

**IDENTIFICAÇÃO DE SISTEMAS NÃO LINEARES
UTILIZANDO MODELOS NARX, FUNÇÕES ORTONORMAIS
E OTIMIZAÇÃO HEURÍSTICA**

ELDER OROSKI

**TESE DE DOUTORADO EM ENGENHARIA DE SISTEMAS
ELETRÔNICOS E AUTOMAÇÃO
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**FACULDADE DE TECNOLOGIA
UNIVERSIDADE DE BRASÍLIA**

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**IDENTIFICAÇÃO DE SISTEMAS NÃO LINEARES
UTILIZANDO MODELOS NARX, FUNÇÕES ORTONORMAIS
E OTIMIZAÇÃO HEURÍSTICA**

ELDER OROSKI

TESE DE DOUTORADO ACADÊMICO SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA ELÉTRICA DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM ENGENHARIA DE SISTEMAS ELETRÔNICOS E AUTOMAÇÃO.

APROVADA POR:

**Prof. Adolfo Bauchspiess, Dr., PGEA/UnB
(Orientador)**

**Prof. Alex da Rosa, Dr., PGEA/UnB
Examinador Interno**

**Prof. Wesley Calixto Pacheco, Dr., PGEA/IFG
Examinador Interno**

**Prof. Gustavo Henrique da Costa Oliveira, Dr., UFPR
Examinador Externo**

**Prof. André Laurindo Maitelli, Dr., UFRN
Examinador Externo**

BRASÍLIA, 20 DE NOVEMBRO DE 2015.

FICHA CATALOGRÁFICA

OROSKI, ELDER

Identificação de Sistemas Não Lineares Utilizando Modelos NARX, Funções Ortonormais e Otimização Heurística [Distrito Federal] 2015.

xi, 124p., 210 x 297 mm (ENE/FT/UnB, Doutor, Engenharia Elétrica, 2015).

Tese de Doutorado – Universidade de Brasília, Faculdade de Tecnologia.

Departamento de Engenharia Elétrica

1. Identificação de Sistemas

2. NARX e Séries de Volterra

3. Funções Ortonormais

4. Otimização Heurística

I. ENE/FT/UnB

II. Título (série)

REFERÊNCIA BIBLIOGRÁFICA

OROSKI, E. (2015). Identificação de Sistemas Não Lineares Utilizando Modelos NARX, Funções Ortonormais e Otimização Heurística, Tese de Doutorado em Engenharia de Sistemas Eletrônicos e Automação, Publicação PGEA.DM-104/15 TD, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 124p.

CESSÃO DE DIREITOS

AUTOR: Elder Oroski

TÍTULO: Identificação de Sistemas Não Lineares Utilizando Modelos NARX, Funções Ortonormais e Otimização Heurística.

GRAU: Doutor

ANO: 2015

É concedida à Universidade de Brasília permissão para reproduzir cópias desta tese de doutorado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte dessa tese de doutorado pode ser reproduzida sem autorização por escrito do autor.

Elder Oroski

Departamento de Eng. Elétrica (ENE) - FT

Universidade de Brasília (UnB)

Campus Darcy Ribeiro

CEP 70919-970 - BRASÍLIA - DF - Brasil

*Dedico este trabalho aos meus pais:
João J. Oroski e Dinalva S. L. Oroski
e principalmente a minha esposa Be-
atriz Pês.*

AGRADECIMENTOS

Agradeço à minha querida esposa, Beatriz Pês, pelo constante apoio, nos bons e maus momentos, e inúmeras revisões deste texto. Ao meu orientador, Prof. Adolfo Bauchspiess, pelo suporte e orientação. Ao amigo Rafael Houdorf Lopes pelo auxílio na confecção desta tese e por discussões produtivas.

Agradeço aos meus pais, João Joanny Oroski e Dinalva Silva Lima Oroski, por sempre me incentivarem a estudar e por terem me motivado a sempre continuar neste processo.

Agradeço ao CINDACTA 1, meu ambiente de trabalho. Agradeço ainda, ao colega Rafael Rodrigues Luz Benevidez pela demonstração de algumas facetas da Matemática que ainda me eram desconhecidas.

RESUMO

IDENTIFICAÇÃO DE SISTEMAS NÃO LINEARES UTILIZANDO MODELOS NARX, FUNÇÕES ORTONORMAIS E OTIMIZAÇÃO HEURÍSTICA

Autor: Elder Oroski

Orientador: Prof. Adolfo Bauchspiess, Dr., PGEA/UnB

Programa de Pós-graduação em Engenharia de Sistemas Eletrônicos e Automação

BRASÍLIA, 20 de novembro de 2015

Identificar um sistema não linear pode constituir uma tarefa complexa. Para tratar tal complexidade uma série de modelos e métodos de identificação de sistemas foram desenvolvidos ao longo das últimas décadas. Neste cenário, dois dos modelos não lineares mais promissores são as chamadas séries de Volterra e os modelos NARX que, embora sejam implementações computacionalmente exigentes, apresentam boa precisão e representatividade do sistema a ser modelado. Tais modelos podem ter seu custo de implementação computacional reduzido pelo uso de funções ortonormais. Assim, nesta tese, tem-se por objetivo identificar um conjunto de sistemas não lineares a partir de modelos NARX, simplificados através do uso de funções ortonormais, conhecidos como modelos NARX-OBF. Para tanto, são expostos métodos de otimização determinísticos e heurísticos, com foco na determinação dos parâmetros de modelo. Foram criados métodos de otimização heurística, como o inédito Algoritmo do Princípio Antrópico, que são utilizados na determinação da estrutura do modelo bem como de seus parâmetros. São apresentados ainda exemplos de identificação de sistemas não lineares reais, dois levitadores magnéticos, baseados nestas propostas. A partir da comparação entre os modelos citados, foi constatado que o modelo NARX-OBF apresentou menor MSE que o modelo Volterra-OBF, para modelos com o mesmo número de termos, superando também os modelos clássicos de identificação de sistemas neste quesito. Como perspectivas futuras para este trabalho, tem-se a possibilidade de utilização de Programação Genética na determinação da estrutura do modelo e a exploração de diferentes funções ortonormais.

ABSTRACT

NONLINEAR SYSTEM IDENTIFICATION USING NARX MODELS, ORTHONORMAL FUNCTIONS AND HEURISTIC OPTIMIZATION

Author: Elder Oroski

Supervisor: Prof. Adolfo Bauchspiess, Dr., PGEA/UnB

Programa de Pós-graduação em Engenharia de Sistemas Eletrônicos e Automação

Identifying a nonlinear system generally is a complex task. In order to manage this complexity many system identification methods were developed over the past decades. Among them, two of the most promising models are the Volterra series and the NARX series. Their implementation normally are computationally costly, but they have good accuracy and representativeness. These kind of models have their identification cost reduced by using orthonormal functions. Thus, this thesis has the objective of identifying a class of nonlinear systems, with Volterra and NARX models, simplified by orthonormal functions. In order to achieve this goal, some optimization methods are presented, deterministic and heuristic, focusing on model parameter determination. Heuristic optimization methods were created, as the novel Antropic Principle Algorithm, which are used in determining the model structure and their parameters. Examples of real nonlinear system identification is presented, as two magnetic levitators, based on these proposals. From the comparison of the above models, it was found that the NARX-OBF model showed lower MSE than the Volterra-OBF model, to models with the same number of terms. As future work, is possible using Genetic Programming in the model structure determination and to explore different kind of orthonormal functions.

SUMÁRIO

1	INTRODUÇÃO	1
1.1	DEFINIÇÃO DO PROBLEMA	4
1.2	OBJETIVOS	4
1.2.1	OBJETIVO GERAL	4
1.2.2	OBJETIVOS ESPECÍFICOS	4
1.3	CONTRIBUIÇÕES DO TRABALHO	5
1.4	PRINCIPAIS DIFICULDADES	5
1.5	ORGANIZAÇÃO DA TESE DE DOUTORADO	6
2	CONCEITOS PRELIMINARES	7
2.1	SISTEMAS E MODELOS	7
2.2	IDENTIFICAÇÃO DE SISTEMAS	9
2.2.1	IDENTIFICAÇÃO CAIXA BRANCA	10
2.2.2	IDENTIFICAÇÃO CAIXA PRETA	10
2.2.3	IDENTIFICAÇÃO CAIXA CINZA	10
2.3	IDENTIFICAÇÃO POR DECONVOLUÇÃO	11
2.4	IDENTIFICAÇÃO BASEADA EM CORRELAÇÃO	12
2.5	MÉTODO DOS MÍNIMOS QUADRADOS	13
2.5.1	HIPÓTESES DE GAUSS-MARKOV	19
2.5.2	ORTOGONALIDADE ENTRE REGRESSORES E RESÍDUOS	20
2.6	O ESTIMADOR ESTENDIDO DE MÍNIMOS QUADRADOS	22
2.7	MÍNIMOS QUADRADOS ORTOGONAIS	24
2.7.1	MÉTODO DE GRAM-SCHMIDT	25
2.8	CONCLUSÃO	26
3	MODELOS NÃO LINEARES	27
3.1	SISTEMAS NÃO LINEARES	27
3.2	ESPAÇO DE HILBERT	28
3.3	PRINCIPAIS MODELOS NÃO LINEARES	29
3.4	MODELOS FIR E ARX	30
3.5	SÉRIES DE VOLTERRA	31
3.6	NARMAX	35
3.7	FUNÇÕES ORTONORMAIS	36
3.7.1	FUNÇÕES DE LAGUERRE	40
3.7.2	FUNÇÕES DE KAUTZ	41
3.8	MODELO VOLTERRA-ObF	42

3.9	MODELO NARX-OBF.....	44
3.9.1	MÉTODOS DE SELEÇÃO DE TERMOS DO MODELO NARX-OBF	48
3.10	CONCLUSÃO	50
4	MÉTODOS DE OTIMIZAÇÃO DETERMINÍSTICA	51
4.1	DEFINIÇÕES INICIAIS	51
4.2	EXPANSÃO DA FUNÇÃO DE AVALIAÇÃO	54
4.3	MECANISMOS DE BUSCA.....	55
4.3.1	BUSCA EM LINHA.....	56
4.3.2	REGIÃO DE CONFIANÇA	57
4.3.3	MÉTODO DE NEWTON.....	58
4.3.4	MÉTODO DE GAUSS NEWTON.....	60
4.3.5	ALGORITMO DE LEVENBERG-MARQUARDT	61
4.4	CONCLUSÃO	62
5	MÉTODOS DE OTIMIZAÇÃO HEURÍSTICA	63
5.1	ALGORITMO GENÉTICO	64
5.1.1	OPERADORES DO ALGORITMO GENÉTICO	65
5.2	ALGORITMO DO PRINCÍPIO ANTRÓPICO	71
5.2.1	PRINCÍPIO ANTRÓPICO.....	71
5.2.2	ESTRUTURA DO APA	73
5.2.3	PSEUDO CÓDIGO DO ALGORITMO DO PRINCÍPIO ANTRÓPICO	79
5.2.4	EXEMPLOS DE APLICAÇÃO	79
5.2.5	ANÁLISE ESTATÍSTICA DO APA	87
5.3	CONCLUSÃO	89
6	METODOLOGIA PROPOSTA PARA IDENTIFICAÇÃO DE SISTEMAS A PARTIR DE MODELOS ORTONORMAIS E ALGORITMOS HEU- RÍSTICOS	91
6.1	IDENTIFICAÇÃO DE SISTEMAS LINEARES COM AG E FUNÇÕES ORTO- NORMAIS	91
6.2	IDENTIFICAÇÃO DE SISTEMAS NÃO LINEARES, UTILIZANDO ALGORIT- MOS GENÉTICOS E FUNÇÕES ORTONORMAIS	92
6.3	MODELOS DE VOLTERRA-OBF UTILIZANDO ALGORITMOS GENÉTICOS....	95
6.4	MODELOS NARX-OBF UTILIZANDO O ALGORITMO DE LEVENBERG MAR- QUARDT E ALGORITMOS GENÉTICOS	96
6.5	SELEÇÃO DE TERMOS DE UM MODELO NARX-OBF	99
6.6	CONCLUSÃO	101
7	EXPERIMENTOS E VALIDAÇÃO DAS METODOLOGIAS PROPOS- TAS.....	103

7.1	MODELAGEM DE SISTEMAS LIT A PARTIR DE FUNÇÕES DE KAUTZ	103
7.1.1	MODELOS DE KAUTZ UTILIZANDO APA	104
7.1.2	MODELOS DE KAUTZ UTILIZANDO AG.....	105
7.1.3	COMPARAÇÕES ENTRE APA E AG NOS MODELOS DE KAUTZ.....	107
7.2	MODELAGEM DE UM LEVITADOR MAGNÉTICO VERTICAL A PARTIR DE MODELOS VOLTERRA-ObF	108
7.2.1	LEVITADOR MAGNÉTICO VERTICAL	108
7.2.2	COMPARAÇÃO DOS MODELOS VOLTERRA-ObF COM OS MODELOS POLINOMIAIS	113
7.3	MODELAGEM DE UM LEVITADOR MAGNÉTICO VERTICAL A PARTIR DE MODELOS NARX-ObF.....	116
7.3.1	COMPARAÇÃO DOS MODELOS NARX-ObF COM OS MODELOS PO- LINOMIAIS	124
7.3.2	COMPARAÇÃO ENTRE OS MODELOS NARX-ObF E VOLTERRA-ObF	124
7.4	IDENTIFICAÇÃO DE UM LEVITADOR MAGNÉTICO HORIZONTAL A PARTIR DE MODELOS NARX-ObF SIMPLIFICADOS	124
7.4.1	RESULTADOS OBTIDOS COM MODELOS NARX-ObF SIMPLIFICA- DOS	125
7.4.2	COMPARAÇÃO ENTRE OS MODELOS NARX-ObF E OUTROS MO- DELOS	131
7.4.3	COMPARAÇÃO ENTRE OS MODELOS NARX-ObF COMPLETOS E SIMPLIFICADOS	132
7.4.4	COMPARAÇÃO ENTRE O APA E O AG NA SIMPLIFICAÇÃO DE MO- DELOS NARX-ObF	132
7.5	CONCLUSÃO	132
8	CONCLUSÃO.....	135
8.1	REALIZAÇÕES DA TESE	135
8.2	PRINCIPAIS CONTRIBUIÇÕES	136
8.3	RECAPITULAÇÃO DE OBJETIVOS.....	136
8.3.1	OBJETIVO GERAL	136
8.3.2	OBJETIVOS ESPECÍFICOS	137
8.4	TRABALHOS FUTUROS	137
8.5	ARTIGOS.....	138
8.6	CONSIDERAÇÕES FINAIS	138
	REFERÊNCIAS BIBLIOGRÁFICAS.....	140

LISTA DE FIGURAS

2.1	Diagrama de blocos de um sistema LIT no domínio da frequência.	8
2.2	Projeção do vetor B no plano formado pelas colunas de A	15
2.3	Ortogonalidade entre os regressores, Ψ , e os resíduos, ξ , (AGUIRRE, 2007).	21
3.1	Esquemático de sistema representado por uma Série de Volterra, (ROSA, 2009).	32
3.2	Modelo OBF com dinâmica de Laguerre, (CAMPELLO; AMARAL; OLIVEIRA, 2007).	40
3.3	Modelo Volterra-OBF, implementado a partir de funções de Kautz.	43
3.4	Modelo NARX-OBF com dinâmica de Kautz.	45
5.1	Ilustração do operador de <i>Crossover</i> com um ponto de corte.	66
5.2	Ilustração do operador de Mutação.	66
5.3	Operações de reflexão, expansão, contração e redução, (ABDEL; FUKUSHIMA, 2003).	69
5.4	Atualização das características, $\mathcal{C} = c_1, c_2, \dots, c_n$, através das leis físicas, $\mathcal{L} = l_1, l_2, \dots, l_n$	77
5.5	Operador de propagação de característica de um universo propagante, \mathbb{U}^* , para um universo qualquer, \mathbb{U} , ou simplesmente: $\mathbb{U}^* \gg \mathbb{U}$	78
5.6	Operador de Armagedom, $\mathcal{A}(\cdot)$, atuando sobre a característica c_1	79
5.7	História de um Multiverso. A evolução das características do melhor universo durante convergência do APA.	81
5.8	Curva de convergência do Algoritmo do Princípio Antrópico.	83
5.9	Sinal de entrada $u(k)$ aplicado ao sistema $H(z)$ e sua saída estimada $y(k)$, em relação ao conjunto de dados de validação.	86
5.10	Curva de convergência do Algoritmo do Princípio Antrópico.	87
6.1	Fluxo de decisão entre os modelos OBF lineares (6.1) ou (6.2) e não lineares do tipo (6.3) ou (6.4). Em que o losango de decisão SA é responsável por testar se o sistema a ser identificado tem ou não características subamortecidas e o MSE representa o erro médio quadrático.	94
6.2	Fluxograma que expressa o comportamento do Algoritmo Genético, com utilização dos operadores de Mínimos Quadrados e de Nelder-Mead, aplicados a busca de parâmetros de modelos num processo de Identificação de Sistemas.	96
6.3	Fluxograma que expressa o comportamento do Algoritmo Genético, em hibridização com o método de Levenberg-Marquardt, aplicado a busca de parâmetros de modelos num processo de Identificação de Sistemas.	98
6.4	Fluxo de decisão entre os modelos Volterra-OBF e NARX-OBF. Em que o losango de decisão NT é responsável por avaliar o número de termos que compõem cada modelo e o MSE representa o erro médio quadrático.	99
7.1	Sinal de entrada $u(k)$ aplicado ao sistema $H(z)$ e sua saída $y(k)$	104

7.2	Convergência do Algoritmo do Princípio Antrópico.	105
7.3	Sinal de entrada $u(k)$ aplicado ao sistema $H(z)$ e sua saída $y(k)$	106
7.4	Convergência do Algoritmo Genético.	107
7.5	Levitador Magnético.....	108
7.6	Levitador Magnético, (ECP, 1999).....	108
7.7	a - Sinal de Entrada e b - Sinal de Saída.	109
7.8	Modelo Volterra-OBF utilizando 2 funções de Kautz ($MSE = 0,0886$).	110
7.9	Modelo Volterra-OBF utilizando 6 funções de Kautz ($MSE = 0,0105$).	111
7.10	Modelo linear polinomial, de ordem 50 nas regressões tanto de entrada quanto de saída. Resultado obtido com a <i>toolbox</i> Ident do Matlab ($MSE = 0,0983$).....	115
7.11	Modelo ARX Não Linear, com ordem igual a 10 tanto nas regressões de entrada quanto nas de saída. Resultado obtido com a <i>toolbox</i> Ident do Matlab ($MSE = 0,1056$).	116
7.12	Resposta do modelo NARX, com <i>kernels</i> expressos a partir de 2 funções de Kautz em linhas tracejadas e a resposta do sistema real, em linhas contínuas ($MSE = 0,0085$).	117
7.13	Resposta do modelo NARX, com <i>kernels</i> expressos a partir de 4 funções de Kautz em linhas tracejadas e a resposta do sistema real, em linhas contínuas ($MSE = 0,0019$).	118
7.14	Implementação no Simulink de um modelo NARX-OBF, com 4 funções de Kautz.	120
7.15	Parte linear, relativa as saídas filtradas, do modelo NARX-OBF, com 4 funções de Kautz (bloco Said. Lineares na figura 7.14).	121
7.16	Parte não linear, relativa as saídas filtradas, do modelo NARX-OBF, com 4 funções de Kautz (bloco Said. Nonlin. na figura 7.14).	121
7.17	Parte não linear, relativa as entradas e saídas filtradas (termos cruzados), do modelo NARX- OBF, com 4 funções de Kautz (bloco Mescladas na figura 7.14).	122
7.18	Parte linear, relativa as entradas filtradas, do modelo NARX-OBF, com 4 funções de Kautz (bloco Ent. Lineares na figura 7.14).	123
7.19	Parte não linear, relativa as entradas filtradas, do modelo NARX-OBF, com 4 funções de Kautz (bloco Ent. Nonlin. na figura 7.14)	123
7.20	Esquemático do levitador magnético, (COELHO, 2012).	124
7.21	Foto do sistema de levitação magnética horizontal, (COELHO, 2012).	125
7.22	Resposta do levitador magnético horizontal a um sinal senoidal.	126
7.23	Amostas dos sinais de entrada e saída utilizados na identificação do levitador magnético horizontal.	127
7.24	Validação da identificação do levitador magnético horizontal, a partir de um modelo NARX- OBF com 4 funções de Kautz na sua versão simplificada, $MSE = 3,6031 \times 10^{-4}$, com 11 termos (o modelo completo possui 44 termos).	128
7.25	Detalhe da figura 7.24 entre os instantes 1 e 4s, ($MSE = 3,6031 \times 10^{-4}$)	128
7.26	Detalhe da figura 7.24 entre os instantes 6 e 9s (identificação do levitador magnético a partir de Modelos NARX-OBF simplificados, $MSE = 3,6031 \times 10^{-4}$).	129
7.27	Detalhe da figura 7.24 entre os instantes 11 e 14s (identificação do levitador magnético a partir de Modelos NARX-OBF simplificados, $MSE = 3,6031 \times 10^{-4}$).	129

7.28	Resultado da identificação do levitador magnético vertical, a partir da <i>toolbox</i> do Matlab, utilizando modelos não lineares Wiener/Hammerstein, ($MSE = 3,9075 \times 10^{-4}$).	130
7.29	Detalhe da figura 7.28 entre os instantes 1 e 4s. Identificação do levitador magnético a partir da <i>toolbox</i> ident do Matlab, com modelos Wiener/Hammerstein, ($MSE = 3,9075 \times 10^{-4}$). ...	130
7.30	Resultado da identificação do levitador magnético vertical, a partir da <i>toolbox</i> do Matlab, utilizando modelos Nonlinear ARX com redes <i>Wavelet</i> de 10 funções ($MSE = 4,5638 \times 10^{-4}$).	131
7.31	Detalhe da figura 7.30 entre os instantes 1 e 4s. Identificação do levitador magnético a partir da <i>toolbox</i> ident do Matlab, com modelos Nonlinear ARX, ($MSE = 4,5638 \times 10^{-4}$).	131

Lista de Símbolos

Símbolos Latinos

A	Matriz de estados de sistema contínuo
\mathcal{A}	Matriz de estimação de parâmetros
B	Matriz de entrada de um sistema contínuo
b	Viés (<i>Bias</i>)
C	Matriz de saída de um sistema contínuo
\mathfrak{C}	Vetor que representa um cromossomo
\mathbb{C}	Conjunto dos números complexos
\mathcal{C}	Característica de um universo \mathbb{U}
\mathcal{C}	Conjunto dos termos candidatos em um modelo NARX-OBF
D	Matriz de alimentação direta de um sistema contínuo
\mathfrak{D}^n	Espaço vetorial composto funções de transferência com ordem menor ou igual a n
$E[.]$	Esperança matemática
$f(.)$	Função estática não linear
H	Matriz de saída de um sistema discreto
\mathcal{H}	Espaço de Hilbert
$H(z)$	Função de transferência
\mathcal{H}	Matriz de estimação da saída
$h(t)$	Resposta ao impulso de um sistema
\mathcal{J}_p	Indivíduo produto
\mathcal{J}_R	Indivíduo de referência
$J(x)$	Função de custo a ser minimizada
i	Unidade imaginária
K	Matriz de ganho dos métodos de estimação recursiva
$K(z)$	Função de Kautz
$L(z)$	Função de Laguerre
$\mathcal{L}\{.\}$	Tranformada de Laplace
\mathfrak{L}	Leis físicas de um universo \mathbb{U}
$M(k)$	Modelo linear dinâmico
\mathbb{M}	Estrutura que representa um multiverso em um Algoritmo do Princípio Antrópico
\mathcal{M}	Matriz de estimação do resíduo
\mathbb{N}	Conjunto do números naturais
\mathcal{P}	Vetor que representa uma população de cromossomos
Q	Matriz com colunas ortonormais
\mathcal{S}	Conjunto dos termos que compõem o modelo NARX-OBF simplificado

Símbolos Latinos

\mathbb{R}	Conjunto dos números reais
r_{uu}	Autocorrelação do sinal u
r_{uy}	Correlação cruzada entre os sinais u e y
\mathbb{U}	Estrutura que representa um universo em um Algoritmo do Princípio Antrópico
\mathbb{U}^*	Universo propagante
u	Vetor de entrada do sistema
$v_i(k)$	Sinal de entrada $u(k)$ filtrado pelo filtro de Laguerre de i -ésima ordem, $L_i(z)$
$w_i(k)$	Sinal de entrada $u(k)$ filtrado pelo filtro de Kautz de i -ésima ordem, $K_i(z)$, para os modelos Volterra-OBF
$w_i^u(k)$	Sinal de entrada $u(k)$ filtrado pelo filtro de Kautz de i -ésima ordem, $K_i(z)$, para os modelos NARX-OBF
$w_i^y(k)$	Sinal de saída $y(k-1)$ filtrado pelo filtro de Kautz de i -ésima ordem, $K_i(z)$, para os modelos NARX-OBF
x	Vetor de estados do sistema
x_e	Vértice de um simplex que sofreu expansão
x_r	Vértice de um simplex que sofreu reflexão
x_{ic}	Vértice de um simplex que sofreu contração interna
x_{oc}	Vértice de um simplex que sofreu contração externa
y	Vetor de saídas do sistema
\mathbb{Z}	Conjunto dos números inteiros
$\mathbb{Z}\{.\}$	Transformada \mathbb{Z}

Símbolos Gregos

α	variável aleatória genérica
β	variável aleatória genérica
$\delta(t)$	Distribuição delta de Dirac
ε	limiar de parada para os algoritmos heurísticos
Γ	Matriz de entrada de um sistema discreto
ξ	Resíduo de um modelo
Ψ	Matriz de Regressores
Ψ_k	Matriz de Regressores filtrados por funções de Kautz
Φ	Matriz de transição de estados para um sistema discreto
σ^2	Variância
θ	Vetor de parâmetros
v	Ruído branco
Ω	Espaço de Busca

Siglas, Acrônimos e Abreviaturas

<i>AG</i>	Algoritmo Genético
<i>APA</i>	Algoritmo do Princípio Antrópico
<i>ACO</i>	<i>Ant Colony Optimization</i>
<i>ARX</i>	<i>Auto Regressive with eXogenous input</i>
<i>ARMAX</i>	<i>Auto Regressive Moving Average with eXogenous input</i>
<i>BLUE</i>	<i>Best Linear Unbiased Estimator</i>
<i>FIR</i>	<i>Finite Impulse Response</i>
<i>GNL</i>	Grau de Não Linearidade
<i>GOBF</i>	<i>Generalized Orthonormal Basis Function</i>
<i>LIT</i>	Linear e Invariante no Tempo
<i>MMQ</i>	Método dos Mínimos Quadrados
<i>MQ</i>	Mínimos Quadrados
<i>MPC</i>	<i>Model Predictive Control</i>
<i>MQE</i>	Mínimos Quadrados Estendidos
<i>MQO</i>	Mínimos Quadrados Ortogonais
<i>MRAC</i>	<i>Model Reference Adaptative Control</i>
<i>MSE</i>	<i>Mean Square Error</i>
<i>NARX</i>	<i>Nonlinear Auto Regressive with eXogenous input</i>
<i>NARMAX</i>	<i>Nonlinear Auto Regressive Moving Average with eXogenous input</i>
<i>NARX</i>	<i>Nonlinear Auto Regressive with eXogenous input</i>
<i>NLARX</i>	<i>NonLinear Auto Regressive with eXogenous input</i> com não linearidades na saída do tipo Sigmóide ou Wavelet
<i>NM</i>	Nelder Mead
<i>OBF</i>	<i>Orthonormal Basis Function</i>
<i>PSO</i>	<i>Particle Swarm Optimization</i>
<i>PDF</i>	<i>Probability Density Function</i>
<i>SISO</i>	<i>Single Input Single Output</i>
<i>SQR</i>	Soma dos Quadrados dos Resíduos

NOTAÇÃO

Neste trabalho, as matrizes, vetores e escalares não apresentarão diferença na notação, sendo definida no texto a natureza de cada variável. As variáveis contínuas serão expressas como função de $t \in \mathbb{R}$, por exemplo $f(t)$, e as variáveis discretas serão expressas como função de $k \in \mathbb{N}$, por exemplo $f(k)$.

As derivadas temporais serão representadas por um ponto na parte superior da variável. Por exemplo, a derivada temporal da variável x será: \dot{x} . As estimativas de variáveis serão expostas com um “circunflexo” na sua parte superior. Se a variável a ser estimada for θ , então sua estimativa será $\hat{\theta}$.

1 INTRODUÇÃO

Wyrð bið ful aræd
Uhtred de Bebbanburg

Nas áreas de Engenharia de Controle e Matemática aplicada, geralmente é necessário encontrar modelos empíricos de sistemas que se deseje analisar ou controlar. Um sistema é o objeto no qual variáveis de diferentes tipos interagem e produzem sinais observáveis, (AGUIRRE, 2007; LJUNG, 1999). Já um modelo é uma relação entre quantidades observadas e permite a predição das propriedades e do comportamento de um objeto, (LJUNG, 1999). Neste contexto, a ciência responsável por relacionar empiricamente um sistema a um modelo é conhecida como Identificação de Sistemas. Em outras palavras, a Identificação de Sistemas é a arte de construir modelos matemáticos que descrevam a dinâmica dos sistemas a partir dos seus dados de entrada e saída, (LJUNG, 1999; ISERMANN; MUNCHHOF, 2011).

Tipicamente, um experimento de Identificação de Sistemas consiste de excitar as entradas com um conjunto "rico" de dados (no mínimo persistentemente excitante), tal como ruído branco, e medir a resposta amostrada. O algoritmo de identificação, então produz uma equação diferencial ou a diferenças relativa à resposta do sistema, (AGUIRRE, 2007; ELKAIN, 2009).

A necessidade de se identificar um sistema repousa no fato de que algumas áreas da Engenharia de Controle, como o MPC, *Model Predictive Control*, têm necessidade de modelos na sua implementação, uma vez que os modelos são utilizados na realização *online* de suas previsões, (MADAKYARU; PATWARDHAN, 2013).

Os sistemas dinâmicos geralmente são divididos em duas grandes classes: os sistemas lineares, que obedecem as propriedades de superposição e homogeneidade, e os sistemas não lineares, que não o fazem, (DORF; BISHOP, 2001). Encontrar modelos para sistemas lineares não constitui atualmente um grande desafio à área de Identificação de Sistemas. Entretanto, para sistemas não lineares a complexidade aumenta consideravelmente, (LANDAU; ZITO, 2006).

Para a identificação de sistemas dinâmicos lineares, discretos no tempo, pode-se recorrer aos modelos polinomiais, como os modelos FIR (*Finite Impulse Response*), ARX (*Auto Regressive with eXogenous input*) e ARMAX (*Auto Regressive Moving Average with eXogenous input*), que apresentam simplicidade de implementação e representam de maneira acurada os sistemas lineares, (AGUIRRE, 2007). Já para tratar a identificação de sistemas não lineares existem inúmeros candidatos a modelos. Destacam-se os modelos de Wiener e Hammerstein, que intercalam modelos dinâmicos lineares e não linearidades estáticas,

(AGUIRRE, 2007; LJUNG, 1999), as séries de Volterra, que embora sejam computacionalmente custosas oferecem boa precisão nas estimativas, (NEMETH; KOLLAR, 2002), e ainda os modelos NARX (*Nonlinear Auto Regressive with exogenous input*), que são modelos de entrada e saída muito utilizados na identificação de sistemas não lineares, além de apresentarem vantagens sobre os modelos de Volterra, pelo potencial de redução do número de termos do modelo, devido a realimentação, (BALDACCHINO TARA; ANDERSON; KADIRKAMANATHAN, 2013).

Uma maneira clássica de se reduzir o número de termos necessários em um modelo é o uso de funções ortonormais, (AGUIRRE, 2007; CAMPELLO R. J. G. B; OLIVEIRA, 2007; SILVA, 2010). Assim, pode-se representar um sistema não linear, a partir de um modelo de Volterra ou NARX, com um número inferior de parâmetros a se determinar.

A ideia de expandir, a partir de funções ortonormais, o conjunto de parâmetros de uma série de Volterra, conhecidos como *Kernels*, data do final da década de 50, proposta por Norbert Wiener, (CAMPELLO; AMARAL; OLIVEIRA, 2007), mas devido às limitações computacionais da época, tais metodologias não foram muito exploradas.

Na década de 80 o interesse nos modelos expandidos em base de funções ortonormais (em inglês: *Orthonormal Basis Functions*) voltou a crescer, (HEUBERGER; HOF; WAHLBERG, 2005), principalmente pela utilização da base FIR, uma vez que tal classe de modelos não apresenta problemas de estabilidade, (OPPENHEIM; SHAFER; BUCK, 1998). Entretanto, ainda eram necessárias quantidades elevadas de termos para proporcionar aproximações precisas utilizando a base FIR, (CAMPELLO; AMARAL; OLIVEIRA, 2007), o que nem sempre era viável, devido as limitações computacionais da época.

Já na década de 90, em (FU; DUMONT, 1993), foi proposta a expansão de funções de transferência estáveis em séries envolvendo as funções de Laguerre e otimização do polo que parametriza tais funções. Neste tipo de modelo tem-se uma generalização do modelo FIR, em que o polo das funções ortonormais não se encontra na origem, (CAMPELLO; AMARAL; OLIVEIRA, 2007). Já (HEUBERGER; HOF; BOSGRA, 1995), deu os fundamentos matemáticos necessários para demonstrar que a utilização de funções de Laguerre poderiam acelerar o processo de convergência de séries de funções ortonormais. Isto com foco na representação de funções de transferência com amortecimento elevado. Em (WAHLBERG, 1994), foi aplicada à área de identificação a extensão dos modelos de Laguerre para os modelos de Kautz. Neste caso as funções ortonormais passam a ser parametrizadas por polos complexos, e que desta forma são mais adequados para modelar sistemas ressonantes, (CAMPELLO; AMARAL; OLIVEIRA, 2007). Com o sucesso destes e de outros trabalhos com o mesmo enfoque, a utilização de funções ortonormais em Identificação de Sistemas voltou à tona e passou a estar presente na literatura das áreas de Controle, Modelagem e Identificação de Sistemas.

Mais recentemente, nos anos 2000, muitos trabalhos seguiram a linha proposta por Nor-

bert Wiener, de descrever os *Kernels* de Volterra a partir de funções ortonormais. Entre eles, pode se citar o trabalho (NEMETH; KOLLAR, 2002), em que os *kernels* de um modelo de Volterra foram determinados por técnicas de interpolação. Já em (CAMPELLO; AMARAL; OLIVEIRA, 2007) e (MADAKYARU; PATWARDHAN, 2013), foram comparados os resultados de modelos baseados em funções ortonormais, com não linearidade do tipo polinomial (modelo de Volterra), baseadas na lógica *Fuzzy* (modelo Takagi Sugeno) e a partir de redes neurais. Tendo sido levantadas as vantagens e as dificuldades apresentadas por cada um destes métodos. Em (ROSA; CAMPELLO; AMARAL, 2009), é exposto um método analítico de seleção dos polos ótimos das funções ortonormais de Kautz e GOBFs (*Generalized Orthonormal Basis Functions*). Isto visando a representação dos *kernels* de Volterra a partir destas funções. Para tanto, o gradiente das saídas dos filtros ortonormais são calculados em relação ao polo das funções, e assim o polo ótimo pode ser encontrado por *Back-Propagation/Levenberg-Marquardt*. Com isso, atinge-se uma redução do número de termos necessários na série de Volterra e, conseqüentemente, um modelo mais simples.

Além de sugerir que modelos de Volterra tivessem seus *kernels* descritos por funções ortonormais, Wiener também sugeriu que modelos ARX, e suas versões não lineares, NARX, tivessem seus parâmetros descritos por funções ortonormais, (AGUIRRE, 2007). Tal sugestão serviu como uma das principais inspirações para o desenvolvimento desta tese e também para alguns trabalhos que serão citados nos próximos parágrafos.

Em 2011, (BADWE; PATWARDHAN; GUDI, 2011) desenvolveu uma metodologia de identificação de sistemas em malha fechada, focando em aplicações de MPC, a partir de modelos ARX-OBF. Para tanto, um sistema, e o ruído a ele aplicado, são identificados a partir de sinais pré-filtrados por funções ortonormais, associadas a modelos ARX. Isto em substituição aos modelos ARX puros, em que são necessárias ordens elevadas, o que torna difícil a determinação dos seus parâmetros. Em (LEMMA; RAMASAMY, 2011), um modelo ARX-OBF é adaptado para modelar sistemas em malha fechada, com incertezas associadas aos seus atrasos no tempo, e demonstrado que os modelos ARX-OBF são mais efetivos que os modelos ARX. Já em (BALDACCHINO TARA; ANDERSON; KADIRKAMANATHAN, 2013), são propostos métodos Bayesianos para a seleção de termos em um modelo NARMAX, em que são estimados as distribuições dos parâmetros de modelo e são feitos testes estatísticos para a inclusão, ou não, de um termo em um modelo NARMAX.

Seguindo a linha exposta nos parágrafos anteriores, em (MADAKYARU; PATWARDHAN, 2013), modelos ARX associados a funções ortonormais foram obtidos e utilizados na área de AMPC (*Adaptative Model Predictive Control*). Em MOHIDEEN, 2013, foi proposta a identificação e controle de um sistema não linear a partir do método MRAC, *Model Reference Adaptative Control*, sendo os parâmetros do controlador sintonizados por Algoritmos Genéticos, (MOHIDEEN et al., 2013). Já (TIELS; SCHOUKENS, 2014), tratou de modelos dinâmicos lineares baseados em funções ortonormais, seguidos por não linearidades do tipo polinomial, conhecidos como modelos Wiener-Schetzen. Neste caso os polos das funções

foram determinados por *Best Linear Approximation* e os coeficientes do polinômio foram buscados por regressão linear.

Pelas exposições acima, é possível perceber que a utilização das funções ortonormais vem sendo bastante explorada na literatura de Identificação de Sistemas nas últimas décadas. Mais do que isso, é possível inferir que trata-se de uma área com significativo potencial de desenvolvimento, e esta premissa foi assumida como verdadeira nesta tese.

1.1 DEFINIÇÃO DO PROBLEMA

A temática tratada nesta tese pode ser definida a partir do seguinte problema de pesquisa:

Investigação de condições para que os modelos NARX possam ter seus *kernels* desenvolvidos a partir de funções ortonormais, gerando modelos que necessitem de um número menor de termos para representar um sistema não linear.

Para que tal problema de pesquisa possa ser tratado, nas próximas seções serão expostos os objetivos desta tese.

1.2 OBJETIVOS

Nesta seção serão apresentados os objetivos, gerais e específicos, deste trabalho.

1.2.1 Objetivo Geral

O objetivo geral deste trabalho consiste em desenvolver uma metodologia de identificação de sistemas baseada em modelos NARX, com número de parâmetros reduzido pelo uso de funções ortonormais.

1.2.2 Objetivos Específicos

Para que o objetivo principal desta tese possa ser atingido, tem-se os seguintes objetivos específicos:

1. Identificar sistemas não lineares através de modelos NARX com seus *kernels* descritos por funções de base ortonormais;
2. Estudar o desempenho de novos operadores heurísticos aplicados a Algoritmos Genéticos destinados a resolução de problemas de identificação de sistemas;

3. Desenvolver métodos de otimização heurística que permitam encontrar parâmetros aceitáveis para o modelo de um sistema não linear.
4. Realizar a seleção dos termos mais relevantes de um modelo NARX-OBF de forma a se obter um modelo mais simples, sem a perda da capacidade de representação do modelo.

Para motivar a leitura deste texto serão apresentadas a seguir as principais contribuições e dificuldades encontradas neste trabalho.

1.3 CONTRIBUIÇÕES DO TRABALHO

Entre as principais contribuições deste trabalho destacam-se as seguintes:

1. Identificação de sistemas não lineares a partir de Modelos NARX com seus *kernels* descritos por funções ortonormais;
2. Proposta de novos operadores para Algoritmos Genéticos.
3. Criação de um paradigma inédito para os algoritmos de otimização heurística: O Algoritmo do Princípio Antrópico.
4. Proposta de um novo método de simplificação, de um modelo NARX-OBF, a partir de seleção de termos utilizando métodos heurísticos.

1.4 PRINCIPAIS DIFICULDADES

Entre as principais dificuldades encontradas neste trabalho podem ser citadas:

1. A grande variedade de comportamentos possíveis em um sistema não linear torna difícil a escolha de uma metodologia de identificação que seja aplicável a qualquer sistema não linear;
2. Os algoritmos heurísticos nem sempre convergem a valores aceitáveis para os parâmetros dos modelos no processo de identificação;
3. A obtenção de dados de processos reais para identificação mostrou-se um processo tão complexo quanto a própria identificação em si.

A divisão de conteúdos desta tese será exposta na próxima seção.

1.5 ORGANIZAÇÃO DA TESE DE DOUTORADO

Este trabalho está organizado em 8 capítulos. Sendo o capítulo 2 responsável por apresentar os principais conceitos necessários à Identificação de Sistemas, o método dos Mínimos Quadrados, os algoritmos de estimação recursiva e suas principais variações. No capítulo 3 são apresentadas algumas das principais características de sistemas não lineares, bem como os modelos de Volterra e NARX, e as funções de base ortonormal. No capítulo 4 são expostos os conceitos relativos à otimização determinística, com foco na otimização de parâmetros de modelos para sistemas não lineares. No capítulo 5 são apresentados os conceitos relativos à otimização heurística, com foco nos Algoritmos Genéticos, sendo ainda apresentado o Algoritmo do Princípio Antrópico. No capítulo 6 são expostas as metodologias propostas nesta tese. Ficando para o capítulo 7 a parte dos experimentos de validação, a partir de um levitador magnético com oscilação vertical e um segundo levitador com oscilação horizontal. No capítulo 8 são apresentadas as principais conclusões do trabalho e as propostas de trabalhos futuros.

2 CONCEITOS PRELIMINARES

Neste capítulo serão apresentados os conceitos preliminares desta tese. São expostas as definições de sistema e identificação de sistemas, bem como as principais metodologias de identificação. Serão tratadas a identificação por deconvolução, por correlação e o método dos mínimos quadrados. São apresentados ainda os principais estimadores recursivos aplicados à Identificação de Sistemas.

2.1 SISTEMAS E MODELOS

Um sistema é uma estrutura na qual diferentes variáveis interagem e produzem sinais observáveis, (LJUNG, 1999). Neste cenário, considere uma família de sistemas em que um determinado número de variáveis dependem do tempo $x_1(t), \dots, x_n(t)$ e têm sua evolução descrita por uma equação da forma:

$$\begin{cases} \dot{x} = F(t, x, u) \\ y = G(t, x, u), \end{cases} \quad (2.1)$$

onde $F : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ e $G : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^l$. As variáveis do vetor $x^T = [x_1(t), \dots, x_n(t)]$ são conhecidas como variáveis de estado, e o vetor $u^T = [u_1(t), \dots, u_m(t)]$ é chamado de vetor de variáveis de controle ou de entradas, (BAUMEISTER; AO, 2014).

A principal característica de um sistema dinâmico é que o futuro depende do passado, (LJUNG, 1999). Assim, a predição da saída torna-se estocasticamente calculável a partir dos entradas anteriores.

Os sistemas genéricos expressos pela equação (2.1) podem ter seu escopo restringido a sistemas conceitualmente mais simples: os sistemas lineares, expressos pela equação (2.2).

$$\begin{cases} \dot{x}(t) = A(t)x(t) + B(t)u(t) \\ y(t) = C(t)x(t) + D(t)u(t) \end{cases} \quad (2.2)$$

em que $y(t)$ é o vetor de variáveis mensuráveis, $A(t)$ é a matriz de estados, as matrizes $B(t)$ e $C(t)$ correspondem às matrizes de entrada e saída, respectivamente, e $D(t)$ é dita matriz de alimentação direta. Já as variáveis x , u e y podem ser definidas da seguinte forma: $x : [t_0, t_1] \rightarrow \mathbb{R}^n$, $u : [t_0, t_1] \rightarrow \mathbb{R}^m$ e $y : [t_0, t_1] \rightarrow \mathbb{R}^l$, em que t_0 e t_1 são os tempos inicial e final de medida. Este sistemas, descritos por (2.2), são conhecidos como sistemas lineares.

No caso de sistemas lineares e invariantes no tempo e sem transmissão direta, as matrizes A , B e C são constantes e D é nula.

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) \end{cases} \quad (2.3)$$

Por simplicidade de notação, neste trabalho, tais sistemas serão referenciados apenas por (A,B,C) .

Se a entrada $u(t)$ e a saída $y(t)$ de um sistema linear e invariante no tempo (A,B,C) forem escalares reais, pode-se recorrer a uma das modelagens mais utilizadas para esta classe de sistemas, a função de transferência.

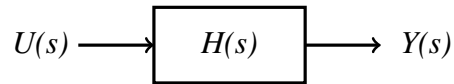


Figura 2.1: Diagrama de blocos de um sistema LIT no domínio da frequência.

Na figura 2.1 é ilustrado um sistema linear e invariante no tempo, LIT, modelado por uma função de transferência $H(s)$. Os sistemas que possuem apenas uma entrada, $u(t) \in \mathbb{R}$, e apenas uma saída, $y(t) \in \mathbb{R}$, são conhecidos como sistemas SISO, *Single Input Single Output*.

Uma função de transferência $H(s)$ é definida como a relação entre a transformada de Laplace da saída, $Y(s) = \mathcal{L}\{y(t)\}$, e da entrada, $U(s) = \mathcal{L}\{u(t)\}$, de um sistema, como expresso pela equação (2.4), (DORF; BISHOP, 2001).

$$H(s) = \frac{Y(s)}{U(s)} = \frac{\mathcal{L}\{y(t)\}}{\mathcal{L}\{u(t)\}}, \quad (2.4)$$

com condições iniciais nulas, sendo a transformada de Laplace definida por:

$$\mathcal{L}\{f(t)\} = \int_{-\infty}^{\infty} f(t)e^{-st} dt, \quad (2.5)$$

em que a função $f(t)$ segue a forma $f: \mathbb{R} \rightarrow \mathbb{R}$ e a variável s definida como: $s = \sigma + iw$, sendo i a unidade imaginária, $\sigma \in \mathbb{R}$ e w a frequência angular, em rad/s.

Neste contexto, se um sistema LIT, modelado por uma função de transferência $H(s)$, fosse excitado por um sinal com a seguinte descrição:

$$\delta(t) = \begin{cases} 0 & \text{se } t \neq 0 \\ \infty & \text{se } t = 0, \end{cases} \quad (2.6)$$

com a restrição: $\int_{-\infty}^{\infty} \delta(t) dt = 1$, obter-se-ia a chamada resposta ao impulso $h(t)$ do sistema $H(s)$:

$$\begin{aligned} h(t) &= \mathcal{L}^{-1}\{H(s)\mathcal{L}[\delta(t)]\} \\ &= \mathcal{L}^{-1}\{H(s)\}. \end{aligned} \quad (2.7)$$

Com isto, tem-se algumas definições iniciais relativas a sistemas e pode-se tratar da identificação de sistemas, que será introduzida na próxima seção.

2.2 IDENTIFICAÇÃO DE SISTEMAS

A Identificação de Sistemas é a área de conhecimento que estuda a modelagem matemática realizada com pouco ou nenhum conhecimento prévio do sistema a ser modelado, (AGUIRRE, 2007). Sendo assim, identificar um sistema significa determinar um modelo que represente a sua dinâmica, a partir dos dados de entrada e saída do mesmo, (LANDAU; ZITO, 2006).

A Identificação de Sistemas trata da construção matemática de modelos para sistemas dinâmicos a partir de dados observados, (LJUNG, 1999). Tal procedimento se justifica pois, nem sempre se tem o conhecimento das leis físicas que regem o sistema e, geralmente, o conhecimento de um modelo é necessário ao projeto e a implementação de controladores de alta performance, (LANDAU; ZITO, 2006).

A Identificação de Sistemas geralmente é dividida em 3 grandes grupos, de acordo com a natureza dos elementos necessários para a determinação do modelo. São eles:

- Identificação Caixa Branca: é baseada nas leis físicas que descrevem a dinâmica do sistema;
- Identificação Caixa Preta: fundamenta-se única e exclusivamente nos dados de entrada e saída do sistema;
- Identificação Caixa Cinza: funde as duas metodologias anteriores, ou seja, faz uso de algum conhecimento prévio, quer parâmetros, quer estrutura ou lei física, e de medidas

de alguns sinais de entrada e saída.

Estes três tópicos serão tratados em mais detalhes nas próximas seções.

2.2.1 Identificação Caixa Branca

A identificação caixa branca de um sistema é feita a partir das leis físicas que descrevem este sistema. Neste caso é necessário um vasto conhecimento sobre os princípios que regem o sistema, e normalmente, o efeito do ruído não é levado em conta neste tipo de modelagem.

Geralmente, na modelagem caixa branca parte-se dos mecanismos físicos para se encontrar uma equação diferencial, caso o sistema seja contínuo, ou uma equação a diferenças, caso o sistema seja discreto, para descrever o sistema.

Ressalta-se que, em situações práticas, dificilmente se tem informações suficientes para se obter um modelo acurado do sistema.

2.2.2 Identificação Caixa Preta

A identificação caixa preta compreende os métodos de modelagem cuja única fonte de informação, a partir da qual o modelo é construído, é o conjunto de dados dinâmicos, (AGUIRRE, 2007).

Numa configuração caixa preta, a ideia é parametrizar uma função dinâmica genérica, de modo flexível, assim o sistema pode ter sua dinâmica aproximada por tal função, desde que os parâmetros adequados sejam escolhidos. A escolha típica é o uso de expansões em funções de base, (LJUNG, 1999), algo semelhante ao que as séries de Taylor e de Fourier fazem com funções estáticas.

Cabe ressaltar que este trabalho focará exclusivamente na identificação caixa preta de sistemas.

2.2.3 Identificação Caixa Cinza

A identificação caixa cinza de um sistema é baseada em modelos híbridos, uma vez que são utilizadas técnicas de identificação caixa preta e caixa branca simultaneamente. Um exemplo prático de identificação caixa cinza pode ser observado nos filtros de Kalman e seus variantes, (ISERMANN; MUNCHHOF, 2011).

por tratar-se de uma operação inversa à convolução, este método de identificação ficou conhecido como Deconvolução.

Cabe ressaltar que, para que a matriz U seja inversível, é necessário que ela seja quadrada e que suas colunas sejam linearmente independentes, (STRANG, 2009; LIMA, 2008). Assim, as dimensões da matriz U devem ser escolhidas de forma a atender tais requisitos.

Este método é uma das formas mais simples de identificação de sistemas, mas possui como inconveniente a falta de robustez ao ruído. Quando se tem uma baixa relação sinal ruído tem-se uma perda significativa de desempenho na identificação, (AGUIRRE, 2007). Assim, apesar do método da Deconvolução ser simples e intuitivo, ele é pouco útil em situações práticas, nas quais o ruído e as não linearidades estão presentes.

2.4 IDENTIFICAÇÃO BASEADA EM CORRELAÇÃO

Como mencionado anteriormente, o método de identificação baseado na Deconvolução é pouco robusto ao ruído e oferece bons resultados apenas nas situações em que se tem uma boa relação sinal/ruído. Como alternativa, existem métodos que levam em consideração o aspecto estocástico dos sinais de entrada e saída amostrados e tornam-se, desta forma, mais resistentes à presença do ruído. Nesta linha de algoritmos, a identificação baseada em correlação é uma das mais utilizadas.

Para tanto, considera-se os sinais de entrada $u(t)$ e saída $y(t)$ de um sistema como variáveis aleatórias. Assim define-se correlação cruzada como:

$$r_{u,y}(\tau,t) = E\{u(t)y(t+\tau)\} \quad (2.12)$$

$$r_{u,y}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T u(t)y(t+\tau)dt,$$

que pode ser expressa, no caso discreto, da seguinte forma:

$$r_{u,y}(k) = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{i=-N}^N u(i)y(i+k). \quad (2.13)$$

Um caso particular de $r_{u,y}$ se dá quando $u = y$, neste caso se tem a autocorrelação de u , expressa por:

$$r_{u,u}(k) = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{i=-N}^N u(i)u(i+k). \quad (2.14)$$

Nos casos práticos, em que se tem um número finito de amostras pode-se utilizar os seguintes estimadores¹ para $\hat{r}_{u,u}$ e $\hat{r}_{u,y}$, (AGUIRRE, 2007):

$$\hat{r}_{u,u}(k) = \frac{1}{N} \sum_{i=1}^{N-k} u(i)u(i+k), \quad \text{para } k = 0, 1, \dots, N-1. \quad (2.15)$$

$$\hat{r}_{u,y}(k) = \begin{cases} \frac{1}{N} \sum_{i=1}^{N-k} u(i)y(i+k), & \text{para } k = 0, 1, \dots, N-1, \\ \frac{1}{N} \sum_{i=1}^{N+k} u(i)y(i-k), & \text{para } k = 0, -1, \dots, -N+1. \end{cases} \quad (2.16)$$

A partir destas definições, pode-se afirmar que a identificação de sistemas baseada em correlação parte do seguinte princípio: a resposta ao impulso h de um sistema linear e invariante no tempo (A,B,C) pode ser obtida a partir do produto entre o inverso da autocorrelação do sinal de entrada r_{uu}^{-1} e correlação cruzada, r_{uy} , (PAPOULIS; PILLAI, 2002):

$$h = r_{uu}^{-1} r_{uy}. \quad (2.17)$$

É importante mencionar que este método é mais robusto que a deconvolução, mas, ainda assim, apresenta problemas quanto a minimizar a variância do erro de saída do modelo obtido, (AGUIRRE, 2007).

2.5 MÉTODO DOS MÍNIMOS QUADRADOS

A maioria dos problemas de identificação de sistemas pode ser simplificada a uma estrutura na qual se tem variáveis desconhecidas, parâmetros do sistema, e variáveis conhecidas, observações do sistema. Neste cenário, existe um número de observações necessárias para a determinação das grandezas desconhecidas, ou seja, o número de observações que torna

¹Uma definição mais apropriada seria utilizar como divisor $N - K$ em vez de N . Pois assim cada termo de autocorrelação/correlação cruzada seria calculado pela média dos pontos efetivamente utilizados.

o sistema de equações possível e determinado. Com isso, teoricamente, os parâmetros do sistema poderiam ser determinados. Entretanto, nas situações práticas, tem-se o ruído associado à medição das variáveis envolvidas, e, portanto são necessárias mais observações, o que torna o sistema de equações sobredeterminado.

Pelo prisma da Álgebra Linear, os problemas mais básicos de identificação de sistemas podem ser entendidos como uma tentativa de resolução do seguinte sistema linear:

$$AX = B, \quad (2.18)$$

sendo $X \in \mathbb{R}^{k \times 1}$ o vetor de parâmetros do sistema a ser identificado, considere ainda que as matrizes $A \in \mathbb{R}^{n \times k}$ e $B \in \mathbb{R}^{n \times 1}$ contenham os sinais de entrada e saída, observações empíricas.

Se $n = k$, é possível encontrar X a partir de $A^{-1}B$, de forma semelhante ao exposto na seção 2.3, mas, se $n > k$, esta operação torna-se inviável, pois a matriz A deixa de ser quadrada. Para sanar esta dificuldade, faz-se uso da pré-multiplicação de ambos os membros da equação (2.18) por A^T :

$$[A^T A]X = A^T B. \quad (2.19)$$

$A^T A$ é uma matriz quadrada. Se as colunas de A forem linearmente independentes então $A^T A$ será inversível, (STRANG, 2009). Tem-se ainda que a multiplicação de A por A^T preserva o número de equações linearmente independentes do sistema (posto da matriz A), pois:

$$\text{Rank}\{A^T A\} = \text{Rank}\{A\},$$

sendo $\text{Rank}\{.\}$ o posto da matriz.

Assim, pode-se isolar X :

$$\hat{X} = [A^T A]^{-1} A^T B, \quad (2.20)$$

utilizando-se o símbolo \hat{X} para expressar que tal valor é uma estimativa de X .

Como, no caso sobredeterminado, X pode não atender todas as equações que são combinadas em (2.18). Assim, busca-se conciliar as diferentes equações da melhor forma possível. A técnica de isolamento do vetor de incógnitas, X , quando A não é inversível é conhecida como método da pseudoinversa, (STRANG, 2009).

Para se justificar (2.20) volta-se ao sistema (2.18). No caso em que $n > k$, geralmente

tem-se um sistema impossível, pois se tem apenas k parâmetros a serem determinados e n medidas contaminadas pelo ruído, o que pode gerar equações conflitantes. Assim, torna-se inviável a resolução do sistema expresso em (2.18), mas a projeção de B no espaço vetorial das colunas de A é possível, o que na determinação de X leva a solução expressa em (2.20), (STRANG, 2009).

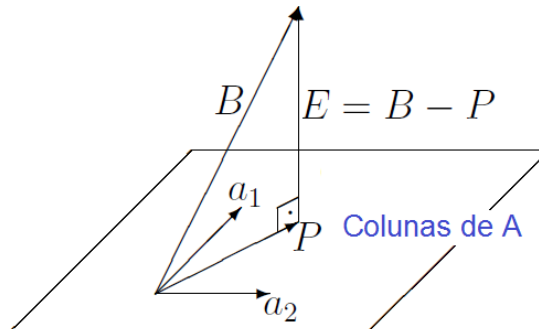


Figura 2.2: Projeção do vetor B no plano formado pelas colunas de A .

No caso de um sistema sobredeterminado, $n > k$, o vetor B não pode ser composto por uma combinação linear dos elementos de A . Assim, faz-se necessária a projeção de B sobre o subespaço vetorial formado pelas colunas de A , como pode ser observado na figura 2.2 (para $n > k = 2$). Desta forma não tem-se a resolução do sistema expresso em (2.18), mas sim do sistema exposto em (2.21):

$$A\hat{X} = P, \quad (2.21)$$

sendo P a projeção de B sobre o hiperplano formado pelas colunas de A . Assim, P é o vetor possível mais próximo de B , sendo esta a solução ótima do sistema $AX = B$, que seria inicialmente irrealizável.

Cabe ressaltar que a diferença vetorial entre B e sua projeção P é dado por E . Este elemento pode ser interpretado como o vetor de erro e o mesmo é perpendicular a A , conforme a figura 2.2.

$$E = B - P. \quad (2.22)$$

Para utilização da equação (2.20) com uma notação mais próxima da área de Engenharia de Controle, redefine-se X e Y como os vetores de entradas e saídas, respectivamente, de um sistema linear e invariante no tempo (LIT) e θ como o vetor que contém seus parâmetros. Assim, tem-se:

$$\hat{\theta} = [X^T X]^{-1} X^T Y, \quad (2.23)$$

que é solução ao seguinte sistema:

$$\hat{Y} = X \hat{\theta}, \quad (2.24)$$

Esta metodologia foi desenvolvida por Gauss em 1809, e ficou conhecida como o estimador dos mínimos quadrados, (STRANG, 2009).

Substituindo-se a equação (2.23) em (2.24), tem-se:

$$\hat{Y} = X[X^T X]^{-1} X^T Y. \quad (2.25)$$

Pode-se ainda estender a equação (2.23) para modelos auto-regressivos, em que a matriz de entrada, X , se torna a matriz de medidas, ψ , e passa a contemplar também as medidas anteriores de Y :

$$\hat{\theta} = [\psi^T \psi]^{-1} \psi^T Y. \quad (2.26)$$

O vetor ψ , chamado de vetor de regressores, pode ser representado da seguinte forma, para um número igual de medidas e parâmetros, $n = k$:

$$\psi = [y(n), y(n-1), \dots, y(n-N_y), x(n), x(n-1), \dots, x(n-N_x)]. \quad (2.27)$$

Tomando-se mais observações de x e de y , ou seja, tornando-se o sistema sobredeterminado, $n > k$, pode-se expandir o vetor ψ para a matriz Ψ :

$$\Psi = \begin{bmatrix} y(0), & y(-1) & \dots, & y(-N_y), & x(0), & x(-1) & \dots, & x(-N_x) \\ \vdots & & & \vdots & & & & \vdots \\ y(n), & y(n-1) & \dots, & y(n-N_y), & x(n), & x(n-1) & \dots, & x(n-N_x) \\ \vdots & & & \vdots & & & & \vdots \end{bmatrix}, \quad (2.28)$$

com isso, a equação (2.24) pode ser escrita como:

$$\hat{Y} = \Psi \hat{\theta}, \quad (2.29)$$

sendo θ o vetor de parâmetros, Ψ a matriz de medidas anteriores e \hat{Y} a estimativa da saída atual do sistema.

Em situações de identificação de sistemas reais as medidas estão “contaminadas” pelo ruído: e . Assim, pode-se esperar uma diferença entre os valores estimados, \hat{Y} , e medidos efetivamente, Y , justamente devido ao ruído.

$$\begin{aligned} Y &= \hat{Y} + e \\ Y &= \Psi \cdot \hat{\theta} + e. \end{aligned} \quad (2.30)$$

Neste cenário, define-se o ruído, e , como o agente responsável por corromper a variável a ser explicada e o resíduo, ξ , como o erro cometido pelo modelo ao explicar a variável Y , (AGUIRRE, 2007).

Assim, para o modelo estimado, a diferença entre o Y real e sua estimativa, $X\theta$, corrompida por e , é dada pelo vetor de resíduos ξ .

$$\xi = Y - \Psi \hat{\theta}, \quad (2.31)$$

ou, ainda, na seguinte forma:

$$Y = \Psi \hat{\theta} + \xi. \quad (2.32)$$

Pode-se minimizar a variância do modelo obtido utilizando-se mais medidas de entrada e saída, ou seja, tornando o sistema sobredeterminado, $n > k$. Sob este prisma, pode-se estabelecer uma métrica para quantificar a qualidade dos parâmetros obtidos pelo processo de identificação. Uma alternativa interessante para isso é a soma dos quadrados dos resíduos, expresso por:

$$f(\xi) = \sum_{i=1}^N \xi(i)^2. \quad (2.33)$$

Para que a qualidade do modelo possa ser maximizada, o que equivale a reduzir o máximo possível os resíduos, a função expressa em (2.33) deve ter seu mínimo encontrado. Portanto,

tem-se:

$$\frac{\partial f(\xi)}{\partial \hat{\theta}} = 0.$$

Apenas por simplicidade, nesta operação será considerado um modelo não auto-regressivo, ou seja: $Y = X\theta$.

$$f(\xi) = (Y - X\hat{\theta})^T \cdot (Y - X\hat{\theta}),$$

$$f(\xi) = Y^T Y - Y^T X \hat{\theta} - \hat{\theta}^T X^T Y + \hat{\theta}^T X^T X \hat{\theta},$$

$$\begin{aligned} \frac{\partial f(\xi)}{\partial \hat{\theta}} &= -(Y^T X^T)^T - X^T Y + (X^T X + X^T X) \hat{\theta}, \\ &= -X^T Y - X^T Y + 2X^T X \hat{\theta}. \end{aligned} \quad (2.34)$$

Fazendo-se a equação (2.34) igual a zero, pois a derivada é nula nos pontos de máximo e mínimo (ver teorema 4.1.1), tem-se:

$$\hat{\theta} = [X^T X]^{-1} X^T Y.$$

com isso, chega-se a equação (2.23), obtida pelo método da pseudoinversa.

A verificação de que este ponto trata-se realmente de um ponto de mínimo depende ainda da análise da derivada segunda de (2.33), conhecida como matriz Hessiana, e esta deve ser semi-definida positiva, (IZMAILOV; SOLODOV, 2014; NOCEDAL; WRIGHT, 1999). Para maiores detalhes em relação a este requisito consulte a seção 4.1.

A solução, apresentada pela equação (2.23), minimiza a função de custo mostrada em (2.33) e sua principal vantagem é ser o mínimo global de tal função, (LJUNG, 1999).

Para o adequado tratamento dos próximos conceitos, considere a definição de esperança

matemática em relação à variável aleatória α , de natureza discreta ou contínua amostrada, que segue uma função de massa de probabilidade $p(\alpha_i)$:

$$E\{\alpha\} = \sum_{i=1}^{\infty} \alpha_i p(\alpha_i), \quad (2.35)$$

Da mesma forma, considere a definição de covariância entre duas variáveis aleatórias α e β , também discretas ou contínuas amostradas:

$$Cov\{\alpha, \beta\} = E\{(\alpha - E\{\alpha\})(\beta - E\{\beta\})\}. \quad (2.36)$$

Um caso particular de $Cov\{\alpha, \beta\}$ acontece quando $\alpha = \beta$. Este caso recebe o nome de variância de α , expressa por $Var\{\alpha\}$:

$$\begin{aligned} Var\{\alpha\} &= E\{(\alpha - E\{\alpha\})(\alpha - E\{\alpha\})\} \\ &= E\{(\alpha - E\{\alpha\})^2\}. \end{aligned} \quad (2.37)$$

Com esta definição, pode ser conceituada a ideia de *bias*, que serve para verificar a presença de erro sistemático na estimativa dos parâmetros. O *bias* é definido por:

$$b = E\{\hat{\theta}\} - \theta. \quad (2.38)$$

Para que um estimador esteja livre de *bias* é necessário que algumas premissas sejam assumidas, e estas serão tratadas na próxima seção.

2.5.1 Hipóteses de Gauss-Markov

Para garantir a ausência de viés na estimação de parâmetros, é necessário assumir a veracidade de algumas suposições nos processos de identificação de sistemas sem tratamento de *bias*. Estas são conhecidas como hipóteses de Gauss-Markov e são expressas abaixo:

- A esperança matemática para o vetor de resíduos, ξ , indexado por $i \in \mathbb{N}$, deve ser nula para qualquer i :

$$E\{\xi_i\} = 0, \quad \forall i, \quad (2.39)$$

sendo $E[.]$ o operador esperança matemática.

- A variância dos resíduos, ξ_i , deve ser constante e igual a σ^2 para qualquer i :

$$\text{Var}(\xi_i) = \sigma^2, \quad \forall i; \quad (2.40)$$

- A covariância entre os termos ξ_i e ξ_j , do vetor de resíduos, deve ser nula exceto quando $i = j$:

$$\text{Cov}(\xi_i, \xi_j) = 0 \quad \forall i \neq j, \quad (2.41)$$

sendo i e $j \in \mathbb{N}$.

Em um estimador é desejável que as condições impostas pelas hipóteses de Gauss-Markov sejam atendidas, mas nem sempre os sistemas a serem identificados atendem tais requisitos.

Os estimadores que atendem as condições impostas pelas hipóteses de Gauss-Markov são conhecidos como BLUE, *Best Linear Unbiased Estimators*, e possuem como principal característica:

$$E\{\theta\} = \theta, \quad (2.42)$$

ou seja, são estimadores livres de viés (bias), (LJUNG, 1999).

2.5.2 Ortogonalidade entre Regressores e Resíduos

Uma das propriedades mais interessantes do estimador dos Mínimos Quadrados, MQ, é a ortogonalidade, conceito que estabelece que, se dois vetores A e B forem ortogonais, $A \perp B$, o produto interno entre estes vetores será nulo, (LIMA, 2008; RORRES, 2000).

$$\langle A, B \rangle = 0.$$

Assim, para que não haja correlação entre os regressores, Ψ , e o ruído, e , o produto interno entre os mesmos deve ser nulo (o vetor A deve possuir projeção nula sobre B).

$$\langle \Psi, e \rangle = 0. \quad (2.43)$$

Consequentemente, a esperança deste produto deve ser também nula.

$$E\{\Psi e\} = 0. \quad (2.44)$$

Uma vez que (2.44) não seja verdadeira, o estimador sofre polarização e isto, geralmente, ocorre quando o ruído de regressão não for branco, (AGUIRRE, 2007; LJUNG, 1999). Para sanar tais problemas, existem algumas metodologias como o operador estendido dos mínimos quadrados, o branqueamento do ruído, entre outras.

Quando um modelo é inconsistente em relação ao parâmetro, existirá um erro sistemático (polarização) nos parâmetros do modelo e este erro não pode ser removido com o aumento da dimensão dos vetores de entrada e saída apresentados ao estimador, (LEMMA; RAMASAMY; SHUHAIMI, 2010).

Para que não haja polarização do estimador MMQ, deve-se atender a condição de que os resíduos sejam ortogonais aos regressores, (AGUIRRE, 2007).

$$\langle \Psi \xi \rangle = 0. \quad (2.45)$$

A interpretação geométrica de tal condição pode ser expressa através da figura 2.3, em que pode-se observar que o valor de y obtido a partir do modelo, \hat{y} , pode ser decomposto em duas componentes, $\hat{\theta}_{p1}$ e $\hat{\theta}_{p2}$. A diferença entre \hat{y} e o vetor que representa o seu valor real, y , é dada por ξ . Assim, o vetor de medidas, representado pelo regressor, Ψ , deve ser perpendicular ao resíduo, ξ , para que seu produto interno possa ser nulo e a equação (2.45) possa ser atendida.

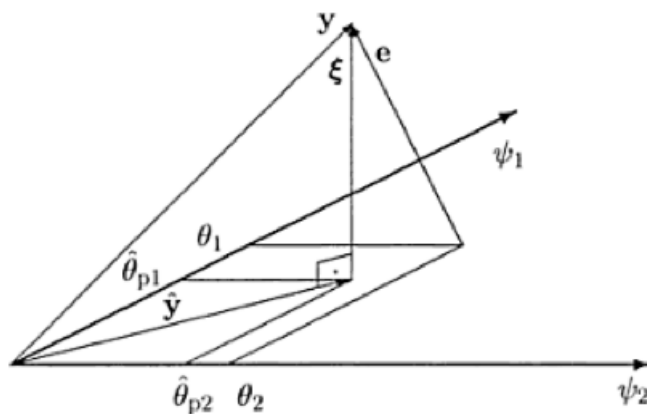


Figura 2.3: Ortogonalidade entre os regressores, Ψ , e os resíduos, ξ , (AGUIRRE, 2007).

Cabe ressaltar que os processos de eliminação de *bias* devem tornar os regressores perpendiculares ao resíduo, quando estes não o sejam.

2.6 O ESTIMADOR ESTENDIDO DE MÍNIMOS QUADRADOS

Na identificação de sistemas reais, nem sempre as hipóteses de Gauss-Markov, mostradas na seção 2.5.1, podem ser atendidas, e nestes casos tem-se um estimador polarizado, ou seja: $b \neq 0$. Uma das possibilidades para se contornar tal problema é a utilização do estimador estendido dos mínimos quadrados (MQE).

Como exemplo, considere o seguinte modelo ARX (autoregressivo com entrada exógena):

$$y(k) = ay(k-1) + bu(k) + e(k), \quad (2.46)$$

o ruído $e(k)$ pode ser equacionado do seguinte forma:

$$e(k) = cv(k-1) + v(k), \quad (2.47)$$

sendo v definido como ruído branco.

O desenvolvimento do estimador MQE segue os seguintes passos: aplica-se (2.47) em (2.46) a fim de obter:

$$y(k) = ay(k-1) + bu(k) + cv(k-1) + v(k), \quad (2.48)$$

ou ainda:

$$y(k-1) = ay(k-2) + bu(k-1) + \underline{cv(k-2)} + \underline{v(k-1)}. \quad (2.49)$$

Quando se substitui (2.49) em (2.48), chega-se a

$$y(k) = a[ay(k-2) + bu(k-1) + \underline{cv(k-2)} + \underline{v(k-1)}] + bu(k) + \underline{cv(k-1)} + v(k). \quad (2.50)$$

onde os termos são sublinhados apenas para destacar a presença de $v(k-1)$.

A equação (2.50) pode ser expressa na forma $y(k) = \Psi^T(k-1)\theta + e(k)$, com a seguinte estrutura:

$$\Psi^T(k-1) = [(ay(k-2) + bu(k-1) + cv(k-2) + \underline{v(k-1)}) \quad u(k-1)],$$

$$\theta = [a \quad b]^T,$$

$$e(k) = [c\underline{v(k-1)} + v(k)]. \quad (2.51)$$

Como os termos sublinhados são idênticos, pode-se perceber que os regressores, Ψ , estão correlacionados com o ruído, $e(k)$, não permitindo que a condição (2.45) seja atendida. Desta forma, se o operador MQ convencional fosse utilizado nesta situação, ele se encontraria polarizado.

Para resolver este problema, pode-se incluir a parte correlacionada do vetor de erro na matriz de regressores, (AGUIRRE, 2007). Assim, garante-se que Ψ não esteja correlacionado com $e(k)$.

$$\begin{bmatrix} y(k) \\ y(k+1) \\ \vdots \\ y(k+N-1) \end{bmatrix} = \begin{bmatrix} y(k-1) & u(k-1) \\ y(k) & u(k) \\ \vdots & \vdots \\ y(k+N-2) & u(k+N-2) \end{bmatrix} \dots$$

$$\dots \begin{bmatrix} v(k-1) \\ v(k) \\ \vdots \\ v(k+N-1) \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} + \begin{bmatrix} v(k) \\ v(k+1) \\ \vdots \\ v(k+N-1) \end{bmatrix}. \quad (2.52)$$

Assim, chega-se a seguinte forma para o estimador:

$$y^* = \Psi^* + e^*, \quad (2.53)$$

sendo: $y^* = y$, $e^* = [v(k) \dots v(k+N-1)]^T$ e

$$\Psi^* = \begin{bmatrix} \vdots & v(k-1) \\ \Psi & v(k) \\ \vdots & \vdots \\ \vdots & v(k+N-2) \end{bmatrix}.$$

Desta forma, chega-se a uma matriz de regressores, Ψ^* , descorrelacionada com o ruído e, conseqüentemente, tem-se $E\{\Psi^* e^*\} = 0$. Desta forma, pode se esperar que as estimativas dadas por (2.53) e (2.54) sejam não polarizadas.

$$\hat{\theta}^* = [\Psi^{*T} \Psi^*]^{-1} \Psi^{*T} y. \quad (2.54)$$

Cabe salientar que, para que tal procedimento seja realizado, é necessário que o modelo do ruído seja conhecido, o que implica em outro processo de estimação/identificação.

2.7 MÍNIMOS QUADRADOS ORTOGONAIS

O método dos mínimos quadrados constitui uma das ferramentas mais básicas nas áreas de estimação e identificação de sistemas, sendo extremamente útil devido a sua simplicidade e aplicabilidade em diferentes problemas, (AGUIRRE, 2007; LJUNG, 1999; LANDAU; ZITO, 2006). Entretanto, se Ψ possuir muitas colunas, o que normalmente ocorre na prática para sistemas não lineares, o uso da matriz pseudoinversa, equação (2.26), não resulta em boas estimativas para os parâmetros desejados, (AGUIRRE, 2007). Para contornar tais problemas, pode ser utilizado o método dos Mínimos Quadrados Ortogonais (MQO), que faz uso da propriedade que define que $\Psi^T \Psi$ é simétrico por construção. Para tanto, a matriz Ψ é fatorada na seguinte forma:

$$\Psi = QA, \quad (2.55)$$

onde: A é uma matriz triangular superior e Q é uma matriz com colunas ortogonais.

Para se chegar à matriz ortogonal Q pode-se recorrer ao método de Gram-Schmidt, que será abordado na próxima seção.

2.7.1 Método de Gram-Schmidt

O método de Gram-Schmidt tem por objetivo fatorar a matriz de regressores, Ψ , que é posto pleno, em uma matriz triangular, A , e uma matriz ortogonal Q , conforme (2.55). Este algoritmo calcula A , coluna por coluna, e ortogonaliza Ψ , da seguinte forma:

$$\begin{cases} q_1 = \psi_1, \\ \alpha_{j,i} = \frac{\langle q_i, \psi_j \rangle}{\langle q_i, q_j \rangle} \quad 1 \leq j < i, \\ q_i = \psi_i - \sum_{j=1}^{i-1} \alpha_{j,i} \cdot q_j, \end{cases} \quad (2.56)$$

considerando q_i a i -ésima coluna de Q e ψ_i a i -ésima coluna de Ψ .

Multiplicando-se a equação (2.29) por Ψ^T , tem-se:

$$\begin{aligned} y &= \Psi \hat{\theta} \\ \Psi^T y &= \Psi^T \Psi \hat{\theta}, \end{aligned} \quad (2.57)$$

aplicando (2.55) em (2.57) tem-se:

$$A^T Q^T y = A^T Q^T Q A \hat{\theta}. \quad (2.58)$$

Como Q é uma matriz com colunas ortogonais, tem-se que $Q^T Q = D$, sendo D uma matriz diagonal, definida positiva, (AGUIRRE, 2007; STRANG, 2013). Desta forma, tem-se:

$$\begin{aligned} A^T Q^T y &= A^T D A \hat{\theta} \\ Q^T y &= D A \hat{\theta} \\ D^{-1} Q^T y &= A \hat{\theta}, \end{aligned} \quad (2.59)$$

fazendo $\hat{g} = A \hat{\theta}$, tem-se:

$$\hat{g} = D^{-1} Q^T y. \quad (2.60)$$

Desta forma pode-se determinar os parâmetros de um sistema a partir da aplicação do método dos mínimos quadrados ortogonais, expresso pela equação (2.61):

$$\hat{\theta} = A^{-1} \hat{g}. \quad (2.61)$$

Alternativamente, pode-se determinar \hat{g} a partir de:

$$\hat{g}_i = \frac{\langle q_i, y \rangle}{\langle q_i, q_i \rangle}, \quad i = 1, 2, 3, \dots, n_\theta. \quad (2.62)$$

Com a aplicação de fatoração $\Psi = QR$, sendo Q uma matriz de colunas ortogonais e R uma matriz triangular superior, a matriz $\Psi^T \Psi$ torna-se igual a $R^T Q^T QR = R^T R$. Assim, o método dos mínimos quadrados clássico, $\Psi^T \Psi \hat{\theta} = \Psi^T y$, pode ser simplificado para $R \hat{\theta} = Q^T y$, (STRANG, 2013).

2.8 CONCLUSÃO

Neste capítulo foram abordados os conceitos de modelo e de sistema, bem como os conceitos iniciais necessários à Identificação de Sistemas. Foram abordadas a identificação de sistemas baseada em deconvolução, em correlação e o método dos mínimos quadrados. Foram exploradas ainda as hipóteses de Gauss-Markov, necessárias a um estimador livre de viés. Também foram apresentadas as versões estendidas do método dos mínimos quadrados, como Mínimos Quadrados Estendidos e Ortogonais, para as situações em que as hipóteses de Gauss-Markov não são válidas.

3 MODELOS NÃO LINEARES

Neste capítulo serão apresentados modelos para diferentes sistemas. Serão abordados o espaço de Hilbert, na seção 3.2 e os modelos FIR e ARX na seção 3.4. As séries de Volterra serão expostas na seção 3.5 e os modelos NARMAX na seção 3.6. As funções ortonormais são apresentadas em 3.7, sendo as funções de Laguerre na seção 3.7.1 e as funções de Kautz em 3.7.2. E finalmente, o modelo Volterra-OBF será apresentado na seção 3.8, assim como os modelos NARX-OBF na seção 3.9.

3.1 SISTEMAS NÃO LINEARES

A linearidade é uma característica que facilita os processos de controle e de identificação de um sistema. Entretanto, não são muitos os casos em que um sistema real pode ter sua dinâmica aproximada por um sistema linear, (HENDRICK; GIRARD, 2010).

Um sistema, L , de entrada $u \in R^n$ e saída $y \in R^m$ é dito linear se satisfaz o princípio da superposição, (MONTEIRO, 2003). Para melhor definir tal princípio, considere os escalares a_1 e $a_2 \in \mathbb{R}$, as entradas $u_1[k]$, $u_2[k]$, as saídas $y_1[k]$ e $y_2[k]$, relativas ao sistema L para $k \in \mathbb{Z}$, considerando este sistema discreto ou contínuo discretizado. Assim, pode-se expressar o princípio da superposição da seguinte forma:

1. Se para uma entrada $u_1[k]$ obtém-se uma saída $y_1[k]$;

Então, para uma entrada $a_1x_1[k]$ obtém-se uma saída $a_1y_1[k]$;

2. Se para uma entrada $x_2[k]$ obtém-se uma saída $y_2[k]$;

Então, para uma entrada $a_2x_2[k]$ obtém-se uma saída $a_2y_2[k]$;

3. Assim, para uma entrada $a_1x_1[k] + a_2x_2[k]$ obtém-se uma saída $a_1y_1[k] + a_2y_2[k]$.

De forma genérica, pode-se dizer que aplicada uma entrada, $\sum_{n=0}^{N-1} a_nx[n]$, a um sistema linear, obtém-se uma saída $\sum_{n=0}^{N-1} a_ny[k]$. Sendo a_n escalares reais e $x[k]$ e $y[k]$ sinais de entrada e saída do sistema linear L .

Cabe ressaltar que o princípio da superposição expressa a conjunção entre duas propriedades: a Homogeneidade, mostrada nos itens 1 e 2, e a Aditividade, mostrada no item 3, (MONTEIRO, 2006). Desta forma, entende-se como não linear o conjunto de todos os sistemas que não satisfazem as condições de superposição, (HENDRICK; GIRARD, 2010).

3.2 ESPAÇO DE HILBERT

Um espaço de Hilbert \mathcal{H} é um espaço vetorial, de dimensão infinita, munido de um produto interno, e completo em relação à norma definida por este produto interno, (LIMA, 2013c). Denota-se por $L^2(S)$ o espaço de Hilbert de funções quadraticamente integráveis em $S \subset \mathbb{R}$ com produto interno (MAESTRELLI, 2010):

$$\langle f(t), g(t) \rangle = \int_S f(t)g^*(t)dt, \quad (3.1)$$

sendo $*$ a operação responsável por conjugar um argumento complexo. E define-se como quadraticamente integráveis as funções que atendem a seguinte condição:

$$\int_S \|f(t)\|^2 dt < \infty. \quad (3.2)$$

O espaço de Hilbert é um meio de se tratar espaços com dimensões infinitas, e, ao mesmo tempo, se manter a geometria do espaço Euclidiano comum, (STRANG, 2013).

Neste contexto, a norma de uma função pertencente ao espaço de Hilbert é dada por:

$$\|f(t)\| = \sqrt{\langle f(t), f(t) \rangle} = \sqrt{\int_S f(t)^2 dt}. \quad (3.3)$$

Para que se possa expor uma das principais propriedades do espaço de Hilbert, é necessário definir uma sequência de Cauchy. Assim sendo, uma sequência (x_n) é chamada sequência de Cauchy quando, para todo $\varepsilon > 0$ dado, existe $n_0 \in \mathbb{N}$ tal que $m, n > n_0 \Rightarrow d(x_m, x_n) < \varepsilon$, (LIMA, 2013c), sendo $d(x_m, x_n)$ a distância entre x_m e x_n . Em outras palavras, os termos da sequência vão se tornando cada vez mais próximos uns dos outros conforme o índice n avança.

Uma característica importante do espaço de Hilbert é que a métrica deste espaço é fechada, ou seja, qualquer sequência de Cauchy das funções converge para uma função do espaço, (MAESTRELLI, 2010).

Um conjunto de funções, $\{\psi_i\}_{i=1}^{\infty}$, do espaço de Hilbert, \mathcal{H} , é considerado ortonormal se as funções deste conjunto forem ortogonais entre si e cada uma das funções do conjunto possuir norma unitária, (MAESTRELLI, 2010).

O conjunto $\{\psi_i\}_{i=1}^{\infty}$ é completo em \mathcal{H} , se e somente se cada função de \mathcal{H} puder ser

arbitrariamente aproximada por uma combinação linear de um número finito de funções do citado conjunto, (MAESTRELLI, 2010).

3.3 PRINCIPAIS MODELOS NÃO LINEARES

Com inspiração no conceito de espaço vetorial da Álgebra Linear, as áreas de Controle e Identificação de Sistemas vêm buscando uma base vetorial \mathfrak{B} que pudesse explorar um espaço de funções dinâmicas, capazes de modelar a dinâmica de diferentes sistemas, (HEUBERGER; HOF; BOSGRA, 1995).

Para melhor definir \mathfrak{B} , considere o conjunto \mathfrak{D}^n composto pelas funções de transferência de ordem menor que n , assim, tem-se que $\mathfrak{B} \in \mathfrak{D}^n$ e os modelos gerados seriam combinações lineares e não lineares dos vetores de \mathfrak{B} . Tal base vetorial seria útil principalmente para modelagem de sistemas não lineares, como descrito pela equação (2.1):

$$\dot{x} = F(t, x, u),$$

em que $F : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$. As variáveis do vetor $x^T = [x_1(t), \dots, x_n(t)]$ são as variáveis de estado, e o vetor $u^T = [u_1(t), \dots, u_m(t)]$ é o vetor de variáveis de controle, (BAUMEISTER; AO, 2014).

Como os sistemas não lineares não são uma única classe de sistemas, com propriedades comuns bem definidas, mas uma coleção de todos os sistemas que não possuem a linearidade como característica, (MONTEIRO, 2003), não é fácil encontrar uma base de funções que sejam suficientemente genéricas para modelar uma ampla gama de sistemas não lineares.

Entre as candidatas mais promissoras a compor a base vetorial \mathfrak{B} , encontram-se as funções de Laguerre, de Kautz e as *GOBFs* - *Generalized Orthornormal Basis Functions*, (HEUBERGER; HOF; WAHLBERG, 2005). Neste cenário, este capítulo aborda a utilização de funções ortonormais discretas de Laguerre e Kautz, implementadas sob a forma de filtros digitais, para a identificação de sistemas.

A fim de utilizar as funções ortonormais para composição de modelos de sistema, é necessário a utilizar métodos de otimização para se encontrar os polos de parametrização e os coeficientes destas funções. Métodos clássicos de otimização como o Gradiente Descendente e de Newton podem executar tal tarefa, porém necessitam de uma modelagem bem detalhada do sistema. Geralmente, é necessário o levantamento das derivadas bem como das matrizes Jacobianas e Hessianas do sistema, o que não se tem a priori, se o objetivo do processo for identificar o sistema.

Para contornar os problemas de otimização determinística, geralmente são utilizados os

métodos heurísticos, que podem obter resultados próximos do ótimo sem a necessidade de obtenção prévia das derivadas do sistema. Os métodos de busca determinística dos parâmetros serão tratados no capítulo 4, já os métodos de busca heurística serão expostos no capítulo 5.

3.4 MODELOS FIR E ARX

Considere um sistema discreto, representado por $y(k) = F(k, u[n-k])$, com entrada $u(k) \in \mathbb{R}$ e saída $y(k) \in \mathbb{R}$, sendo k e $n \in \mathbb{N}$. Neste cenário, os modelos *Finite Impulse Response*, FIR, são derivados de modelos com seguinte estrutura:

$$y(k) = \sum_{n=1}^{\infty} c_n u(k-n), \quad (3.4)$$

sendo que os coeficientes c_n assumem valores reais. A partir de um modelo com esta estrutura, pode-se modelar uma ampla gama de sistemas. Cabe salientar que a série envolve infinitos termos, entretanto, na prática o somatório é truncado em um valor finito de N , em que a exatidão necessária ao modelo é atingida.

$$y(k) = \sum_{n=1}^N c_n u(k-n). \quad (3.5)$$

No espectro, fazendo uso da transformada \mathcal{Z} em (3.4), pode-se afirmar que: se um sistema linear puder ser expresso por uma função de transferência $G(z) = Y(z)/U(z)$, então tal sistema pode ser expresso como, (HEUBERGER; HOF; BOSGRA, 1995):

$$G(z) = \sum_{n=1}^{\infty} c_n z^{-n}, \quad (3.6)$$

esta modelagem é conhecida como expansão de Laurent e foi muito utilizada nas décadas passadas como modelo de sistema. Uma vez que não há realimentação da saída, este modelo é sempre estável, (ISERMANN; MUNCHHOF, 2011).

Como generalização dos modelos FIR, os modelos ARX (*AutoRegressive with eXogenous input*) são constituídos por combinações lineares tanto das entradas quanto das saídas de um sistema, como expresso em:

$$y(k) = \sum_{n=1}^{N_y} a_n y(k-n) + \sum_{n=1}^{N_u} b_n u(k-n) + e(k), \quad (3.7)$$

em que $e(k)$ é o termo associado ao ruído.

Apesar da simplicidade e da grande utilização dos modelos polinomiais, tanto os modelos FIR como os ARX resultam em parâmetros inconsistentes, para a maioria dos sistemas práticos, (LEMMA; RAMASAMY; SHUHAIMI, 2010).

Quando um modelo é inconsistente, existirá um erro sistemático (*bias*) no modelo estimado que não poderá ser removido com o aumento do número de pontos, aumento da dimensão da matriz de medidas, (LEMMA; RAMASAMY; SHUHAIMI, 2010).

A fim de solucionar o problema da inconsistência, foram propostas algumas classes de modelos diferentes. Dentre estas, as de interesse para este trabalho serão tratadas nas próximas seções.

3.5 SÉRIES DE VOLTERRA

As séries de Volterra são funcionais, expressos através de séries infinitas, que generalizam o conceito de convolução para sistemas não lineares, (ISERMANN; MUNCHHOF, 2011). As séries de Volterra tornam possível usar alguns conceitos, definidos somente para os sistemas lineares, para representar também os processos não-lineares, (EYKHOFF, 1974). Pode-se observar sua estrutura, para um sistema causal, na equação a seguir:

$$y(k) = h_0 + \sum_{t_1=1}^{\infty} h_1(t_1)u(k-t_1) + \sum_{t_1=1}^{\infty} \sum_{t_2=1}^{\infty} h_2(t_1, t_2)u(k-t_1)u(k-t_2) + \dots + \sum_{t_1=1}^{\infty} \dots \sum_{t_n=1}^{\infty} h_n(t_1, \dots, t_n)u(k-t_1) \dots u(k-t_n) + \dots, \quad (3.8)$$

sendo que as funções $h_n(t_1, \dots, t_n)$ são os *kernels* de Volterra de ordem n e h_0 é um termo responsável por representar o nível constante não nulo de $y(k)$.

Desconsiderando-se o termo h_0 , nível contínuo, pode-se chegar a:

$$y(k) = \sum_{t_1=1}^{\infty} \sum_{t_2=1}^{\infty} \cdots \sum_{t_n=1}^{\infty} h_n(t_1, \dots, t_n) \prod_{l=1}^n u(k-t_l). \quad (3.9)$$

Um modelo de Volterra descreve a saída de um sistema não-linear como a soma das respostas dos operadores de primeira ordem, de segunda ordem, de terceira ordem, e assim por diante, (ROSA; CAMPELLO; AMARAL, 2009):

$$y(k) = \mathcal{H}_1[u(k)] + \mathcal{H}_2[u(k)] + \cdots + \mathcal{H}_\eta[u(k)] + \cdots. \quad (3.10)$$

Cabe ressaltar que os componentes não lineares, na maioria das vezes, desempenham um papel secundário em comparação com o componente linear (*kernel* de primeira ordem), que é dominante no sistema, (NEMETH; KOLLAR, 2002). Assim, η pode ser estabelecido como um valor finito em que se trunca a aproximação. Em Boyd and Chua, 1985, (BOYD; CHUA, 1985), foi mostrado que qualquer sistema não linear, causal, invariante no tempo, com memória finita, pode ser representado por uma expansão finita de uma série de Volterra.

A estrutura de um modelo de Volterra pode ser observada na figura 3.1.

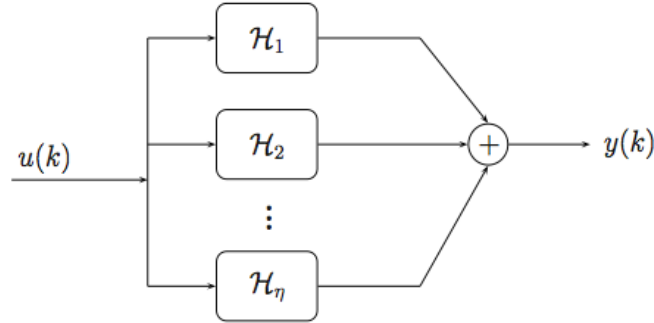


Figura 3.1: Esquemático de sistema representado por uma Série de Volterra, (ROSA, 2009).

Na figura 3.1, os operadores de ordem n , \mathcal{H}_n , são matematicamente descritos pela equação:

$$\mathcal{H}_n[u(k)] = \sum_{t_1=1}^{\infty} \cdots \sum_{t_n=1}^{\infty} h_n(t_1, \dots, t_n) u(k-t_1) \cdots u(k-t_n). \quad (3.11)$$

A saída de cada bloco pode ser escrita na forma vetorial. Assim, o bloco de primeira ordem pode ser expresso por:

$$y_1(k) = h_1^T(k)u_1(k), \quad (3.12)$$

sendo o vetor $u_1(k)$ a entrada de primeira ordem em um modelo de Volterra, e a resposta ao impulso do sistema, h_1 , com seus elementos armazenados em forma de um vetor unidimensional.

Para os blocos de ordem elevada, pode-se definir algumas operações que permitem a generalização da equação (3.12) para ordens mais elevadas.

Para definir uma entrada u de ordem p , considere a seguinte operação:

$$u_p(k) = u_1(k) \otimes u_{p-1}(k), \quad (3.13)$$

sendo \otimes um produto de Kronecker e $p \in \mathbb{N}$. Assim, pode-se definir uma entrada generalizada u_v a partir de:

$$u_v(k) = [u_1(k)^T \quad u_2(k)^T \quad \cdots \quad u_p(k)^T]. \quad (3.14)$$

De forma semelhante, faz-se necessário organizar as respostas ao impulso generalizadas, de modo que a saída de um sistema, $y(k)$, possa ser expressa a partir de um produto interno entre suas entradas generalizadas, $u_v(k)$, e as suas respostas ao impulso generalizadas, $h_v(k)$. Por simplicidade, considere um *kernel* de segunda ordem:

$$h_2 = [h_2(1,1) \quad h_2(1,2) \cdots h_2(1,N) \quad \cdots \quad h_2(2,1) \quad h_2(2,2) \cdots h_2(2,N) \quad \cdots \\ h_2(N,1) \quad h_2(N,2) \cdots h_2(N,N)]^T. \quad (3.15)$$

Pode-se notar que tal operação converte uma matriz $N \times N$ em um vetor de \mathbb{R}^n , sendo $n = N^2$. Esta conversão ocorre através da concatenação das colunas da matriz, resultando em um vetor coluna. Pode-se generalizar o processo de geração de h_2 para h_p pelo mesmo processo.

A partir da definição exposta em (3.15), pode-se chegar a uma resposta ao impulso generalizada, no formato adequado ao produto interno exposto em (3.12):

$$h_v = [h_1^T \quad h_2^T \cdots h_p^T]^T. \quad (3.16)$$

Assim, pode-se chegar a equação que permite determinar de forma vetorial a saída, $y_v(k)$, de um sistema não linear a partir de um modelo de Volterra.

$$y_v(k) = h_v^T(k)u_v(k), \quad (3.17)$$

para $v \in \mathbb{N}$.

Observa-se, a partir da equação (3.17), que a interação entre uma entrada, $u(k)$, e um sistema não linear, representado por um modelo de Volterra, pode ser representada por um produto interno, (BATISTA, 2009).

A série dada pela equação (3.10) representa uma generalização do conceito de resposta ao impulso, (ROSA; CAMPELLO; AMARAL, 2009). Desta forma, aos sistemas lineares (A, B, C) pode-se aplicar apenas o operador de primeira ordem da série de Volterra, resultando:

$$y(k) = \sum_{t_1}^{\infty} h_1(t_1)u(k - t_1), \quad (3.18)$$

tal equação é conhecida como equação de convolução e tem seu escopo de aplicação restrito aos sistemas lineares.

Como o termo $\mathcal{H}_n[u(k)]$ de (3.10) contém contribuições de grau n para a entrada $u(k)$ tem-se que o modelo de Volterra é uma generalização não linear polinomial do modelo linear (3.18), (ROSA; CAMPELLO; AMARAL, 2009).

Segundo Eykhoff, (EYKHOFF, 1974), os *kernels* de Volterra possuem 3 propriedades a se destacar:

- $h_n(t_1, t_2, \dots, t_n)$ podem ser construídos de forma a serem simétricos;
- $\lim_{t_i \rightarrow \infty} h_n(t_1, t_2, \dots, t_n) = 0$, para $i = 1, \dots, n$, expressando a estabilidade dos *kernels*;
- $h_n(t_1, t_2, \dots, t_n) = 0 \quad \forall t_i < 0$, para $i = 1, \dots, n$, expressando a causalidade dos *kernels*.

Um modelo de Volterra pode ser interpretado como um modelo de Wiener no qual os termos são descritos a partir de funções ortonormais, z^{-n} , (SILVA, 2010), onde: $z \in \mathbb{C}$ e $n \in \mathbb{N}$.

As principais vantagens dos modelos de Volterra são: a capacidade de descrição do modelo e a não necessidade de realimentação das saídas, $y(k)$. A principal desvantagem desta classe de modelos é o grande número de parâmetros a se determinar, e o grande número de termos necessários na série para que se possa aproximar com precisão um dado sistema, (SILVA, 2010). Cabe ressaltar que modelos de Volterra não podem modelar comportamentos como histerese e caos, (NEMETH; KOLLAR, 2002).

Como extensão das Séries de Volterra tem-se os modelos NARMAX polinomiais, que serão abordados na próxima seção.

3.6 NARMAX

Os modelos NARMAX¹ (*Nonlinear AutoRegressive Moving Average with eXogenous input*) descrevem um sistema a partir de um funcional não linear, composto por entradas, saídas e ruídos estimados, sendo estes termos defasados no tempo, (RAHROOH; SHEPARD, 2009).

$$\begin{aligned}
 y(k) = f^l(u(k), u(k-1), u(k-2), \dots, u(k-n_u), \\
 y(k-1), y(k-2), y(k-3), \dots, y(k-n_y) \\
 e(k), e(k-1), e(k-2), \dots, e(k-n_e)), \quad (3.19)
 \end{aligned}$$

em que $y(k)$, $u(k)$ e $e(k)$ são os sinais de saída, entrada e ruído estimado, amostrados no instante de tempo k , respectivamente. Já n_u , n_y e n_e são as ordens de regressão dos vetores de entrada, saída e ruído. A função $f^l(\cdot)$ é uma função não linear dos seus argumentos, com grau de não linearidade $l \in \mathbb{N}$. Na implementação polinomial do modelo NARMAX, o grau de não linearidade, l , se define como o número máximo de termos, $u(k)$, $y(k)$ ou $e(k)$, multiplicados entre si, na função $f^l(\cdot)$.

Em uma simplificação deste modelo, em que o ruído não aparece nos regressores, pode-se representar a função $f^l(\cdot)$ pela seguinte estrutura:

$$y(k) = \sum_{m=0}^l \sum_{p=0}^m \sum_{n_1, n_m}^{n_y, n_u} h_{p, m-p} \prod_{i=1}^p y(k-n_i) \prod_{i=p+1}^m u(k-n_i), \quad (3.20)$$

sendo $h_{p, m-p}(n_1, \dots, n_m)$ os *kernels* do modelo e

$$\sum_{n_1, n_m}^{n_y, n_u} = \sum_{n_i=1}^{n_y} \dots \sum_{n_m}^{n_u}.$$

A versão dos modelos NARMAX sem os termos associados ao ruído é conhecida como NARX (*Nonlinear AutoRegressive with eXogenous input*).

Os modelos NARMAX possibilitam a representação para uma grande classe de sistemas

¹Neste trabalho descrevem-se modelos NARMAX polinomiais, existem, porém inúmeras outras opções. Por exemplo, na *toolbox* do software Matlab, se utilizam redes neurais como função $f(\cdot)$ e combinação entre entradas e saídas, $u(k)$ e $y(k)$, realizadas através de funções $\tanh(x)$.

não lineares e possuem vantagens sobre as séries de Volterra no quesito de precisão, (RAHROOH; SHEPARD, 2009). Na verdade, tais modelos, na sua versão polinomial, podem ser considerados como uma generalização dos modelos de Volterra, nos quais a série passa a ser composta também pelos valores do erro, $e(k)$, e da saída, $y(k)$, do sistema a ser identificado.

Tanto os modelos de Volterra quanto os modelos NARMAX sofrem com grande número de termos necessários à série, para que o modelo represente adequadamente um sistema a ser identificado. Entretanto, este problema é menos crítico nos modelos NARMAX, uma vez que neste as saídas passadas, $y(k-n)$ e o erro estimado, $e(k)$, entram no equacionamento. Esta limitação, necessidade de um número elevado de termos no modelo, é conhecida como problema da dimensionalidade. Neste contexto, mesmo com o elevado peso computacional necessário em sua implementação, é possível se obter modelos NARMAX polinomiais que ajustem dados com boa exatidão, desde que estes dados não apresentem variações abruptas, (AGUIRRE, 1998).

Devido ao elevado número de possibilidades de combinação de regressores para a formação da matriz de regressão, Ψ , na prática não é usual trabalhar com conjuntos completos de regressores, (AGUIRRE, 2007). Geralmente recorre-se a alguma métrica de desempenho para seleção dos regressores mais relevantes na modelagem do comportamento do sistema a ser descrito. Em outras palavras, o primeiro passo na modelagem de um sistema a partir de modelos NARMAX é a seleção das combinações de entradas e saídas mais apropriadas ao modelo, (AKANYETI et al., 2009).

Uma das medidas possíveis para a redução do custo computacional da implementação de modelos NARMAX é mostrada na seção 6.5, em que os termos mais relevantes da série são selecionados. Uma outra possibilidade é a utilização de funções ortonormais para representar os *kernels* destes modelos. Na próxima seção serão apresentadas as principais funções ortonormais.

3.7 FUNÇÕES ORTONORMAIS

As funções ortonormais podem ser interpretadas como vetores, em espaços vetoriais de funções, que apresentam produto interno nulo entre si, (STRANG, 2013). Tal condição é expressa em (3.21):

$$\langle \psi_p(k), \psi_q(k) \rangle = \begin{cases} 0 & p \neq q \\ 1 & p = q, \end{cases} \quad (3.21)$$

sendo p e $q \in \mathbb{N}$. Ainda, $\psi_p(k)$ e $\psi_q(k)$ são funções ortonormais e $\langle \cdot \rangle$ é o produto interno, adequado a tais funções, definido pela equação (3.22), segundo (HEUBERGER; HOF; WAHL-

BERG, 2005):

$$\langle \psi_p(z), \psi_q(z) \rangle = \frac{1}{2\pi i} \oint_c \psi_p(z) \psi_q^*(1/z^*) \frac{dz}{z}, \quad (3.22)$$

em que i é a unidade imaginária e c é o círculo de raio unitário: $c = \{z \in \mathbb{C} : |z| \leq 1\}$. Outra possibilidade de produto interno pode ser definida da seguinte forma:

$$\langle \psi_p(k), \psi_q(k) \rangle = \sum_{k=0}^{\infty} \psi_p(k) \psi_q(k). \quad (3.23)$$

E assim pode-se definir a norma de uma função por:

$$\|\psi_n(k)\| = \sqrt{\langle \psi_n(k), \psi_n(k) \rangle}. \quad (3.24)$$

Assim, as funções ortonormais atenderão aos seguintes requisitos:

- $\psi_p(k) \perp \psi_q(k)$ para $p \neq q$, ou seja, as funções serão ortogonais entre si;
- $\|\psi_n(k)\| = 1, \forall n$, isto é, as funções apresentarão norma unitária.

A ortogonalidade é interessante para a base de funções, pois além de as funções serem linearmente independentes, elas também possuem projeções nulas umas sobre as outras. Assim, tais funções são indicadas para compor a base vetorial, \mathfrak{B} , responsável por explorar o espaço vetorial composto por funções dinâmicas, que representam determinadas classes de sistemas, (MACHADO; CAMPELLO; AMARAL, 2010).

Neste contexto, as funções OBF (*Orthonormal Basis Function*), $\psi(k)$, são funções ortonormais que integram as bases vetoriais para um espaço vetorial de Hilbert, L^2 . Tal espaço vetorial é conhecido como espaço de Hilbert em \mathbb{N} e é definido como o espaço das funções discretas, $\psi(k)$, quadraticamente somáveis, (BELT, 1997; MAESTRELLI, 2010), ou seja, funções que atendem a equação (3.25).

$$\sum_{k=1}^{\infty} \psi^2(k) \leq \infty. \quad (3.25)$$

A ideia da representação OBF em modelos dinâmicos é desenvolver a resposta ao impulso, $h(k)$, de um determinado sistema através de uma base de funções ortonormais, $\psi(k)$, (CAMPELLO; AMARAL; OLIVEIRA, 2007). Devido à completude das bases ortonormais, é possível aumentar a capacidade de representação apenas aumentando o número de funções

ortonormais da base, (HEUBERGER; HOF; WAHLBERG, 2005). Assim, a resposta ao impulso de um sistema pode ser aproximada com precisão arbitrária por:

$$\hat{h}(k) = \sum_{j=1}^n c_j \psi_j(k), \quad (3.26)$$

ou utilizando a transformada $Z\{\cdot\}$ chega-se a função de transferência do citado sistema:

$$\hat{H}(z) = \sum_{j=1}^n c_j \Phi_j(z). \quad (3.27)$$

Para um sistema linear, pode-se determinar a saída do sistema, $y(k)$, pela convolução de $\hat{h}(k)$ e uma determinada entrada, $u(k)$, como expresso em (3.28):

$$\hat{y}(k) = \sum_{i=0}^{\infty} \hat{h}(i) u(k-i). \quad (3.28)$$

Aplicando a equação (3.26) em (3.28), com o somatório truncado em m termos, obtém-se:

$$\begin{aligned} \hat{y}(k) &= \sum_{i=0}^m \sum_{j=1}^n c_j \psi_j(k) u(k-i) \\ &= \sum_{j=1}^n c_j \sum_{i=0}^m \psi_j(k) u(k-i). \end{aligned}$$

Considerando $\Psi_i(k)$ como a convolução entre $u(k)$ e a i -ésima função ortonormal $\psi(k)$, $\Psi_j(k) = \sum_{i=0}^m \psi_j(k) u(k-i)$, tem-se:

$$\hat{y}(k) = \sum_{j=0}^n c_j \Psi_j(k), \quad (3.29)$$

sendo que as $\psi_n(z)$ funções ortonormais compõem uma base \mathfrak{B} para o espaço de funções \mathfrak{D}^n :

$$\mathfrak{B} : [\psi_1(z), \psi_2(z), \dots, \psi_n(z)]. \quad (3.30)$$

Neste cenário, considere um sistema linear, SISO, discreto, expresso pela função de

transferência $H(z)$. Considere ainda a representação deste sistema a partir de uma combinação linear de funções ortonormais:

$$H(z) = \sum_{i=1}^N c_i \psi_i(z). \quad (3.31)$$

Para a determinação dos coeficientes, c_i , considere a multiplicação da equação (3.31) por ψ_j :

$$H(z) \psi_j(z) = \sum_{i=1}^N c_i \psi_i(z) \psi_j(z). \quad (3.32)$$

Aplicando-se o produto interno em ambos os lados da equação (3.32) tem-se:

$$\langle H(z), \psi_j(z) \rangle = \sum_{i=1}^N c_i \langle \psi_i(z), \psi_j(z) \rangle.$$

Uma vez que as funções são ortogonais, o produto interno $\langle \psi_i, \psi_j \rangle$ só possui valores não nulos para $i = j$ e o somatório pode ser ignorado. E como as funções ψ_i e ψ_j são normalizadas, $\|\psi_i\| = \|\psi_j\| = 1$, tem-se a determinação dos coeficientes, c_i , dada por:

$$c_i = \langle H(z), \psi_j(z) \rangle. \quad (3.33)$$

Existe uma infinidade de funções ortonormais, de acordo com a definição de produto interno utilizada. As principais são: as funções de Hermite, Jacobi, Laguerre, Legendre, Kautz e as Funções Ortonormais Generalizadas, em inglês: GOBFs, (BELT, 1997). Algumas das funções ortonormais mais utilizadas na área de Identificação de Sistemas são apresentadas nas próximas seções.

3.7.1 Funções de Laguerre

As funções de Laguerre são funções ortonormais, parametrizadas por um polo real, (LEMMA; RAMASAMY; SHUHAIMI, 2010), e podem ser expressas pela equação:

$$L_n(z) = \frac{\sqrt{1-p^2}}{z-p} \left(\frac{1-pz}{z-p} \right)^{n-1}, \quad (3.34)$$

sendo $p \in \mathbb{R}$ o polo das funções de Laguerre e z a variável complexa associada à transformada $\mathcal{Z}\{\cdot\}$.

A funções da base de Laguerre podem ser implementadas em forma de filtros digitais, sendo cada função associada a um filtro como exposto na figura 3.2.

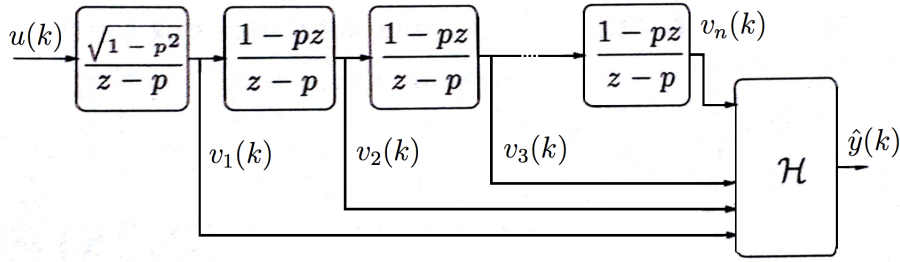


Figura 3.2: Modelo OBF com dinâmica de Laguerre, (CAMPELLO; AMARAL; OLIVEIRA, 2007).

No caso de sistemas Lineares e Invariantes no Tempo (LIT), o operador \mathcal{H} pode ser definido como apenas a combinação linear das saídas dos filtros de Laguerre v_1, v_2, \dots, v_n .

$$\begin{aligned} y_{nL}(k) &= c_1 v_1(k) + c_2 v_2(k) + \dots + c_n v_n(k) \\ &= \sum_{i=1}^n c_i v_i(z), \end{aligned} \quad (3.35)$$

sendo \hat{y} a saída estimada e $v_i(k)$ a entrada $u(k)$ filtrada pelo filtro de Laguerre de i -ésima ordem, $L_i(z)$, para $i = 1, 2, \dots, n$.

A base de funções de Laguerre é caracterizada pela utilização de funções de transferência com apenas um polo real, (HEUBERGER; HOF; WAHLBERG, 2005). Além disso, as funções de Laguerre podem ser caracterizadas como modelos FIR com polos fora da origem, sendo estes um caso particular de base de Laguerre, (CAMPELLO; AMARAL; OLIVEIRA,

2007). Normalmente a base FIR requer um número elevado de termos para representar sistemas com dinâmica predominantemente lenta, (MACHADO; CAMPELLO; AMARAL, 2010).

A base de Laguerre é apropriada para representar sistemas com polos puramente reais ou com parte imaginária de valor reduzido, (MACHADO; CAMPELLO; AMARAL, 2010). Outro ponto relevante é que em modelos que utilizam a abordagem OBF não existe realimentação da saída, isto é, não existe realimentação dos erros de previsão, o que normalmente resulta em modelos mais precisos, (NELLES, 2001).

Os modelos OBF permitem a incorporação do conhecimento prévio da dinâmica do sistema ao modelo, e o faz capturando a dinâmica do sistema com apenas um pequeno número de parâmetros, (LEMMA; RAMASAMY; SHUHAIMI, 2010). Assim sendo, tem-se uma redução da ordem de representação do modelo, simplificando os problemas de identificação e controle associados, (MACHADO; CAMPELLO; AMARAL, 2010).

Outra forma de representar um modelo de Laguerre é através de espaço de estados:

$$\begin{cases} \Psi(k+1) = A_f \Psi(k) + B_f u(k) \\ \hat{y}(k) = C_f \Psi(k), \end{cases} \quad (3.36)$$

sendo A_f e B_f dependentes dos polos do modelo, C_f contém os termos da expansão da série e $\Psi(k) = [\psi_1(k), \dots, \psi_{n_f}(k)]^T$ é o vetor de estados ortonormais. Cabe ressaltar que a ortonormalidade é uma propriedade das funções ortonormais e não dos estados em si, (MACHADO; CAMPELLO; AMARAL, 2010), mas como o vetor de estados é representado por tais funções, por abuso de linguagem, se diz que os estados são ortonormais.

3.7.2 Funções de Kautz

As funções de Kautz são funções ortonormais, parametrizadas por polo complexo, (LEMMA; RAMASAMY; SHUHAIMI, 2010), e podem ser expressas pelas equações (3.37) e (3.38):

$$K_{2m}(z) = \frac{z\sqrt{(1-c^2)(1-b^2)}}{z^2 + b(c-1)z - c} \left[\frac{-cz^2 + b(c-1)z + 1}{z^2 + b(c-1)z - c} \right]^{m-1}, \quad (3.37)$$

$$K_{2m-1}(z) = \frac{z(z-b)\sqrt{1-c^2}}{z^2 + b(c-1)z - c} \left[\frac{-cz^2 + b(c-1)z + 1}{z^2 + b(c-1)z - c} \right]^{m-1}, \quad (3.38)$$

sendo $K_{2m}(z)$ e $K_{2m-1}(z)$ as funções pares e ímpares de Kautz, respectivamente. Já p e \bar{p} são os polos, complexos conjugados, que parametrizam as funções e os termos b e c podem ser expressos por (3.39) e (3.40):

$$b = (p + \bar{p}) / (1 + p\bar{p}), \quad (3.39)$$

$$c = -p\bar{p}. \quad (3.40)$$

O agrupamento de n funções de Kautz pode formar uma base vetorial, K , num espaço de funções $L_2(\mathbb{N}^*)$. Esta base é conhecida como base de Kautz e pode representar sistemas lineares da seguinte forma:

$$\begin{aligned} y_{nK}(k) &= c_1 w_1(k) + c_2 w_2(k) + \dots + c_n w_n(k) \\ &= \sum_{i=1}^n c_i w_i(k), \end{aligned} \quad (3.41)$$

sendo $w_i(k)$ a entrada $u(k)$ filtrada pelo filtro de Kautz de i -ésima ordem, $K_i(z)$, para $i = 1, 2, \dots, n$.

3.8 MODELO VOLTERRA-OBF

Os modelos de Volterra-OBF são modelos de Volterra com seus *kernels* expressos por funções ortonormais. Assim, em vez de se determinar cada um dos elementos dos *kernels*, apenas se determinam os coeficientes das funções ortonormais e os polos que as parametrizam, (CAMPELLO; AMARAL; OLIVEIRA, 2007).

Desta forma, a regra de formação da base de funções ortonormais de Kautz pode ser observada na figura (3.3).

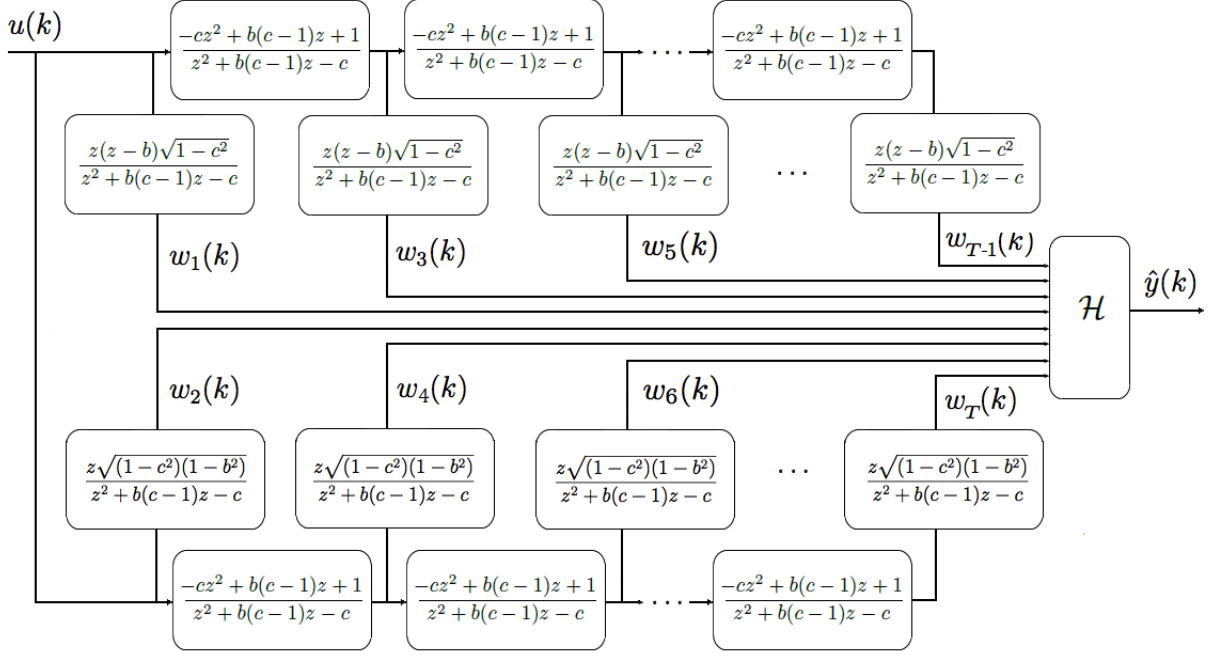


Figura 3.3: Modelo Volterra-OBF, implementado a partir de funções de Kautz.

Na figura 3.3, o mapeamento $\mathcal{H}\{\cdot\}$ é uma função não linear de $w_1(k), w_2(k), \dots, w_n(k)$. Salienta-se o fato de que estes elementos correspondem à entrada $u(k)$ filtrada em cada uma das direções perpendiculares.

Expandindo o operador $\mathcal{H}\{\cdot\}$ para uma série de Volterra, tem-se uma equação com a seguinte forma:

$$\begin{aligned}
y_{nK}(k) &= c_1 w_1(k) + c_2 w_2(k) + \dots + c_n w_n(k) + \\
&+ c_{1,1} w_1(k) w_1(k) + c_{1,2} w_1(k) w_2(k) + c_{1,3} w_1(k) w_3(k) + \dots + c_{1,n} w_1(k) w_n(k) + \\
&+ c_{2,2} w_2(k) w_2(k) + c_{2,3} w_2(k) w_3(k) + \dots + c_{2,n} w_2(k) w_n(k) + \\
&+ c_{3,3} w_3(k) w_3(k) + \dots + c_{3,n} w_3(k) w_n(k) + \\
&\vdots \\
&c_{n,n} w_n(k) w_n(k) = \\
&= \sum_{i=1}^n c_i w_i(k) + \sum_{p=1}^n \sum_{q=p}^n c_{p,q} w_p(k) w_q(k),
\end{aligned} \tag{3.42}$$

sendo $w_i(k)$ a entrada $u(k)$ filtrada pelo filtro de Kautz de i -ésima ordem, $K_i(z)$, para $i = 1, 2, \dots, n$.

A equação (3.42) corresponde a um modelo Volterra-OBF, com *kernels* modelados por

funções de Kautz, com truncamento no *kernel* de segunda ordem. As não linearidades tratadas nas saídas dos filtros de Kautz, de forma polinomial, caracterizam tal estrutura como um modelo de Wiener, (AGUIRRE, 2007). Ressalta-se o fato de que os modelos de Wiener são definidos como uma composição de um modelo dinâmico linear, $M(k)$, em cascata com uma função estática não linear, $f(\cdot)$, (ISERMANN; MUNCHHOF, 2011).

Grande parte dos sistemas não lineares podem ser representados a partir dos *kernels* de primeira e segunda ordem, sendo desprezados no modelo os *kernels* de ordem superior, (NEMETH; KOLLAR, 2002). Cabe ressaltar que um *kernel* de primeira ordem pode ser expresso por um vetor C em \mathbb{R}^n . Já um *kernel* de segunda ordem pode ser expresso por uma matriz simétrica de dimensão $m \times m$, que é parte de um espaço vetorial composto pelas matrizes simétricas com a mesma dimensão. Assim, conforme a ordem do *kernel* aumenta, cresce a dimensão da matriz que o representa.

Existe um isomorfismo entre o espaço vetorial das matrizes simétricas de dimensão $m \times m$ e o espaço composto pelos vetores de $\mathbb{R}^{\frac{m}{2} \cdot (m+1)}$, (LIMA, 2008), permitindo que matrizes simétricas, que representam os *kernels* de segunda ordem, possam ser representadas como vetores de $\mathbb{R}^{\frac{m}{2} \cdot (m+1)}$. Desta forma, pode-se representar os *kernels* de primeira e segunda ordem de um sistema não linear a partir de vetores de dimensão n e $\frac{m}{2} \cdot (m+1)$, respectivamente.

Em relação à estabilidade, os modelos de Volterra não possuem realimentação da saída. Assim, considerando seus *kernels* estáveis, pode-se garantir a estabilidade dos mesmos, (KIBANGU; FAVIER; HASSANI, 2005). Em outras palavras, a estabilidade do modelo está associada ao fato de os *kernels* serem absolutamente somáveis, em $[0, \infty)$. Se tal requisito é atendido, o modelo é estável, (CAMPELLO; AMARAL; OLIVEIRA, 2007).

Desta forma, foram expostas as principais características dos modelos OBF-Volterra. Uma generalização deste, o modelo NARX-OBF polinomial, é mostrada na próxima seção.

3.9 MODELO NARX-OBF

Nesta tese é proposta a utilização de funções ortonormais para representar os *kernels* de um modelo NARX, esta estrutura recebe o nome de modelo NARX-OBF. Ressalta-se que a recursão de saída, neste tipo não linear de modelo OBF, é inédito na literatura, até onde vão os conhecimentos do autor.

A estrutura de formação dos modelos NARX, a partir das funções ortonormais de Kautz, é apresentada na figura 3.4.

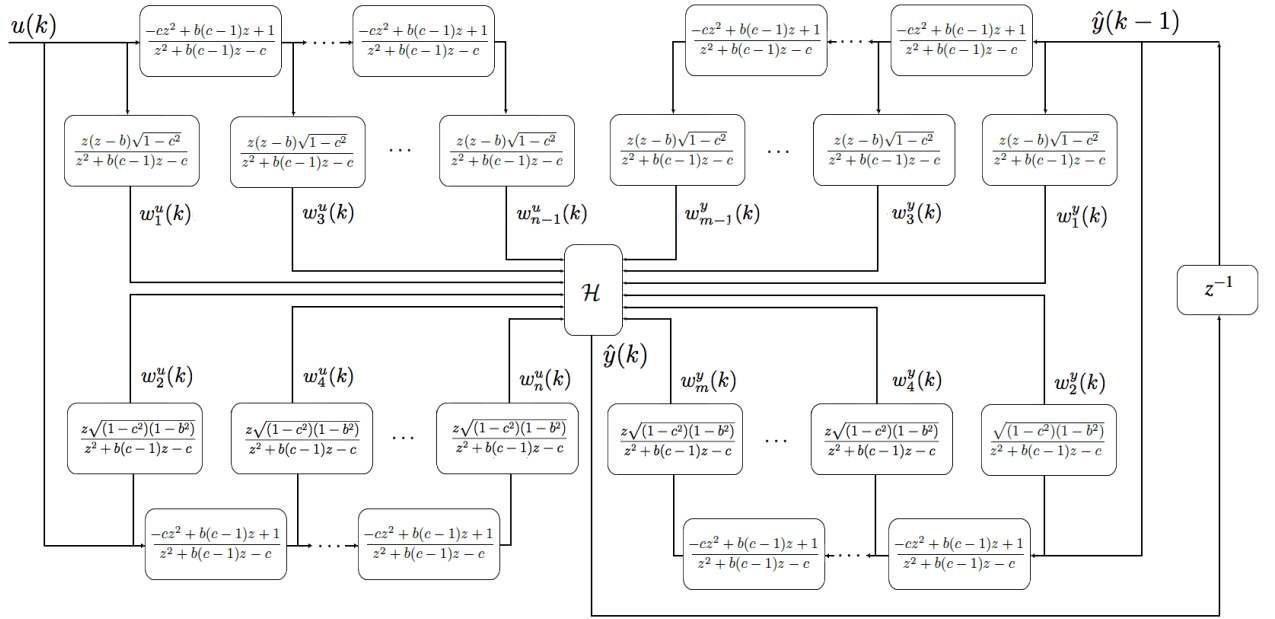


Figura 3.4: Modelo NARX-OBF com dinâmica de Kautz.

Como no modelo Volterra-OBF, o mapeamento $\mathcal{H}\{\cdot\}$ permanece sendo uma função não linear. Esta função faz o somatório de todas as combinações possíveis entre os elementos, $w_n^u(k)$ e $w_m^y(k)$, que são compostos pela entrada, $u(k)$, e pela saída, $\hat{y}(k-1)$, filtradas pelas funções de Kautz:

$$\mathfrak{B}_u = [K_1^u(z), K_2^u(z), K_3^u(z), \dots, K_n^u(z)],$$

sendo $K_n^u(z)$ as funções de Kautz responsáveis por filtrar a entrada $u(k)$, e ainda:

$$\mathfrak{B}_y = [K_1^y(z), K_2^y(z), K_3^y(z), \dots, K_m^y(z)],$$

sendo $K_n^y(z)$ as funções de Kautz responsáveis por filtrar a saída, $\hat{y}(k-1)$. Com estas funções pode-se compor a base de vetorial \mathfrak{B} :

$$\mathfrak{B} = [\mathfrak{B}_u \quad \mathfrak{B}_y].$$

Esta expansão pode ainda ser representada, de maneira mais compacta, pela equação (3.44).

$$\begin{aligned}
y_{nK}(k) = & \sum_{i=1}^n c_i^u w_i^u(k) + \sum_{j=1}^m c_j^y w_j^y(k) + \\
& \sum_{p=1}^n \sum_{q=p}^N c_{p,q}^u w_p^u(k) w_q^u(k) + \sum_{r=1}^m \sum_{s=r}^N c_{r,s}^y w_r^y(k) w_s^y(k) + \\
& \sum_{v=1}^m \sum_{w=1}^n c_{v,w}^{uy} w_v^u(k) w_w^y(k), \tag{3.44}
\end{aligned}$$

Na primeira linha da equação (3.44), tem-se a combinação linear das saídas dos filtros que compõem a base vetorial \mathfrak{B} . Já na segunda linha desta equação, tem-se os termos que representam as combinações não lineares dos sinais relativos à entrada filtrada, $w_1^u(k), w_2^u(k), \dots, w_n^u(k)$, e combinações não lineares dos sinais relativos à saída filtrada, $w_1^y(n), w_2^y(n), \dots, w_n^y(n)$. E, finalmente, na terceira linha da citada equação, tem-se as combinações não lineares dos termos cruzados, sinais oriundos da filtragem tanto dos sinais de entrada como de saída.

A equação (3.44) corresponde a um modelo NARX-OBF, com *kernels* modelados por funções de Kautz, sendo a série truncada no *kernel* de segunda ordem. Assim como nos modelos Volterra-OBF, nos modelos NARX-OBF as não linearidades são tratadas nas saídas dos filtros ortonormais, de forma polinomial e, portanto, também são modelos de Wiener.

Em forma vetorial, pode-se expressar os parâmetros dos *kernels* de primeira ordem do modelo NARX-OBF da seguinte forma:

$$C_u = [c_1^u \ c_2^u \ c_3^u \ \dots \ c_n^u]^T, \tag{3.45}$$

$$C_y = [c_1^y \ c_2^y \ c_3^y \ \dots \ c_m^y]^T, \tag{3.46}$$

e os parâmetros dos *kernels* de segunda ordem, relativos à entrada e à saída, C_{uu} e C_{yy} , respectivamente, podem ser compostos da seguinte forma:

$$C_{uu} = \begin{bmatrix} c_{1,1}^{uu} & c_{1,2}^{uu} & c_{1,3}^{uu} & \dots & c_{1,n}^{uu} \\ 0 & c_{2,2}^{uu} & c_{2,3}^{uu} & \dots & c_{2,n}^{uu} \\ 0 & 0 & c_{3,3}^{uu} & \dots & c_{3,n}^{uu} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & 0 & c_{n,n}^{uu} \end{bmatrix}, \tag{3.47}$$

$$C_{yy} = \begin{bmatrix} c_{1,1}^{yy} & c_{1,2}^{yy} & c_{1,3}^{yy} & \cdots & c_{1,m}^{yy} \\ 0 & c_{2,2}^{yy} & c_{2,3}^{yy} & \cdots & c_{2,m}^{yy} \\ 0 & 0 & c_{3,3}^{yy} & \cdots & c_{3,m}^{yy} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & c_{m,m}^{yy} \end{bmatrix}, \quad (3.48)$$

já os parâmetros do *kernel* de segunda ordem, relativo aos termos cruzados é definido por:

$$C_{uy} = \begin{bmatrix} c_{1,1}^{uy} & c_{1,2}^{uy} & c_{1,3}^{uy} & \cdots & c_{1,m}^{uy} \\ c_{2,1}^{uy} & c_{2,2}^{uy} & c_{2,3}^{uy} & \cdots & c_{2,m}^{uy} \\ c_{3,1}^{uy} & c_{3,2}^{uy} & c_{3,3}^{uy} & \cdots & c_{3,m}^{uy} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ c_{n,1}^{uy} & c_{n,2}^{uy} & c_{n,3}^{uy} & \cdots & c_{n,m}^{uy} \end{bmatrix}. \quad (3.49)$$

Como pode-se notar, o número de parâmetros a serem determinados em um modelo NARX ainda pode ser elevado. Entretanto, se a escolha dos polos das funções ortonormais se aproximar das dinâmicas apresentadas pelo sistema, pode-se reduzir consideravelmente o número de termos necessários na série para representar um sistema não linear, (HEUBERGER; HOF; WAHLBERG, 2005).

Cabe ressaltar que a grande desvantagem dos modelos NARX-OBF é que não se pode sempre garantir a convergência do modelo, uma vez que existem regressões de saída, que podem gerar realimentações positivas, levando o modelo a instabilidade.

3.9.1 Métodos de Seleção de termos do Modelo NARX-OBF

Como os modelos NARX polinomiais puros possuem um número muito elevado de termos, (CAMPELLO; AMARAL; OLIVEIRA, 2007; LEMMA; RAMASAMY; SHUHAIMI, 2010), sua implementação computacional torna-se muito exigente. Assim, geralmente são utilizadas funções ortonormais, para reduzir o número de termos da série que implementa tal modelo, NARX-OBF, (OROSKI; HOLDORF; BAUCHSPIESS, 2015). Mas ainda neste cenário, existem termos mais representativos que outros. Assim, é possível selecionar os termos mais relevantes no modelo NARX-OBF, compondo o modelo NARX-OBF simplificado.

Nos modelos NARX polinomiais, a complexidade do modelo pode ser analisada a partir da ordem do modelo e do número de termos que a compõem. Nos modelos NARX-OBF, o conceito puro de ordem deixa de fazer sentido, uma vez que as amostras não são só deslocadas no tempo, e sim filtradas pelas funções ortonormais. Assim, a complexidade dos

modelos NARX-OBF pode ser analisada em função dos seguintes fatores:

- Número de funções ortonormais utilizadas no modelo;
- Ordem das funções ortonormais envolvidas;
- Número de termos que compõem a série que implementa o modelo NARX-OBF.

Nos modelos NARX polinomiais, se a ordem utilizada for muito menor que a ordem efetiva do sistema real, o modelo não possuirá a complexidade estrutural necessária para reproduzir a dinâmica do sistema. Já se a ordem deste modelo for muito maior que a necessária, a estimação dos parâmetros pode ser mal condicionada, (AGUIRRE, 2007). De forma semelhante, para os modelos NARX-OBF é preciso encontrar uma complexidade de equilíbrio, que fuja destes óbices.

Ao conjunto de todos os termos que compõem o modelo NARX-OBF dá-se o nome de conjunto de termos candidatos \mathcal{C} . O modelo NARX-OBF simplificado é composto por um subconjunto \mathcal{S} de \mathcal{C} , ou seja: $\mathcal{S} \subset \mathcal{C}$. Isto de forma que os termos excluídos não representem uma mudança significativa no erro médio quadrático do modelo em relação aos dados amostrados do sistema a ser identificado.

No método de simplificação do modelo NARX-OBF, proposto nesta seção, o foco se direciona para a seleção dos termos mais relevantes a minimização do erro médio quadrático do modelo. As justificativas para tal escolha repousam nos fatos de que em sistemas não lineares a existência de termos redundantes destrói a dinâmica original do sistema, (AGUIRRE, 2007), e que a escolha dos termos adequados já seleciona a ordem das funções ortonormais que são mais representativas para descrever a dinâmica do sistema a ser identificado, desde que a mesma esteja disponível dentro do conjunto de termos candidatos.

Para que diferentes composições do conjunto \mathcal{S} possam ser comparadas, algumas métricas foram definidas. Entre elas, encontram-se:

- Critério de Akaike (AIC):

$$AIC = N \ln[\sigma_{erro}^2(n_\theta)] + 2n_\theta, \quad (3.50)$$

sendo N o número de amostras dos sinais de entrada e saída e n_θ o número de coeficientes (termos) na série que implementa o modelo.

- Critério de Bayes (BIC):

$$BIC = N \ln[\sigma_{erro}^2(n_\theta)] + n_\theta \ln N \quad (3.51)$$

- Critério do Erro final de Predição (FPE):

$$FPE = N \ln[\sigma_{erro}^2(n_\theta)] + N \ln \frac{N + n_\theta}{N - n_\theta} \quad (3.52)$$

Tais métricas podem ser utilizadas para a seleção de termos de um modelos NARX-OBF simplificado. Neste caso, a ideia seria fazer uso dos recursos da otimização para encontrar uma combinação de termos de \mathcal{S} que minimizem o critério escolhido.

Especificamente nesta tese, foi adotado uma versão adaptada do Critério de Akaike, que será melhor descrita na seção 6.5, sendo utilizado um algoritmo Genético para buscar os componentes de \mathcal{S} que minimizem o *AIC*. Desta forma, pode-se chegar a um modelo mais simples, o que o torna menos exigente em termos de processamento, e isto sem uma mudança significativa no seu erro médio quadrático.

3.10 CONCLUSÃO

Neste capítulo foi discutida a modelagem de sistemas não lineares a partir de modelos FIR, das Séries de Volterra e dos Modelos NARX. Como alternativa ao elevado custo computacional destas implementações, foi explorada a utilização das funções de base ortonormal para redução do custo computacional das mesmas. Foram apresentadas ainda métricas para seleção de termos dos modelos NARX-OBF, visando simplificar tal modelo.

4 MÉTODOS DE OTIMIZAÇÃO DETERMINÍSTICA

Neste capítulo serão tratados alguns fundamentos matemáticos necessários aos métodos de otimização determinística, como Gradiente Descendente, método de Newton, de Quasi-Newton e Levenberg-Marquardt. Ressalta-se que os primeiros métodos apenas são apresentados para se alcançar a base conceitual necessária ao o método de Levenberg-Marquardt, que efetivamente é empregado nesta tese.

4.1 DEFINIÇÕES INICIAIS

Dados os pontos $a \in \mathbb{R}^n$ e o número real $r > 0$, a bola aberta de centro em a e raio r é o conjunto $B(a; r)$ dos pontos $x \in \mathbb{R}^n$ cuja distância ao ponto a é menor que r , (LIMA, 2013b).

$$B(a; r) = \{x \in \mathbb{R}^n; \|x - a\| < r\},$$

seja ainda $a \in X \subset \mathbb{R}^n$, diz-se que o ponto a é interior ao conjunto X quando, para algum $r > 0$, tem-se $B(a; r) \subset X$. Isto significa que todos os pontos suficientemente próximos de a também pertencem a X , (LIMA, 2013b).

Assim, pode-se definir os pontos de mínimo global, x_g^* , e local, x_l^* , para uma função f em \mathbb{R}^n . Seja $D \subset \mathbb{R}^n$ e uma função $f : D \rightarrow \mathbb{R}$, então o ponto $x_g^* = [x_{1g}, x_{2g}, x_{3g}, \dots, x_{ng}]$ é ponto de mínimo global, se $f(x_g^*) \leq f(x)$, $\forall x \in D$. Entretanto, se apenas existir uma bola aberta B , de centro em a , com raio positivo, tal que $f(x_l^*) \leq f(x)$, $\forall x \in B$, então diz-se que $x_l^* = [x_{1l}, x_{2l}, x_{3l}, \dots, x_{nl}]$ é um ponto de mínimo local para f , (LIMA, 2013a, 2007).

Uma vez definidos os conceitos de mínimo local, x_l^* , e global, x_g^* , considere o seguinte teorema que auxilia na busca por tais pontos:

Teorema 4.1.1 (Teorema de Máximos e Mínimos) *Seja x um ponto interior de D , suponha que $\frac{\partial}{\partial x_1} f(x_1, x_2, \dots, x_n)$, $\frac{\partial}{\partial x_2} f(x_1, x_2, \dots, x_n)$, ..., $\frac{\partial}{\partial x_n} f(x_1, x_2, \dots, x_n)$ existem. Assim, se x é um ponto de máximo ou mínimo de f , então ∇f é o vetor nulo.*

Considere ainda dois conceitos auxiliares para definição dos métodos de busca por pontos de máximos e mínimos:

- Seja $X \subset \mathbb{R}^n$. Diz-se que um subconjunto $A \subset X$ é aberto em X quando cada ponto $a \in A$ é centro de uma bola aberta $B(a, r)$, tal que $B(a, r) \cap X \subset A$.
- Diz-se que um conjunto é fechado se o limite de toda a sequência convergente de pontos de F é ainda um ponto de F .

Tais conceitos são necessários à determinação dos mínimos de uma função $f : D \rightarrow \mathbb{R}$ porque os métodos de busca diferem, se D for aberto ou fechado.

Neste contexto, diz-se que x_c é um ponto crítico no interior do domínio de f , se $\nabla f(x_c) = (0,0,\dots,0)^T$. Assim, em uma função f , diferenciável e definida em um domínio aberto, D_a , os únicos candidatos a pontos de máximo ou mínimo são os pontos críticos, (LIMA, 2013b). Já as funções definidas a partir de um domínio fechado, D_f , precisam ser analisadas em duas etapas: a análise dos pontos críticos, no interior do domínio, e a análise dos pontos que compõem a fronteira do domínio.

Para cada problema de otimização tem-se uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ a ser minimizada. Especificamente na área de identificação de sistemas, tal função é a soma dos quadrados dos resíduos, SQR, exposta em seções anteriores, através da equação (2.33):

$$SQR = \sum_{i=1}^n [\hat{f}_i(x) - f_i(x)]^2.$$

Assim, seguindo-se as definições de sistema e modelo expostas na seção 2.1, pode-se quantificar o quão adequado se encontra o modelo, expresso por $\hat{f}(x)$, ao sistema em si, representado por $f(x)$. Entretanto, se fosse utilizada uma notação mais usual à área de Controle, $\hat{y}(t)$ e $y(t)$ representariam as saídas estimada e medida, respectivamente. Assim, a equação para o SQR pode ser escrita como se segue:

$$SQR = \sum_{i=1}^n [\hat{y}_i(t) - y_i(t)]^2, \quad (4.1)$$

considerando $t = kT_s$, sendo T_s o período de amostragem, tem-se:

$$SQR = \sum_{k=1}^n [\hat{y}(k) - y(k)]^2. \quad (4.2)$$

Dividindo o SQR pelo número de amostras, n , da saída $y(k)$, tem-se:

$$f(x) = MSE = \frac{1}{n} \sum_{k=1}^n [\hat{y}(k) - y(k)]^2, \quad (4.3)$$

sendo MSE o *Mean Square Error* e x o vetor que contém os parâmetros do modelo $\hat{y}(k)$.

Através da equação (4.3), pode-se mensurar a aderência do modelo aos dados amostrados das entradas e saídas do sistema. Esta métrica é conhecida como erro médio quadrático, ou no inglês: *Mean Square Error* (MSE).

A princípio, qualquer métrica poderia ser utilizada na verificação da aderência dos dados ao modelo, entre as possibilidades tem-se:

$$g(x) = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i(x) - y_i(x)|, \quad (4.4)$$

ou mesmo algo como:

$$h(x) = \frac{1}{n} \sum_{i=1}^n [\hat{y}_i(x) - y_i(x)]^6, \quad (4.5)$$

entretanto, existem vantagens estatísticas na escolha da métrica baseada no erro médio quadrático, utilizando a soma dos quadrados dos resíduos, expressa na equação (2.33), (NOCEDAL; WRIGHT, 1999).

Desta forma, grande parte dos problemas de otimização atrelados à área de identificação de sistemas podem ser enunciados da seguinte forma:

$$x_g^* = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} f(x), \quad (4.6)$$

ou seja, busca-se um vetor x_g^* , de dimensão n , que minimiza a função $f(x)$.

Faz-se necessário ainda definir o gradiente da função de avaliação:

$$\nabla f(x) = \frac{\partial f(x)}{\partial x}. \quad (4.7)$$

Como o domínio de f tem natureza vetorial, $f : \mathbb{R}^n \rightarrow \mathbb{R}$, a equação (4.7) pode ser melhor apresentada pela Jacobiana da função de avaliação: $Jf(x)$.

$$J_f(x) = \nabla f(x) = \left[\frac{\partial f(x)}{\partial x_1} \quad \frac{\partial f(x)}{\partial x_2} \quad \dots \quad \frac{\partial f(x)}{\partial x_n} \right]^T, \quad (4.8)$$

sendo $x_1, x_2, x_3, \dots, x_n$ componentes do vetor de parâmetros x .

Com isso, pode-se definir a Hessiana da função de avaliação: $H_f(x)$.

$$H_f(x) = \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ & & \vdots & \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}. \quad (4.9)$$

Feitas estas definições, as formas de minimização de uma função de custo $f(x)$ podem ser classificadas em relação ao tipo de informação, relativa à função f , necessária na busca por x_g^* , (ISERMANN; MUNCHHOF, 2011). Assim, as classes de algoritmos de otimização podem ser separadas em três ordens diferentes, explicitadas a seguir:

- Ordem zero: dependem apenas da avaliação de $f(x)$;
- Primeira ordem: são empregados $f(x)$ e seu gradiente, $\nabla f(x)$;
- Segunda ordem: utilizam $f(x)$, o gradiente de $f(x)$, $\nabla f(x)$ e sua Hessiana, $\nabla^2 f(x)$, ou aproximações desta.

Os algoritmos de otimização podem recair em mínimos locais, x_l^* , antes de encontrar o seu objetivo, o mínimo global, x_g^* . Assim, se $f(x)$ for duas vezes diferenciável, pode-se avaliar se x é um ponto de mínimo local analisando apenas o gradiente $\nabla f(x)$ e a Hessiana $\nabla^2 f(x)$ de tal função, (NOCEDAL; WRIGHT, 1999). Isto é, se x for um mínimo local, $x = x_l^*$, as seguintes condições serão atendidas:

- $\nabla f(x_l^*) = 0$;
- $\nabla^2 f(x_l^*)$ é uma matriz semidefinida positiva;

Para que tais informações sejam melhor justificadas, considere a expansão da função de avaliação expressa na próxima seção.

4.2 EXPANSÃO DA FUNÇÃO DE AVALIAÇÃO

Uma das ferramentas matemáticas mais usadas na minimização de funções é o teorema de Taylor, expresso em 4.2.1.

Teorema 4.2.1 (Teorema de Taylor) *Suponha que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é continuamente diferenciável e que $p \in \mathbb{R}^n$. então tem-se que:*

$$f(x+p) = f(x) + \nabla f(x + \alpha p)^T p, \quad (4.10)$$

sendo $\alpha \in (0,1)$. Entretanto, como f é continuamente diferenciável, pode-se aplicar o operador ∇ uma vez mais:

$$\nabla f(x+p) = \nabla f(x) + \int_0^1 \nabla^2 f(x + \alpha p) p dt, \quad (4.11)$$

e assim tem-se:

$$f(x+p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x + \alpha p) p. \quad (4.12)$$

Tal expansão tem o objetivo de aproximar a função de avaliação por estruturas que sejam computacionalmente mais leves. Neste cenário, e equação (4.12) ou simplificações da mesma serão aplicadas nas próximas seções, visando reduzir a complexidade computacional do cálculo das derivadas da função de avaliação.

4.3 MECANISMOS DE BUSCA

Na maioria dos métodos de otimização segue-se o modelo de convergência iterativa no seguinte formato:

$$x_{k+1} = x_k + \alpha_k \Delta x_k, \quad (4.13)$$

em que Δx_k é o passo ou direção de busca e o escalar α_k é o "tamanho" do passo, ambos na iteração k . Com a escolha adequada destes parâmetros espera-se que:

$$f(x_{k+1}) < f(x_k), \quad (4.14)$$

com isso, tem-se um processo em que x_k , com $k = 1, 2, \dots$, converge para o vetor x_l^* . Assim, de maneira genérica, pode-se expressar tal processo pelo algoritmo 1.

Algoritmo 1: Métodos de direção descendente.

Data: Dado um ponto inicial $x \in \text{dom } f$;

while $\nabla f(x_k) \neq 0$ **do**

1. Determina-se a direção de busca: Δx ;
 2. *Line search*: Escolhe-se o tamanho de passo, α , sendo $\alpha > 0$;
 3. Atualiza-se: $x_{k+1} = x_k + \alpha \Delta x$;
-

Geralmente, os métodos de minimização de funções alternam entre duas etapas: a determinação de uma direção de busca e uma busca na reta formada por tal direção, sendo esta segunda também conhecida como *line search*, (BOYD; VANDENBERGHE, 2009). Os principais métodos de busca em linha serão detalhados na próxima seção.

4.3.1 Busca em Linha

A busca em linha é uma técnica de otimização utilizada nos casos em que já se tem uma definição da direção de busca. Assim, procura-se estabelecer um mínimo local na reta definida por tal direção. Nesta seção serão abordados os dois tipos mais comuns de busca em linha: a busca exata em linha e o *Backtracking*.

4.3.1.1 Busca Exata em Linha

Assumindo-se que a direção de busca esteja estabelecida, a partir de um método qualquer, parte-se para a busca em linha. A busca exata é um dos métodos de busca em linha. É baseada na escolha de um valor de $\alpha \in \mathbb{R}$ que minimize a função f ao longo do raio $\{x + \alpha \Delta x\}$.

$$\alpha = \underset{\alpha \geq 0}{\operatorname{argmin}} f(x + \alpha \Delta x), \quad (4.15)$$

sendo $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

Cabe ressaltar que a resolução de (4.15) pela busca exata pode ser computacionalmente custosa, e geralmente, recorre-se a versões simplificadas da mesma. A maioria das buscas em linha, realizadas na prática, são inexatas, (BOYD; VANDENBERGHE, 2009), ou seja, o passo é escolhido de forma a aproximadamente minimizar f , ao longo do raio $\{x + \alpha \Delta x\}$, sendo $\alpha > 0$.

4.3.1.2 *Backtracking*

Um dos métodos mais simples e eficientes de busca inexata em linha é o *Backtracking*. Neste método de busca em linha são necessários dois parâmetros: α e β , e a dinâmica deste tipo de busca pode ser observada no algoritmo 2.

Algoritmo 2: Backtraking Search.

Data: Dado uma direção de busca Δx , sendo $x \in \text{dom } f$, $\alpha \in (0,0.5)$ e $\beta \in (0,1)$
 $u=1$;
while $f(x + \Delta x) > f(x) + \alpha u \nabla f(x)^T \Delta x \neq 0$ **do**
 $u = \beta u$;

Este tipo de busca em linha é chamada de *Backtracking* porque o valor do passo começa com valor unitário e se reduz em um fator β até que a condição de parada seja atingida, (BOYD; VANDENBERGHE, 2009): $f(x + \Delta x) > f(x) + \alpha u \nabla f(x)^T \Delta x \neq 0$. Desde que Δx seja uma direção descendente, tem-se que $\nabla f(x)^T \Delta x < 0$, então para passos de pequeno módulo, u , tem-se:

$$f(x + u\Delta x) \approx f(x) + u \nabla f(x)^T \Delta x < f(x) + \alpha u \nabla f(x)^T \Delta x, \quad (4.16)$$

o que mostra que a busca *Backtracking* eventualmente termina. A constante α pode ser interpretada como uma fração aceitável do decréscimo em f previsto por extrapolação linear, (BOYD; VANDENBERGHE, 2009).

4.3.2 Região de Confiança

O método da Região de Confiança é uma alternativa aos métodos que alternam a busca de direção e a busca em linha. Neste método a direção e o passo são escolhidos simultaneamente, (NOCEDAL; WRIGHT, 1999). Tal método é bastante discutido por ser uma das estratégias utilizadas no método de Levenberg-Marquardt, um dos mais eficientes e mais utilizados métodos de otimização, principalmente na área de Identificação de Sistemas, (ISERMANN; MUNCHHOF, 2011).

Os métodos de Região de Confiança definem regiões em volta da iteração atual, na qual se gera um modelo, $m_k(p)$, que seja uma representação adequada para a função objetivo, f , e então escolhe-se um passo que aproximadamente minimize o modelo na região de confiança.

$$\min_{p \in \mathbb{R}^n} m_k(p) = f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B_k p, \quad (4.17)$$

em que B_k é alguma matriz simétrica e $\Delta_k > 0$ na região de confiança.

Normalmente, a região de confiança é uma bola definida por $\|p\| \leq \Delta$, no qual o escalar $\Delta > 0$ é chamado de raio da Região de Confiança, (NOCEDAL; WRIGHT, 1999).

Para avaliar a concordância entre o modelo, m_k , e a função objetivo, f , pode-se recorrer ao coeficiente ρ , expresso pela equação (4.18).

$$\rho = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)}. \quad (4.18)$$

Cabe ressaltar que as buscas em Linha e por Região de Confiança diferem nas formas como cada uma escolhe a direção de busca e o passo em cada iteração, (NOCEDAL; WRIGHT, 1999). Enquanto a busca em linha inicia o seu processo identificando uma distância apropriada em uma direção fixa, a busca por Região de Confiança fixa uma distância máxima, raio da região de confiança, e depois procura uma direção e um passo apropriado para as melhorias possíveis.

O mecanismo básico de um algoritmo de Região de Confiança pode ser observado no algoritmo 3.

Algoritmo 3: Método de Região de Confiança.

Data: Dados: $\bar{\Delta} > 0$, $\Delta_0 \in (0, \bar{\Delta})$ e $\eta \in [0, \frac{1}{4}]$;

for $k = 0, 1, 2, \dots$ **do**

 Obter p_k para resolver aproximadamente a equação (4.17);

$\min_p \quad m_p(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T B_k p$;

 Avaliar a equação (4.18);

$\rho = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)}$

if $\rho < \frac{1}{4}$ **then**

$\Delta_{k+1} = \frac{1}{4} \|p_k\|$

else

if $\rho > \frac{3}{4}$ e $\|p_k\| = \Delta_k$ **then**

$\Delta_{k+1} = \min(2 \cdot \Delta_k, \bar{\Delta})$

else

$\Delta_{k+1} = \Delta_k$

if $\rho_k > \eta$ **then**

$x_{k+1} = x_k + p_k$;

else

$x_{k+1} = x_k$;

4.3.3 Método de Newton

O método de Newton é um algoritmo de otimização, de direção descendente, de segunda ordem, (BOYD; VANDENBERGHE, 2009), que visa minimizar a função f a partir do algoritmo 4. Para tanto, considere a expansão, em séries de Taylor, numa aproximação da função f .

$$f(x_0 + \Delta x) \approx f(x_0) + \nabla f(x_0)\Delta x + \frac{1}{2}\Delta x^T \nabla^2 f(x_0)\Delta x. \quad (4.19)$$

Aplicando o gradiente à equação (4.19) tem-se:

$$\frac{\partial f(x_0 + \Delta x)}{\partial \Delta x} = \nabla f(x_0) + \nabla^2 f(x_0)\Delta x. \quad (4.20)$$

Considerando x_g^* o ponto que minimiza a função $f(x)$, então $\nabla f(x_g^*) = 0$, assim tem-se que:

$$0 = \nabla f(x_0) + \nabla^2 f(x_0)\Delta x. \quad (4.21)$$

Assim, pode-se chegar ao passo de Newton:

$$\Delta x = - [\nabla^2 f(\bar{x})]^{-1} \nabla f(\bar{x}). \quad (4.22)$$

Utilizando a notação recursiva tem-se:

$$x_{k+1} = x_k - [\nabla^2 f(\bar{x})]^{-1} \nabla f(\bar{x}). \quad (4.23)$$

Algoritmo 4: Método de Newton.

Define-se $x_0 \in \mathbb{R}^n$;
 Calcule Δx solução de
 $\nabla^2 f(x_0)\Delta x = -\nabla f(x_0)$;
 $x_1 = x_0 + \Delta x$;
 $k = 1$;
while $\nabla f(x_k) \neq 0$ **do**
 Calcule Δx solução de
 $\nabla^2 f(x_k)\Delta x = -\nabla f(x_k)$;
 Determine x_{k+1} :
 $x_{k+1} = x_k + \Delta x$;
 $k = k + 1$;

As vantagens dos métodos de segunda ordem, como o método de Newton, são sua eficiência e velocidade de convergência. Já as desvantagens deste método são: a sua dependência do conhecimento de $\nabla^2 f(x)$ e a necessidade de inversão da mesma, que pode ser proble-

mática, por questões de mal condicionamento numérico, para pontos distantes do ótimo, (ISERMANN; MUNCHHOF, 2011). Mas o principal ponto fraco do método de Newton ainda é o custo computacional empregado no cálculo da matriz Hessiana e do novo passo do algoritmo, (BOYD; VANDENBERGHE, 2009).

4.3.4 Método de Gauss Newton

O método de Gauss Newton é um método de segunda ordem, mas busca simplificações para a inversão de $\nabla^2 f(x)$, que consiste na etapa mais problemática do método de Newton.

Assim, considere a aplicação do operador ∇ à equação (2.33):

$$\nabla f(x) = J_\xi(x)\xi(x). \quad (4.24)$$

Aplicando o operador ∇ mais uma vez na equação (4.24), tem-se:

$$\nabla^2 f(x) = \sum_{i=1}^n \nabla \xi_i(x) \nabla \xi_i(x)^T + \sum_{i=1}^n \xi_i(x) \nabla^2 \xi_i(x). \quad (4.25)$$

A primeira parcela desta equação, $\sum_{i=1}^n \nabla \xi_i(x) \nabla \xi_i(x)^T$, que também pode ser escrita como $J_\xi(x)J_\xi(x)^T$, é sempre mais importante que a segunda parcela desta equação, (NOCEDAL; WRIGHT, 1999). Assim, no método de Gauss-Newton considera-se a seguinte simplificação:

$$\nabla^2 f(x) \approx J_\xi(x)^T J_\xi(x), \quad (4.26)$$

assim, em vez de se resolver a equação de Newton: $\nabla^2 f(x)\Delta x = -\nabla f(x)$, pode-se recorrer a esta simplificação. E, mesmo considerando a omissão do segundo termo de (4.25), o método de Gauss-Newton possui desempenho muito similar ao método de Newton, (NOCEDAL; WRIGHT, 1999).

Na prática, o cálculo computacional do gradiente como um todo, $\nabla f(x)$, é mais simples que o cálculo individual das derivadas parciais de primeira ordem, para formar posteriormente o gradiente, equação (4.24), (NOCEDAL; WRIGHT, 1999).

Por razões de economia, renomeia-se $\nabla^2 f(x)$ simplificado, equação (4.26), para $B(x)$.

$$B(x) = J_\xi(x)J_\xi(x)^T \approx \nabla^2 f(x). \quad (4.27)$$

O método de Gauss-Newton é mostrado a partir do algoritmo 5.

Algoritmo 5: Método de Gauss Newton.

Define-se $x_0 \in \mathbb{R}^n$;

Calcule Δx_0 solução de

$$B(x_0)\Delta x_0 = -\nabla f(x_0);$$

onde: $B(x) = J_\xi(x)J_\xi(x)^T$

$$x_1 = x_0 + \Delta x_0 ;$$

$$k = 1;$$

while $\nabla f(x_k) \neq 0$ **do**

 Calcule Δx_k solução de

$$B(x_k)\Delta x_k = -\nabla f(x_k);$$

 Determine x_{k+1} ;

$$x_{k+1} = x_k + \Delta x_k ;$$

$$k = k + 1;$$

4.3.5 Algoritmo de Levenberg-Marquardt

O algoritmo de Levenberg-Marquardt é um método de segunda ordem, que compartilha com os métodos baseados em gradiente a habilidade de convergir rapidamente a partir de um ponto inicial qualquer, ainda que este ponto esteja fora da região de convergência de outros métodos, (MARQUARDT, 1963).

O método de Levenberg-Marquardt pode ser derivado substituindo a estratégia de busca em linha pela estratégia de Região de Confiança. O uso da estratégia de Região de Confiança evita o principal problema do método de Gauss Newton, que ocorre quando a Jacobiana $J_\xi(x)$ deixa de ser posto pleno ou se aproxima disto, (NOCEDAL; WRIGHT, 1999).

Pode-se entender o algoritmo de Levenberg-Marquardt como o método de Gauss-Newton com a seguinte modificação:

$$\begin{cases} x_{k+1} = x_k + \Delta x_k \\ \Delta x_k = -[J_\xi(x)^T J_\xi(x) + \lambda I]^{-1} [J_\xi(x)^T \xi(x)], \end{cases} \quad (4.28)$$

em que $\lambda \in \mathbb{R}$ e I é a matriz identidade.

Neste cenário, é razoável considerar o uso de algoritmos híbridos, que se comportem como Levenberg-Marquardt, se os resíduos forem pequenos, e comutem seu comportamento para o método de Newton, se os resíduos se tornarem grandes, (NOCEDAL; WRIGHT, 1999). A direção de Levenberg-Marquardt interpola as direções dos métodos de Newton e do gradiente descendente, (SEBER; WILD, 2003). A ideia do algoritmo de Levenberg-Marquardt é perturbar a matriz $B(x)$, considerando $B(x) + \rho \cdot I$ para $\rho > 0$.

Como em alguns dos métodos de Gauss-Newton, o algoritmo de Levenberg-Marquardt é baseado em uma expansão em séries de Taylor, (MARQUARDT, 1963).

O mecanismo de busca utilizado pelo algoritmo de Levenberg-Marquardt pode ser observado no algoritmo 6.

Algoritmo 6: Algoritmo de Levenberg-Marquardt.

Define-se $x_0 \in \mathbb{R}^n$;
Calcule d_0 solução de
 $[B(x_0) + \lambda I] \Delta x_0 = -\nabla f(x_0)$;
onde: $B(x) = J_\xi(x) J_\xi(x)^T$;
 $x_1 = x_0 + \Delta x_0$;
 $k=1$;
while $\nabla f(x_k) \neq 0$ **do**
 Calcule Δx_k solução de:
 $[B(x_k) + \lambda . I] . \Delta x_k = -\nabla f(x_k)$;
 Determine x_{k+1} ;
 $x_{k+1} = x_k + \Delta x_k$;
 $k=k+1$
end while

4.4 CONCLUSÃO

Neste capítulo foram apresentados os principais métodos de otimização determinística. Foram expostos os métodos do gradiente descendente, de Newton, de Quasi-Newton, de Gauss-Newton e de Levenberg-Marquardt. Também foram apresentados os meios matemáticos necessários a compreensão de cada um destes, bem como os algoritmos que os implementam.

5 MÉTODOS DE OTIMIZAÇÃO HEURÍSTICA

Nas últimas décadas, a busca por soluções de problemas de otimização tem motivado grandes avanços na Matemática e na Engenharia. Métodos como o de Newton, *Steepest Descent* e Levenberg-Marquardt, expostos no capítulo anterior, vem tornando possível a solução de uma série de problemas de otimização, (NOCEDAL; WRIGHT, 1999). Entretanto, estes métodos requerem requisitos muito restritos para sua aplicação, como convexidade, disponibilidade dos gradientes, entre outros, (BOYD; VANDENBERGHE, 2009). É importante salientar que em muitas aplicações industriais, o projetista tem que lidar com muitas peculiaridades como: não linearidades, não convexidade, existência de múltiplos mínimos locais, presença de variáveis discretas, entre outras, (MOCKUS JONAS; EDDY; REKLAITIS, 1997).

Uma das classes de métodos de otimização que potencialmente podem contornar estes problemas são os algoritmos heurísticos. Algumas vantagens destes algoritmos incluem: (i) o fato destes métodos não requererem as informações de gradiente e poderem ser aplicados a problemas em que o gradiente é difícil de se obter ou simplesmente não são definidos; (ii) esses algoritmos não ficam “presos” em mínimos locais, se corretamente sintonizados; (iii) esses métodos podem ser aplicados a funções pouco suaves ou descontínuas; (iv) Esses algoritmos fornecem um conjunto de soluções sub-ótimas ao invés de uma única solução, dando ao projetista um conjunto de opções a serem escolhidas; e (v) Esses métodos podem ser facilmente empregados para resolver problemas de otimização com variáveis mistas (discretas e contínuas), (KOZA, 1998). Entre os algoritmos heurísticos mais populares estão os Algoritmos Genéticos (AG), (KOZA, 1998), o *Ant Colony Optimization* (ACO), (SOLNON, 2010), e o *Particle Swarm Optimization* (PSO), (OLSSON, 2011), todos estes inspirados por princípios biológicos.

Outros princípios têm sido empregados no desenvolvimento de algoritmos heurísticos de otimização, como o *Imperialist Competitive Algorithm*, (ATASHPAZ; LUCAS, 2007), baseado nas políticas imperialistas de estender os poderes e a área de influência de um governo além de suas fronteiras, *Group Search Optimizer* inspirado pelos comportamentos de busca animal, (HE; WU; SAUNDERS, 2009), e o *Biogeography-Based Optimization*, que faz uso da distribuição geográfica de organismos biológicos como inspiração para algoritmos na área de otimização, (SIMON, 2008).

Neste capítulo serão apresentados os principais conceitos relativos ao Algoritmo Genético, assim como seus operadores, aplicáveis a problemas de Identificação de Sistemas. Será apresentado ainda um algoritmo de otimização heurística proposto nesta tese. Trata-se do Algoritmo do Princípio Antrópico, APA, aplicado a uma série de problemas de otimização e identificação de sistemas dinâmicos.

5.1 ALGORITMO GENÉTICO

Nas décadas de 1950 e 1960, muitos cientistas da computação estudaram sistemas evolutivos com a ideia de que a evolução pudesse ser utilizada como uma ferramenta de otimização para problemas de engenharia, (MICHELL, 1999).

O Algoritmo Genético, AG, foi desenvolvido por John Holland nos anos 1960, (MICHELL, 1999), e suas ideias foram apresentadas em seu livro: *Adaptation in Natural and Artificial Systems*, em 1975, (COLEY, 1999). Inspirado nas idéias evolucionistas de Darwin, Holland criou um método de resolução de problemas de otimização que dispensa o equacionamento das matrizes Jacobianas e Hessianas da função de avaliação do problema, (KOZA, 1998).

Em analogia ao mecanismo de reprodução de um conjunto de organismos, e da evolução dele decorrente, o AG extrai um comportamento emergente de convergência. Comportamentos emergentes envolvem a aplicação de regras simples, repetidas vezes, que geram comportamentos complexos, (KOZA, 1998). No caso do AG, a ação de operadores, como a Mutação e o *Crossover*, e a maior probabilidade de reprodução dos indivíduos com melhor imagem na função de avaliação, geram, na maioria dos casos, o comportamento convergente do algoritmo.

Em um AG cada indivíduo é modelado como um conjunto de constantes: c_1, c_2, \dots, c_n , chamadas de genes, formando um vetor, conhecido como cromossomo: \mathcal{C} .

$$\mathcal{C} : \begin{array}{|c|c|c|c|c|} \hline c_1 & c_2 & c_3 & \dots & c_n \\ \hline \end{array}$$

tais constantes podem ser binárias ou reais dependendo do problema a ser tratado.

Em um AG, não se tem um único cromossomo, mas toda uma população, \mathcal{P} , de cromossomos:

$$\mathcal{P} = \{\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \dots, \mathcal{C}_m\}, \quad (5.1)$$

sendo que cada cromossomo é analisado por uma função de avaliação, conhecida como função de *Fitness*, $\mathcal{F}_{it}(\mathcal{C}) : \mathbb{R}^n \rightarrow \mathbb{R}$, e os cromossomos com melhores *Fitness* possuirão maior probabilidade de se reproduzir nas próximas gerações do AG.

O Algoritmo Genético simula o processo de evolução Darwiniana a partir de operações sobre os genes em um cromossomo. Em outras palavras, um AG é um algoritmo matemático que transforma uma população de objetos matemáticos, com função de avaliação bem definida, \mathcal{F}_{it} , em uma nova população de objetos matemáticos, seguindo os princípios Darwinianos de reprodução dos mais adaptados. Assim, pode-se definir o Algoritmo Ge-

nético como um método heurístico de otimização, baseado nos princípios da evolução de Darwin, (KOZA, 1998).

Conforme o algoritmo 7, em um AG clássico tem-se a criação aleatória de uma população, \mathcal{P} . Esta população é composta por possíveis soluções, \mathcal{C} , ao problema tratado. Na sequência, tais soluções são avaliadas, pela função $\mathcal{F}_{it}(\mathcal{C})$, segundo sua adequação como solução do problema proposto. Após estes passos, o algoritmo entra em um ciclo no qual os melhores indivíduos (cromossomos com menor imagem na função de avaliação, \mathcal{F}_{it}) são probabilisticamente selecionados para serem progenitores da próxima geração. Os operadores de Cruzamento, *Crossover* e Mutação são aplicados à população e os novos indivíduos gerados são avaliados. Este ciclo repete-se até que o critério de parada seja alcançado.

Algoritmo 7: Algoritmo Genético

```
Inicializa-se a população,  $\mathcal{P}$ , de soluções;  
Simula-se o modelo gerado por cada cromossomo;  
Calcula-se o Fitness para cada cromossomo,  $\mathcal{F}_{it}(\mathcal{C})$ ;  
while  $Fitness_{min} > criterio_{parada}$  do  
    Salva-se o melhor indivíduo;  
    Seleção dos progenitores;  
    Aplicação do Operador de Cruzamento;  
    Aplicação do Operador de Crossover;  
    Aplicação do Operador de Mutação;  
    Simula-se o modelo gerado por cada cromossomo;  
    Calcula-se o Fitness para cada cromossomo;
```

Seguindo os princípios deste algoritmo, espera-se que os indivíduos sofram um processo de seleção natural e os melhores indivíduos da população se aproximem da solução ótima do problema, (KOZA, 1998).

Cabe ressaltar que o AG não garante a convergência para o ponto ótimo do problema, podendo acabar confinado a uma região de mínimo local, (COLEY, 1999). Como mencionado no capítulo 4, um ponto x_l é um mínimo local se existir uma vizinhança \mathcal{V} de x_l tal que $f(x_l) \leq f(x)$ para $x \in \mathcal{V}$, (NOCEDAL; WRIGHT, 1999).

5.1.1 Operadores do Algoritmo Genético

Nesta seção, são apresentados os operadores clássicos de um AG, como *Crossover* e Mutação, e alguns operadores adaptados para resolução de problemas de Identificação de Sistemas, como o operador de mínimos quadrados.

5.1.1.1 Operador de *Crossover*

O operador de *Crossing Over*, ou simplesmente *Crossover*, é um dos operadores que já compunham o AG clássico. Tal operador foi inspirado no processo bioquímico de *Crossing-*

Over, em que partes de dois cromossomos são permutadas no processo de reprodução sexual de seres vivos, (SASTRY; GOLDBERG; KENDALL, 2005).

Na figura 5.1 é mostrada a ação do operador de *Crossover* sobre os cromossomos $f = [f_1, f_2, \dots, f_n]$ e $g = [g_1, g_2, \dots, g_n]$, gerando um novo cromossomo com partes gênicas tanto de f como de g .

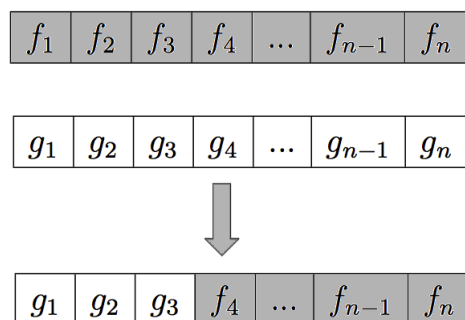


Figura 5.1: Ilustração do operador de *Crossover* com um ponto de corte.

O ponto de cisão nos cromossomos é escolhido aleatoriamente, e as partes dos cromossomos dos progenitores são permutadas formando um novo cromossomo, que é denominado de prole, (MICHELL, 1999).

Cabe ressaltar que a função deste operador, em um processo de otimização heurística, é a de busca global no espaço de busca, (KOZA, 1998).

5.1.1.2 Operador de Mutação

O operador de Mutação, assim como o *Crossover*, está entre os operadores que compunham as primeiras versões do Algoritmo Genético. Para que buscas locais pudessem ser realizadas, o operador de mutação insere uma constante aleatória em uma posição gênica também aleatória, como expresso na figura 5.2.

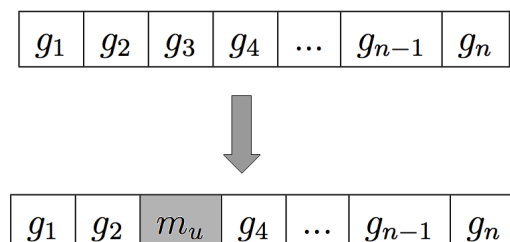


Figura 5.2: Ilustração do operador de Mutação.

Este operador tem função essencialmente de busca local, sendo complementado pelo

operador de *crossover* que realiza buscas globais, compondo o mecanismo básico do Algoritmo Genético clássico. Em algumas versões, o operador de mutação apenas soma, ao gene a ser mutado, uma constante aleatória de pequeno módulo. Diminuindo, dessa forma, o efeito de busca aleatória realizada por tal operador, (KOZA, 1998).

5.1.1.3 Hibridização com o Método dos Mínimos Quadrados

O método dos Mínimos Quadrados, MQ, apresentado na equação (2.26), foi uma das metodologias pioneiras na área de Identificação de Sistemas. Entretanto, tal metodologia só é aplicável a modelos que apresentem linearidade em relação aos seus parâmetros, (AGUIRRE, 2007). Para sistemas não lineares, a aplicação deste método geraria apenas aproximações para os parâmetros que efetivamente minimizariam a função de avaliação, expressa na equação (4.3). Assim, este método pode ser utilizado conjuntamente com um Algoritmo Genético, tendo como responsabilidade encontrar um cromossomo que seja uma aproximação razoável ao cromossomo ótimo. E uma vez que este seja incorporado à população de soluções, \mathcal{P} , o AG continua seu processo evolutivo.

O método MQ consiste na aplicação do método dos mínimos quadrados, expresso pela equação (2.26), a um cromossomo \mathcal{C} , escolhido aleatoriamente dentre a população de possíveis soluções, \mathcal{P} .

Por exemplo, em problemas de identificação de sistemas, nos quais o modelo é expresso a partir de funções ortonormais, $\psi_1, \psi_2, \dots, \psi_n$, o AG seria responsável por encontrar os coeficientes c_1, c_2, \dots, c_n adequados ao modelo, bem como os polos, que parametrizam as funções ortonormais. Um cromossomo deste AG teria a seguinte composição de genes:

c_1	c_2	c_3	c_4	...	c_n	a	b
-------	-------	-------	-------	-----	-------	-----	-----

Se as funções ortonormais utilizadas, $\psi_1(z), \psi_2(z), \dots, \psi_n(z)$, fossem as funções de Kautz, $K_1(z), K_2(z), \dots, K_n(z)$, os genes relativos ao polo, $a + bi$, que parametriza as funções ortonormais, seriam utilizados para gerá-las. E, assim, a matriz Ψ_k poderia ser formada a partir de versões filtradas, $w_n(k)$, do vetor de entrada $u(k)$.

$$w_n(k) = \sum_{\tau=0}^k K_n(\tau) \cdot u(k - \tau). \quad (5.2)$$

A estrutura da matriz de Regressão é expressa na equação (5.3).

$$\Psi_k = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ w_1(k-2) & w_2(k-2) & \cdots & w_n(k-2) \\ w_1(k-1) & w_2(k-1) & \cdots & w_n(k-1) \\ w_1(k) & w_2(k) & \cdots & w_n(k) \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}, \quad (5.3)$$

onde $w_n(k)$ corresponde a entrada $u(k)$ filtrada por um filtro de Kautz de ordem n .

E assim, a equação (2.26) pode ser reescrita para o operador MQ como:

$$\hat{\theta} = (\Psi_k^T \Psi_k)^{-1} \Psi_k^T Y, \quad (5.4)$$

sendo Ψ_k a matriz de regressão e θ o vetor de parâmetros do modelo:

$$\theta = [c_1, c_2, \dots, c_n]^T. \quad (5.5)$$

Cabe ressaltar que, apesar do nome, a matriz Ψ_k não possui versões regressivas do sinal de entrada, $u(n)$, e sim versões filtradas do mesmo.

Após a ação do método MQ, os genes relativos aos coeficientes das funções de Kautz são substituídos pelos elementos do vetor $\hat{\theta}$ e o AG segue seu ciclo normal de convergência, uma vez que a solução encontrada é apenas uma aproximação linear do vetor de parâmetros ótimo. Tal operador é utilizado no AG a cada ciclo de C_{MQ} gerações, sendo interessante impor que: $C_{MQ} > 5$. Isto para evitar a convergência prematura das soluções a um mínimo local.

5.1.1.4 Hibridização com o Método de Nelder Mead

O algoritmo de Nelder Mead, NM, foi desenvolvido em 1965 por John Nelder e Roger Mead. Consiste em um método heurístico para minimizar uma função de avaliação $f : \mathbb{R}^n \rightarrow \mathbb{R}$, em um espaço multidimensional, (NELDER; MEAD, 1965).

O algoritmo de Nelder Mead pode ser aplicado em conjunto com o Algoritmo Genético, com objetivo de acelerar o processo de convergência do AG, (ABDEL; FUKUSHIMA, 2003). Para tanto, são utilizadas estruturas geométricas conhecidas como *simplex*, que consistem em poliedros de $n + 1$ pontos, equidistantes dos seus vértices vizinhos, (ISERMANN; MUNCHHOF, 2011). Na busca por um ponto de mínimo, x^* , da função de avaliação, $f(x)$, os *simplex* sofrem transformações geométricas, em um espaço de n dimensões, (NELDER; MEAD, 1965).

As transformações geométricas, em \mathbb{R}^2 , aplicadas a um simplex $[x_1, x_2, x_3]$, podem ser visualizadas através da figura 5.3.

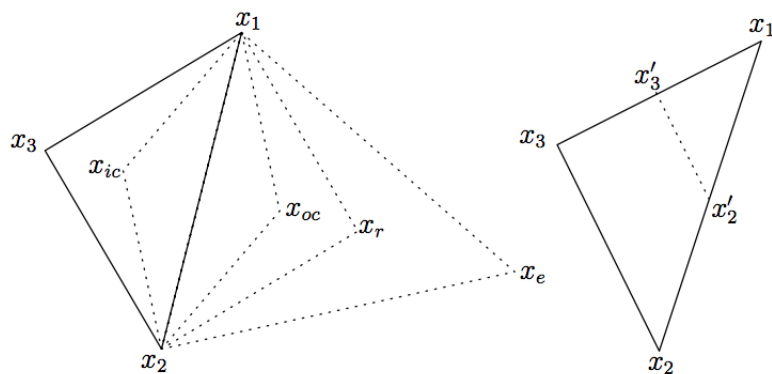


Figura 5.3: Operações de reflexão, expansão, contração e redução, (ABDEL; FUKUSHIMA, 2003).

As transformações geométricas aplicadas sobre o simplex $[x_1, x_2, x_3]$, mostradas na figura 5.3 podem ser nomeadas como:

- Reflexão, gerando o simplex $[x_1, x_2, x_r]$;
- Expansão, formando o simplex $[x_1, x_2, x_e]$;
- Contração interna, gerando o simplex $[x_1, x_2, x_{ic}]$;
- Contração externa, formando o simplex $[x_1, x_2, x_{oc}]$;
- Redução, gerando o simplex $[x_1, x_2', x_3']$.

Para o ordenamento necessário ao algoritmo de Nelder Mead, considera-se que: $f(x_1) \leq f(x_2) \leq f(x_3)$, ou seja, a imagem do vértice x_1 é menor que a de x_2 e a imagem de x_2 , por sua vez, é menor que a do vértice x_3 . O método completo pode ser descrito pelos passos expostos no algoritmo 8, (ABDEL; FUKUSHIMA, 2003).

Algoritmo 8: Nelder Mead

1. Ordenar os $n + 1$ vértices: x_1, x_2, \dots, x_{n+1} , tal que: $f(x_1) \leq f(x_2) \leq \dots \leq f(x_{n+1})$;
 2. Computar a reflexão: $x_r = \bar{x} + \rho \cdot (\bar{x} - x_{n+1})$,
sendo \bar{x} é o centroide dos n melhores vértices: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$;
Avaliar $f(x_r)$;
if $f(x_1) \leq f(x_r) < f(x_n)$ **then**
 Compute: $x_{n+1} = x_r$;
 Ir para o passo 6;
 3. **if** $x_r < x_1$ **then**
 Compute o ponto de expansão: $x_e = \bar{x} + \chi \cdot (x_r - \bar{x})$;
 Avaliar $f(x_e)$;
 if $f(x_e) < f(x_r)$ **then**
 Compute: $x_{n+1} = x_e$;
 else
 $x_{n+1} = x_r$;
 Ir para o passo 6;
 4. **if** $x_r \geq x_n$ **then**
 - 4.1. **if** $x_n \geq x_r \geq x_{n+1}$ **then**
 Computar a contração externa ao Simplex: $x_{oc} = \bar{x} + \gamma \cdot (\bar{x} - x_{n+1})$;
 Avaliar $f(x_{oc})$;
 if $f(x_{oc}) \leq x_r$ **then**
 Compute: $x_{n+1} = x_{oc}$;
 Ir para o passo 6;
 else
 Ir para o passo 5;
 - 4.2. **if** $x_r \geq x_{n+1}$ **then**
 Computar a contração interna ao Simplex: $x_{ic} = \bar{x} + \gamma \cdot (\bar{x} - x_{n+1})$;
 Avaliar $f(x_{ic})$;
 if $f(x_{ic}) \leq f(x_{n+1})$ **then**
 Compute: $x_{n+1} = x_{ic}$;
 Ir para o passo 6;
 else
 Ir para o passo 5;
 5. Aplicar a Redução ao Simplex: $x'_i = x_1 + \sigma(x_i - x_1)$, $i = 2, \dots, n + 1$,
Substituir os vértice x_1, \dots, x_{n+1} por x'_1, \dots, x_{n+1} ;
 6. Critério de Parada: Ordenar os $n + 1$ vértices: x_1, x_2, \dots, x_{n+1} , de forma que:
 $f(x_1) \leq f(x_2) \leq \dots \leq f(x_{n+1})$;
if $f(x_{n+1}) - f(x_1) < \varepsilon$ **then**
 Parar o algoritmo;
else
 Ir para o passo 2.
-

sendo: ρ , χ , γ e σ os coeficientes de reflexão, expansão, contração e encolhimento, respectivamente. E ε é o erro tolerado, utilizado como critério de parada para o algoritmo.

A utilização do método de Nelder Mead como um acelerador de convergência local de um AG repousa na ideia de que $n + 1$ cromossomos, de dimensão n , seriam utilizados para compor um simplex, de dimensão $n + 1$, (ABDEL; FUKUSHIMA, 2003). E após a aplicação do operador NM, descrito pelo algoritmo 8, o vértice com melhor desempenho, x_1 , é devolvido à população \mathcal{P} do Algoritmo Genético, e o ciclo de iterações continua a ser empregado até que se obtenha o critério de parada do AG.

Cabe salientar que os algoritmos heurísticos, como o AG, respondem bem ao problema da dimensionalidade e dos mínimos locais, apesar de não se poder garantir a determinação do ótimo global, (KOZA, 1998).

5.2 ALGORITMO DO PRINCÍPIO ANTRÓPICO

Seguindo a linha dos algoritmos inspirados por princípios naturais, nesta tese é proposto um algoritmo heurístico de otimização baseado no Princípio Antrópico, que é originado na Física. De acordo com o Princípio Antrópico, incontáveis fatores tiveram que convergir, na história do universo, para tornar a existência da humanidade possível, (HAWKING; MLODINOW, 2012). Assim, a evolução do universo pode ser vista como um processo de otimização cuja função objetivo visa minimizar os efeitos contrários a existência da humanidade.

Aspectos como a baixa excentricidade da órbita da Terra, a relação entre a massa do Sol e a distância e sua distância a Terra são exemplos no conjunto de condições adequadas (e necessárias) a existência da humanidade, (HAWKING; MLODINOW, 2012). Se apenas um elemento deste conjunto fosse diferente, a vida, como se conhece, provavelmente não teria se desenvolvido. Em outras palavras, as constantes do universo parecem ter evoluído de tal maneira a assumir valores ideais para tornar a vida possível, (HAWKING; MLODINOW, 2012). Este conceito é bastante semelhante a uma "sintonia-fina" realizada por um algoritmo de otimização heurística.

Nas próximas seções serão apresentados, em mais detalhes, o Princípio Antrópico, bem como o algoritmo heurístico nele baseado.

5.2.1 Princípio Antrópico

O princípio Antrópico foi formalmente definido pelo astrofísico Brandon Carter em 1974, (CARTER, 1973). Na sequência, John D. Barrow e Frank J. Tipler ampliou este conceito e o compilou suas ideias no livro: *Cosmological Principle*, (BARROW; TIPLER, 1986).

Segundo o Princípio Antrópico, para que condições adequadas para existência da hu-

manidade sejam atingidas, existe uma série de requisitos a serem atendidos. Coisas como uma diferente relação entre as massas do elétron e do próton poderia impedir a existência de estruturas mais complexas de matéria, (HAWKING; MLODINOW, 2012). Se o módulo da força eletromagnética fosse alterado por uma pequena quantidade, as moléculas orgânicas não poderiam se agrupar, (HAWKING; MLODINOW, 2012). Se, por exemplo, a distância entre o Sol e a Terra fosse diferente, a humanidade provavelmente não teria se desenvolvido. Se a força gravitacional fosse minimamente alterada, as órbitas dos planetas não teriam se formado e, conseqüentemente, não haveria vida, (COMITTI, 2011).

Na realidade, o princípio Antrópico parte da premissa de que dada a existência da humanidade, apenas as histórias do universo compatíveis com este fato podem ser consideradas para os modelos físicos deste universo. Este princípio é conhecido como Princípio Antrópico Fraco (WAP, *Weak Anthropic Principle*), (HAWKING; MLODINOW, 2012). A partir deste princípio, entende-se que as características do universo e suas leis físicas permitiram a existência da humanidade.

Devido ao grande número de coincidências necessárias para que a vida pudesse se desenvolver no universo, e a baixa probabilidade de que isso pudesse ocorrer, alguns físicos desenvolveram a teoria que afirma que o universo evoluiu com o objetivo de permitir a existência de observadores conscientes. Este conceito é conhecido como Princípio Antrópico Forte (SAP, *Strong Anthropic Principle*), e este é um princípio bastante controverso na Física, (KALLOSH; LINDE, 2003).

De acordo com o SAP, o universo evoluiu de tal forma que fosse possível a existência da Humanidade, (HAWKING; MLODINOW, 2012). Em outras palavras, as quantidades físicas e cosmológicas evoluíram de forma a permitir a criação da vida.

O Princípio Antrópico na sua versão forte é bastante controverso. Existem muitos argumentos contrários ao mesmo, como o fato de que valores das constantes fundamentais incompatíveis com o desenvolvimento da vida inteligente nunca poderiam ser observados, (STARKMAN; TROTTA, 2006). Entretanto, nesta tese, o Princípio Antrópico é utilizado apenas como uma inspiração para o desenvolvimento de um novo algoritmo de otimização heurística.

A ideia por trás do Princípio Antrópico parece ser promissora, como inspiração, para um algoritmo heurístico. Neste contexto, propõem-se de maneira inédita, até onde vão os conhecimentos do autor, o Algoritmo do Princípio Antrópico (APA), que faz uso das ideias apresentadas acima e as aplica em um *framework* de otimização, com foco inicial de aplicação na área de Identificação de Sistemas. Os conceitos necessários a este novo algoritmo serão apresentados na próxima seção.

5.2.2 Estrutura do APA

O Algoritmo do Princípio Antrópico possui o universo, \mathbb{U} , como elemento básico. A sua estrutura pode ser definida como:

$$\mathbb{U} = \{\mathcal{C}, \mathcal{L}\}, \quad (5.6)$$

sendo: \mathcal{C} a característica e \mathcal{L} as leis físicas de um universo \mathbb{U} . Tais conceitos podem ser melhor definidos da seguinte forma:

- Característica: conjunto de atributos que caracterizam um universo, por exemplo,

$$\mathcal{C} = \{c_1, c_2, \dots, c_n\}. \quad (5.7)$$

- Leis Físicas: Conjunto \mathcal{L} de equações que atualizam as características \mathcal{C} de um universo:

$$\mathcal{L} = \{l_1, l_2, \dots, l_n\}. \quad (5.8)$$

A relação entre \mathcal{C} e \mathcal{L} pode ser expressa de várias formas, entre elas tem-se:

$$c_i^{(k+1)} = l_i \{c_i^{(k)}\}, \quad (5.9)$$

em que k representa a k -ésima iteração do algoritmo. A equação (5.9) propõem que a i -ésima característica de um universo seja atualizada pela i -ésima lei física, usando apenas os valores correntes das características.

A “maneira” com a qual a característica de um universo \mathbb{U} é atualizada ao longo das iterações é chamada Historia de \mathbb{U} . A mesma é representada por \mathcal{H} e expressa a evolução de um universo entre o seu estado inicial, $\mathbb{U}^{(0)}$, e atual, $\mathbb{U}^{(k)}$:

$$\mathcal{H}: \quad \mathbb{U}_i^{(0)} \rightsquigarrow \mathbb{U}^{(k)}, \quad (5.10)$$

em outras palavras, a Historia de um universo descreve o desenvolvimento das características do melhor universo, através de suas leis e operadores:

$$\{\mathcal{C}^{(0)}, \mathcal{L}\} \rightsquigarrow \{\mathcal{C}^{(k)}, \mathcal{L}\}.$$

Um indivíduo \mathcal{J} pode ser gerado em um universo, \mathbb{U} , se as características \mathcal{C} deste universo forem favoráveis para a criação deste tipo específico de vida. A notação utilizada para representar a geração de um indivíduo em um universo é:

$$\mathbb{U} \Rightarrow \mathcal{J}.$$

Um indivíduo gerado por um universo \mathbb{U} é chamado indivíduo produto: $\mathbb{U} \Rightarrow \mathcal{J}_{\mathcal{P}}$. Aqui, pode-se introduzir o conceito de indivíduo de referência $\mathcal{J}_{\mathcal{R}}$. Isto é, um indivíduo idealizado gerado por um universo ideal \mathbb{U}_R , cujas características \mathcal{C}_R são as mais favoráveis para a geração de um tipo específico de vida, desta forma tem-se: $\mathbb{U}_R \Rightarrow \mathcal{J}_{\mathcal{R}}$. Assim, pode-se considerar que existe um tipo específico de vida em um universo \mathbb{U} , se seu indivíduo produto $\mathcal{J}_{\mathcal{P}}$ é similar ao indivíduo de referência $\mathcal{J}_{\mathcal{R}}$. Propõem-se medir esta similaridade com auxílio da função $\mathcal{F}(\cdot)$. Isto é, o universo \mathbb{U} gera este tipo específico de vida se sua imagem na função $\mathcal{F}(\mathcal{J}_{\mathcal{R}}, \mathcal{J}_{\mathcal{P}})$ romper um limiar α_v . Em outras palavras, se $\mathcal{F}(\mathcal{J}_{\mathcal{R}}, \mathcal{J}_{\mathcal{P}}) < \alpha_v$, passa a existir um tipo específico de vida no universo \mathbb{U} e, conseqüentemente, o universo pode ser chamado "universo vivo", que é designado por \mathbb{U}^V .

No APA, o indivíduo de referência, $\mathcal{J}_{\mathcal{R}}$, é representado por um conjunto de requisitos almeçados para a solução ótima do problema. Por exemplo, se o problema de otimização for um problema de minimização da função $f(x_1, x_2, \dots, x_n)$, sendo $f: \mathbb{R}^n \rightarrow \mathbb{R}$, o $\mathcal{J}_{\mathcal{R}}$ é representado pela imagem mínima de $f(x_1, x_2, \dots, x_n)$. Se o problema é de minimização de uma função multiobjetivo, o conjunto dos valores desejados para as variáveis, envolvidas no processo de otimização, representam as restrições que caracterizam o indivíduo de referência. Em um problema de identificação de sistemas, cada par de amostras (entrada e saída do sistema a ser identificado) é considerado como uma restrição para a existência da vida, e o conjunto destas restrições representa o indivíduo de referência.

Agora, o desenvolvimento acima pode ser associado a um algoritmo de otimização heurística. No APA, as variáveis a serem otimizadas são representadas pelas características de um universo, por exemplo, o valor da solução global de um problema de otimização é representado pelas características \mathcal{C}_R do universo de referência \mathbb{U}_R . A função $\mathcal{F}(\cdot)$ pode ser

definido como a diferença entre a função de avaliação de um dado indivíduo produto e:

- 1- A solução global provida por $J_{\mathcal{R}}$ no caso de um problema de otimização. ou,
- 2- A melhor solução encontrada na corrente iteração do algoritmo na minimização de uma dada função.

O APA foi desenvolvido como um algoritmo de otimização multi-agente, assim, consequentemente, existe um conjunto de universos candidatos para criação da vida. neste contexto, o conjunto \mathbb{M} de universos, $\{\mathbb{U}_1, \mathbb{U}_2, \dots, \mathbb{U}_m\}$, é chamado Multiverso e pode ser expresso como:

$$\mathbb{M} = \{\mathbb{U}_1, \mathbb{U}_2, \dots, \mathbb{U}_m\}. \quad (5.11)$$

Na inicialização do algoritmo as características $\mathcal{C} = [c_1, c_2, \dots, c_n]$, de cada universo, são geradas aleatoriamente, e as leis físicas de cada universo, \mathcal{L} , podem ser regras de atualização de diferentes tipos. Por exemplo, considere como possível lei a equação a diferenças na seguinte forma:

$$l_i: \quad c_i^{(k+1)} = a_1 c_i^{(k)} + a_2 c_i^{(k-1)} + \dots + a_n c_i^{(k-n+1)}, \quad (5.12)$$

em que os coeficientes a_1, a_2, \dots, a_n são inicializados randomicamente. As condições iniciais para a equação (5.12) podem ser definidas a partir dos valores prévios da característica, c_i . Se estas variáveis estiverem indisponíveis, seus valores são assumidos como sendo zero.

Um outro exemplo de uma possível lei física pode ser observada na equação (5.13),

$$l_i: \quad \begin{cases} \Delta c_i^{(k+1)} = a_1 c_i^{(k)} \Delta c_i^{(k)} + a_2 \Delta \mathcal{F}^{(k)}(.) \\ c_i^{(k+1)} = c_i^{(k)} + \Delta c_i^{(k+1)} \end{cases} \quad (5.13)$$

em que os coeficientes $a_1, a_2 \in \mathbb{R}^{*+}$ são também aleatoriamente inicializados e $\Delta \mathcal{F}^{(k)} = \mathcal{F}^{(k)} - \mathcal{F}^{(k-1)}$. Neste tipo de leis, a atualização das características é indireta, utilizando passos Δc_i , e existe uma realimentação, implementada pela última imagem da função de avaliação, $\mathcal{F}(\cdot)$, visando o ajuste do tamanho de passo mais adequado.

Os coeficientes a_1 e a_2 são inicialmente todos positivos, mas seus sinais são alterados,

como mostrado na tabela 5.1, em cada iteração. Isso é valido se a característica, c_i , é positiva. Em caso contrário os sinais de a_1 e a_2 são invertidos.

Tabela 5.1: Sinais de a_1 e a_2 relativas a lei (5.13).

$\Delta c^{(k)}$	$\Delta \mathcal{F}^{(k)}$	a_1	a_2	$\Delta c^{(k+1)}$
-	-	-	-	-
-	+	-	+	+
+	-	+	-	+
+	+	-	-	+

O objectivo desta troca de sinais é manter $\Delta \mathcal{F}^{(k)}$ negativo. Os sinais dos coeficientes a_1 e a_2 são escolhidos dependendo dos sinais das duas primeiras colunas da tabela 5.1, visando obter o correto sinal para $\Delta c^{(k+1)}$, mostrado na terceira coluna da mesma tabela.

É importante mencionar que apenas dois tipos de leis físicas foram apresentadas aqui, mas muitas outras são possíveis.

Para que a estrutura do algoritmo proposto possa se tornar mais clara, são definidas três diferentes classes de Universos. O universo \mathbb{U}_i de \mathbb{M} que apresente a melhor característica \mathcal{C} , por exemplo, o menor valor na função de avaliação, é chamado Universo propagante, representado por: \mathbb{U}^* . Este universo recebe este nome porque o mesmo propaga suas características para os universos menos desenvolvidos, como será visto na seção 5.2.2.2.

Duas outras classes de universos podem ser definidas: os universos promissores \mathbb{U}^\diamond e os universos estagnados \mathbb{U}^\dagger . Os universos promissores possuem leis físicas favoráveis para a evolução destes universos, em outras palavras, são universos com leis físicas que melhoram, em uma iteração, a imagem da função de avaliação, $\mathcal{F}^{(k+1)}(.) < \mathcal{F}^{(k)}(.)$, em um problema de minimização. No sentido contrário, os universos com leis físicas que vão contra a evolução do universo em p iterações, ou seja, $\mathcal{F}^{(k+p)}(.) > \mathcal{F}^{(k)}(.)$, são chamados universos estagnados \mathbb{U}^\dagger .

É importante mencionar que no APA o indivíduo de referência, $\mathcal{J}_{\mathcal{R}}$, é fixo e o “meio”, \mathbb{U} , é “adaptado” a ele, enquanto no Algoritmo Genético, (AG) o meio é fixo e o indivíduo (cromossomo) evolui.

Para que o APA possa ser definido como algoritmo, os principais operadores propostos para o mesmo serão apresentados nas próximas seções.

5.2.2.1 Atualização de Característica

O operador de atualização de característica tem o objectivo de sistematicamente alterar a característica \mathcal{C} do universo \mathbb{U} , através de suas leis físicas \mathcal{L} . Isto é feito de tal forma que a característica, c_i , é atualizada pela sua lei física correspondente, l_i , como ilustrado na figura 5.4.

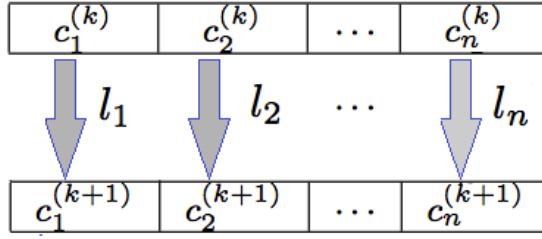


Figura 5.4: Atualização das características, $\mathcal{C} = c_1, c_2, \dots, c_n$, através das leis físicas, $\mathcal{L} = l_1, l_2, \dots, l_n$.

O Operador de Atualização Característica pode ser implementado de diferentes formas, dependendo da estrutura das leis físicas utilizadas. Para exemplificar isto, dois tipos de atualização de características são apresentadas a seguir:

- 1- Atualização direta: em que a lei, l_i , age diretamente sobre a característica, c_i , como mostrado na equação (5.12). Neste tipo de lei física, cada característica tem sua própria dinâmica, e cabe ao algoritmo selecionar as leis que possuam a tendência de levar a função de avaliação a valores menores;
- 2- Atualização por Δc : em que a lei, l_i , gera uma certa quantia Δc_i , que é somada a característica c_i , como mostrado na equação (5.13). Neste tipo de lei física, o passo $\Delta c_i^{(k+1)}$ é calculado levando em conta o último valor de $\Delta c_i^{(k)}$ e o último valor da função de avaliação $\mathcal{F}(\cdot)$. Assim, a lei cria mecanismo de controle, uma realimentação, que visa evitar a divergência das características.

5.2.2.2 Operador de Propagação de Característica

O universo que atinge a condição de universo propagante pode transferir suas características integralmente aos universos menos evoluídos. O operador de propagação de característica é representado pelo símbolo \gg e expressa a propagação de todas as características, $\mathcal{C} = c_1, c_2, \dots, c_n$, do universo propagante, \mathbb{U}^* , para o universo \mathbb{U} :

$$\{\mathcal{C}^*, \mathcal{L}^*\} \gg \{\mathcal{C}, \mathcal{L}\},$$

ou simplesmente:

$$\mathbb{U}^* \gg \mathbb{U}.$$

A ação deste operador pode ser observada na figura 5.5.

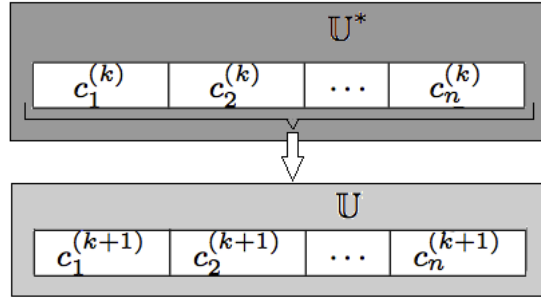


Figura 5.5: Operador de propagação de característica de um universo propagante, \mathbb{U}^* , para um universo qualquer, \mathbb{U} , ou simplesmente: $\mathbb{U}^* \gg \mathbb{U}$.

É importante mencionar que, desde que as leis físicas de cada universo não são as mesmas, as características propagadas são atualizadas de formas diferentes, varrendo o espaço de busca de diferentes formas.

Cabe salientar também que o este operador age apenas sobre as características \mathcal{C} de um universo \mathbb{U} , não alterando as leis físicas deste universo. Neste contexto, este operador também pode ser aplicado de um universo vivo \mathbb{U}^V para um universo promissor \mathbb{U}^\diamond :

$$\mathbb{U}^V \gg \mathbb{U}^\diamond.$$

Este procedimento visa propagar todas as boas características, \mathcal{C}^V , dos universos vivos, \mathbb{U}^V , para os universos, \mathbb{U}^\diamond , que possuem leis físicas favoráveis, \mathcal{L}^\diamond . Assim, as características que apresentam potencial são levadas à leis físicas que também se mostrem promissoras, e seus valores podem ser evoluídos.

5.2.2.3 Operador de “Big Bang”

A partir dos conceitos da Física, pode-se entender que o Universo foi gerado por um “Big Bang”, (HAWKING; MLODINOW, 2012). No APA, o operador de “Big Bang”, $\mathcal{B}(\cdot)$, randomicamente gera um novo universo \mathbb{U} , e substitui um universo estagnado \mathbb{U}^\dagger . Em outras palavras, as regiões não promissoras do espaço de busca são descartadas e a busca se inicia em uma nova região aleatória.

Para se evitar o efeito de busca aleatória gerado pelo operador de “Big Bang”, a taxa de ocorrência deste operador deve ser inferior a 10% dos universos do multiverso, em cada iteração.

5.2.2.4 Operador de Armagedom

Similarmente a uma catástrofe natural, o operador de Armagedom, $\mathcal{A}(\cdot)$, perturba algumas características de um universo. Esta perturbação ocorre a partir de constantes aleatórias, $|\Delta c|$, somadas aos valores correntes das características do universo que sofre ação deste operador, como pode ser observado na figura 5.6.

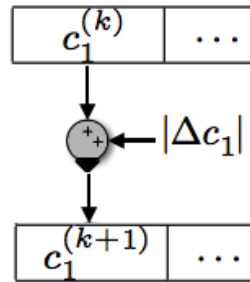


Figura 5.6: Operador de Armagedom, $\mathcal{A}(\cdot)$, atuando sobre a característica c_1 .

O efeito deste operador pode ser sintonizado pela taxa de Armagedom, permitindo que o algoritmo realize buscas locais em torno dos valores atuais das características de um universo, sem chegar-se a busca aleatória.

5.2.3 Pseudo Código do Algoritmo do Princípio Antrópico

Para que os conceitos visto até então possam ser concatenados, pode-se visualizar a dinâmica do APA no algoritmo 9.

É importante mencionar que a ação de cada operador do APA pode ocorrer em alguns universos, no multiverso, em algum instante de tempo. A taxa de ocorrência destes operadores pode tornar a convergência do algoritmo mais rápida ou mais lenta. O intervalo de valores possíveis na inicialização dos coeficientes das leis também é muito importante para estabilidade do algoritmo.

A escolha destes parâmetros pode ser vista com uma “sintonia” empírica do algoritmo, e é um processo necessário para cada problema a ser resolvido.

5.2.4 Exemplos de Aplicação

Nesta seção alguns exemplos de otimização e identificação utilizando o APA serão apresentados. Os experimentos foram realizados em um Mac com 2.2 GHz, processador Intel, 4.0 GB RAM. Os códigos foram escritos e executados no Matlab R2013. O sistema operacional utilizado foi o Mac OS Lion.

Algoritmo 9: Algoritmo do Princípio Antrópico.

Data: Conjunto de Universos $\mathbb{U}_n \in \mathbb{M}$;
 $\mathbb{U}_R \Rightarrow \mathcal{J}_R$;
for $i=1$ *until* n **do**
 Inicializar aleatoriamente \mathcal{C} para \mathbb{U}_i ;
 Inicializar aleatoriamente \mathcal{L} para \mathbb{U}_i ;
 $\mathbb{U}_i \Rightarrow \mathcal{J}_P$;
 Calcular $\mathcal{F}_i(\mathcal{J}_R, \mathcal{J}_P)$ for \mathbb{U}_i ;
 $\mathbb{U}^* \leftarrow \min \{\mathcal{F}(\mathcal{J}_R, \mathcal{J}_P)\}$;
while $\mathcal{F}_{\min}(\mathcal{J}_R, \mathcal{J}_P) > \text{Critério_Parada}$ **do**
 for $j=1$ *until* n **do**
 Atualizar \mathcal{C} for \mathbb{U}_j ;
 for $k=1$ *until* n_{prop} **do**
 $\mathbb{U}^* \gg \mathbb{U}_k$;
 $\mathbb{U}^V \gg \mathbb{U}_k^\diamond$;
 $\mathbb{U}^A \leftarrow \mathcal{A}(\mathbb{U}^\dagger)$;
 $\mathbb{U}^B \leftarrow \mathcal{B}(\mathbb{U}^A)$;
 $\mathbb{U} \leftarrow \mathbb{U}^B$;
 for $l=1$ *until* n **do**
 $\mathbb{U}_i \Rightarrow \mathcal{J}_P$;
 Calcular $\mathcal{F}_l(\mathcal{J}_R, \mathcal{J}_P)$ for \mathbb{U}_l ;
 $\mathbb{U}^* \leftarrow \min \{\mathcal{F}(\mathcal{J}_R, \mathcal{J}_P)\}$;

5.2.4.1 APA Aplicado a Função de Rosenbrock

Para exemplificar a utilização do APA na solução de um problema de otimização, considere a função de Rosenbrock:

$$f(x) = (a - x_1)^2 + b(x_2 - x_1^2)^2, \quad \text{para } x = [x_1, x_2]^T, \quad (5.14)$$

em que f é uma função não convexa, $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, com $a = 1$ e $b = 100$ nos testes de otimização global. Neste cenário, o objetivo da otimização consiste em encontrar um vetor $x^* = [x_1, x_2]$ que minimiza $f(x)$:

$$x^* = \operatorname{argmin} \{f(x)\}, \quad (5.15)$$

sendo que o mínimo global da função $f(x)$ localizado em $x^* = [1, 1]$, valor para o qual a função de avaliação tem o valor $f(x^*) = 0$.

Neste exemplo, para o APA, cada universo possui 2 características, $[c_1, c_2]$, que representam os valores $[x_1, x_2]$ da função de Rosenbrock, e duas leis, $[l_1, l_2]$, responsáveis pela atualização destas características. É importante mencionar que, neste problema, a função de avaliação é a função $f(x)$ e que as leis físicas, \mathcal{L} , usadas neste exemplo, são classes de equações com a estrutura apresentada em (5.13).

Para ilustrar o processo de convergência do algoritmo, o mínimo de uma função de Rosenbrock foi buscado pelo APA. Em todas as execuções do algoritmo nesta seção, a taxa de propagação de característica foi iniciada em 50% e terminada em 30%, a taxa de Big-Bang iniciou-se em 5% e finalizando em 10%, já a taxa de Armagedon iniciou-se em 5% terminando em 15%, sendo estas taxas variadas linearmente ao longo das iterações.

Para efeito de ilustração, o algoritmo foi executado por 10 iterações. A melhor solução encontrada em cada iteração foi plotada na figura 5.7 e detalhada na tabela 5.2.

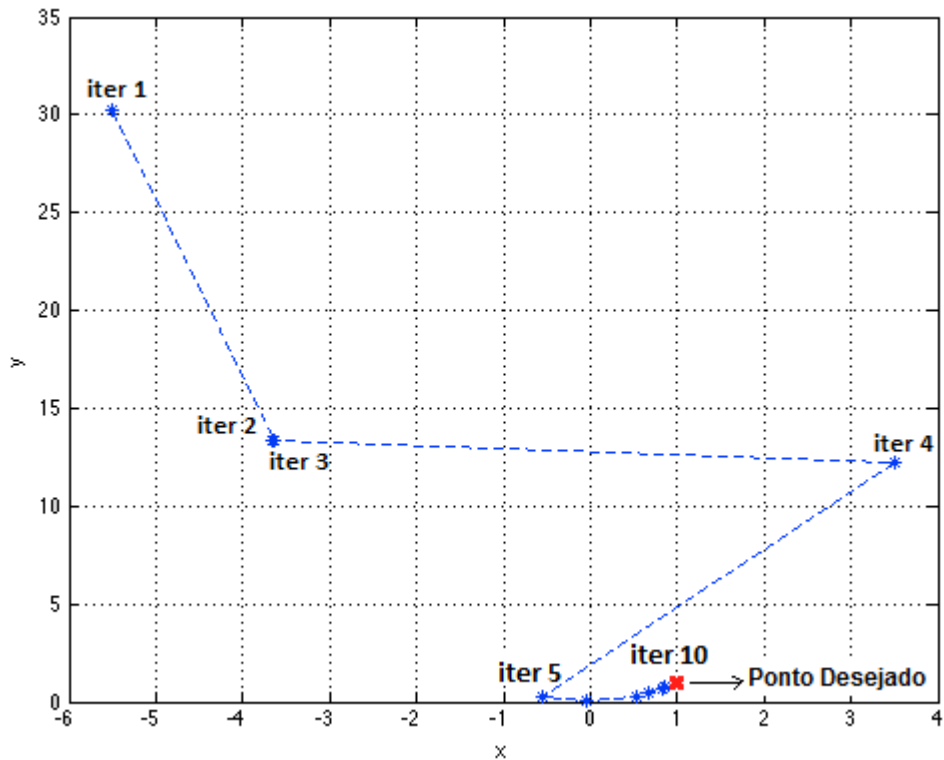


Figura 5.7: História de um Multiverso. A evolução das características do melhor universo durante convergência do APA.

Tabela 5.2: Características dos melhores universos em cada iteração.

Iter.	c_1	c_2	Δc_1	Δc_2	$\Delta \mathcal{F}$
1	-5,5024	30,2659	NA	NA	NA
2	-3,6278	13,3869	-39,2095	-7,2623	$-1,5511 \times 10^8$
3	-3,6582	13,3894	0,0305	0,0025	-72,4581
4	3,4940	12,2212	-7,1522	-1,1682	$-8,0588 \times 10^7$
5	-0,5403	0,2111	-0,4129	-2,9837	-523,4355
6	-0,0485	0,0877	-0,9072	-2,3368	-230,3496
7	0,5232	0,2301	-0,6340	-0,3256	-1,9760
8	0,6736	0,4385	0,0023	0,2083	$-1,9823 \times 10^4$
9	0,8378	0,7159	0,1642	0,1693	$-1,3187 \times 10^4$
10	0,8448	0,7276	0,0070	0,0117	$-1,2819 \times 10^5$

Da tabela 5.2, pode-se observar que o APA não atingiu o ponto ótimo [1, 1] em apenas 10 iterações. Entretanto, o processo de convergência do algoritmo pode ser observado claramente.

Tabela 5.3: Leis físicas dos melhores universos em cada iteração.

	l_1		l_2	
Iter.	a_1	a_2	a_1	a_2
1	0,0556	$2,0082 \times 10^{-8}$	0,6309	$6,5930 \times 10^{-8}$
2	0,3287	$9,9367 \times 10^{-8}$	0,1148	$2,3732 \times 10^{-8}$
3	0,7401	$1,6827 \times 10^{-8}$	0,3792	$2,0312 \times 10^{-8}$
4	0,3889	$9,225 \times 10^{-8}$	0,4464	$1,5073 \times 10^{-8}$
5	0,2505	$5,4600 \times 10^{-8}$	0,6141	$2,7179 \times 10^{-8}$
6	0,4815	$1,5400 \times 10^{-8}$	0,2903	$7,9859 \times 10^{-8}$
7	0,5318	$5,4355 \times 10^{-8}$	4,1028	$7,9859 \times 10^{-8}$
8	0,1386	$7,8980 \times 10^{-8}$	0,4280	$7,8184 \times 10^{-8}$
9	0,4242	$3,9651 \times 10^{-8}$	0,2491	$5,8548 \times 10^{-8}$
10	0,4428	$7,6638 \times 10^{-8}$	0,4401	$7,2683 \times 10^{-8}$

A maioria dos universos apresentados na tabela 5.2 não são totalmente conectados no sentido evolutivo. Em outras palavras, estes universos evoluíram de outros universos, no multiverso, que não eram o melhor universo na iteração anterior do algoritmo.

Na tabela 5.2, em 4 iterações o melhor universo evoluiu de si próprio. Isto ocorreu nas iterações 2 para 3, 3 para 4, 8 para 9 e 9 para 10. Nestes casos, é possível calcular manualmente a evolução destes universos usando os valores mostrados nas tabelas 5.2 e 5.3, uma vez que as leis físicas são responsáveis pela atualização das características. Nas outras iterações, o melhor universo foi atingido pela ação de operadores do APA. Na transição das iterações 4 para 5, 6 para 7 e 7 para 8 o melhor universo foi atingido pela propagação das melhores características para universos com leis físicas que permitiram a redução do valor da função de avaliação. Na transição da iteração 1 para 2, o melhor universo foi atingido pelo operador de Armagedom. Neste exemplo o operador de "Big Bang" não gerou nenhum

universo que estivesse na situação de melhor universo.

Para comparar o APA com os algoritmos heurísticos mais conhecidos, o mesmo foi executado por 50 iterações, na resolução da minimização da função de Rosenbrock. Os resultados atingidos foram comparados com os resultados obtidos a partir de um Algoritmo Genético e de um *Particle Swarm Optimization*. Neste caso, o algoritmo proposto atingiu a melhor solução $x^* = [1.0001, 1.0002]$. Como é sabido que o mínimo da função de Rosenbrock é $[1, 1]$, (ROSENBROCK, 1960), pode-se considerar que o resultado obtido é bastante próximo do ótimo, atingindo uma imagem na função igual a 1.1125×10^{-8} . A curva de convergência do APA para esta execução é ilustrada na figura 5.8.

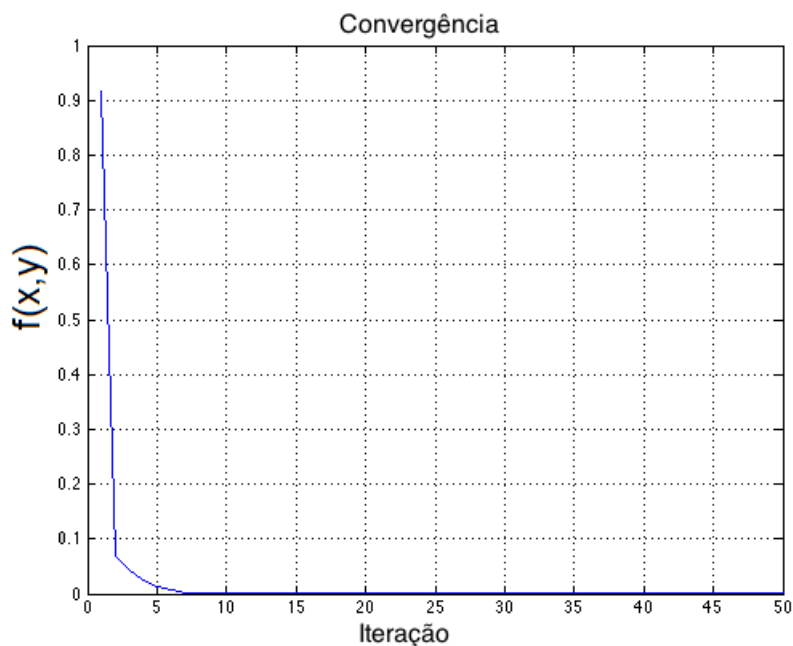


Figura 5.8: Curva de convergência do Algoritmo do Princípio Antrópico.

Para concatenar os resultados da comparação entre o APA, o GA e o PSO, a tabela 5.4 mostra o valor da função objetivo atingida por estes algoritmos em uma única execução. O critério de parada para todos os algoritmos foi de 50 iterações. Os resultados do APA e do AG foram simulados neste trabalho, já o resultado do PSO foi buscado na literatura, (RAVAL; MAKWANA, 2011).

Tabela 5.4: Resultado da minimização da função de Rosenbrock.

Método	APA	GA	PSO
Valor da Função	1,1125e-8	0,0018	0,001341

Pode-se observar na tabela 5.4 que o menor valor da função de avaliação, para a função de Rosenbrock foi obtida a partir do APA, superando, neste caso, os outros algoritmos heurísticos.

5.2.4.2 Exemplos de Identificação de Sistemas

Para testar a capacidade de aplicação do APA em um processo de identificação de sistemas, considere o seguinte sistema estável, linear e invariante no tempo, com entrada $u(k)$ e saída $y(k)$:

$$H(z) = \frac{\mathcal{Z}\{y(k)\}}{\mathcal{Z}\{u(k)\}} = \frac{Y(z)}{X(z)} = \frac{2z - 1}{z^2 - 0.2z + 0.26}. \quad (5.16)$$

em que \mathcal{Z} representa a transformada Z e z é a variável complexa associada a esta transformada. Considere ainda que o modelo buscado faz uso de combinações lineares das funções de Laguerre, como expresso em (5.17).

$$\begin{aligned} M(z) &= d_1L_1(z) + d_2L_2(z) + \dots + d_rL_r(z) \\ &= \sum_{i=1}^r d_iL_i(z), \end{aligned} \quad (5.17)$$

em que $M(z)$ seria o modelo aproximado para o sistema $H(z)$, baseado nas funções de Laguerre, apresentadas a partir da equação (3.34).

Assim, para identificar o sistema $H(z)$, expresso em (5.16), é necessário encontrar o polo, p , que parametriza as funções de Laguerre, e seus coeficientes d_i , que minimizem $\mathcal{F}(\cdot)$, como expresso pela equação (5.18),

$$x^* = \operatorname{argmin} \{\mathcal{F}(\mathcal{J}_{\mathcal{R}}, \mathcal{J}_{\mathcal{P}}(x))\} \quad (5.18)$$

em que $\mathbf{x} = [d_1, \dots, d_r, p]$ é o vetor de parâmetros a serem buscados, com r coeficientes d_i e um polo p . Já $\mathcal{F}(\mathcal{J}_{\mathcal{R}}, \mathcal{J}_{\mathcal{P}})$ é a função de avaliação do algoritmo, expressa pelo erro médio quadrático (MSE), da saída do modelo em relação aos dados observados (medidos) a partir da saída do sistema. É importante mencionar aqui que, em um problema de identificação de sistema, os requisitos necessários ao indivíduo de referência, $\mathcal{J}_{\mathcal{R}}$, são conhecidos a partir dos dados experimentais de entrada e saída. Em outras palavras, cada par de dados de entrada e saída, representa um requisito a existência do $\mathcal{J}_{\mathcal{R}}$, e o conjunto de todos estes pares passa a representar o próprio $\mathcal{J}_{\mathcal{R}}$.

Para identificar o sistema $H(z)$, pode-se recorrer a uma entrada $u(k)$, do tipo PRBS, *Pseudo Random Binary Signal*. Para a etapa de identificação, foram utilizadas 128 amostras de entrada/saída. Outras 128 amostras foram reservadas para a etapa de validação. Neste exemplo, foi composto um modelo utilizando três funções de Laguerre, isto é: $r = 3$ na equação (5.17). Assim, o problema de otimização resultante possui 4 variáveis: 3 coeficientes d_i e um polo p .

Para a resolução deste problema foi utilizado um APA com multiverso de 200 universos, rodando por 50 iterações. O algoritmo foi executado com taxa de propagação de característica começando com 50% e terminando em 30%, Já a taxa de *Big-Bang* começou em 5% e terminou em 10%, enquanto a taxa de Armagedom iniciou-se com 5% e terminou com 10%, variando linearmente. Também, é relevante mencionar que as leis físicas, \mathcal{L} , utilizadas neste exemplo, foram classes de equações a diferenças de primeira ordem, em que a_1 e $a_2 \in \mathbb{R}$ são gerados aleatoriamente para cada lei de cada universo. A estrutura geral destas leis é do tipo:

$$l_i: \quad c_i^{(k+1)} = a_1 c_i^{(k)} + a_2, \quad (5.19)$$

em que $c_i^{(k+1)}$ é a i -ésima característica de um universo, atualizada pela equação (5.19), na iteração $(k + 1)$ do algoritmo. A melhor característica, $\mathfrak{C} = [d \quad p]$, obtida pelo APA é dada em (5.20), que resultou em $\text{MSE} = 6.011 \times 10^{-4}$.

$$d = [1,4350 \quad 2,1601 \quad -0,1512] \\ p = 0,1868. \quad (5.20)$$

Utilizando a informação provida pelo APA, através do polo p , pode-se gerar as funções de Laguerre, $[L_1(z), L_2(z), L_3(z)]$, utilizadas no modelo. (5.17):

$$L_1(z) = \frac{0,982}{z - 0,187}, \quad (5.21)$$

$$L_2(z) = \frac{-0,184z + 0,982}{z^2 - 0,374z + 0,035}, \quad (5.22)$$

$$L_3(z) = \frac{0,034z^2 - 0,367z + 0,982}{z^3 - 0,560z^2 + 0,105z - 0,0065}, \quad (5.23)$$

e finalmente pode-se construir o modelo aproximado $M(z) \approx \frac{Y(z)}{X(z)}$, como mostrado em (5.24).

$$y_{3L}(z) = 1,4350L_1(z) + 2,1601L_2(z) - 0,1512L_3(z), \quad (5.24)$$

sendo o polo, p , que parametriza $L_1(z)$, $L_2(z)$ e $L_3(z)$ igual 0,1868, como pode ser visto em (5.20).

A saída medida do sistema $H(z)$ e a saída obtida pelo modelo aproximado $M(z)$ são mostradas na figura 5.9. Desta figura, pode-se observar que o modelo aproximado representa adequadamente os dados experimentais, resolvendo o problema de identificação proposto nesta seção.

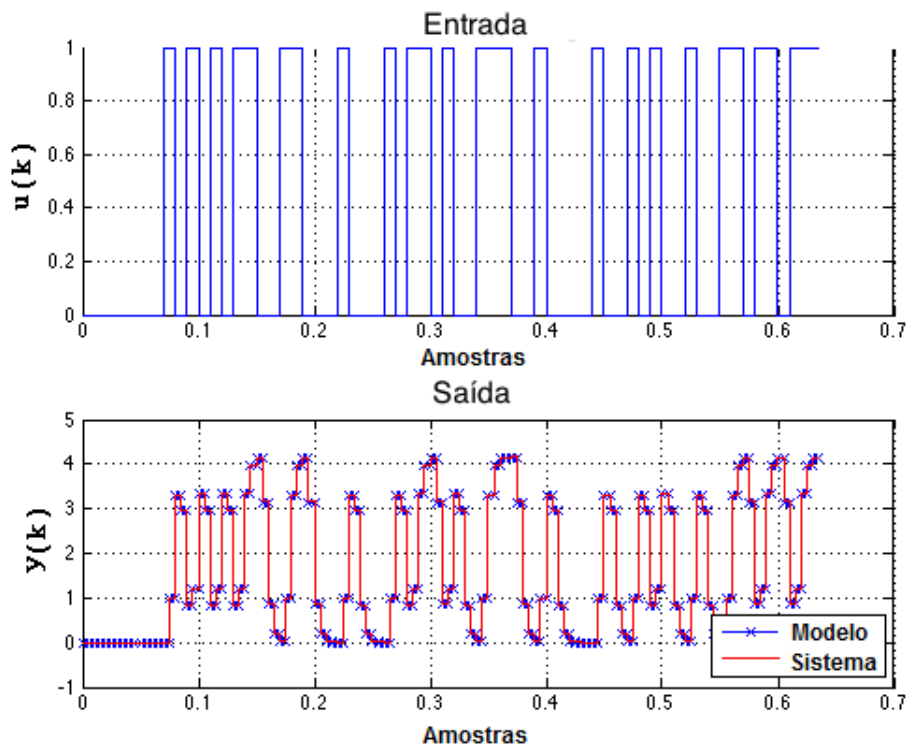


Figura 5.9: Sinal de entrada $u(k)$ aplicado ao sistema $H(z)$ e sua saída estimada $y(k)$, em relação ao conjunto de dados de validação.

A curva de convergência do APA para este problema é mostrada na figura 5.10.

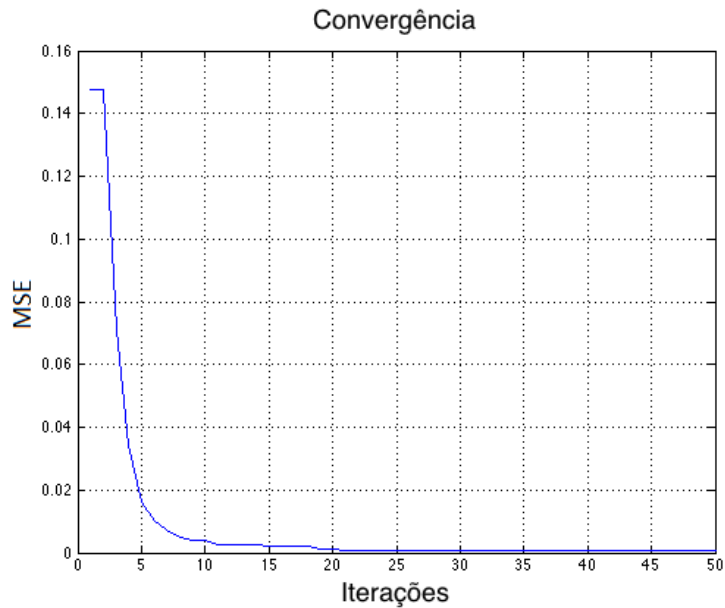


Figura 5.10: Curva de convergência do Algoritmo do Princípio Antrópico.

Na figura 5.10, pode-se observar que o processo de convergência do APA é mais intenso nas primeiras iterações e então a evolução passar a ocorrer mais lentamente até atingir seu valor final.

Para efeitos de comparação o desempenho do APA foi comparado com o AG clássico. Este problema de identificação foi resolvido por ambos os algoritmos usando o mesmo critério de parada: 100 iterações. O MSE resultante dos modelos obtidos pelo APA e pelo GA são mostrados na tabela 5.5.

Tabela 5.5: Resultado da minimização do MSE no problema de identificação de sistemas com funções de Laguerre.

Método	APA	GA
MSE	$6,011 \times 10^{-4}$	$7,110 \times 10^{-4}$

Como pode ser observado na tabela 5.5, o APA atingiu um MSE, *Mean Square Error*, definido em (4.3), sensivelmente melhor que o AG neste exemplo.

5.2.5 Análise Estatística do APA

Com o objetivo de realizar uma comparação justa entre o APA e os outros algoritmos heurísticos, como AG e PSO, cada um deles foi executado por 50 vezes em diferentes problemas de otimização. Cada uma das execuções foi realizada de forma independente, com critério de parada igual a 100 iterações para cada algoritmo. As estatísticas destas execuções são calculadas para que a natureza estocástica destes algoritmos possa ser considerada na

comparação entre os mesmos.

Para realização da comparação estatística entre os algoritmos heurísticos foram escolhidas seis funções de teste, sendo as mesmas mostradas na tabela 5.7. Tais funções são utilizadas como “*benchmark*” para a comparação do APA com os demais algoritmos. As funções f_1 , f_2 e f_3 são unimodais, com dimensão igual a 30. Já a função f_3 é uma função randômica. As funções f_4 e f_5 são multimodais, com 30 dimensões e alguns mínimos locais, enquanto a função f_6 é do tipo polinomial com 2 dimensões, (HE; WU; SAUNDERS, 2009).

Os resultados da comparação entre o APA e os outros algoritmos são apresentados na tabela 5.6. Cabe mencionar que os resultados do AG e do PSO foram buscados na literatura, em He et. al., 2009, (HE; WU; SAUNDERS, 2009). Também é importante comentar que o domínio e o correto mínimo global destas funções é mostrado na tabela 5.7 na terceira e na quarta coluna, respectivamente. Na inicialização das características dos universos do APA, foram seguidos os ranges expressos nestes domínios.

Tabela 5.6: Estatísticas dos Resultados do Problema da Minimização da Função de Rosenbrock.

Função	Algoritmo	Média	Desvio Padrão	COV
f_1	GA	3,1711	1,6621	0,5241
	PSO	3,6927e-37	$2,4598 \times 10^{-36}$	6,6612
	APA	1,5859e-10	$1,6380 \times 10^{-10}$	1,0329
f_2	GA	9749,9145	2594,9593	0,2661
	PSO	1,1979e-3	$2,1109 \times 10^{-3}$	1,7622
	APA	7,4209e-06	$1,3054 \times 10^{-5}$	1,7591
f_3	GA	0,1045	$3,6217 \times 10^{-2}$	0,3466
	PSO	9,9024e-3	$3,5380 \times 10^{-2}$	3,5729
	APA	8,9354	0,5094	$5,7009 \times 10^{-2}$
f_4	GA	0,6509	0,3594	0,5521
	PSO	20,7863	5,9400	0,2858
	APA	34,6328	36,8309	1,0635
f_5	GA	1,0038	$6,7545 \times 10^{-2}$	$6,7289 \times 10^{-2}$
	PSO	0,2323	0,4434	1,9087
	APA	2,1888e-04	$2,4709 \times 10^{-4}$	1,5857
f_6	GA	-1,0298	$3,1314e \times 10^{-3}$	NA
	PSO	-1,0160	$1,2786e \times 10^{-2}$	NA
	APA	-1,0316	$3,8328 \times 10^{-16}$	NA

sendo $COV=(\text{Desvio Padrão})/\text{Média}$, o coeficiente de variação das variáveis avaliadas e NA representa os casos não aplicáveis.

A partir da tabela 5.6, pode-se observar que o APA atingiu valores mais interessantes que AG nas funções f_1 , f_2 , f_5 e f_6 . Na comparação com o PSO, o APA mostrou resultados melhores nas funções f_2 , f_5 e f_6 . O AG e o PSO apresentaram melhores resultados nos casos das funções f_3 e f_4 . A natureza estocástica da função f_3 pode explicar porquê o APA não retornou o melhor resultado neste caso. Em relação a função f_4 , uma vez que ela é

Tabela 5.7: Seis funções de *Benchmark*, para as quais n expressa a dimensão da função, S representa o seu domínio e f_{min} é o mínimo global da função.

Funções de teste	n	S	f_{min}
$f_1(x) = \sum_{i=1}^n x_i^2$	30	$[-100,100]^n$	0
$f_2(x) = \sum_{i=1}^n (\sum_{j=1}^i x_j)^2$	30	$[-100,100]^n$	0
$f_3(x) = \sum_{i=1}^n ix_i^4 + rand[0,1)$	30	$[-1.28,1.28]^n$	0
$f_4(x) = \sum_{i=1}^n (x_i^2 - 10\cos(2\pi x_i) + 10)^2$	30	$[-5.12,5.12]^n$	0
$f_5(x) = \frac{1}{4000} \sum_{i=1}^{30} (x_i - 100)^2 - \prod_{i=1}^n \cos(\frac{x_i - 100}{\sqrt{i}}) + 1$	30	$[-600,600]^n$	0
$f_6(x) = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4$	2	$[-5,5]^n$	-1.0316285

multimodal, é possível que o APA tenha ficado “preso” em um mínimo local, tendo atingido um valor próximo ao apresentado pelo PSO.

Em síntese, o APA atingiu resultados razoáveis na maioria dos casos testados, superando os outros algoritmos em muitos casos.

5.3 CONCLUSÃO

Neste capítulo foram apresentados os principais conceitos relativos aos algoritmos heurísticos, como o clássico Algoritmo Genético e o Algoritmo do Princípio Antrópico, proposto nesta tese.

Inicialmente, foram mostrados os operadores clássicos do AG, como *crossover* e mutação e também operadores mais específicos como o Nelder Mead e operador de Mínimos Quadrados, estes dois últimos com objetivo de acelerar o processo de convergência do Algoritmo.

Foram ainda apresentados os conceitos básicos relativos ao Princípio Antrópico, bem como um algoritmo heurístico nele baseado. Foram detalhados os operadores do APA e o seu mecanismo básico de otimização heurística.

Ressalta-se que o APA difere do AG no quesito de definição de suas direções de busca. No APA a propagação da característica, de um universo propagante para os demais universos, gera uma gama de “direções” de busca. Assim, as características do universo propagante serão atualizadas de diferentes formas, varrendo o espaço de busca de forma diversa na evolução de cada universo.

A partir dos casos estudados pode-se observar que o APA é pouco sensível aos seus parâmetros de controle. Em outras palavras, o algoritmo é robusto a alterações nas taxas dos seus operadores. A exceção a este padrão apresentou-se no range de inicialização dos coeficientes das leis físicas, que podem levar o algoritmo a instabilidade.

Com o objetivo de testar a capacidade do APA em encontrar o mínimo de uma função não

convexa, o algoritmo foi aplicado a função de Rosenbrock e seus resultados foram comparados, em uma única execução, com o AG e o PSO. Desta comparação pode-se observar que o APA atingiu um valor muito próximo do mínimo global da função, superando os resultados apresentados pelo AG e pelo PSO.

Para testar o algoritmo em uma aplicação prática de engenharia, o APA foi aplicado a um problema de identificação de sistemas. Em outras palavras: o APA foi utilizado para encontrar os parâmetros de um modelo de Laguerre para um sistema dinâmico. O resultado obtido foi comparado com o resultado fornecido por um AG. Nesta comparação o APA superou sensivelmente o AG. Pelo baixo MSE alcançado pelo modelo pode-se perceber que o modelo aproximado modela perfeitamente os dados experimentais, como pode-se observar na figura 5.9.

Nas figuras 5.8 e 5.10, pode-se observar que, tipicamente, a convergência do APA é mais intensa nas primeiras iterações, passando então a “evoluir” mais lentamente, mas quase que de forma contínua.

Para comparar as estatísticas do APA, AG e PSO, um conjunto de seis funções de *benchmark* foram empregadas. Cada algoritmo foi executado por 50 vezes. Da tabela 5.6, pode-se observar que o APA atingiu melhores resultados que o AG nas funções f_1 , f_2 , f_5 e f_6 . Na comparação com o PSO, os resultados do APA foram melhores nas funções f_2 , f_5 e f_6 . O AG e o PSO apresentaram melhores resultados nos casos expressos pelas funções f_3 e f_4 . A natureza estocástica da função f_3 pode explicar porque o APA não retornou o melhor resultado neste caso. Em relação a função f_4 , como trata-se de uma função multimodal, pode-se inferir que o APA ficou “preso” em um mínimo local, uma vez que seu resultado é um valor muito próximo do valor também retornado pelo PSO.

Cabe ressaltar o fato de que o APA possui algumas características “simétricas” em relação ao AG. Enquanto um fixa o indivíduo e busca um meio que possa gerá-lo, o outro fixa o meio e evolui os indivíduos. Ambos baseados em princípios evolutivos. Em trabalhos futuros espera-se gerar um algoritmo de otimização heurística que seja uma generalização destes, em que se tenha tanto o meio quanto os indivíduos como variáveis. Considerando que o APA está em seu estado inicial de desenvolvimento, é razoável assumir que muitas melhorias ainda são possíveis. Futuramente, é possível também o uso de métodos baseados no gradiente do erro como leis físicas para o APA, misturando otimização heurística e determinística.

Em relação a aplicabilidade do APA, pode-se constatar que o mesmo pode ser aplicado a uma gama de problemas de otimização e apresenta potencial para resolver problemas do mundo real. Na comparação entre os resultados obtidos com a aplicação do Algoritmo do Princípio Antrópico e outros algoritmos heurísticos, pode-se concluir que o APA apresentou resultados razoáveis em todos os casos estudados, tendo atingido os melhores resultados e estatísticas na maioria dos casos testados.

6 METODOLOGIA PROPOSTA PARA IDENTIFICAÇÃO DE SISTEMAS A PARTIR DE MODELOS ORTONORMAIS E ALGORITMOS HEURÍSTICOS

Neste capítulo serão sistematizadas as metodologias de identificação de sistemas desenvolvidas nesta tese. Inicialmente, na seção 6.1 é mostrada a identificação de sistemas lineares a partir do APA e de funções ortonormais. Na seção 6.2 é apresentado um método de identificação de sistemas não lineares, baseado em funções ortonormais e não linearidades do tipo polinomial. Na sequência, na seção 6.3 é mostrada uma metodologia de identificação de sistemas não lineares, que consiste na utilização de Algoritmos Genéticos na determinação dos parâmetros de um modelo de Volterra-OFB. Em seguida, na seção 6.4 é apresentada uma metodologia de identificação de sistemas não lineares a partir de modelos NARX-OFB, sendo seus polos buscados por um AG e seus coeficientes, pelo método de Levenberg-Marquardt. Finalmente, é apresentado um método de simplificação dos modelos NARX-OFB a partir de Algoritmos Genéticos.

6.1 IDENTIFICAÇÃO DE SISTEMAS LINEARES COM AG E FUNÇÕES ORTONORMAIS

Como mencionado nas seções 3.7.1 e 3.7.2, um sistema SISO e LIT estável, pode ser descrito como uma combinação linear de funções ortonormais. Se as funções ortonormais forem as de Laguerre, o modelo pode ser descrito da seguinte forma:

$$\begin{aligned} y_{nL}(k) &= c_1 v_1(k) + c_2 v_2(k) + \dots + c_n v_n(k) \\ &= \sum_{i=1}^n c_i v_i(k), \end{aligned} \quad (6.1)$$

sendo $y_{nL}(k)$ um modelo composto por n funções ortonormais de Laguerre e $v_i(k)$ a entrada $u(k)$ filtrada pelo filtro de Laguerre de i -ésima ordem, $L_i(z)$, para $i = 1, 2, \dots, n$. Cabe ressaltar que tais modelos são mais adequados a sistemas lineares sobreamortecidos.

Para determinação dos modelos lineares de Laguerre, é necessário encontrar o polo real, $p \in \mathbb{R}$, que parametriza estas funções, assim como os seus coeficientes $[c_1, c_2, \dots, c_n]$.

Se o sistema a ser identificado for linear com características subamortecidas, as funções ortonormais de Kautz são mais adequadas e o modelo pode ser descrito da seguinte forma:

$$\begin{aligned}
y_{nK}(k) &= c_1 w_1(k) + c_2 w_2(k) + \dots + c_n w_n(k) \\
&= \sum_{i=1}^n c_i w_i(k),
\end{aligned} \tag{6.2}$$

sendo $y_{nK}(k)$ um modelo composto por n funções ortonormais de Kautz e $w_i(k)$ a entrada $u(k)$ filtrada pelo filtro de Kautz de i -ésima ordem, $K_i(z)$, para $i = 1, 2, \dots, n$.

Para a determinar os parâmetros do modelo linear de Kautz é necessário encontrar o polo complexo, $p \in \mathbb{C}$, que parametriza tais funções e os coeficientes, $[c_1, c_2, \dots, c_n]$, que as multiplicam.

A busca pelos parâmetros destes modelos pode ser realizada por diferentes métodos, tanto de otimização determinística quanto heurística. No método proposto serão focados os métodos heurísticos, uma vez que estes não precisam do equacionamento exato do sistema e de suas derivadas. Mais especificamente, propõem-se a utilização de um Algoritmo Genético, utilizando seleção por torneio e função de *fitness* baseada na minimização do erro médio quadrático do modelo. A codificação de um cromossomo deste AG, é mostrada a seguir:

c_1	c_2	\dots	c_n	p
-------	-------	---------	-------	-----

em que c_1, c_2, \dots, c_n são os coeficientes das funções ortonormais de Laguerre, e p é o polo que parametriza estas funções. Desta forma, uma população de soluções seria evoluída com objetivo de se encontrar os parâmetros do modelo (3.41), de forma a minimizar o seu erro médio quadrático.

Cabe ressaltar que, na resolução de tal problema, poderia se fazer uso outros métodos de otimização heurística, como o APA, PSO, ACO, entre outros, na determinação dos parâmetros do modelo.

6.2 IDENTIFICAÇÃO DE SISTEMAS NÃO LINEARES, UTILIZANDO ALGORITMOS GENÉTICOS E FUNÇÕES ORTONORMAIS

Os modelos apresentados na seção anterior, através das equações (6.1) e (6.2), restringem-se a modelar sistemas lineares, ou a aproximar sistemas com não linearidade suave. Para modelar sistemas com não linearidade significativa é necessário inserir elementos não lineares no modelo. No método expresso nesta seção foram utilizadas não linearidades do tipo polinomial.

Neste cenário, considere um sistema não linear, invariante no tempo, SISO, com en-

trada $u(k) \in \mathbb{R}$ e saída $y(k) \in \mathbb{R}$, sendo $k \in \mathbb{Z}$, e um conjunto de funções ortonormais, $\psi_1(z), \psi_2(z), \dots, \psi_n(z)$. Considere ainda que a entrada $u(k)$ é filtrada por cada um dos filtros ortonormais. Na sequência, as saídas de cada um dos filtros $\psi_1, \psi_2, \dots, \psi_n$ são elevados aos expoentes g_1, g_2, \dots, g_n , sendo $g_n \in \mathbb{N}$. Estes expoentes expressarão o Grau de Não Linearidade, GNL, do modelo em relação a cada termo.

Se, por exemplo, as funções ortonormais $\psi_1(z), \psi_2(z), \dots, \psi_n(z)$, fossem as funções de Laguerre, $L_1(z), L_2(z), \dots, L_n(z)$, teria-se um modelo com a seguinte estrutura:

$$y_{nL}(k) = c_1[v_1(k)]^{g_1} + c_2[v_2(k)]^{g_2} + \dots + c_n[v_n(k)]^{g_n}, \quad (6.3)$$

sendo $v_i(k)$ a entrada $u(k)$ filtrada pelo filtro de Laguerre de i -ésima ordem, $L_i(z)$, para $i = 1, 2, \dots, n$.

Se fossem utilizadas as funções de Kautz, $K_1(z), K_2(z), \dots, K_n(z)$, teria-se o seguinte modelo:

$$y_{nK}(z) = c_1[w_1(k)]^{g_1} + c_2[w_2(k)]^{g_2} + \dots + c_n[w_n(k)]^{g_n}, \quad (6.4)$$

sendo $w_i(k)$ a entrada $u(k)$ filtrada pelo filtro de Kautz de i -ésima ordem, $K_i(z)$, para $i = 1, 2, \dots, n$.

A parametrização dos modelos (6.3) e (6.4) depende da determinação do polo real, $p \in \mathbb{R}$, que parametriza as funções de Laguerre, ou do polo complexo, $p = a + bi$, que parametriza as funções de Kautz, dos coeficientes de cada termo do modelo, c_1, c_2, \dots, c_n , e de seus respectivos expoentes, g_1, g_2, \dots, g_n .

Para encontrar os parâmetros necessários aos modelos (6.3) e (6.4), propõe-se utilizar um Algoritmo Genético, com objetivo de minimizar o MSE do modelo. A codificação dos parâmetros de um modelo de Kautz, nos genes de um cromossomo do AG, pode ser expressa por:

c_1	c_2	\dots	c_n	a	b
-------	-------	---------	-------	-----	-----

g_1	g_2	\dots	g_n
-------	-------	---------	-------

sendo c_1, c_2, \dots, c_n os coeficientes das funções ortonormais de Kautz e $a + bi$ o polo que as parametriza. No caso de um modelo baseado em funções de Laguerre, tem-se $b = 0$. Para

cada uma das n funções ortonormais utilizadas, tem-se n genes: g_1, g_2, \dots, g_n , responsáveis por considerar a não linearidade em cada uma das n direções ortonormais.

Os genes g_1, g_2, \dots, g_n podem assumir os seguintes valores: $0, 1, 2, \dots, n$. O efeito de cada um destes genes é expresso na tabela 6.1:

Tabela 6.1: Codificação gênica para as não linearidades do sistema.

Valor de g_n	Saída do n-ésimo filtro ortonormal
0	é elevada a 0 (torna-se o <i>offset</i> do modelo)
1	Mantém-se
2	é elevada ao quadrado
\vdots	\vdots
n	é elevada a n

É conveniente mencionar que, no caso da identificação de um sistema linear, os genes, g_n , provavelmente seriam levados, pelo processo evolucionário do AG, aos valores unitários: $[1, 1, \dots, 1]$. A seleção do GNL do modelo foi codificada nos cromossomos do AG, de forma que a convergência do algoritmo defina qual o melhor GNL para o modelo.

Assim, o AG fica responsável por definir quais as direções ortonormais e quais graus de não linearidade são mais interessantes para modelar um determinado sistema. Isto além da busca pelos coeficientes de cada termo do modelo.

A escolha entre os modelos OBF expressos pelas equações (6.1), (6.2), (6.3) e (6.4), ocorre segundo fluxograma mostrado na figura 6.1.

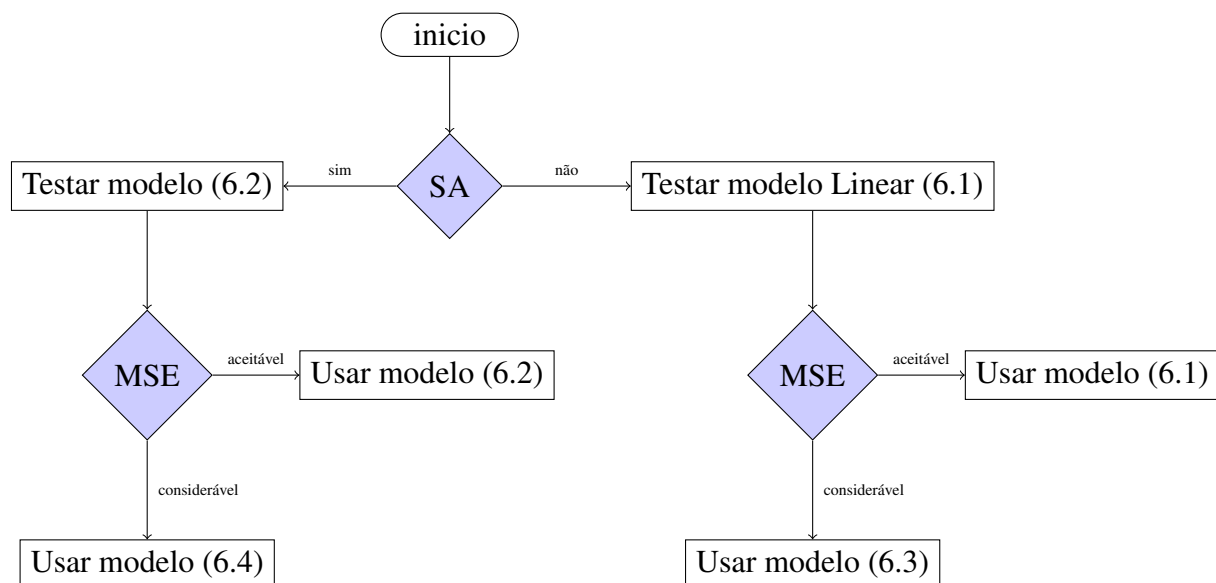


Figura 6.1: Fluxo de decisão entre os modelos OBF lineares (6.1) ou (6.2) e não lineares do tipo (6.3) ou (6.4). Em que o losango de decisão SA é responsável por testar se o sistema a ser identificado tem ou não características subamortecidas e o MSE representa o erro médio quadrático.

Caso não se obtenha resultados razoáveis com os modelos (6.1) ou (6.2), nem com os modelos (6.3) ou (6.4), sugere-se a aplicação dos modelos mostrados nas próximas seções.

6.3 MODELOS DE VOLTERRA-OBF UTILIZANDO ALGORITMOS GENÉTICOS

O modelo expresso na seção anterior é uma opção interessante para aproximar sistemas não lineares, mas pode não representar adequadamente os sistemas que apresentem não linearidades mais severas.

Neste contexto, considere um sistema não linear, invariante no tempo, SISO, com entrada $u(k) \in \mathbb{R}$ e saída $y(k) \in \mathbb{R}$, sendo $k \in \mathbb{Z}$, e um conjunto de funções ortonormais de Kautz, $K_1(z), K_2(z), \dots, K_n(z)$.

Os modelos de Volterra-OBF, como expresso em (3.42), consistem, nesta tese, em séries de Volterra truncadas no *kernel* de segunda ordem, com seus *kernels* expressos através de funções ortonormais. Na metodologia exposta nesta seção, foram escolhidas as funções de Kautz para representar os *kernels* do citado modelo de sistema não linear.

A busca pelos parâmetros de modelos foi realizada através de Algoritmos Genéticos, auxiliados pelos otimizadores locais de Nelder Mead, NM, e Mínimos Quadrados, MQ. Este AG é responsável por buscar o polo complexo, $p \in \mathbb{C}$, que parametriza as funções de Kautz, e os coeficientes reais do modelo (3.42).

Com a ação do operador MQ é possível determinar a melhor solução linear ao problema de identificação. Este operador consiste na aplicação do método dos mínimos quadrados à matriz que possui suas colunas compostas pela entrada, $u(k)$, filtrada pelas funções ortonormais, conforme equação (5.3). Entretanto, como o sistema a ser identificado é de natureza não linear, a solução linear pode ser insuficiente para modelar a dinâmica do sistema em questão. Para contornar tal problema, a solução oferecida pelo operador MQ é repassada ao AG, em forma de um cromossomo, dentro da população de soluções, para que a dinâmica evolutiva do AG tente melhorar a resposta fornecida pelo operador MQ.

Cabe ressaltar que a ação dos operadores NM e MQ já pressupõe a definição dos polos das funções ortonormais, agindo somente na busca pelos coeficientes do modelo. Inicialmente, o AG define um polo complexo e aleatório para cada cromossomo. As funções de Kautz são geradas a partir deste polo, de forma que cada cromossomo possua seu conjunto de funções ortonormais. Só a partir deste ponto são utilizados os operadores NM e MQ. Estes operadores são utilizados apenas a cada ciclo de iterações do AG, buscando evitar a convergência prematura do Algoritmo Genético. Suas funções são, basicamente, as de otimizadores locais.

A operação do Algoritmo Genético apoiado pelos operadores MMQ e de Nelder Mead, podem ser observada no fluxograma da figura 6.2.

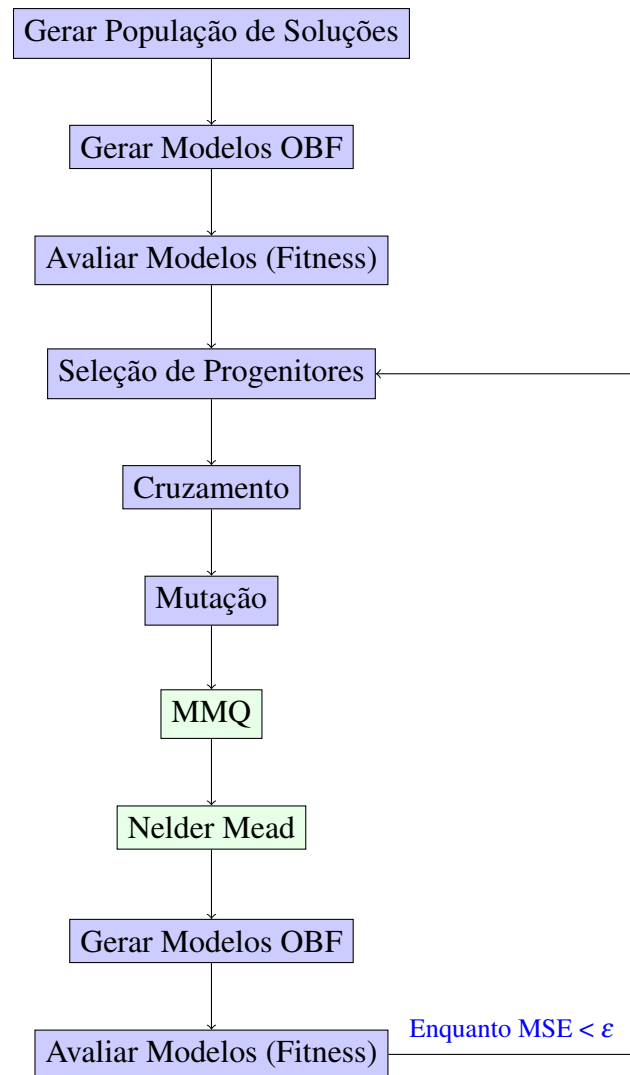


Figura 6.2: Fluxograma que expressa o comportamento do Algoritmo Genético, com utilização dos operadores de Mínimos Quadrados e de Nelder-Mead, aplicados a busca de parâmetros de modelos num processo de Identificação de Sistemas.

Para se estabelecer uma relação de compromisso entre exatidão e complexidade dos modelos Volterra-OBF, o número de funções de Kautz, que compõem a base de funções ortonormais, é variado de um pequeno número até que seja atingida a exatidão necessária ao modelo.

6.4 MODELOS NARX-OBF UTILIZANDO O ALGORITMO DE LEVENBERG MARQUARDT E ALGORITMOS GENÉTICOS

O modelo apresentado na seção anterior é um modelo não linear bastante utilizado devido a sua garantia de estabilidade, entre outras características. Entretanto, os modelos de Volterra sofrem muito com o problema da dimensionalidade, limitação esta que é amenizada, em parte, pelo uso de funções ortonormais, mas ainda assim o problema persiste parcialmente.

Para contornar esta dificuldade, os modelos NARX incluem termos de saída no equacionamento, reduzindo potencialmente o número de termos necessários ao modelo.

Para esta seção, considera-se um sistema não linear idêntico ao descrito na seção 6.3, ou seja, invariante no tempo, SISO, com entrada $u(k) \in \mathbb{R}$ e saída $y(k) \in \mathbb{R}$, sendo $k \in \mathbb{Z}$. Considera-se ainda um conjunto de funções ortonormais de Kautz, $K_1(z), K_2(z), \dots, K_n(z)$, que serão utilizadas na composição de um modelo NARX-OBF.

Como mencionado na seção 3.9, os modelos NARX-OBF, expressos em (3.44), consistem em modelos NARX com seus *kernels* expressos a partir de funções ortonormais. Na metodologia exposta nesta seção, a busca pelos coeficientes do modelo foi realizada através de um algoritmo de Levenberg-Marquardt. Já o polo complexo, $p \in \mathbb{C}$, que parametriza as funções de Kautz, foi encontrado a partir de um Algoritmo Genético.

A escolha do Algoritmo de Levenberg-Marquardt, na determinação dos coeficientes dos modelos NARX, se deve ao fato de que estes modelos, mesmo truncados no *kernel* de segunda ordem, ainda sofrem com o grande número de termos necessários, e, conseqüentemente, de coeficientes a se determinar. Como o método Levenberg-Marquardt possui bom desempenho em otimização com funções não lineares e com número elevado de parâmetros, o mesmo foi escolhido para esta tarefa.

Em relação à busca pelos polos, como estes possuem restrição nos seus valores possíveis, $\|p\| \leq 1$, o Algoritmo Genético mostrou-se uma das formas mais simples de determinação dos mesmos. É importante mencionar que, como dito na seção 5.1, os Algoritmos Genéticos não garantem a determinação de um mínimo global. Assim, os polos encontrados por este método, normalmente, representarão mínimos locais na otimização dos polos das funções de Kautz. Nos modelos NARX-OBF, isto não é um grande problema, uma vez que estes modelos possuem sensibilidade reduzida em relação aos polos das funções ortonormais, se comparada com sua sensibilidade aos coeficientes dos termos que o compõem. Como a busca pelo polo antecede a busca pelos coeficientes, os pequenos desvios do polo obtido pelo AG, em relação ao polo ótimo, podem ser compensados por coeficientes adequados.

Ressalta-se que a utilização dos AGs na determinação dos polos das funções de Kautz tem como função secundária rejeitar possíveis modelos instáveis gerados na população de soluções. Neste contexto, a própria dinâmica evolutiva do AG se encarrega de excluir soluções que levem a modelos instáveis, uma vez que o AG visa minimizar o MSE da sua população de soluções.

Os detalhes da hibridização entre o Algoritmo Genético e o método de Levenberg-Marquardt, com objetivos de acelerar a convergência do processo de busca por parâmetros de um modelo OBF, podem ser observados no fluxograma expresso na figura 6.3.

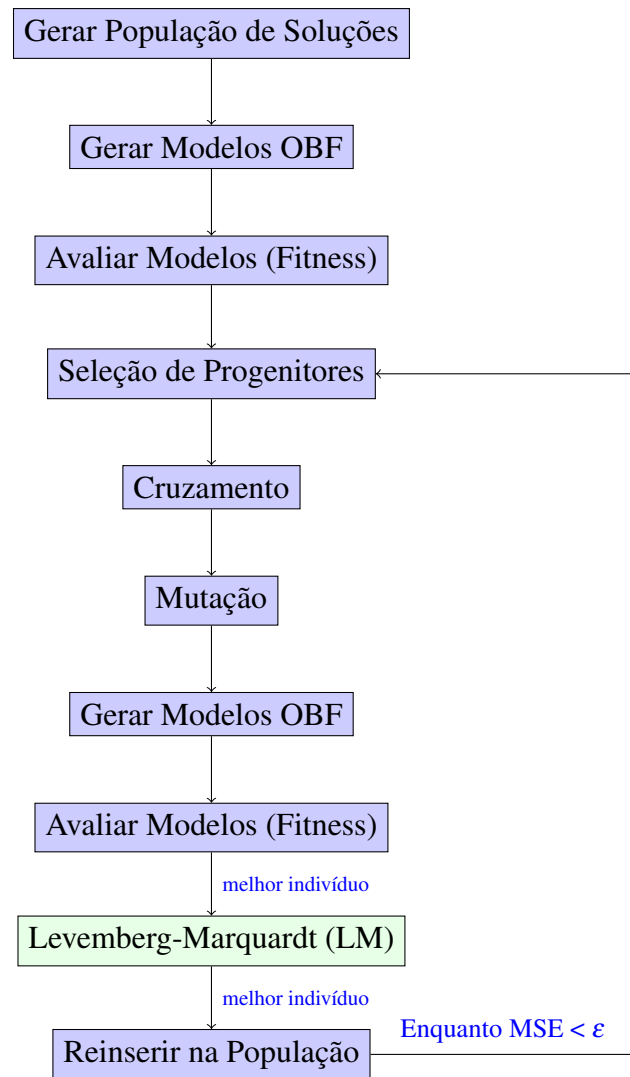


Figura 6.3: Fluxograma que expressa o comportamento do Algoritmo Genético, em hibridização com o método de Levenberg-Marquardt, aplicado a busca de parâmetros de modelos num processo de Identificação de Sistemas.

Cabe ressaltar que no fluxograma mostrado na figura 6.3, o algoritmo possui uma dinâmica muito próxima do AG clássico, que conta com os operadores de Cruzamento e Mutaçao. Entretanto, o melhor indivíduo obtido pelo AG é conduzido ao método de Levenberg-Marquardt, que calcula os coeficientes do modelo visando minimizar o MSE do mesmo. Na sequência um novo cromossomo é formado com os parâmetros fornecidos pelo LM e devolvido a população de soluçoes do AG. Este ciclo se repete até que o MSE do modelo se torne menor que o limiar ϵ , sendo este o critério de parada do algoritmo.

Para auxiliar a escolha do modelo, em identificação de sistemas não lineares, pode-se recorrer ao fluxograma representado a partir da figura 6.4.

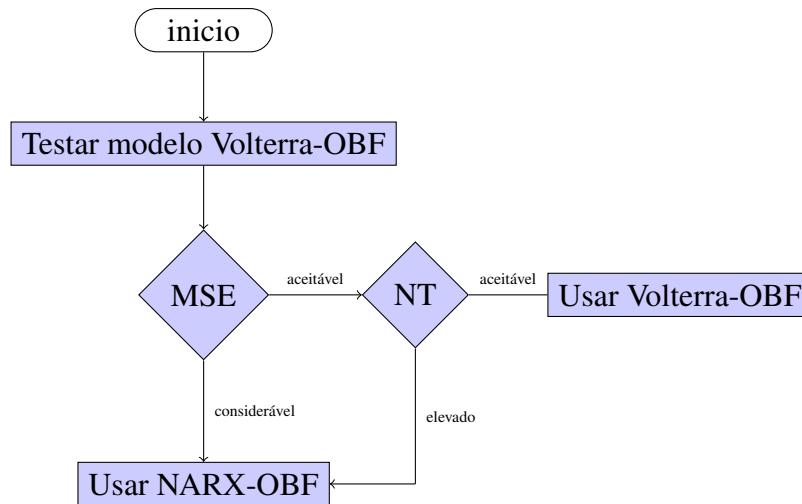


Figura 6.4: Fluxo de decisão entre os modelos Volterra-OBF e NARX-OBF. Em que o losango de decisão NT é responsável por avaliar o número de termos que compõem cada modelo e o MSE representa o erro médio quadrático.

6.5 SELEÇÃO DE TERMOS DE UM MODELO NARX-OBF

Um modelo NARX-OBF pode ser mais compacto que um modelo Volterra-OBF, devido ao uso da realimentação das saídas do sistema. Entretanto, nem todos os termos que compõem o modelo NARX-OBF necessariamente precisam estar no modelo para que este seja representativo. Assim, os termos menos significativos do modelo podem ser excluídos e o modelo se torna mais simples. Algo muito semelhante ao processo de redução de ordem, que pode ser realizado nos modelos NARX polinomiais, em que os regressores mais relevantes são selecionados para compor um modelo simplificado.

Neste cenário, conforme já expresso na seção 3.9.1, para a determinação da estrutura e dos parâmetros de um modelo NARX-OBF propõem-se a utilização de um Algoritmo Genético (AG) combinado com o método de Levenberg-Marquardt (LM). A partir destas idéias, o AG é utilizado na definição da estrutura do modelo, ou seja, na determinação de quais termos do modelo NARX-OBF devem compor o modelo simplificado. O AG é utilizado também na busca do polo que parametriza as funções ortonormais, enquanto o método de Levenberg-Marquardt é utilizado para encontrar os coeficientes deste modelo.

A redução do número de termos de um modelo NARX-OBF se faz necessária porque este modelo na sua forma completa, embora seja mais compacto que seus concorrentes, ainda possui um número elevado de termos, (CAMPELLO; AMARAL; OLIVEIRA, 2007; LEMMA; RAMASAMY; SHUHAIMI, 2010), tornando sua implementação computacional muito exigente. Assim, propõem-se que o AG realize uma seleção dos termos mais relevantes. Para tanto, a função de avaliação (*fitness*) utilizada considera tanto a minimização do erro médio quadrático (*MSE*) quanto a redução do número de termos do modelo, n_{θ} . A

função de *fitness* utilizada no AG foi inspirada pelo critério de AKAIKE, (AKAIKE, 1974), e é expressa pela equação (6.5).

$$fitness = 2 \times 10^{-4} N \ln(MSE) + 2n_{\theta}, \quad (6.5)$$

em que N é o número de amostras realizadas sobre os sinais de entrada e saída do sistema, MSE é o erro médio quadrático e n_{θ} é o número de parâmetros envolvidos no modelo. Ressalta-se que a constante 2×10^{-4} foi escolhida empiricamente de forma a conciliar a ordem de grandeza dos dois termos da equação (6.5).

No contexto deste método, para cada termo do modelo NARX-OBF tem-se um gene, p_n , que pode assumir os valores $\{0,1\}$. Caso este gene possua valor 1, o termo NARX-OBF a ele associado é mantido no modelo, caso o gene seja 0 o termo é desconsiderado. Assim, cada Cromossomo deste AG é composto pelos seguintes elementos:

$$\begin{array}{|c|c|} \hline Real(\mathbf{p}) & Imag(\mathbf{p}) \\ \hline \end{array},$$

$$\begin{array}{|c|c|c|c|} \hline p_1 & p_2 & \cdots & p_n \\ \hline \end{array},$$

sendo \mathbf{p} o polo das funções de Kautz, e os genes do vetor $[p_1, p_2, p_3, \dots, p_n]$ representam a presença ou não de cada termo do modelo NARX-OBF, na sua versão simplificada. Por exemplo, se o gene p_1 for 0 o primeiro termo da série, que compõem o modelo NARX-OBF, é desconsiderado. Se o gene p_2 for 1, o segundo termo da série é mantido. Assim, a dinâmica evolutiva do AG torna-se responsável por selecionar quais os termos do modelos são representativos para o sistema a ser identificado. Realizando uma espécie de seleção dos termos mais relevantes, permitindo que o modelo NARX-OBF simplificado represente um determinado sistema sem um aumento significativo no MSE do modelo.

É importante mencionar que os modelos NARX-OBF são modelos realimentados, e por isso não são sempre estáveis. Isto constituiria um problema para os métodos de convencionais de determinação dos parâmetros deste tipo de modelo. Entretanto, quando a busca pelos parâmetros é feita a partir de uma população de soluções, como no Algoritmo Genético, isto deixa de ser um problema. Os modelos instáveis obtidos são desprezados pela dinâmica de seleção natural do AG. Uma vez que a função de avaliação é baseada no erro médio quadrático, uma solução que implique em um modelo instável, teria um avaliação muito ruim e, conseqüentemente, baixa probabilidade de se reproduzir.

Neste cenário, a partir do método apresentado nesta seção é possível simplificar os modelos NARX-OBF sem uma redução muito significativa no MSE do modelo. Cabe ressaltar que o papel do AG na seleção de termos deste tipo de modelo poderia ser desempenhado por outros algoritmos heurísticos, como PSO ou mesmo o APA.

6.6 CONCLUSÃO

Neste capítulo foram apresentadas as metodologias de identificação propostas para modelos lineares a partir de combinações lineares de funções ortonormais de Laguerre e de Kautz. Foi ainda apresentado um método de identificação de sistemas não lineares, a partir de funções ortonormais, com não linearidade do tipo polinomial na saída, com parâmetros sintonizados por Algoritmo Genético.

Na sequência, foram apresentadas metodologias de identificação de sistemas não lineares a partir do modelo Volterra-OBF e NARX-OBF, sendo os parâmetros do primeiro buscados por Algoritmo Genético e os parâmetros do segundo buscados por Levenberg-Marquardt em conjunto com um AG.

Foi apresentada ainda uma metodologia para seleção dos termos mais relevantes de um modelo NARX-OBF, a partir de um Algoritmo Genético, que ficou responsável por encontrar o polo das funções de Kautz e selecionar os termos mais significativos do modelo, enquanto o método de Levenberg Marquardt foi utilizado para determinar os coeficientes do mesmo.

7 EXPERIMENTOS E VALIDAÇÃO DAS METODOLOGIAS PROPOSTAS

Neste capítulo são apresentados os experimentos realizados neste trabalho com objetivo de validar a metodologia proposta no capítulo 6. Inicialmente, são expostos os resultados da identificação de um sistema linear, realizada a partir de funções de Kautz, com a busca dos parâmetros de modelo realizada pelo APA e posteriormente por um AG. Na sequência, é exposta a identificação de um levitador Magnético vertical a partir de modelos de Volterra e NARX, ambos auxiliados por funções ortonormais, e com parâmetros sintonizados pelo método de Levenberg-Marquardt em conjunto com um AG. Finalmente, é realizada a identificação de um levitador magnético horizontal, a partir de modelos NARX-OBF simplificados a partir de um AG.

7.1 MODELAGEM DE SISTEMAS LIT A PARTIR DE FUNÇÕES DE KAUTZ

Para exemplificar a otimização heurística a partir do Algoritmo do Princípio Antrópico (APA) considere o seguinte sistema LIT e estável:

$$G(z) = \frac{Y(z)}{X(z)} = \frac{2z - 1}{z^2 - 0,2z + 0,26}. \quad (7.1)$$

Considere ainda que busca-se um modelo para tal sistema a partir de combinações lineares das funções de Kautz, como expresso em (6.2). Ressalta-se o fato de que o sistema, expresso em (7.1), é subamortecido. Isto favorece a utilização das funções de Kautz na composição de um modelo que possa representá-lo, (MACHADO; CAMPELLO; AMARAL, 2010).

Para a identificação do sistema $G(z)$, recorre-se a uma entrada, $u(k)$, do tipo PRBS, *Pseudo Random Binary Signal*. Para a etapa de identificação do sistema foram utilizadas 256 amostras de entrada e saída, tendo sido reservados outras 128 amostras para a validação do modelo. Estes dados foram gerados e processados a partir do software Matlab, R2013.

A busca dos parâmetros do modelo expresso em 6.2 será realizada utilizando o APA e o AG, o primeiro procedimento é explicado na seção 7.1.1, e o segundo na seção 7.1.2. A comparação entre os resultados obtidos para cada procedimento é apresentada na seção 7.1.3.

7.1.1 Modelos de Kautz Utilizando APA

Para realizar a identificação do sistema $G(z)$, expresso em (7.1), é preciso encontrar o polo, $p = a + bi$, que parametriza as funções de Kautz, e seus coeficientes, $c = [c_1, c_2, \dots, c_n]$. Os resultados desta identificação, obtidos a partir do APA, são mostrados na figura 7.1.

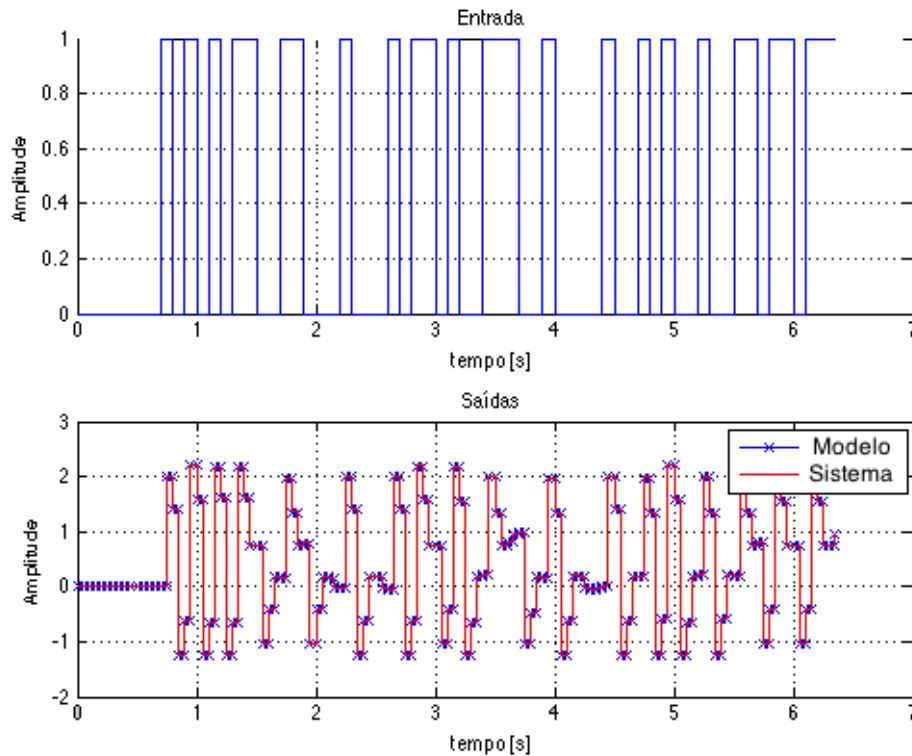


Figura 7.1: Sinal de entrada $u(k)$ aplicado ao sistema $H(z)$ e sua saída $y(k)$.

O melhor resultado obtido pelo APA apresentou um MSE de $1,397 \times 10^{-4}$ e é composto pela característica $\mathcal{C} = [c \ r]$, exposta em (7.2).

$$\begin{aligned} c &= [-0,7162 \quad 2,0690] \\ r &= [0,1006 \quad -0,5000], \end{aligned} \quad (7.2)$$

em que c é composto pelos coeficientes das funções de Kautz e r por seu polo.

A partir do polo das funções de Kautz, expresso em r , pode-se gerar a base de funções, $\mathfrak{B} = [K_1, K_2]$, que servirá ao modelo:

$$K_1(z) = \frac{0,9532z}{z^2 - 0,2012z + 0,2601}, \quad (7.3)$$

$$K_2(z) = \frac{0,9656z^2 - 0,1542z}{z^2 - 0,2012z + 0,2601}, \quad (7.4)$$

Com isto, chega-se ao seguinte modelo:

$$\hat{y}_{2K}(k) = -0,7162w_1(k) + 2,0690w_2(k),$$

sendo $w_i(k)$ a entrada $u(k)$ filtrada pelo filtro de Kautz de i -ésima ordem, $K_i(z)$, para $i = 1$ e 2 .

A curva de convergência do APA, na busca pelos parâmetros em um modelo de Kautz, é mostrada na figura 7.2.

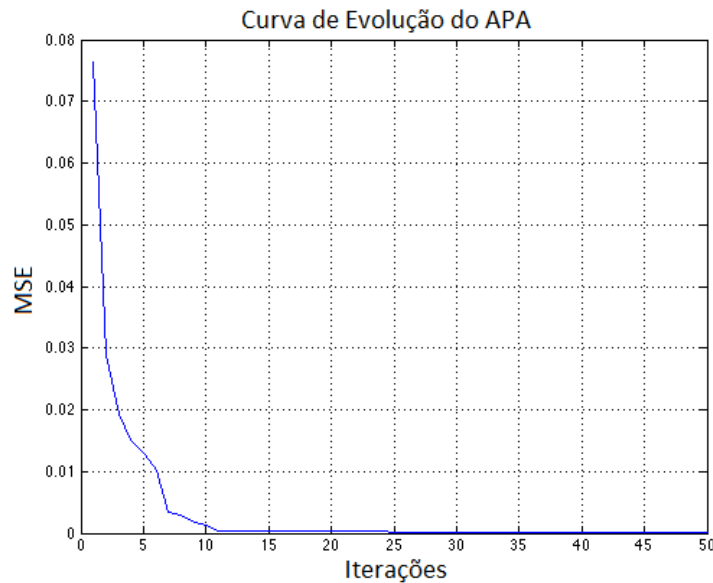


Figura 7.2: Convergência do Algoritmo do Princípio Antrópico.

7.1.2 Modelos de Kautz Utilizando AG

Utilizando-se os mesmos dados do problema exposto na seção 7.1.1 e aplicando-se Algoritmo Genético na determinação dos parâmetros do modelo, chega-se à resposta mostrada na figura 7.3:

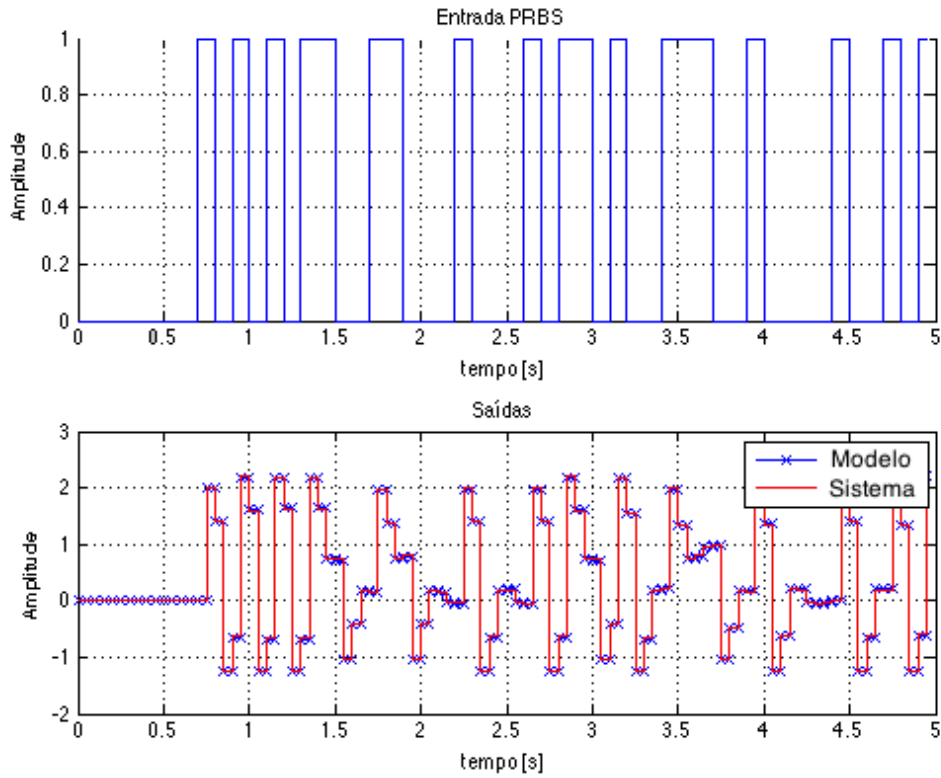


Figura 7.3: Sinal de entrada $u(k)$ aplicado ao sistema $H(z)$ e sua saída $y(k)$.

O melhor resultado obtido pelo AG apresentou um MSE de $1,211 \times 10^{-3}$. Este é representado pelo seguinte cromossomo:

$$c = [-0,703 \quad 2,065 \quad 0,104 \quad -0,510],$$

sendo os dois primeiros genes relativos aos coeficientes das funções de Kautz, e os dois últimos, 0,104 e 0,510, correspondem ao polo que parametriza estas funções ($0,104 - 0,510i$).

A partir do polo das funções de Kautz, pode-se gerar a base de funções, $\mathfrak{B} = [K_1, K_2]$:

$$K_1(z) = \frac{0,950z}{z^2 - 0,208z + 0,271}, \quad (7.5)$$

$$K_2(z) = \frac{0,963z^2 - 0,157z}{z^2 - 0,208z + 0,271}. \quad (7.6)$$

Com isto, chega-se ao seguinte modelo:

$$\hat{y}_{2K}(k) = -0,703w_1(k) + 2,065w_2(k),$$

sendo $w_i(k)$ a entrada $u(k)$ filtrada pelo filtro de Kautz de i -ésima ordem, $K_i(z)$, para $i = 1$ e 2 .

A curva de convergência do AG é expressa na figura 7.4.

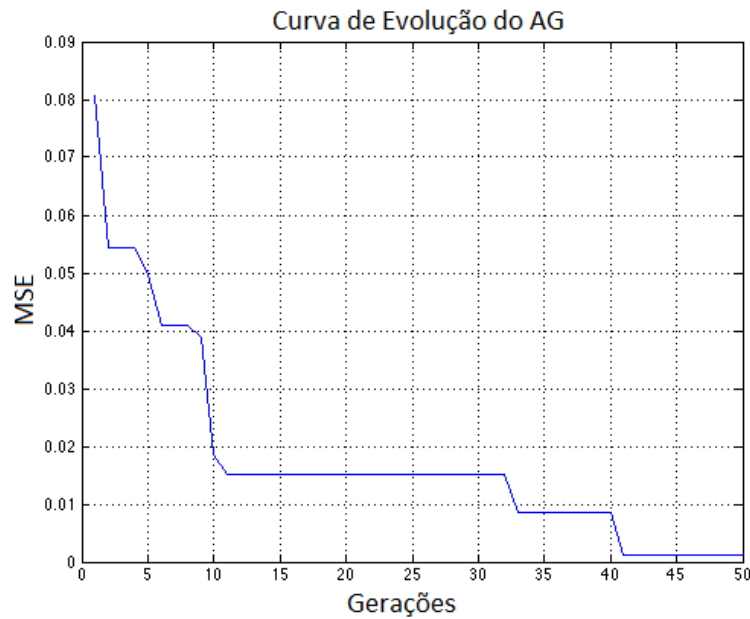


Figura 7.4: Convergência do Algoritmo Genético.

7.1.3 Comparações entre APA e AG nos Modelos de Kautz

Na comparação entre a busca de parâmetros do modelo de Kautz, a partir dos métodos heurísticos, AG e APA, pode-se concluir:

- Assim como nos modelos de Laguerre, seção 5.2.4.2, os resultados obtidos pelo APA são sensivelmente melhores que os obtidos pelo AG, no sentido de menor MSE, em um processo evolutivo de 50 iterações.
- Tanto no APA como no AG, quando aplicados à Identificação de Sistemas, o tempo mais relevante em cada iteração é o tempo gasto na avaliação de cada solução. Assim, quando estes algoritmos tratam problemas iguais, pode-se afirmar que o tempo gasto numa iteração do APA é aproximadamente igual ao de uma iteração do AG. Neste contexto, a partir dos gráficos mostrados nas figuras 7.2 e 7.4, pode-se afirmar que a velocidade média de convergência do APA é superior a do AG clássico.

7.2 MODELAGEM DE UM LEVITADOR MAGNÉTICO VERTICAL A PARTIR DE MODELOS VOLTERRA-OBF

Nesta seção será apresentado um sistema de levitação magnética e a sua identificação a partir de séries de Volterra-OBF.

7.2.1 Levitador Magnético Vertical

O sistema a ser identificado é composto por duas bobinas magnéticas e um disco magnético móvel, cuja posição é função das correntes que percorrem as bobinas. Tal sistema é mostrado de forma esquemática na figura 7.5.

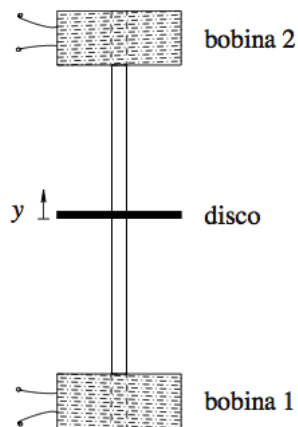


Figura 7.5: Levitador Magnético.

Uma foto deste sistema é mostrada ainda na figura 7.6.



Figura 7.6: Levitador Magnético, (ECP, 1999).

Neste sistema, o disco movimentar-se por uma haste de vidro que minimiza o atrito, permitindo grandes deslocamentos verticais. Tal sistema foi excitado a partir da forma de onda exposta na figura 7.7-a, enquanto que a saída obtida é mostrada na figura 7.7-b:

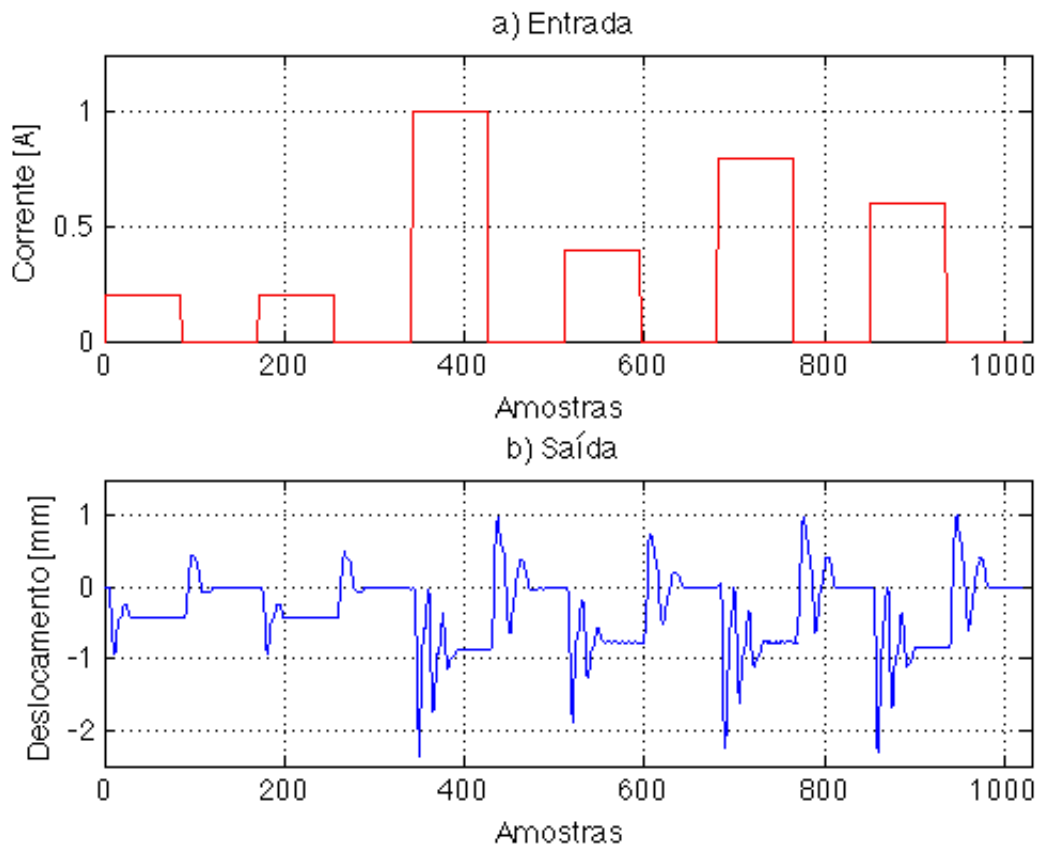


Figura 7.7: a - Sinal de Entrada e b - Sinal de Saída.

No processo de identificação do levitador, os modelos gerados foram obtidos a partir de vetores de entrada e saída de 1020 amostras do sistema. A validação foi realizada em vetores de dimensão 680.

7.2.1.1 Resultados obtidos a partir do modelo Volterra-OBF

Os modelos de Volterra-OBF obtidos para o sistema de levitação magnética, tem seus desempenhos expressos na tabela 7.1:

Tabela 7.1: O erro médio quadrático, MSE , em relação ao n^o de funções de Kautz que compõem o modelo.

NT	NF	polo	$MSE_{ident.}$	$MSE_{valid.}$
5	2	$0,7782 \pm 0,3295i$	0,0759	0,0886
14	4	$0,8381 \pm 0,2698i$	0,0212	0,0255
27	6	$0,7936 \pm 0,2827i$	0,0085	0,0105
	\vdots	\vdots	\vdots	\vdots
230	20	$0,7906 \pm 0,2800i$	$1,7761 \times 10^{-4}$	$6,1266 \times 10^{-4}$

Na tabela 7.1 o $MSE_{indent.}$ e o $MSE_{valid.}$ são os erros médios quadráticos na etapa de identificação e validação, respectivamente. Estes resultados foram obtidos após 200 gerações do AG, com auxílio dos operadores NM e MMQ.

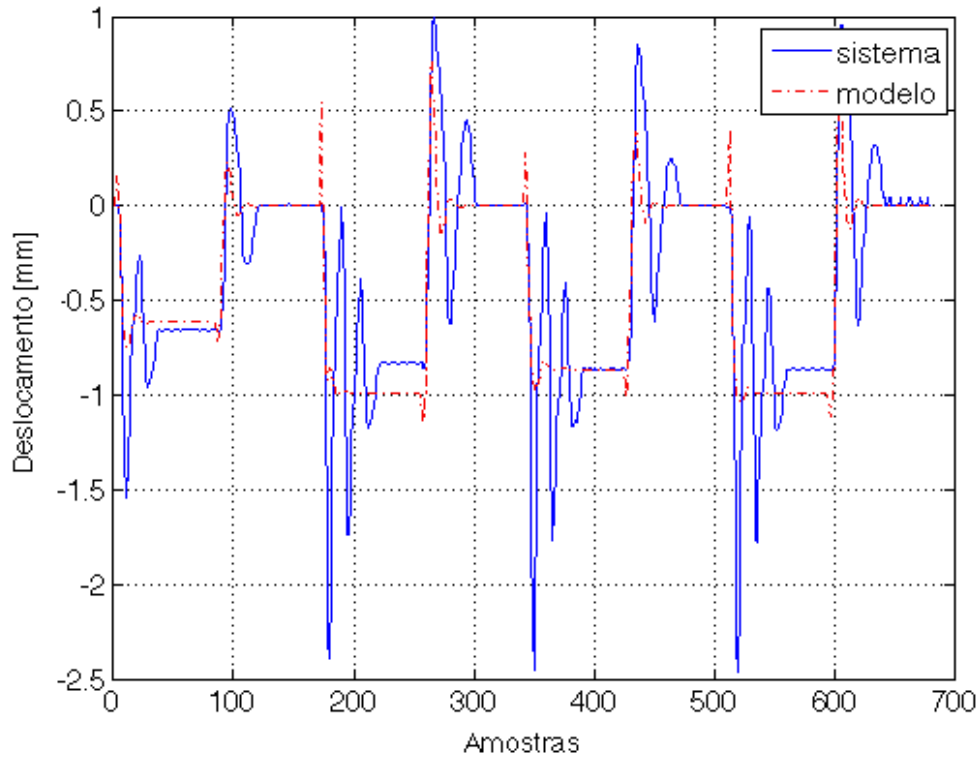


Figura 7.8: Modelo Volterra-OBF utilizando 2 funções de Kautz ($MSE = 0,0886$).

A simulação do modelo apontado na 1ª linha numérica da tabela 7.1, utilizando 2 funções de Kautz, é expresso na equação (7.7) e pode ser visualizado na figura 7.8.

$$\hat{y}_{2K}(k) = -6,8077w_1(k) + 3,5903w_2(k) + 2,9948 [w_1(k)]^2 + \\ -2,1374w_1(k)w_2(k) + 2,3279 [w_2(k)]^2, \quad (7.7)$$

sendo $w_i(k)$ a entrada $u(k)$ filtrada pelo filtro de Kautz de i -ésima ordem, $K_i(z)$, para $i = 1$ e 2 .

Já a simulação do modelo apontado na 3ª linha numérica da tabela 7.1, utilizando 6 funções de Kautz, pode ser visualizado na figura 7.9.

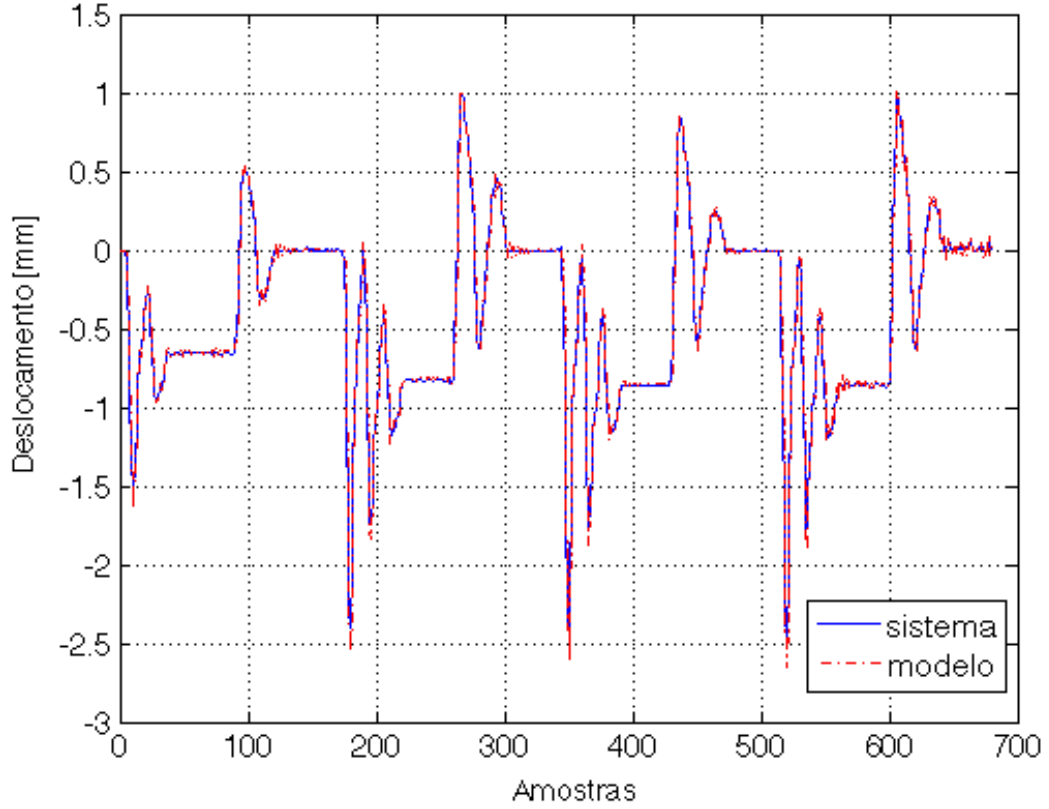


Figura 7.9: Modelo Volterra-OBF utilizando 6 funções de Kautz ($MSE = 0,0105$).

Tal modelo, composto por 6 funções de Kautz, é expresso pela equação (7.8).

$$\hat{y}_{6K}(k) = \sum_{i=1}^6 c_i w_i(k) + \sum_{p=1}^6 \sum_{q=p}^6 c_{p,q} w_p(k) w_q(k), \quad (7.8)$$

sendo $w_i(k)$ a entrada $u(k)$ filtrada pelo filtro de Kautz de i -ésima ordem, $K_i(z)$, para $i = 1, \dots, 6$, e os parâmetros do *kernel* de primeira ordem são expressos pelo seguinte vetor:

$$C_6 = \begin{bmatrix} -6,0409 & 2,7393 & -0,8848 & -4,9437 & 2,2708 & -2,5762 \end{bmatrix},$$

já os parâmetros do *kernel* de segunda ordem são mostrados na matriz que segue:

$$C_{66} = \begin{bmatrix} -1,5398 & -6,9036 & 8,3240 & 0,9579 & -5,2464 & 5,5990 \\ 0 & 1,3529 & 15,7666 & -1,2997 & -11,3885 & 1,8267 \\ 0 & 0 & 0,4880 & -2,5214 & -1,7015 & -5,5351 \\ 0 & 0 & 0 & 1,7455 & 7,1258 & -0,5095 \\ 0 & 0 & 0 & 0 & 1,1335 & -0,3490 \\ 0 & 0 & 0 & 0 & 0 & 2,1309 \end{bmatrix}.$$

A configuração do AG utilizado na busca dos parâmetros deste modelo é detalhada na próxima seção.

7.2.1.2 Parâmetros do Algoritmo Genético

Os parâmetros do AG responsável pela otimização dos polos e dos coeficientes das funções ortonormais são expressos na tabela 7.2.

Tabela 7.2: Parâmetros do AG

Parâmetro	Valor
População	200 Cromossomos
Met. de Seleção	Recombinação por Torneio
Taxa de Mutação	Variável (5 à 20 %)
Taxa de <i>Crossover</i>	Variável (80 à 20 %)

Neste algoritmo o *Fitness*, métrica utilizada para medição do desempenho de um cromossomo, é baseado no MSE, *Mean Square Error*. As taxas de mutação e *crossover* são variáveis em relação ao número de gerações realizadas. A taxa de *crossover* se inicia em 80% buscando enfoque em buscas locais, sendo reduzida ao longo das gerações. Já a taxa mutação incia-se em 20%, e é elevada com o passar das gerações, visando enfoque em buscas globais.

7.2.2 Comparação dos Modelos Volterra-OBF com os Modelos Polinomiais

Para efeitos de *Benchmark*, os dados do levitador magnético foram utilizados em uma identificação realizada a partir da *toolbox* Ident do Matlab, fazendo uso de um modelo linear, polinomial, de ordem 10, tanto na regressão dos sinais de entrada quanto dos sinais de saída. Este modelo é mostrado na equação (7.9).

$$A_1(z)Y(z) = B_1(z)U(z), \quad (7.9)$$

sendo $A_1(z)$:

$$A_1(z) = 1 - 2,16z^{-1} + 1,35z^{-2} + 0,282z^{-3} - 0,437z^{-4} - 0,194z^{-5} + \\ + 0,113z^{-6} + 0,158z^{-7} - 0,0301z^{-8} - 0,0107z^{-9} - 0,0378z^{-10},$$

e $B_1(z)$ pode ser expresso por:

$$B_1(z) = 0,0119 - 0,023z^{-1} + 0,0231z^{-2} - 0,0694z^{-3} - 0,106z^{-4} + \\ - 0,073z^{-5} + 0,0732z^{-6} + 0,122z^{-7} + 0,060z^{-8} - 0,0601z^{-9},$$

este modelo obteve um MSE igual a 0,1060.

Já para um modelo polinomial linear, de ordem 50 nas regressões dos sinais de entrada e também 50 nas regressões do sinal de saída, tem-se: (7.10):

$$A_2(z)Y(z) = B_2(z)U(z), \quad (7.10)$$

em que $A_2(z)$ é igual a:

$$\begin{aligned}
A_2(z) = & 1 - 2,12z^{-1} + 1,268z^{-2} + 0,347z^{-3} - 0,448z^{-4} - 0,163z^{-5} + \\
& -0,011z^{-6} + 0,264z^{-7} + 0,103z^{-8} - 0,323z^{-9} + 0,188z^{-10} + \\
& -0,051z^{-11} + 0,113z^{-12} - 0,172z^{-13} + 0,040z^{-14} - 0,064z^{-15} + \\
& +0,027z^{-16} + 0,144z^{-17} - 0,080z^{-18} - 0,082z^{-19} + 0,007z^{-20} + \\
& +0,096z^{-21} - 0,007z^{-22} - 0,012z^{-23} - 0,094z^{-24} + 0,080z^{-25} + \\
& -0,041z^{-26} - 0,024z^{-27} + 0,061z^{-28} - 0,049z^{-29} - 0,007z^{-30} + \\
& +0,128z^{-31} - 0,159z^{-32} + 0,041z^{-33} + 0,067z^{-34} - 0,055z^{-35} + \\
& +0,004z^{-36} - 0,012z^{-37} + 0,005z^{-38} + 0,023z^{-39} - 0,051z^{-40} + \\
& +0,053z^{-41} + 0,001z^{-42} - 0,048z^{-43} + 0,084z^{-44} - 0,109z^{-45} + \\
& +0,079z^{-46} + 0,006z^{-47} - 0,047z^{-48} + 0,012z^{-49} + 0,012z^{-50},
\end{aligned}$$

e $B_2(z)$ pode ser expresso por:

$$\begin{aligned}
B_2(z) = & 0,013 - 0,023z^{-1} + 0,022z^{-2} - 0,068z^{-3} - 0,109z^{-4} + \\
& -0,076z^{-5} + 0,071z^{-6} + 0,121z^{-7} + 0,064z^{-8} - 0,039z^{-9} + \\
& -0,062z^{-10} + 0,002z^{-11} + 0,093z^{-12} - 0,068z^{-13} + 0,025z^{-14} + \\
& -0,057z^{-15} - 0,030z^{-16} + 0,022z^{-17} - 0,012z^{-18} + 0,036z^{-19} + \\
& +0,080z^{-20} - 0,014z^{-21} + 0,049z^{-22} - 0,030z^{-23} - 0,019z^{-24} + \\
& -0,022z^{-25} + 0,019z^{-26} + 0,004z^{-27} - 0,057z^{-28} + 0,045z^{-29} + \\
& +0,011z^{-30} - 0,013z^{-31} + 0,062z^{-32} - 0,058z^{-33} + 0,041z^{-34} + \\
& -0,028z^{-35} - 0,007z^{-36} - 0,007z^{-37} - 0,041z^{-38} + 0,030z^{-39} + \\
& +0,044z^{-40} - 0,048z^{-41} + 0,018z^{-42} + 0,006z^{-43} + 0,041z^{-44} + \\
& -0,041z^{-45} - 0,026z^{-46} + 0,021z^{-47} + 0,006z^{-48} - 0,0082z^{-49},
\end{aligned}$$

Este modelo obteve um MSE de 0,0983 e pode-se observar seu resultado graficamente a partir da figura 7.10.

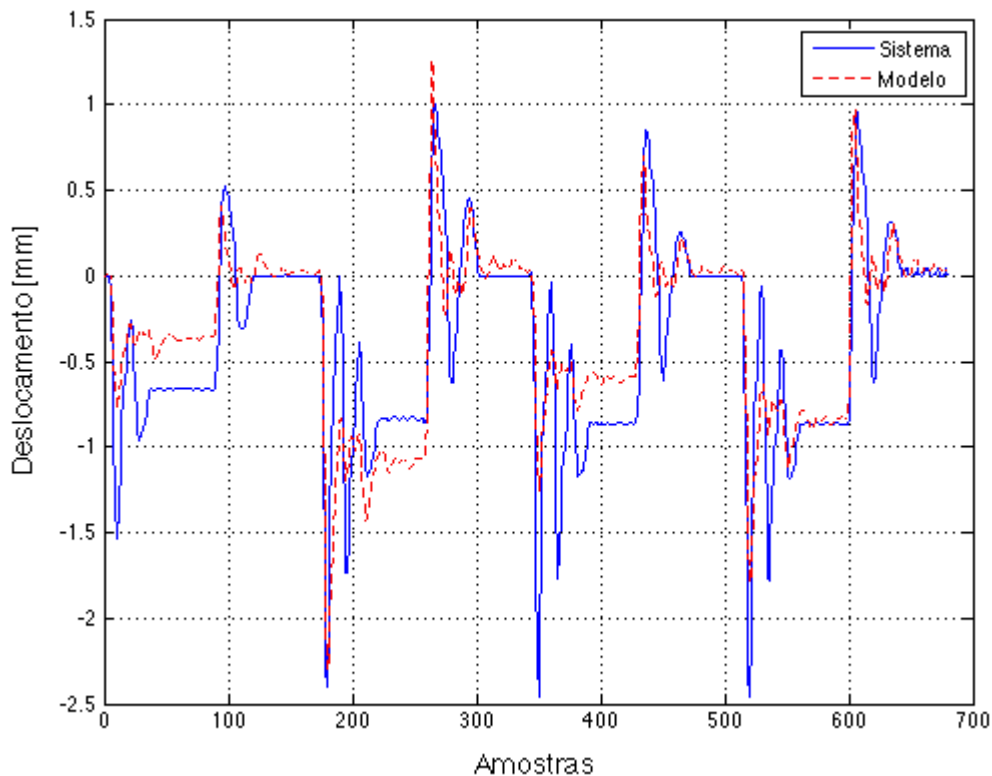


Figura 7.10: Modelo linear polinomial, de ordem 50 nas regressões tanto de entrada quanto de saída. Resultado obtido com a *toolbox* Ident do Matlab ($MSE = 0,0983$).

Comparando as figuras 7.9 e 7.10 pode-se perceber que o modelo de Volterra-OBF envolvendo 6 funções de Kautz é uma aproximação mais exata do que o modelo polinomial, com ordem 50, tanto para a entrada quanto para a saída.

O modelo polinomial linear apresentou erros consideráveis, tanto nas respostas de regime transiente quanto nas de regime permanente, tornando inviável sua aplicação como modelo do levitador magnético.

Os resultados obtidos a partir do modelo linear, (7.10), servem para mostrar que modelos lineares, mesmo com ordens extremamente elevadas, não servem como aproximação para o levitador magnético, confirmando o fato de que trata-se de um sistema não linear.

Partindo-se para modelos não lineares foi utilizado o *Benchmark* dos modelos ARX não lineares, também da *toolbox* Ident do Matlab. Trata-se de um modelo de ordem 10, tanto na regressão de entrada como na saída, com não linearidade do tipo sigmóide.

Os resultados obtidos com o modelo ARX não linear são apresentados na figura 7.11. Com este modelo obteve-se um MSE de 0,1056 na identificação do levitador magnético. Este modelo mostrou-se uma aproximação pouco exata da resposta do sistema em regime transiente e apresentou consideráveis erros na resposta em regime permanente.

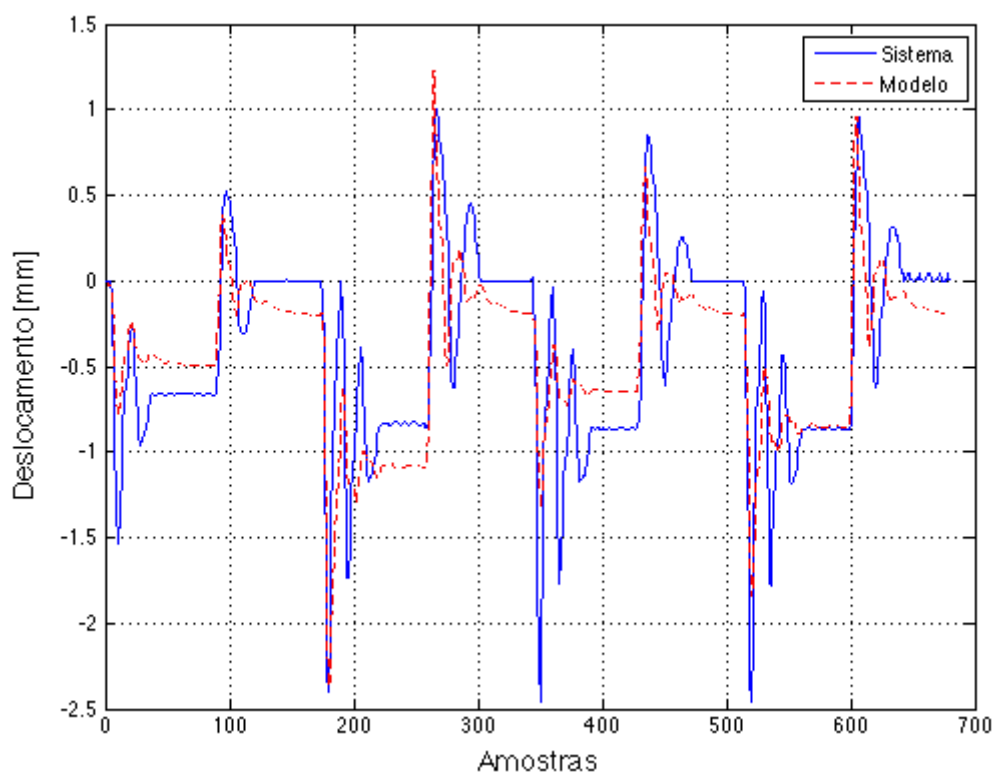


Figura 7.11: Modelo ARX Não Linear, com ordem igual a 10 tanto nas regressões de entrada quanto nas de saída. Resultado obtido com a *toolbox* Ident do Matlab ($MSE = 0,1056$).

Comparando as figuras 7.9 e 7.11 pode-se perceber que o modelos de Volterra-OBF são modelos mais representativos da dinâmica do sistema a ser identificado. E que o modelo ARX não linear não é capaz de representar a dinâmica do levitador magnético.

7.3 MODELAGEM DE UM LEVITADOR MAGNÉTICO VERTICAL A PARTIR DE MODELOS NARX-OBF

Nesta seção serão apresentados os resultados da identificação do levitador magnético, mostrado na figura 7.5, a partir de modelos NARX-OBF. Os resultados obtidos podem ser visualizados na tabela 7.3.

Tabela 7.3: O erro médio quadrático, MSE , em relação ao n° de funções de Kautz que compõem o modelo.

NT	NF	polo	$MSE_{ident.}$	$MSE_{valid.}$
14	2	$-0,1242 \pm 0,7427i$	$6,9571 \times 10^{-3}$	0,0085
44	4	$0,3749 \pm 0,5336i$	$1,7192 \times 10^{-3}$	0,0019
90	6	$0,4577 \pm 0,4036i$	$8,0988 \times 10^{-4}$	$8,2609 \times 10^{-4}$

Na tabela 7.3, NT é o número de termos que compõem o modelo, NF é o número de funções que formam a base ortonormal, e $MSE_{ident.}$ e $MSE_{valid.}$ são os erros médios quadráticos

de identificação e validação, respectivamente.

Estes resultados foram obtidos a partir da excitação do sistema pela forma de onda exposta na figura 7.7a. O resultado gráfico da identificação tratada nesta seção, expressa pelo modelo NARX de segunda ordem, (3.44), fazendo uso de 2 funções ortonormais de Kautz, é mostrado na figura 7.12. Já o modelo (3.44), fazendo uso de 4 funções ortonormais de Kautz é mostrado na figura 7.13.

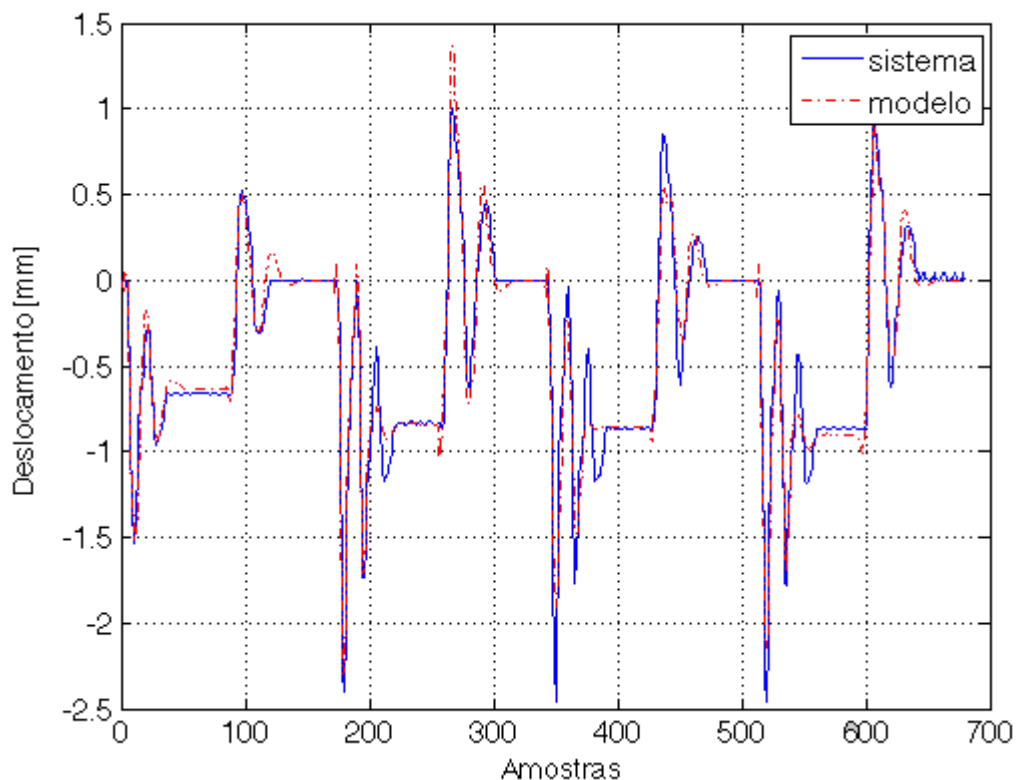


Figura 7.12: Resposta do modelo NARX, com *kernels* expressos a partir de 2 funções de Kautz em linhas tracejadas e a resposta do sistema real, em linhas contínuas ($MSE = 0,0085$).

Estes resultados foram obtidos pelo método de Levenberg-Marquardt na determinação dos elementos que compõem os *kernels* do modelo NARX-OBF. Já os polos das funções ortonormais foram buscados com auxílio de um Algoritmo Genético com os mesmos parâmetros expressos na tabela 7.2.

Para exemplificar a estrutura dos *kernels* NARX-OBF, o modelo que possui seus resultados de identificação expressos na segunda linha da tabela 7.3, terá seus *kernels* expressos pela equação (7.11).

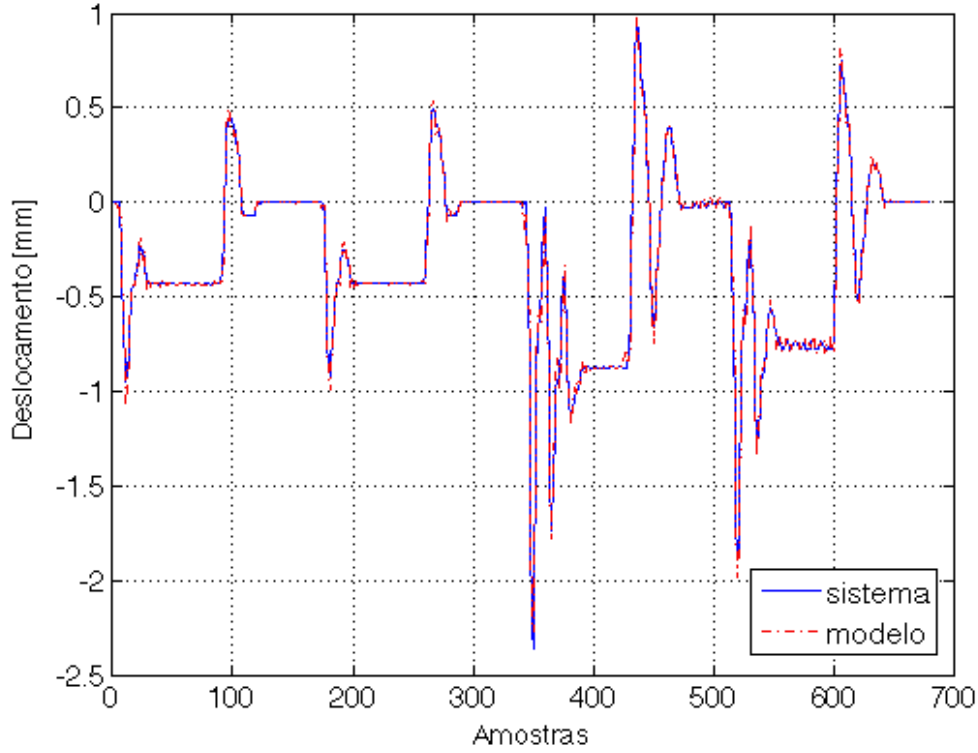


Figura 7.13: Resposta do modelo NARX, com *kernels* expressos a partir de 4 funções de Kautz em linhas tracejadas e a resposta do sistema real, em linhas contínuas ($MSE = 0,0019$).

$$\begin{aligned}
 \hat{y}_{4K}(k) = & \sum_{i=1}^4 C_i^u w_i^u(k) + \sum_{j=1}^4 C_j^y w_j^y(k) + \\
 & \sum_{p=1}^4 \sum_{q=p}^4 C_{p,q}^u w_p^u(k) w_q^u(k) + \sum_{r=1}^4 \sum_{s=r}^4 C_{r,s}^y w_r^y(k) w_s^y(k) + \\
 & \sum_{v=1}^4 \sum_{w=1}^4 C_{v,w}^{uy} w_v^u(k) w_w^y(k), \tag{7.11}
 \end{aligned}$$

sendo $w_i^u(k)$ a entrada $u(k)$ filtrada pelo filtro de Kautz de i -ésima ordem, $K_i(z)$, para $i = 1, 2, \dots, 4$. Já $w_i^y(k)$ é a saída $\hat{y}(k-1)$ filtrada pelo filtro de Kautz de i -ésima ordem, $K_i(z)$, para $i = 1, 2, \dots, 4$.

Os parâmetros do *kernel* de primeira ordem, associados aos termos de entrada, são mostrados no seguinte vetor:

$$C^u = \begin{bmatrix} -0,0797 & -0,5700 & 1,2751 & -2,0884 \end{bmatrix},$$

já os parâmetros do *kernel*, também de primeira ordem, associados aos termos de saída, são

mostrados no vetor que segue:

$$C^y = \begin{bmatrix} 5,5325 & -3,7228 & 3,6844 & 0,5203 \end{bmatrix}.$$

Os parâmetros do *kernel* de segunda ordem, associados aos termos de saída, são expostos na seguinte matriz:

$$C^{yy} = \begin{bmatrix} 0,1443 & 0,1277 & -0,3493 & 0,2973 \\ 0 & 0,5153 & -1,6918 & 0,6472 \\ 0 & 0 & 1,6086 & -0,9046 \\ 0 & 0 & 0 & -0,1022 \end{bmatrix}, \quad (7.12)$$

já os parâmetros do *kernel* de segunda ordem, associados aos termos de entrada, são mostrados na matriz que segue:

$$C^{uu} = \begin{bmatrix} 0,8457 & 1,9896 & -2,6491 & -2,0168 \\ 0 & -11,6411 & 27,5607 & 9,0143 \\ 0 & 0 & -18,3848 & -7,9737 \\ 0 & 0 & 0 & -1,0980 \end{bmatrix},$$

os parâmetros do *kernel* de segunda ordem, associados aos termos cruzados, são dados por:

$$C^{uy} = \begin{bmatrix} 11,4558 & 1,8338 & 11,1570 & -17,2212 \\ 5,2553 & -15,6086 & 3,4056 & 14,0388 \\ -9,5983 & -5,4970 & -16,3187 & 19,2909 \\ -5,7227 & 14,0593 & 4,1261 & -15,1342 \end{bmatrix}.$$

A implementação dos modelos NARX-OBF utilizados nesta tese foi realizada a partir do software Matlab, mais especificamente a partir do Simulink. Um exemplo de modelo NARX-OBF, com 4 funções de Kautz pode ser visualizada na figura 7.14, sendo seus blocos internos mostrados nas figuras 7.15, 7.16, 7.17, 7.18 e 7.19. Na figura 7.15 mostram-se as contribuições das saídas filtradas na sua forma linear, já na figura 7.16 se expressa a não linearidade associada as saídas filtradas. Na figura 7.17 mostram-se as contribuições não lineares dos termos cruzados (entradas e saídas filtradas), na figura 7.18 expressa-se a contribuição das entradas filtradas associadas de forma linear e na figura 7.19 as entradas filtradas são associadas de forma não linear.

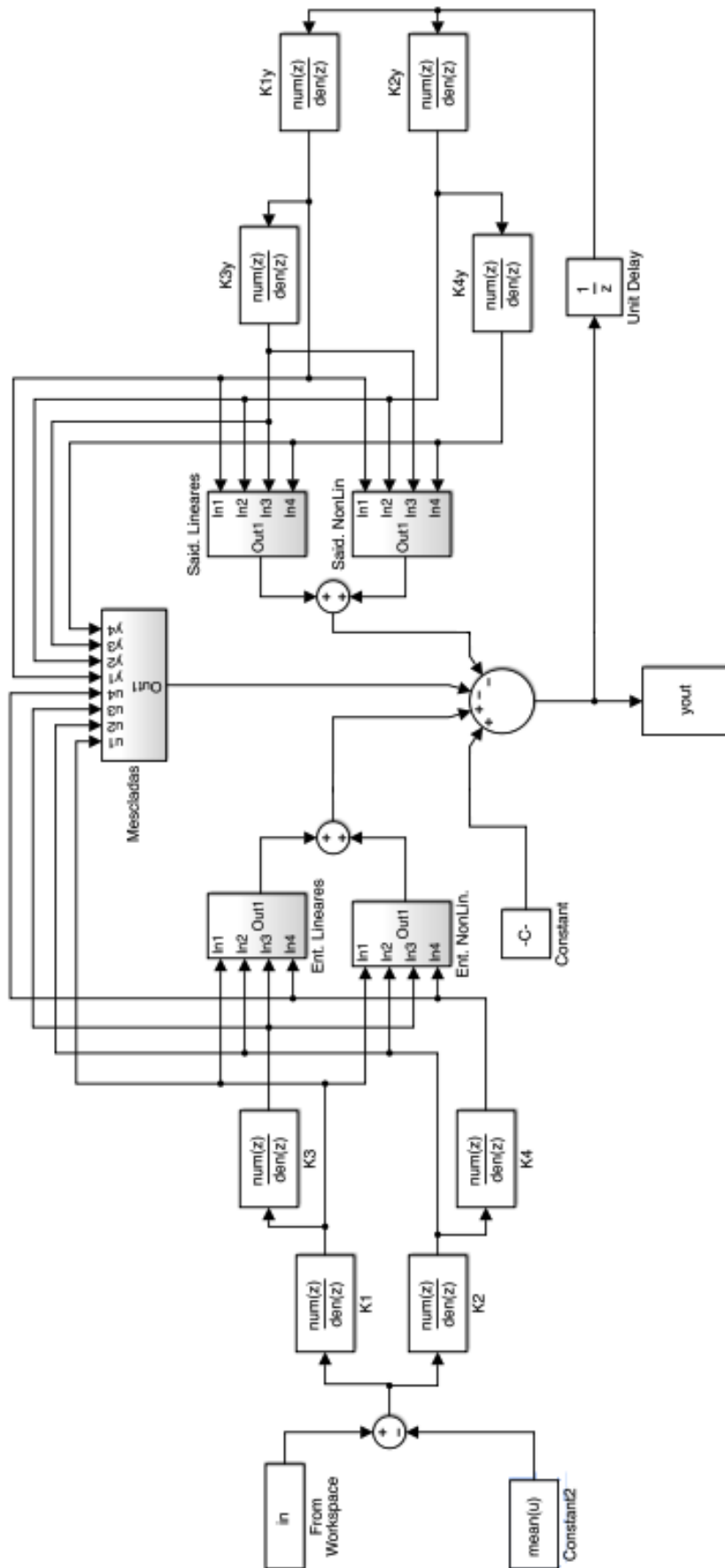


Figura 7.14: Implementação no Simulink de um modelo NARX-OBF, com 4 funções de Kautz.

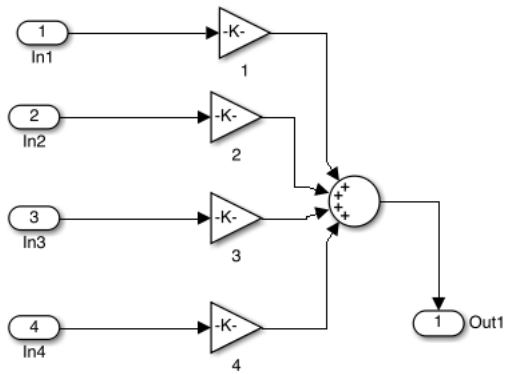


Figura 7.15: Parte linear, relativa as saídas filtradas, do modelo NARX-OBF, com 4 funções de Kautz (bloco Said. Lineares na figura 7.14).

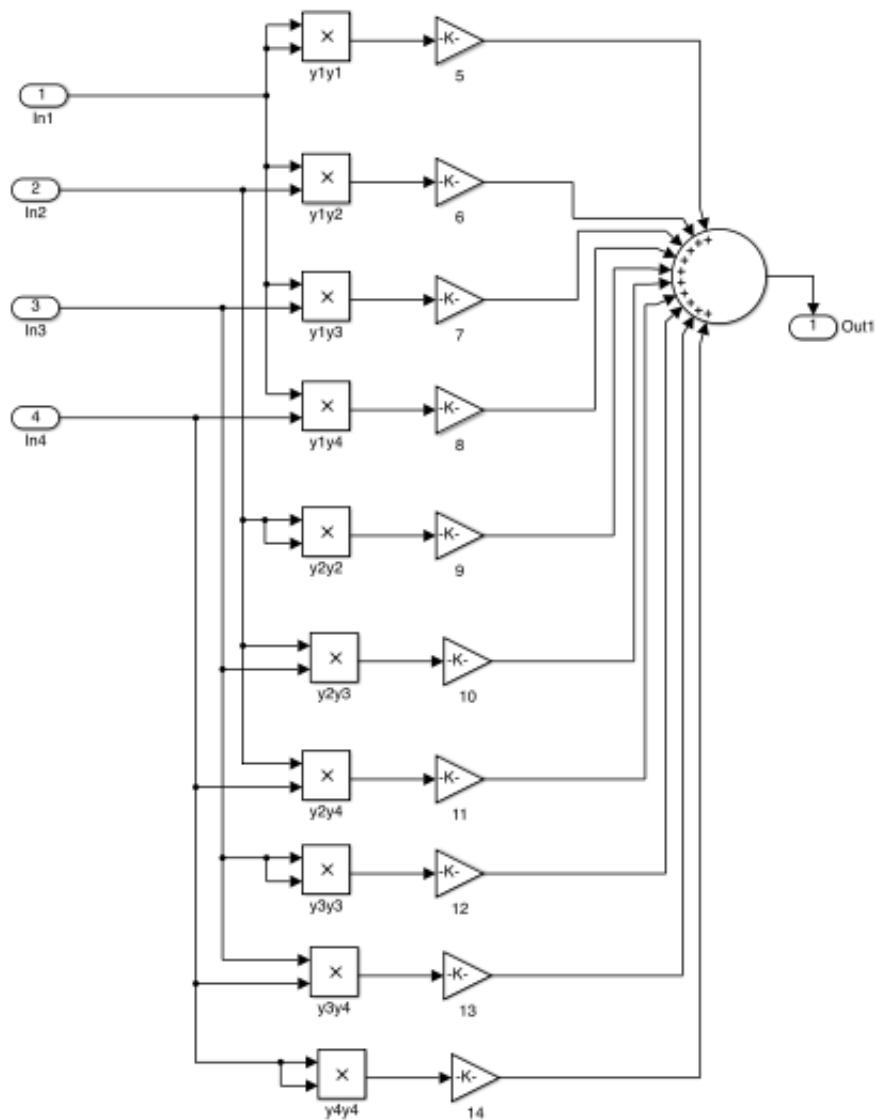


Figura 7.16: Parte não linear, relativa as saídas filtradas, do modelo NARX-OBF, com 4 funções de Kautz (bloco Said. Nonlin. na figura 7.14).

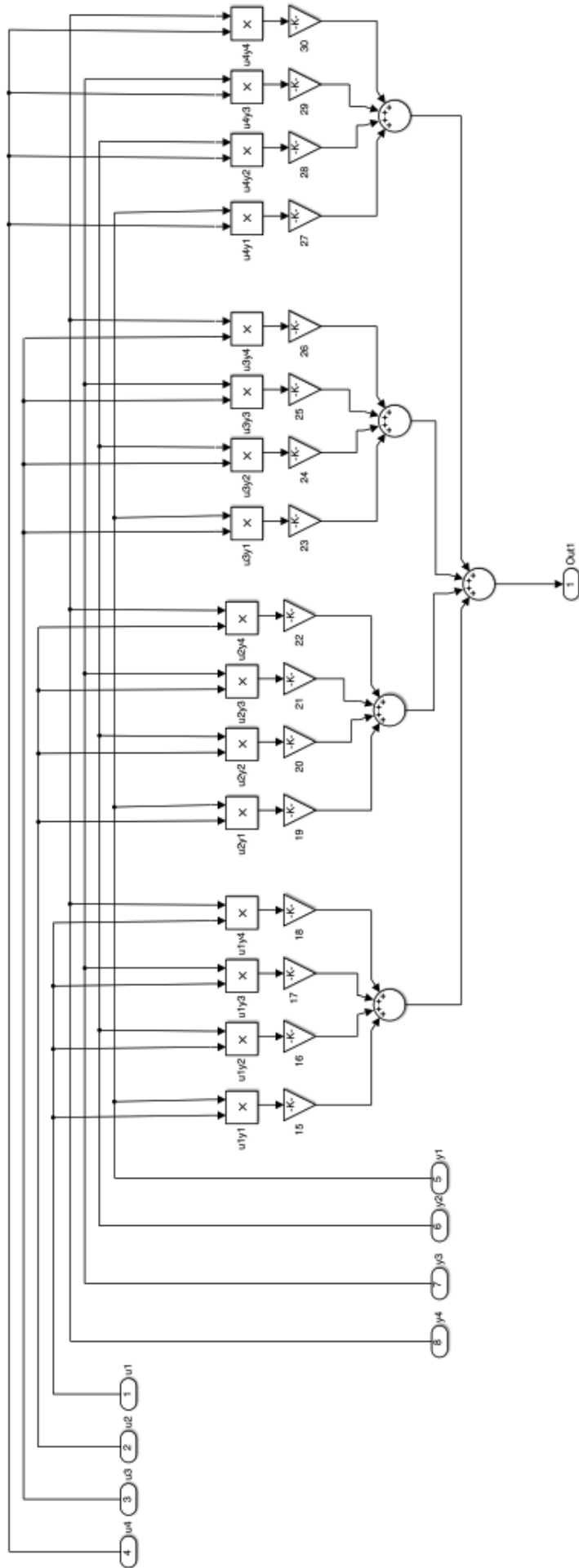


Figura 7.17: Parte não linear, relativa as entradas e saídas filtradas (termos cruzados), do modelo NARX-OBF, com 4 funções de Kautz (bloco Mescladas na figura 7.14).

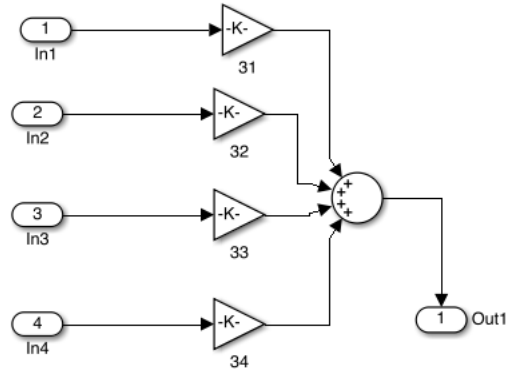


Figura 7.18: Parte linear, relativa as entradas filtradas, do modelo NARX-OBF, com 4 funções de Kautz (bloco Ent. Lineares na figura 7.14).

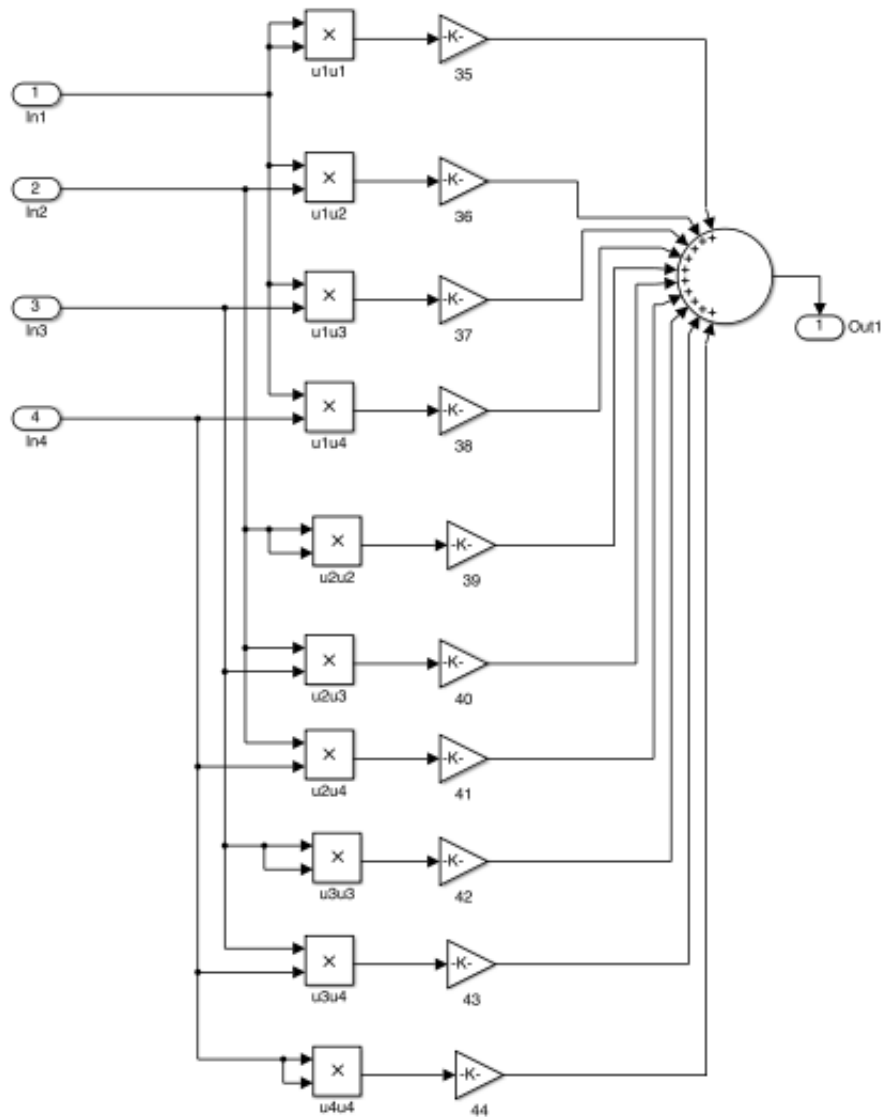


Figura 7.19: Parte não linear, relativa as entradas filtradas, do modelo NARX-OBF, com 4 funções de Kautz (bloco Ent. Nonlin. na figura 7.14)

7.3.1 Comparação dos Modelos NARX-OBF com os Modelos Polinomiais

Na comparação com os métodos clássicos de identificação, modelos polinomiais lineares e ARX não lineares, expressos nas figuras 7.10 e 7.11, pode-se afirmar que o modelo NARX-OBF apresenta resultados mais exatos que seus concorrentes.

7.3.2 Comparação Entre os Modelos NARX-OBF e Volterra-OBF

Na comparação entre as linhas 1 da tabela 7.3 e 2 da tabela 7.1, que expressam ambos modelos com 14 termos, pode-se observar que o modelo NARX-OBF apresenta menor MSE, para o exemplo tratado.

Em síntese, pode-se afirmar que entre os modelos analisados, o modelo NARX-OBF apresentou os melhores índices de MSE, mostrando-se o modelo mais preciso para representar a dinâmica do levitador magnético.

7.4 IDENTIFICAÇÃO DE UM LEVITADOR MAGNÉTICO HORIZONTAL A PARTIR DE MODELOS NARX-OBF SIMPLIFICADOS

Para validação da metodologia exposta na seção 6.5, foi utilizado um outro sistema de levitação magnética, composto por 2 ímãs permanentes e um disco magnético móvel. Para atuar sobre o disco móvel o sistema conta com quatro bobinas magnéticas, sendo utilizadas apenas duas delas neste experimento. Assim, apenas o movimento do disco no eixo x é considerado, sendo o movimento nos eixos y e z desprezíveis, (COELHO, 2013). O esquemático do sistema é mostrado na figura 7.20, já a sua implementação física pode ser observada na figura 7.21. Uma descrição mais detalhada sobre este sistema pode ser encontrada em (COELHO, 2012).

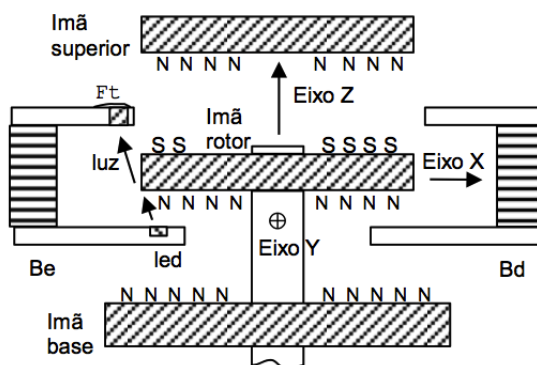


Figura 7.20: Esquemático do levitador magnético, (COELHO, 2012).

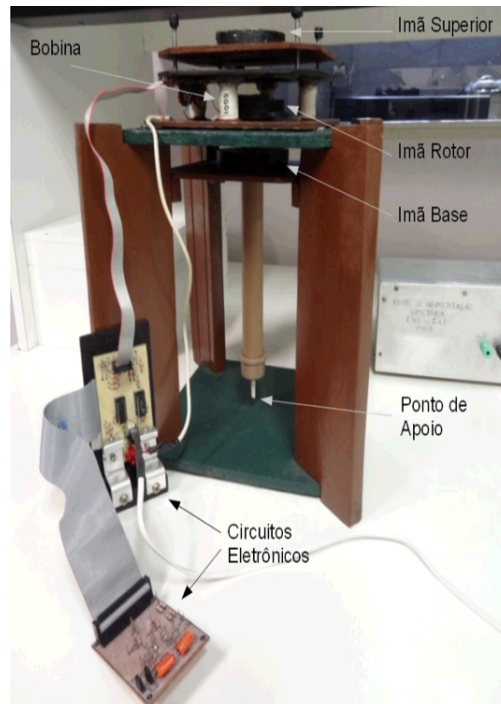


Figura 7.21: Foto do sistema de levitação magnética horizontal, (COELHO, 2012).

Também cabe salientar que tal sistema é não linear, como pode ser observado na figura 7.22. Tendo o sistema sido excitado por uma entrada senoidal, figura 7.22a, a sua saída apresentou-se como uma senoide distorcida, como mostrado na figura 7.22b, o que demonstra a característica não linear do sistema, (AGUIRRE, 2007; ISERMANN; MUNCHHOF, 2011).

É importante mencionar que tal sistema é instável, segundo o teorema de Earnshaw, (EARNSHAW, 1842). Assim, para contornar o problema da instabilidade, o sistema opera em malha fechada, sob ação de um controlador PD. Para excitar a dinâmica deste sistema, foi utilizado o sinal de entrada do tipo PRMLS (*Pseudo Random Multi Level Signal*), expresso pela figura 7.23-a, tendo sido obtido o sinal de saída mostrado na figura 7.23-b, para a identificação do mesmo. Para a identificação foram utilizados sinais de entrada e saída compostas por 100000 amostras, sendo o período de amostragem de $1ms$. Para a etapa de validação foram utilizados vetores com 15000 amostras.

7.4.1 Resultados Obtidos com Modelos NARX-OBF Simplificados

No Algoritmo Genético utilizado na busca pelos parâmetros do modelo NARX-OBF simplificado, foram utilizadas populações de 200 cromossomos, sendo o torneio o método de seleção empregado. As taxas de *Crossover* e mutação, variaram de inicialmente de 80 a 20% e de 5 a 20%, respectivamente.

O erro médio quadrático, *MSE*, obtido na identificação, em malha fechada, do levitador magnético horizontal, para modelos NARX-OBF com diferentes números de funções

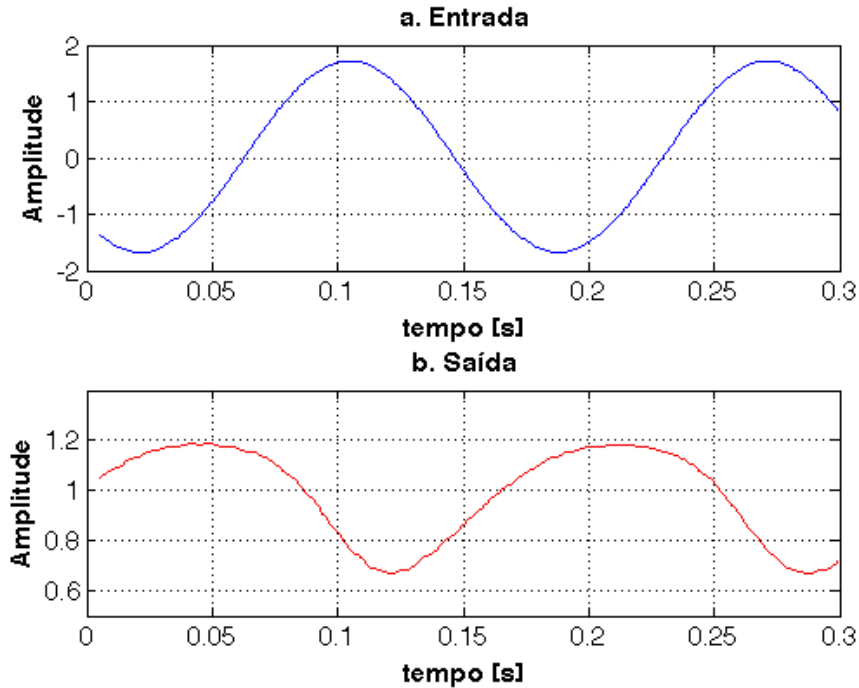


Figura 7.22: Resposta do levitador magnético horizontal a um sinal senoidal.

ortonormais, pode ser observado na tabela 7.4:

Tabela 7.4: MSE para modelos NARX-OBF completos.

NF	NT	Polo	$MSE_{ident.}$	$MSE_{valid.}$
2	14	$-0,2979 \pm 0,0012i$	$1,2416 \times 10^{-3}$	$1,2957 \times 10^{-3}$
4	44	$0,0521 \pm 0,0012i$	$3,5212 \times 10^{-4}$	$3,589 \times 10^{-4}$

em que NF representa o número de funções ortonormais utilizadas e NT corresponde ao número de termos (coeficientes) que compõem cada modelo NARX-OBF.

Já os MSE s obtidos a partir de modelos NARX-OBF, simplificados por AG, podem ser observados na tabela 7.5:

Tabela 7.5: MSE para modelos NARX-OBF com identificação a partir do AG.

NF	NT	Polo	$MSE_{ident.}$	$MSE_{valid.}$	$fitness$
2	9	$-0,3833 \pm 0,7870$	$1,6059 \times 10^{-3}$	$1,6245 \times 10^{-3}$	37,2677
4	11	$0,4964 \pm 0,2187i$	$3,5817 \times 10^{-4}$	$3,6197 \times 10^{-4}$	45,7718

na tabela 7.5, pode-se ainda observar o $fitness$, expresso pela equação (6.5), utilizado como função de avaliação na redução do número de termos dos modelos NARX-OBF.

Na tabela 7.6 podem ser vistos os MSE s para os modelos NARX-OBF quando a simplificação dos mesmos é realizada a partir do APA.

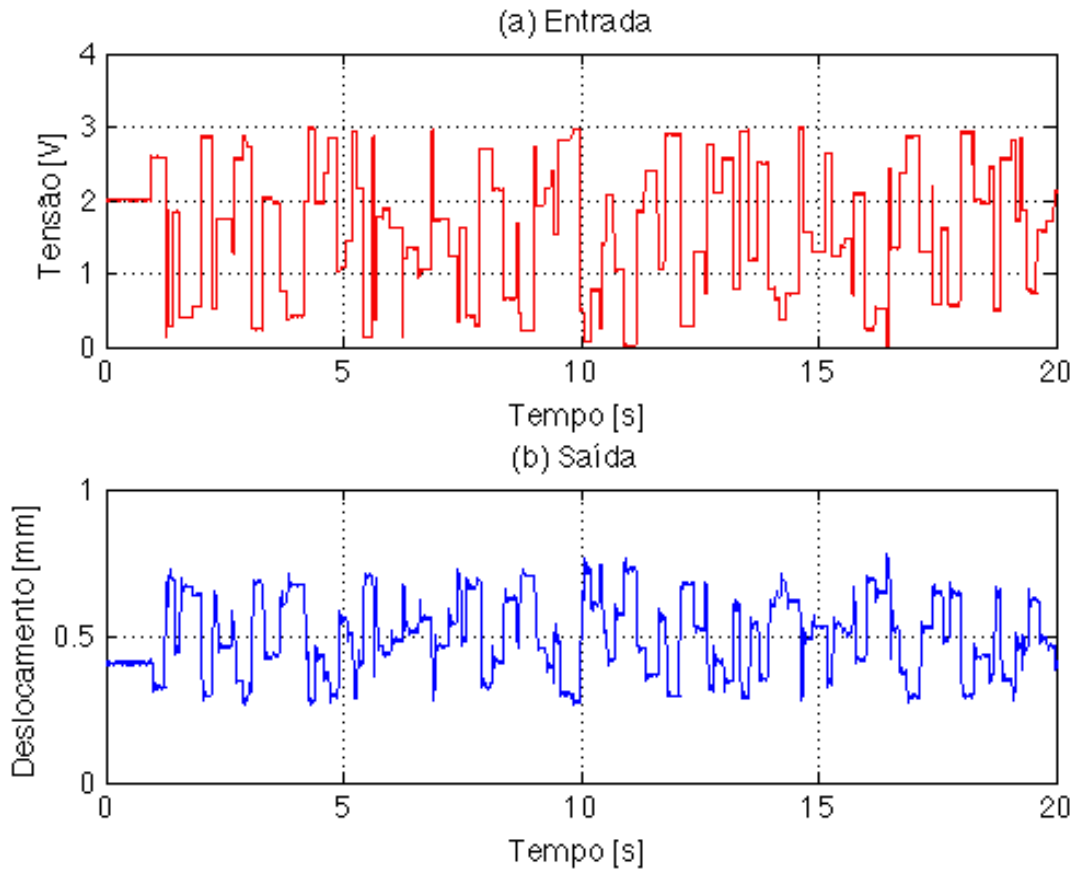


Figura 7.23: Amostras dos sinais de entrada e saída utilizados na identificação do levitador magnético horizontal.

Tabela 7.6: MSE para modelos NARX-OBF com identificação a partir do APA.

NF	NT	Polo	$MSE_{ident.}$	$MSE_{valid.}$	$fitness$
2	9	$-0,3837 \pm 0,6712$	$1,6202 \times 10^{-3}$	$1,6286 \times 10^{-3}$	37,2601
4	10	$0,5943 \pm 0,2197i$	$3,5229 \times 10^{-4}$	$3,6031 \times 10^{-4}$	43,7856

Um exemplo dos resultados obtidos para modelos NARX-OBF simplificados, com 4 funções ortonormais, 11 termos, pode ser visualizado na figura 7.24. A equação que expressa tal modelo, NARX-OBF simplificado, é expressa em (7.13).

$$\begin{aligned}
 \hat{y}_{4K} = & -6,4514 \times 10^{-3} w_2^y - 3,1029 \times 10^{-2} w_4^y - 6,6609 \times 10^{-4} w_1^u w_1^y + \\
 & + 1,3954 \times 10^{-3} w_1^u w_2^y + 5,7851 \times 10^{-4} w_1^u w_4^y - 1,5393 \times 10^{-3} w_2^u w_1^y + \\
 & + 7,0438 \times 10^{-4} w_2^u w_4^y - 7,7240 \times 10^{-4} w_3^u w_1^y + 8,0864 \times 10^{-4} w_3^u w_2^y + \\
 & + 2,6677 \times 10^{-4} w_3^u w_4^y + 0,5067, \quad (7.13)
 \end{aligned}$$

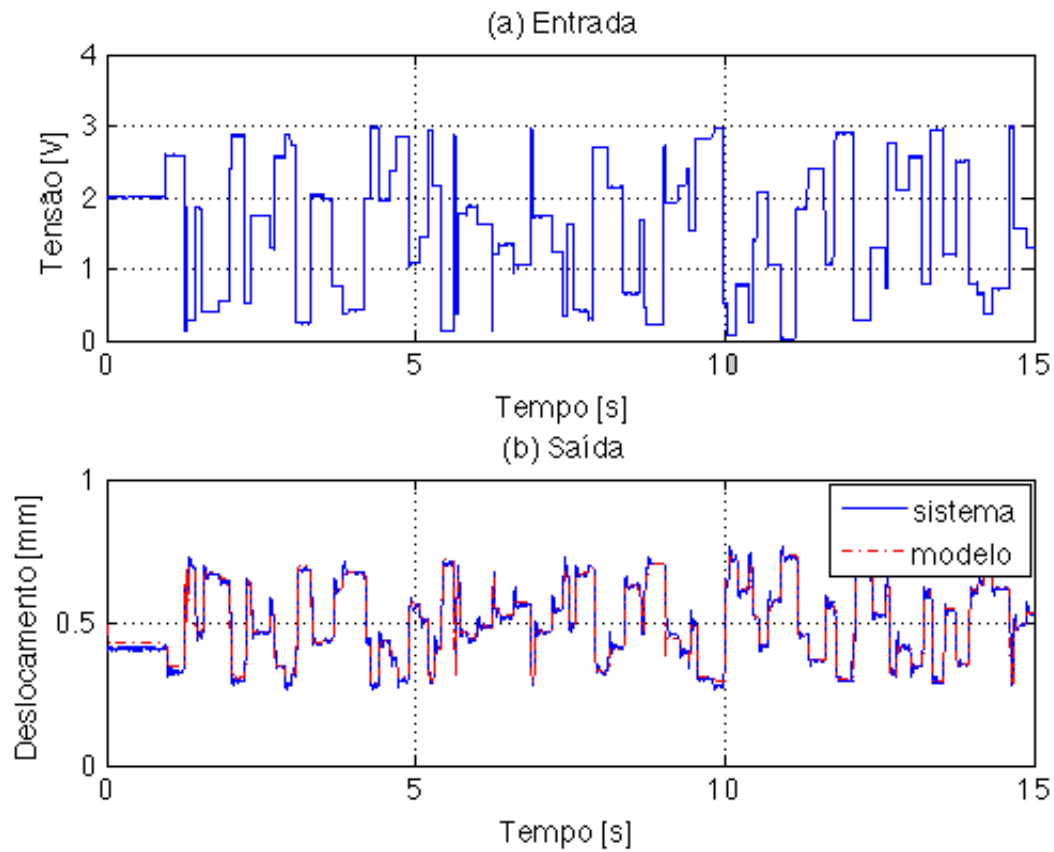


Figura 7.24: Validação da identificação do levitador magnético horizontal, a partir de um modelo NARX-OBF com 4 funções de Kautz na sua versão simplificada, $MSE = 3,6031 \times 10^{-4}$, com 11 termos (o modelo completo possui 44 termos).

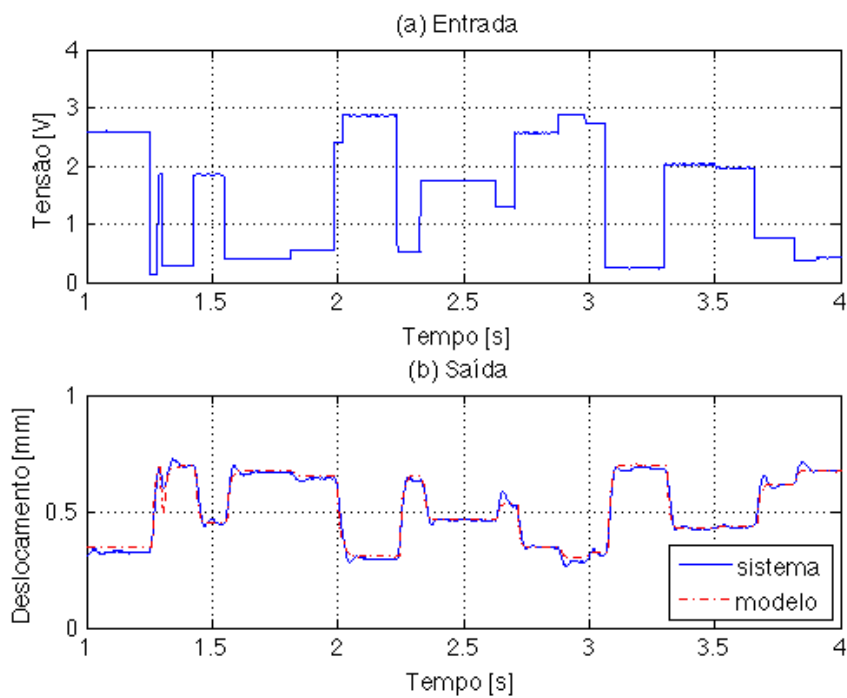


Figura 7.25: Detalhe da figura 7.24 entre os instantes 1 e 4s, ($MSE = 3,6031 \times 10^{-4}$).

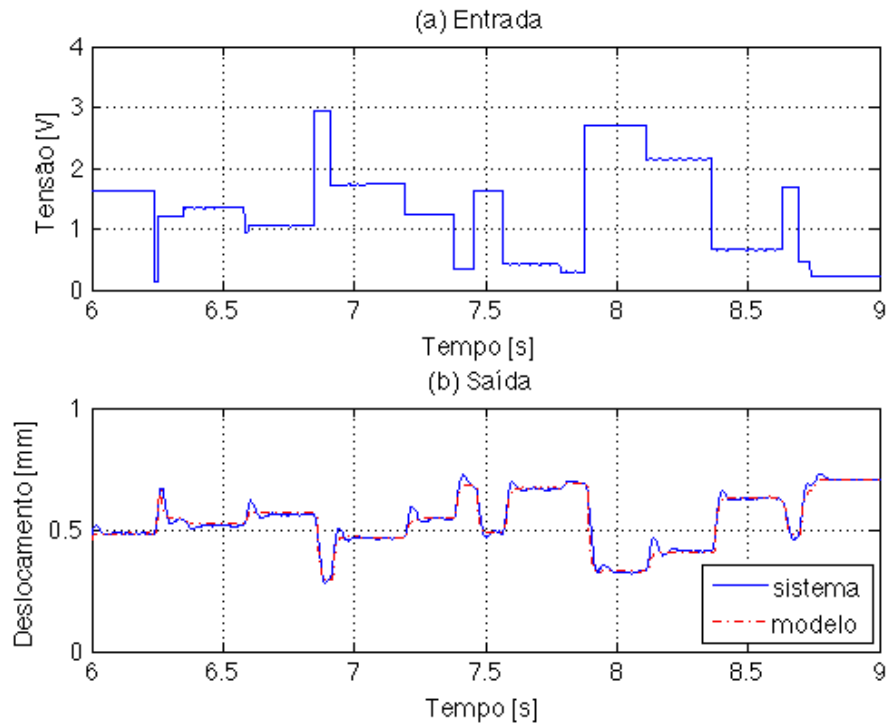


Figura 7.26: Detalhe da figura 7.24 entre os instantes 6 e 9s (identificação do levitador magnético a partir de Modelos NARX-OBF simplificados, $MSE = 3,6031 \times 10^{-4}$).

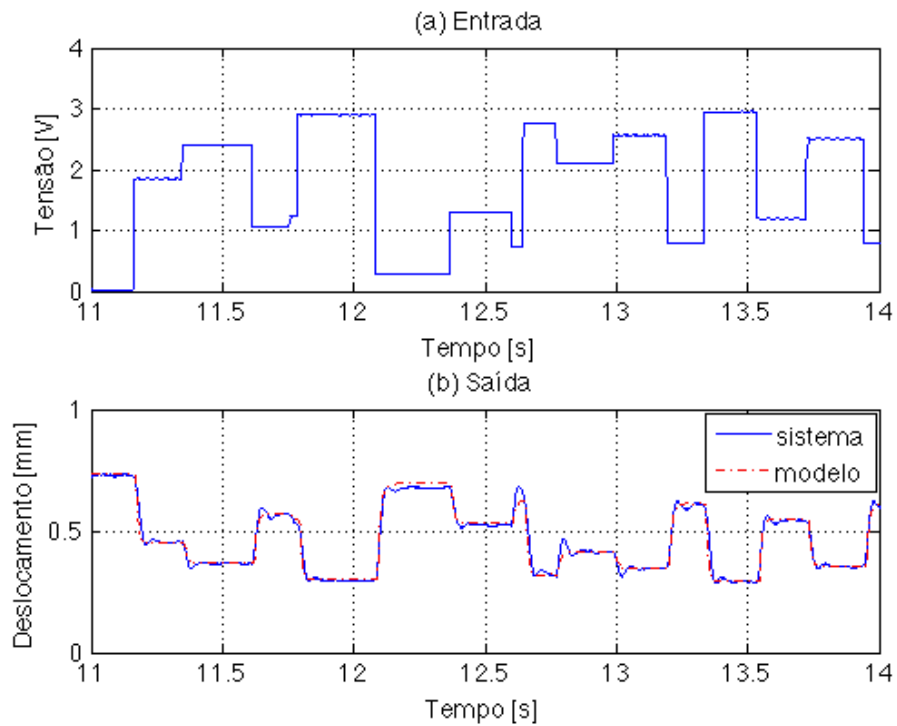


Figura 7.27: Detalhe da figura 7.24 entre os instantes 11 e 14s (identificação do levitador magnético a partir de Modelos NARX-OBF simplificados, $MSE = 3,6031 \times 10^{-4}$).

Para efeitos de comparação, os dados de entrada e saída do levitador magnético hori-

zontal foram levados à *toolbox* ident, do Matlab. Os resultados obtidos são mostrados nas figuras 7.28 e 7.30.

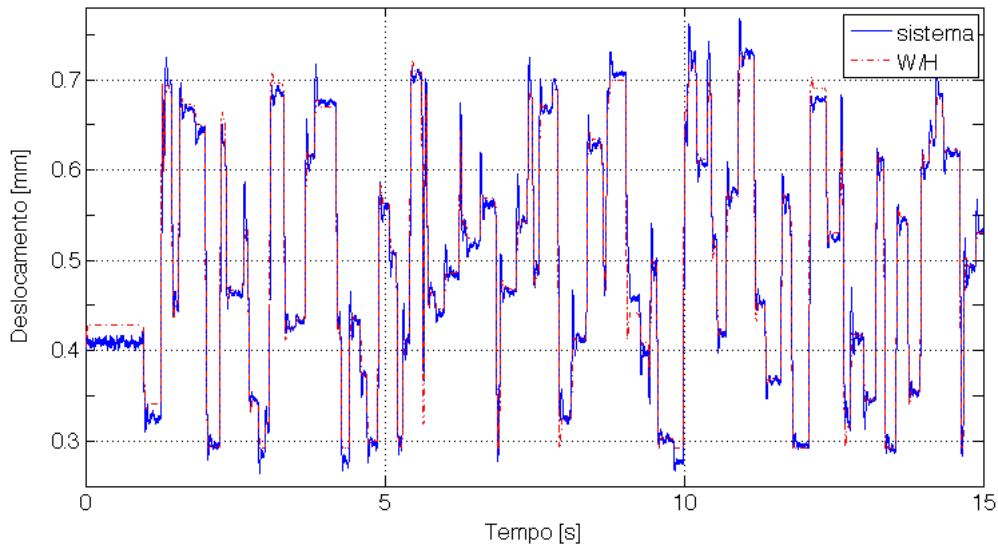


Figura 7.28: Resultado da identificação do levitador magnético vertical, a partir da *toolbox* do Matlab, utilizando modelos não lineares Wiener/Hammerstein, ($MSE = 3,9075 \times 10^{-4}$).

O modelo mostrado na figura 7.28 obteve um MSE de $3,9075 \times 10^{-4}$ e um detalhe desta figura, entre os instantes de 1 a 4s, pode ser observado na figura 7.29.

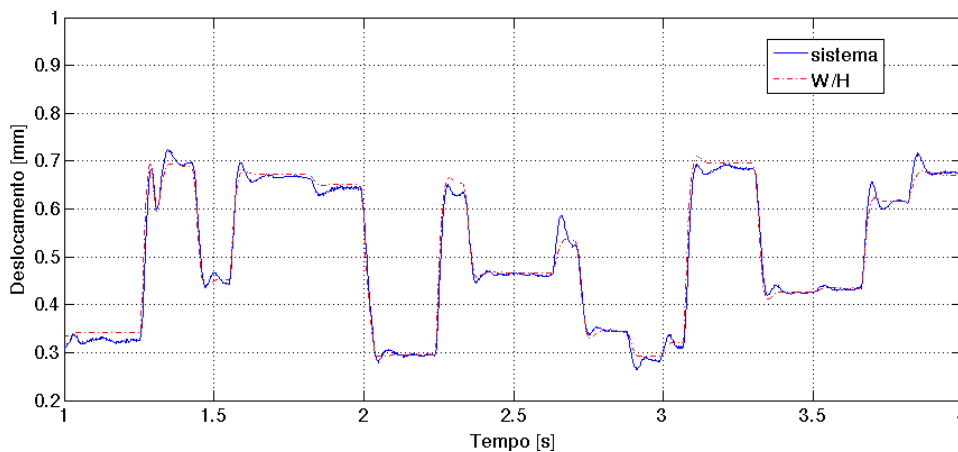


Figura 7.29: Detalhe da figura 7.28 entre os instantes 1 e 4s. Identificação do levitador magnético a partir da *toolbox* ident do Matlab, com modelos Wiener/Hammerstein, ($MSE = 3,9075 \times 10^{-4}$).

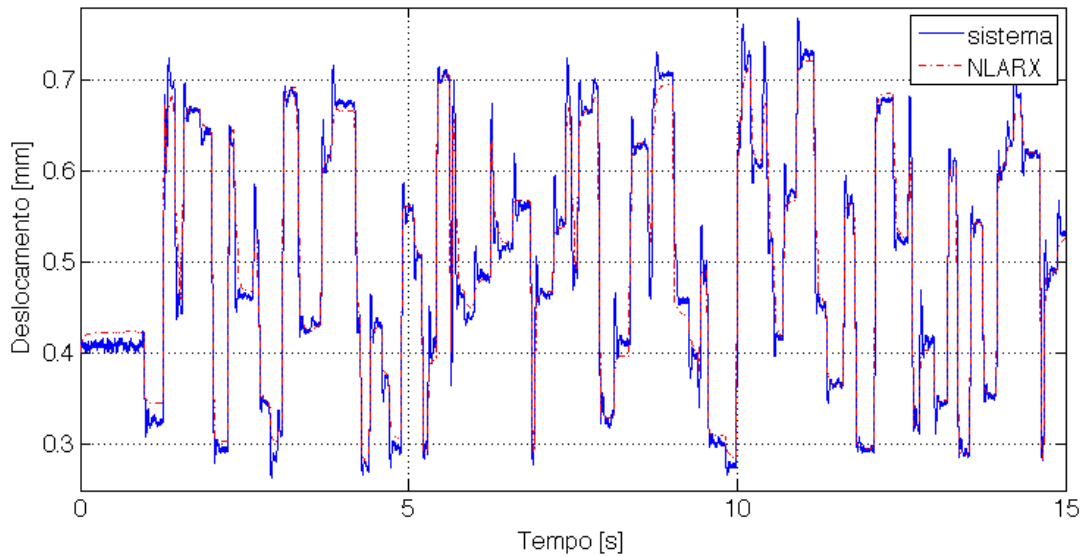


Figura 7.30: Resultado da identificação do levitador magnético vertical, a partir da *toolbox* do Matlab, utilizando modelos Nonlinear ARX com redes *Wavelet* de 10 funções ($MSE = 4,5638 \times 10^{-4}$).

O modelo mostrado na figura 7.30 obteve um MSE de $4,5638 \times 10^{-4}$ e um detalhe desta figura, entre os instantes de 1 a 4s, pode ser observado na figura 7.31.

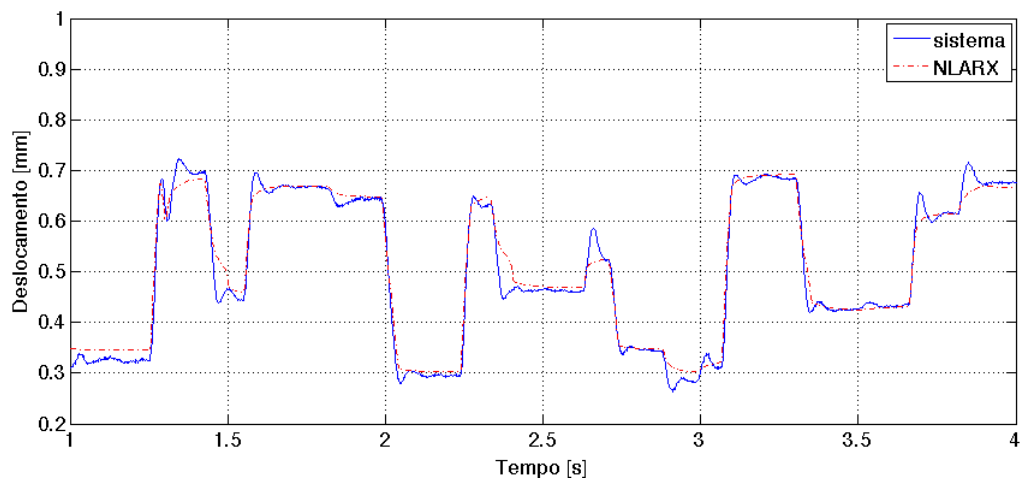


Figura 7.31: Detalhe da figura 7.30 entre os instantes 1 e 4s. Identificação do levitador magnético a partir da *toolbox* ident do Matlab, com modelos Nonlinear ARX, ($MSE = 4,5638 \times 10^{-4}$).

7.4.2 Comparação entre os Modelos NARX-OBF e Outros Modelos

A partir da tabela 7.7 e da comparação entre as figuras 7.24-b, 7.28 e 7.30, pode-se concluir que os modelos NARX-OBF, ainda que simplificados, ainda possuem menor MSE que seus concorrentes.

Tabela 7.7: Comparação entre modelos aplicados na identificação do levitador magnético horizontal.

Modelo	detalhes	$MSE_{valid.}$
NARX-OBF Completo	44 termos	$3,5212 \times 10^{-4}$
NARX-OBF Simplificado	11 termos	$3,6031 \times 10^{-4}$
NLARX/Wavelet	Rede Wavelet de 10 funções	$4,5638 \times 10^{-4}$
Wiener/Hammerstein	Rede de Sigmoides de 10 unidades	$3,9075 \times 10^{-4}$

7.4.3 Comparação entre os Modelos NARX-OBF Completos e Simplificados

A partir da comparação entre os modelos NARX-OBF completos e simplificados (mostrados nas tabelas 7.4 e 7.5 e nas duas primeiras linhas da tabela 7.7), pode-se concluir que é possível reduzir o número de termos de um modelo NARX-OBF completo, sem afetar significativamente o erro médio quadrático obtido na identificação do sistema proposto.

Pode-se concluir também que grande parte dos termos que compõem os modelos NARX-OBF completos não é relevante para a modelar o sistema em questão, uma vez que sua supressão não alterou significativamente o desempenho do modelo simplificado.

7.4.4 Comparação entre o APA e o AG na Simplificação de Modelos NARX-OBF

A partir da comparação entre as tabelas 7.5 e 7.6, pode-se inferir que os desempenhos do APA e do AG foram muito próximos na obtenção dos parâmetros de um modelo NARX-OBF simplificado.

Ressalta-se que devido à natureza da inicialização aleatória dos dois algoritmos, os parâmetros obtidos para os modelos NARX-OBF simplificados são diversos, entretanto, produziram MSEs muito próximos. Mostrando que modelos NARX-OBF diferentes podem servir como modelos adequados a um mesmo sistema.

7.5 CONCLUSÃO

Nesta seção foram expostos os resultados dos experimentos relativos à identificação de sistemas lineares simulados. Inicialmente modelos de Laguerre e Kautz foram obtidos a partir de AGs clássicos e na sequência a partir de um APA. Os resultados atingidos foram comparados e o APA obteve resultados mais exatos. No caso do modelo de Laguerre, o APA obteve MSE de $1,39 \times 10^{-4}$, contra $1,21 \times 10^{-3}$ do AG. Já no caso do modelo de Kautz, o MSE foi de $6,01 \times 10^{-4}$ para o APA, contra $7,01 \times 10^{-4}$ do AG. O APA obteve ainda uma convergência mais rápida, como pode ser observado nas figuras 7.2 e 7.4.

Na sequência, foram apresentados os resultados da identificação, a partir de dados reais, de um Levitador Magnético vertical. Os modelos utilizados foram as séries de Volterra-OBF

e modelos NARX-OBF. Seus *kernels* foram determinados por um Algoritmo de Levenberg-Marquardt e um AG clássico foi utilizado na determinação do polo que parametriza as funções ortonormais. Na comparação entre os modelos Volterra-OBF e NARX-OBF, pode-se concluir, a partir das tabelas 7.1 e 7.3, que os NARX-OBF apresentam menor MSE para modelos com o mesmo número de termos. Pode-se observar ainda que a utilização da recursão da saída, filtrada por funções ortonormais, implica em uma redução do número de termos necessários no modelo, em relação aos modelos Volterra-OBF.

Por último, foram apresentados os resultados da identificação de um levitador magnético horizontal, utilizando um método de seleção de termos mais relevantes do modelo NARX-OBF, visando a simplificação do mesmo. A partir das tabelas 7.4, 7.5 e 7.6 pode-se concluir que é possível reduzir o número de termos de um modelo NARX-OBF completo, sem afetar significativamente o erro médio quadrático, obtido na identificação do sistema proposto. Pode-se inferir também que grande parte dos termos que compõem os modelos NARX-OBF completos não é relevante para a modelar o sistema em questão, uma vez que sua supressão não alterou significativamente o desempenho do modelo simplificado.

8 CONCLUSÃO

Neste capítulo serão apresentadas as realizações desta tese, suas principais conclusões e a recapitulação dos objetivos geral e específicos. Serão apresentadas ainda as propostas de trabalhos futuros, os artigos produzidos durante o processo de confecção desta tese e as suas principais conclusões.

8.1 REALIZAÇÕES DA TESE

Sobre as realizações e objetivos alcançados nesta tese, pode-se afirmar que:

- Foi proposto um operador híbrido de busca local aplicado a Algoritmos Genéticos. Este operador foi inspirado no método dos mínimos quadrados, que em conjunto com os operadores clássicos, tornou o processo de convergência deste algoritmo heurístico mais rápido que o do AG clássico.
- Foi proposto um método inédito de otimização heurística, baseado no conceito físico do Princípio Antrópico, o APA. Na comparação entre o APA e um AG clássico, pôde-se concluir que a convergência do APA se mostrou mais rápida que a do AG. Isto devido à diversidade de modos de exploração do espaço de busca, gerada pela diversidade de leis físicas presentes nos vários universos do APA.
- Na comparação entre o modelo NARX-OBF e os modelos clássicos de identificação, modelos lineares e modelos de Volterra, pode-se afirmar que o modelo NARX-OBF apresenta resultados mais exatos que seus concorrentes clássicos, para os casos testados.
- Comparando-se os modelos de Volterra-OBF e NARX-OBF, pode-se afirmar que os modelos NARX-OBF apresentam menor MSE, para modelos com o mesmo número de termos. Desta forma, pode-se inferir que estes modelos apresentam uma maior capacidade de representação do que aqueles, isto devido à possibilidade de recursão das saídas filtradas. Cabe ressaltar que a grande desvantagem dos modelos NARX-OBF é que não se pode sempre garantir a convergência do modelo, uma vez que existem regressões da saída, que podem gerar realimentações positivas, levando o modelo à instabilidade (todavia, estes modelos são descartados pelo AG).
- Entre os modelos analisados, o modelo NARX-OBF apresentou os melhores índices de MSE, mostrando-se o modelo mais adequado a representar a dinâmica dos levitadores magnéticos, bem como a de sistemas não lineares com características semelhantes.

- Na comparação entre os modelos NARX-OBF completos e suas versões simplificadas, ou seja, compostos apenas pelos termos mais relevantes do modelo completo, pode-se concluir que a utilização dos modelos simplificados implica num aumento pouco significativo do MSE, isto para os sistemas abordados. Desta forma, pode-se trabalhar com modelos mais compactos sem perdas consideráveis na representatividade do mesmo.

8.2 PRINCIPAIS CONTRIBUIÇÕES

Podem ser citadas como contribuições desta tese de doutorado:

- Um novo paradigma de busca heurística: o Algoritmo do Princípio Antrópico, sendo o mesmo aplicado à área de Otimização e Identificação de Sistemas;
- Um operador heurístico, MQ, baseado no método dos Mínimos Quadrados, aplicado aos Algoritmos Genéticos;
- Modelos NARX, simplificados a partir de funções ortonormais de Kautz, aplicados à identificação de sistemas não lineares;
- Uma metodologia de seleção dos termos mais relevantes dos modelos NARX-OBF a partir de algoritmos heurísticos.

8.3 RECAPITULAÇÃO DE OBJETIVOS

Nesta seção serão revistos e comentados os objetivos geral e específicos desta tese.

8.3.1 Objetivo Geral

O objetivo geral deste trabalho consistiu em desenvolver uma metodologia de identificação de sistemas baseada em modelos NARX, com número de parâmetros reduzido pelo uso de funções ortonormais.

A partir dos resultados apresentados no capítulo 7 pode-se constatar que a metodologia proposta no capítulo 6, identificação de sistemas a partir de modelos NARX-OBF, mostrou-se consistente e uma alternativa interessante em relação aos seus concorrentes. Desta forma pode-se concluir que o objetivo geral desta tese foi alcançado.

8.3.2 Objetivos Específicos

Os objetivos específicos desta tese foram:

1. Identificar sistemas não lineares através de modelos NARX com seus *kernels* descritos por funções de base ortonormais;

A partir da identificação dos levitadores magnéticos, vertical e horizontal, utilizando modelos NARX-OBF, com resultados apresentados nas seções 7.3 e 7.4, pode-se concluir que este objetivo específico foi atingido.

2. Estudar o desempenho de novos operadores heurísticos aplicados a Algoritmos Genéticos destinados a resolução de problemas de identificação de sistemas;

Por meio da utilização do método dos Mínimos Quadrados como um acelerador de convergência para algoritmos heurísticos, apresentado na seção 5.1.1.3, pode-se concluir que este objetivo específico foi alcançado.

3. Desenvolver métodos de otimização heurística que permitam encontrar parâmetros aceitáveis para o modelo de um sistema não linear.

A partir do desenvolvimento (criação) do Algoritmo do Princípio Antrópico (APA), apresentado na seção 5.2, com resultados expressos nas seções 5.2.5, 7.1 e 7.4.1, pode-se concluir que este objetivo específico foi atingido.

4. Realizar a seleção dos termos mais relevantes de um modelo NARX-OBF de forma a se obter um modelo mais simples, sem a perda da capacidade de representação do modelo.

Por meio da identificação do levitador magnético horizontal utilizando modelos NARX-OBF, simplificados a partir da utilização de apenas seus termos mais relevantes, apresentados na seção 7.4.1, pode-se concluir que este objetivo específico foi alcançado.

8.4 TRABALHOS FUTUROS

Nesta seção serão expostas as seguintes possibilidades de trabalhos futuros:

- Aplicação de Programação Genética (PG) para a determinação da estrutura do modelo não linear, bem como dos parâmetros do mesmo;
- Inserir os operadores de busca local, como Nelder Mead e Mínimos Quadrados, no APA, aplicado à área de Identificação de Sistemas. Tal medida visa reduzir o tempo de convergência deste algoritmo heurístico;

- Aplicação dos métodos determinísticos de otimização como leis físicas do APA, visando obter o melhor de cada método durante um processo de otimização.
- Estudo sobre a estabilidade dos modelos NARX-OBF aplicados à modelagem de sistemas não lineares;
- Avaliar a simplificação de modelos NARX a partir de GOBFs, ao invés da utilização de funções de Laguerre ou de Kautz. Como as GOBFs tendem a ter maior poder de representação, devido ao polo variável, é possível que seja necessário um número menor de termos na série que compõem o modelo NARX-OBF.

8.5 ARTIGOS

Durante o processo de confecção deste trabalho, os seguintes artigos foram submetidos e publicados em congressos:

- OROSKI, E. DEROCO, J. e BAUCHSPIESS, A. Identificação de um levitador magnético não linear a partir de funções de Kautz e Algoritmos Genéticos. Congresso Nacional de Matemática Aplicada à Indústria, 2014.
- OROSKI, E. LOPEZ, R. H. and BAUCHSPIESS, A. Nonlinear Buck Circuit Identification Using Orthonormal Functions with Heuristic Optimization. XX Congresso Brasileiro de Automática, pp. 798 - 805, 2014.
- OROSKI, E. BAUCHSPIESS, A. COELHO, M. E. Redução de Ordem de Modelos NARX-OBF Utilizando Algoritmos Genéticos: Identificação de um Levitador Magnético. Conferência Brasileira de Dinâmica, Controle e Aplicações - DINCON, 2015.

Os seguintes artigos foram submetidos à revistas internacionais:

- OROSKI, E. LOPES, R. H. and BAUCHSPIESS, A. Identification of Nonlinear Systems using NARX-OBF Models and Heuristic Optimization. *International Journal of Modeling, Identification and Control*, 2015.
- OROSKI, E. LOPES, R. H. and BAUCHSPIESS, A. Anthropic Principle Algorithm: A new Heuristic Optimization Method. *Journal of Optimization Theory and Applications*, 2015.

8.6 CONSIDERAÇÕES FINAIS

Nesta tese foram desenvolvidos métodos de Identificação de Sistemas, baseados em funções ortonormais, focando em identificação de sistemas não lineares. Foram criados ainda alguns algoritmos heurísticos com objetivo de encontrar os parâmetros para os modelos desenvolvidos.

Sobre a aplicação dos modelos NARX-OBF, pode-se afirmar que estes podem ser uma alternativa interessante aos modelos Volterra-OBF, nos cenários em que os últimos apresentam um número muito elevado de termos ou o MSE acima do aceitável, segundo um determinado critério de projeto.

Acerca da aplicação dos algoritmos heurísticos pode-se afirmar que os mesmos se comportam bem diante de problemas como a dimensionalidade e os mínimos locais, mesmo não garantindo a determinação do ótimo global. E na sua aplicação à área de Identificação de Sistemas tem-se ainda as vantagens de que tais métodos desprezam os modelos instáveis e podem auxiliar na escolha da estrutura do modelo.

Sobre o APA, pode-se afirmar que trata-se de um algoritmo heurístico que oferece resultados próximos aos de seus principais concorrentes. Como este método ainda está em estágio inicial de desenvolvimento, é possível afirmar que o mesmo ainda pode ser melhorado com auxílio de leis físicas mais adequadas a cada tipo específico de busca. Outra possibilidade, neste sentido, seria a inserção de aceleradores de convergência, como os métodos dos Mínimos Quadrados ou de Nelder Mead.

No processo de simplificação dos modelos NARX-OBF, pode-se concluir que os mesmos podem ter o seu número de termos reduzidos consideravelmente, e isto sem um aumento expressivo do MSE para o modelo simplificado. Com modelos mais simples pode-se atingir uma implementação computacionalmente mais eficiente e, em muitos casos, sem que o poder de representação do modelo seja perdido.

Um aspecto interessante deste trabalho, assim como outros trabalhos em Identificação de Sistemas, é que pode-se chegar a um modelo matemático para um sistema real. Na área de Controle, geralmente parte-se da premissa de que se tem disponível um modelo para o sistema que se almeja controlar. Na prática, isto geralmente não é verdade. Neste contexto, os trabalhos em Identificação de Sistemas podem servir como ponte entre as técnicas de controle, que pressupõem a existência de um modelo, e o mundo real, em que não se conhece os parâmetros matemáticos do sistema que se deseja controlar.

O autor acredita que o trabalho desenvolvido nesta tese tem potencial para ser aplicado, de forma prática, nas áreas de Engenharia de Controle, não se restringindo apenas às contribuições teóricas. E particularmente, na futura área de atuação do autor, a docência, é possível a aplicação direta dos conceitos desenvolvidos nesta tese.

REFERÊNCIAS BIBLIOGRÁFICAS

ABDEL, H.; FUKUSHIMA, M. Minimizing multimodal functions by simplex coding genetic algorithm. *Optimization Methods and Software.*, v. 18, n. 3, p. 265–282, 2003.

AGUIRRE, L. A. Identificação de sistemas não lineares utilizando modelos narmax polinomiais - uma revisão e novos resultados. *SBA Controle e Automação*, v. 09, n. 02, p. 90–106, 1998.

AGUIRRE, L. A. *Introdução à Identificação de Sistemas - Técnicas Lineares e Não Lineares Aplicadas à Sistemas Reais*. Belo Horizonte MG: Editora UFMG, 2007.

AKAIKE, H. A new look at the statistical model identification. *IEEE transactions on Automatic Control*, v. 19, issue 6, p. 716–723, 1974.

AKANYETI, O. et al. An application of lyapunov stability analysis to improve the performance of narmax models. *Elsevier - Robotics and Autonomous Systems*, v. 58, p. 229–238, 2009.

ATASHPAZ, E.; LUCAS, C. An algorithm for optimization inspired by imperialistic competition. *IEEE Congress on Evolutionary Computation*, p. 4661–4667, 2007.

BADWE, A.; PATWARDHAN, S.; GUDI, R. Closed loop identification using direct approach and high order arx/gobf-arx. *Elsevier - Journal of Process Control*, v. 21, p. 1056–1071, 2011.

BALDACCHINO TARA; ANDERSON, S.; KADIRKAMANATHAN, V. Computational system identification for bayesian narmax modelling. *Elsevier - Automatica*, n. 49, p. 2441–2651, 2013.

BARROW, J. D.; TIPLER, F. J. *The Anthropic Cosmological Principle*. 1st. Oxford, UK: Oxford Univ. Press, 1986.

BATISTA, E. L. O.

Filtros de Volterra: Estruturas Interpoladas e Modelos Estocásticos, Outubro 2009.

BAUMEISTER, J.; AO, A. L. *Introdução à teoria de Controle e Programação Dinâmica*. 1st. Rio de Janeiro, RJ: Editora do IMPA, 2014.

BELT, H. W.

Orthonormal Bases for Adaptive Filtering, May 1997.

BOYD, S.; CHUA, L. O. Fading memory and the problem of approximating nonlinear operators with volterra series. *IEEE Trans. on Circuits and Systems*, v. 32, n. 11, p. 1150–1161, 1985.

BOYD, S.; VANDENBERGHE, L. *Convex Optimization*. Cambridge University, UK: Cambridge Press, 2009.

- CAMPELLO, R.; AMARAL, W. C.; OLIVEIRA, G. H. C. Identificação de controle de processos via desenvolvimento de séries ortonormais. *Revista Controle e Automação*, Vol. 18, pp 301-319. Jan 2007.
- CAMPELLO R. J. G. B; OLIVEIRA, G. H. C. *Modelos Não Lineares*. São Paulo - SP: Enciclopédia de Automática. Editora Edgar Blucher. Vol. 3, Cap. 4., 2007.
- CARTER, B. Confrontation of cosmological theories with observational data. *IAU Symposium*, Krakow, n. 63, 1973.
- COELHO, M. A. E. Kit para levitação magnética com ímãs permanentes. Congresso Brasileiro de Automação. Setembro 2012.
- COELHO, M. A. E. Identificação em kit didático para levitação magnética. SBAI-Simpósio Brasileiro de Automação Inteligente. Outubro 2013.
- COLEY, A. D. *An Introduction to Genetic Algorithms for Scientists and Engineers*. London, Uk: Would Scientific, 1999.
- COMITTI, V. S. Princípio antrópico cosmológico. *Revista Brasileira de Ensino da Física*, v. 33, n. 1, p. 1504, 2011.
- DORF, R.; BISHOP, R. *Sistemas de Controle Moderno*. 8^a. [S.l.]: LTC, 2001.
- EARNSHAW, S. On the nature on the molecular forces with regulate the constitution of luminiferous ether. *Trans. Camb. Phill. Soc.*, v. 7, p. 97–112, 1842.
- ECP, M. *Manual for Model 730, Magnetic Levitation System, Educational Control Products*. California, USA, 1999.
- ELKAIN, G. H. System identification-based control of an unmanned autonomous wind-propelled catamaran. *Elsevier - Control Engeneering Practice*, v. 17, n. 02, p. 158–169, 2009.
- EYKHOFF, P. *System Identification: Parameter and State Estimation*. New York: John Wiley Sons, 1974.
- FU, Y.; DUMONT, G. A. An optimum time scale for discrete laguerre network. *IEEE Transactions on Automatic Control*, v. 38, n. 6, p. 934–938, 1993.
- HAWKING, S.; MLODINOW, L. *O Grande Projeto*. 1st. [S.l.]: Editora Nova Fronteira, 2012.
- HE, S.; WU, K. H.; SAUNDERS, J. R. Group search optimizer: An optimization algorithm inspired by animal searching behavior. *IEEE Transactions on Evolutionary Computation*, v. 13, n. 5, p. 973–990, 2009.
- HENDRICK, P. J. K.; GIRARD, A. *Control of Nonlinear Dynamic Systems: Theory and Applications*. 1st. [S.l.]: Berkley, 2010.
- HEUBERGER, P.; HOF, P. V. D.; BOSGRA, O. H. A generalized orthonormal basis for linear dynamical systems. *IEEE Transactions*, v. 40, n. 3, p. 451–465, 1995.

- HEUBERGER, P.; HOF, P. V. D.; WAHLBERG, B. *System Identification: Theory for the User*. 1st. [S.l.]: Springer Press, 2005.
- ISERMANN, R.; MUNCHHOF, M. *Identification of Dynamic Systems - An Introduction with Applications*. 1st. [S.l.]: Springer, 2011.
- IZMAILOV, A.; SOLODOV, M. *Otimização- Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade*. 3rd. [S.l.]: Editora do IMPA, 2014.
- KALLOSH, R.; LINDE, A. M theory, cosmological constant and anthropic principle. *Physical Review D* 67, 023510, 2003.
- KIBANGU, A.; FAVIER, G.; HASSANI, M. Iterative optimization method of gob-volterra filters. *Proceedings of the 16th IFAC World Congress*, v. 16, p. 129–134, 2005.
- KOZA, J. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. 6th. [S.l.]: MIT Press, 1998.
- LANDAU, I.; ZITO, G. *Digital Control - Design Identification and Implementation*. Germany: Springer Press, 2006.
- LEMMA, D. T.; RAMASAMY; SHUHAIMI. System identification using orthonormal basis filters. *Journal of Applied Sciences*, v. 10, n. 21, p. 2516–2522, 2010.
- LEMMA, D. T.; RAMASAMY, M. Closed-loop identification of systems with uncertain time delays using arxDoBF structure. *Elsevier - Journal of Process Control*, v. 21, n. doi:10.1016/j.jprocont.2011.06.021, p. 1148–1154, 2011.
- LIMA, E. L. *Análise Real, Vol. 3, Análise Vetorial*. 2nd. [S.l.]: Editora do IMPA, 2007.
- LIMA, E. L. *Algebra Linear*. Rio de Janeiro, RJ: Editora do IMPA, 2008.
- LIMA, E. L. *Análise Real, Vol. 1, Funções de uma Variável*. 6th. [S.l.]: Editora do IMPA, 2013.
- LIMA, E. L. *Análise Real, Vol. 2, Funções de n Variáveis*. 6th. [S.l.]: Editora do IMPA, 2013.
- LIMA, E. L. *Espaços Métricos*. 2th. [S.l.]: Editora do IMPA, 2013.
- LJUNG, L. *System Identification - Theory For The User*. Upper Saddle River - New Jersey: Prentice Hall PTR, 1999.
- MACHADO, J. B.; CAMPELLO, J. G. B.; AMARAL, W. C. Modelos lineares gobf com funções internas. *XVIII Congresso Brasileiro de Automática*, v. 10, n. 21, p. 2516–2522, 2010.
- MADAKYARU, M.; PATWARDHAN, S. Adaptive predictive control using gobf-arx models: An experimental case study. *IFAC - International Symposium on Dynamics and Control of Process Systems*, 2013.
- MAESTRELLI, R. *Funções Ortonormais em Tempo Contínuo com Seleção Ótima das Dinâmicas do Modelo na Identificação de Sistema no Domínio da Frequência*. mathesis, 2010.

- MARQUARDT, D. W. An algorithm for least squares estimation of nonlinear parameters. *Journal of Society for Industrial and Applied Mathematics*, v. 11, n. 2, p. 431–441, 1963.
- MICHELL, M. *An Introduction to Genetic Algorithms*. Massachusetts: MIT Press, 1999.
- MOCKUS JONAS; EDDY, W.; REKLAITIS, G. *Bayesian Heuristic Approach to Discrete and Global Optimization: Algorithms, Visualization, Software and Applications*. [S.l.]: Kluwer Academic Publishers, 1st edition, 1997.
- MOHIDEEN, K. et al. Real-coded genetic algorithm for system identification and tuning of a modified model reference adaptive controller for hybrid tank system. *Elsevier- Applied Mathematical Modelling*, n. 37, p. 3829–3847, 2013.
- MONTEIRO, G. V. d. S. *Controlo Não Linear. 2^a*. Setúbal - Portugal: Escola Superior de Tecnologia, 2003.
- MONTEIRO, L. H. A. *Sistemas Dinâmicos*. São Paulo: Editora da Física, 2006.
- NELDER, J. A.; MEAD, R. A simplex for function minimization. *Computer Journal*, Vol. 7, pp 308-313. Jan 1965.
- NELLES, O. *Nonlinear System Identification*. New York: Springer, 2001.
- NEMETH, J.; KOLLAR, I. Identification of volterra kernels using interpolation. *IEEE Transactions on Instrumentation and Measurements*, Vol. 51. Jan 2002.
- NOCEDAL, J.; WRIGHT, S. *Numerical Optimization*. New York: Springer, 1999.
- OLSSON, A. E. *Particle Swarm: Theory, Techniques and Applications*. 1st. [S.l.]: New Science Publishers, 2011.
- OPPENHEIM, A.; SHAFER, R.; BUCK, J. *Discrete Time Signal Processing*. New Jersey: Prentice Hall, 1998.
- OROSKI, E.; HOLDORF, R.; BAUCHSPIESS, A. Identification of a magnetic levitator using narx-obf models and genetic algorithm. *International Journal of Modeling Identification and Control*, 2015.
- PAPOULIS, A.; PILLAI, S. U. *Probability, Random Variables and Stochastic Processes*. 4th. Singapore: McGraw-Hill Higher, 2002.
- RAHROOH, A.; SHEPARD, S. Identification of nonlinear systems using narmax models. *Elsevier - Nonlinear Analysis*, n. 71, p. 1198–1202, 2009.
- RAVAL, F.; MAKWANA, J. International journal of advances in engineering and technology. *The Computer Journal*, v. 01, p. 99–106, 2011.
- RORRES, A. *Algebra linear com Aplicações*. São Paulo - SP: Artmed Editora, 2000.
- ROSA, A. *Identificação de Sistemas Não Lineares Usando Modelos de Volterra Baseados em Funções Ortonormais de Kautz e Generalizadas*. mathesis, 2009.
- ROSA, A.; CAMPELLO, R.; AMARAL, W. Exact search directions for optimizations of linear and nonlinear models based on generalized orthonormal functions. *IEEE Transactions on Automatic Control*, Vol. 54, N 12. Jan 2009.

- ROSENBROCK, H. H. An automatic method for finding the greatest or least value of a function. *The Computer Journal*, v. 3, p. 175–184, 1960.
- SASTRY, K.; GOLDBERG, D.; KENDALL, G. *Genetic Algorithms*. 2005.
- SEBER, G.; WILD, C. J. *Nonlinear Regression*. New Jersey, USA: Wiley, 2003.
- SILVA, S. Nonlinear mechanical system identification using discrete-time volterra models and kautz filter. *9th Brazillian Conference on Dynamics, Control and their Applications*, 2010.
- SIMON, D. Biogeography-based optimization. *IEEE Transactions on Evolutionary Computation*, v. 12, n. 6, p. 702–713, 2008.
- SOLNON, C. *Ant Colony Optimization and Constraint Programming*. 1st. [S.l.]: Wiley Press, 2010.
- STARKMAN, G. D.; TROTTA, R. Why anthropic principle cannot predict λ . *Physical Review Letters*, *PRL 97, 201301.*, 2006.
- STRANG, G. *Introduction to Linear Algebra*. Wellesley, MA: Wellesley-Cambridge Press, 2009.
- STRANG, G. *Álgebra Linear e suas Aplicações*. 4^a. [S.l.]: Cengage Learning, 2013.
- TIELS, K.; SCHOUKENS, J. Wiener system identification with generalized orthonormal basis function. *Elsevier - Automatica*, v. 50, p. 3147–3154, 2014.
- WAHLBERG, B. System identification using kautz models. *IEEE Transactions on Automatic Control*, v. 39, n. 6, p. 1276–1282, 1994.