuados seguiram a mesma linha de raciocínio delineada nas seções precedentes, com uma pequena alteração: num primeiro momento utilizaram-se as equações apresentadas na Seção 3.2, gradientes recursivos, e, posteriormente, as equações não-recursivas desenvolvidas em Machado (2011) e da Rosa et al. (2009) (Apêndice A) para as derivadas das saídas dos filtros ortonormais. Apenas os resultados para a modelagem assimétrica serão apresentados, pois para a modelagem simétrica a relação entre os tempos de processamento foram semelhantes à modelagem assimétrica.

Iniciando-se com as funções de Kautz, na Tabela 3.23, apresentam-se os tempos de processamento para a identificação dos valores ótimos dos polos e coeficientes do exemplo ilustrativo 1, cujo *kernel* é definido em (3.74), e do exemplo ilustrativo 2, com *kernel* dado por (3.80), cujos valores ótimos já foram apresentados nas Tabelas 3.1, 3.2 e 3.5, 3.6, respectivamente.

Enquanto que na Tabela 3.24, encontram-se os tempos de processamento, quando as equações utilizadas são as desenvolvidas em Machado (2011), conforme apresentadas no Apêndice A na Seção A.1.

Além disso, para mostrar que ambas as abordagens geram resultados similares, na Tabela 3.25, fornecem-se os valores ótimos dos polos encontrados pela utilização das equações presentes na Seção A.1 para cada um dos exemplos ilustrativos.

Tab. 3.23: Tempo de processamento para os Exemplos Ilustrativos quando utilizadas as equações recursivas (Funções de Kautz).

Número de Funções (M)	Tempo de Processamento (Minutos)	
	Exemplo Ilustrativo 1	Exemplo Ilustrativo 2
8	5, 1568	5, 5620
6	3, 9562	4, 7060
4	0, 5690	0, 5948
2	0, 1400	0, 1980

Tab. 3.24: Tempo de processamento para os Exemplos Ilustrativos utilizando as equações não recursivas (Funções de Kautz).

Número de funções (M)	Tempo (Minutos)	
	Exemplo Ilustrativo 1	Exemplo ilustrativo 2
8	49, 3625	48, 6316
6	31, 6462	34, 1529
4	15, 6862	14, 8957
2	0, 2019	0, 2144

Tab. 3.25: Valores ótimos dos polos de Kautz utilizando as equações não recursivas.

Número de funções (M)	Direção de h_2	Polos ótimos	
		Exemplo Ilustrativo 1	Exemplo Ilustrativo 2
8	$l = 1$	0, 5842	$0, 2768 \pm 0, 4089i$
	$l = 2$	$0, 2847 \pm 0, 7489i$	$0, 4741 \pm 0, 7494i$
6	$l = 1$	0, 6300	$0, 4412 \pm 0, 4683i$
	$l = 2$	$0, 3143 \pm 0, 7819i$	$0, 4842 \pm 0, 7811i$
4	$l = 1$	0, 6866	$0, 3181 \pm 0, 4325i$
	$l = 2$	$0, 2994 \pm 0, 7986i$	$0, 5107 \pm 0, 7814i$
2	$l = 1$	0, 7002	$0, 4068 \pm 0, 7558i$
	$l = 2$	$0, 3000 \pm 0, 8000i$	$0, 4067 \pm 0, 7557i$

Perceba que ao comparar as Tabelas 3.1 e 3.25, os valores ótimos dos polos são similares a menos de uma pequena variação em torno da terceira casa decimal. Trata-se de um comportamento natural, já que está envolvido um procedimento de otimização numérico. Resultados próximos já eram esperados, uma vez que conforme discutido no Apêndice A, as equações (3.20), (3.21), (3.24) e (3.25) possuem o mesmo resultado que as equações (A.3) e (A.5).

Quando as bases GOBF são utilizadas para modelar os exemplos ilustrativos da Seção 3.3, tem-se como resultado para o tempo de processamento os dados resumidos na Tabela 3.26. Os valores

ótimos, polos e EQS, são os que já foram apresentados nas Tabelas 3.3, 3.4, 3.7 e 3.8.

Tab. 3.26: Tempo de processamento para os Exemplos Ilustrativos quando utilizadas as equações recursivas (GOBF).

Número de funções (M)	Tempo de Processamento (Minutos)	
	Exemplo Ilustrativo 1	Exemplo ilustrativo 2
2	0, 2205	0, 2169
3	3, 6820	4, 9591
4	7, 8319	8, 7542

Se forem utilizadas as equações não recursivas (Seção A.2 do Apêndice A), encontram-se os tempos de processamento apresentados na Tabela 3.24. Novamente, por questão de conveniência, a medida de tempo escolhida para esta tabela é minutos.

Tab. 3.27: Tempo de processamento para os Exemplos Ilustrativos utilizando as equações não recursivas (GOBF).

Número de funções (M)	Tempo (Minutos)	
	Exemplo Ilustrativo 1	Exemplo ilustrativo 2
2	0, 3571	0, 3316
3	17, 9501	16, 2311
4	35, 6068	37, 4860

Assim como no caso em que a base de Kautz foi utilizada, os resultados obtidos - polos ótimos - foram semelhantes tanto para a abordagem recursiva, discutida neste capítulo, quanto para as equações apresentadas no Apêndice A, e, por questão de simplicidade, não serão aqui apresentados. Novamente, estes resultados já eram esperados, pois as equações atuais produzem os mesmos resultados que aquelas apresentadas no Apêndice A.

Analisando os resultados apresentados nas tabelas anteriores, percebe-se a superioridade da abordagem proposta aqui quando comparada com as abordagens propostas em da Rosa et al. (2009) e Machado (2011) no que tange ao tempo de processamento.

3.6 Resumo e Contribuições

Este capítulo apresentou uma extensão para a técnica de otimização numérica dos polos das bases ortonormais em modelos OBF-Volterra de qualquer ordem cujos *kernels* são assimétricos. Desta maneira, cada direção do *kernel* do modelo é decomposto em um conjunto independente de bases ortonormais, em vez de uma única base de funções. Cada base é parametrizada por conjunto independente de polos que representam a dinâmica dominante do *kernel* ao longo de uma direção particular. Foram desenvolvidas, também, as equações para o cálculo dos gradientes de forma recursiva. A

principal vantagem desta metodologia é que os gradientes calculados só dependem dos dados entrada/saída medidos a partir do sistema a ser modelado, sem a necessidade de informação *a priori* dos *kernels* do sistema. Além disso, por serem recursivas, o tempo de processamento necessário para o cálculo dos valores ótimos tiveram uma redução significativa quando comparados com as abordagens propostas em da Rosa et al. (2009) e Machado (2011).

A aplicabilidade da técnica foi ilustrada por meio de dois exemplos numéricos. A princípio, a base utilizada foi a de Kautz e, posteriormente, as funções generalizadas foram utilizadas para a identificação do processo. Os exemplos ilustrativos deixaram claro que o uso de bases independentes para representar cada direção dos *kernels* pode reduzir o erro de truncamento quando o *kernel* apresenta dinâmicas distintas em cada uma de suas múltiplas direções. No caso da base de Kautz, houve uma redução significativa no número de parâmetros a serem estimados, em especial no que diz respeito ao exemplo ilustrativo 1, pois na abordagem assimétrica foram necessários 20 parâmetros enquanto que na abordagem assimétrica 38 parâmetros tiveram de ser estimados e mesmo assim possuindo um erro 3 vezes maior do que a metodologia assimétrica, significando redução não só no erro de truncamento quanto no número de parâmetros. No exemplo 2, foi desenvolvida uma comparação onde ao fixar a precisão do modelo e aumentar o número de funções, na modelagem, até que a precisão fosse atingida, foram necessários 173 parâmetros na abordagem simétrica para descrever o mesmo *kernel* que na abordagem assimétrica foi descrito utilizando 68 parâmetros. Tal procedimento também foi aplicado nas Funções Ortonormais Generalizadas (GOBF) e um resultado semelhante foi encontrado. Embora esta base, por possuir a propriedade de incorporar uma variedade de polos, consiga descrever *kernels* assimétricos com um menor número de termos, mesmo quando a abordagem simétrica é utilizada, ainda assim a abordagem assimétrica obteve melhores resultados, redução do erro de truncamento, quando as bases estavam subparametrizadas e, quando fixada a precisão, redução no número de termos a serem estimados. O exemplo 2, em especial, demonstrou que ao aumentar a ordem do modelo simétrico para atingir a precisão definida foram necessários 27 parâmetros, 7 parâmetros a mais que na abordagem assimétrica. Por outro lado, destaca-se que caso o *kernel* seja simétrico, ambas as abordagens produzem resultados similares. Por isso, caso deseje-se identificar um modelo, cujo *kernel* seja simétrico, então a utilização de uma base independente para cada direção não apresenta melhorias nos resultados, apenas acarretando aumento no número de parâmetros a estimar. Por outro lado, caso o modelo seja assimétrico, o uso de uma base independente para cada direção do *kernel* pode produzir melhores resultados com um menor número de parâmetros ou um desempenho melhor com um número de parâmetros similar.

Outro importante fator foi a redução do tempo de processamento para a obtenção dos valores ótimos. Isto foi devido a utilização das equações dos gradientes na forma recursiva. Esta melhoria permite agora utilizar esta metodologia em aplicações em tempo real (*on-line*), objeto de estudos futuros.

Capítulo 4

Estimação Robusta de Regiões Paramétricas com Modelos OBF-Volterra

Diversos processos reais apresentam comportamento incerto, seja devido a perturbações externas associadas às informações coletadas (medidas realizadas), decorrentes, por exemplo, da precisão dos sensores utilizados, ou de variações nos parâmetros do modelo. Portanto, é importante levar em consideração essas incertezas, incorporando-as aos parâmetros do modelo quando da obtenção de um conjunto de parâmetros que descreva o comportamento do sistema real (Milanese and Belforte, 1982; Walter and Piet-Lahanier, 1990; Zhou and Doyle, 1999).

Neste capítulo, apresentam-se resultados referentes a duas abordagens para a estimação robusta dos parâmetros de modelos OBF-Volterra. Na Seção 4.1, assume-se o conhecimento de um conjunto de dados de entrada e saída do processo, cuja saída do processo acha-se corrompida por um erro aleatório aditivo. Neste caso, utilizam-se métodos de identificação robusta baseados na abordagem UBBE – *Unknown But Bounded Error Approach* (Milanese and Belforte, 1982; Walter and Piet-Lahanier, 1990), em que o erro associado às medidas é descrito por meio de limites inferior e superior, ou seja, supõe-se que a perturbação presente no processo é limitada. Trabalhos anteriores utilizaram estes métodos para a estimação de parâmetros em modelos lineares (Milanese and Belforte, 1982; Vicino and Milanese, 1989; Arruda et al., 1990; Mo and Norton, 1990). Aqui, estende-se esta ideia para tratar da estimação robusta dos parâmetros em sistemas não lineares descritos por modelos OBF-Volterra. Na segunda abordagem, Seção 4.3, assume-se o conhecimento *a priori* da incerteza do modelo associada à resposta ao impulso do processo. Trabalhos anteriores (Oliveira, 1997; Oliveira et al., 2000; Moreira, 2006; da Rosa, 2009; da Rosa et al., 2010) abordaram esta questão tanto para sistemas lineares quanto para sistemas não lineares. Porém, todos estes trabalhos impuseram restrições sobre a base de funções utilizada para a solução do problema. Assim, o objetivo na segunda parte deste capítulo é apresentar uma nova solução para o cálculo de limitantes das incertezas baseado em um conjunto de *kernels* de Volterra incertos, sem as restrições mencionadas anteriormente. Apresentam-se também dois conjuntos de exemplos ilustrativos para as abordagens, nas respectivas seções em que são estudadas.

Usualmente, a representação das incertezas do modelo é aproximada por formas geométricas no espaço dos parâmetros, como por exemplo um polítopo, um ortotopo ou um elipsoide (Walter and Piet-Lahanier, 1990). Neste trabalho, as metodologias apresentadas, tanto para a abordagem com incertezas sobre a saída quanto para a metodologia com incertezas sobre o *kernel*, representam o espaço

de incertezas paramétricas por meio de um ortotopo, i.e., a incerteza relativa ao modelo está associada a um valor mínimo e máximo para cada parâmetro. Portanto, se θ é o vetor paramétrico do modelo, composto por P componentes, então cada componente de θ é definida como $\theta_i \in [\theta_i^{min}, \theta_i^{max}]$, com $i = 1, \dots, P$.

4.1 Estimação Robusta de Processos com Incertezas sobre a Saída

A abordagem mais comumente utilizada para caracterizar as incertezas associadas às informações coletadas (sinal de saída) é assumir que o conjunto de medidas foi corrompido por um ruído aleatório aditivo com função de densidade de probabilidade (*fdp*) conhecida, por exemplo, um ruído branco gaussiano. Quando as incertezas possuem este tipo de caracterização estatística, vários modelos podem ser utilizados para estimar os parâmetros do modelo, como o estimador de máxima-verossimilhança, mínimos quadrados e o estimador de Bayes (Schweppe, 1968; Eykhoff, 1974; Ljung, 1987; Walter and Piet-Lahanier, 1990).

Apesar do grande esforço teórico dedicado à solução deste problema por meio de uma abordagem estocástica, importantes problemas práticos não podem ser apropriadamente resolvidos por envolverem alguns dos seguintes inconvenientes (Clement and Gentil, 1988; Milanese and Belforte, 1982; Walter and Piet-Lahanier, 1990):

1. o número de medidas disponíveis deve ser suficientemente grande para validar o comportamento assintótico do modelo;
2. os modelos estocásticos podem ser construídos apenas para uma escolha particular da função densidade de probabilidade (*fdp*) assumida para a incerteza, como exemplo as distribuições normal, Poisson e outras poucas;
3. incertezas de natureza determinística e que correspondem a erros estruturais não são adequadamente descritas por uma função densidade de probabilidade.

Outro inconveniente que merece ser citado ocorre em situações reais quando o sistema que gera os dados é complexo, possuindo não-linearidades, parâmetros variantes no tempo e mesmo *outliers* (dados suficientemente dissimilares ou aberrantes do restante da massa de dados). Isso implica na necessidade de se utilizar modelos simplificados onde as incertezas são caracterizadas de forma mais simples (Schweppe, 1968, 1973; Belforte et al., 1992).

Uma alternativa atrativa à descrição estocástica do erro que supera algumas das desvantagens citadas anteriormente é caracterizá-lo apenas por meio de limites inferiores e superiores. Esta abordagem, denominada UBBE (*Unknown But Bounded Error Approach*), é mais realista e menos exigente que a abordagem estatística. Foi introduzida no domínio da estimação de estados por Witsenhausen (1968) e Schweppe (1968), e desde então vários trabalhos foram publicados na área (Bertsekas and Rhodes, 1971; Lin, 1970). Já no contexto de estimação paramétrica, existem na literatura trabalhos que supõem que a perturbação atuando no processo é limitada (Milanese and Belforte, 1982; Vicino and Milanese, 1989; Arruda et al., 1990; Mo and Norton, 1990).

Diferentemente da abordagem estocástica em que os métodos de estimação fornecem um vetor de parâmetros únicos para o sistema, o método UBBE calcula uma região de pertinência (*membership set*) para os parâmetros do modelo, devendo tal região ser compatível com as medidas e os limites superiores e inferiores das perturbações. Isto tem significado prático para aplicações de controle, quando a informação disponível sobre a perturbação é fornecida como limites sobre seu valor absoluto, em aplicações nas quais os dados medidos não são suficientes para garantir resultados assintóticos confiáveis, e para diagnósticos de falhas. Neste último caso, as perturbações correspondem a flutuações dos valores dos componentes em torno de seus valores nominais (Arruda, 1992).

A abordagem UBBE já vem sendo utilizada em diversas aplicações: farmacocinética, previsão de séries temporais, interpolação numérica, química, biologia, processamento de sinais, controle robusto e adaptativo, restaurações de imagens, etc (Arruda, 1992; Clement and Gentil, 1988).

4.1.1 Formulação do Problema

Seja um sistema representado por um modelo com P parâmetros desconhecidos e que sejam realizadas N medidas da saída do sistema. O modelo para o sistema real pode ser escrito matematicamente em termos de uma relação linear ou não, entre um vetor paramétrico real $\theta^0 = [\theta_1^0 \ \theta_2^0 \ \dots \ \theta_P^0]$ no espaço P -dimensional e um vetor de medidas da saída do sistema $y = [y(1) \ y(2) \ \dots \ y(N)]$ no espaço N -dimensional. Assuma que este modelo é corrompido por um vetor de incertezas $e = [e(1) \ e(2) \ \dots \ e(N)]$ que pode representar os erros de modelagem, erros estruturais, erros de medidas, ruídos externos, etc. Portanto, tem-se:

$$y(k) = F(\theta^0) + e(k), \quad k = 1, \dots, N \quad (4.1)$$

sendo $y(k)$, a k -ésima medida realizada, $F(\theta^0)$, uma família de modelos de θ capaz de representar os dados medidos no sistema real, e $e(k)$ o valor da incerteza (erro) no instante k .

Na abordagem UBBE, o erro $e(k) \in \mathbb{R}^N$ satisfaz o seguinte critério:

$$e(k) \in [e^{\min}(k), \ e^{\max}(k)] \quad (4.2)$$

em que $e^{\min}(k)$ e $e^{\max}(k)$ são os limites (inferior e superior) para o vetor de incertezas das saídas no instante k . Em princípio, esses limites são assumidos conhecidos, ou seja, passíveis de determinação. Por exemplo, a maioria dos fabricantes de sensores fornecem regras para calcular os erros máximos e mínimos de medição em um dado intervalo de operação (Walter and Piet-Lahanier, 1990). Entretanto, caso os limites não sejam conhecidos, pode-se arbitrariamente, com os seguintes riscos (Campello et al., 2007a):

- do problema de identificação robusta dos parâmetros não ter solução, caso os limites sejam subestimados, ou;
- da solução ser excessivamente conservativa para limites superestimados (conservativos).

Supondo que Ω seja o conjunto admissível (domínio) *a priori* dos parâmetros, um método de estimação robusta sob hipótese de erro desconhecido, mas limitado, calcula o conjunto S de todos os valores admissíveis de θ que sejam consistentes com as entradas e saídas dadas por (4.1) e que

satisfaça a condição (4.2). O conjunto S é então definido como a interseção entre Ω e o conjunto $R(k)$ das soluções para θ das $2 \cdot N$ desigualdades:

$$R(k) = \{\theta \mid y(k) - e^{max}(k) \leq F(\theta) \leq y(k) - e^{min}(k)\}, \quad k = 1, \dots, N \quad (4.3)$$

$$S = \Omega \cap R(k), \quad k = 1, \dots, N. \quad (4.4)$$

Se nenhum conhecimento *a priori* estiver disponível sobre os parâmetros a serem estimados, o conjunto Ω será o próprio conjunto \mathbb{R}^P , e o conjunto S definido pela expressão (4.4) passa a ser $S = \cap R(k)$.

4.1.2 Método UBBE para Estimação de Modelos Lineares

Se o sistema em (4.1) for representado por uma relação linear, então $F(\theta^0)$ pode ser escrito como um modelo de regressão linear:

$$F(\theta^0) = \phi^T(k) \cdot \theta^0 \quad (4.5)$$

onde $\phi^T(k)$ é um vetor linha de Φ , uma matriz de regressão contendo as medidas do sistema real com dimensão $N \times P$. Portanto, a equação (4.1) torna-se:

$$y(k) = \phi^T(k) \cdot \theta^0 + e(k). \quad (4.6)$$

A substituição de (4.5) em (4.3), resulta em:

$$R(k) = \{\theta \mid y(k) - e^{max}(k) \leq \phi^T(k) \cdot \theta \leq y(k) - e^{min}(k)\}, \quad k = 1, \dots, N. \quad (4.7)$$

Todos os parâmetros θ que são consistentes com a estrutura do modelo (4.7), com os limites *a priori* dos erros (4.2) e com as medidas $y(k)$ pertencem a um conjunto que pode ser visto como a região do espaço paramétrico que é delimitada por k pares de hiperplanos $H_1(k)$ e $H_2(k)$, definidos por:

$$H_1(k) = \{\theta \mid \phi^T(k) \cdot \theta = y(k) - e^{min}(k)\} \quad (4.8)$$

$$H_2(k) = \{\theta \mid \phi^T(k) \cdot \theta = y(k) - e^{max}(k)\}. \quad (4.9)$$

Além disso, cada hiperplano $H_i(k)$, $i = 1, 2$, divide o espaço paramétrico em dois semiespaços $H_i^+(k)$ e $H_i^-(k)$, tal que:

$$H_1^+(k) = \{\theta \mid \phi^T(k) \cdot \theta \leq y(k) - e^{min}(k)\} \quad (4.10)$$

$$H_1^-(k) = \{\theta \mid \phi^T(k) \cdot \theta > y(k) - e^{min}(k)\} \quad (4.11)$$

e

$$H_2^+(k) = \{\theta \mid \phi^T(k) \cdot \theta \geq y(k) - e^{max}(k)\} \quad (4.12)$$

$$H_2^-(k) = \{\theta \mid \phi^T(k) \cdot \theta < y(k) - e^{max}(k)\}. \quad (4.13)$$

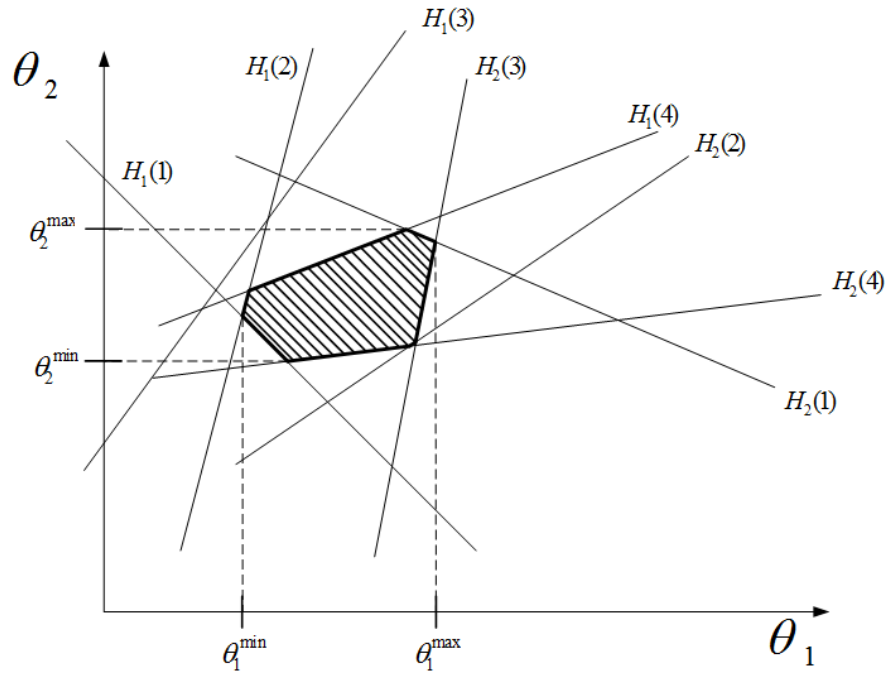


Fig. 4.1: Região de pertinência paramétrica para modelos lineares

Então, o conjunto S é dado por:

$$S = \Omega \cap \left(\bigcap_{k=1}^N R(k) \right), \quad \text{onde } R(k) = H_1^+(k) \cap H_2^+(k). \quad (4.14)$$

A expressão (4.14) representa uma sequência monótona não crescente de regiões de forma convexa, conforme ilustrado na Figura 4.1. Esta região (região hachurada na figura) é denominada Região de Pertinência Paramétrica ou Região Politópica, no caso dos modelos lineares. Se um vetor de parâmetros θ pertencente ao conjunto S então é uma estimação válida de θ^0 . Na prática, o centro da região S em algum sentido geométrico é tomado como o “melhor” estimador dos parâmetros reais. De fato, toda a região de pertinência paramétrica S pode ser vista como o conjunto de todos os estimadores verdadeiros e admissíveis para o modelo que representa o sistema real.

Outra importante definição relacionada à região convexa S é o conceito de Intervalos de Incerteza Paramétrica. Estes intervalos, conforme observado na Figura 4.1, correspondem aos valores mínimo (θ_i^{min}) e máximo (θ_i^{max}) associados a cada parâmetro dentro da região de pertinência. Estes intervalos, na abordagem UBBE, são utilizados para caracterizar o grau de confiança associado aos estimadores: quanto menor o intervalo de incerteza associado a um parâmetro, menor será o erro de estimação cometido na substituição do parâmetro real por seu valor estimado.

Com o aumento do número de medidas utilizadas no problema e, em consequência, do número de vértices, a determinação da região politópica S pode tornar-se complexa. Este é o objetivo de muitos métodos que tentam aproximar a região S por uma forma mais simples, usualmente um elipsoide ou um ortotopo (Walter and Piet-Lahanier, 1990). A necessidade desta aproximação pode ser verificada a partir do exemplo apresentado a seguir.

Considere que qualquer interseção de hiperplanos linearmente independentes ($H_i(k)$) seja um vértice de S . Então o número máximo de vértices ($n_{vértices}$) é dado por (da Silva, 1995):

$$n_{vértices} = 2 \cdot P \cdot \binom{N}{P} \quad (4.15)$$

onde:

$$\binom{N}{P} = \frac{N!}{P!(N-P)!}, \quad N = \text{número de medidas e } P = \text{número de parâmetros.}$$

No caso de modelos com 4 parâmetros e 10 observações, o número máximo de vértices será de 1680. Se o número de medidas for aumentado para 15 observações, este número vai a 10920 vértices; com 20 observações, obtém-se 38760 vértices.

Na literatura, três aproximações distintas para o cálculo da região paramétrica S tem sido investigada. A primeira delas utiliza *ortotopos* (hipercubos) para descrever a região paramétrica (Milanese and Belforte, 1982); a segunda utiliza *elipsoides* (Fogel and Huang, 1982; Belforte and Bona, 1985). Em ambos os casos, tais aproximações são feitas de forma que as regiões aproximadas possam conter (*outer-bounding*) ou estar contidas (*inner-bounding*) na região paramétrica S e são consideradas escolhas “ótimas” (Milanese and Belforte, 1982; Vicino and Milanese, 1989). Por fim, a terceira aproximação, que determina exatamente a região S , é conhecida como politópica-exata.

Neste trabalho, duas abordagens serão apresentadas: solução por meio de hiper cubo externo (Milanese and Belforte, 1982), a qual é apresentada na Seção 4.1.3, e o método recursivo de descrição exata da região politópica (Mo and Norton, 1990), descrito na Seção 4.1.4.

Essa escolha foi feita com base nos seguintes argumentos: os métodos que aproximam a região paramétrica S por elipsoides interno (*Algorithms Ellipsoidal-Inner-Bounding* - EIB) ou externo (*Algorithms Ellipsoidal-Outer-Bounding* - EOB) apresentam estrutura simples e recursiva, o que permite a sua utilização em aplicações em tempo real; possuem uma estrutura muito semelhante ao conhecido algoritmo de Mínimos Quadrados Recursivo, facilitando a análise das propriedades de convergência e estabilidade; permitam a incorporação de um fator de esquecimento. Por outro lado, apresentam as seguintes desvantagens.

Algoritmos EOB:

- O elipsoide resultante da aplicação do algoritmo não é ótimo, mesmo para N (número de medidas) $\gg P$ (número de parâmetros). A região elipsoidal é muito maior que a região politópica real, o que gera um intervalo de incerteza muito pessimista (Walter and Piet-Lahanier, 1990);
- Nem todos os estimadores contidos na região elipsoidal podem ser considerados como uma estimativa válida para os parâmetros.

Algoritmos EIB:

- Embora todos os estimadores contidos na região elipsoidal sejam considerados uma estimação válida para os parâmetros, os intervalos de incerteza paramétricas são menores se comparados com os intervalos reais da região de pertinência paramétrica;

- Apresentam também a tendência do elipsoide resultante encolher e desaparecer (Walter and Piet-Lahanier, 1990).

4.1.3 Solução por Hipercubo Externo

A seguir trata-se o caso em que a região paramétrica é descrita por meio de um hipercubo externo (HE) à região S alinhado com os eixos paramétricos (Milanese and Belforte, 1982). Esta abordagem permite fornecer para cada parâmetro θ_i do modelo um valor mínimo e máximo, além de reduzir a complexidade da estimação da região politópica exata e fornecer exatamente os intervalos de incerteza paramétrica (Walter and Piet-Lahanier, 1990). A principal vantagem desta metodologia em relação à de hipercubo interno, que calcula o ortotopo de máximo volume interno a região S , é que nesta última os intervalos de incertezas calculados são menores que os intervalos fornecidos pela região S , e alguns possíveis estimadores paramétricos não estão contidos no ortotopo calculado (Arruda, 1992).

Seja o Intervalo de Incerteza Paramétrica (*Parameter Uncertainty Interval - PUI*), denotado por PUI_i , como:

$$PUI_i = [\theta_i^{min}, \theta_i^{max}], \quad (4.16)$$

e o critério de erro apresentado na equação (4.2), como:

$$|e(k)| \leq e^M(k), \quad k = 1, \dots, N \quad (4.17)$$

onde $e^M(k)$ é o limite máximo da incerteza associada à saída no instante k .

Considere a matriz Φ composta pelos vetores linha $\phi^T(k)$ apresentados em (4.5), uma matriz de *rank* (posto) completo¹. Como Φ é de *rank* completo, existe uma submatriz Φ_P de Φ , não-singular, de dimensões $P \times P$. Sem perda de generalidade, suponha que Φ_P seja composta pelas primeiras P linhas linearmente independentes da matriz Φ . Assim, pode-se definir uma matriz Λ como a inversa de Φ_P , ou ainda:

$$\Lambda = \Phi_P^{-1}. \quad (4.18)$$

Além disso, seja $z \in \mathbb{R}^P$ um vetor definido por:

$$z(i) = y(i) \quad i = 1, \dots, P, \quad (4.19)$$

ou seja, z é um subconjunto de y com $P \leq N$.

Define-se ainda ϕ_i^T e λ_i^T como sendo a i -ésima linha de Φ e Λ , respectivamente. Os valores mínimo (θ_i^{min}) e máximo (θ_i^{max}) de θ_i em (4.16) são soluções dos seguintes problemas de programação linear:

$$\theta_i^{min} = \min_{z \in Z} \lambda_i^T z \quad (4.20)$$

$$\theta_i^{max} = \max_{z \in Z} \lambda_i^T z \quad (4.21)$$

¹Exigência necessária para que o sistema seja identificável (Eykhoff, 1974; Aguirre, 2007).

em que

$$Z = \{z \in \mathbb{R}^P : y(j) - e^M(j) \leq z(j) \leq y(j) + e^M(j), \quad j = 1, \dots, P$$

$$y(k) - e^M(k) \leq \phi_k^T \Lambda z \leq y(k) + e^M(k), \quad k = P + 1, \dots, N\}. \quad (4.22)$$

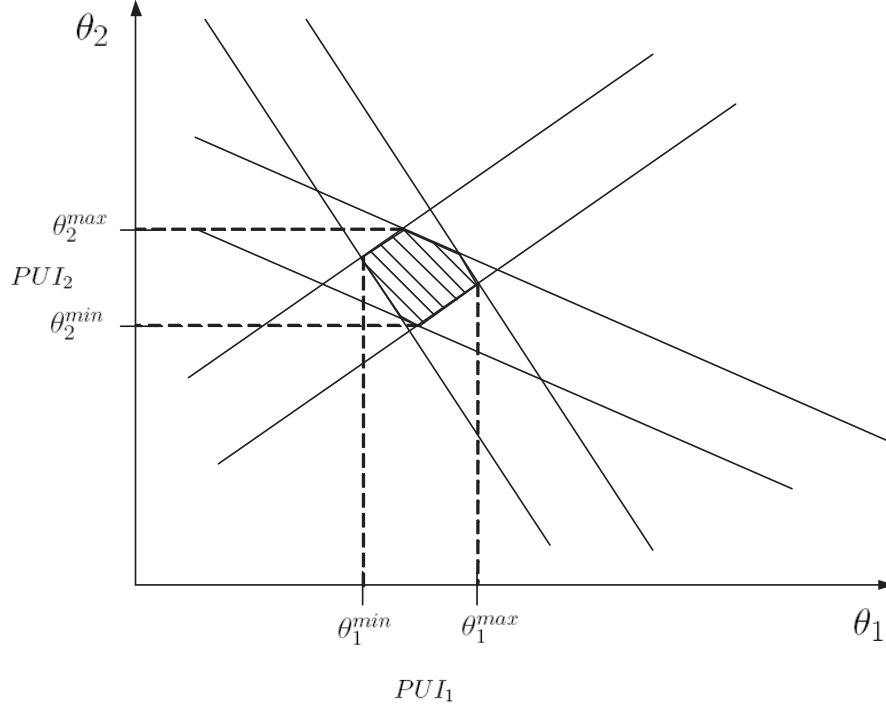


Fig. 4.2: Região de pertinência paramétrica para modelos lineares com $|e(k)| = e^M(k)$

Essa abordagem exige a solução de $2P$ problemas de programação linear com P variáveis e $2N$ restrições. Na Figura 4.2, ilustra-se o caso em que $P = 2$ (número de parâmetros), $N = 3$ (número de medidas efetuadas), e $|e(k)| \leq e^M(k)$. Como pode ser observado na figura, os hiperplanos gerados pelos pares de restrições Z são paralelos. Isto ocorre por causa da escolha dos limitantes do erro onde $-e^{min}(k) = e^{max}(k) = e^M(k)$, diferentemente do que ocorre na Figura 4.1.

É digno de nota que segundo Milanese and Tempo (1985), o melhor estimador de θ^0 é o estimador central, definido como:

$$\theta_i^{central} = \frac{(\theta_i^{max} + \theta_i^{min})}{2}, \quad i = 1, \dots, P. \quad (4.23)$$

4.1.4 Descrição Exata da Região Politópica

O método de descrição exata (DE) da região politópica S , proposto por Mo and Norton (1990), fornece uma caracterização exata da região S , é recursivo e gera limites de incerteza paramétrica

exatos. Este método representa a região por dois conjuntos de índices associados a cada vértice \mathbf{v} da região. O primeiro conjunto lista todos os vértices adjacentes ao vértice \mathbf{v} - lista $(\mathbf{v}-\mathbf{v})$ - e o segundo conjunto lista os hiperplanos (faces) que cortam o vértice \mathbf{v} , lista $(\mathbf{v}-\mathbf{f})$.

Como comentado, este método descreve a região politópica de maneira exata por meio de seus vértices e faces. No entanto neste trabalho o interesse é em encontrar os limites superior e inferior da incerteza paramétrica associados a cada parâmetro do modelo. Assim, após a obtenção do politopo final $S(k)$, este será substituído pelo hipercubo (ortotopo) que o circunscreve.

A seguir, algumas definições e propriedades dos politopos necessárias para a compreensão do algoritmo são apresentadas.

Definição 1. Um conjunto poliédrico $S \in \mathbb{R}^P$ é a interseção de uma família finita de semiespaços do \mathbb{R}^P (Mo and Norton, 1990).

Definição 2. Um politopo \mathcal{P} é a envoltória convexa de seus vértices:

$$\mathcal{P} = \text{conv}(\mathcal{V}), \quad \text{onde } \mathcal{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_x\} \quad \text{com } \mathbf{v}_i \in \mathbb{R}^P, i = 1, 2, \dots, x. \quad (4.24)$$

Definição 3. Um conjunto S é convexo se o segmento de reta que une dois pontos do conjunto também pertence ao mesmo conjunto, ou ainda (Bazaraa et al., 2006):

$$(1 - \lambda)\boldsymbol{\theta}_j + \lambda\boldsymbol{\theta}_k \in S, \quad \forall \lambda \in [0, 1] \quad \text{e} \quad \forall \boldsymbol{\theta}_j, \boldsymbol{\theta}_k \in S. \quad (4.25)$$

Uma importante análise é a de como o politopo \mathcal{P} é gerado pelos limites apresentados em (4.5). Partindo de $R(k)$, definido em (4.7), pode-se reescrevê-la como:

$$R(k) = \{\boldsymbol{\theta} \mid c^+(k) \geq \boldsymbol{\phi}^T(k) \cdot \boldsymbol{\theta} \geq c^-(k)\}, \quad (4.26)$$

sendo: $c^+(k) = y(k) - e^{\max}(k)$ e $c^-(k) = y(k) - e^{\min}(k)$.

Pela Definição 1, $S = \bigcap_{k=1}^N R(k)$ é um conjunto poliédrico. Todo conjunto poliédrico fechado é um politopo, então S é um politopo se e somente se $N \geq P$ e P elementos de $\boldsymbol{\phi}^T(k)$ são linearmente independentes. S também é convexo, uma vez que a interseção de qualquer família de conjuntos convexos é convexa (Bazaraa et al., 2006). De acordo com a Definição 3, todo hiperplano $R(k)$ é convexo.

Este método consiste em calcular, a cada iteração, o politopo $S(k)$ resultante da interseção entre o politopo $S(k-1)$, obtido com as $k-1$ medidas anteriores, e o hiperplano $R(k) = H_1^+(k) \cap H_2^+(k)$, gerados pelos semiespaços $H_1^+(k)$ e $H_2^+(k)$, com a obtenção da nova medida k .

O método descreve o politopo por meio do conjunto de seus vértices, a cada nova iteração do algoritmo as listas $\mathbf{v}-\mathbf{v}$ e $\mathbf{v}-\mathbf{f}$ devem ser atualizadas. Para isto é preciso identificar os vértices que serão rejeitados e os novos vértices formados com a introdução da nova medida.

Introdução de novas faces

Seja $H_1^+(k)$ e $H_2^+(k)$ os semiespaços gerados pela introdução da nova medida $y(k)$. Então:

$$\begin{aligned} H_1^+(k) &= \{\boldsymbol{\theta} \mid \boldsymbol{\phi}^T(k) \cdot \boldsymbol{\theta} \leq y(k) - e^{\min}(k) = c^+(k)\} \\ H_2^+(k) &= \{\boldsymbol{\theta} \mid \boldsymbol{\phi}^T(k) \cdot \boldsymbol{\theta} \geq y(k) - e^{\max}(k) = c^-(k)\} \end{aligned} \quad (4.27)$$

cujas fronteiras são denotadas por:

$$\begin{aligned} H_1(k) &= \{\boldsymbol{\theta} \mid \boldsymbol{\phi}^T(k) \cdot \boldsymbol{\theta} = c^+(k)\} \\ H_2(k) &= \{\boldsymbol{\theta} \mid \boldsymbol{\phi}^T(k) \cdot \boldsymbol{\theta} = c^-(k)\}. \end{aligned} \quad (4.28)$$

As posições de todos os vértices em \mathcal{V} com relação a H_1 são denotadas por α^+ , em que:

$$\boldsymbol{\alpha}^+ = [\alpha_1^+ \ \alpha_2^+ \ \dots \ \alpha_x^+]^T, \quad \alpha_i^+ = c^+(k) - \boldsymbol{\phi}^T(k) \mathbf{v}_i, \quad i = 1, 2, \dots, x. \quad (4.29)$$

Se α_i^+ é não-negativo, \mathbf{v}_i está em $H_1^+(k)$; se for negativo, fora de $H_1^+(k)$; se zero, em $H_1(k)$. Quando o politopo $S(k-1)$ é atualizado para $S(k)$, três situações distintas podem ocorrer:

- $S(k-1)$ não é modificado por $H_1^+(k)$ se $\boldsymbol{\alpha}^+$ é inteiramente não-negativo;
- $S(k)$ será vazio se todos os elementos de $\boldsymbol{\alpha}^+$ forem negativos, e;
- $H_1^+(k)$ intercepta e modifica $S(k-1)$ se alguns, mas não todos os elementos de $\boldsymbol{\alpha}^+$ forem negativos.

Para $H_2^+(k)$, o mesmo procedimento é realizado com as devidas alterações de sinais. Assim, por questão de praticidade, apenas $H_1^+(k)$ será considerado daqui por diante.

Atualiza-se o politopo da seguinte maneira. O hiperplano $H_1^+(k)$ intercepta $S(k-1)$ formando uma nova face de $S(k)$, denominada $H^*(k)$. Esta face $H^*(k)$ pode ser considerada a envoltória convexa de seu conjunto de vértices \mathcal{V}^* , sendo cada vértice, um vértice de $S(k)$, mas não de $S(k-1)$. Além disso, cada aresta $\mathcal{E}(j, k)$ de $S(k-1)$ é uma combinação convexa de dois vértices adjacentes \mathbf{v}_j e \mathbf{v}_k , o que permite escrever:

$$\mathcal{E}(j, k) = \{\boldsymbol{\theta} \mid \boldsymbol{\theta} = (1 - \lambda)\mathbf{v}_j + \lambda\mathbf{v}_k\}, \quad \lambda \in [0, 1], \quad \mathbf{v}_j, \mathbf{v}_k \in \mathcal{V}_{k-1} \quad (4.30)$$

os vértices adjacentes \mathbf{v}_j e \mathbf{v}_k podem ser encontrados na lista $\mathbf{v}-\mathbf{v}$. Se uma aresta é cortada por $H_1(k)$, a interseção é um vértice em \mathcal{V}^* . Pode-se analisar onde $\mathcal{E}(j, k)$ intercepta $H_1(k)$ ao examinar as posições dos vértices \mathbf{v}_j e \mathbf{v}_k com relação a $H_1(k)$. $\mathcal{E}(j, k)$ corta $H_1(k)$ se e somente se $\mathbf{v}_j \in H_1^+(k)$ e $\mathbf{v}_k \notin H_1^+(k)$, i.e., se e somente se $\alpha_j^+ \geq 0$ e $\alpha_k^+ < 0$, como ilustrado na Figura 4.3.

O novo vértice \mathbf{v}_i^* satisfaz as duas equações seguintes:

$$\boldsymbol{\phi}^T(k) \mathbf{v}_i^* = c^+(k), \quad (4.31)$$

uma vez que $\mathbf{v}_i^* \in H_1$, e

$$\mathbf{v}_i^* = (1 - \lambda)\mathbf{v}_k + \lambda\mathbf{v}_j, \quad \lambda \in [0, 1]; \quad \alpha_j \geq 0, \quad \alpha_k < 0 \quad (4.32)$$

pois $\mathbf{v}_i^* \in \mathcal{E}(j, k)$.

Daí, substituindo (4.32) em (4.31), obtém-se:

$$\lambda = \frac{\boldsymbol{\phi}^T(k) \mathbf{v}_k - c^+(k)}{\boldsymbol{\phi}^T(k) (\mathbf{v}_k - \mathbf{v}_j)}. \quad (4.33)$$

O conjunto \mathcal{V}^* de novos vértices é encontrando aplicando-se (4.33) a todos os pares de vértices de $\mathcal{V}(k-1)$ que passam no teste em (4.32). Os elementos negativos de $\boldsymbol{\alpha}^+$ mostram quais vértices de $\mathcal{V}(k-1)$ devem ser deletados.

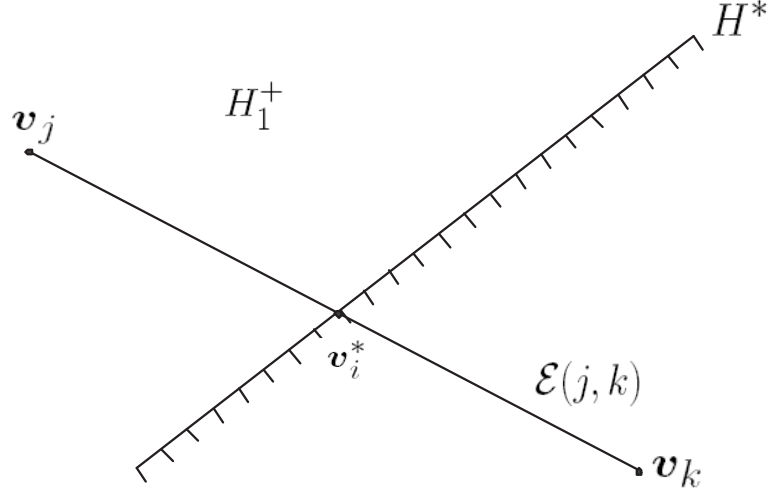


Fig. 4.3: Interpolação de um novo vértice com a introdução de uma nova observação.

Atualização da lista vértice-vértice adjacente

A adjacência $v_j \leftrightarrow v_k$ de vértices $v_j \in H_1^+$ e $v_k \notin H_1^+$ é substituída por $v_j \leftrightarrow v_i^*$ quando S é atualizado. O novo vértice v_i^* é também adjacente a outro novo vértice, que deve ser identificado. A adjacência de dois vértices é detectada ao se checar se eles têm $P - 1$ hiperplanos (faces) em comum.

Atualização da lista vértice-face

O hiperplano H_1 de H_1^+ , interceptando $S(k - 1)$, aparecerá na lista $(v - f)$ dos vértices v_i^* em \mathcal{V}^* desde que $\mathcal{V}^* \subset H_1$. Além disso, $v_i^* \in \mathcal{E}(j, k)$, a aresta formada pelos vértices adjacentes v_j e v_k de $S(k - 1)$, e $\mathcal{E}(j, k)$ é a interseção das $P - 1$ faces comuns à lista $(v - f)$ de v_j e v_k . Estas são as faces de $S(k - 1)$ que continuam na lista $(v - f)$ de \mathcal{V}^* .

Formação do politopo inicial

A determinação de $S(0)$ baseia-se no conhecimento *a priori* do sistema a ser identificado. Entretanto, se nenhum conhecimento estiver disponível, pode-se arbitrar uma região politópica $S(0)$, como um hipercubo P -dimensional, suficientemente grande tal que $S(k) \subset S(0)$ para todo k .

Resumo do Algoritmo

- (1) Formar $S(0)$ e alterar k para 1
- (2) Obter a medida $y(k)$ e formar o vetor $\phi(k)$
- (3) PARA $H_1^+(k)$ e $H_2^+(k)$ FAÇA

```

(3.1) PARA  $i = 1$  A  $x$  FAÇA {onde  $x$  = número de vértices de  $S(k-1)$ }
      SE  $H_1^+(k)$  ENTÃO  $\alpha_i^+ = c^+(k) - \phi^T(k)v_i$ 
      SENÃO  $\alpha_i^- = \phi^T(k)v_i - c^-(k)$ 

(3.2) Formar  $\alpha = [SE \ H_1^+(k) \ \text{ENTÃO} \ \alpha^+ \ \text{SENÃO} \ \alpha^-]$ 

(3.3) SE todos os elementos de  $\alpha$  são não negativos
      ENTÃO  $S(k) = S(k-1)$ ; VÁ PARA (3.4)
      SENÃO SE todos os elementos de  $\alpha$  são negativos
        ENTÃO  $S(k) = 0$ ; VÁ PARA (5)
        SENÃO (3.3.1) Encontrar  $\mathcal{V}^*$ 
              (3.3.2) Formar listas  $(v - f)$  para  $\mathcal{V}^*$ 
              (3.3.3) Formar listas  $(v - v)$  para  $\mathcal{V}^*$ 
              (3.3.4) Atualizar  $S(k-1)$ 
                    Acrescentar  $\mathcal{V}$  a  $\mathcal{V}^*$ 
                    Deletar vértices  $\notin H_+^*(k)$ 

(3.4) CONTINUE

(4) SE existem mais medidas
      ENTÃO incrementar  $k$  e VÁ PARA (2)
      SENÃO PARE

(5) PARE

```

Este algoritmo permite determinar de forma exata a região politópica S . No entanto, neste trabalho, o politopo final $S(k)$ será substituído pelo ortotopo ou hipercubo que o circunscreve. Esta aproximação permite descrever para cada parâmetro θ o limite de incerteza paramétrica PUI associado a esse parâmetro. Este procedimento, embora prático, torna, em alguns casos, a estimação mais conservadora, já que pode acrescentar possíveis estimadores que não pertenciam à região politópica original. No entanto, assim como os métodos que aproximam a região politópica por uma região ortotópica (hipercubo externo) conduzem ao cálculo de intervalos de incerteza idênticos ao da região S (Walter and Piet-Lahanier, 1990; Arruda, 1992), a substituição do politopo S pelo hipercubo que o circunscreve segue o mesmo princípio e, portanto, também produz intervalos de incerteza paramétrica idênticos ao da região politópica.

A principal vantagem do método utilizado é o fato de ser recursivo, o que facilita sua implementação em tempo real (*online*), além da economia em tempo de processamento e baixa sensibilidade a erros de arredondamento (Mo and Norton, 1990). Como desvantagem, pode-se citar o aumento da complexidade de descrição dos vértices com a ampliação do conjunto de medidas e o fato de que os limitantes não são atualizados recursivamente.

4.1.5 Estimação Robusta via Modelos de Volterra

Os sistemas não lineares são normalmente modelados por meio de representações matemáticas não lineares nos parâmetros. Tal modelagem, embora seja muito eficiente na representação desses

sistemas, não é diretamente aplicável na formulação UBBE descrita na Seção (4.1.2). Isto ocorre porque a região de pertinência de tais modelos não possui forma politópica como no caso linear (Figura 4.1). Essa constatação deve-se ao seguinte fato, a região produzida pela equação (4.3) não é mais limitada por hiperplanos e sim por hipersuperfícies, o que pode gerar uma região de pertinência não convexa, ou composta por sub-regiões, como ilustrado na Figura 4.4.

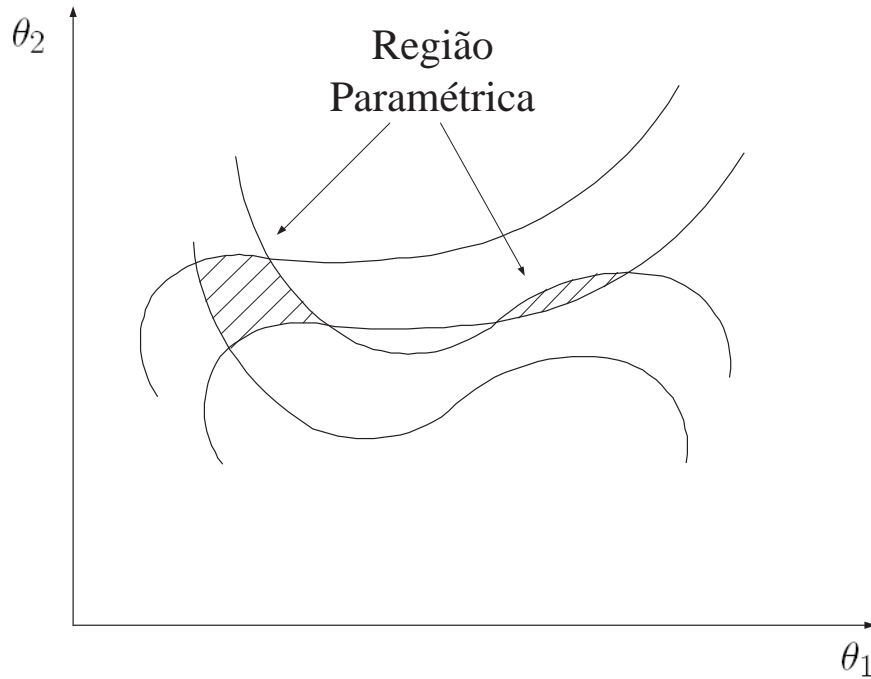


Fig. 4.4: Região de pertinência paramétrica para modelos não lineares.

Assim, os modelos OBF-Volterra apresentam-se como uma representação matemática viável, pois além de permitirem a representação de sistemas não lineares, possuem a característica de serem lineares nos parâmetros da expansão ortonormal, conforme discutido no Capítulo 2. Portanto, podem ser descritos na forma da equação (4.5) como um modelo de regressão linear.

Como exemplo, considere o modelo OBF-Volterra (Eq. 2.3). Por razões práticas, os modelos de segunda ordem ($\eta = 2$) são aqueles usualmente adotados tanto em problemas de natureza acadêmica como em problemas reais (Billings, 1980; Dumont and Fu, 1993). Ainda por razões práticas, são limitados às quantidades M_1 e M_2 de funções ortonormais para os *kernels* de primeira e segunda ordem, respectivamente:

$$\hat{y}(k) = \sum_{i_1=1}^{M_1} \alpha_{i_1} \check{\psi}_{1,i_1}(k) + \sum_{i_1=1}^{M_2} \sum_{i_2=1}^{M_2} \alpha_{i_1,i_2} \check{\psi}_{21,i_1}(k) \check{\psi}_{22,i_2}(k) \quad (4.34)$$

em que $\check{\psi}_{l,i_l}(k) = \sum_{\tau=0}^k \psi_{l,i_l}(\tau) u(k-\tau)$, como já definido em (3.1).

Considerando os coeficientes $\alpha_{(\cdot)}$ como parâmetros a serem estimados, o modelo (4.34) pode ser reescrito na forma (4.5) definindo-se os vetores $\phi^T(k)$ e θ como:

$$\begin{aligned} \phi^T(k) = & [\check{\psi}_{1,1}(k) \check{\psi}_{1,2}(k) \cdots \check{\psi}_{1,M_1}(k) \check{\psi}_{21,1}(k)\check{\psi}_{22,1}(k) \check{\psi}_{21,1}(k)\check{\psi}_{22,2}(k) \check{\psi}_{21,1}(k)\check{\psi}_{22,3}(k) \\ & \cdots \check{\psi}_{21,M_2}(k)\check{\psi}_{22,1}(k) \check{\psi}_{21,M_2}(k)\check{\psi}_{22,2}(k) \cdots \check{\psi}_{21,M_2}(k)\check{\psi}_{22,M_2}(k)] \end{aligned} \quad (4.35)$$

$$\theta^T = [\alpha_1 \alpha_2 \cdots \alpha_{M_1} \alpha_{1,1} \alpha_{1,2} \alpha_{1,3} \cdots \alpha_{M_2,1} \alpha_{M_2,2} \cdots \alpha_{M_2,M_2}]. \quad (4.36)$$

Portanto, como os modelos OBF-Volterra podem ser reescritos na forma (4.5), a técnica UBBE apresentada neste capítulo pode ser diretamente aplicada. É necessário observar que os modelos OBF-Volterra são dependentes da parametrização das bases ortonormais. Isto significa que é necessária uma etapa de identificação dos polos ótimos a serem utilizados na parametrização da base.

Neste trabalho, a aplicação da abordagem UBBE nos modelos OBF-Volterra pode ser resumida nos seguintes passos:

- As medidas efetuadas (dados coletados com ruído) são tratadas como exatas para o procedimento de identificação dos polos ótimos que parametrizam as bases ortonormais, segundo a metodologia proposta no Capítulo 3;
- Estimação dos limites de incerteza por meio da abordagem UBBE de hipercubos externos (Seção 4.1.3) ou descrição exata da região politópica (Seção 4.1.4).

4.2 Exemplos Ilustrativos - Modelos com Incerteza na Saída

Nesta seção será apresentado um conjunto de exemplos para ilustrar a utilização da técnica UBBE nos modelos OBF-Volterra. O primeiro deles é um modelo de Volterra composto por *kernels* de primeira e segunda ordem, sendo este último simétrico. Já o segundo exemplo é o modelo apresentado na Seção 3.3.1, cujo *kernel* possui dinâmicas diferenciadas ao longo de suas direções. Em ambos os exemplos, as abordagens apresentadas nas Seções 4.1.3 e 4.1.4 serão utilizadas.

4.2.1 Modelo de Volterra simétrico de segunda ordem

Considere um modelo OBF-Volterra de segunda ordem (Kibangou et al., 2005) sujeito a um ruído aditivo na saída, cujos *kernels* de primeira e segunda ordem, respectivamente, são dados por:

$$h_1(k_1) = \mathcal{Z}^{-1} \left[\frac{z + 0,5}{(z - 0,3)(z - 0,2)} \right] \quad (4.37)$$

$$h_2(k_1, k_2) = 0,25g(k_1)g(k_2) \quad \text{onde} \quad g(k) = \mathcal{Z}^{-1} \left[\frac{z + 1}{(z - 0,8)(z + 0,8)} \right] \quad (4.38)$$

sendo \mathcal{Z}^{-1} é a transformada \mathcal{Z} inversa unilateral. O cálculo das sequências discretas (4.37) e (4.38) resulta em:

$$h_1(k_1) = \frac{80}{3} \cdot (0, 3)^{k_1} - 35 \cdot (0, 2)^{k_1}$$

$$h_2(k_1, k_2) = 0,25 \cdot \left[\frac{45}{32} \cdot (0, 8)^{k_1} + \frac{5}{32} \cdot (-0, 8)^{k_1} \right] \cdot \left[\frac{45}{32} \cdot (0, 8)^{k_2} + \frac{5}{32} \cdot (-0, 8)^{k_2} \right]$$

para $k_1, k_2 \geq 1$. As bases GOBF com polos reais - equação (2.24) com $\beta_{l,n} = \bar{\beta}_{l,j} = p_{l,j}$ - são adotadas para a expansão dos kernels de primeira e segunda ordem. Além disso, assume-se que o *kernel* de segunda ordem do modelo é simétrico, o que pode ser feito sem perda de generalidade. Isto significa que $h_2(k_1, k_2) = h_2(k_2, k_1)$ e, portanto, os coeficientes da expansão satisfazem $\alpha_{i_1, i_2} = \alpha_{i_2, i_1}$. O modelo resultante é o seguinte:

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

$$y(k) = \hat{x}(k) + e(k) \quad (4.39)$$

em que as funções $\{\check{\psi}_{2,i_1}(k)\}$ e $\{\check{\psi}_{2,i_2}(k)\}$ denotam o resultado da filtragem do sinal $u(k)$ pelas funções ortonormais $\{\psi_{2,i_1}(k)\}$ e $\{\psi_{2,i_2}(k)\}$, respectivamente, $u(k)$ é um sinal aleatório uniformemente distribuído no intervalo $[0 ; 1]$ e $e(k)$ um ruído uniformemente distribuído no intervalo $[-0,5 ; 0,5]$, ou ainda, $|e(k)| \leq 0,5$, adicionado à saída.

Como já mencionado, o primeiro passo para a utilização da abordagem UBBE é a identificação dos polos que irão parametrizar as bases ortonormais. Para esta etapa, as medidas efetuadas foram tratadas como exatas e a metodologia utilizada é aquela apresentada no Capítulo 3. Testou-se diferentes números de funções na base ortonormal. Como o modelo apresentado em (4.39) é completamente descrito por duas funções ortonormais generalizadas para cada *kernel*, primeira e segunda ordem, foram feitos testes utilizando 2 e 4 funções na base. Os resultados (polos encontrados) para a etapa de identificação são apresentados na Tabela 4.1.

Tab. 4.1: Valores ótimos dos polos encontrados na etapa de identificação - **Exemplo 1**.

Nº de funções (M)	Ordem do <i>kernel</i> (η)	Polos ótimos
2	1	0,6253 0,8006
	2	-0,6484 0,8137
4	1	0,7090 0,2143 0,8721 0,2143
	2	-0,4076 0,6146 -0,4077 0,8503

Após a obtenção dos polos que parametrizam as bases, realiza-se o procedimento de identificação dos intervalos de incerteza paramétrica.

É necessário agora definir o valor dos limitantes do erro $e(k)$, de acordo com o critério de erro (4.17). Se o valor dos limitantes de erro não for conhecido ou passível de determinação, realizam-se diversos procedimentos de identificação para valores decrescentes do limitante do erro e^M . Na

abordagem por hipercubo externo, inicia-se com um valor suficientemente grande para e^M até um valor para o qual os problemas de programação linear, (4.20) e (4.21), tornem-se infactíveis (Clement and Gentil, 1988). No método de descrição exata da região politópica, o valor do limitante do erro e^M é encontrado por meio de um procedimento similar. Entretanto, o critério de parada é atingido quando o politopo resultante é vazio.

Os testes foram divididos em duas partes. Inicialmente, foi estudado o caso em que utilizaram-se duas funções na base. Tanto para a metodologia por hipercubo externo quanto para o método de descrição exata, o valor do limitante de erro encontrado, utilizando o procedimento apresentado acima, foi $e^M = 0,57$, tendo como resultado os mesmos intervalos de incerteza paramétrica. Isto já era esperado, uma vez que a região politópica final foi substituída pelo hipercubo que a circunscreve. Os resultados são apresentados na Tabela 4.2, compreendendo os intervalos de incerteza paramétrica (PUI) para cada um dos coeficientes do modelo (4.39), tendo os valores “reais” dos parâmetros contidos nos intervalos, e a largura dos intervalos encontrados.

Tab. 4.2: Limites de Incerteza Paramétrica dos parâmetros α do modelo $y(k)$ com 100 medidas e 2 funções na base - **Exemplo 1**.

Nº de funções	Ordem do <i>kernel</i> (η)	Parâmetro α	PUI	Largura PUI
2	1	1,6475	[1,2909 ; 1,9060]	0,6151
		-0,2617	[-0,3066 ; -0,1513]	0,1554
	2	-0,0121	[-0,3426 ; 0,4803]	0,8229
		0,0700	[-0,1224 ; 0,0913]	0,2137
		0,7916	[0,6975 ; 0,8891]	0,1916

Ao se utilizar 4 funções na base, a primeira implicação prática é o aumento no número de parâmetros do modelo a serem estimados. O valor encontrado para o limite máximo da incerteza associada à saída foi $e^M = 0,54$. Os resultados obtidos nesta etapa podem ser observados na Tabela 4.3.

Uma observação importante é que o aumento do número N de medidas tende a reduzir os limites de incertezas. Isto ocorre porque quanto menor for o conjunto de medidas, mais sensível ao ruído será o algoritmo utilizado. Por outro lado, quanto maior o conjunto de medidas utilizado, melhor é a capacidade do algoritmo de superar os efeitos do ruído (Xue et al., 2002), reduzindo os limites de incerteza paramétricas, que se aproximam de zero.

Na Tabela 4.4, apresentam-se os intervalos de incerteza paramétrica quando o número de dados utilizados é igual a 50 e o critério de erro e^M é mantido o mesmo em ambos os casos, ou seja, igual a 0,57, para 2 funções na base, e 0,54, quando 4 funções são utilizadas para modelar o sistema, juntamente com a largura dos intervalos. Isto permite confrontar a largura dos intervalos com aqueles apresentados nas Tabelas 4.2 e 4.3. Percebe-se claramente a redução dos intervalos de incerteza com o aumento do número de medidas utilizadas na etapa de estimação.

4.2.2 Modelo de Volterra assimétrico de segunda ordem

O modelo apresentado a seguir é aquele estudado no Capítulo 3 na Seção 3.3.1. Lá foi apresentado o *kernel* de segunda ordem assimétrico, que no domínio do tempo é descrito por:

Tab. 4.3: Limites de Incerteza Paramétrica dos parâmetros α do modelo $y(k)$ com 100 medidas e 4 funções na base - **Exemplo 1**.

Nº de funções	Ordem do <i>kernel</i> (η)	Parâmetro α	<i>PUI</i>	Largura <i>PUI</i>
4	1	1,4698	[0,6330 ; 2,5922]	1,9592
		0,0975	[-0,6938 ; 1,1534]	1,8472
		-0,3720	[-1,4407 ; 0,1483]	1,5890
		0,3125	[-0,2624 ; 1,1407]	1,4031
	2	0,1264	[-0,8991 ; 1,0955]	1,9945
		0,5874	[-0,5734 ; 1,6134]	2,1867
		0,5676	[-0,3885 ; 1,4700]	1,8585
		0,0105	[-0,7536 ; 0,9944]	1,7480
		0,0493	[-0,8913 ; 0,7039]	1,5952
		-0,0225	[-0,8165 ; 0,7179]	1,5343
		0,1539	[-0,3992 ; 0,7179]	1,2592
		0,4382	[-0,1904 ; 1,0121]	1,2025
		0,0498	[-0,6704 ; 0,7258]	1,3962
		0,0758	[-0,1135 ; 0,4419]	0,5553

$$h(k_1, k_2) = 0,25 \cdot \left[\frac{5}{3} \cdot \left((0,7)^{k_1} - (0,1)^{k_1} \right) \right] \cdot \left[(0,3 + 0,8i)^{k_2} + (0,3 - 0,8i)^{k_2} \right] \quad (4.40)$$

para $k_1, k_2 > 1$.

Neste caso, assume-se a assimetria do *kernel* e a nulidade da componente linear do modelo. Tais hipóteses permitem descrever o modelo como:

$$\begin{aligned} \hat{x}(k) &= \sum_{i_1=1}^M \sum_{i_2=1}^M \alpha_{i_1, i_2} \check{\psi}_{21, i_1}(k) \check{\psi}_{22, i_2}(k) \\ y(k) &= \hat{x}(k) + e(k) \end{aligned} \quad (4.41)$$

em que as funções $\{\check{\psi}_{n,i}(k)\}$ representam o resultado da filtragem de um sinal de distribuição normal, média nula e variância unitária, $u(k)$ pelas funções $\{\psi_{n,i}(k)\}$. Além disso, $e(k)$ é considerado um ruído aleatório uniformemente distribuído no intervalo $[-0,5 ; 0,5]$. Para efeito de avaliação e comparação, tanto a base de Kautz quanto as funções generalizadas (GOBF) serão utilizadas.

O procedimento permanece o mesmo: etapa de identificação dos polos ótimos que parametrizarão as bases, de acordo com a metodologia proposta no Capítulo 3, considerando as medidas efetuadas como exatas, seguido da etapa de estimação dos limites de incerteza, utilizando a abordagem UBBE. Foi utilizado um conjunto de 100 medidas tanto para a etapa de identificação quanto para a estimação dos intervalos.

Tab. 4.4: Limites de Incerteza Paramétrica *PUI* para os parâmetros α do modelo simétrico $y(k)$ com 50 dados - **Exemplo 1**.

Nº de funções (M)	Ordem do <i>kernel</i> (η)	<i>PUI</i>	Largura <i>PUI</i>
2	1	[1, 0129 ; 2, 1273]	1, 1144
		[-0, 5154 ; 0, 0072]	0, 5226
	2	[-1, 5010 ; 1, 1121]	2, 6131
		[-0, 2319 ; 0, 6714]	0, 9033
		[0, 5791 ; 1, 0789]	0, 4998
4	1	[-0, 1121 ; 3, 2293]	3, 3414
		[-1, 8639 ; 1, 6953]	3, 5592
		[-1, 9669 ; 0, 5763]	2, 5432
		[-0, 7222 ; 1, 5763]	2, 2985
	2	[-2, 0548 ; 2, 2675]	4, 3222
		[-1, 5398 ; 2, 0643]	3, 6042
		[-0, 7845 ; 1, 7941]	2, 5786
		[-2, 0230 ; 1, 4961]	3, 5191
		[-1, 0342 ; 1, 6793]	2, 7135
		[-2, 1428 ; 1, 0452]	3, 1880
		[-0, 8487 ; 1, 5798]	2, 4285
		[-0, 6370 ; 1, 3647]	2, 0017
		[-0, 9795 ; 1, 0277]	2, 0073
		[-0, 1410 ; 0, 5386]	0, 6796

Nas Tabelas 4.5 e 4.6, apresentam-se os polos ótimos encontrados para as bases de Kautz e GOBF, respectivamente. Novamente, utilizaram-se 2 e 4 funções em cada base. No caso das funções ortogonais generalizadas, variou-se também os tipos de polos utilizados, um par de polos complexos conjugados quando 2 funções foram utilizadas e um par de polos complexos conjugados e 2 polos reais quando utilizaram-se 4 funções.

Tab. 4.5: Valores ótimos dos polos encontrados na etapa de identificação – Base de Kautz - **Exemplo 2**.

Nº de funções (M)	Direção l do <i>kernel</i>	Polos ótimos
2	1	0, 6880
	2	$0, 2996 \pm 0, 8020i$
4	1	$0, 1862 \pm 0, 3741i$
	2	$0, 3239 \pm 0, 4889i$

O valor de e^M fixado, conforme procedimento descrito anteriormente, foi de 0, 57, tanto para 2 quanto 4 funções na base. A redução desse valor implicou na infactibilidade dos problemas de programação linear ou na nulidade do politopo. Na Tabela 4.7, apresentam-se os intervalos de incerteza

Tab. 4.6: Valores ótimos dos polos encontrados na etapa de identificação – GOBF - **Exemplo 2**.

Nº de funções (M)	Direção l do $kernel$	Polos ótimos
2	1	$0,2999 \pm 0,8021i$
	2	$0,5853$
4	1	$-0,6582 \quad -0,6582 \quad 0,3106 \pm 0,8007i$
	2	$-0.4296 \quad 0,7270 \quad 0,8665 \pm 0,3993i$

paramétrica e a largura dos respectivos intervalos quando a base de Kautz é utilizada para modelar o sistema $y(k)$.

Tab. 4.7: Limites de Incertezas Paramétricas PUI para os parâmetros α do modelo $y(k)$ com 100 dados e $e^M = 0,57$ – Bases de Kautz - **Exemplo 2**.

Nº de funções (M)	Parâmetro α	PUI	Largura PUI
2	0,2539	[0,1830 ; 0,2904]	0,1075
	-0,5798	[-0,5732 ; -0,5087]	0,0645
	0,2149	[0,1582 ; 0,2771]	0,1189
	-0,4668	[-0,5468 ; -0,4420]	0,1048
4	0,2571	[0,1022 ; 0,3676]	0,2654
	-0,5722	[-0,6756 ; -0,3821]	0,2935
	-0,0013	[-0,1477 ; 0,1236]	0,2713
	-0,0063	[-0,1519 ; 0,1900]	0,3419
	0,2106	[0,0947 ; 0,3180]	0,2233
	-0,4659	[-0,5722 ; -0,3913]	0,1809
	0,0015	[-0,1639 ; 0,1134]	0,2773
	-0,0049	[-0,1550 ; 0,1635]	0,3184
	-0,0120	[-0,1354 ; 0,1506]	0,2859
	0,0347	[-0,0687 ; 0,1328]	0,2016
	-0,0031	[-0,1435 ; 0,1128]	0,2562
	-0,0002	[-0,1404 ; 0,1271]	0,2675
	-0,0020	[-0,1921 ; 0,1292]	0,3213
	-0,0007	[-0,1566 ; 0,1010]	0,2576
	-0,0008	[-0,1418 ; 0,0668]	0,2086
	0,0005	[-0,1157 ; 0,1186]	0,2343

A modelagem com funções ortonormais generalizadas apresentaram como limitante máximo do erro o mesmo valor encontrado para as funções de Kautz, ou seja $e_M = 0,57$. O resumo dos resultados são apresentados na Tabela 4.8, a qual contém os valores dos parâmetros reais do modelo, os valores mínimos e máximos e a largura dos intervalos de incerteza paramétrica.

Deve-se observar que ambas as abordagens utilizadas, hipercubo externo e descrição exata, produziram os mesmos intervalos de incerteza paramétrica, já que nesse último a região politópica foi

Tab. 4.8: Limites de Incertezas Paramétricas *PUI* para os parâmetros α do modelo $y(k)$ com 100 dados e $e^M = 0,57$ – GOBF - **Exemplo 2**.

Nº de funções (M)	Parâmetro α	<i>PUI</i>	Largura <i>PUI</i>
2	0,0543	$[-0,0038 ; 0,1098]$	0,1136
	0,0144	$[-0,0310 ; 0,0551]$	0,0861
	0,7786	$[0,7091 ; 0,8142]$	0,1050
	0,2148	$[0,1483 ; 0,2647]$	0,1163
4	0,1110	$[0,0285 ; 0,2054]$	0,1768
	0,2320	$[0,0501 ; 0,3731]$	0,3229
	-0,0038	$[-0,3165 ; 0,3535]$	0,6700
	0,0000	$[-0,2092 ; 0,2108]$	0,4199
	0,0779	$[-0,1019 ; 0,2170]$	0,3189
	0,1515	$[0,0364 ; 0,2858]$	0,2494
	-0,0037	$[-0,3330 ; 0,3249]$	0,6579
	-0,0004	$[-0,1926 ; 0,1799]$	0,3725
	0,1388	$[-0,0930 ; 0,3941]$	0,4871
	0,2722	$[0,0662 ; 0,4412]$	0,3750
	0,0049	$[-0,4026 ; 0,5123]$	0,9149
	0,0036	$[-0,2349 ; 0,2815]$	0,5165
	0,2550	$[0,0951 ; 0,3722]$	0,2771
	0,5205	$[0,4109 ; 0,6049]$	0,1939
	-0,0038	$[-0,2490 ; 0,1967]$	0,4458
	0,0020	$[-0,1472 ; 0,1190]$	0,2662

substituída pelo hipercubo que a circunscreve. Este resultado já era esperado, uma vez que, como já discutido, este procedimento conduz ao cálculo de intervalos de incerteza paramétrica idênticos aos valores reais da região S , assim como o método do hipercubo externo o faz. Esta constatação permite que ambas as abordagens sejam aplicadas para a estimação dos limites de incerteza dos modelos OBF-Volterra. No entanto, o último método possui algumas características que tornam a sua aplicação mais atrativa, pois é recursivo, o que significa dizer que não é necessário calcular todo o politopo novamente com a inclusão de uma nova medida, apenas acrescentar esta nova restrição de forma a alterar a região quando necessário. Esta recursão permite uma possível aplicação em tempo real.

4.3 Estimação Robusta de Processos com Incertezas sobre o *Kernel*

Nas seções anteriores, discutiu-se a estimação robusta dos parâmetros do modelo OBF-Volterra com base na abordagem UBBE (*Unknown But Bounded Error*) (Milanese and Belforte, 1982; Walter and Piet-Lahanier, 1990), em que assume-se que as incertezas associadas ao sinal de saída estão

limitadas ao intervalo $[e^{\min}, e^{\max}]$. Entretanto, quando se dispõe de informações sobre as incertezas nos parâmetros do modelo, outra abordagem deve ser utilizada. Se as incertezas associadas a estes parâmetros forem definidas por intervalos de pertinência (limites de incerteza) e não houver dúvidas relativas à ordem do modelo, tal incerteza é dita do tipo estruturada (Oliveira, 1997; Moreira, 2006).

Na literatura, trabalhos anteriores já abordaram a estimação dos intervalos de pertinência dos coeficientes do modelo OBF no contexto dos sistemas lineares (Oliveira, 1997; Oliveira et al., 2000; Moreira, 2006) e dos coeficientes do modelo OBF-Volterra em sistemas não lineares (da Rosa, 2009), utilizando um conjunto de respostas ao impulso. Contudo, naqueles trabalhos uma exigência imposta à solução dos problemas era a de que as funções ortonormais utilizadas na expansão fossem sempre positiva para todo instante de tempo k , o que não pode ser garantido para as funções de Laguerre, Kautz e GOBF. Portanto, o objetivo desta seção é apresentar uma solução para este problema, eliminando aquela restrição.

4.3.1 Cálculo dos limitantes dos coeficientes de modelos com incertezas

Apresenta-se a formulação de modelos lineares (OBF) e não lineares (OBF-Volterra) quando estes possuem incerteza sobre os coeficientes ou *kernels* do modelo, juntamente com a solução para o cálculo dos limitantes de incerteza.

Modelos OBF lineares incertos

Considere, inicialmente, um modelo de Volterra (2.1) de primeira ordem, e suponha que haja incerteza sobre seus parâmetros, aqui representados pelas respostas ao impulso do processo, ou seja, $h_{1_{\min}}(k) \leq \{h_1(k)\} \leq h_{1_{\max}}(k)$, para $k = 0, 1, 2, \dots$. Este conjunto pode ser reescrito de forma exata em função das variações dos valores da resposta ao impulso em cada instante de tempo k em torno do valor médio (Oliveira, 1997; Oliveira et al., 2000; Moreira, 2006):

$$\{h_1(k)\} = \{\bar{h}_1(k) + \rho_1(k) \Delta h_1(k)\} \quad (4.42)$$

onde

- $\bar{h}_1(k) = \frac{h_{1_{\max}} + h_{1_{\min}}}{2}$ é o valor médio do conjunto de respostas ao impulso;
- $\rho_1(k)$ é a incerteza sobre $h_1(k)$, tal que $\rho_1 \in [-1, 1]$;
- $\Delta h_1(k) = h_{1_{\max}}(k) - \bar{h}_1(k)$ representa o módulo da máxima variação das respostas ao impulso contidas em $\{h_1(k)\}$ em torno de $\bar{h}_1(k)$.

Os coeficientes da expansão em uma base de funções ortonormais $\{\psi_m\}$ de uma resposta ao impulso, conforme a equação (2.16) são dados por:

$$\alpha_m = \sum_{k=0}^{\infty} h_1(k) \psi_m(k), \quad m = 1, 2, \dots$$

Logo, para calcular o conjunto de coeficientes incertos do conjunto de respostas ao impulso $\{h_1(k)\}$, basta substituir (4.42) em (2.16):

$$\begin{aligned}\{\alpha_m\} &= \sum_{k=0}^{\infty} \{h_1(k)\} \psi_m(k) \\ &= \sum_{k=0}^{\infty} \{\bar{h}_1(k) + \rho_1(k) \Delta h_1(k)\} \psi_m(k), \quad m = 1, 2, \dots\end{aligned}\quad (4.43)$$

Como o sistema deve ser estável, para valores suficientemente grandes de k , a resposta ao impulso se anula ou pode ser desprezada, ou seja, $h_1(k) = 0$ para $k > N$. Assim, pode-se reescrever o conjunto de coeficientes incertos (4.43), como:

$$\{\alpha_m\} = \sum_{k=0}^N \bar{h}_1(k) \psi_m(k) + \sum_{k=0}^N \{\rho_1(k)\} \Delta h_1(k) \psi_m(k), \quad m = 1, \dots, M. \quad (4.44)$$

Embora não seja possível determinar analiticamente a máxima incerteza associada a cada coeficiente do desenvolvimento ortonormal $\alpha_{(\cdot)}$ do modelo incerto, é possível obter uma estimativa desta incerteza. Para isto, tomam-se os valores extremos dos coeficientes como sendo a máxima incerteza associada a cada parâmetro.

Portanto, o objetivo é encontrar os valores mínimo e máximo de cada um dos M conjuntos incertos de coeficientes $\{\alpha_m\}$, com $m = 1, \dots, M$. Ou seja, para $m = 1, \dots, M$, deseja-se encontrar:

$$\alpha_m^{\min} = \min \alpha_m \quad | \quad \alpha_m \in \{\alpha_m\},$$

e

$$\alpha_m^{\max} = \max \alpha_m \quad | \quad \alpha_m \in \{\alpha_m\}.$$

As únicas variáveis em (4.44), que descrevem os conjuntos de coeficientes incertos $\{\alpha_m\}$, são os escalares $\rho_1(k)$, com $k = 0, \dots, N$, os quais devem estar contidos no intervalo $[-1, 1]$. Além disso, o primeiro termo da soma em (4.44), que representa a expansão em série da resposta ao impulso média, é constante. Com base nestas duas observações, pode-se formular os dois problemas de otimização a seguir para a obtenção dos valores mínimo e máximo de cada um dos $m = 1, \dots, M$ coeficientes incertos, respectivamente:

$$\begin{aligned}\min_{\rho(0), \dots, \rho(N)} \quad & \sum_{k=0}^N \rho_1(k) \Delta h_1(k) \psi_m(k) \\ \text{s.a.} \quad & -1 \leq \rho_1(k) \leq 1 \quad k = 0, \dots, N\end{aligned}\quad (4.45)$$

e

$$\begin{aligned}\max_{\rho(0), \dots, \rho(N)} \quad & \sum_{k=0}^N \rho_1(k) \Delta h_1(k) \psi_m(k) \\ \text{s.a.} \quad & -1 \leq \rho_1(k) \leq 1 \quad k = 0, \dots, N,\end{aligned}\quad (4.46)$$

isto significa resolver $2M$ problemas (2 para cada $m = 1, \dots, M$) de programação linear, cada um dos quais com N variáveis de decisão. Em princípio, se N for suficientemente “grande” (sistema “lento” em tempo discreto), a solução numérica dos $2M$ problemas de otimização linear pode se tornar computacionalmente oneroso, senão impraticável. Note, porém, que os termos $\Delta h_1(k)$ na função objetivo de cada problema de programação linear são não negativos e as restrições sobre as variáveis $\rho_1(k)$, $\forall k = 0, \dots, N$ não envolvem interdependência entre essas variáveis. Assim, é possível calcular analiticamente as soluções ótimas para os problemas de otimização (4.45) e (4.46), respectivamente, como:

$$\rho_1^{\min}(k) = -\text{sinal}(\psi_m(k)) \quad k = 0, \dots, N \quad (4.47)$$

e

$$\rho_1^{\max}(k) = \text{sinal}(\psi_m(k)) \quad k = 0, \dots, N, \quad (4.48)$$

em que, a função sinal é a função que retorna o sinal de um número real, ou seja,

$$\text{sinal}(x) = \begin{cases} -1, & \text{se } x < 0 \\ 0, & \text{se } x = 0 \\ 1, & \text{se } x > 0. \end{cases}$$

Logo, substituindo as soluções (4.47) e (4.48) em (4.44) é possível calcular os valores mínimos e máximos do conjunto de coeficientes incertos $\{\alpha_m\}$, i.e.:

$$\alpha_m^{\min} = \sum_{k=0}^N \bar{h}_1(k) \psi_m(k) - \sum_{k=0}^N \text{sinal}(\psi_m(k)) \Delta h_1(k) \psi_m(k), \quad m = 1, \dots, M, \quad (4.49)$$

e

$$\alpha_m^{\max} = \sum_{k=0}^N \bar{h}_1(k) \psi_m(k) + \sum_{k=0}^N \text{sinal}(\psi_m(k)) \Delta h_1(k) \psi_m(k), \quad m = 1, \dots, M. \quad (4.50)$$

Os valores máximos (4.49) e mínimos (4.50) podem, portanto, ser utilizados para descrever as envoltórias superior e inferior, respectivamente, que contêm o conjunto de respostas ao impulso $\{h_1(k)\}$.

Observação 1. Em Oliveira (1997) e Oliveira et al. (2000), propõe-se uma solução para encontrar os coeficientes do modelo incerto, a qual é sucintamente descrita a seguir, para efeito de comparação.

Dado que $\rho_1(k)$ pertence ao intervalo $[-1, 1]$ e considerando que seja válida a condição:

$$\psi_m(k) > 0, \quad \forall k, \quad m = 1, \dots, M, \quad (4.51)$$

pode-se reescrever (4.44) utilizando a seguinte desigualdade (Oliveira, 1997; Oliveira et al., 2000):

$$-\sum_{k=0}^N \Delta h_1(k) \psi_m(k) \leq \sum_{k=0}^N \rho_1(k) \Delta h_1(k) \psi_m(k) \leq \sum_{k=0}^N \Delta h_1(k) \psi_m(k), \quad (4.52)$$

$m = 1, \dots, M.$

Consequentemente, os extremos da desigualdade dada em (4.52) são tomados como a máxima incerteza (α_m^{min} e α_m^{max}) sobre os coeficientes da expansão ortonormal do modelo linear.

Embora esta solução forneça uma estimativa para a incerteza associada aos parâmetros $\alpha_{(\cdot)}$, exige que a condição (4.51) seja atendida, o que em geral não se verifica para as bases ortonormais de Laguerre, Kautz e GOBF, tornando impeditiva a sua aplicação de maneira genérica. A solução proposta neste trabalho não só elimina esta restrição, mas também contém a solução proposta em Oliveira (1997) e Oliveira et al. (2000), pois assumindo-se em (4.45) e (4.46), hipoteticamente, que as funções ortonormais $\psi_m(k)$ são positivas para todo k , então as soluções encontradas em (4.49) e (4.50) são exatamente as mesmas soluções encontrada em (4.52). Trata-se, portanto, de uma solução exata, não de uma aproximação, para o problema de estimação dos limites dos coeficientes do modelo incerto. Apesar disso, a solução apresentada em (4.47) e (4.48) ainda pode ser conservadora, já que aproxima externamente o conjunto de coeficientes incertos por um hiper-retângulo no espaço dos coeficientes.

Modelos OBF não lineares incertos

Os conceitos apresentados na seção anterior podem ser estendidos para o contexto dos modelos não lineares (OBF-Volterra). Assim, considere um modelo de Volterra de ordem η , como o expresso em (2.23), possuindo incerteza sobre seus parâmetros, representados aqui pelos *kernels* do modelo. Similarmente, ao raciocínio utilizado na seção anterior, as incertezas podem ser descritas matematicamente por um conjunto de *kernels* do tipo:

$$\{h_\eta(k_1, \dots, k_\eta)\} = \{\bar{h}_\eta(k_1, \dots, k_\eta) + \rho_\eta(k_1, \dots, k_\eta)\Delta h_\eta(k_1, \dots, k_\eta)\} \quad (4.53)$$

onde

- $\bar{h}_\eta(k_1, \dots, k_\eta)$ é o valor médio do conjunto de *kernels* de Volterra;
- $\rho_\eta(k_1, \dots, k_\eta)$ é a incerteza sobre $h_\eta(k_1, \dots, k_\eta)$, tal que $\rho_\eta(k_1, \dots, k_\eta) \in [-1, 1]$;
- $\Delta h_\eta(k_1, \dots, k_\eta)$ representa o módulo da máxima variação das respostas ao impulso contidas em $\{h_\eta(k_1, \dots, k_\eta)\}$ em torno de $\bar{h}_\eta(k_1, \dots, k_\eta)$.

Os coeficientes da expansão em séries ortonormais do modelo são, portanto, calculados por meio de (2.16), ou seja:

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

Utilizando o conjunto incerto dado em (4.53) para reescrever a equação acima, e assumindo que o sistema é estável, ou seja, $h_\eta(k_1, \dots, k_\eta) = 0$ para $k > N$, tem-se o seguinte conjunto de coeficientes incertos:

$$\begin{aligned}
\{\alpha_{i_1, \dots, i_\eta}\} &= \sum_{k_1=0}^N \cdots \sum_{k_\eta=0}^N \{h_\eta(k_1, \dots, k_\eta)\} \prod_{l=1}^{\eta} \psi_{l, i_l}(k_l) \\
&= \sum_{k_1=0}^N \cdots \sum_{k_\eta=0}^N \{\bar{h}_\eta(k_1, \dots, k_\eta) + \rho_\eta(k_1, \dots, k_\eta) \Delta h_\eta(k_1, \dots, k_\eta)\} \prod_{l=1}^{\eta} \psi_{l, i_l}(k_l).
\end{aligned} \tag{4.54}$$

Novamente, o objetivo é encontrar os valores mínimo e máximo de cada um dos conjuntos incertos de coeficientes $\{\alpha_{i_1, \dots, i_\eta}\}$. Assim como no caso linear, as únicas variáveis em (4.54), que descreve o conjunto de coeficientes incertos, são os escalares $\rho_\eta(k_1, \dots, k_\eta)$, os quais estão contidos no intervalo $[-1, 1]$. Além disso, o termo que representa a expansão em série do valor médio do conjunto de *kernels* é constante. Com base nestas duas observações, pode-se formular os dois problemas de otimização apresentados a seguir:

$$\begin{aligned}
\min_{\rho_\eta(k_1, \dots, k_\eta)} \quad & \sum_{k_1=0}^N \cdots \sum_{k_\eta=0}^N \rho_\eta(k_1, \dots, k_\eta) \Delta h_\eta(k_1, \dots, k_\eta) \prod_{l=1}^{\eta} \psi_{l, i_l}(k_l) \\
\text{s.a.} \quad & -1 \leq \rho_\eta(k_1, \dots, k_\eta) \leq 1 \quad k_l = 0, \dots, N, \quad i_l = 1, \dots, M,
\end{aligned} \tag{4.55}$$

e

$$\begin{aligned}
\max_{\rho_\eta(k_1, \dots, k_\eta)} \quad & \sum_{k_1=0}^N \cdots \sum_{k_\eta=0}^N \rho_\eta(k_1, \dots, k_\eta) \Delta h_\eta(k_1, \dots, k_\eta) \prod_{l=1}^{\eta} \psi_{l, i_l}(k_l) \\
\text{s.a.} \quad & -1 \leq \rho_\eta(k_1, \dots, k_\eta) \leq 1 \quad k_l = 0, \dots, N, \quad i_l = 1, \dots, M,
\end{aligned} \tag{4.56}$$

o que significa resolver $2M^2$ problemas (modelos assimétricos) ou $M(M+1)$ (modelos simétricos), cada um com N variáveis de restrição. Como os termos $\Delta h_\eta(k_1, \dots, k_\eta)$ na função objetivo de (4.55) e (4.56) são não negativos e as restrições $\rho_\eta(k_1, \dots, k_\eta)$ não envolvem interdependência entre as variáveis, é possível determinar analiticamente as soluções ótimas dos problemas de otimização como:

$$\rho_\eta^{\min}(k_1, \dots, k_\eta) = -\text{sinal} \left(\prod_{l=1}^{\eta} \psi_{l, i_l}(k_l) \right), \quad k_l = 0, \dots, N, \quad i_l = 1, \dots, M, \tag{4.57}$$

e

$$\rho_\eta^{\max}(k_1, \dots, k_\eta) = \text{sinal} \left(\prod_{l=1}^{\eta} \psi_{l, i_l}(k_l) \right), \quad k_l = 0, \dots, N, \quad i_l = 1, \dots, M. \tag{4.58}$$

Substituindo as soluções (4.57) e (4.58) em (4.54), obtêm-se os valores máximos e mínimos do coeficiente incerto $\alpha_{i_1, \dots, i_\eta}$.

Observação 2. Em da Rosa (2009), propõe-se uma formulação para encontrar os limites das incertezas em modelos não lineares, similar àquela encontrada para o caso linear em Oliveira (1997) e Oliveira et al. (2000).

Dado que $\rho_\eta(k_1, \dots, k_\eta)$ pertence ao intervalo $[-1, 1]$ e considerando que seja válida uma extensão da condição (4.51), i.e.,

$$\prod_{l=1}^{\eta} \psi_{l,i_l}(k_l) > 0, \quad \forall k_l, \quad i_l = 1, \dots, M \quad (4.59)$$

e com base em (4.54), pode-se escrever a seguinte desigualdade (da Rosa, 2009):

$$\begin{aligned} - \sum_{k_1=0}^N \cdots \sum_{k_\eta=0}^N \Delta h_\eta(k_1, \dots, k_\eta) \prod_{l=1}^{\eta} \psi_{l,i_l}(k_l) &\leq \\ \sum_{k_1=0}^N \cdots \sum_{k_\eta=0}^N \rho_\eta(k_1, \dots, k_\eta) \Delta h_\eta(k_1, \dots, k_\eta) \prod_{l=1}^{\eta} \psi_{l,i_l}(k_l) &\leq \quad (4.60) \\ \sum_{k_1=0}^N \cdots \sum_{k_\eta=0}^N \Delta h_\eta(k_1, \dots, k_\eta) \prod_{l=1}^{\eta} \psi_{l,i_l}(k_l) &\end{aligned}$$

sendo os extremos da desigualdade (4.60) tomados como a máxima incerteza associada aos coeficientes da expansão ortonormal do modelo não linear. No entanto, assim como acontecia no caso linear, é preciso assegurar que a condição (4.59) seja atendida para os modelos não lineares, o que em geral não é verdade para as bases de Laguerre, Kautz e GOBF. A solução proposta aqui não só elimina esta restrição, como também fornece uma solução exata para o problema de estimação dos coeficientes incertos, o que não ocorria na solução proposta em da Rosa (2009).

4.3.2 Relação com a Solução de Sobre-estimação do Conjunto de Respostas ao Impulso do Processo

Em Moreira (2006), propõe-se uma formulação para o cálculo dos limitantes das incertezas sobre os coeficientes em modelos dinâmicos lineares por meio da solução de um problema de minimização que sobre-estima o conjunto de respostas ao impulso e, em da Rosa (2009) e da Rosa et al. (2010), estende-se esta ideia para contemplar os processos não lineares (Veja Apêndice B).

Considere apenas a solução para a estimação do conjunto de coeficientes incertos com base nas respostas ao impulso do processo linear. Busca-se encontrar o hiper-retângulo (no espaço dos coeficientes) com volume mínimo que contenha os coeficientes capazes de representar os limitantes superior e inferior para as incertezas na resposta ao impulso. Ou seja, busca-se encontrar o menor hiper-retângulo que contenha os coeficientes α'_m e α''_m , $m = 1, 2, \dots$ tais que, para todo k , tem-se:

$$\sum_{m=1}^{\infty} \alpha'_m \psi_m(k) \leq h_{1_{min}}(k) \quad (4.61)$$

$$\sum_{m=1}^{\infty} \alpha''_m \psi_m(k) \geq h_{1_{max}}(k).$$

Uma solução factível (não necessariamente ótima) é um hiper-retângulo qualquer que contenha os coeficientes α'_m e α''_m , $m = 1, 2, \dots$ que representam os próprios envelopes $h_{1_{min}}(k)$ e $h_{1_{max}}(k)$, respectivamente, ou seja, aqueles coeficientes que tornam igualdades as desigualdades em (4.61). Esses coeficientes são dados por:

$$\begin{aligned}\alpha'_m &= \sum_{k=0}^{\infty} h_{1_{min}}(k) \psi_m(k) \\ \alpha''_m &= \sum_{k=0}^{\infty} h_{1_{max}}(k) \psi_m(k).\end{aligned}\tag{4.62}$$

Como $h_{1_{min}}(k) = \bar{h}_1(k) - \Delta h_1(k)$ e $h_{1_{max}}(k) = \bar{h}_1(k) + \Delta h_1(k)$, pode-se reescrever os coeficientes como:

$$\begin{aligned}\alpha'_m &= \sum_{k=0}^{\infty} (\bar{h}_1(k) - \Delta h_1(k)) \psi_m(k) \\ \alpha''_m &= \sum_{k=0}^{\infty} (\bar{h}_1(k) + \Delta h_1(k)) \psi_m(k).\end{aligned}\tag{4.63}$$

Lembrando que $\Delta h_1(k) \geq 0$, é evidente a partir de (4.63) que, para todo $m = 1, 2, \dots$ individualmente, as seguintes desigualdades são válidas:

$$\begin{aligned}\sum_{k=0}^{\infty} [\bar{h}_1(k) - \text{sign}(\psi_m(k)) \Delta h_1(k)] \psi_m(k) &\leq \alpha'_m \\ \sum_{k=0}^{\infty} [\bar{h}_1(k) + \text{sign}(\psi_m(k)) \Delta h_1(k)] \psi_m(k) &\geq \alpha''_m\end{aligned}\tag{4.64}$$

o que significa que α'_m e α''_m estão contidos no intervalo dado precisamente pela solução ótima apresentada na Seção 4.3.1, ou seja, α'_m e $\alpha''_m \in [\alpha_m^{min}, \alpha_m^{max}]$ ².

Resumindo, o resultado em (4.64) diz que coeficientes α'_m e α''_m que representam, para todo m , soluções factíveis para o problema de otimização formulado em Moreira (2006) e da Rosa (2009), estão contidos no hiper-retângulo de incertezas dado pelas soluções ótimas corrigidas apresentadas na Seção 4.3.1. Logo, um hiper-retângulo de volume mínimo que contenha esses coeficientes (o que é buscado pelo problema de otimização) será um subconjunto do hiper-retângulo de incertezas calculados na Seção 4.3.1, sendo que ambos possuem centros comuns em $\bar{\alpha}_m$, ($m = 1, 2, \dots$)³. Isso confirma a percepção de que a formulação em Moreira (2006) e da Rosa (2009) pode produzir

²Observe de (4.63) e (4.64) que os extremos do intervalo (igualdade em (4.64)) só são atingidos se a função ortonormal referente aos coeficientes em questão (m -ésima função $\psi_m(k)$) for sempre positiva ou negativa para todo k , o que em geral não se pode garantir para as funções de Laguerre, Kautz e GOBF.

³Os termos $\bar{\alpha}_m$ são os coeficientes da expansão da resposta média $\bar{h}(k)$.

resultados menos conservativos. Entretanto, esse menor conservadorismo tem um preço. De fato, note que as seguintes respostas ao impulso:

$$h_1(k) = \bar{h}_1(k) - \text{sinal}(\psi_m(k)) \Delta h_1(k) \quad (4.65)$$

$$h_1(k) = \bar{h}_1(k) + \text{sinal}(\psi_m(k)) \Delta h_1(k) \quad (4.66)$$

para um dado m qualquer, (4.65) e (4.66) incluem o conjunto incerto $\{h_1(k)\}$ e são mapeadas nos limitantes do intervalo $[\alpha_m^{\min}, \alpha_m^{\max}]$ do m -ésimo coeficiente. Essas duas funções estariam mapeadas fora do hiper-retângulo mínimo que contém α'_m e α''_m . Como esses dois coeficientes foram arbitrados em (4.62), apenas como uma solução factível - de acordo com as restrições em (4.61) - para o problema formulado em Moreira (2006) e da Rosa (2009), não se pode, em princípio, garantir que a solução ótima do problema não seja um hiper-retângulo de volume total igual ou ainda menor, mas que seja mais amplo nesta (m -ésima) dimensão em particular a ponto de compreender ao mesmo tempo outros coeficientes α'_m e α''_m que satisfaçam (4.61), os extremos do intervalo $[\alpha_m^{\min}, \alpha_m^{\max}]$ e, por conseguinte, as respostas ao impulso correspondentes em (4.65) e (4.66). O alargamento do hiper-retângulo na m -ésima dimensão, no entanto, exigiria um estreitamento em ao menos uma das demais dimensões, o que certamente implicaria que respostas ao impulso dadas por (4.65) e (4.66) com valores de m referentes a essa(s) outra(s) dimensão(ões) (e que também fazem parte do conjunto incerto $\{h_1(k)\}$) não estariam representadas neste hiper-retângulo.

É evidente que as respostas (4.65) e (4.66) para todo m incluem o conjunto $\{h_1(k)\}$ e precisam todas estar contidas no hiper-retângulo de incertezas, constituindo uma sobre-estimação de $\{h_1(k)\}$. Isso só seria possível com um hiper-retângulo maior que aquele correspondente à solução factível que utilizamos em todo o raciocínio acima. Por definição, este hiper-retângulo maior não poderia nunca ser solução ótima para o problema formulado em Moreira (2006) e da Rosa (2009). Logo, conclui-se que a formulação em Moreira (2006) e da Rosa (2009) não produz um mapeamento completo do espaço $\{h_1(k)\}$ em α_m , $m = 1, 2, \dots$. Em outras palavras, dada uma combinação qualquer de coeficientes contidos no hiper-retângulo que define as incertezas, não é possível garantir que a resposta ao impulso correspondente estará contida no conjunto $\{h_1(k)\}$.

Um raciocínio similar pode ser aplicado para a estimação do conjunto de coeficientes incertos com base nas respostas ao impulso do processo não linear, como desenvolvido em da Rosa (2009), levando à mesma conclusão, ou seja, dada uma combinação qualquer de coeficientes contidos no hiper-retângulo que define as incertezas, encontrado pelo problema de otimização (B.5), não é possível garantir que o *kernel* correspondente aos coeficientes escolhidos estará contido no conjunto $\{h_\eta(k_1, \dots, k_\eta)\}$.

Nota 1: Observe que as respostas ao impulso em (4.65) e (4.66) são bastante particulares e portanto improváveis na prática, o que ilustra bem como a solução apresentada na Seção 4.3.1 é de fato conservadora. O fato da solução do problema formulado em Moreira (2006) e da Rosa (2009) não contemplar essas respostas, não implica que não possa ser uma boa aproximação em sistemas reais (representando as dinâmicas prováveis) que, por ser menos conservativa, seja até mais útil e eficaz em muitas aplicações práticas. Portanto, pode-se escolher entre um menor conservadorismo, aplicando as soluções propostas em Moreira (2006) e da Rosa (2009), correndo o risco de produzir uma resposta

ao impulso ou *kernel* de Volterra, utilizando um conjunto genérico de coeficientes contidos na solução $[\alpha'_m, \alpha''_m]$ ou $[\alpha'_{i_1, \dots, i_\eta}, \alpha''_{i_1, \dots, i_\eta}]$, que não esteja contido no conjunto $\{h_1(k)\}$ ou $\{h_\eta(k_1, \dots, k_\eta)\}$, respectivamente.

Nota 2: Tendo-se em vista que os problemas de otimização (B.4) e (B.5) são não lineares, soluções iniciais diferentes podem conduzir a soluções ótimas diferentes. Assim, pode ser difícil obter numericamente o mínimo global daqueles problemas, podendo apresentar como soluções mínimos locais (Moreira, 2006). Além disso, a solução daqueles problemas podem exigir um grande esforço computacional, diferentemente do que ocorre com a solução apresentada na Seção 4.3.1, uma vez que fornece a solução exata e analítica reduzindo o esforço computacional requerido. Entretanto, como já comentado, esta leva a soluções mais conservadoras, conforme será discutido na Seção 4.4.

4.4 Exemplo Ilustrativo - Modelos com Incertezas no *Kernel*

O objetivo desta seção é ilustrar a formulação proposta na Seção 4.3, cujo objetivo é obter os limites das incertezas dos parâmetros nos modelos OBF-Volterra (incerteza estruturada). Assim, suponha um sistema não linear, como aquele descrito pelas equações (4.37) e (4.38), composto pelos *kernels* incertos de Volterra de primeira e segunda ordem:

$$h_1(k_1) = \mathcal{Z}^{-1} \left[\frac{\epsilon_1 (z + 0.5)}{(z - 0, 3)(z - 0, 2)} \right] \quad (4.67)$$

$$h_2(k_1, k_2) = 0, 25 g_{i_1}(k_1) g_{i_2}(k_2) \quad \text{onde} \quad g_{i_l}(k) = \mathcal{Z}^{-1} \left[\frac{\epsilon_{i_l} (z + 1)}{(z - 0, 8)(z + 0, 8)} \right] \quad (4.68)$$

onde ϵ_1 e $(\epsilon_{i_1}, \epsilon_{i_2})$ são parâmetros incertos do processo e estão restritos aos seguintes intervalos: $\epsilon_1 \in [0, 85; 1, 15]$ e $(\epsilon_{i_1}, \epsilon_{i_2}) \in [0, 87; 1, 14] \times [0, 87; 1, 14]$. Para a simulação computacional, os intervalos foram discretizados com uma taxa de amostragem igual a 0, 01. Tem-se, portanto:

$$\begin{cases} \epsilon_1(j) = 0, 85 + \kappa \times 0, 01 & \kappa = 0, 1, \dots, 30 \\ \epsilon_{i_1}(j') = 0, 87 + \kappa' \times 0, 01 & \kappa' = 0, 1, \dots, 27 \\ \epsilon_{i_2}(j'') = 0, 87 + \kappa'' \times 0, 01 & \kappa'' = 0, 1, \dots, 27. \end{cases}$$

Os conjuntos de respostas ao impulso $\{h_1(k_1)\}$ e $\{h_2(k_1, k_2)\}$ é calculado simulando-se o processos (4.67) e (4.68) para os diversos valores factíveis de ϵ_1 e $(\epsilon_{i_1}, \epsilon_{i_2})$. Além disso, 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 calculados como sendo a média de todos os *kernels* factíveis de $h_1(k_1)$ e $h_2(k_1, k_2)$ (Moreira, 2006; da Rosa, 2009). Portanto, os *kernels* centrais são obtidos por:

$$\bar{h}_1(k_1) = \frac{1}{31} \sum_{\kappa=1}^{31} h_1(k_1, \epsilon_1(j)) \quad (4.69)$$

$$\bar{h}_2(k_1, k_2) = \frac{1}{28 \times 28} \sum_{\kappa'=1}^{28} \sum_{\kappa''=1}^{28} h_2(k_1, k_2, \epsilon_{i_1}(j'), \epsilon_{i_2}(j'')). \quad (4.70)$$

Apenas para efeito de ilustração, o conjunto de *kernels* de Volterra de primeira ordem $\{h_1(k_1)\}$ gerado para cada valor do parâmetro incerto ϵ_1 em (4.67) é apresentado na Figura 4.5. Enquanto que na Figura 4.6, ilustram-se os limites superior e inferior do conjunto de *kernels* de segunda ordem $h_2(k_1, k_2)$ gerado por (4.68).

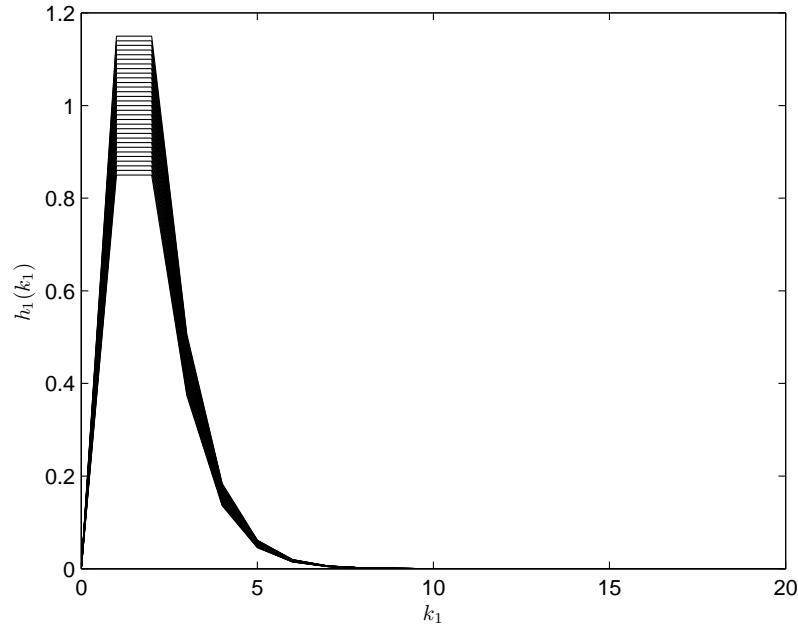


Fig. 4.5: Conjunto de *kernels* de primeira ordem.

Para fins de simulação, o conjunto de funções ortonormais utilizado para modelar o sistema incerto OBF-Volterra de primeira ordem, em (4.67), é a Base de Funções Ortonormais Generalizadas (GOBF). Para o exemplo em questão, utilizam-se 4 funções, $M = 4$, e as respostas ao impulso foram truncadas para $k_1 > 20$. Usando os polos ótimos do *kernel* central $\bar{h}_1(k_1)$, identificados de acordo com a metodologia apresentada no Capítulo 3, os limitantes superior e inferior dos coeficientes incertos $\{\alpha_{i_1}\}$ encontrados são apresentados na Tabela 4.9.

Tab. 4.9: Limites dos coeficientes α dos *kernels* incertos $\{h_1(k_1)\}$ com 4 funções (GOBF) na base.

Ordem do <i>kernel</i> (η)	α^{min}	α^{max}
1	0,2597	0,3513
	1,2013	1,6252
	0,1195	0,5205
	-0,3149	-0,0152

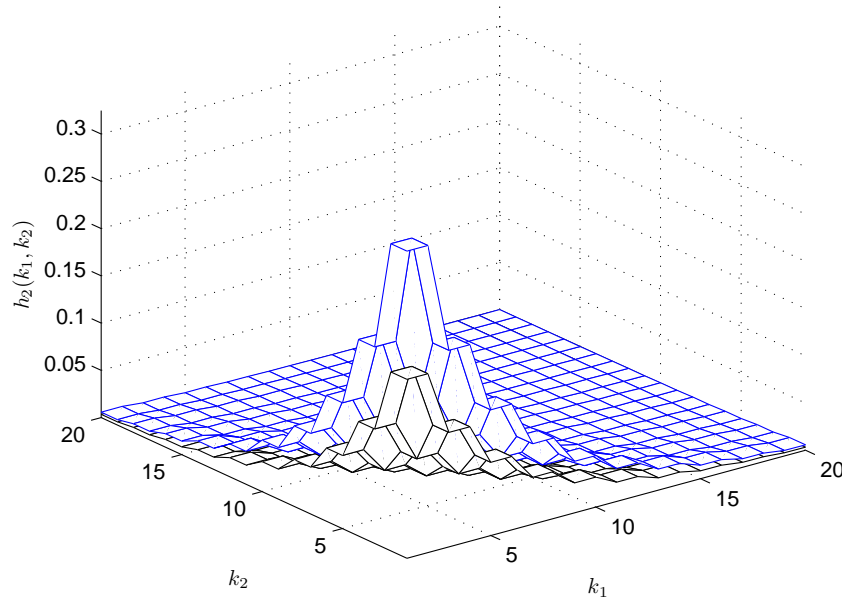


Fig. 4.6: Limites inferior (preto) e superior (azul) do conjunto de *kernels* de segunda ordem.

Na Figura 4.7, apresentam-se os limites superior e inferior do conjunto de *kernels* de primeira ordem gerado por (4.67), juntamente com os envelopes superior e inferior do modelo obtidos com a metodologia proposta neste trabalho. Note que o conjunto incerto de respostas ao impulso está contido no conjunto de respostas ao impulso mapeadas pelos coeficientes incertos obtidos como solução.

Estes resultados permitem comparar a metodologia apresentada aqui com outra já existente na literatura, mais especificamente as metodologias apresentadas em Oliveira (1997), Oliveira et al. (2000) e da Rosa (2009), conforme discutido na Observação 1 da Seção 4.3. A análise da Figura 4.8 permite concluir que aquela metodologia não é capaz de representar os limites de $\{h_1\}$ em todo o intervalo. Pode-se observar que entre os instantes $1,8 < k_1 < 2,2$, o limite superior do modelo não contém o limitante superior do processo. Além disso, a partir de $k_1 > 4,8$, novamente o limite superior do modelo não contém o limitante superior do conjunto de respostas ao impulso dos *kernels* incertos do processo. Isto ocorre porque a condição (4.51) não foi satisfeita. Como já discutido, esta condição exige que as funções ortonormais sejam positivas para todo k , o que em geral não ocorre para as bases de Laguerre, Kautz e, inclusive, GOBF, utilizada neste exemplo.

Na Figura 4.9, apresenta-se uma comparação entre os limites superior e inferior da metodologia proposta aqui e aquela desenvolvida em trabalhos anteriores (Oliveira, 1997; Oliveira et al., 2000; da Rosa, 2009) e os limitantes superior e inferior do conjunto de respostas ao impulso do processo. O objetivo é mostrar que a abordagem anterior não consegue representar os limites do *kernel* incerto em todo instante de tempo. Apenas para efeito de melhor visualização, será apresentado o intervalo que compreende os instantes $k = 0, 5$ e $k = 4$.

Obtém-se agora os limites superior e inferior dos coeficientes incertos $\{\alpha_{i_1, i_2}\}$ do modelo (4.68). Novamente, utilizam-se $M = 4$ funções ortonormais generalizadas na expansão dos *kernels* incertos, as respostas ao impulso foram truncadas para $k_l > 20$, para $l = 1, 2$, e foram utilizados os polos

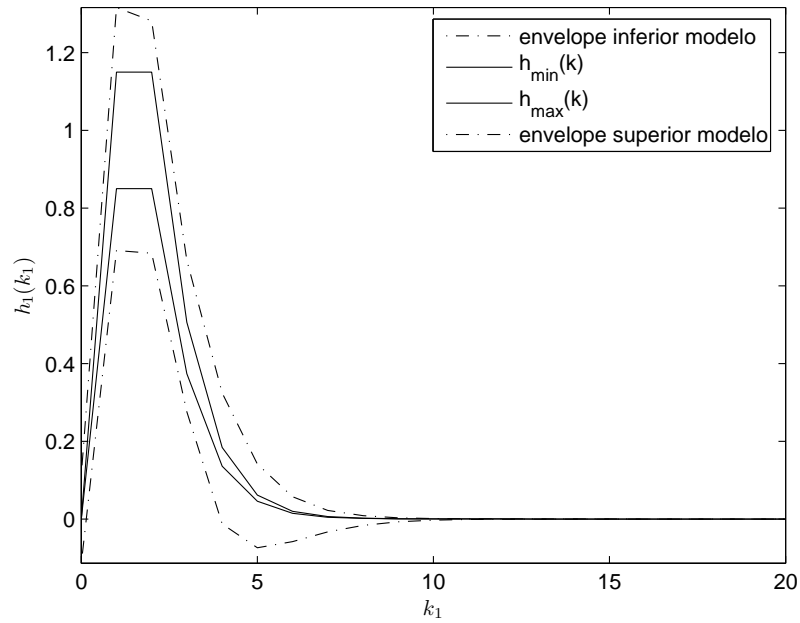


Fig. 4.7: Limites superior e inferior do modelo (linhas pontilhadas) e limitantes superior e inferior (linhas contínuas) do conjunto de repostas ao impulso dos *kernels* incertos de primeira ordem do processo.

ótimos do *kernel* central $\bar{h}_2(k_1, k_2)$, obtidos segundo a metodologia apresentada no Capítulo 3. Além disso, supõe-se a simetria do *kernel* de segunda ordem. Os valores obtidos são apresentados na Tabela 4.10.

Tab. 4.10: Limites dos coeficientes α dos *kernels* incertos $\{h_2(k_1, k_2)\}$ com 4 funções (GOBF) na base.

Ordem do <i>kernel</i> (η)	α^{min}	α^{max}
2	0,2854	0,4900
	0,2161	0,4658
	-0,0904	0,1401
	0,1683	0,3620
	0,1474	0,4522
	-0,1188	0,1625
	0,1149	0,3514
	-0,1282	0,1314
	-0,0921	0,1261
	0,0896	0,2730

De forma similar, na Figura 4.10, apresenta-se uma comparação para os limites obtidos para o

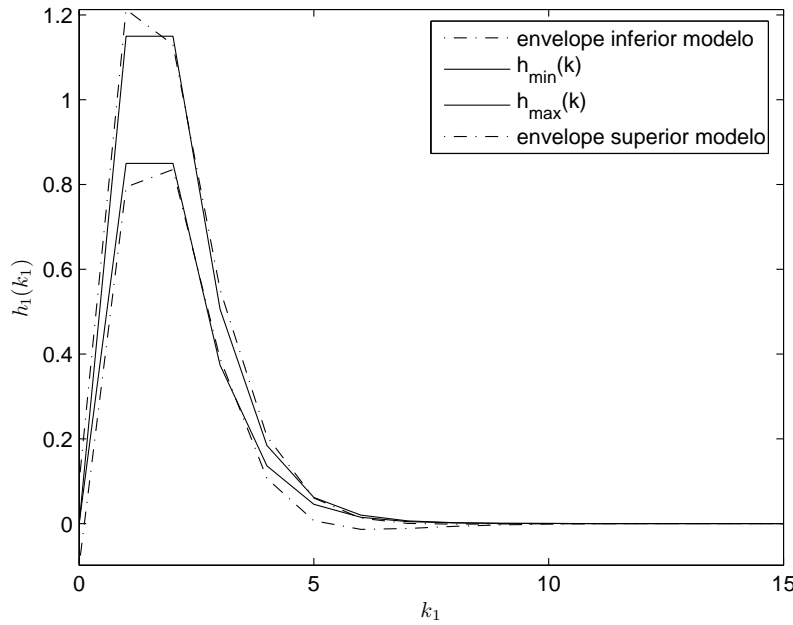


Fig. 4.8: Limites superior e inferior do modelo (linhas pontilhadas) e limitantes superior e inferior (linhas contínuas) do conjunto de repostas ao impulso dos *kernels* incertos de primeira ordem do processo utilizando a metodologia proposta em Oliveira et al. (2000) e da Rosa (2009).

kernel de segunda ordem, a partir de uma projeção no plano $k_2 = 5$. Observa-se que o método anterior proposto em da Rosa (2009) não é capaz de representar os limites de $\{h_2\}$ para todo instante de tempo k . A explicação para este fato é que a condição (4.59) não é satisfeita para todo instante de tempo de observação.

Como pode ser observado pela análise, em especial, das Figuras 4.9 e 4.10, embora a solução proposta garanta que o modelo encontrado consegue representar os *kernels* do processo incerto, ela pode ser considerada conservadora, já que aproxima externamente o conjunto de coeficientes incertos por um hiper-retângulo no espaço dos coeficientes.

4.5 Resumo e Conclusão

A abordagem UBBE para estimação paramétrica é uma alternativa interessante à abordagem estocástica, pois supera algumas desvantagens desta última, como o não conhecimento da função densidade de probabilidade que caracteriza o ruído, entre outras. Esta abordagem calcula uma região do espaço paramétrico denominada região de incerteza paramétrica, que é compatível com a estrutura do modelo, com os limites conhecidos do erro e com todos os dados medidos. A caracterização das incertezas é por meio de um intervalo, limites superior e inferior do erro, supostamente conhecido. Se a estimação da região politópica for feita por um algoritmo do tipo hipercubo, para cada parâmetro do modelo fornece-se um valor mínimo e máximo.

Entretanto, esta abordagem só é aplicável a modelos lineares nos parâmetros, o que pode ser

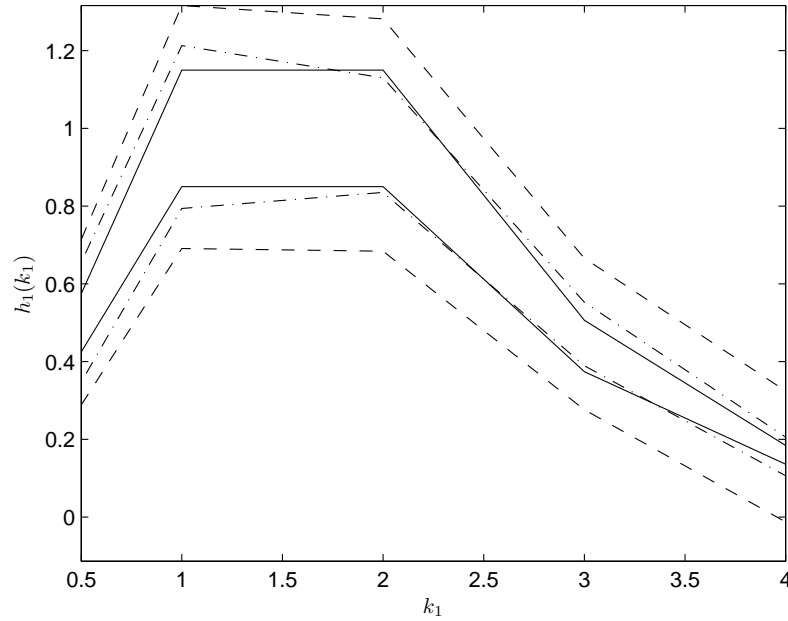


Fig. 4.9: Comparação entre os limites inferior e superior do *kernel* de primeira ordem obtidos por meio da abordagem proposta neste trabalho (linhas tracejadas), da abordagem em Oliveira et al. (2000) (linhas pontilhadas) e os limitantes do *kernel* incerto de primeira ordem (linhas contínuas).

um empecilho para a modelagem de sistemas não lineares incertos. Desta feita, os modelos OBF-Volterra apresentam-se como uma representação matemática viável, uma vez que são lineares nos parâmetros da expansão ortonormal. A linearidade dos parâmetros permite escrevê-los numa estrutura de regressão linear e, portanto, aplicar a abordagem UBBE.

Apresentou-se a descrição de dois métodos de estimação robusta, hipercubo externo (Milanese and Belforte, 1982) e descrição exata (Mo and Norton, 1990). O primeiro método descreve a região por meio de um hipercubo externo à região. Embora este método forneça exatamente os intervalos de incerteza paramétrica, o método não é recursivo e exige a solução de $2P$ problemas de programação linear com P variáveis e $2N$ restrições. O segundo método identifica recursivamente a região politópica de forma exata, o que o torna viável para aplicações em tempo real. Por outro lado, sua complexidade aumenta quando o número de medidas cresce.

Outro problema investigado foi a estimação robusta de processos com incerteza estruturada. Nestes processos, as incertezas sobre os parâmetros do modelo são definidas por intervalos de pertinência e conhece-se a ordem do modelo. A estratégia apresentada neste trabalho permite obter os limitantes das incertezas com base no conjunto de respostas ao impulso do processo. Diferentemente de outras abordagens na literatura (Oliveira, 1997; Oliveira et al., 2000; da Rosa, 2009), eliminou-se a exigência de que as funções ortonormais fossem positivas para todo instante de tempo, uma vez que tal condição, em geral, não se verifica para as bases de funções ortonormais mais utilizadas na literatura, Laguerre, Kautz e GOBF. A eliminação desta restrição permite, então, obter um modelo com incertezas que representa todos os *kernels* do processo incerto.

Os exemplos ilustram que as abordagens são aplicáveis aos sistemas não lineares descritos por

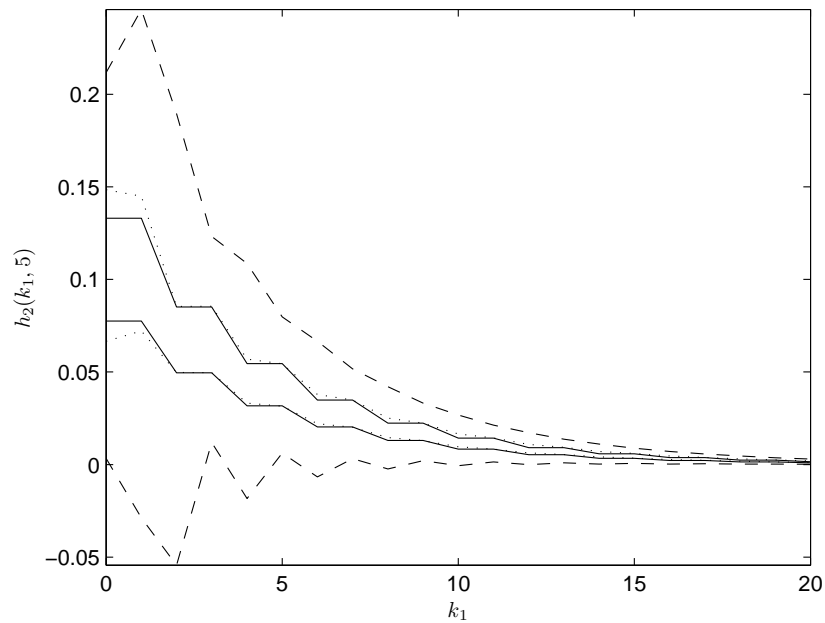


Fig. 4.10: Comparação entre os limites inferior e superior do *kernel* de segunda ordem obtidos por meio da abordagem proposta neste trabalho (linhas tracejadas), da abordagem em da Rosa (2009) (linhas pontilhadas) e os limitantes do *kernel* incerto de primeira ordem (linhas contínuas).

modelos OBF-Volterra. Para a modelagem UBBE foram utilizados modelos de segunda ordem de Volterra simétricos e assimétricos. Com relação à processos com incerteza estruturada, o exemplo apresentado permitiu concluir que os limites calculados para o modelo compreendem todos os possíveis *kernels* do processo. Conhecer os limites de incerteza é importante em problemas de controle preditivo robusto, pois o sinal de entrada é calculado considerando-se o pior desempenho possível do sinal de referência.

Capítulo 5

Conclusões

5.1 Conclusões e Contribuições

A modelagem de sistemas não lineares tem sido objeto de estudo de muitos pesquisadores nas últimas décadas. Enquanto, anteriormente, as abordagens para descrever tais sistemas eram superficiais e aplicáveis a uma classe restrita de problemas da engenharia, hoje com as melhorias na informática e as técnicas de otimização, uma grande variedade de sistemas não-lineares com propriedades diferentes podem ser melhor estudados. Apoiando-se nestes conceitos, este trabalho tratou da identificação de processos não-lineares por meio de modelos de Volterra. Tais modelos possuem diversas características que os tornam atrativos para aplicações práticas, como o fato de serem lineares nos parâmetros, que permite estimar os parâmetros utilizando técnicas de regressão linear. Entretanto, como não possuem a realimentação da saída, os modelos de Volterra tendem a exigir um grande número de termos para representar os *kernels*. No entanto, esta desvantagem, como detalhado no Capítulo 2, pode ser reduzida por expandir cada *kernel* do modelo por meio de uma base de funções ortonormais. Esta representação, também chamada OBF-Volterra, reduz o conjunto de parâmetros a ser estimado simplificando os problemas de identificação e controle. Outra vantagem desta estrutura tem que ver com a capacidade de incorporar conhecimento prévio sobre a dinâmica do sistema, como por exemplo atraso de transporte. Como as bases são parametrizadas por um conjunto de polos (parâmetros de projeto), o número de termos pode ser ainda mais reduzido se os polos escolhidos para parametrizar a base estiverem próximos dos polos reais do processo. Esta ideia foi abordada em (da Rosa, 2009; da Rosa et al., 2009), no entanto, abordou-se apenas os casos onde os *kernels* eram simétricos e, portanto, todas as direções eram parametrizadas por um único conjunto de funções ortonormais. Este trabalho, portanto, abordou o problema da seleção ótima dos polos da base, quando os *kernels* são assimétricos e podem ser parametrizados por múltiplos conjuntos de bases ortonormais, ou ainda cada direção do *kernel* pode ser descrita por uma base de funções ortonormais.

Como principal contribuição, pode-se citar a obtenção de um procedimento de otimização numérica de bases ortonormais cujos *kernels* de Volterra sejam assimétricos. Este procedimento baseia-se no cálculo das expressões analíticas para as derivadas das saídas dos filtros ortonormais com relação aos polos da base, as quais fornecem direções de buscas a serem seguidas pelo algoritmo de otimização. As derivadas das saídas dos filtros ortonormais foram obtidas de forma recursiva, resultando em um método mais eficiente computacionalmente, pois permitiu a redução do tempo de processamento e uma possível implementação *on-line*. Ênfase especial foi dada as funções ortonormais de

Kautz e GOBF. Esta metodologia mostrou-se eficaz para a utilização em modelos OBF-Volterra que apresentam *kernels* assimétricos, dinâmicas dominantes diferenciadas ao longo de suas múltiplas direções, reduzindo o erro de truncamento quando as bases estão subparametrizadas ou ainda redução no número de termos quando fixada uma mesma precisão para as abordagens. Outro fator importante foi a redução no tempo de processamento pela utilização das equações dos cálculos dos gradientes das saídas dos filtros de forma recursiva. Outra vantagem desta metodologia é a redução no tempo de processamento com a utilização das equações recursivas para o cálculo dos gradientes dos filtros ortonormais.

No Capítulo 4, estudou-se o problema da estimação robusta sob a hipótese de erro limitado mas desconhecido (UBBE). Esta abordagem, diferentemente da abordagem estocástica, calcula uma região de pertinência para os parâmetros do modelo que é compatível com as medidas e os limites das perturbações. Como importante característica pode-se citar o fato desta abordagem ser mais realista, pois caracterizam-se as incertezas apenas por meio dos limites inferior e superior. Apresentaram-se também dois métodos de identificação robusta: solução por Hipercubo Externo (Milanese and Belforte, 1982) e o algoritmo de descrição exata da região politópica (Mo and Norton, 1990). E mostrou-se que embora tais abordagens e métodos tenham sido desenvolvidos para modelos lineares, podem ser aplicados nos modelos OBF-Volterra já que estes modelos embora utilizados para descrever sistemas não-lineares possuem a propriedade de serem lineares nos parâmetros da expansão ortonormal. A análise dos exemplos ilustrativos, permitiram concluir que a solução, representada pelas envoltórias superior e inferior dos modelos compreende o valor real da saída do sistema.

Abordou-se ainda a modelagem de sistemas não-lineares incertos com incerteza estruturada utilizando modelos OBF-Volterra. Como principal contribuição foi o desenvolvimento de uma metodologia que permite calcular de forma exata os intervalos de pertinência dos coeficientes do modelo. Esta metodologia eliminou as restrições imposta por soluções anteriores (Oliveira, 1997; Oliveira et al., 2000; da Rosa, 2009) de que as bases ortonormais deveriam ser positivas para todo instante de observação, condição que não é satisfeita, em geral, para as principais bases de funções ortonormais utilizadas na literatura, Laguerre, Kautz e GOBF. O exemplo explorado naquele capítulo mostrou que os limites calculados para o modelo compreendem todos os possíveis *kernels* do processo.

5.2 Perspectivas e Trabalhos Futuros

Como futuras perspectivas de trabalhos, relacionadas ao tema deste trabalho, sugere-se, entre outras, as seguintes:

- Estimação adaptativa, em tempo real, dos modelos OBF-Volterra para utilização em sistemas de controle. Neste caso, busca-se estabelecer uma técnica recursiva de obtenção dos modelos OBF em tempo real para modelagem de sistemas variantes no tempo e possível implementação em sistemas de controle adaptativos.
- Cálculo recursivo dos limitantes de incerteza na abordagem UBBE, uma vez que o método de descrição exata da região politópica (Mo and Norton, 1990), utilizado no Capítulo 3, é recursivo apenas na obtenção do politopo.
- Estudo de modelos de Volterra que apresentem erros de medida nos sinais de saída e na entrada.

Bibliografia

- Aguirre, L. A. (2007). *Introdução à Identificação de Sistemas - Técnicas Lineares e Não-Lineares*. Editora UFMG, Belo Horizonte, 3ª edition.
- Anton, H. (2007). *Calculus, A New Horizon*. Jonh Wiley & Son, INC., New York, 7 edition.
- Arruda, L., Favier, G., and do Amaral, W. (1990). A comparative study of robust parameter estimation methods. In *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on*, volume 5, pages 2967 –2970.
- Arruda, L. V. R. (1992). *Etude d'Algorithmes d'Estimation Robust et Développement d'un Système à Base de Connaissance pour l'Identification*. PhD thesis, Université de Nice, Sophia Antipolis, França.
- Bazaraa, M. S., Sherali, H. D., and Shetty, C. M. (2006). *Nonlinear Programming: Theory and Algorithms*. Jonh Wiley & Sons, 3ª edição. edition.
- Belforte, G. and Bona, B. (1985). An improved parameter identification algorithm for signals with unknown but bounded errors. In *Proc. 7th IFAC/IFORS Symp. Identification Syst. Parameter Estimation*, York, U. K.
- Belforte, G., Teo, Y., and Tay, T. (1992). Recursive parameter estimation for time-varying systems in presence of unknown but bounded measurement noise. In *Proceedings of the International Conference on Intelligent Control and Instrumentation, 1992. SICICI '92.*, volume 2, pages 945 –950, Singapore.
- Bertsekas, D. and Rhodes, I. (1971). Recursive state estimation for a set-membership description of uncertainty. *IEEE Transactions on Automatic Control*, 16(2):117–128.
- Billings, S. (1980). Identification of nonlinear systems - a survey. *Control Theory and Applications, IEE Proceedings D*, 127(6):272 –285.
- Bokor, J. and Schipp, F. (1998). Approximate identification in laguerre and kautz basis. *Automatica*, 34(4):463–468.
- Boyd, S. and Chua, L. (1985). Fading memory and the problem of approximating nonlinear operators with volterra series. *IEEE Transactions on Circuits and Systems*, 32(32):1150–1161.
- Broome, P. W. (1965). Discrete orthonormal sequences. *Journal of the Association for Computing Machinery*, 12(2):151–168.

- Campello, R. J., do Amaral, W. C., and 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. (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. and Oliveira, G. H. C. (2007). Modelos não-lineares. In Aguirre, L. A., Silva, A. P. A., Campos, M. F. M., and Amaral, W. C., editors, *Enciclopédia de Automática*, volume 3. Edgar Blücher.
- Campello, R. J. G. B., Oliveira, G. H. C., and Amaral, W. C. (2007a). Identificação e controle de processos via desenvolvimentos em série ortonormais: Parte a: Identificação. *Controle & Automação*, 18(3):301–321.
- Campello, R. J. G. B., Oliveira, G. H. C., and Amaral, W. C. (2007b). Identificação e controle de processos via desenvolvimentos em série ortonormais: Parte b: Controle. *Controle & Automação*, 18(3):321–336.
- Clement, T. and Gentil, S. (1988). Reformulation of parameter identification with unknown-but-bounded errors. *Mathematics and Computers in Simulation*, 30(3):257–270.
- da Rosa, A., , Campello, R. J. G. B., Ferreira, P. A. V., Oliveira, G. H. C., and Amaral, W. C. (2010). Robust expansion of uncertain volterra kernels into orthonormal series. In *Proc. American Control Conference*, Baltimore, EUA.
- da Rosa, A. (2009). *Identificação de Sistemas Não-Lineares Usando Modelos de Volterra Baseados em Funções Ortonormais de Kautz e Generalizadas*. Tese de doutorado, Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas (UNICAMP), Campinas-SP, Brasil.
- da Rosa, A., Amaral, W. C., and Campello, R. J. G. B. (2005). Choice of free parameters in expansions of discrete-time volterra models using kautz functions. In *Proc. 16th IFAC World Congress*, Praga, República Tcheca.
- da Rosa, A., Campello, R. J., and 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., and Amaral, W. C. (2008). An optimal expansion of volterra models using independent kautz basis for each kernel dimension. *International Journal of Control*, 81(6):962–975.
- da Rosa, A., Campello, R. J. G. B., and Amaral, W. C. (2009). Exact search directions for optimization of linear and nonlinear models based on generalized orthonormal functions. *IEEE Transactions on Automatic Control*, 54(12):2757–2772.
- da Silva, I. N. (1995). *Estimação Paramétrica Robusta Através de Redes Neurais Artificiais*. Tese de doutorado, Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas (UNICAMP), Campinas, Brasil.

- Doyle III, F. J., Pearson, R. K., and Ogunnaike, B. A. (2002). *Identification and Control Using Volterra Models*. Springer-Verlag, Berlin.
- Dumont, G. A. and Fu, Y. (1993). Non-linear adaptive control via laguerre expansion of volterra kernels. *International Journal of Adaptive Control and Signal Processing*, 7(5):367 – 382.
- Eykhoﬀ, P. (1974). *System Identification: Parameter and State Estimation*. John Wiley & Sons, New York.
- Fogel, E. and Huang, Y. F. (1982). On the value of information in system identification–bounded noise case. *Automatica*, 18(2):229–238.
- Fu, Y. and Dumont, G. (1993). An optimum time scale for discrete laguerre network. *IEEE Transactions on Automatic Control*, 38(6):934 –938.
- Heuberger, P. S. C., Van den Hof, P. M. J., and Bosgra, O. H. (1995). A generalized orthonormal basis for linear dynamical systems. *IEEE Transactions on Automatic Control*, 40(3):451–465.
- Kibangou, A. Y., Favier, G., and Hassani, M. M. (2005). Selection of generalized orthormal 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. *Quartely of Applied Mathematics*, 2(2):164–168.
- Lin, J. (1970). Determination of reachable set for a linear discrete system. *IEEE Transactions on Automatic Control*, 15(3):339–342.
- Ljung, L. (1987). *System identification — Theory for the user*. Prentice Hall, Englewood Cliff, NJ.
- Machado, J. B. (2011). *Modelagem de sistemas não-lineares por base de funções ortonormais generalizadas com funções internas*. Tese de doutorado, Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas (UNICAMP), Campinas-SP, Brasil.
- Marquardt, D. W. (1963). An algorithm for the least-squares estimation of nonlinear parameters. *SIAM Journal of Applied Mathematics*, 11(2):431–441.
- Mbarek, A., Messaoud, H., and Favier, G. (2003). Robust predictive control using kautz model. In *ICECS 2003. Proceedings of the 2003 10th IEEE International Conference on Electronics, Circuits and Systems, 2003.*, volume 1, pages 184 – 187.
- Milanese, M. and Belforte, G. (1982). Estimation theory and uncertainty intervals evaluation in presence of unknown but bounded errors: Linear families of models and estimators. *IEEE Transactions on Automatic Control*, 27(2):408–414.
- Milanese, M. and Tempo, R. (1985). Optimal algorithms theory for robust estimation and prediction. *IEEE Transactions on Automatic Control*, 30(8):730–738.
- Mo, S. H. and Norton, J. P. (1990). Fast and robust algorithm to compute exact polytope parameter bounds. *Mathematics and Computers in Simulation*, 32(5-6):481 – 493.

- 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.
- Nalbantoglu, V., Bokor, J., Balas, G., and Gaspar, P. (2003). System identification with generalized orthonormal basis functions: an application to flexible structures. *Automatica*, 11(3):245–259.
- Nelles, O. (2001). *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*. Springer-Verlag, Berlin.
- Ninness, B. and Gustafsson, F. (1997). A unifying construction of orthonormal bases for systems identification. *IEEE Transactions on Automatic Control*, 42(4):515–521.
- Nocedal, J. and Wright, S. J. (1999). *Numerical Optimization*. Springer-Verlag, New York.
- Ogata, K. (1987). *Discrete-time control systems*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- Ogunfunmi, T. (2007). *Adaptive Nonlinear System Identification*. Springer, New York.
- Oliveira, G. H., Amaral, W. C., Favier, G., and Dumont, G. A. (2000). Constrained robust predictive controller for uncertain processes modeled by orthonormal series functions. *Automatica*, 36(4):563–571.
- Oliveira, G. H. C. (1997). *Controle Preditivo para Processos com Incertezas Estruturadas baseado em Séries de Funções Ortonormais*. Tese de doutorado, Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas (UNICAMP), Campinas-SP, Brasil.
- Oliveira e Silva, T. (1994). Optimality conditions for truncated laguerre networks. *Signal Processing, IEEE Transactions on*, 42(9):2528–2530.
- Oliveira e Silva, T. (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. (1995b). Optimality conditions for truncated kautz networks with two periodically repeating complex conjugate poles. *Automatic Control, IEEE Transactions on*, 40(2):342–346.
- Rugh, W. J. (1981). *Nonlinear Systems Theory - The Volterra/Wiener Approach*. The Johns Hopkins University Press, Baltimore.
- Schetzen, M. (1980). *The Volterra and Wiener Theories of Nonlinear Systems*. Robert Krieger Publishing Company.
- Schweppe, F. (1968). Recursive state estimation: Unknown but bounded errors and system inputs. *IEEE Transactions on Automatic Control*, 13(1):22–28.
- Schweppe, F. C. (1973). *Uncertain Dynamic Systems*. Prentice-Hall, NJ.

- Takenaka, S. (1925). On the orthogonal functions and a new formula of interpolation. *Japanese Journal of Mathematics (II)*, 2:129–145.
- Tanguy, N., Morvan, R., Vilbe, P., and Calvez, L.-C. (2002). Pertinent choice of parameters for discrete kautz approximation. *IEEE Transactions on Automatic Control*, 47(5):783–787.
- Tripathi, A. N. (1998). *Linear Systems Analysis*. New Age International Publishers.
- Van den Hof, P. M. J., Heuberger, P. S., and Bokor, J. (1995). System identification with generalized orthonormal basis functions. *Automatica*, 31(12):1821–1834.
- Vicino, A. and Milanese, M. (1989). Optimal inner bounds of feasible parameter set in linear estimation with bounded noise. In *Proceedings of the 28th IEEE Conference on Decision and Control, 1989*, volume 3, pages 2576–2580.
- Volterra, V. (1959). *Theory of Functionals and of Integral and Integro-Differential Equations*. Dover Publications, Inc., New York.
- 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. and Mäkilä, P. M. (1996). Approximation of stable linear dynamical systems using laguerre e kautz functions. *Automatica*, 32(5):693–708.
- Walter, E. and Piet-Lahanier, H. (1990). Estimation of parameter bounds from bounded-error data: a survey. *Mathematics and Computers in Simulation*, 32(5 - 6):449–468.
- Wiener, N. (1942). Response of a nonlinear device to noise. Technical Report 129, Radiation Laboratory, M.I.T.
- Wiener, N. (1958). *Nonlinear Problems in Random Theory*. The MIT Press, New York.
- Witsenhausen, H. (1968). Sets of possible states of linear systems given perturbed observations. *IEEE Transactions on Automatic Control*, 13(13):556–558.
- Xue, Y.-C., Yang, Q.-W., and Qian, J.-X. (2002). Improved robust minmax estimation algorithm with moving time window. In *Machine Learning and Cybernetics, 2002. Proceedings. 2002 International Conference on*, volume 3, pages 1529–1532 vol.3.
- Zadeh, L. (1962). From circuit theory to system theory. *Proceedings of the IRE*, 50(5):856–865.
- Zadeh, L. A. (1953). A contribution to the theory of nonlinear systems. *Journal of the Franklin Institute*, 255:387–408.
- Zhou, K. and Doyle, J. C. (1999). *Essentials of Robust Control*. Prentice Hall, New York.
- Ziaei, K. and Wang, D. W. L. (2006). Applications of ortonormal basis functions for identification of flexible-link manipulators. *Control Engineering Prattice*, 14(2):99–106.

Apêndice A

Equações dos gradientes das saídas dos filtros ortonormais

Este apêndice apresenta as equações das saídas dos filtros ortonormais conforme desenvolvidas em Machado (2011) para a Base de Kautz e em da Rosa et al. (2009) para a Base de Funções Ortonormais Generalizadas - GOBF.

A.1 Funções de Kautz

Assim como feito no Capítulo 3, Seção 3.2.1, para a base de Kautz adota-se o vetor \mathbf{p}_l (3.18) contendo os parâmetros reais que descrevem os polos de Kautz, ao invés de utilizar a sua forma complexa.

Utilizando o modelo das funções de Kautz em espaço de estados, como descrito na equação (2.31), então as saídas dos filtros $\check{\psi}_{l,i_l}(k)$ são descritas por:

$$\begin{aligned}\mathcal{X}(k+1) &= \mathcal{A}_K \mathcal{X}(k) + \mathcal{B}_K \mathcal{U}(k) \\ \check{\Psi}_l(k) &= \mathcal{C}_{K_m} \mathcal{X}(k)\end{aligned}\tag{A.1}$$

onde $\mathcal{X}(k)$, $\mathcal{U}(k)$, \mathcal{A}_K e \mathcal{B}_K são as mesmas matrizes definidas na Seção 2.3.3 para as funções de Kautz e \mathcal{C}_{K_m} e $\check{\Psi}_l(k)$ são aquelas definidas Seção 3.2.1.

A solução de uma equação em espaço de estados, como a apresentada em (A.1), é bem conhecida na literatura, sendo definida para $\check{\Psi}_l(k)$ como (Ogata, 1987; Tripathi, 1998):

$$\check{\Psi}_l(k) = \mathcal{C}_{K_m} \mathcal{A}_K^k \mathcal{X}(0) + \sum_{i=0}^{k-1} \mathcal{C}_{K_m} \mathcal{A}_K^i \mathcal{B}_K \mathcal{U}(k-1-i).\tag{A.2}$$

O gradiente de $\nabla_{\mathbf{p}_l} \check{\Psi}_l(k)$ é então calculado tomando as derivadas da equação (A.2) com relação aos parâmetros b_l e c_l que parametrizam as bases de Kautz, ou seja, $\nabla_{\mathbf{p}_l} \check{\Psi}_l(k) = [\partial \check{\Psi}_l(k) / \partial b_l \quad \partial \check{\Psi}_l(k) / \partial c_l]^T$, cuja derivada com relação ao parâmetro b_l é dada por (Machado, 2011):

$$\frac{\partial \check{\Psi}_l(k)}{\partial b_l} = \mathcal{C}_{K_m} \frac{\partial(\mathcal{A}_K^k)}{\partial b_l} \mathcal{X}(0) + \mathcal{C}_{K_m} \sum_{i=0}^{k-1} \left[\frac{\partial(\mathcal{A}_K^i)}{\partial b_l} \mathcal{B}_K + \mathcal{A}_K^i \frac{\partial \mathcal{B}_K}{\partial b_l} \right] \mathcal{U}(k-1-i) \quad (\text{A.3})$$

sendo $\partial \mathcal{A}_K^k / \partial b_l$ dado por (da Rosa, 2009; Machado, 2011):

$$\frac{\partial(\mathcal{A}^k)}{\partial b_l} = \sum_{j=1}^k \mathcal{A}^{j-1} \frac{\partial \mathcal{A}}{\partial b_l} \mathcal{A}^{k-j} \quad (\text{A.4})$$

onde as derivadas das matrizes \mathcal{A}_K e \mathcal{B}_K , ou seja, $\partial \mathcal{A}_K / \partial b_l$ e $\partial \mathcal{B}_K / \partial b_l$, são as definidas no Capítulo 3 na Seção 3.2.1.

A derivada de $\check{\Psi}_l(k)$ com relação ao parâmetro c_l é descrita pela seguinte equação:

$$\frac{\partial \check{\Psi}_l(k)}{\partial c_l} = \mathcal{C}_{K_m} \frac{\partial(\mathcal{A}_K^k)}{\partial c_l} \mathcal{X}(0) + \mathcal{C}_{K_m} \sum_{i=0}^{k-1} \left[\frac{\partial(\mathcal{A}_K^i)}{\partial c_l} \mathcal{B}_K + \mathcal{A}_K^i \frac{\partial \mathcal{B}_K}{\partial c_l} \right] \mathcal{U}(k-1-i) \quad (\text{A.5})$$

sendo $\partial \mathcal{A}_K^k / \partial c_l$ dado por (da Rosa, 2009; Machado, 2011):

$$\frac{\partial(\mathcal{A}^k)}{\partial c_l} = \sum_{j=1}^k \mathcal{A}^{j-1} \frac{\partial \mathcal{A}}{\partial c_l} \mathcal{A}^{k-j}. \quad (\text{A.6})$$

As derivadas das matrizes \mathcal{A}_K e \mathcal{B}_K com relação ao parâmetro c_l são as mesmas apresentadas no Capítulo 3 na Seção 3.2.1.

O fato é que as equações (3.20) e (3.21) ou (3.24) e (3.25) do Capítulo 3 geram o mesmo resultado que as equações (A.3) e (A.5), respectivamente, bastando agrupar aquelas numa única equação e proceder o cálculo até o instante $k = 0$.

A.2 Funções Ortonormais Generalizadas (GOBF)

O processo de desenvolvimento é semelhante àquele apresentado no Capítulo 3, Seção 3.2.2. Primeiro, o desenvolvimento para o caso onde somente polos reais são utilizados, depois para o caso onde apenas um par complexo parametriza a base e, por fim, quando dois pares de polos complexos são utilizados.

A.2.1 Caso GOBF com polos reais

Partindo da equação (3.31), cuja solução é dada por:

$$\check{\psi}_{l,1}(k) = p_{l,1}^k \check{\psi}_{l,1}(0) + \sqrt{1 - p_{l,1}^2} \sum_{i=0}^{k-1} p_{l,1}^i u(k-i) \quad (\text{A.7})$$

a derivada da saída do primeiro filtro com relação ao polo $p_{l,1}$ é dada por:

$$\frac{\partial \check{\psi}_{l,1}(k)}{\partial p_{l,1}} = kp_{l,1}^{k-1} \check{\psi}_{l,1}(0) + \frac{1}{\sqrt{1-p_{l,1}^2}} \sum_{i=0}^{k-1} [ip_{l,1}^{i-1} - (i+1)p_{l,1}^{i+1}] u(k-i). \quad (\text{A.8})$$

Como o polo $p_{l,1}$ parametriza apenas a base $\check{\psi}_{l,1}(k)$, então $\partial \check{\psi}_{l,1}(k)/\partial p_{l,k} = 0$ para $k = 2, \dots, M$. Para os demais filtros, utiliza-se a equação (3.34) que apresenta como solução para $\check{\psi}_{l,m}(k)$:

$$\check{\psi}_{l,m}(k) = p_{l,m}^k \check{\psi}_{l,m}(0) + \sqrt{\frac{1-p_{l,m}^2}{1-p_{l,m-1}^2}} \sum_{i=0}^{k-1} p_{l,m}^i [\check{\psi}_{l,m-1}(k-1-i) - p_{l,m-1} \check{\psi}_{l,m-1}(k-i)]. \quad (\text{A.9})$$

E as derivadas das saídas dos demais filtros ortonormais são dadas por (da Rosa et al., 2009):

$$\begin{aligned} \frac{\partial \check{\psi}_{l,m}(k)}{\partial p_{l,m}} &= kp_{l,m}^{k-1} \check{\psi}_{l,m}(0) + \frac{1}{\sqrt{(1-p_{l,m}^2)(1-p_{l,m-1}^2)}} \cdot \\ &\cdot \sum_{i=0}^{k-1} [ip_{l,m}^{i-1} - (i+1)p_{l,m}^{i+1}] \cdot [\check{\psi}_{l,m-1}(k-1-i) - p_{l,m-1} \check{\psi}_{l,m-1}(k-i)] \end{aligned} \quad (\text{A.10})$$

$$\begin{aligned} \frac{\partial \check{\psi}_{l,m}(k)}{\partial p_{l,m-1}} &= \sum_{i=0}^{k-1} \left[\sqrt{\frac{1-p_{l,m}^2}{(1-p_{l,m-1}^2)^3}} p_{l,m-1} p_{l,m}^i (\check{\psi}_{l,m-1}(k-1-i) - p_{l,m-1} \check{\psi}_{l,m-1}(k-i)) \right. \\ &+ \sqrt{\frac{1-p_{l,m}^2}{1-p_{l,m-1}^2}} p_{l,m}^i \left(\frac{\partial}{\partial p_{l,m-1}} \check{\psi}_{l,m-1}(k-1-i) - \check{\psi}_{l,m-1}(k-1-i) \right. \\ &\left. \left. - \frac{\partial}{\partial p_{l,m-1}} \check{\psi}_{l,m-1}(k-1-i) \right) \right] \end{aligned} \quad (\text{A.11})$$

$$\begin{aligned} \frac{\partial \check{\psi}_{l,m}(k)}{\partial p_{l,k}} &= \sqrt{\frac{1-p_{l,m}^2}{1-p_{l,m-1}^2}} p_{l,m} \sum_{i=0}^{k-1} p_{l,m}^i \left[\frac{\partial}{\partial p_{l,k}} \check{\psi}_{l,m-1}(k-1-i) \right. \\ &\left. - p_{l,m-1} \frac{\partial}{\partial p_{l,k}} \check{\psi}_{l,m-1}(k-i) \right]. \end{aligned} \quad (\text{A.12})$$

A.2.2 Caso GOBF com um par de polos complexos

Para o caso em que a base possui um polo complexo, o desenvolvimento das equações das derivadas parte da equação (3.49) cuja solução para $\check{\psi}'_{l,M}(k)$ é:

$$\begin{aligned} \check{\psi}'_{l,M}(k) &= CA^k x'_{l,M}(0) + C \sum_{i=0}^{k-1} A^i \left[B_1 \check{\psi}_{l,M-1}(k+1-i) \right. \\ &\quad \left. + B_2 \check{\psi}_{l,M-1}(k-i) + B_3 \check{\psi}_{l,M-1}(k-1-i) \right] \end{aligned} \quad (\text{A.13})$$

onde $x'_{l,M}(0) = [\check{\psi}'_{l,M}(0) \ \check{\psi}'_{l,M}(1)]$ e as matrizes A, B_1, B_2, B_3, C são aquelas definidas no Capítulo 3, Seção 3.2.2.

Se, novamente, o polo $\beta_{l,M}$ for tratado como um vetor bidimensional real, ou seja, definindo $\beta_{l,m} = [\mathbf{Re}(\beta_{l,M}) \ \mathbf{Im}(\beta_{l,M})]^T$, o gradiente da função ortonormal passa a ser definido como $\nabla_{\beta_{l,M}} \check{\psi}'_{l,M}(k) := [\partial \check{\psi}'_{l,M}(k) / \partial \mathbf{Re}(\beta_{l,M}) \ \partial \check{\psi}'_{l,M}(k) / \partial \mathbf{Im}(\beta_{l,M})]^T$, cujo primeiro elemento (derivada com relação a parte real) é definido pela equação (A.14) (da Rosa et al., 2009):

$$\begin{aligned} \frac{\partial \check{\psi}'_{l,M}(k)}{\partial \mathbf{Re}(\beta_{l,M})} = & C \frac{\partial(A^k)}{\partial \mathbf{Re}(\beta_{l,M})} x_{l,M}(0) + C \sum_{i=0}^{k-1} \frac{\partial(A^k)}{\partial \mathbf{Re}(\beta_{l,M})} [B_1 \check{\psi}_{l,M-1}(k+1-i) \\ & + B_2 \check{\psi}_{l,M-1}(k-i) + B_3 \check{\psi}_{l,M-1}(k+1-i)] + A^i \left[\frac{\partial B_1}{\partial \mathbf{Re}(\beta_{l,M})} \check{\psi}_{l,M-1}(k+1-i) \right. \\ & \left. + \frac{\partial B_2}{\partial \mathbf{Re}(\beta_{l,M})} \check{\psi}_{l,M-1}(k-i) + \frac{\partial B_3}{\partial \mathbf{Re}(\beta_{l,M})} \check{\psi}_{l,M-1}(k-1-i) \right]. \end{aligned} \quad (\text{A.14})$$

A derivada com relação à parte imaginária do polo $\beta_{l,M}$ é obtida da equação (A.14), substituindo-se $\mathbf{Re}(\beta_{l,M})$ por $\mathbf{Im}(\beta_{l,M})$.

Finalmente, as derivadas com relação a $p_{l,k}$ ($k = 1, \dots, M-1$), polos reais, podem ser calculadas usando a seguinte equação (da Rosa et al., 2009):

$$\begin{aligned} \frac{\partial \check{\psi}'_{l,M}(k)}{\partial p_{l,k}} = & C \sum_{i=0}^{k-1} A^i \left[B_1 \frac{\partial \check{\psi}_{l,M-1}(k+1-i)}{\partial p_{l,k}} + \frac{\partial B_1}{\partial p_{l,k}} \check{\psi}_{l,M-1}(k+1-i) \right. \\ & + B_2 \frac{\partial \check{\psi}_{l,M-1}(k-i)}{\partial p_{l,k}} + \frac{\partial B_2}{\partial p_{l,k}} \check{\psi}_{l,M-1}(k-i) \\ & \left. + B_3 \frac{\partial \check{\psi}_{l,M-1}(k-1-i)}{\partial p_{l,k}} + \frac{\partial B_3}{\partial p_{l,k}} \check{\psi}_{l,M-1}(k-1-i) \right]. \end{aligned} \quad (\text{A.15})$$

As derivadas complementares de $\check{\Psi}''_{l,M}(z)$ são obtidas por um desenvolvimento matemático análogo ao apresentado, bastando substituir λ' e γ' por λ'' e γ'' , respectivamente nas equações.

A.2.3 Caso GOBF com dois pares de polos complexos

Neste caso, a equação-base é a (3.60), a qual incorpora dois polos complexos à base ortonormal. Essa equação apresenta como solução para $\check{\psi}_{l,M+1}(k)$:

$$\begin{aligned} \check{\psi}'_{l,M+1}(k) = & C_1 A_1^k x'_{l,M+1}(0) + C_1 \sum_{i=0}^{k-1} [B_4 \check{\psi}'_{l,M}(k+2-i) + B_5 \check{\psi}'_{l,M}(k+1-i) \\ & + B_6 \check{\psi}'_{l,M}(k-i) + B_7 \check{\psi}'_{l,M}(k-1-i)] \end{aligned} \quad (\text{A.16})$$

onde $x'_{l,M+1}(0) = [\check{\psi}'_{l,M+1}(0) \ \check{\psi}'_{l,M+1}(1) \ \check{\psi}'_{l,M+1}(2)]^T$ e as matrizes A_1, B_4, B_5, B_6, B_7 e C_1 já foram definidas no Capítulo 3, Seção 3.2.2.

Portanto, os elementos do gradiente $\nabla_{\mathbf{p}_l} \check{\psi}'_{l,M+1}(k)$ são:

$$\frac{\partial \check{\psi}'_{l,M+1}(k)}{\partial p_{l,k}} = C_1 \sum_{i=0}^{k-1} A_1^i \left[B_4 \frac{\partial \check{\psi}_{l,M}(k+2-i)}{\partial p_{l,k}} + B_5 \frac{\partial \check{\psi}_{l,M}(k+1-i)}{\partial p_{l,k}} + B_6 \frac{\partial \check{\psi}_{l,M}(k-i)}{\partial p_{l,k}} + B_7 \frac{\partial \check{\psi}_{l,M}(k-1-i)}{\partial p_{l,k}} \right] \quad k = 1, \dots, M-1 \quad (\text{A.17})$$

$$\begin{aligned} \frac{\partial \check{\psi}'_{l,M+1}(k)}{\partial \text{Re}(\beta_{l,M})} &= C_1 \sum_{i=0}^{k-1} \left\{ \frac{\partial(A_1^i)}{\partial \text{Re}(\beta_{l,M})} [B_4 \check{\psi}'_{l,M}(k-1-i) \right. \\ &+ B_5 \check{\psi}'_{l,M}(k+1-i) + B_6 \check{\psi}'_{l,M}(k-i) + B_7 \check{\psi}'_{l,M}(k-1-i)] \\ &+ A_1^i \left[\frac{\partial B_4}{\partial \text{Re}(\beta_{l,M})} \check{\psi}'_{l,M}(k+2-i) + B_4 \frac{\check{\psi}'_{l,M}(k+2-i)}{\text{Re}(\beta_{l,M})} \right. \\ &+ \frac{\partial B_5}{\partial \text{Re}(\beta_{l,M})} \check{\psi}'_{l,M}(k+1-i) + B_5 \frac{\check{\psi}'_{l,M}(k+1-i)}{\text{Re}(\beta_{l,M})} \\ &+ \frac{\partial B_6}{\partial \text{Re}(\beta_{l,M})} \check{\psi}'_{l,M}(k-i) + B_6 \frac{\check{\psi}'_{l,M}(k-i)}{\text{Re}(\beta_{l,M})} \\ &\left. \left. + \frac{\partial B_7}{\partial \text{Re}(\beta_{l,M})} \check{\psi}'_{l,M}(k-1-i) + B_7 \frac{\check{\psi}'_{l,M}(k-1-i)}{\text{Re}(\beta_{l,M})} \right] \right\} \quad (\text{A.18}) \end{aligned}$$

$$\begin{aligned} \frac{\partial \check{\psi}'_{l,M+1}(k)}{\partial \text{Re}(\beta_{l,M+1})} &= C_1 \frac{\partial(A_1^k)}{\partial \text{Re}(\beta_{l,M+1})} x'_{l,M+1}(0) + C_1 \sum_{i=0}^{k-1} \left\{ \frac{\partial(A_1^i)}{\partial \text{Re}(\beta_{l,M+1})} [B_4 \check{\psi}'_{l,M}(k+2-i) \right. \\ &+ B_5 \check{\psi}'_{l,M}(k+1-i) + B_6 \check{\psi}'_{l,M}(k-i) + B_7 \check{\psi}'_{l,M}(k-1-i)] \\ &+ A_1^i \left[\frac{\partial B_4}{\partial \text{Re}(\beta_{l,M+1})} \check{\psi}'_{l,M}(k+2-i) + \frac{\partial B_5}{\partial \text{Re}(\beta_{l,M+1})} \check{\psi}'_{l,M}(k+1-i) \right. \\ &\left. \left. + \frac{\partial B_6}{\partial \text{Re}(\beta_{l,M+1})} \check{\psi}'_{l,M}(k-i) + \frac{\partial B_7}{\partial \text{Re}(\beta_{l,M+1})} \check{\psi}'_{l,M}(k-1-i) \right] \right\}. \quad (\text{A.19}) \end{aligned}$$

onde $x'_{l,M+1}(0) = [\check{\psi}'_{l,M+1}(0) \ \check{\psi}'_{l,M+1}(1) \ \check{\psi}'_{l,M+1}(2)]$ e as matrizes $A_1, B_4, B_5, B_6, B_7, C_1$ são aquelas definidas no Capítulo 3, Seção 3.2.2.

As derivadas com relação às partes imaginárias de $\beta_{l,M}$ e $\beta_{l,M+1}$ são obtidas pela substituição direta de $\text{Re}(\beta_{l,M})$ e $\text{Re}(\beta_{l,M+1})$ por $\text{Im}(\beta_{l,M})$ e $\text{Im}(\beta_{l,M+1})$ em (A.18) e (A.19). As derivadas das funções complementares $\check{\Psi}''_{l,M+1}(k)$ são obtidas procedendo a respectiva substituição das variáveis ρ' e μ' por ρ'' e μ'' .

Apêndice B

Cálculo dos Limitantes de Incertezas em Modelos OBF-Volterra

Neste apêndice, apresentam-se as formulações para o cálculo dos limitantes de incertezas em modelos dinâmicos baseados em funções ortonormais, conforme desenvolvidas em Moreira (2006), para sistemas lineares, e sua extensão em da Rosa (2009) e da Rosa et al. (2010), para sistemas não-lineares.

B.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, o qual é apresentado a seguir.

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

O critério de custo do problema (B.1) é o volume do hiper-retângulo formado pelo prolongamento da incerteza sobre os coeficientes da expansão ortonormal, ou seja, $\sigma_m \Delta \alpha_m$. As parcelas incertas dos coeficientes do modelo, $\Delta \alpha_m (m = 1, 2, \dots, M)$, são 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 variáveis do problema de otimização (B.1) são os coeficientes σ_m ($m = 1, 2, \dots, M$). 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. Assim:

$$\alpha_m^{min} \leq \alpha_m \leq \alpha_m^{max} \quad (\text{B.2})$$

com:

$$\begin{cases} \alpha_m^{min} = \bar{\alpha}_m - \sigma_{m,\acute{o}timo} \cdot \Delta\alpha_m \\ \alpha_m^{max} = \bar{\alpha}_m + \sigma_{m,\acute{o}timo} \cdot \Delta\alpha_m \end{cases} \quad (\text{B.3})$$

onde $\sigma_{m,\acute{o}timo}$ são as soluções do problema (B.1).

B.2 Modelos Não-Lineares

Em da Rosa (2009) e da Rosa et al. (2010), estende-se a ideia apresentada na seção anterior para modelos não-lineares.

Considera-se um modelo OBF-Volterra de ordem η , como aquele em (2.23). As incertezas sobre os coeficientes do modelo são representados pelos termos $\sigma_{i_1} \Delta\alpha_{i_1}$, e $\sigma_{i_1, \dots, i_\eta} \Delta\alpha_{i_1, \dots, i_\eta}$, 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 haja uma interdependência entre h_1, \dots, h_η , deve-se solucionar os seguintes η problemas de otimização separadamente, como a seguir:

Kernel de primeira ordem:

$$\begin{aligned} \min_{\sigma_{i_1}} \quad & \prod_{i_1=1}^M \sigma_{i_1} \Delta\alpha_{i_1} & (\text{B.4}) \\ \text{sujeito a} \quad & \{h_1(k_1)\} \subseteq \{\check{h}_1(k_1)\} \\ & \check{h}_1(k_1) = \sum_{i_1=1}^M \left[\bar{\alpha}_{i_1} + \sigma_{i_1} \tilde{\Delta}\alpha_{i_1} \right] \psi_{i_1}(k_1), \\ & |\tilde{\sigma}_{i_1}| \leq \sigma_{i_1}, \quad i_1 = 1, \dots, M \\ & \bar{\alpha}_{i_1} = \sum_{k_1=0}^{\infty} \bar{h}_1(k_1) \psi_{i_1}(k_1) \\ & \Delta\alpha_{i_1} = \sum_{k_1=0}^{\infty} \Delta h_1(k_1) \psi_m(k) \end{aligned}$$

Kernel de η -ésima ordem:

$$\begin{aligned}
 & \min_{\sigma_{i_1, \dots, i_\eta}} \prod_{i_1=1}^M \cdots \prod_{i_\eta=1}^M \sigma_{i_1, \dots, i_\eta} \cdot \Delta \alpha_{i_1, \dots, i_\eta} \\
 & \text{sujeito a} \quad \{h_\eta(k_1, \dots, k_\eta)\} \subseteq \{\check{h}_\eta(k_1, \dots, k_\eta)\} \\
 & \check{h}_\eta(k_1, \dots, k_\eta) = \sum_{i_1=1}^M \cdots \sum_{i_\eta=1}^M [\bar{\alpha}_{i_1, \dots, i_\eta} + \tilde{\sigma}_{i_1, \dots, i_\eta} \Delta \alpha_{i_1, \dots, i_\eta}] \prod_{l=1}^{\eta} \psi_{i_l}(k_l), \\
 & |\tilde{\sigma}_{i_1, \dots, i_\eta}| \leq \sigma_{i_1, \dots, i_\eta}, \quad i_l = 1, \dots, M \\
 & \bar{\alpha}_{i_1, \dots, i_\eta} = \sum_{k_1=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} \bar{h}_\eta(k_1, \dots, k_\eta) \prod_{l=1}^{\eta} \psi_{i_l}(k_l) \\
 & \Delta \alpha_{i_1, \dots, i_\eta} = \sum_{k_1=0}^{\infty} \cdots \sum_{k_\eta=0}^{\infty} \Delta h_\eta(k_1, \dots, k_\eta) \prod_{l=1}^{\eta} \psi_{i_l}(k_l)
 \end{aligned} \tag{B.5}$$

Assim, como no caso linear, a interpretação para a solução dos problemas acima é minimizar os volumes gerados pelo escalonamento das incertezas sobre os coeficientes $\alpha_{(\cdot)}$ do modelo ortonormal. Como assume-se que as 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 $\{\check{h}_l(k_1, \dots, k_l)\}$.

As variáveis dos problemas de otimização (B.4) e (B.5) são os coeficientes σ_{i_1} e $\sigma_{i_1, \dots, i_\eta}$, respectivamente. No 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.

Apêndice C

Trabalhos Publicados

O seguinte trabalho foi publicado durante o mestrado:

- Márcio Feliciano Braga, Jeremias Barbosa Machado, Ricardo José Gabrielli Barreto Campello e Wagner Caradori do Amaral. Optimization of Volterra Models with Assymetrical Kernels based on Generalized Orthonormal Functions. 19th Mediterranean Conference on Control and Automation, Corfu, Grécia, 2011.