

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

HIGOR DE SOUZA SERAFIN

**IDENTIFICAÇÃO DE SISTEMAS NÃO-LINEARES COM MODELO
VOLTERRA-KAUTZ**

CURITIBA

2022

HIGOR DE SOUZA SERAFIN

**IDENTIFICAÇÃO DE SISTEMAS NÃO-LINEARES COM
MODELO VOLTERRA-KAUTZ**

Identification of nonlinear systems with the Volterra-Kautz model

Dissertação apresentada como requisito para obtenção do título de Mestre em Engenharia Elétrica e Informática Industrial, do Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial, da Universidade Tecnológica Federal do Paraná (UTFPR).

Orientador: Prof. Dr. Andre Eugenio Lazzaretto

Coorientador: Prof. Dr. Elder Oroski

CURITIBA

2022



[4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es).

Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.



HIGOR DE SOUZA SERAFIN

IDENTIFICAÇÃO DE SISTEMAS NÃO-LINEARES COM MODELO VOLTERRA-KAUTZ

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestre em Ciências da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Engenharia De Automação E Sistemas.

Data de aprovação: 12 de Agosto de 2022

Dr. Andre Eugenio Lazzaretti, Doutorado - Universidade Tecnológica Federal do Paraná

Dr. Elder Oroski, Doutorado - Universidade Tecnológica Federal do Paraná

Dr. Gustavo Henrique Da Costa Oliveira, Doutorado - Universidade Federal do Paraná (Ufpr)

Dr. Victor Baptista Frend, Doutorado - Universidade Tecnológica Federal do Paraná

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 12/08/2022.

Dedico este trabalho aos meus pais, Claudemir
Serafin e Tatiane de Souza Serafin, que me
proporcionaram toda a base para ir além.

AGRADECIMENTOS

À minha esposa Karina por todo apoio, paciência e amor que me dispusera para que eu conquistasse mais este sonho.

Aos meus orientadores, por suas incontáveis revisões, que me mostram os caminhos a serem seguidos, e pela confiança depositada.

Enfim, a Deus por ter colocado todos esses no meu caminho e ter alcançado por mim onde minhas mãos não chegaram.

“Todos os modelos estão errados,
mas alguns são úteis.”
(George E. P. Box, 1976)

RESUMO

SERAFIN, Higor. **Identificação de sistemas não-lineares com modelo Volterra-Kautz**. 2022. 86 f. Dissertação (Mestrado em Engenharia Elétrica e Informática Industrial) – Universidade Tecnológica Federal do Paraná. Curitiba, 2022.

Este trabalho apresenta o problema da estimação dos polos de Kautz na expansão do *kernel* com modelos de Volterra-Kautz. A solução ótima para os parâmetros que modelam os polos ainda está em aberto na literatura. Assim, este trabalho traz duas abordagens de otimização, uma usando o algoritmo de Levenberg-Marquardt e outra a Otimização Bayesiana para encontrar os parâmetros de Kautz. As bases de funções para os *kernels* são construídas por meio de uma estrutura de filtros digitais. Na validação dos algoritmos implementados são utilizados dados coletados de um sistema eletricamente acoplado e, por meio do *software* MATLAB®, diversos experimentos foram executados para estudar os impactos dos diferentes parâmetros de cada método empregado, também foi observado como o aumento das funções na base influenciam o modelo. Como resultado, os modelos gerados atingiram um erro médio quadrático na ordem de 10^{-4} , enquanto que trabalhos da literatura, que utilizaram o mesmo conjunto de dados para a modelagem, atingiram um erro médio quadrático com ordem entre 10^{-2} e 10^{-3} .

Palavras-chave: Identificação de Sistemas. Volterra-Kautz. Kautz. Levenberg-Marquardt. Otimização Bayesiana.

ABSTRACT

SERAFIN, Higor. **Identification of nonlinear systems with the Volterra-Kautz model**. 2022. 86 p. Dissertation (Master's Degree in Engenharia Elétrica e Informática Industrial) – Universidade Tecnológica Federal do Paraná. Curitiba, 2022.

This work presents the problem of estimating the Kautz poles in the kernel expansion with Volterra-Kautz models. The optimal solution for the parameters that model the poles is still open in the literature. Thus, this work brings two optimization approaches, one using the Levenberg-Marquardt algorithm and the other using Bayesian Optimization to find the Kautz parameters. Function bases for kernels are built through a structure of digital filters. To validate the implemented algorithms, data collected from an electrically coupled system was used and, on the MATLAB® software, several experiments were performed to study the impacts of the different parameters of each method used, it was also observed how the increase of the functions in the base influence the model. As a result, the generated models reached a mean square error of the order of 10^{-4} , while works in the literature, which used the same dataset for modeling, reached a mean square error in the order between 10^{-2} and 10^{-3} .

Keywords: System Identification. Volterra-Kautz. Kautz. Levenberg-Marquardt. Bayesian optimization.

LISTA DE ALGORITMOS

Algoritmo 1 – Método de LM.	42
Algoritmo 2 – Descrição do funcionamento da OB.	49

LISTA DE ILUSTRAÇÕES

Figura 1 – Estrutura do modelo de Volterra-Kautz.	31
Figura 2 – Dados amostrados para o exemplo a partir da função $y(t)$ em (71) com $x_1 = 12,0$; $x_2 = 37,0$; $x_3 = 0,600$; $x_4 = -2,00$	43
Figura 3 – Comparação entre os dados amostrados com os x_n reais e os dados amostrados com os x_n iniciais.	44
Figura 4 – Comparação entre os dados amostrados e os dados estimados com o método de LM.	44
Figura 5 – Comportamento do MSE e do parâmetro λ durante a minimização para estimar os parâmetros com o LM.	45
Figura 6 – Comparação entre os dados amostrados e os dados estimados com o algoritmo de LM para pontos iniciais distantes dos pontos reais x_n	46
Figura 7 – Exemplificação do funcionamento da OB.	49
Figura 8 – Exemplificação do funcionamento da PI.	53
Figura 9 – Dados amostrados para o exemplo a partir da função (93) considerando $x = 8,00$	56
Figura 10 – Resultado da estimação do parâmetro x com a OB.	57
Figura 11 – Comportamento do MSE durante a OB para encontrar o valor de x	57
Figura 12 – Evolução da estimativa de x durante a OB.	58
Figura 13 – Comportamento do modelo com a OB durante a busca por x	58
Figura 14 – Fluxograma de visão geral dos algoritmos implementados.	61
Figura 15 – Sistema usado na coleta de dados.	62
Figura 16 – Sinal de excitação do sistema.	63
Figura 17 – Sinal de resposta do sistema.	63
Figura 18 – Base de Kautz como filtro digital.	64
Figura 19 – Comportamento do MSE com aumento das funções de Kautz no modelo Volterra-Kautz utilizando o algoritmo de LM para otimização.	66
Figura 20 – Resposta do Modelo de Volterra-Kautz de um sistema composto por dois motores e uma polia, com <i>kernels</i> modelados por 2 funções de Kautz, utilizando otimização por método de LM.	67
Figura 21 – Resposta do Modelo de Volterra-Kautz de um sistema composto por dois motores e uma polia, com <i>kernels</i> modelados por 20 funções de Kautz, utilizando otimização por método de LM.	68
Figura 22 – Comportamento do MSE durante a busca pelos polos de Kautz no modelo de Volterra-Kautz utilizando o algoritmo de LM para otimização.	69
Figura 23 – Comportamento do parâmetro λ durante a busca pelos polos de Kautz no modelo de Volterra-Kautz utilizando o algoritmo de LM para otimização.	69
Figura 24 – Comportamento dos parâmetros b e c durante a execução da otimização com o algoritmo de LM para encontrar os polos de Kautz no modelo de Volterra-Kautz.	70
Figura 25 – Comportamento do MSE durante o aumento das funções de Kautz na base do modelo de Volterra-Kautz utilizando o algoritmo de OB para encontrar os polos.	71
Figura 26 – Resposta do modelo de Volterra-Kautz de um sistema composto por dois motores e uma polia com <i>kernels</i> modelados por 20 funções de Kautz, utilizando OB.	72

Figura 27 – Comportamento do MSE durante a busca pelos polos de Kautz no modelo de Volterra-Kautz utilizando o algoritmo de OB.	73
Figura 28 – Modelo gerado durante a OB.	73

LISTA DE TABELAS

Tabela 1 – Parâmetros do algoritmo de LM.	64
Tabela 2 – Resultados da otimização com o algoritmo de LM para estimar os polos de Kautz no modelo de Volterra-Kautz.	65
Tabela 3 – Resultados do MSE para 20 funções de Kautz na base utilizando o algoritmo de LM para estimar os polos.	66
Tabela 4 – Parâmetros do algoritmo de OB.	71
Tabela 5 – Resultados do algoritmo de OB na estimação dos polos de Kautz.	72
Tabela 6 – Comparação entre os resultados das modelagens do <i>benchmark</i> com diferentes métodos.	74

LISTA DE ABREVIATURAS, SIGLAS E ACRÔNIMOS

SIGLAS

ARX	<i>Auto-Regressive with eXogenous inputs</i>
OBF	<i>Orthonormal Bases of Functions</i>
GOBF	<i>Generalized Orthonormal Bases of Functions</i>
NARX	<i>Nonlinear Auto-Regressive with eXogenous inputs</i>
IIR	<i>Infinite Impulse Response</i>
MMQ	Método dos Mínimos Quadrados
LM	Levenberg-Marquardt
MSE	<i>Mean Square Error</i>
OB	Otimização Bayesiana
EGO	<i>Efficient Global Optimization</i>
PI	<i>Probability of Improvement</i>
EI	<i>Expected Improvement</i>
CDF	<i>Cumulative Distribution Function</i>
PDF	<i>Probability Density Function</i>
NARMAX	<i>Nonlinear Autoregressive Moving Average with eXogenous input</i>
RBFNNs	<i>Radial Basis Functions Neural Networks</i>

LISTA DE SÍMBOLOS

LETRAS GREGAS

α	Ganhos Escalares
ϕ	Funções Ortonormais
ξ	Polos das Funções Ortonormais
τ	Operador de Atraso
ψ	Matriz de Regressores
θ	Vetor de Parâmetros Escalares
λ	Fator de Regularização
μ	Média
σ	Variância
Ψ	Função de Distribuição Acumulada
φ	Função de Densidade de Probabilidade
π	Quantidade de Permutações

SOBRESCRITOS

$N + 1$	Passo de Tempo Posterior
$\bar{\xi}$	Polo Complexo Conjugado de ξ
\hat{y}	Valor Aproximado/Estimado de y
h^{sim}	Kernel simétrico de h

NOTAÇÕES

$\ \cdot\ $	Norma
$\langle \cdot \rangle$	Produto Interno
\mathcal{N}	Distribuição Normal
\mathcal{PG}	Processo Gaussiano
\mathbb{E}	Esperança Matemática
∇f	Matriz Jacobiana de f
$\nabla^2 f$	Matriz Hessiana de f

SUMÁRIO

1	INTRODUÇÃO	15
1.1	OBJETIVOS	17
1.1.1	Objetivo Geral	17
1.1.2	Objetivos Específicos	17
1.2	ORGANIZAÇÃO DA DISSERTAÇÃO	17
2	FUNÇÕES ORTONORMAIS	19
2.1	MODELAGEM DE SISTEMAS COM FUNÇÕES ORTONORMAIS	20
2.2	FUNÇÕES ORTONORMAIS GENERALIZADAS	22
2.3	FUNÇÕES DE LAGUERRE	23
2.4	FUNÇÕES DE KAUTZ	24
2.5	SÍNTESE	25
3	MODELO VOLTERRA-KAUTZ	26
3.1	SÉRIE DE VOLTERRA	27
3.2	MODELO VOLTERRA-KAUTZ	30
3.3	SÍNTESE	33
4	ESTIMAÇÃO DE PARÂMETROS	35
4.1	MÍNIMOS QUADRADOS	35
4.2	LEVENBERG-MARQUARDT	38
4.2.1	Método de Levenberg-Marquardt	39
4.2.2	Exemplo Ilustrativo	42
4.3	OTIMIZAÇÃO BAYESIANA	47
4.3.1	Método de Otimização Bayesiana	47
4.3.2	Processo Gaussiano	50
4.3.3	Função de Aquisição	52
4.3.3.1	Probabilidade de Melhoria	53
4.3.3.2	Melhoria Esperada	54
4.3.4	Exemplo Ilustrativo	56
4.4	SÍNTESE	59
5	ESTUDO DE CASO	60
5.1	FORMULAÇÃO DO PROBLEMA	60
5.2	ANÁLISES E DISCUSSÕES	62
6	CONSIDERAÇÕES FINAIS	75
6.1	PUBLICAÇÕES	77
	REFERÊNCIAS	78

1 INTRODUÇÃO

A modelagem dos fenômenos e dos sistemas que rodeiam o ser humano por meio de representações matemáticas não é um desafio novo. Desde a Antiguidade, matemáticos como Arquimedes, Euclides e Pitágoras elaboravam teorias, métodos e técnicas na busca por descrições e soluções dos problemas que os cercavam. Apesar do tempo decorrido, a modelagem matemática ainda é um problema na atualidade, tendo sua complexidade incrementada pelo desenvolvimento tecnológico. Nesse cenário surgem sistemas mais complexos que passaram a requer estratégias de modelagem cada vez mais elaboradas, dentre elas, a Identificação de Sistemas.

Em Aguirre (2007), a Identificação de Sistemas é tratada como uma área do conhecimento que investiga as formas de representar e analisar sistemas dinâmicos a partir das observações dos seus dados de entrada e de saída. Os sistemas dinâmicos são divididos em dois grandes grupos: sistemas lineares e sistemas não-lineares. Esta divisão é feita de acordo com as propriedades de homogeneidade e de superposição, sendo que os sistemas lineares são aqueles que apresentam estas duas características e os não-lineares são aqueles que não obedecem a uma ou as duas propriedades.

A modelagem de sistemas dinâmicos lineares é geralmente desenvolvida com base nas estruturas ARX (*Auto-Regressive with eXogenous inputs*), em que a saída de um sistema em um determinado momento do tempo discreto é descrita por meio da utilização das amostras anteriores dos sinais de entrada e de saída (CAMPELLO; OLIVEIRA, 2007). Um problema presente nesta abordagem é que uma estrutura autorregressiva, em geral, é sensível à ordem do modelo e ocasiona um acúmulo dos erros da estimação, principalmente para horizontes longos de previsão (CAMPELLO *et al.*, 2007).

Como uma das estratégias para contornar este problema, surge a abordagem que utiliza-se de modelos sem realimentação de saída, dotados de uma estrutura OBF (*Orthonormal Bases of Functions*) (CAMPELLO *et al.*, 2007). Esses modelos são desenvolvidos por meio de bases de funções ortonormais (NINNESS *et al.*, 1995), tais como as funções de Laguerre (WAHLBERG, 1991), as funções de Kautz (WAHLBERG, 1994) ou as Funções de Bases Ortonormais Generalizadas (*GOBF - Generalized Orthonormal Bases of Functions*) (HEUBERGER *et al.*, 2005). Estes modelos apresentam algumas características que os tornam interessantes nas aplicações de identificação de sistemas (FU; DUMONT, 1993; HEUBERGER *et al.*, 2005), tais como:

- Ausência dos erros de previsão, devido à realimentação da saída;
- Não há necessidade do conhecimento dos termos relevantes de entrada e de saída do sistema;
- Possibilidade do aumento da capacidade de representação do modelo com o aumento do número de funções na base;
- Representação estável de um sistema estável;
- Tolerância às dinâmicas não modeladas;
- Baixa sensibilidade a variações nos parâmetros estimados.

Para a modelagem de sistemas não-lineares existem diversas estruturas que podem ser citadas, por exemplo, os modelos de Wiener e Hammerstein que arranjam modelos dinâmicos lineares e estáticos não-lineares para representar o sistema. Os modelos que usam estruturas NARX (*Nonlinear Auto-Regressive with eXogenous inputs*) podem ser entendidos como uma generalização da estrutura ARX para o caso não-linear e os modelos de Volterra, que se baseiam em *kernels*, podem ser compreendidos como uma generalização do modelo de resposta ao impulso. O uso da série de Volterra para modelar sistemas não-lineares possibilita uma ideia do significado físico do processo (MORTENSEN, 1983). Dessa forma tem-se o benefício conceitual de realizar a descrição de um processo não-linear como uma generalização do caso linear, ou seja, a série de Volterra possibilita o uso de conceitos de sistemas lineares para identificar sistemas não-lineares (EYKHOFF, 1979).

Contudo, a modelagem de sistemas não-lineares utilizando modelos de Volterra implica no problema da estimação dos *kernels* do modelo (CAMPELLO *et al.*, 2007). Uma alternativa para contornar essa situação foi proposta em Wiener (1966), na qual os *kernels* de Volterra são desenvolvidos por meio de base de funções ortonormais. Neste cenário, o desafio da modelagem está na determinação dos polos que parametrizam as funções da base. No caso das funções de Laguerre, a determinação do polo ótimo já está solucionada na literatura (MASNADI-SHIRAZI; AHMED, 1991). Entretanto, para as funções de Kautz, a determinação dos polos ótimos ainda é um problema em aberto. Desta forma, o presente trabalho concentra os esforços na estimação dos polos de Kautz para identificação de sistemas não-lineares utilizando os modelos de Volterra com *kernels* desenvolvidos por meio das funções de Kautz. Para isso, serão utilizadas duas abordagens de estimação de parâmetros: uma com o algoritmo de Levenberg-Marquardt e outra usando Otimização Bayesiana

1.1 OBJETIVOS

1.1.1 Objetivo Geral

O objetivo deste trabalho é estimar os polos das funções ortonormais de Kautz nos modelos Volterra-Kautz, por meio do uso de técnicas de otimização para sistemas não-lineares, como método de Levenberg-Marquardt e a Otimização Bayesiana.

1.1.2 Objetivos Específicos

Para que o objetivo geral deste trabalho possa ser atingido, os seguintes objetivos específicos precisam ser trabalhados:

- Modelar o *kernel* de Volterra com uma base de funções ortonormais de Kautz, gerando o modelo Volterra-Kautz.
- Otimizar os parâmetros do modelo Volterra-Kautz, recorrendo ao Método dos Mínimos Quadrados para estimar os parâmetros lineares.
- Otimizar os parâmetros do modelo Volterra-Kautz, utilizando o algoritmo de Levenberg-Marquardt e Otimização Bayesiana para os parâmetros não-lineares.
- Aplicar os algoritmos propostos de estimação de parâmetros a um sistema de *benchmark*.
- Analisar o desempenho de cada algoritmo na estimação dos parâmetros do modelo Volterra-Kautz.
- Comparar o desempenho dos algoritmos propostos com os resultados presentes na literatura.

1.2 ORGANIZAÇÃO DA DISSERTAÇÃO

Esta dissertação está organizada em seis Capítulos, sendo que no Capítulo 2 será abordado o desenvolvimento matemático das funções ortonormais e algumas funções específicas receberão um maior destaque, como as GOBF, as funções de Laguerre e as funções de Kautz, por terem um interesse significativo para o presente trabalho. No Capítulo 3 será apresentada a série de Volterra e como o modelo Volterra-Kautz surgiu como uma alternativa para contornar o

problema da quantidade de parâmetros necessários a serem estimados. Na sequência, o Capítulo 4 irá discorrer sobre alguns métodos para estimação de parâmetros, sendo eles o Método dos Mínimos Quadrados, o algoritmo de Levenberg-Marquardt e a Otimização Bayesiana. No Capítulo 5, será realizado um estudo de caso, a partir de dados coletados de um sistema real, para validar os algoritmos implementados e a identificação deste sistema com o modelo Volterra-Kautz. Por fim, no Capítulo 6 serão abordadas as considerações finais sobre o presente trabalho.

2 FUNÇÕES ORTONORMAIS

Neste capítulo serão apresentadas as funções ortonormais. Primeiramente será realizado um levantamento bibliográfico dos trabalhos mais relevantes da área. Na Seção 2.1 serão expostos os referenciais matemáticos que regem os desenvolvimentos das funções ortonormais. Na sequência são apresentadas as funções ortonormais mais utilizadas, iniciando pelas GOBF na Seção 2.2. Depois, na Seção 2.3, será apresentado o caso particular para polo real, isto é, as Funções de Laguerre. Por último, na Seção 2.4, será descrito o caso particular para os polos complexos, ou seja, as Funções de Kautz.

Os primeiros trabalhos que tratam das funções ortonormais apareceram em meados da década de 1920 em Takenaka (1925) e Malmquist (1925). Esses estudos apresentam as funções ortonormais generalizadas com aplicações de aproximações por interpolações. Na década seguinte, Walsh (1935) foi um dos principais trabalhos com o uso das funções ortonormais, ainda no contexto de aproximação por interpolação.

Neste período, Wiener passou a estudar o uso das redes contínuas de Laguerre para construção de previsores ótimos. Porém, uma apresentação mais completa dos seus trabalhos são apresentadas mais tardiamente em Wiener (1933) e Wiener (1949). Sob a mentoria de Wiener, Lee apresentou os trabalhos que usaram as funções de Laguerre para trabalhos na área de redes elétricas (LEE, 1932).

Na década de 1950, Kautz revisou a formulação das GOBFs (KAUTZ, 1954; KAUTZ, 1952) e suas versões no tempo contínuo para aplicações em síntese de rede de parâmetros finitos de sistemas transitórios. Nesta mesma época foi produzida a referência base dos polinômios ortonormais em Sansone (1959) e o primeiro trabalho de identificação de sistemas por meio de *kernels* foi produzido em Aronszajn (1950).

Nas décadas seguintes, 1960 e 1970, trabalhos com parametrização ortonormal para sistemas discretos e aplicações em controle foram feitas nos trabalhos de Broome (1965) e Mendel (1966). As primeiras implementações de funções ortonormais, a partir de estruturas de filtros digitais, aparecem nos trabalhos de Fettweis (1970) e Fettweis (1971) e o uso das bases de Laguerre para identificação de sistemas também é apresentada em King e Paraskevopoulos (1977).

Com o avanço computacional nas décadas de 1980 e 1990, houve um aumento no interesse da comunidade científica pelo uso das funções ortonormais nas áreas de processamento

de sinais (PEREZ; TSUJII, 1991; BRINKER, 1994), controle baseado em identificação de sistemas (BODIN; WAHLBERG, 1994; GUNNARSSON; WAHLBERG, 1991; NURGES, 1987), modelagem (PARTINGTON, 1991), controle adaptativo (ZERVOS *et al.*, 1988), entre outras áreas (VAIDYANATHAN, 1985; WAHLBERG; HANNAN, 1993).

Entre o final do século XX e início do século XXI muitos trabalhos voltados para identificação de sistemas podem ser citados. Um livro bastante importante para a área é o de Heuberger *et al.* (2005). Outros trabalhos que apresentam relevância são Akçay e Heuberger (2001), no qual é apresentado um algoritmo de identificação iterativo no domínio da frequência. Já Chou *et al.* (1999) realiza a identificação de sistemas SISO (*Single Input Single Output*) com funções de Laguerre e Nalbantoğlu *et al.* (2003), em que é feita a identificação de sistemas com GOBFs.

Importantes trabalhos em português também foram produzidos nas décadas de 2000 e 2010, entre eles cabe citar: Campello *et al.* (2007), Campello e Oliveira (2007), Braga *et al.* (2011), Oroski e Bauchspiess (2014) e Oroski (2015), podendo ser citado com mais destaque o trabalho de Rosa (2005) no qual é apresentado um modelo analítico para encontrar o polo sub-ótimo das funções de Kautz, e o trabalho de Machado *et al.* (2011) que apresenta a identificação de sistemas não-lineares por GOBF com funções internas. Dessa forma, a Seção 2.1 irá abordar um detalhamento das Funções Ortonormais e suas aplicações.

2.1 MODELAGEM DE SISTEMAS COM FUNÇÕES ORTONORMAIS

O uso de famílias de funções ortonormais para a modelagem de sistemas é uma escolha conhecida na literatura, pois, apesar do uso de somas finitas de exponenciais para modelagem de um sistema linear e invariante no tempo ser uma estratégia que simplifica o trabalho de identificação, o emprego de filtros de resposta infinita ao impulso (IIR - *Infinite Impulse Response*) com bases de funções ortonormais podem produzir estruturas mais eficientes (HEUBERGER *et al.*, 2005). Estes filtros são parametrizados a partir de polos pré-especificados, possibilitando a incorporação de informações *a priori*. Essas estruturas são aplicadas, principalmente, em identificação de sistemas, porque a parametrização de modelos em termos da expansão infinita de coeficientes implica em modelos lineares nos parâmetros, permitindo o uso de técnicas de estimação lineares para determinar os parâmetros.

Segundo Strang (2010), as funções ortonormais são vetores em um espaço vetorial de funções com um produto interno nulo entre si, o qual é definido em Wahlberg (1999) como:

$$\langle \phi_n(z), \phi_m(z) \rangle = \frac{1}{2\pi i} \oint_C \phi_n(z) \bar{\phi}_m\left(\frac{1}{\bar{z}}\right) \frac{1}{z} dz, \quad (1)$$

em que $\phi_n(z)$ e $\phi_m(z)$ são funções ortonormais, C é o círculo de raio unitário, i é a unidade imaginária, $n, m \in \mathbb{N}$ e $z \in \mathbb{C}$. Em sua forma discreta, este produto interno pode ser representado por meio da equação (2), sendo $k \in \mathbb{N}$.

$$\langle \phi_n(k), \phi_m(k) \rangle = \sum_{k=0}^{\infty} \phi_n(k) \phi_m(k). \quad (2)$$

As funções ortonormais são quadraticamente somáveis, pertencendo ao espaço L^2 (CAMPELLO *et al.*, 2007), ou seja, são funções que atendem a condição expressa em (3).

$$\sum_{k=0}^{\infty} \phi(k)^2 \leq \infty. \quad (3)$$

Como consequência, dada uma função $h(k) : \mathbb{N} \rightarrow \mathbb{R}$ quadraticamente somável em $[0, \infty)$, haverá um número inteiro $\eta > 0$ que, para o erro $\epsilon > 0$ (CAMPELLO *et al.*, 2007; LEMMA *et al.*, 2010), tem-se:

$$\sum_{k=0}^{\infty} \left(h(k) - \sum_{n=1}^{\eta} \alpha_n \phi_n(k) \right)^2 < \epsilon, \quad (4)$$

em que $\{\phi_1(k), \dots, \phi_\eta(k)\}$ são as η funções ortonormais da base e $\alpha_1, \dots, \alpha_\eta \in \mathbb{R}$ são ganhos escalares. Assim, é possível mostrar que, para todo $k \in \mathbb{N}$, na equação (5) a função $\hat{h}(k)$ converge à função original, $h(k)$, dado um η infinito, ou seja:

$$\hat{h}(k) = \sum_{n=0}^{\eta} \alpha_n \phi_n(k), \forall k. \quad (5)$$

Para estimar os coeficientes α_n é necessário multiplicar a equação (5) por $\sum_{k=0}^{\eta} \phi_m(k)$:

$$\sum_{k=0}^{\eta} \hat{h}(k) \phi_m(k) = \sum_{n=0}^{\eta} \alpha_n \phi_n(k) \sum_{k=0}^{\eta} \phi_m(k). \quad (6)$$

Manipulando o lado direito da equação (6), tem-se:

$$\begin{aligned} \sum_{k=0}^{\eta} \hat{h}(k) \phi_m(k) &= \sum_{n=0}^{\eta} \alpha_n \sum_{k=0}^{\eta} \phi_n(k) \phi_m(k) \\ &= \sum_{n=0}^{\eta} \alpha_n \langle \phi_n(k), \phi_m(k) \rangle. \end{aligned} \quad (7)$$

Dado que as funções são ortonormais, o produto interno que surge na equação (7) não é nulo apenas quando $n = m$. Assim, os coeficientes são encontrados a partir de:

$$\alpha_m = \sum_{k=0}^{\eta} \widehat{h}(k) \phi_m(k), \quad (8)$$

ou ainda,

$$\alpha_m = \langle \widehat{h}(k), \phi_m(k) \rangle. \quad (9)$$

A representação de sistemas dinâmicos lineares com as OBFs é possível somente para sistemas BIBO (*Bound Input Bound Output*) estáveis, uma vez que sua resposta ao impulso é quadraticamente somável (CAMPELLO *et al.*, 2007). Para o caso não-linear, a modelagem com OBF exige um mapeamento estático não linear. Um dos operadores para este mapeamento é o modelo Volterra-OBF que será tratado mais adiante neste trabalho.

Há várias funções ortonormais, cada uma definida de acordo com produto interno empregado. As mais usadas são: funções de Laguerre, funções de Kautz e as GOBF (BELT, 1997). Nas seções a seguir haverá um detalhamento maior sobre cada uma delas.

2.2 FUNÇÕES ORTONORMAIS GENERALIZADAS

A família de bases de funções ortonormais GOBF são desenvolvidas por meio da conexão de n filtros passa-tudo de n -ésima ordem. Desta forma, os filtros terão n polos, podendo ser reais ou complexos (HEUBERGER *et al.*, 2005). Essa classe de base de funções também é conhecida como Takenaka-Malmquist e é expressa pela equação (10).

$$F_n(z) = \frac{\sqrt{1 - |\xi_n|^2}}{z - \xi_n} \prod_{m=1}^{n-1} \left[\frac{1 - \bar{\xi}_m z}{z - \xi_m} \right], \quad n = 1, 2, \dots, \quad |\xi_m| < 1, \quad (10)$$

no qual ξ_m e $\bar{\xi}_m \in \mathbb{C}$ são os polos da base e os conjugados complexo, respectivamente.

As funções $F_n(z)$ geralmente são complexas, o que é indesejável na maioria dos sistemas físicos reais por gerar uma resposta ao impulso complexa. No trabalho de Ninness e Gustafsson (1997) foi desenvolvido um método para contornar esta desvantagem. Este método consiste em fazer as $n - 1$ primeiras funções serem formadas a partir de polos reais, ou seja, $\xi_m \in \mathbb{R}$ para $m = 1, 2, \dots, n - 1$. Para inserir um polo complexo é necessário que duas novas funções sejam formadas a partir de uma combinação linear de $F_n(z)$ e $F_{n+1}(z)$. Assim, a base formada possui apenas funções com resposta ao impulso real e não perdem a ortonormalidade. Dada a

possibilidade da combinação de polos reais e imaginários, o que possibilita a representação de sistemas mais complexos, esta família de funções é indicada para a modelagem de sistemas de ordem superior a 2. Existem dois casos particulares das GOBFs, são eles: as funções de Laguerre e as funções de Kautz, ambas serão discutidas nas seções a seguir.

2.3 FUNÇÕES DE LAGUERRE

As funções de Laguerre podem ser entendidas como filtros de primeira ordem ou como uma particularidade da GOBF, em que $\xi_m = p, \forall m$, sendo p um polo real (LEMMA *et al.*, 2010). Estas funções são descritas por meio da equação (11).

$$F_n(z) = \frac{\sqrt{1-p^2}}{z-p} \left(\frac{1-pz}{z-p} \right)^{(n-1)}, \quad n = 1, 2, \dots, \quad p \in \mathbb{R}, \quad |p| < 1. \quad (11)$$

Assim, dado um conjunto de n funções de Laguerre, pode-se definir uma base ortonormal como:

$$\{F_1(z), F_2(z), \dots, F_n(z)\}. \quad (12)$$

No uso de base de funções de Laguerre para modelar sistemas lineares e invariantes no tempo, o operador é definido como a combinação linear das n saídas dos filtros de Laguerre, como é expresso na equação (13) (OROSKI, 2015).

$$\hat{y}(k) = \sum_{n=1}^{\eta} \alpha_n L_n(k), \quad (13)$$

sendo $\hat{y}(k)$ a saída estimada, α_n os ganhos escalares e $L_n(k)$ corresponde à entrada $u(k)$ filtrada pela n -ésima função de Laguerre. Contudo, para sistemas não-lineares, é necessário um mapeamento estático não-linear com as saídas do filtro para obter uma modelagem adequada (OROSKI; BAUCHSPIESS, 2014).

Um dos benefícios das funções de Laguerre é que, por serem parametrizadas por um único polo real, suas transformadas tendem a ser funções racionais com uma forma repetitiva simples. Outra vantagem é que, devido à influência do polo real, estas funções são adequadas para reproduzir sistemas com dinâmicas bem amortecidas, ou seja, sistemas com a parte imaginária do polo pouco significativa em relação a parte real ou um polo puramente real (MACHADO; AMARAL, 2010). Por outro lado, para a modelagem de sistemas que possuem uma dinâmica predominantemente subamortecida, é necessário um elevado número de funções de Laguerre para conseguir uma representação adequada (CAMPELLO *et al.*, 2007). Este comportamento é

consequência da distância entre o polo real das funções de Laguerre e os polos complexos desta categoria de sistemas (BRAGA *et al.*, 2011). Uma alternativa para contornar este problema é o uso das funções de Kautz para a modelagem. Na Seção 2.4 será realizada uma explanação mais detalhada sobre estas funções.

2.4 FUNÇÕES DE KAUTZ

As funções de Kautz são um caso particular das GOBF, em que ξ_m é um par de polos complexo conjugados (LEMMA *et al.*, 2010; WAHLBERG, 1994). Em Wahlberg (1994), as funções discretas de Kautz são definidas como em (14) e (15).

$$F_{2n}(z) = \frac{\sqrt{(1-c^2)(1-b^2)}}{z^2 + b(c-1)z - c} \left(\frac{-cz^2 + b(c-1)z + 1}{z^2 + b(c-1)z - c} \right)^{(n-1)}, \quad (14)$$

$$F_{2n-1}(z) = \frac{(z-b)\sqrt{1-c^2}}{z^2 + b(c-1)z - c} \left(\frac{-cz^2 + b(c-1)z + 1}{z^2 + b(c-1)z - c} \right)^{(n-1)}, \quad (15)$$

sendo que $F_{2n}(z)$ e $F_{2n-1}(z)$ são as funções de Kautz pares e ímpares, respectivamente, b e c correspondem aos parâmetros descritos em (16) e (17), sendo que p é o polo e \bar{p} é o seu complexo conjugado.

$$b = \frac{(p + \bar{p})}{1 + p\bar{p}}, \quad |b| < 1. \quad (16)$$

$$c = -p\bar{p}, \quad |c| < 1. \quad (17)$$

Além disso, para garantir que os polos sejam complexos conjugados, é preciso que:

$$b^2(1-c)^2 + 4c \leq 0, \quad (18)$$

e, como é exposto em Braga *et al.* (2011), para certificar a estabilidade do sistema é necessário que:

$$-1 < c < 0. \quad (19)$$

Dessa forma, dado um conjunto de n funções de Kautz, uma base ortonormal é definida como em (20).

$$\{F_1(z), F_2(z), \dots, F_n(z)\}. \quad (20)$$

Assim como nas funções de Laguerre, para o uso de funções de Kautz na modelagem de sistemas lineares e invariantes no tempo, o operador é definido como sendo a combinação linear das n saídas do filtro de Kautz, como expresso em (21) (OROSKI, 2015).

$$\hat{y}(k) = \sum_{n=1}^{\eta} \alpha_n w_n(k), \quad (21)$$

em que $\hat{y}(k)$ são as saídas estimadas, α_n são os ganhos escalares e $w_n(k)$ corresponde à entrada $u(k)$ filtrada pela n -ésima função de Kautz. Para sistemas não-lineares, o mapeamento não-linear das saídas será tratado nas seções seguintes deste trabalho.

Como descrito em Wahlberg e Mäkila (1996), as funções de Kautz apresentam uma vantagem sobre as funções de Laguerre na modelagem de sistemas subamortecidos, precisando de um menor número de funções para a representação. Isso ocorre pelo fato das funções de Kautz serem parametrizadas por um par de polos complexos conjugados, ou seja, a dinâmica oscilatória é inerente às funções. Outra vantagem é a possibilidade de ajustar o fator de amortecimento do modelo e a frequência de oscilação (BRAGA *et al.*, 2011). Desta forma, o estudo destas funções tem recebido a atenção de diversos trabalhos, como Wahlberg (1994), Rosa (2005), Braga *et al.* (2011), Machado *et al.* (2011) e Oroski (2015).

2.5 SÍNTESE

As funções ortonormais são vetores que compõem um espaço vetorial de funções com um produto interno nulo e norma unitária, sendo as principais as funções de Laguerre, as funções de Kautz e as GOBFs.

As GOBFs são funções ortonormais parametrizadas por polos reais e complexos e são mais indicadas para modelagem de sistemas com ordem superior a 2. As funções de Laguerre podem ser interpretadas como um caso particular das GOBFs para quando as funções são parametrizadas por um polo puramente real. Devido a essa característica, o uso de funções de Laguerre são mais apropriadas para identificação de sistemas superamortecidos. Por sua vez, as funções de Kautz também podem ser entendidas como um caso particular das GOBFs para quando as funções são parametrizadas por um par de polos complexos conjugados, o que dá a elas um comportamento subamortecido. Esse aspecto leva esse conjunto de funções a ser mais adequado para a modelagem de sistemas subamortecidos.

Em razão de suas propriedades, as funções ortonormais são empregadas recorrentemente na identificação de sistemas lineares. Entretanto, para a modelagem de sistemas não-lineares são necessárias estruturas mais robustas, uma delas é a Série de Volterra. No capítulo a seguir é abordada a fundamentação matemática para o uso destas estruturas na modelagem de sistemas dinâmicos não-lineares.

3 MODELO VOLTERRA-KAUTZ

Neste capítulo será apresentada a série de Volterra. Inicialmente, um breve levantamento histórico de suas aplicações será realizado. Em seguida, na Seção 3.1, será descrito o desenvolvimento matemático e as propriedades da série, assim como, sua aplicação em identificação de sistemas. Na sequência, a Seção 3.2 abordará uma estratégia de aproximação do *kernel* de Volterra com funções de Kautz para contornar o problema do elevado número de parâmetros necessários na modelagem de sistemas dinâmicos não-lineares.

O desenvolvimento da série de Volterra foi apresentado no trabalho do matemático italiano Vito Volterra (VOLTERRA, 1932). Contudo, as primeiras aplicações em análise de sistemas apareceram nos trabalhos de Wiener (1942) e Ikehara (1951). Com o uso da série de Volterra muitos fenômenos puderam ser modelados. Porém, até a década de 1990 as aplicações práticas destas séries caminharam lentamente devido às limitações de processamento computacional do período (CHENG *et al.*, 2017).

Atualmente a série de Volterra é aplicada em diversas análises e projetos de sistemas não-lineares, como sistemas polinomiais não-lineares (JING *et al.*, 2009), sistemas bilineares (CHATTERJEE, 2010) e sistemas com saturações (MASTROMAURO *et al.*, 2007). Outras aplicações são encontradas no trabalho de Mirri *et al.* (2003) que descrevem a modelagem de sistemas não-lineares e diversas variações da série de Volterra com um número de operadores reduzidos em relação à versão clássica. Em Chatterjee e Vyas (2003a) e Chatterjee e Vyas (2004) são desenvolvidos métodos recursivos para identificação de parâmetros não-lineares, em que estas técnicas são aplicadas à estimação de parâmetros de um rotor em Chatterjee e Vyas (2003b).

Aplicações da série de Volterra em sistemas de controle também são comuns. Em Gruber *et al.* (2012) foi projetado um controlador preditivo baseado em um modelo não-linear para controlar a taxa de fluxo de ar em uma célula a combustível. Já no trabalho de Yoon e Sun (2014) foi desenvolvido um método de controle robusto para rastreamento de um atuador de válvula para motores sem *came*¹. Todavia, as aplicações não estão restritas às áreas das Engenharias, podendo ser citada a identificação de sistemas biológicos realizados em Marmarelis (1993) e em Korenberg e Hunter (1996).

¹ Peça no formato excêntrico que possui movimento giratório que juntamente com uma peça denominada “seguidor” transforma o movimento circular em linear.

Como a série de Volterra normalmente demanda a estimação de muitos parâmetros, alguns trabalhos se dedicaram a expor abordagens para amenizar esse problema, como é o caso do trabalho de Chen *et al.* (2010) que faz uso das teorias do espaço de Hilbert para modelar o *kernel* de Volterra com o modelo de Wiener. Outro trabalho que utiliza desta mesma teoria é Braga *et al.* (2011), mas neste é feita a expansão do *kernel* com as OBF. Nas seções a seguir, serão apresentados o desenvolvimento matemático da série de Volterra e uma estratégia de redução do seu número de parâmetros.

3.1 SÉRIE DE VOLTERRA

O uso da série de Volterra para modelar sistemas não-lineares possibilita uma ideia do significado físico do processo (MORTENSEN, 1983). Desta forma, tem-se o benefício conceitual de realizar a descrição de um processo não-linear como uma generalização do caso linear, ou seja, as séries de Volterra possibilitam o uso de conceitos de sistemas lineares para identificar sistemas não-lineares (EYKHOFF, 1979).

A série de Volterra modela um sistema a partir da soma das respostas dos operadores de primeira ordem até o operador de ordem η . Estes operadores são descritos por uma função multivariável denominada *kernel*. Na equação (22) é possível visualizar a descrição de um sistema por uma série de Volterra.

$$\begin{aligned}
 y(k) = & \sum_{\tau_1=-\infty}^{+\infty} h_1(\tau_1)u(k - \tau_1) + \sum_{\tau_1=-\infty}^{+\infty} \sum_{\tau_2=-\infty}^{+\infty} h_2(\tau_1, \tau_2)u(k - \tau_1)u(k - \tau_2) \\
 & + \dots + \sum_{\tau_1=-\infty}^{+\infty} \dots \sum_{\tau_\eta=-\infty}^{+\infty} h_\eta(\tau_1, \dots, \tau_\eta)u(k - \tau_1) \dots u(k - \tau_\eta),
 \end{aligned} \tag{22}$$

no qual, $h_\eta(\tau_1, \dots, \tau_\eta)$ é chamado de *kernel* de ordem η .

A série de Volterra é uma generalização do modelo linear de convolução, sendo que o *kernel* de primeira ordem apresenta o mesmo sentido matemático que a resposta ao impulso para sistemas lineares. Já o *kernel* de η -ésima ordem corresponde ao grau da não-linearidade do sistema (ROSA, 2005), podendo ser interpretado como uma resposta ao impulso generalizada com dimensão η . Logo, um conjunto de η *kernels* pode ser expresso como:

$$\{h_1(\tau_1), h_2(\tau_1, \tau_2), \dots, h_\eta(\tau_1, \dots, \tau_\eta)\}. \tag{23}$$

Para a representação de sistemas reais e estáveis, os *kernels* de Volterra possuem as seguintes propriedades (EYKHOFF, 1979):

- $h_\eta(\tau_1, \dots, \tau_\eta)$ é, ou pode ser feito, simétrico;
- $\lim_{\tau_l \rightarrow \infty} h_\eta(\tau_1, \dots, \tau_\eta) = 0 \quad l = 1, 2, \dots, \eta;$
- $h_\eta(\tau_1, \dots, \tau_\eta) = 0 \quad \forall \tau_i < 0.$

A propriedade de simetria é interessante porque permite que um *kernel* assimétrico possa ser sintetizado de maneira a tornar-se simétrico sem perda de generalidade (SCHETZEN, 1980). A principal intenção por trás da simetrização de um *kernel* é a otimização do número de parâmetros a serem estimados, conseqüentemente, simplificando as análises computacionais. A simetrização de *kernels* assimétricos é descrita em Rugh (1981) e apresentada na equação (24).

$$h_\eta^{sim}(\tau_1, \dots, \tau_\eta) = \frac{1}{\eta!} \sum_{\pi(\cdot)} h_\eta(\tau_{\pi(1)}, \dots, \tau_{\pi(\eta)}), \quad (24)$$

em que $\pi(\cdot)$ representa todas as permutações dos termos τ_1 a τ_η . Assim, um *kernel* é simétrico se $h_2(\tau_1, \tau_2) = h_2(\tau_2, \tau_1)$.

Exemplo 1. A seguir, é desenvolvido um exemplo do processo de simetrização considerando um *kernel* de terceira ordem de um sistema dado por (25) (RUGH, 1981).

$$h_3(\tau_1, \tau_2, \tau_3) = g(\tau_1)g(\tau_2)g(\tau_3)f(\tau_1 + \tau_3). \quad (25)$$

Aplicando (24) em (25), tem-se que:

$$\begin{aligned} h_3^{sim} &= \frac{1}{3!} [h_3(\tau_1, \tau_2, \tau_3) + h_3(\tau_2, \tau_1, \tau_3) + h_3(\tau_3, \tau_2, \tau_1) + h_3(\tau_1, \tau_3, \tau_2) + \\ &\quad h_3(\tau_2, \tau_3, \tau_1) + h_3(\tau_3, \tau_1, \tau_2)] \\ &= \frac{1}{6} [2g(\tau_1)g(\tau_2)g(\tau_3)f(\tau_1 + \tau_3) + 2g(\tau_1)g(\tau_2)g(\tau_3)f(\tau_2 + \tau_3) + \\ &\quad 2g(\tau_1)g(\tau_2)g(\tau_3)f(\tau_1 + \tau_2)] \\ h_3^{sim} &= \frac{1}{3} g(\tau_1)g(\tau_2)g(\tau_3) [f(\tau_1 + \tau_2) + f(\tau_2 + \tau_3) + f(\tau_1 + \tau_3)]. \end{aligned} \quad (26)$$

□

Como pode ser observado em (26), a expressão do *kernel* simétrico é matematicamente mais complexa e, usualmente, é o que ocorre (RUGH, 1981). Contudo, a simetrização se justifica porque, em vários casos, as propriedades de um sistema podem estar mais relacionadas com as

propriedades do *kernel* simétrico do que com as propriedades do *kernel* assimétrico (BRAGA *et al.*, 2011; ROSA, 2005). Ademais, permite uma diminuição significativamente nos parâmetros a serem estimados e a representação na forma simétrica é única, enquanto que a forma assimétrica do *kernel* não o é (RUGH, 1981).

A segunda propriedade é decorrente da identificação de sistemas estáveis (EYKHOFF, 1979), em que a resposta à entrada vai se atenuando com o passar do tempo; estes sistemas com esta característica são conhecidos como *fading memory*. Segundo Braga *et al.* (2011), esta propriedade permite truncar a série em um termo de ordem M e considerar a contribuição do *kernel* desprezível a partir de um certo escalar $N > M$.

Como resultado dessa propriedade, o truncamento da série em *kernels* que forneça uma representação satisfatória é muito empregada, pois, este recurso diminui o esforço computacional das análises, permitindo ainda uma boa representação para a maioria dos operadores não-lineares. Desta forma, considerando um sistema causal e uma entrada $u(k)$, a resposta do sistema pode ser aproximada por (27).

$$\hat{y}(k) \cong \sum_{\eta=1}^M \sum_{\tau_1=0}^N \cdots \sum_{\tau_\eta=0}^N h_\eta(\tau_1, \dots, \tau_\eta) \prod_{m=1}^{\eta} u(k - \tau_m). \quad (27)$$

A última propriedade é válida para sistemas causais e com condições iniciais nulas, nos quais a resposta do sistema depende apenas dos sinais do passado do sinal de entrada (SCHETZEN, 1980). Devido à série de Volterra ser uma soma infinita, é necessário que haja condições de convergência, e estas devem satisfazer a garantia de que a representação da série permaneça com um sentido físico. Na maioria dos casos estas condições levam à requisição de uma entrada $u(k)$ limitada (ROSA, 2005).

Apesar dos modelos de Volterra poderem ser utilizados na identificação de sistemas não-lineares, existe uma grande dificuldade na estimação dos *kernels*, por serem funções multidimensionais não parametrizadas, nas quais sua estimativa é possível apenas com a individualização da sua contribuição na resposta total do sistema (SCHETZEN, 1980; RUGH, 1981). Uma técnica para contornar este problema é generalizar as abordagens dos modelos FIR, considerando cada componente do *kernel* como uma variável a ser determinada (AGUIRRE, 2007). Todavia, ainda para este caso, o número de parâmetros necessários para ter uma estimação satisfatória são elevados. Sendo assim, uma abordagem de particular interesse é apresentada na Seção 3.2 a seguir.

3.2 MODELO VOLTERRA-KAUTZ

Como exposto na Seção 3.1, é possível modelar sistemas não-lineares com as séries de Volterra. Contudo, é necessário um elevado número de termos para representar os seus *kernels*. Uma alternativa para contornar este problema é utilizar uma base de funções ortonormais para modelar tais *kernels* (ROSA, 2005). O uso das funções ortonormais possibilita também a incorporação do conhecimento prévio da dinâmica do sistema (CAMPELLO *et al.*, 2007), como já exposto no Capítulo 2. Nesta seção será apresentado o desenvolvimento matemático para que seja possível fazer esta representação dos *kernels* considerando as funções de Kautz.

No primeiro momento, assume-se uma entrada $u(k - \tau)$, $\tau \in \mathbb{N}$ e seus regressores como sendo quadraticamente somáveis. Também, supõe-se que um conjunto de funções de Kautz $\{\phi_1(\tau), \phi_2(\tau), \dots, \phi_n(\tau)\}$, $n \in \mathbb{N}$, formam uma base ortonormal. Assim, um sinal de entrada pode ser descrito por:

$$u(k - \tau) = \sum_{n=1}^{\infty} u_n(k) \phi_n(\tau). \quad (28)$$

Como decorrência da ortogonalidade das funções de Kautz, tem-se que:

$$u_n(k) = \sum_{\tau=0}^{\infty} u(k - \tau) \phi_n(\tau). \quad (29)$$

Substituindo (29) em (22), obtém-se:

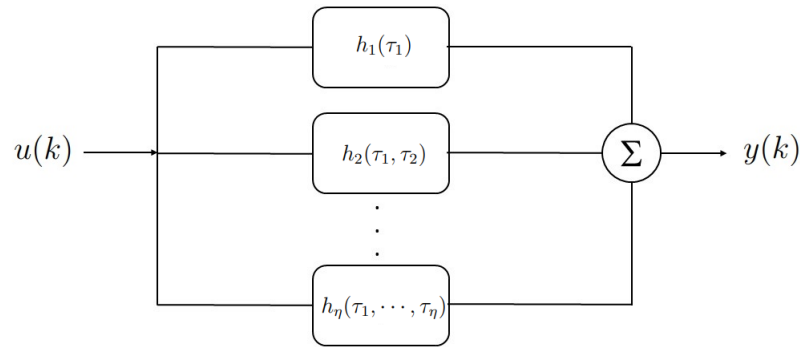
$$\begin{aligned} y(k) &= \sum_{n=1}^{\infty} \sum_{\tau=0}^{\infty} \alpha_n u(k - \tau) \phi_n(\tau) + \\ &+ \sum_{n_1=1}^{\infty} \sum_{n_2=1}^{\infty} \sum_{\tau_1=0}^{\infty} \sum_{\tau_2=0}^{\infty} \alpha_{n_1, n_2} u(k - \tau_1) \phi_{n_1}(\tau_1) u(k - \tau_2) \phi_{n_2}(\tau_2) + \dots \end{aligned} \quad (30)$$

A partir de (30), pode-se representar os *kernels* da seguinte forma:

$$\begin{aligned} h_1(\tau_1) &= \sum_{n=1}^{\infty} \sum_{\tau=0}^{\infty} u_n(k - \tau) \phi_n(\tau) \\ h_2(\tau_1, \tau_2) &= \sum_{n_1=1}^{\infty} \sum_{n_2=1}^{\infty} \sum_{\tau_1=0}^{\infty} \sum_{\tau_2=0}^{\infty} \alpha_{n_1, n_2} u(k - \tau_1) \phi_{n_1}(\tau_1) u(k - \tau_2) \phi_{n_2}(\tau_2) \\ &\vdots \\ h_\eta(\tau_1, \dots, \tau_\eta) &= \sum_{n_1=1}^{\infty} \dots \sum_{n_\eta=1}^{\infty} \alpha_{n_1, \dots, n_\eta} \prod_{m=1}^{\eta} \phi_{n_m}(\tau_m) \end{aligned} \quad (31)$$

A Figura 1 ilustra a estrutura do modelo de Volterra com os *kernels* desenvolvidos com as funções ortonormais de Kautz.

Figura 1 – Estrutura do modelo de Volterra-Kautz.



Fonte: Adaptado de Rosa (2005).

A modelagem dos *kernel* com OBF tem a vantagem de utilizar menos termos para atingir a mesma precisão (ROSA, 2005). Nos sistemas com *fading memory* as somas infinitas dos *kernels* podem ser truncadas, como já mencionado nas propriedades da série de Volterra. O ponto em que o sistema será truncado depende do sistema que está sendo modelado. Desta forma, irá existir uma troca entre a qualidade da representação do modelo e a quantidade de parâmetros a serem estimados.

Assim sendo, o número de funções da base é limitada em um certo N para cada *kernel*. Com isto, o *kernel* de ordem η da equação (31) pode ser reescrito como:

$$\hat{h}_\eta(\tau_1, \dots, \tau_\eta) \cong \sum_{n_1=1}^N \cdots \sum_{n_\eta=1}^N \alpha_{n_1, \dots, n_\eta} \prod_{m=1}^{\eta} \phi_{n_m}(\tau_m). \quad (32)$$

Este truncamento limita o número de funções para expansão do *kernel*. Ademais, é necessário realizar o truncamento da ordem do *kernel* em um número finito M , o que também é possível em decorrência da segunda propriedade da série de Volterra. A única exigência necessária é que a entrada $u(k)$ seja limitada (FU; DUMONT, 1993). Considerando um sistema causal e a aproximação do *kernel* descrita em (32), a equação (30) pode ser reformulada como:

$$\hat{y}(k) \cong \sum_{\eta=1}^M \left[\sum_{n_1=0}^N \cdots \sum_{n_\eta=0}^N \alpha_{n_1, \dots, n_\eta} \prod_{m=1}^{\eta} \left(\sum_{\tau_m=0}^k \phi_{n_m}(\tau_m) u(k - \tau_m) \right) \right]. \quad (33)$$

Com os ajustes matemáticos realizados, pode-se notar que o modelo de Volterra, que é aproximado em (33), não deixou de ser linear nos parâmetros e o desafio da estimação dos *kernels* torna-se um problema de estimação linear destes parâmetros, podendo ser realizada por meio dos algoritmos clássicos, como o método dos mínimos quadrados.

A limitação mais crítica nos modelos Volterra-Kautz advém do truncamento da ordem do modelo (BRAGA *et al.*, 2011). Na literatura, os modelos com esta abordagem são limitados a *kernels* de segunda ordem ($M = 2$), independente da aplicação (FU; DUMONT, 1993; ROSA, 2005; CAMPELLO; OLIVEIRA, 2007; BRAGA *et al.*, 2011). Embora o truncamento implique na redução da quantidade de parâmetros estimados, este artifício acarreta na limitação da capacidade de representação do modelo.

Dessa forma, propõe-se um sistema dinâmico, causal, subamortecido e não-linear com uma entrada $u(k) \in \mathbb{R}$ e a saída $y(k) \in \mathbb{R}$. Para realizar a modelagem desse sistema com um modelo de Volterra-Kautz serão utilizados apenas os dois primeiros *kernels* simétricos. Assim, o sistema pode ser aproximado por (34):

$$\hat{y}(k) \cong \sum_{n=1}^N \alpha_n u(k - \tau_n) \phi_n(\tau_n) + \sum_{n_1=1}^N \sum_{n_2=n_1}^N \alpha_{n_1, n_2} u(k - \tau_{n_1}) \phi_{n_1}(\tau_{n_1}) u(k - \tau_{n_2}) \phi_{n_2}(\tau_{n_2}). \quad (34)$$

Considerando que o sistema possui um atraso unitário:

$$\tau_n = \tau_{n_1} = \tau_{n_2} = \tau, \quad (35)$$

e

$$u(k - \tau_n) \phi_n(\tau_n) = w_n(k). \quad (36)$$

Logo, a equação (34) pode ser reformulada como:

$$\hat{\mathbf{y}}(k) \cong \boldsymbol{\gamma}^\top \boldsymbol{\psi}(k) + \boldsymbol{\psi}^\top(k) \boldsymbol{\Gamma} \boldsymbol{\psi}(k), \quad (37)$$

sendo que,

$$\boldsymbol{\gamma}^\top = [\alpha_1 \quad \alpha_2 \quad \dots \quad \alpha_N], \boldsymbol{\gamma}^\top \in \mathbb{R}^{1 \times N}. \quad (38)$$

$$\boldsymbol{\Gamma} = \begin{bmatrix} \alpha_{1,1} & 0 & 0 & \dots & 0 \\ \alpha_{2,1} & \alpha_{2,2} & 0 & \dots & 0 \\ \vdots & & & & \\ \alpha_{N,1} & \alpha_{N,2} & \alpha_{N,3} & \dots & \alpha_{N,N} \end{bmatrix}, \boldsymbol{\Gamma} \in \mathbb{R}^{N \times N}. \quad (39)$$

$$\boldsymbol{\psi}(k) = \begin{bmatrix} w_1(k) \\ \vdots \\ w_N(k) \\ w_{1,1}(k) \\ w_{2,1}(k) \\ w_{2,2}(k) \\ \vdots \\ w_{N,1}(k) \\ \vdots \\ w_{N,N}(k) \end{bmatrix}^\top, \boldsymbol{\psi}(k) \in \mathbb{R}^{N \times \left(N + \frac{N(N-1)}{2}\right)}. \quad (40)$$

A equação (37) pode ser reescrita para ser linear nos parâmetros, como pode ser visto em (41):

$$\widehat{\mathbf{y}}(k) \cong \boldsymbol{\theta}^\top \boldsymbol{\psi}^\top(k), \quad (41)$$

em que,

$$\boldsymbol{\theta} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \\ \alpha_{1,1} \\ \alpha_{2,1} \\ \alpha_{2,2} \\ \vdots \\ \alpha_{N,1} \\ \vdots \\ \alpha_{N,N} \end{bmatrix}^\top, \boldsymbol{\theta} \in \mathbb{R}^{\left(N + \frac{N(N-1)}{2}\right) \times 1}. \quad (42)$$

Portanto, os parâmetros escalares α_n e α_{n_1, n_2} podem ser encontrados por intermédio do uso do Método dos Mínimos Quadrados.

3.3 SÍNTESE

A série de Volterra pode ser entendida como uma generalização do modelo linear de convolução, tal característica permite que conceitos de identificação de sistemas lineares possam ser estendidos para a modelagem de sistemas não-lineares. Na série de Volterra, a representação de um sistema é feita a partir da soma de funções multivariáveis chamadas de *kernels*. Apesar de

sua capacidade de representação, a estimação destes *kernels* requer a determinação de centenas de parâmetros.

Para amenizar o problema da dimensionalidade dos parâmetros, é necessário lançar mão de alguns recursos matemáticos. O primeiro deles é o uso da propriedade de simetria dos *kernels* que, apesar de gerar uma representação mais complexa, possibilita uma diminuição na quantidade de parâmetros a serem estimados. Outro recurso que pode ser empregado é a modelagem do *kernel* com funções ortonormais. A vantagem do uso destas funções é o emprego do conhecimento *a priori* do sistema, pois, com isso, é possível escolher qual função que apresenta um comportamento mais adequado para representar o sistema. Desta forma, o número de parâmetros a serem estimados é reduzido. As técnicas para a determinação dos parâmetros dos modelos constituem o assunto do Capítulo 4.

4 ESTIMAÇÃO DE PARÂMETROS

Neste capítulo serão apresentadas algumas técnicas para estimação dos parâmetros das funções ortonormais de Kautz. A primeira delas será apresentada na Seção 4.1, que abordará a clássica estimação realizada pelo Método dos Mínimos Quadrados (MMQ). Como uma progressão natural dos estudos, a Seção 4.2 apresentará o método de Levenberg-Marquardt (LM), que aborda o problema dos mínimos quadrados nos contextos não-lineares. Em seguida, a Seção 4.3 abordará a Otimização Bayesiana (OB) como uma forma alternativa aos métodos determinísticos apresentados nas seções anteriores. Em ambas as seções será realizado um breve levantamento bibliográfico sobre os assuntos tratados.

4.1 MÍNIMOS QUADRADOS

O desenvolvimento do MMQ foi consequência das buscas entre a combinação das observações de experimentos astronômicos e a precisão das medidas coletadas (CUNHA, 2014). Segundo Marinelli *et al.* (2002), o primeiro problema do gênero que se tem registro é de 1740, em que Jacques Cassini listou dados astronômicos coletados desde 140 a.C.. Neste levantamento, os resultados de um mesmo evento apresentavam divergências significativas. Então, o problema levantado era como elaborar uma curva para satisfazer os dados.

Nesta busca para elaborar uma curva satisfatória, inúmeras soluções foram apresentadas. A solução que foi considerada a mais eficaz foi o MMQ, aparecendo pela primeira vez em 1805 na obra de L'égrende intitulada “*Nouvelles Méthodes pour La Détermination des Orbites des Comète*”. Entretanto, em 1809, Friderico Gauss em sua obra “*Theoria Motus Corporum Coelestium*” também apresentava o MMQ, citando a sua aplicação desde 1795, sendo hoje a prioridade da criação do método a ele atribuída. Mais tarde, Gauss desenvolveu a generalização do método, fundamentando melhor as bases teóricas e um modelo funcional para minimização dos erros.

O tipo de problema em que o MMQ busca resolver tem como base a equação (43).

$$Ax = B, \quad (43)$$

em que $A \in \mathbb{R}^{n \times m}$ é a matriz com os sinais de entrada do sistema, $x \in \mathbb{R}^{m \times 1}$ é o vetor de parâmetros do sistema e $B \in \mathbb{R}^{n \times 1}$ é a matriz com os sinais de saída.

No caso em que $m = n$, \mathbf{x} pode ser calculado a partir da multiplicação à esquerda pela inversa de \mathbf{A} na equação (43), resultando na equação (44).

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{B}, \quad (44)$$

em que \mathbf{A}^{-1} é a matriz inversa de \mathbf{A} .

Porém, se $m < n$, a matriz \mathbf{A} não é uma matriz quadrada e o sistema da equação (43) é impossível. Uma técnica para contornar o problema é multiplicar à esquerda a equação (43) pela matriz transposta de \mathbf{A} , ou seja:

$$\mathbf{A}^\top \mathbf{A} \mathbf{x} = \mathbf{A}^\top \mathbf{B}. \quad (45)$$

A multiplicação $\mathbf{A}^\top \mathbf{A}$ resulta em uma matriz quadrada de dimensão m , e para o caso de \mathbf{A} ter posto completo, esta nova matriz é inversível (STRANG, 2010). Desta forma, pode-se estimar os valores de \mathbf{x} da mesma forma que foi feito em (44), obtendo o resultado expresso em (46).

$$\hat{\mathbf{x}} = [\mathbf{A}^\top \mathbf{A}]^{-1} \mathbf{A}^\top \mathbf{B}, \quad (46)$$

no qual $\hat{\mathbf{x}}$ é o vetor de parâmetros estimados.

Este é o método conhecido como estimador dos mínimos quadrados desenvolvido por Gauss (STRANG, 2010).

Para verificar se o estimador dos mínimos quadrados é o que minimiza o erro da estimação, será considerado um processo contendo uma entrada $u(k)$, uma saída $y(k)$, uma perturbação $r(k)$ e podendo ser representado pela seguinte equação a diferenças:

$$A(q^{-1})y(k) = B(q^{-1})u(k) + r(k), \quad (47)$$

em que,

$$A(q^{-1}) = 1 + a_1q^{-1} + \dots + a_{na}q^{-na},$$

$$B(q^{-1}) = b_0 + b_1q^{-1} + \dots + b_{nb}q^{-nb},$$

e q^{-1} é o operador de atraso.

A partir disso, define-se um vetor de medidas, $\boldsymbol{\psi}(k)$, e um vetor de parâmetros, $\boldsymbol{\theta}(k)$, como mostrado nas equações (48) e (49), respectivamente.

$$\boldsymbol{\psi}(k) = \begin{bmatrix} -y(k-1) & -y(k-2) & \dots & -y(k-na) & u(k) & \dots & u(k-nb) \end{bmatrix}. \quad (48)$$

$$\boldsymbol{\theta}^\top(k) = \begin{bmatrix} a_1 & a_2 & \dots & a_{na} & b_0 & b_1 & \dots & b_{nb} \end{bmatrix}. \quad (49)$$

Logo, a saída do sistema pode ser obtida a partir da equação (50).

$$y(k) = \boldsymbol{\psi}(k)\boldsymbol{\theta}(k) + r(k). \quad (50)$$

Considerando um vetor $\hat{\boldsymbol{\theta}}$ cujos elementos são os parâmetros estimados e \mathbf{y} um vetor contendo as saídas $y(k)$, o vetor $\hat{\mathbf{y}}$ contendo as saídas estimadas é dado por:

$$\hat{\mathbf{y}} = \boldsymbol{\psi}\hat{\boldsymbol{\theta}}, \quad (51)$$

em que o erro da estimativa ou resíduos é dado por (52).

$$\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \boldsymbol{\psi}\hat{\boldsymbol{\theta}}. \quad (52)$$

A fim de minimizar o erro da estimação de $\hat{\boldsymbol{\theta}}$, define-se a função J e se faz o uso do Teorema 1.

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) = \arg \min_{\boldsymbol{\theta}} \|\mathbf{y} - \boldsymbol{\psi}\boldsymbol{\theta}\|^2 \quad (53)$$

Teorema 1. *Considere um conjunto $D \subseteq \mathbb{R}^n$, uma função $f : D \rightarrow \mathbb{R}$ e os pontos $x_n \in D$. Supondo que as derivadas parciais $\frac{\partial f(x_1, x_2, \dots, x_n)}{\partial x_i}$, $i = 1, 2, \dots, n$ existam. Então, $\nabla f(x) = 0$ se x_n é um ponto de máximo ou mínimo de f .*

Aplicando o Teorema 1 em (53), tem-se:

$$\begin{aligned} \frac{\partial J}{\partial \boldsymbol{\theta}} &= \frac{\partial [(\mathbf{y} - \boldsymbol{\psi}\boldsymbol{\theta})^\top (\mathbf{y} - \boldsymbol{\psi}\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} \\ &= \mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \boldsymbol{\psi}\boldsymbol{\theta} - \boldsymbol{\psi}^\top \boldsymbol{\theta}^\top \mathbf{Y} + \boldsymbol{\psi}^\top \boldsymbol{\theta}^\top \boldsymbol{\psi}\boldsymbol{\theta} \\ &= -\mathbf{Y}^\top \boldsymbol{\psi} - \boldsymbol{\psi}^\top \mathbf{Y} + \boldsymbol{\psi}^\top \boldsymbol{\psi}\boldsymbol{\theta} + \boldsymbol{\psi}^\top \boldsymbol{\theta}\boldsymbol{\psi}. \end{aligned} \quad (54)$$

Fazendo $\left. \frac{\partial J}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0$, tem-se:

$$\hat{\boldsymbol{\theta}} = [\boldsymbol{\psi}^\top \boldsymbol{\psi}]^{-1} \boldsymbol{\psi}^\top \mathbf{y}. \quad (55)$$

Com esta análise, pode-se notar que a equação (55) é a mesma igualdade encontrada em (46). Isto leva a conclusão que (46) minimiza a função (53). Contudo, este método é empregado para os casos lineares, uma abordagem para os modelos não-lineares é apresentado na Seção 4.2.

4.2 LEVENBERG-MARQUARDT

A primeira versão do conhecido algoritmo de LM foi apresentada no trabalho de Levenberg (1944). Neste período, Levenberg fez uso das vantagens dos métodos do Gradiente Descendente e do Gauss-Newton para formular uma proposta para a minimização dos resíduos nos problemas de mínimos quadrados não-lineares. Para melhorar o desempenho do fator de regularização e a representação da matriz hessiana, Marquardt desenvolveu em seu trabalho (MARQUARDT, 1963) as bases matemáticas que fundamentaram a proposta do que ficou conhecida como a forma clássica do método de LM.

O método de LM apresenta um desempenho bem interessante para problemas não-lineares, desta forma, existe uma vasta literatura sobre o assunto (MORÉ, 1978; MORÉ, 1983; YUAN, 1994; YAMASHITA; FUKUSHIMA, 2001; FAN, 2003; FAN; YUAN, 2005; MA *et al.*, 2007; MA; JIANG, 2007; MA; TANG, 2008). Apesar do desempenho do método, o custoso esforço computacional exigido levou ao desenvolvimento de diferentes trabalhos que buscam técnicas para acelerar o algoritmo e analisar a convergência do método para diferentes cenários.

No trabalho de Coleman e Li (1996) é apresentada uma abordagem para controlar a dimensão do passo dentro da região de confiança, a fim de melhorar o desempenho da busca. Em Chan e Szeto (1999) é utilizada uma decomposição de matrizes juntamente com a inclusão de um peso de atualização para o cálculo da matriz inversa. Com isto, ele conseguiu melhorar o desempenho e a estabilidade do passo de busca. Já em Wilamowski *et al.* (1999) é abordada uma estratégia de reformulação da função de avaliação e, a partir disto, foi possível minimizar os fatores a serem calculados na matriz inversa.

Das obras que trabalharam com a análise de convergência do método pode-se citar Yamashita e Fukushima (2001), em que é demonstrado que, para mínimos locais, o algoritmo de LM apresenta uma convergência quadrática. Nesta mesma linha de estudos, Dan *et al.* (2002) apresenta as propriedades da convergência do método e, a partir dos resultados de Yamashita, Fan e Yuan (2005), mostram em seu trabalho que a performance de convergência quadrática permanece para problemas com singularidades. Desta forma, a Seção 4.2.1 apresentará o desenvolvimento clássico do método de LM e algumas análises sobre o método serão feitas.

4.2.1 Método de Levenberg-Marquardt

O método de LM trata o problema de minimização dos mínimos quadrados não-lineares (RANGANATHAN, 2004). Desta forma, a função de avaliação, ou função objetivo, pode ter o seguinte formato:

$$f(x) = \frac{1}{2} \sum_{n=1}^N r_n^2(x), \quad (56)$$

em que $r_n(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ é definido por:

$$r(x) = \begin{bmatrix} r_1(x) & r_2(x) & \cdots & r_N(x) \end{bmatrix}. \quad (57)$$

As derivadas de primeira ordem de $f(x)$ são expressas na matriz jacobiana, exibida em (58), e as derivadas de segunda ordem são obtidas pela matriz hessiana, apresentada em (59).

$$\mathbf{J}_{f(x)} = \nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} & \frac{\partial f(x)}{\partial x_2} & \cdots & \frac{\partial f(x)}{\partial x_n} \end{bmatrix}^T, \quad (58)$$

$$\mathbf{H}_{f(x)} = \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}, \quad (59)$$

em que x_1, x_2, \dots, x_n são elementos do vetor de parâmetros $x \in \mathbb{R}^n$.

Para realizar a minimização da função é necessário fazer a expansão da função via série de Taylor, de acordo com o Teorema 2.

Teorema 2. *Seja $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ continuamente diferenciável e $\delta \in \mathbb{R}^n$, então:*

$$f(x + \delta) = f(x) + \nabla f(x + \alpha\delta)^T \delta, \quad (60)$$

em que $\alpha \in [0, 1]$.

Como $f(x)$ é continuamente diferenciável, aplica-se a derivada em relação a x , obtendo:

$$\nabla f(x + \delta) = \nabla f(x) + \int_0^1 \nabla^2 f(x + \alpha\delta) \delta d\alpha. \quad (61)$$

Fazendo o truncamento da série até o termo de segunda ordem, tem-se:

$$f(x + \delta) \cong f(x) + \nabla f(x)^T \delta + \frac{1}{2} \delta^T \nabla^2 f(x + \alpha\delta) \delta. \quad (62)$$

Desta forma, fazendo a expansão de (56) em série de Taylor, tem-se:

$$f(x + \delta) \cong f(x) + \nabla f(x)^\top \delta + \frac{1}{2} \delta^\top \nabla^2 f(x) \delta. \quad \delta \in \mathbb{R}^n. \quad (63)$$

Fazendo uso do Teorema 1 em (63), obtém-se:

$$\begin{aligned} \frac{\partial f(x + \delta)}{\partial \delta} &= \nabla f(x) + \nabla^2 f(x) \delta = 0 \\ \delta &= - [\nabla^2 f(x)]^{-1} \nabla f(x). \end{aligned} \quad (64)$$

O resultado encontrado em (64) é o conhecido método de Newton, a vantagem que ele apresenta é a rápida convergência. Contudo, esta convergência é fortemente dependente do ponto de partida e a inversão da matriz hessiana pode tornar-se um problema, devido ao mal condicionamento numérico (ISERMANN; MÜNCHHOF, 2011).

Uma estratégia para contornar este problema é fazer uma simplificação na matriz hessiana. Para isto aplica-se a derivada primeira em relação a x na equação (56), obtendo a equação (65).

$$\nabla f(x) = \mathbf{J}_{r(x)} r(x). \quad (65)$$

Derivando (65) novamente em relação a x , obtém-se a matriz hessiana, dada em (66).

$$\nabla^2 f(x) = \mathbf{J}_{r(x)} \mathbf{J}_{r(x)}^\top + r(x) \mathbf{H}_{r(x)}. \quad (66)$$

Sob a hipótese dos elementos de $r(x)$ serem muito pequenos, a segunda parcela do segundo membro da equação (66) pode ser ignorada (ISERMANN; MÜNCHHOF, 2011). Assim, a matriz hessiana é aproximada por:

$$\nabla^2 f(x) \cong \mathbf{J}_{r(x)} \mathbf{J}_{r(x)}^\top. \quad (67)$$

Este resultado leva a uma simplificação do cálculo da equação (64), ficando como visto na equação (68).

$$\delta \cong - \left[\mathbf{J}_{r(x)} \mathbf{J}_{r(x)}^\top \right]^{-1} \mathbf{J}_{r(x)} r(x). \quad (68)$$

A equação (68) é conhecida como método de Gauss-Newton, no qual o cálculo do gradiente como um todo é mais simples que a obtenção das derivadas parciais de primeira ordem. Segundo Wright (1999), o desempenho do método é similar ao método de Newton. Uma das desvantagens deste método se dá quando os resíduos $r(x)$ não são muitos pequenos, levando a uma convergência mais lenta. Também o método pode não convergir caso δ seja ortogonal a $\nabla f(x)$ ou a matriz hessiana pode ser indefinida e direcionar para um ponto de maximização local (MEDEIROS *et al.*, 2008).

Aproveitando dos desenvolvimentos dos métodos anteriores, Levenberg (1944) propôs uma estratégia para garantir o decréscimo da função em (68). Para isto, ele considerou o acréscimo de um termo escalar de regularização $\lambda \geq 0$. Desta forma, (68) é reescrita como:

$$\delta \cong - \left[\mathbf{J}_{r(x)} \mathbf{J}_{r(x)}^\top + \lambda \mathbf{I} \right]^{-1} \mathbf{J}_{r(x)} r(x), \quad (69)$$

em que \mathbf{I} é a matriz identidade de dimensão adequada.

O fator λ é atualizado em cada nova iteração. Caso o erro da aproximação diminua, λ é reduzido, diminuindo a influência do gradiente e (69) tende ao método de Gauss-Newton. Porém, se λ é muito grande, o gradiente perde representatividade. Uma melhoria para isto é apresentada no trabalho de Marquardt (1963), no qual a matriz identidade \mathbf{I} é substituída pela diagonal do gradiente $\mathbf{J}_{r(x)} \mathbf{J}_{r(x)}^\top$, como pode ser observado na equação (70), chegando assim, no método de Levenberg-Marquardt.

$$\delta \cong - \left\{ \mathbf{J}_{r(x)} \mathbf{J}_{r(x)}^\top + \lambda \text{diag} \left[\mathbf{J}_{r(x)} \mathbf{J}_{r(x)}^\top \right] \right\}^{-1} \mathbf{J}_{r(x)} r(x). \quad (70)$$

A partir de então, dois teoremas são propostos e provados no trabalho de Marquardt (1963).

Teorema 3. *Seja $\delta(\lambda)$ a solução de (70) para um certo λ . Então $\|\delta(\lambda)\|^2$ é uma função contínua decrescente de λ , sendo que $\|\delta(\lambda)\| \rightarrow 0$ se, e somente se, $\lambda \rightarrow +\infty$.*

Teorema 4. *Defindo o ângulo entre $\delta(0)$ e $\nabla f(x)$ como ς . Então, ς é uma função contínua monótona decrescente de λ , sendo que $\varsigma \rightarrow 0$ se, e somente se, $\lambda \rightarrow +\infty$. Já δ rotaciona para a direção de $\nabla f(x)$ quando $\lambda \rightarrow +\infty$.*

O Teorema 3 implica que a medida que o método de LM minimiza a função (56), o parâmetro λ deve ser diminuído, forçando método a se comportar como o método de Gauss-Newton. Para o caso em que não ocorre uma minimização significativa, o Teorema 4 indica que o aumento de λ força a direção de δ para a direção de máxima descida (MEDEIROS *et al.*, 2008). Em outras palavras, neste método o gradiente move-se em passos largos em direções quando a curvatura do erro é pequena e em pequenos passos quando a curvatura for acentuada.

O modelo do método de LM é descrito no pseudo-código do Algoritmo 1. Embora este método funcione bem na prática, o custo computacional da matriz inversa pode tornar-se inviável para milhares de parâmetros (RANGANATHAN, 2004).

Algoritmo 1 – Método de LM.

Definem-se os parâmetros iniciais;

enquanto critério de parada não atingido **faça**

$$x_{n+1} = x_n - \left\{ \mathbf{J}_{r(x)} \mathbf{J}_{r(x)}^\top + \lambda \text{diag} \left[\mathbf{J}_{r(x)} \mathbf{J}_{r(x)}^\top \right] \right\}^{-1} \mathbf{J}_{r(x)} r(x)$$

Calcula-se o erro ϵ para o novo vetor de parâmetros

se ϵ aumentou **então**

$$x_{n+1} = x_{n-1}$$

$$\lambda = 10 \times \lambda$$

senão,

$$\lambda = \frac{\lambda}{10}$$

finaliza se

finaliza enquanto

Fonte: Autoria própria.

4.2.2 Exemplo Ilustrativo

O algoritmo do LM pode ser aplicado em diversos contextos de otimização. No exemplo a seguir será realizada a estimação dos parâmetros de uma função por meio da redução do Erro Quadrático Médio (MSE - *Mean Squared Error*) com o LM. Inicialmente considera-se a seguinte função:

$$y(t) = 2x_1 e^{\left(\frac{-t}{x_2}\right)} - x_3 \cos\left(\frac{t}{x_4}\right). \quad (71)$$

A partir da função expressa em (71) é simulada a resposta da mesma no intervalo de $t = [0, 100]$ segundos, $x_1 = 12,0$; $x_2 = 37,0$; $x_3 = 0,600$; $x_4 = -2,00$ e foram coletadas 100 amostras, ilustradas na Figura 2.

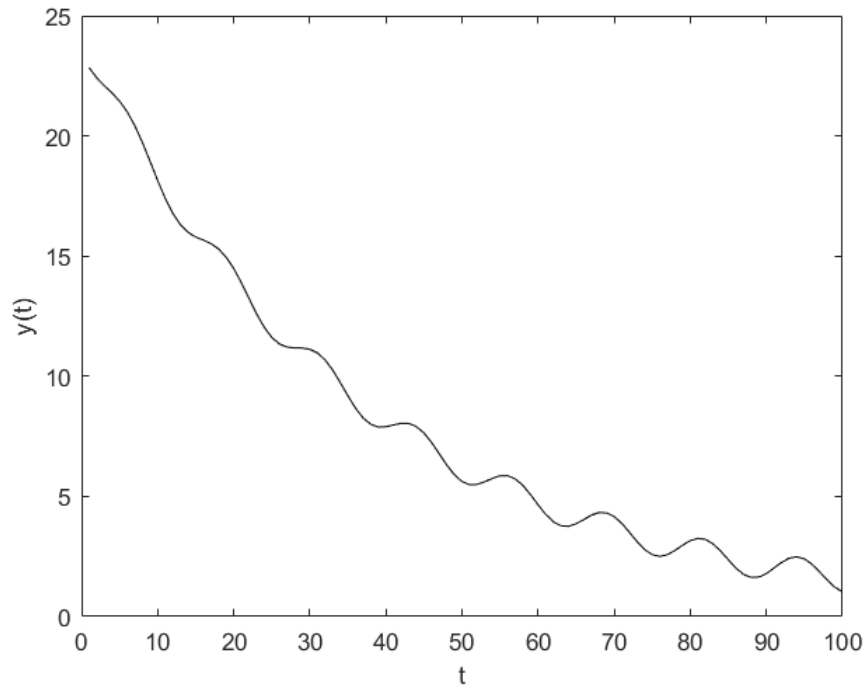
Supondo que x_1, x_2, x_3 e x_4 são desconhecidos e tem-se em mãos os pontos coletados, será aplicado o LM para estimar os parâmetros x_n que minimizam o MSE, definido em (72).

$$MSE = \frac{1}{N} \sum_{n=1}^N (\hat{y}_n(t) - y_n(t))^2, \quad (72)$$

em que $\hat{y}_n(t)$ são os dados estimados a partir da própria estrutura $y(t)$ com os valores de x_n encontrados pelo LM e N é a quantidade de amostras.

Para realizar a otimização é necessário definir quatro pontos iniciais para o método de LM fazer a busca, sendo eles, $x_n = (16,0; 20,0; 1,50; -4,00)$. Com esses pontos, a resposta do sistema é ilustrada na Figura 3.

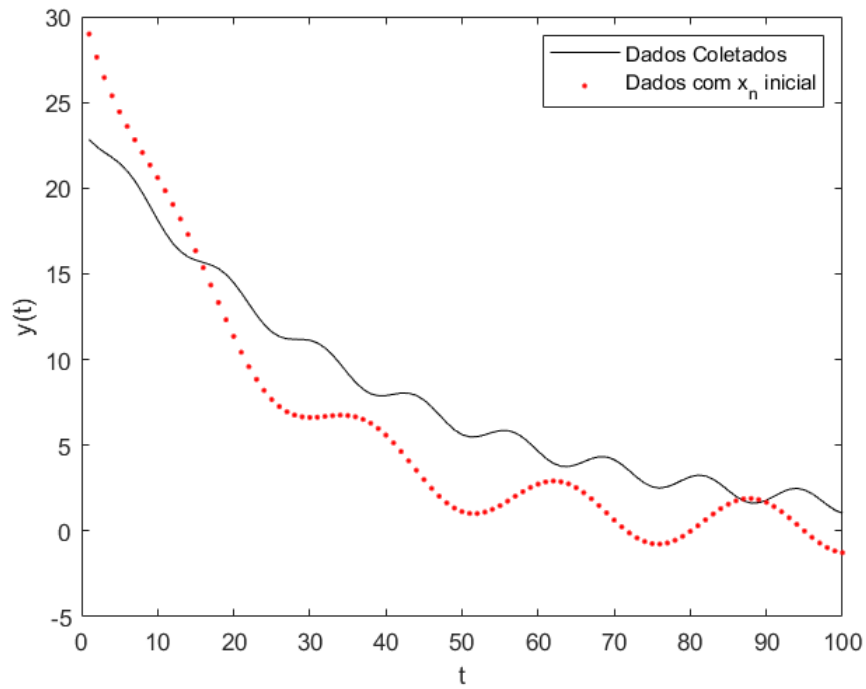
Figura 2 – Dados amostrados para o exemplo a partir da função $y(t)$ em (71) com $x_1 = 12,0$; $x_2 = 37,0$; $x_3 = 0,600$; $x_4 = -2,00$.



Fonte: Autoria Própria.

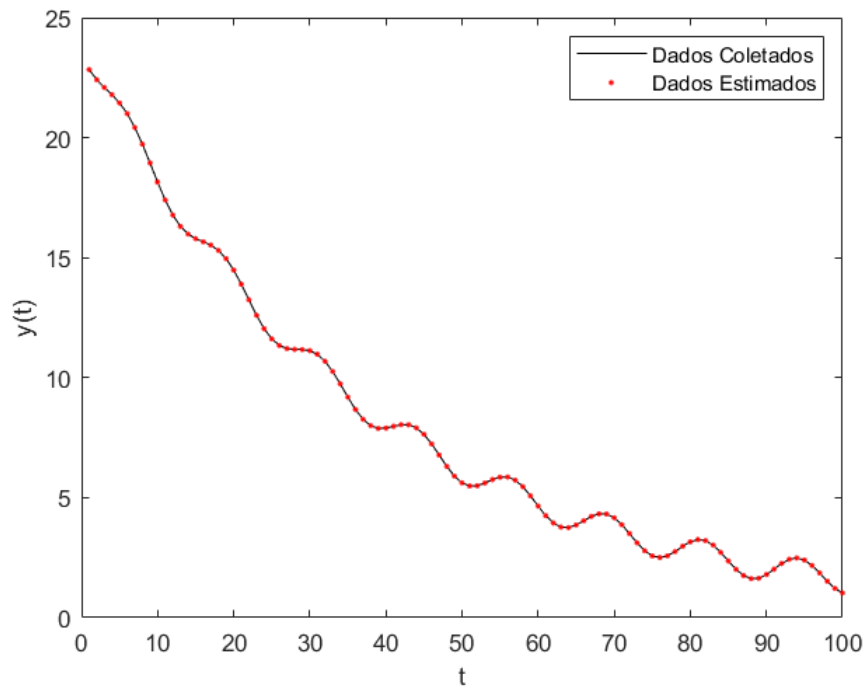
A partir de então, aplica-se o método de LM usando como critério de parada um MSE inferior a 10^{-3} ou 125 iterações, o que ocorrer primeiro. O algoritmo implementado foi interrompido ao atingir o número máximo de iterações. O resultado alcançado pode ser visto na Figura 4, o MSE atingido foi de 3×10^{-3} e os valores dos parâmetros encontrados foram $x_n = (12,0; 36,9; 0,600; -2,00)$.

Figura 3 – Comparação entre os dados amostrados com os x_n reais e os dados amostrados com os x_n iniciais.



Fonte: Autoria Própria.

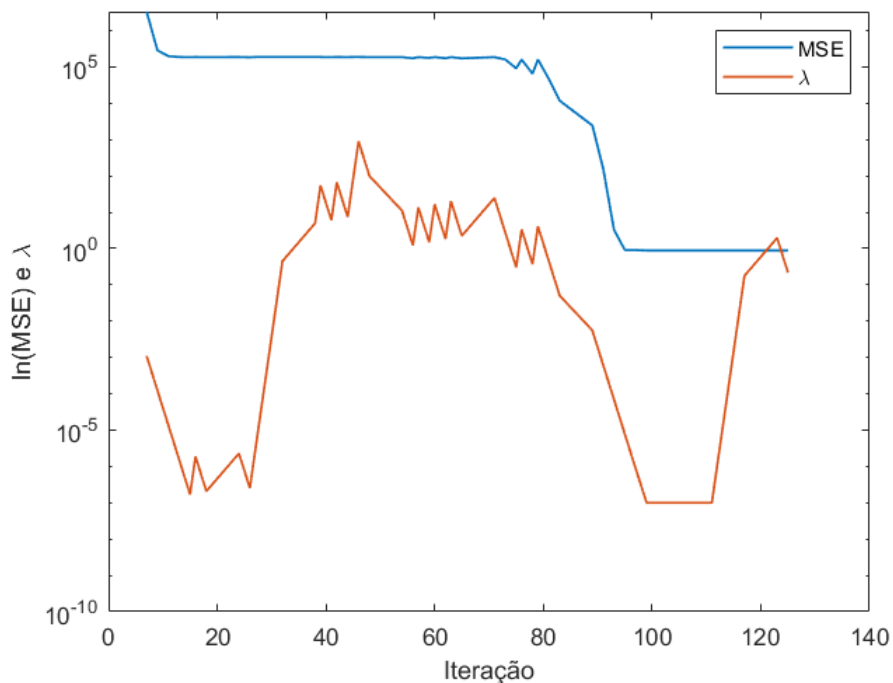
Figura 4 – Comparação entre os dados amostrados e os dados estimados com o método de LM.



Fonte: Autoria Própria.

Uma observação importante que pode ser feita é o comportamento do parâmetro λ durante a busca pela minimização do MSE. Na Figura 5 pode ser visto esse comportamento em que λ diminui rapidamente quando o MSE está diminuindo rapidamente (como ocorre entre as iterações 85 a 100) e, aumenta rapidamente quando o MSE está diminuindo lentamente (como acontece entre as iterações 115 a 120). Assim, o método de LM caminha com pequenos passos quando a curvatura do MSE foi mais intensa e com passos largos quando o MSE teve pouca variação.

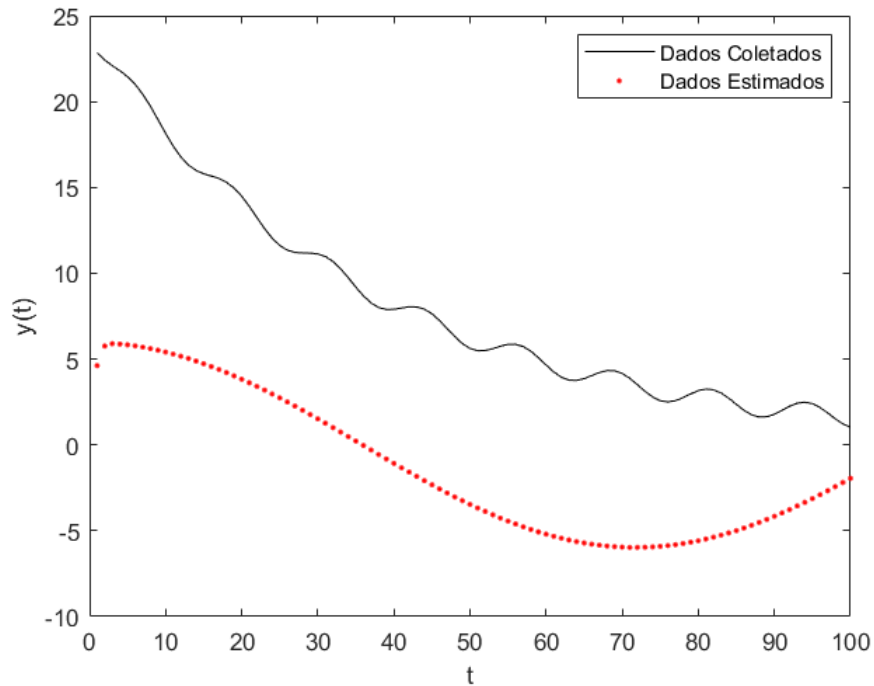
Figura 5 – Comportamento do MSE e do parâmetro λ durante a minimização para estimar os parâmetros com o LM.



Fonte: Autoria Própria.

Com esses resultados, pode se inferir que os parâmetros estimados se aproximaram muito dos valores reais dos parâmetros, permitindo uma modelagem que se assemelhasse aos dados reais. Contudo, o método de LM é dependente dos parâmetros iniciais inseridos. Para ilustrar essa dependência, um novo teste foi executado utilizando os mesmos critérios utilizados no caso anterior, mas dessa vez com o seguinte conjunto de parâmetros iniciais $x_n = (-20,0; 10,0; -9,00; 22,0)$.

Figura 6 – Comparação entre os dados amostrados e os dados estimados com o algoritmo de LM para pontos iniciais distantes dos pontos reais x_n .



Fonte: Autoria Própria.

Os parâmetros alcançados dessa vez foram $x_n = (-4,77; 0,512; -5,97; 22,8)$ e como pode ser visto na Figura 6, o modelo não conseguiu convergir para os dados reais, evidenciando uma dificuldade do método de LM quando os parâmetros iniciais estão muito distantes dos parâmetros reais do sistema.

Assim sendo, uma alternativa para contornar esse problema do método de LM é a utilização de métodos que não dependam do ponto de partida. Nesse sentido, a Seção 4.3, apresenta o método de OB que se baseia em modelos probabilísticos como uma alternativa ao método de LM.

4.3 OTIMIZAÇÃO BAYESIANA

A origem da OB remete a meados da década de 1960 no trabalho de Kushner (1964). Esse artigo trata de processos de Wiener¹ para problemas unidimensionais sem restrições, no qual o modelo desenvolvido consiste em maximizar a probabilidade de melhoria. O modelo proposto também possui um parâmetro para fazer o balanço entre otimização local e global, precedendo o que mais tarde seria chamado de *trade-off* de *exploration* (busca global) e *exploitation* (busca local). Outro trabalho que apresenta relevância no desenvolvimento inicial da otimização bayesiana é Mockus *et al.* (1978), o qual descreve a primeira versão da função de aquisição de melhoria esperada, dessa vez, para aplicações de combinações lineares para os campos de Wiener.

A otimização bayesiana passou a receber mais atenção depois da popularização do trabalho de Jones *et al.* (1998), que apresentou o algoritmo *Efficient Global Optimization* (EGO). Seguindo esse trabalho, uma vasta literatura foi desenvolvida para, por exemplo, resolver problemas de otimização multi-objetivo (KEANE, 2006; KNOWLES, 2006), estudos de faixa de convergência (CALVIN; ŽILINSKAS, 1999; CALVIN; ZILINSKAS, 2003), treinamento de redes neurais profundas (SNOEK *et al.*, 2012) e técnicas de otimização paralela (GINSBOURGER *et al.*, 2007; WU; FRAZIER, 2016).

Algumas aplicações interessantes de OB podem ser encontradas no trabalho de Lizotte (2008), no qual se otimiza um conjunto de parâmetros da caminhada do robô Sony AIBO ERS-7 de forma a maximizar a velocidade de deslocamento. Outra aplicação pode ser encontrada no trabalho de Cora (2008), em que o sistema utiliza OB para aprender as regras de direção para conduzir um veículo dentro de um ambiente. Como pode ser notado, a OB pode ser utilizada em diferentes contextos de aplicações. Sendo assim, a Seção 4.3.1 abordará o desenvolvimento do método em sua forma clássica.

4.3.1 Método de Otimização Bayesiana

A OB é uma abordagem proeminente para os casos em que a função de avaliação é custosa. Este custo pode decorrer de situações em que não há acesso às derivadas, ou ainda, para casos em que o problema é não convexo. Essa técnica de otimização é uma das abordagens mais

¹ Processo de Wiener é um processo estocástico contínuo no tempo, também conhecido como movimento browniano padrão.

eficientes no que diz respeito ao critério da quantidade de avaliações necessárias para obter um resultado satisfatório (JONES, 2001; SASENA, 2002).

Parte da eficiência da OB é decorrente da possibilidade de inserção do conhecimento *a priori* sobre o problema e o uso de *exploration* e *exploitation* no espaço de busca. Esta abordagem é fundamentada no Teorema de Bayes. Assim, assumindo x_n como a n -ésima amostra, $x_{1:n} = \{x_1, \dots, x_n\}$, $f(x_n)$ como o valor observado da função objetivo em x_n e $D_{1:n} = \{x_{1:n}, f(x_{1:n})\}$ um conjunto de dados coletados, então, a probabilidade *a posteriori* (probabilidade condicional) do modelo f dado $D_{1:n}$, é diretamente proporcional à probabilidade *a posteriori* de $D_{1:n}$ dado f multiplicada pela probabilidade *a priori* de f e inversamente proporcional à probabilidade *a priori* de $D_{1:n}$, como mostra (73):

$$P(f|D_{1:n}) \propto \frac{P(D_{1:n}|f)P(f)}{P(D_{1:n})}, \quad (73)$$

Na OB o conhecimento *a priori* pode ser interpretado como o que se sabe sobre as características do sistema, por exemplo, suavidade, velocidade de resposta, entre outras. Essas informações fazem com que algumas funções de avaliação sejam mais factíveis do que outras e isso pode ser incorporado na modelagem.

De forma geral, a metodologia da OB pode ser explicada por meio da Figura 7. No momento $t = 1$ são necessárias duas amostras observadas inicialmente a partir do modelo probabilístico *a priori* de f , x_1 e x_2 . A partir destas amostras, no momento $t = 2$, é calculada a distribuição *a posteriori* e gerado o modelo probabilístico $P(f|D_{x_1,x_2})$ (região em azul), que contém as candidatas à futura amostra x_3 . Para decidir onde estará a amostra x_3 , é utilizada uma função auxiliar, denominada Função de Aquisição (região em verde), na qual seu ponto de maximização corresponde a amostra x_3 . Então, no momento $t = 3$, é calculada novamente a distribuição *a posteriori*, $P(f|D_{x_1,x_2,x_3})$, agora com o novo conjunto de dados, e maximizada a função de aquisição para definir a amostra x_4 . De forma análoga, de $t = 4$ em diante, repete-se esse processo até a função objetivo estimada (linha preta contínua) se aproximar da função objetivo real (linha tracejada) ou até atingir o critério de parada do algoritmo. Com a função objetivo em mãos, pode-se determinar qual é o mínimo global do problema; o Algoritmo 2 generaliza os passos descritos anteriormente.

Algoritmo 2 – Descrição do funcionamento da OB.

Aplique o modelo estatístico para obter *a priori* de f ;
 Obtenha n amostras iniciais no espaço de busca.

enquanto critério de parada não atingido **faça**

Atualize a distribuição de probabilidade *a posteriori* de f usando todas as amostras avaliadas.

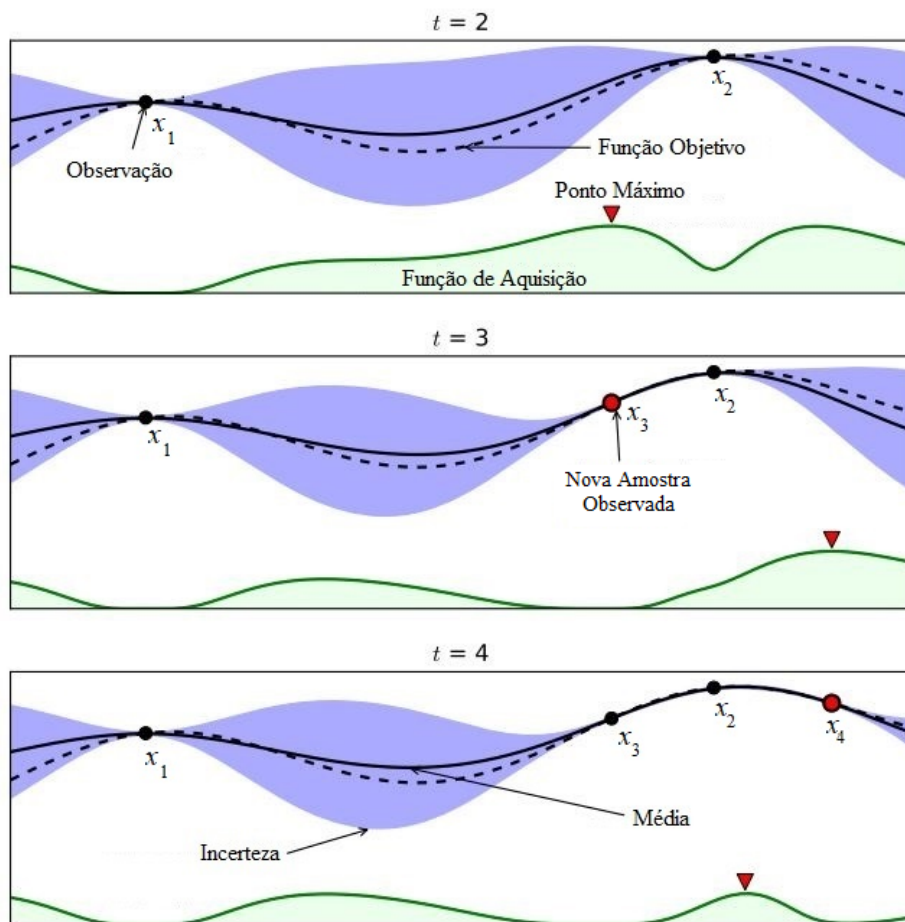
Maximize a função de aquisição usando a distribuição *a posteriori* atual.

Faça $y_n = f(x_n)$.

finaliza enquanto

Fonte: Autoria própria.

Figura 7 – Exemplificação do funcionamento da OB.



Fonte: Adaptado de Brochu *et al.* (2010).

Dessa forma, o problema tratado com a OB pode ser descrito pela seguinte minimização:

$$\min_{x \in A} f(x) \quad (74)$$

em que A é um hiper-retângulo, $x \in \mathbb{R}^m$, $f(x)$ é uma função contínua da qual não serão avaliadas as derivadas, não possui ruídos e as informações sobre linearidade e concavidade de $f(x)$ são desconhecidas. Assim, o foco da otimização é a busca global.

Imposta tais condições, o método de OB irá convergir para o ponto ótimo se (MOCKUS, 1982):

- a função de aquisição for contínua;
- a covariância convergir para zero.

Como pode ser observado na Figura 7, a OB possui duas componentes principais. A primeira é o modelo estatístico bayesiano (*surrogate function*), que irá modelar a função de avaliação, e a segunda é a função de aquisição para decidir qual será o próximo ponto. Para a obtenção do modelo estatístico será utilizado neste trabalho o Processo Gaussiano (\mathcal{PG}), pois o mesmo é recorrente na literatura e, de acordo com Mockus (1982), o processo gaussiano apresenta bons resultados para determinação da distribuição *a priori*.

Para a escolha da função de aquisição, sempre irá ocorrer um *trade-off* na escolha do ponto x_n em que a função objetivo é muito incerta (*exploration*) e onde espera-se que o ponto x_n minimize a função objetivo (*exploitation*). Sendo assim, as Seções 4.3.2 e 4.3.3 irão discorrer sobre o \mathcal{PG} e duas opções para função de aquisição.

4.3.2 Processo Gaussiano

O \mathcal{PG} pode ser entendido como uma extensão da distribuição gaussiana multivariável para um processo estocástico de dimensão infinita (BROCHU *et al.*, 2010). Em outras palavras, o \mathcal{PG} é uma distribuição de funções, $\{f(u)_{u \in \Lambda}\}$, determinada pela função média, $m : \Lambda \rightarrow \mathbb{R}$, e função de covariância, $k : \Lambda \times \Lambda \rightarrow \mathbb{R}$, podendo ser representado por

$$f(u) \sim \mathcal{PG}(m(u), k(u, u^\top)). \quad (75)$$

Considerando um \mathcal{PG} e os conjuntos de dados disponíveis

$$U = \{u_n\}, \quad n = 1, \dots, N \quad \forall u \in \Lambda$$

e

$$Y = \{y_n\}, \quad n = 1, \dots, N \quad \forall y \in \Lambda$$

a matriz de média é dada por:

$$\mathbf{M}(U, Y) = \mathbb{E}\{U, Y\} = \begin{bmatrix} m(u_1) & m(y_1) \\ m(u_2) & m(y_2) \\ \vdots & \vdots \\ m(u_n) & m(y_n) \end{bmatrix}. \quad (76)$$

De forma similar, a matriz de covariância de $f(U)$ e $f(Y)$ é definida como:

$$\mathbf{K}(U, Y) = \text{cov}\{U, Y\} = \mathbb{E}\{[f(U) - m(U)][f(Y) - m(Y)]\}. \quad (77)$$

$$\mathbf{K}(U, Y) = \begin{bmatrix} k(u_1, y_1) & k(u_1, y_2) & \dots & k(u_1, y_n) \\ k(u_2, y_1) & k(u_2, y_2) & \dots & k(u_2, y_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(u_n, y_1) & k(u_n, y_2) & \dots & k(u_n, y_n) \end{bmatrix}. \quad (78)$$

Então, o \mathcal{PG} para esse conjunto de dados é definido como:

$$\begin{bmatrix} f(U) \\ f(Y) \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} m(U) \\ m(Y) \end{bmatrix}, \begin{bmatrix} \mathbf{K}(U, U) & \mathbf{K}(U, Y) \\ \mathbf{K}(Y, U) & \mathbf{K}(Y, Y) \end{bmatrix} \right). \quad (79)$$

Por questões de praticidade, serão considerados dados sem ruídos. Desta forma, a função média pode tomar o valor zero, isto é $m(U) = m(Y) = 0$, sem que haja prejuízo nas análises. Já a função de covariância precisa seguir alguns critérios para seleção, que são:

- $k(u, y) = k(y, u), \quad \forall u, y \in \Lambda;$
- $x^\top \mathbf{K}(U, U)x \geq 0, \quad x \in \mathbb{R}^m, \quad \forall x \neq 0.$

Isto significa que $k(u, y)$ precisa ser simétrica e $\mathbf{K}(U, U)$ precisa ser positiva semidefinida. Logo, o desafio é determinar a função de covariância, k . Uma função frequentemente utilizada na literatura é a função exponencial quadrática, dada por:

$$k(u_n, y_m) = e^{\left(-\frac{1}{2}\|u_n - y_m\|^2\right)}. \quad (80)$$

O uso da função em (80) permite a interpretação de que valores muito similares vão levar a covariância a se aproximar de 1, enquanto valores muito distantes levarão a covariância a

se aproximar de 0. Ou seja, pontos que estejam próximos um do outro vão ter maior influência entre si do que pontos muito distantes.

Dessa forma, considera-se $U = \{u_n\}, n = 1, \dots, N \forall u \in \Lambda$ como um conjunto de dados de entrada e $Y = \{y_n\}, n = 1, \dots, N \forall y \in \mathbb{R}$ como um conjunto de dados de saída disponíveis e que exista uma relação que leva u_n a y_n , ou seja, $f : \Lambda \rightarrow \mathbb{R}$. Deseja-se estimar f . Para tal, define-se $U^{N+1} \subset \Lambda$ como um conjunto de dados de entrada de testes dos quais deseja-se estimar o conjunto de dados correspondentes, $Y^{N+1} \subset \mathbb{R}$, ou seja, deseja-se saber $f(U^{N+1})|U^{N+1}, U, Y$.

Assim sendo, utiliza-se da propriedade de condicionamento da distribuição gaussiana, resultando em:

$$\begin{aligned} f(U^{N+1})|U^{N+1}, U, Y &\sim \mathcal{N}(m(U^{N+1}) + \mathbf{K}(U^{N+1}, U)[\mathbf{K}(U, U)]^{-1}(Y - m(Y)), \\ &\mathbf{K}(U^{N+1}, U^{N+1}) - \mathbf{K}(U^{N+1}, U)[\mathbf{K}(U, U)]^{-1}\mathbf{K}(U, U^{N+1})). \end{aligned} \quad (81)$$

Como está sendo considerado que $m(U) = m(Y) = 0$, então,

$$m(f(U^{N+1})) = \mathbf{K}(U^{N+1}, U)[\mathbf{K}(U, U)]^{-1}Y \quad (82)$$

e

$$\mathbf{K}(f(U^{N+1})) = \mathbf{K}(U^{N+1}, U^{N+1}) - \mathbf{K}(U^{N+1}, U)[\mathbf{K}(U, U)]^{-1}\mathbf{K}(U, U^{N+1}). \quad (83)$$

Portanto, a distribuição *a posteriori* de $f(U^{N+1})$ pode ser encontrada de forma analítica por meio das informações de média e covariância, sendo esse o conjunto central de equações para realizar regressão por meio de \mathcal{PG} . Essas expressões analíticas são possíveis porque foi escolhida uma gaussiana para $P(D_{1:N}|f)$, pois, para casos mais gerais, isso pode não ocorrer. Apresentado o \mathcal{PG} , a Seção 4.3.3 irá abordar as Funções de Aquisição e como elas auxiliam na otimização.

4.3.3 Função de Aquisição

A finalidade da função de aquisição é direcionar o caminho de busca da otimização. Geralmente, a função de aquisição é definida de maneira a maximizar a função objetivo. Na sequência serão abordadas duas funções de aquisição mais usuais, a Probabilidade de Melhoria (PI - *Probability of Improvement*) e Melhoria Esperada (EI - *Expected Improvement*).

4.3.3.1 Probabilidade de Melhoria

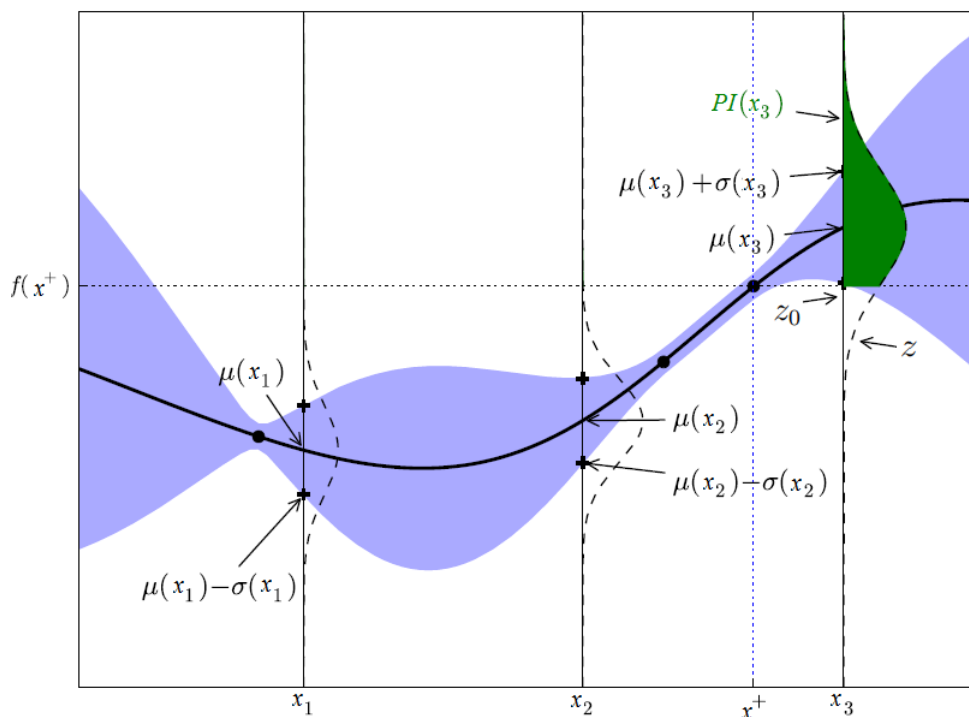
A PI foi apresentada no trabalho de Kushner (1964), ela consiste em buscar o ponto que maximiza a probabilidade de melhoria por meio da análise da média e covariância. Dado que cada ponto x está ligado a uma distribuição gaussiana de $f(x)$ com uma média $\mu(x)$ e variância $\sigma^2(x)$, então:

$$f(x) \sim \mathcal{N}(\mu(x), \sigma^2(x)). \quad (84)$$

Fazendo uso da Figura 8 para exemplificar o funcionamento do método, suponha que deseja-se maximizar $f(x)$, x^+ seja a melhor solução encontrada até o momento e são analisados os possíveis candidatos para ser o novo ponto (x_1 , x_2 ou x_3). Logo, o ponto a ser escolhido é o que possui a maior área acima de $f(x^+)$ (área em verde na Figura 8).

Em outras palavras, o ponto escolhido irá refletir a incerteza da região, pois, ou aquela região possui uma alta incerteza (área em azul na Figura 8), e então é necessário amostrar um ponto naquela região para reduzir a incerteza (*exploration*), ou o ponto deve estar muito próximo da média (*exploitation*) e a área dele será maior do que as outras áreas no espaço de busca.

Figura 8 – Exemplificação do funcionamento da PI.



Fonte: Adaptado de Brochu *et al.* (2010).

Isso posto, define-se a função de melhoria $I(x)$ como em (85). Dessa forma, é avaliado

se cada ponto x é capaz de gerar um $I(x) \geq 0$.

$$I(x) = \max(f(x) - f(x^+), 0). \quad (85)$$

Agora, considerando um certo $z \sim \mathcal{N}(0, 1)$, a equação (84) pode ser reescrita de forma a $f(x)$ continuar com uma média $\mu(x)$ e variância $\sigma^2(x)$, como mostra (86).

$$f(x) = \mu(x) + \sigma(x)z. \quad (86)$$

Conseqüentemente, (85) pode ser reescrita como:

$$I(x) = \max(\mu(x) + \sigma(x)z - f(x^+), 0). \quad (87)$$

Retornando à Figura 8, a $PI(x)$ nada mais é do que a área em verde abaixo da curva tracejada relacionada à variável aleatória z . Para calculá-la, é subtraída a área sob uma gaussiana com distribuição normal da área de (87), para um $z < z_0$ e obtém-se:

$$\begin{aligned} PI(x) &= \int_{-\infty}^{+\infty} e^{\left(\frac{-z^2}{2}\right)} dz - \int_{-\infty}^{z_0} e^{\left(\frac{-z^2}{2}\right)} dz \\ &= \int_{z_0}^{+\infty} e^{\left(\frac{-z^2}{2}\right)} dz \\ &= \Psi(-z_0) = \Psi\left(\frac{\mu(x) - f(x^+)}{\sigma(x)}\right), \end{aligned} \quad (88)$$

em que $\Psi(\cdot)$ é uma função de distribuição acumulada (CDF - *Cumulative Distribution Function*). Dessa forma, um ponto x que possua uma maior probabilidade de $f(x) > f(x^+)$, ou seja, $\sigma(x) < 1$, será sobreposto a outros pontos que possam gerar ganhos maiores, mas que possuam maiores incertezas, isto é $\sigma(x) > 1$, mesmo que essa melhoria seja muito pequena. Isso pode levar o método a realizar uma busca exaustivamente local, em torno do melhor ponto atual, e só depois realizar uma busca global (JONES, 2001). Uma abordagem para atenuar esse problema é apresentada na Seção 4.3.3.2.

4.3.3.2 Melhoria Esperada

O algoritmo da EI foi apresentado no trabalho de Močkus (1975) e foi popularizado depois de ter sido utilizado nos estudos de Jones *et al.* (1998) sobre otimização global. O método foi proposto como uma melhoria ao PI, pois este algoritmo está considerando apenas a probabilidade de melhoria e não a magnitude da melhoria que o ponto pode produzir. Dessa

forma, o $EI(x)$ pode ser compreendido como a magnitude de $I(x)$ e pode ser calculada por meio de:

$$EI(x) = \int_{-\infty}^{+\infty} I(x) e^{\left(\frac{-z^2}{2}\right)} dz, \quad (89)$$

note que está sendo considerado $z \sim \mathcal{N}(0, 1)$ e um *kernel* exponencial quadrático, como definido em (80).

Fazendo a substituição de (87) em (89), tem-se:

$$EI(x) = \int_{-\infty}^{+\infty} (\mu(x) + \sigma(x)z - f(x^+)) e^{\left(\frac{-z^2}{2}\right)} dz \quad (90)$$

Divide-se os limites de integração da equação (90) em duas partes por meio do ponto z_0 . Diante disso, obtém-se:

$$EI(x) = \int_{-\infty}^{z_0} (\mu(x) + \sigma(x)z - f(x^+)) e^{\left(\frac{-z^2}{2}\right)} dz + \int_{z_0}^{+\infty} (\mu(x) + \sigma(x)z - f(x^+)) e^{\left(\frac{-z^2}{2}\right)} dz. \quad (91)$$

Como $I(x)$ para $z < z_0$ é nula, a primeira parte da equação (91) é igual a zero. Portanto, tem-se:

$$\begin{aligned} EI(x) &= \int_{z_0}^{+\infty} (\mu(x) + \sigma(x)z - f(x^+)) e^{\left(\frac{-z^2}{2}\right)} dz \\ &= \int_{z_0}^{+\infty} (\mu(x) - f(x^+)) e^{\left(\frac{-z^2}{2}\right)} dz + \int_{z_0}^{+\infty} \sigma(x)z e^{\left(\frac{-z^2}{2}\right)} dz \\ &= (\mu(x) - f(x^+)) \int_{z_0}^{+\infty} e^{\left(\frac{-z^2}{2}\right)} dz + \sigma(x) \int_{z_0}^{+\infty} z e^{\left(\frac{-z^2}{2}\right)} dz \\ &= (\mu(x) - f(x^+)) \Psi\left(\frac{\mu(x) - f(x^+)}{\sigma(x)}\right) + \sigma(x) \varphi\left(\frac{\mu(x) - f(x^+)}{\sigma(x)}\right), \end{aligned} \quad (92)$$

sendo $\varphi(\cdot)$ uma função de densidade de probabilidade (PDF - *Probability Density Function*) e $\sigma(x) > 0$. Para a situação em que $\sigma(x) = 0$, $EI(x) = 0$.

Dessa forma, a $EI(x)$ é alta quando: $\mu(x) > f(x^+)$, o que significa que o valor da média do processo gaussiano é alto em x (*exploitation*); ou quando $\sigma(x) > 1$, indicando que há muita incerteza na região (*exploration*).

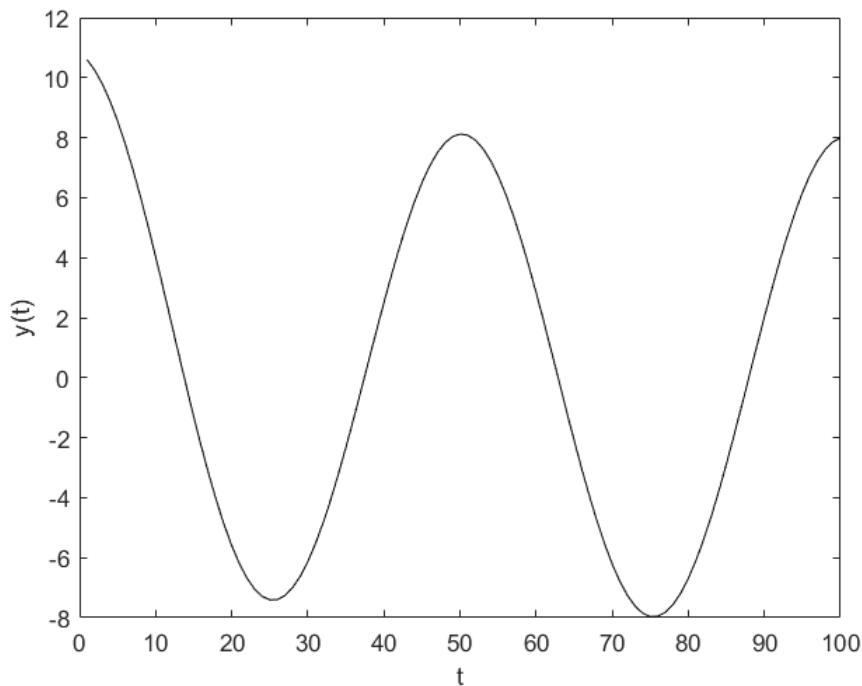
4.3.4 Exemplo Ilustrativo

A OB pode ser utilizada em diferentes problemas de otimização. Na sequência será apresentado um exemplo de problema de estimação de parâmetro. Considere a seguinte função:

$$y(t) = \sqrt{x e^{\left(\frac{-t}{x}\right)}} + x \cos\left(\frac{t}{x}\right). \quad (93)$$

Em seguida, define-se $x = 8,00$. A resposta do sistema para o intervalo de $t = [0, 100]$ pode ser vista na Figura 9. A partir de então, foram inseridos esses dados amostrados no algoritmo e foi executada uma simulação para estimar o valor de x que minimizasse o MSE entre os dados estimados e os dados coletados. Como função de aquisição foi utilizada a função EI. Na Figura 10 pode ser visto o resultado alcançado.

Figura 9 – Dados amostrados para o exemplo a partir da função (93) considerando $x = 8,00$.

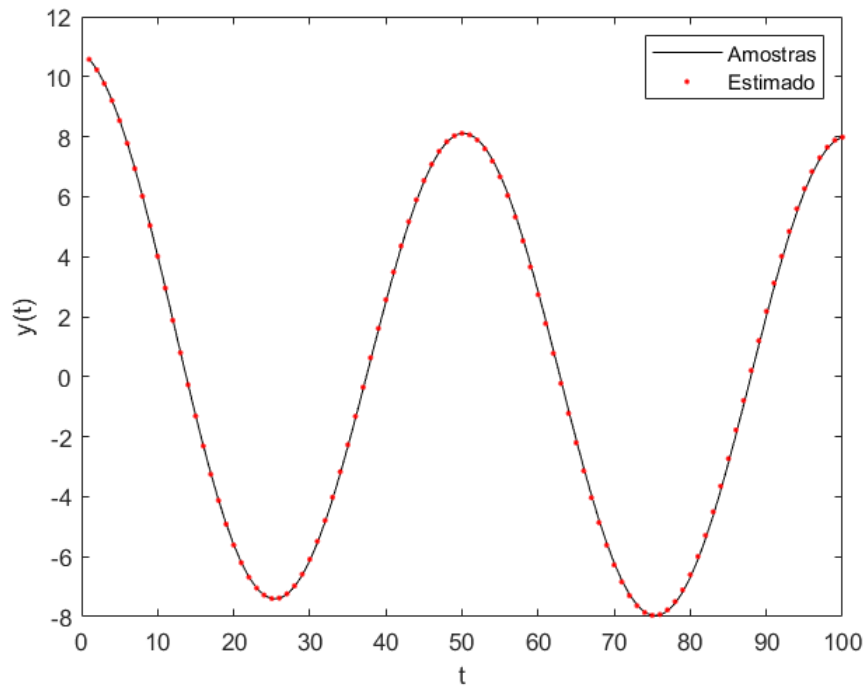


Fonte: Autoria Própria.

O valor do parâmetro x encontrado foi de 7,98. Com esse valor foi possível atingir um valor de MSE de $1,099 \times 10^{-3}$. A Figura 11 ilustra como o MSE se comportou durante a busca, nota-se que o MSE decaiu durante as 10 primeiras iterações, mas depois ficou estagnado. Para entender o motivo, pode-se observar a Figura 12 em que é mostrado como a evolução da estimativa de x se comportou. Então, entende-se que depois da décima iteração o algoritmo não

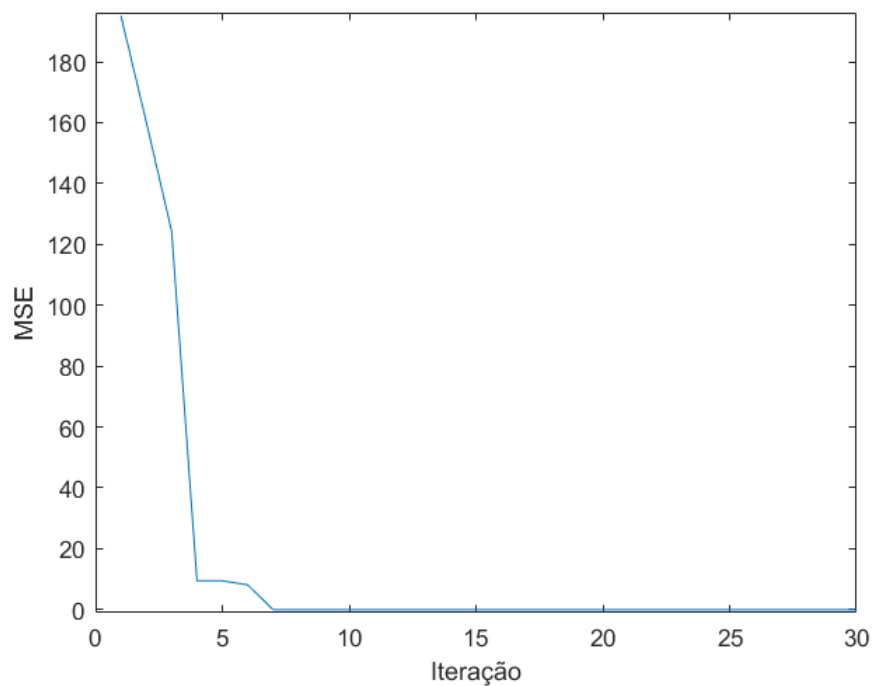
conseguiu um ponto para x que otimizasse mais o problema, dado que o valor encontrado é um mínimo local.

Figura 10 – Resultado da estimação do parâmetro x com a OB.



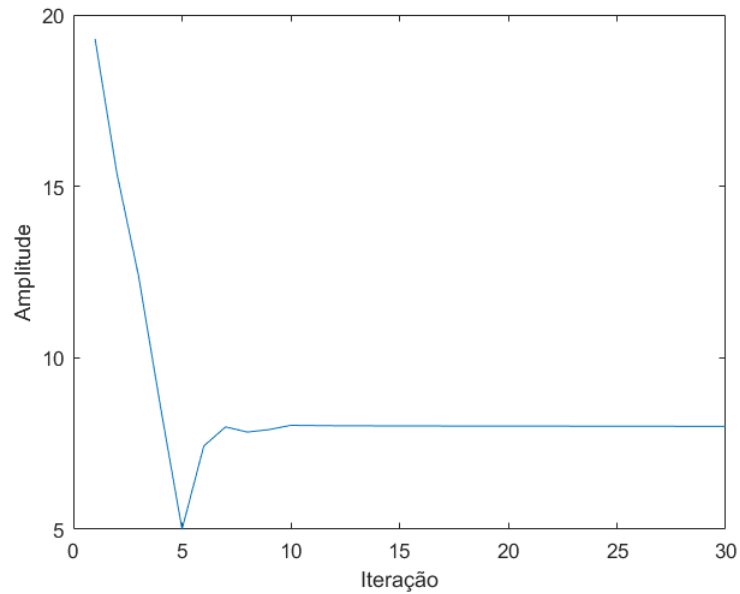
Fonte: Autoria Própria.

Figura 11 – Comportamento do MSE durante a OB para encontrar o valor de x .



Fonte: Autoria Própria.

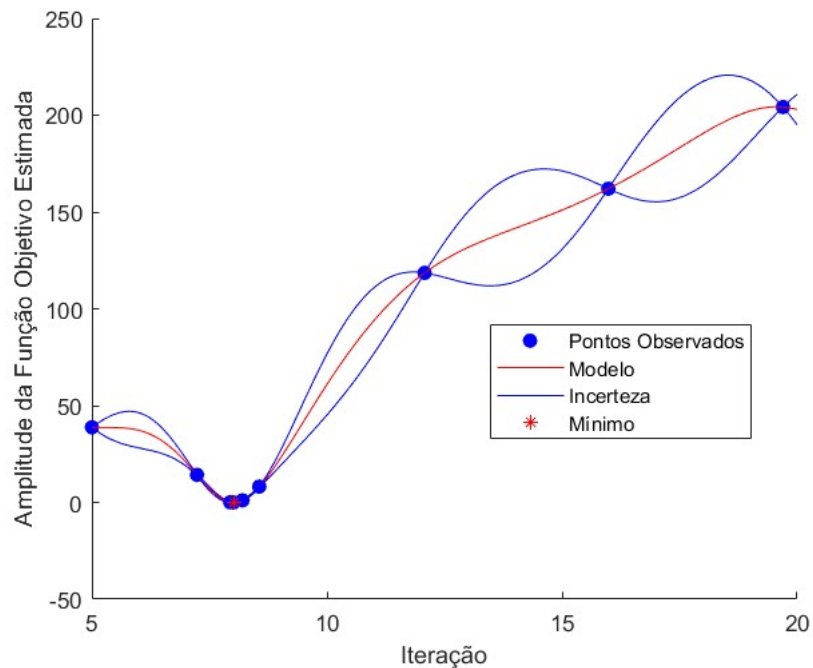
Figura 12 – Evolução da estimativa de x durante a OB.



Fonte: Autoria Própria.

Também pode ser avaliado como se comportou o modelo de avaliação durante a busca, este resultado pode ser observado na Figura 13. Percebe-se que mesmo após a décima iteração o algoritmo ficou buscando pontos para diminuir a incerteza ao longo do modelo, mas nenhum deles conseguiu gerar um resultado melhor do que o já encontrado.

Figura 13 – Comportamento do modelo com a OB durante a busca por x .



Fonte: Autoria Própria.

4.4 SÍNTESE

O MMQ remonta aos problemas de ajustes de curvas para dados astronômicos do século XVIII, empregado para soluções de problemas de minimização linear devido a sua simplicidade e eficiência. Para os casos não-lineares existem outras ferramentas que apresentam um melhor desempenho. Um método que pode ser citado é o LM.

O LM foi inicialmente desenvolvido por Levenberg, que se aproveitou dos conceitos já estabelecidos nos métodos de Gauss-Newton e Gradiente Descendente. Porém, o controle da representatividade do fator de regularização ocasionava uma diminuição do peso do gradiente no cálculo do passo do algoritmo. Para contornar tal problema, Marquardt propôs uma melhoria com a troca da matriz identidade pela diagonal do gradiente. A partir disto, conseguiu-se garantir que o fator de regularização force o passo para a direção de máxima descida.

Outras características que o LM apresenta são a dependência dos pontos iniciais e a possibilidade de apresentar problemas para a inversão de matrizes com muitos parâmetros. Sendo assim, foi exposto nesse capítulo o método de OB que se baseia em modelos probabilísticos, com o uso do processo gaussiano e uma função de aquisição. Esse método é uma abordagem promissora em situações em que a função objetivo é custosa computacionalmente, sendo uma das técnicas mais proeminentes para o critério da quantidade de avaliações necessárias para atingir um resultado satisfatório.

Apresentadas as técnicas para modelagem de sistemas e os métodos para estimação de parâmetros, o Capítulo 5, a seguir, abordará um estudo de caso para validação das teorias apresentadas neste trabalho.

5 ESTUDO DE CASO

Neste capítulo será realizado um estudo de caso e a comparação entre o algoritmo de LM com o algoritmo de OB na solução do problema da estimação dos polos de Kautz no modelo Volterra-Kautz. Na Seção 5.1 será apresentado o tipo de sistema abordado e a formulação do problema. Na Seção 5.2 são apresentados os testes das implementações com um conjunto de dados reais e as discussões e análises dos resultados obtidos.

5.1 FORMULAÇÃO DO PROBLEMA

Como estudo de caso será considerado um sistema dinâmico, causal, subamortecido e não-linear, no qual a entrada $u(k)$ e a saída $y(k)$ estão disponíveis. Esse sistema terá sua dinâmica representada por um modelo Volterra-Kautz com *kernels* de primeira e de segunda ordem. Desta forma, o sistema pode ser aproximado por (94).

$$\hat{y}(k) \cong \sum_{n=1}^N \alpha_n u(k - \tau_n) \phi_n(\tau_n) + \sum_{n_1=1}^N \sum_{n_2=1}^N \alpha_{n_1, n_2} u(k - \tau_{n_1}) \phi_{n_1}(\tau_{n_1}) u(k - \tau_{n_2}) \phi_{n_2}(\tau_{n_2}). \quad (94)$$

Como foi exposto na Seção 3.2, esse modelo pode ser reescrito para ser linear nos parâmetros, como visto na equação (55), sendo que com o uso do MMQ, θ pode ser estimado com a equação (95).

$$\theta = [\psi^\top(k) \psi(k)]^{-1} \psi^\top(k) \mathbf{y}(k). \quad (95)$$

Definida a solução para os parâmetros lineares, falta a estimação de b e c que parametrizam os polos de Kautz. Para tanto, define-se a função $\hat{y}(b, c)$ que corresponde a equação (41) com as funções da base formadas com os parâmetros b e c .

Assim sendo, a função de avaliação escolhida é a que minimiza o MSE, dada por (96). As restrições impostas são as mesmas apresentadas na Seção 2.4 para os valores de b e c .

$$\min \quad \frac{1}{N} \sum_{n=1}^N (\hat{y}_n(b, c) - y_n)^2 \quad (96)$$

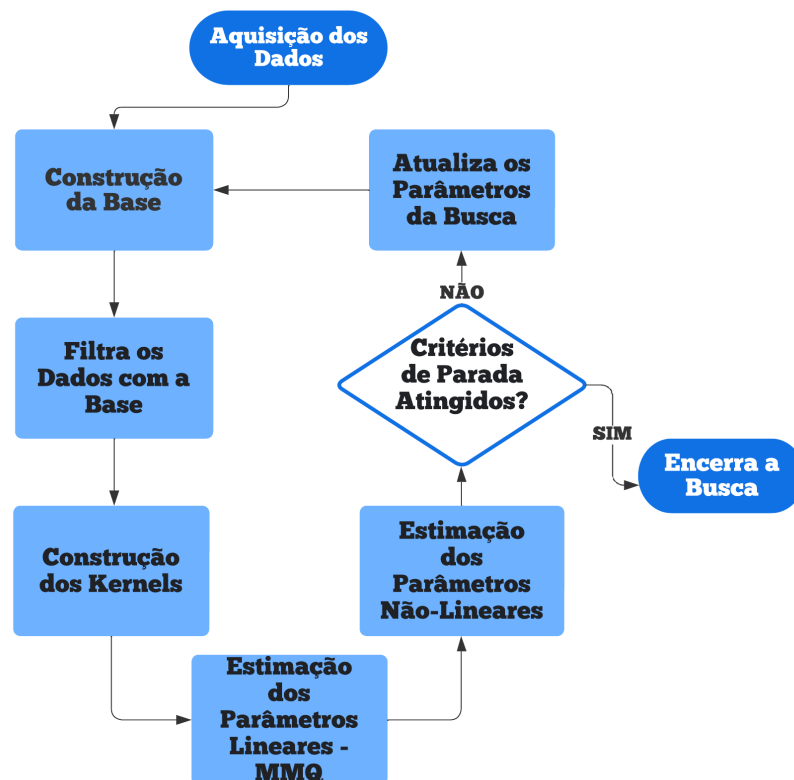
$$s.a. \quad -1 < c < 0; \quad (97)$$

$$|b| < 1; \quad (98)$$

$$b^2(1 - c)^2 + 4c \leq 0. \quad (99)$$

A partir do problema definido, é possível aplicar o algoritmo de LM com restrições e a OB para estimar os parâmetros que otimizem a modelagem do sistema. Na OB foram utilizados o \mathcal{PG} e a técnica de EI como função de aquisição, na Figura 14 é possível ver o fluxograma com a visão geral dos algoritmos implementados. A partir dessas implementações, analisou-se os resultados obtidos com o LM e a OB e foram comparados com os resultados das modelagens de outros dois estudos encontrados na literatura que utilizaram o mesmo *benchmark* que o presente trabalho. O primeiro trabalho é o de Zakaria *et al.* (2021), no qual foi utilizada uma estrutura NARMAX (*Nonlinear Autoregressive Moving Average with exogenous input*) para realizar a identificação do sistema proposto. O diferencial do deste estudo é o uso de um algoritmo evolucionário diferencial para otimização multi-objetiva, chamado de MOODE (*Multi-Objective Optimization Differential Evolution*). Com isso, os autores conseguiram uma modelagem melhor do que a estrutura NARMAX convencional. Já no trabalho de Ayala *et al.* (2014), os autores realizaram a identificação do sistema com modelos de Redes Neurais de Funções de Base Radial (RBFNNs - *Radial Basis Functions Neural Networks*) e um algoritmo evolutivo em cascata para encontrar os coeficientes. As simulações e discussões dos algoritmos implementados serão abordadas na Seção 5.2.

Figura 14 – Fluxograma de visão geral dos algoritmos implementados.

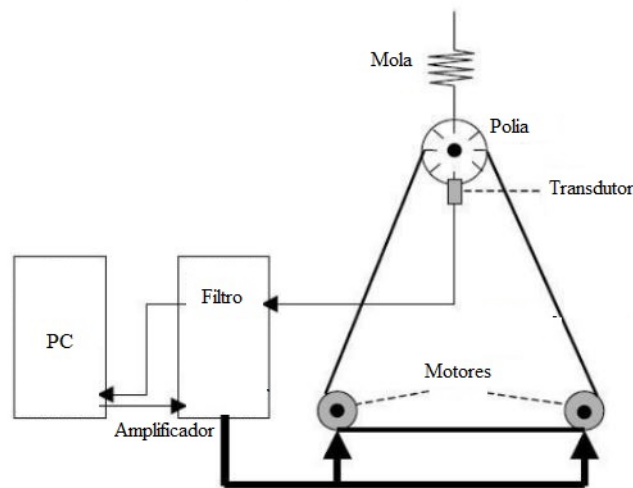


Fonte: Autoria Própria.

5.2 ANÁLISES E DISCUSSÕES

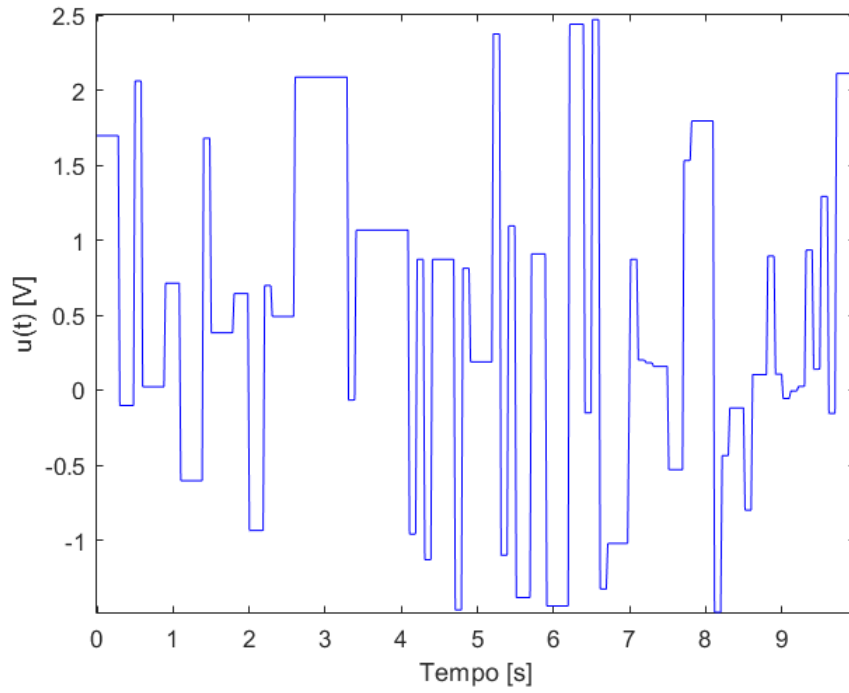
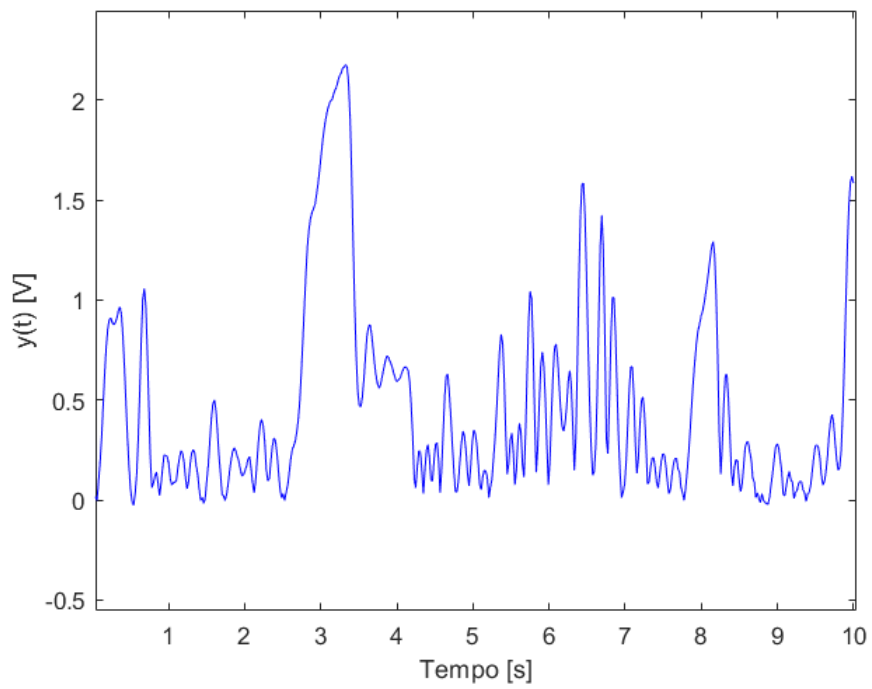
Para validar os algoritmos implementados foi utilizado um conjunto de dados, disponível em Wigren e Schoukens (2017), de um sistema de acoplamento composto por dois motores elétricos, interligados em uma polia, por uma correia. A polia é fixada em uma mola, o que possibilita oscilação no sistema. A tensão mecânica e a velocidade da correia são controladas pelo acionamento individual dos motores. O objetivo deste experimento é coletar a velocidade da oscilação. Então foi instalado um sensor de pulso na polia e seu sinal passa por filtros passa-baixa e *anti-aliasing* e transmitido a um computador, a Figura 15 ilustra o sistema.

Figura 15 – Sistema usado na coleta de dados.



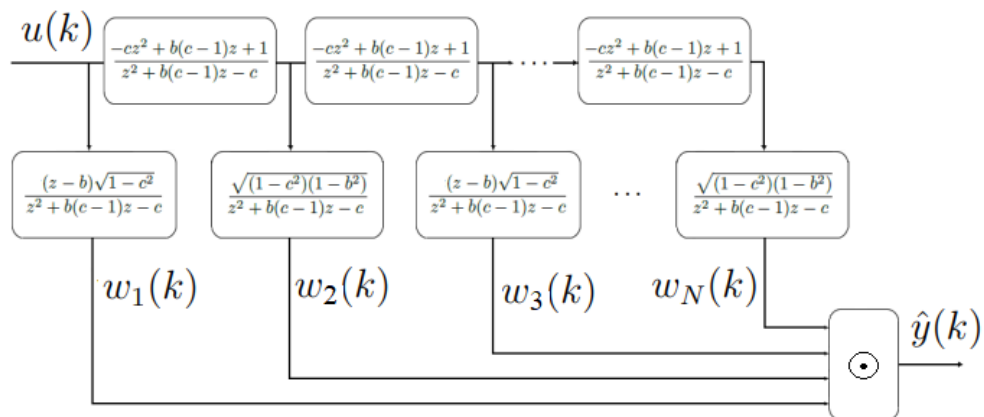
Fonte: Adaptado de Wigren e Schoukens (2017).

Para excitar o sistema, um sinal de tensão elétrica $u(k)$, com amplitude variando entre $-1,5V$ e $+2,5V$, foi aplicado aos motores, o qual foi multiplicado por um escalar aleatório uniformemente distribuído entre 0 e 1. Com isso, foi possível explorar as não linearidades do sistema, tanto em frequência quanto em amplitude. O resultado da velocidade de oscilação da polia foi convertido em um sinal de tensão $y(k)$, e as respostas do sistema foram coletadas. O intervalo de experimentação foi de 10 s com um período de amostragem de 20 ms, resultando em 500 amostras. O sinal de entrada e a resposta do sistema podem ser vistos nas Figuras 16 e 17, respectivamente.

Figura 16 – Sinal de excitação do sistema.**Fonte: Autoria Própria.****Figura 17 – Sinal de resposta do sistema.****Fonte: Autoria Própria.**

Para a implementação do algoritmo de modelagem foi utilizado o *software* MATLAB®. Nele foi executado o algoritmo para o método de LM e o desenvolvimento da base de Kautz na estrutura de filtros digitais, conforme mostrado na Figura 18. Nesta representação, as entradas $u(k)$ são filtradas por cada função de Kautz, par e ímpar, resultando em $w_n(k)$ e, em sequência, o operador \odot que simboliza o mapeamento estático, é realizado por meio da série de Volterra, com *kernels* de primeira e de segunda ordem. Assim, a saída estimada $\hat{y}(k)$ será representada usando o modelo de Wiener-Volterra.

Figura 18 – Base de Kautz como filtro digital.



Fonte: Adaptado de Oroski (2015).

Os parâmetros iniciais do algoritmo de LM são apresentados na Tabela 1, todos esses parâmetros foram ajustados empiricamente. Em seguida foi executado o algoritmo com 20 pontos iniciais diferentes e considerando duas funções de Kautz na base. Os resultados alcançados são mostrados na Tabela 2, em que b_0 e c_0 são os valores iniciais para os parâmetros e b_f e c_f são os parâmetros encontrados pelo algoritmo. Os valores iniciais já atendem as três restrições para minimização.

Tabela 1 – Parâmetros do algoritmo de LM.

Parâmetros	Valores
Iterações	200
Tolerância de Convergência do Gradiente	10^{-4}
Tolerância de Convergência dos Parâmetros	10^{-3}
Tolerância do MSE	10^{-4}
Tamanho Mínimo do Passo	10^{-6}
Valor Inicial de λ	10^{-3}
Fator de Crescimento de λ	10
Fator de Decrescimento de λ	10

Fonte: Autoria Própria.

Tabela 2 – Resultados da otimização com o algoritmo de LM para estimar os polos de Kautz no modelo de Volterra-Kautz.

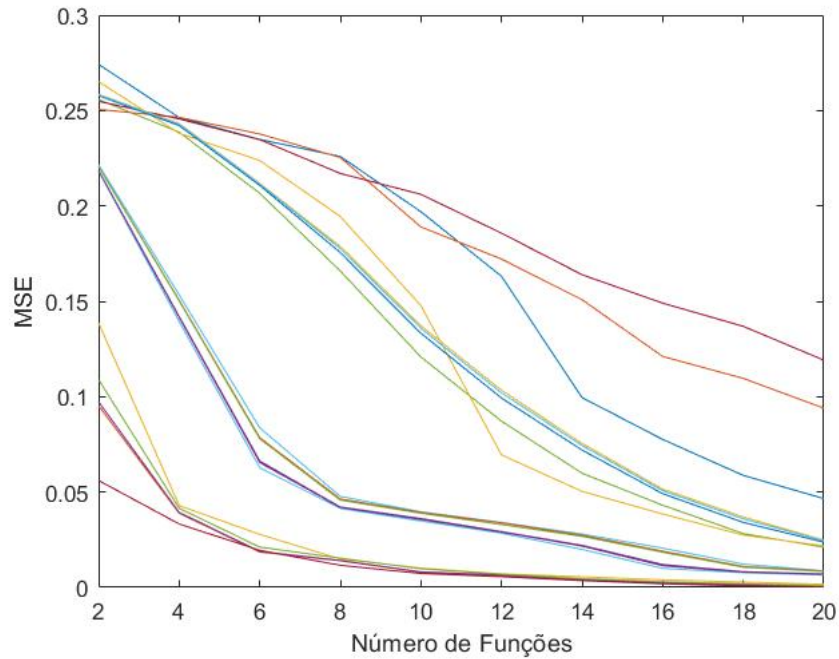
Teste	b_0	c_0	b_f	c_f	Polos	MSE
1	0	-0,300	0,270	-0,060	$0,140 \pm i0,190$	0,218
2	-0,300	-0,200	0,900	-0,400	$0,630 \pm i0,020$	0,095
3	0,600	-0,150	0,820	-0,280	$0,530 \pm i0,020$	0,139
4	0,370	-0,260	0,890	-0,390	$0,620 \pm i0,010$	0,097
5	-0,620	-0,200	-0,640	-0,130	$-0,360 \pm i0,002$	0,256
6	0,250	-0,430	0,330	-0,090	$0,180 \pm i0,240$	0,218
7	0,920	-0,680	0,950	-0,540	$0,740 \pm i0,050$	0,056
8	-0,650	-0,750	-0,490	-0,860	$-0,450 \pm i0,810$	0,274
9	-0,600	-0,600	-0,100	-0,920	$-0,100 \pm i0,950$	0,251
10	0,500	-0,900	0,460	-0,840	$-0,430 \pm i0,810$	0,256
11	-0,200	-0,100	0,270	-0,060	$0,140 \pm i0,190$	0,219
12	0,850	-0,400	0,880	-0,360	$0,600 \pm i0,030$	0,109
13	-0,300	-0,350	0,110	-0,020	$0,060 \pm i0,120$	0,221
14	-0,400	-0,300	0,004	-0,680	$-0,003 \pm i0,830$	0,255
15	-0,750	-0,500	-0,69	-0,160	$-0,400 \pm i0,004$	0,258
16	0,800	-0,900	0,100	-0,003	$0,050 \pm i0,020$	0,221
17	-0,700	-0,400	-0,700	-0,170	$-0,410 \pm i0,002$	0,259
18	0,200	-0,400	0,290	-0,070	$0,150 \pm i0,210$	0,219
19	-0,100	-0,500	0,140	0,020	$0,070 \pm i0,120$	0,221
20	-0,800	-0,300	-0,700	-0,170	$-0,410 \pm i0,002$	0,258

Fonte: Autoria Própria.

Como pode ser observado, os pontos iniciais que possuem valores próximos convergiram para polos com valores próximos. Isto ocorre devido ao algoritmo de LM convergir para mínimos locais. Outra observação pode ser feita são os valores de MSE encontrado, pois, os mesmos não são expressivos, uma vez que, na literatura, os resultados da modelagem tornam-se interessantes a partir de valores na ordem de grandeza de 10^{-3} .

Fazendo uso do que é expresso em (4), com o aumento do número das funções que compõem a base, o erro da aproximação tende a diminuir. Sendo assim, foi realizada uma sequência de testes aumentando o número das funções na base de 2 até 20, a partir dos mesmos polos já encontrados pelo LM. Os resultados são expostos na Figura 19 e os valores dos MSE para 20 funções na base são expostos na Tabela 3.

Figura 19 – Comportamento do MSE com aumento das funções de Kautz no modelo Volterra-Kautz utilizando o algoritmo de LM para otimização.



Fonte: Autoria Própria.

Tabela 3 – Resultados do MSE para 20 funções de Kautz na base utilizando o algoritmo de LM para estimar os polos.

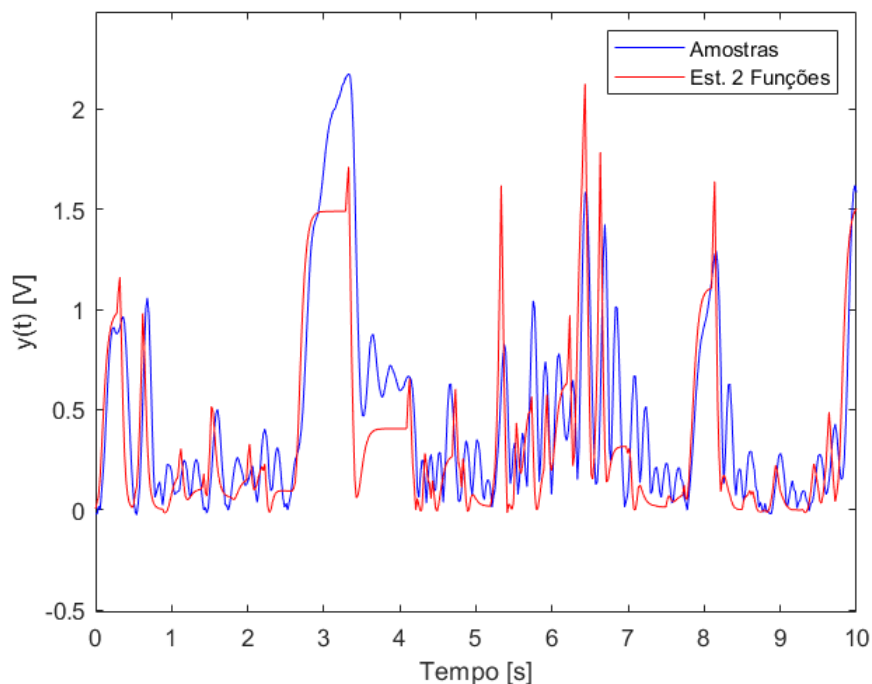
Teste	Polos	MSE
1	$0,140 \pm i0,190$	$7,10 \times 10^{-3}$
2	$0,630 \pm i0,020$	$6,54 \times 10^{-4}$
3	$0,530 \pm i0,020$	$1,60 \times 10^{-3}$
4	$0,620 \pm i0,010$	$6,70 \times 10^{-4}$
5	$-0,360 \pm i0,002$	$2,11 \times 10^{-2}$
6	$0,180 \pm i0,240$	$6,50 \times 10^{-3}$
7	$0,740 \pm i0,050$	$3,88 \times 10^{-4}$
8	$-0,450 \pm i0,810$	$4,66 \times 10^{-2}$
9	$-0,100 \pm i0,950$	$9,40 \times 10^{-2}$
10	$-0,430 \pm i0,810$	$2,18 \times 10^{-2}$
11	$0,140 \pm i0,190$	$7,10 \times 10^{-3}$
12	$0,600 \pm i0,030$	$8,87 \times 10^{-4}$
13	$0,060 \pm i0,120$	$8,70 \times 10^{-3}$
14	$-0,003 \pm i0,830$	$1,19 \times 10^{-1}$
15	$-0,400 \pm i0,004$	$2,37 \times 10^{-2}$
16	$0,050 \pm i0,020$	$8,60 \times 10^{-3}$
17	$-0,410 \pm i0,002$	$2,48 \times 10^{-2}$
18	$0,150 \pm i0,210$	$6,90 \times 10^{-3}$
19	$0,070 \pm i0,120$	$8,40 \times 10^{-3}$
20	$-0,410 \pm i0,002$	$2,44 \times 10^{-2}$

Fonte: Autoria Própria.

Confrontando os valores exibidos nas Tabelas 2 e 3, nota-se a melhora de todos os valores de MSE dos experimentos. Pode-se observar que em alguns casos, como nos testes 4, 6, 10, 19 e 20, a modelagem não conseguiu chegar a MSE inferiores a 10^{-2} , este fato ressalta a importância dos parâmetros iniciais para o algoritmo de LM que ficou preso em um mínimo local não satisfatório, levando a uma modelagem pouco representativa do sistema.

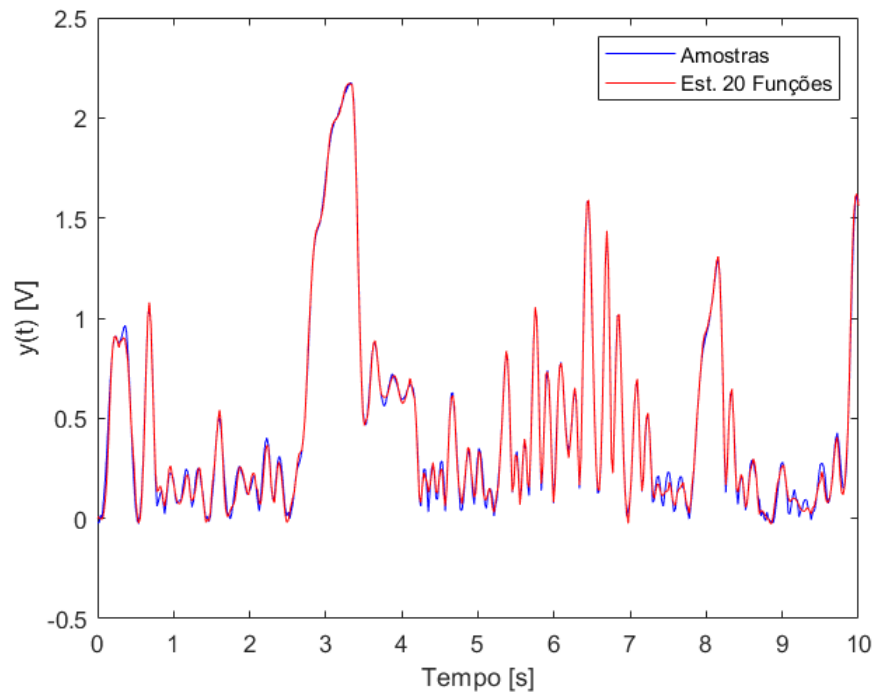
Para um aprofundamento das análises, utilizaram-se os dados do teste 7. Nas Figuras 20 e 21 são ilustrados os resultados das modelagens com 2 e com 20 funções de Kautz, respectivamente. Como pode ser observado, a estrutura com duas funções teve dificuldade para modelar as oscilações nos intervalos de 1 s a 3 s e 5 s a 7 s. Com o uso de mais funções, aumentou-se a capacidade de representação e estes problemas foram amenizados. Contudo, visualmente, é possível notar que para as representações dos picos das oscilações consecutivas, como ocorre no intervalo de 7 s a 8 s e 9 s a 10 s, mesmo com 20 funções, o modelo apresentou um desvio em relação aos dados coletados. Apesar de melhorar a capacidade da modelagem, o aumento do número de funções na base também implica no aumento da quantidade de parâmetros a serem estimados, ou seja, sempre irá ocorrer um *trade-off* entre quantidade de parâmetros a serem estimados e a qualidade do modelo.

Figura 20 – Resposta do Modelo de Volterra-Kautz de um sistema composto por dois motores e uma polia, com kernels modelados por 2 funções de Kautz, utilizando otimização por método de LM.



Fonte: Autoria Própria.

Figura 21 – Resposta do Modelo de Volterra-Kautz de um sistema composto por dois motores e uma polia, com *kernels* modelados por 20 funções de Kautz, utilizando otimização por método de LM.

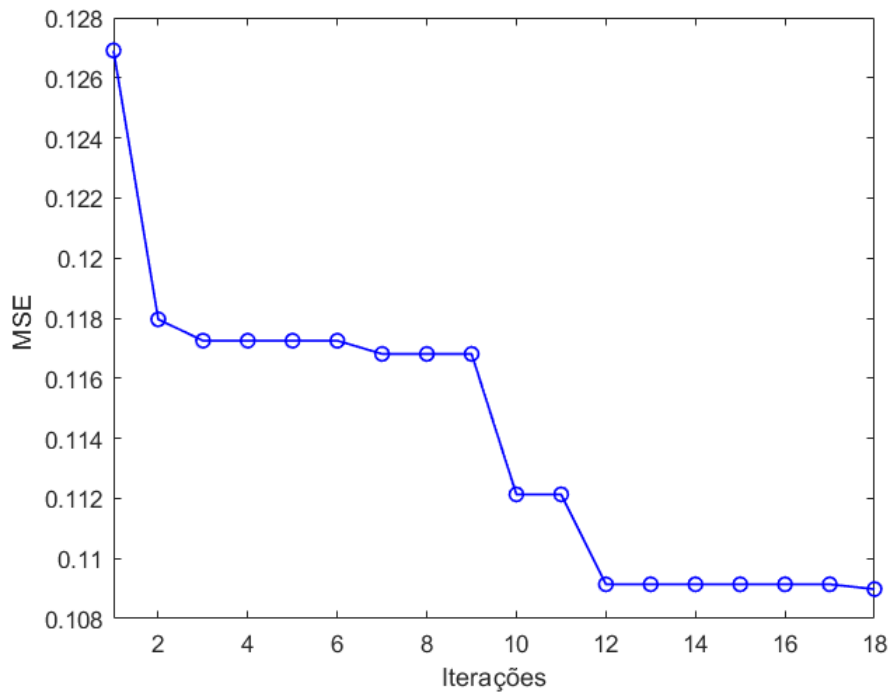


Fonte: Autoria Própria.

Outro comportamento que pode ser analisado é a dinâmica do algoritmo de LM na busca pelos pontos no espaço de parâmetros. Na Figura 22 pode ser visto o comportamento do MSE durante essa busca. Nota-se que o algoritmo não executou as 200 iterações, sendo o critério de convergência do parâmetro atingido, ou seja, a diferença entre os parâmetros encontrados na iteração $n - 1$ e os parâmetros encontrados na iteração n foi inferior a 10^{-3} .

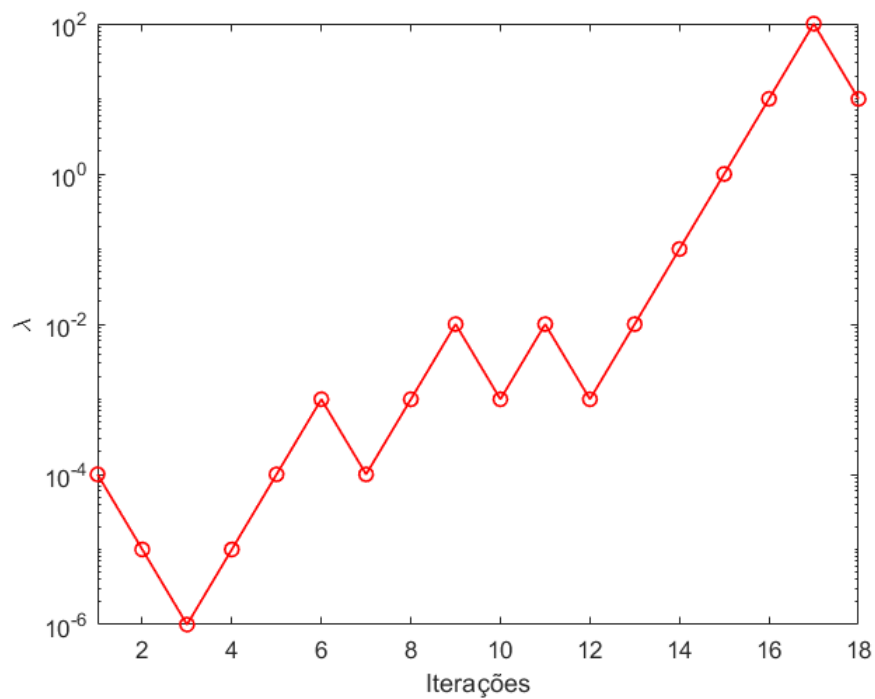
Ao ser contrastado o resultado da Figura 22 com o da Figura 23, que respresenta o comportamento do parâmetro λ durante a busca, pode-se notar que durante as primeiras três iterações e entre a 9ª e 12ª iterações o MSE decresce rapidamente. Como consequência, o parâmetro λ tende a diminuir, levando o algoritmo há um comportamento semelhante ao do método de Gauss-Newton. Já nas outras iterações, a variação do MSE não é muito expressiva, o que levou há um crescimento no valor do λ . Desta forma, o algoritmo caminhou com passos pequenos quando a curvatura do MSE foi acentuada e com largos passos quando o MSE teve pouca variação.

Figura 22 – Comportamento do MSE durante a busca pelos polos de Kautz no modelo de Volterra-Kautz utilizando o algoritmo de LM para otimização.



Fonte: Autoria Própria.

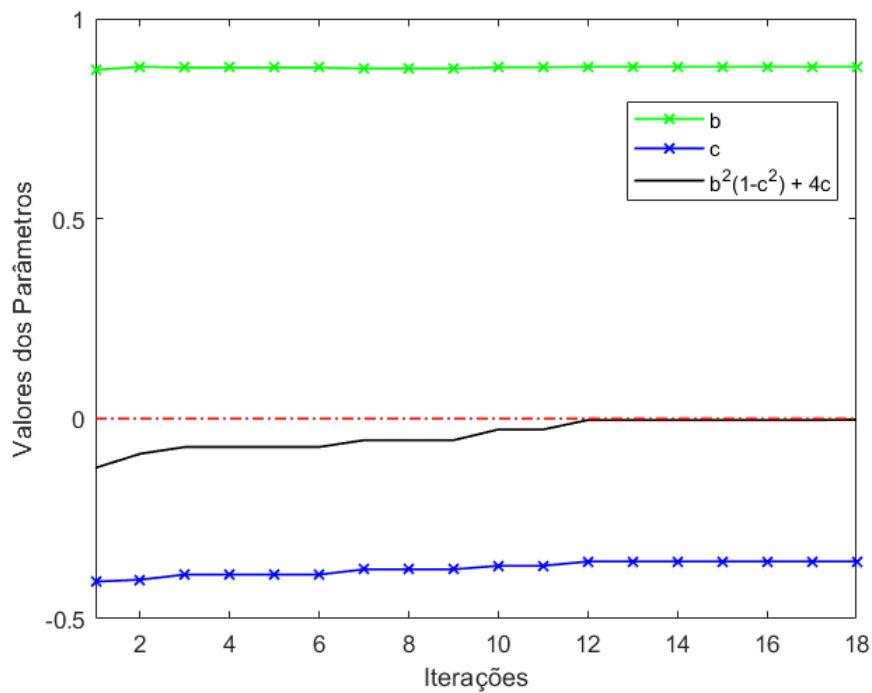
Figura 23 – Comportamento do parâmetro λ durante a busca pelos polos de Kautz no modelo de Volterra-Kautz utilizando o algoritmo de LM para otimização.



Fonte: Autoria Própria.

Pode-se analisar também como os parâmetros se comportaram dentro do espaço de busca, a Figura 24 ilustra este comportamento. Como pode ser observado, para o caso particular do teste 12, os parâmetros não tiveram uma grande variação do valor inicial. Isto decorre devido às restrições que foram impostas ao algoritmo e, como foi exposto na Figura 24, o algoritmo otimizou os parâmetros sem desrespeitar nenhuma das três restrições. Também é notório que foram otimizados os parâmetros até o ponto máximo da terceira restrição, expressa na equação (99), que por sua vez, já estava dentro do espaço de busca da primeira restrição, expressa na equação (97).

Figura 24 – Comportamento dos parâmetros b e c durante a execução da otimização com o algoritmo de LM para encontrar os polos de Kautz no modelo de Volterra-Kautz.



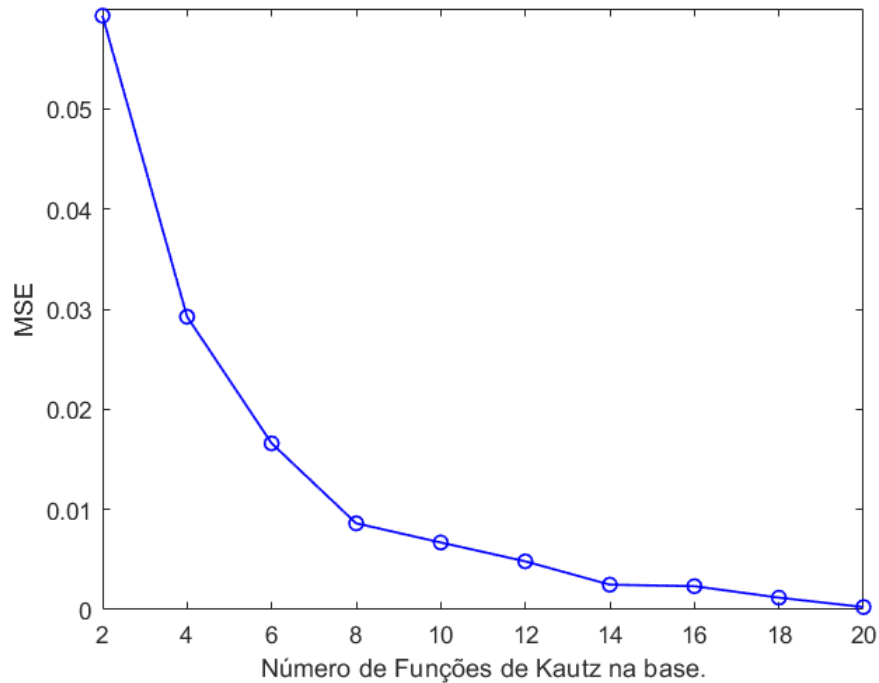
Fonte: Autoria Própria.

O mesmo conjunto de dados foi utilizado para os experimentos com o algoritmo de OB implementado no MATLAB®. Como função de aquisição, foi utilizada a EI por considerar a magnitude da melhoria do ponto estimado. Os parâmetros do algoritmo são expostos na Tabela 4. O teste foi realizado aplicando o algoritmo no conjunto de dados 10 vezes e, em cada uma das vezes, aumentou-se duas funções na base, indo de 2 a 20 funções de Kautz. O resultado do MSE, em cada um dos cenários, são apresentados na Figura 25.

Tabela 4 – Parâmetros do algoritmo de OB.

Parâmetros	Valores
Iterações	40
Tolerância do MSE	10^{-6}

Fonte: Autoria Própria.

Figura 25 – Comportamento do MSE durante o aumento das funções de Kautz na base do modelo de Volterra-Kautz utilizando o algoritmo de OB para encontrar os polos.

Fonte: Autoria Própria.

Como já era esperado, com o aumento do número de funções na base, o MSE tende a diminuir. Os valores dos parâmetros b , c e, conseqüentemente, os valores dos polos encontrados são mostrados na Tabela 5. Pode-se notar que, diferentemente dos polos encontrados com o LM, não há grandes variações entre os valores. Isso ocorre porque a OB não sofre influência de um valor inicial de polo para a otimização.

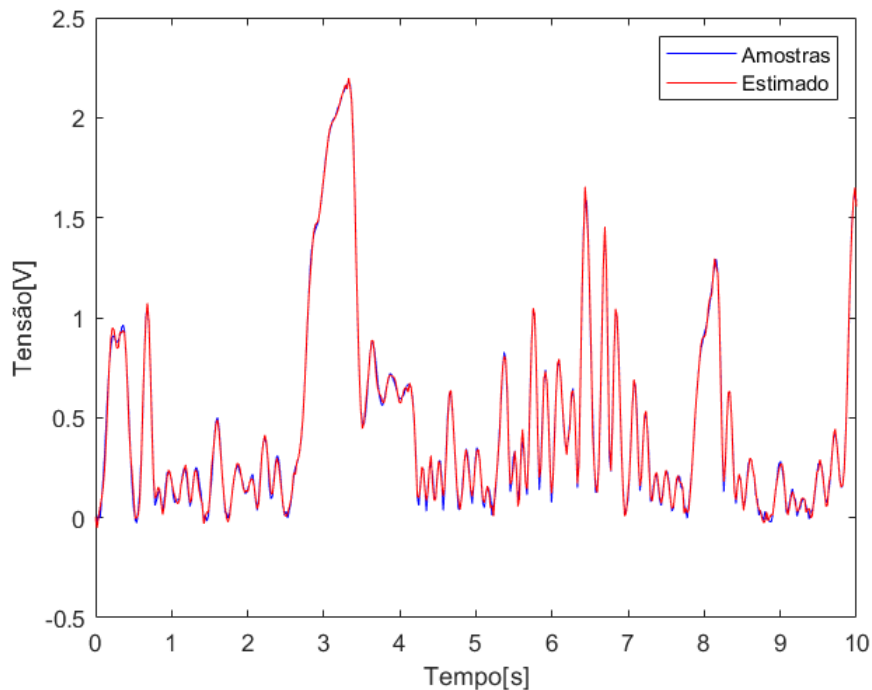
Outra observação que pode ser feita é a quantidade de iterações necessárias para obter os resultados da Tabela 5. Na OB com 40 iterações já se consegue resultados de mesma ordem de grandeza aos resultados do LM, se os valores de MSE forem comparados.

Para efeito analítico, foi considerado um modelo com 20 funções na base e utilizada a OB para estimar os polos de Kautz. O resultado da modelagem é exibido na Figura 26. Como pode ser observado, a resposta gráfica é tão interessante quanto o resultado do LM e conseguiu atingir um MSE de $2,00 \times 10^{-4}$.

Tabela 5 – Resultados do algoritmo de OB na estimação dos polos de Kautz.

Funções	b	c	Polo	MSE
2	0,946	-0,369	$0,648 \pm i0,226$	$5,90 \times 10^{-2}$
4	0,983	-0,744	$0,857 \pm i0,096$	$2,90 \times 10^{-2}$
6	0,916	-0,505	$0,689 \pm i0,173$	$1,70 \times 10^{-2}$
8	0,928	-0,681	$0,780 \pm i0,270$	$9,00 \times 10^{-3}$
10	0,907	-0,584	$0,719 \pm i0,260$	$7,00 \times 10^{-3}$
12	0,906	-0,568	$0,711 \pm i0,251$	$5,00 \times 10^{-3}$
14	0,877	-0,578	$0,692 \pm i0,315$	$3,00 \times 10^{-3}$
16	0,894	-0,432	$0,640 \pm i0,150$	$2,00 \times 10^{-3}$
18	0,928	-0,397	$0,648 \pm i0,151$	$1,00 \times 10^{-3}$
20	0,921	-0,687	$0,777 \pm i0,289$	$2,00 \times 10^{-4}$

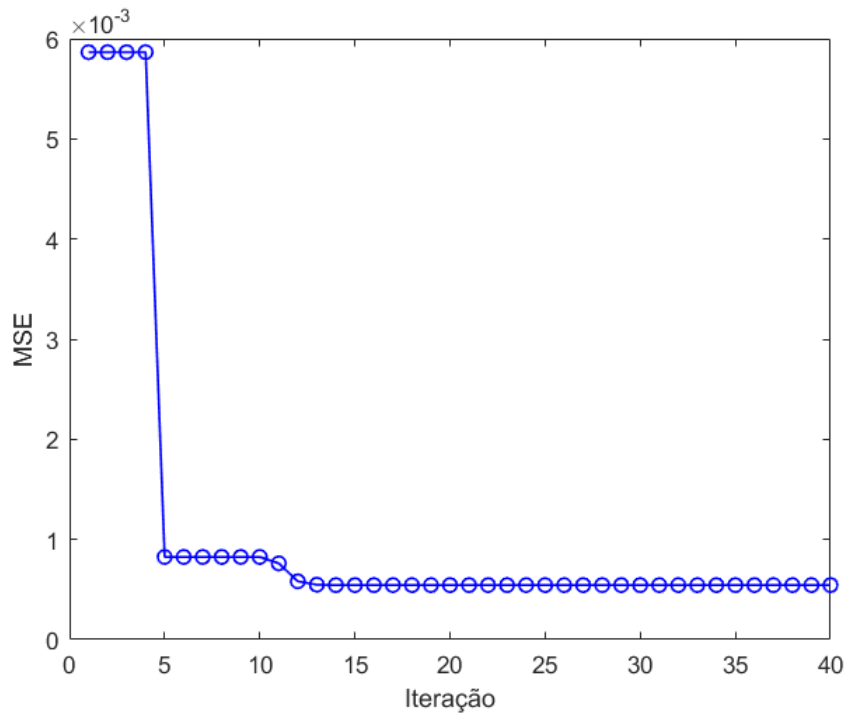
Fonte: Autoria Própria.

Figura 26 – Resposta do modelo de Volterra-Kautz de um sistema composto por dois motores e uma polia com kernels modelados por 20 funções de Kautz, utilizando OB.

Fonte: Autoria Própria.

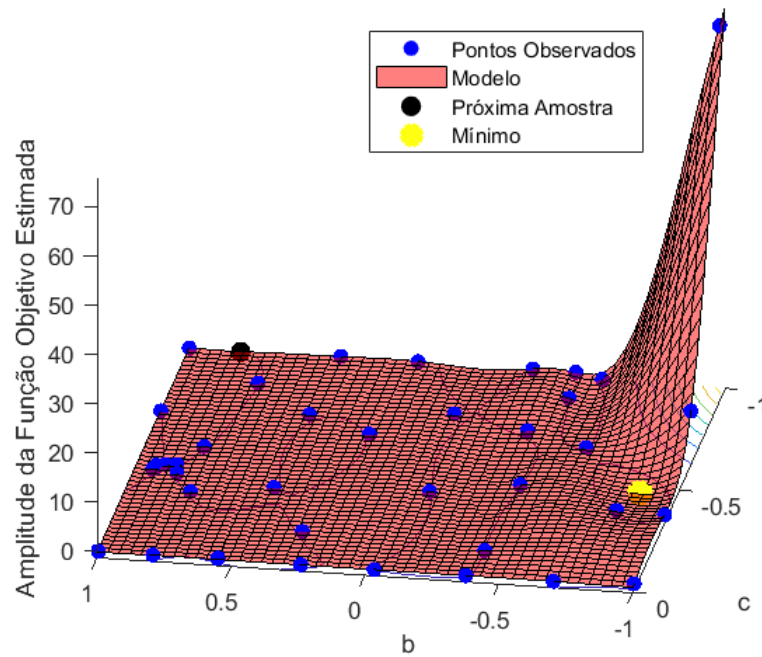
O comportamento do MSE durante a busca evidencia que o algoritmo convergiu rapidamente para um mínimo local e ficou preso nessa região, como é possível concluir a partir da Figura 27. Essa conclusão é reforçada quando o gráfico do MSE é contrastado com a Figura 28 que apresenta o modelo da busca do algoritmo de OB e os pontos observados. Como se nota, os vários pontos observados não conseguiram gerar uma melhoria significativa no modelo e produziram uma região quase plana de pontos e com poucas variações de resultados.

Figura 27 – Comportamento do MSE durante a busca pelos polos de Kautz no modelo de Volterra-Kautz utilizando o algoritmo de OB.



Fonte: Autoria Própria.

Figura 28 – Modelo gerado durante a OB.



Fonte: Autoria Própria.

A fim de comparar os resultados alcançados com as abordagens propostas neste trabalho foi elaborada a Tabela 6, na qual podem ser vistos os valores dos MSEs atingidos por outros dois trabalhos encontradas na literatura que utilizaram o mesmo *benchmark* disponibilizado por Wigren e Schoukens (2017).

Tabela 6 – Comparação entre os resultados das modelagens do *benchmark* com diferentes métodos.

Metodologia	MSE
Volterra-Kautz (LM)	$2,44 \times 10^{-4}$
Volterra-Kautz (OB)	$2,00 \times 10^{-4}$
MOODE-NARMAX	$1,35 \times 10^{-2}$
RBFNNs	$1,60 \times 10^{-3}$

Fonte: Autoria Própria.

Em comparação com os dois trabalhos disponíveis na literatura, as duas estratégias propostas atingiram resultados mais promissores, na ordem de 10^{-4} , sendo esse valor uma ordem de grandeza menor que em relação ao trabalho de Ayala *et al.* (2014), 10^{-3} , e duas ordens de grandeza menor em relação ao trabalho de Zakaria *et al.* (2021), 10^{-2} . Já a comparação entre as técnicas do presente trabalho, o uso da OB, em termos de MSE, proporcionou um resultado 22% melhor se comparado com o uso do LM.

6 CONSIDERAÇÕES FINAIS

A modelagem de sistemas dinâmicos com a série de Volterra é recorrente na literatura, pois esta estrutura matemática permite a descrição de uma gama de modelos lineares e não-lineares. No entanto, o uso desta série implica na estimação de vários parâmetros, principalmente para sistemas não-lineares. Uma estratégia presente na literatura é a utilização de bases de funções ortonormais para estimar o *kernel* do modelo Volterra, resultando em uma redução significativa em termos a serem estimados. As funções ortonormais aplicáveis neste contexto variam segundo as necessidades da dinâmica do sistema a ser modelado, uma vez que essas funções permitem a inserção de conhecimentos *a priori* no modelo. As principais funções utilizadas são as de Laguerre e Kautz, sendo as funções de Laguerre indicadas para sistemas superamortecidos e as funções de Kautz para sistemas subamortecidos. Tais características tornam essas funções interessantes para serem utilizadas na aproximação de *kernels* de Volterra, pois é possível modelar com um número menor de parâmetros em relação ao *kernel* real, gerando assim o conhecido modelo de Volterra-Kautz.

A dificuldade presente nos modelos de Volterra-Kautz é determinar o polo ótimo das funções para a expansão do *kernel*. Para funções de Laguerre, já existe uma solução analítica na literatura para encontrar o polo ótimo que parametriza a função. No entanto, para as funções de Kautz, esse problema ainda está em aberto. No presente trabalho, o algoritmo de LM e a OB foram propostos para encontrar os dois termos que parametrizam os polos de Kautz. Para validar a metodologia proposta, foram utilizados dados de um sistema de motores acoplados a uma correia e medidas de velocidade de oscilação de uma polia fixada com uma mola. Em seguida, os algoritmos de LM e de OB foram implementados no MATLAB® para realizar a otimização do MSE entre o valor medido e a estimativa de saída do modelo de Wiener-Volterra, com os valores de b e c encontrados.

Para a otimização dos parâmetros do modelo de Volterra-Kautz, utilizando o LM, foi necessário determinar os valores iniciais dos parâmetros e, posteriormente, o algoritmo atualizou esses valores de acordo com o gradiente até a otimização. O comportamento deste algoritmo está fortemente ligado as suas condições iniciais, como pode ser analisado nas Tabelas 2 e 3. Devido a isso, foram testados 20 pontos encontrando valores de MSE na ordem de grandeza de 10^{-3} . Outro fator que impactou no resultado da representação foi a quantidade de funções na base, visto que com o aumento das funções, aumentou-se a capacidade de representação e melhorou-se

o resultado da identificação, diminuindo o MSE. Contudo, é importante ressaltar que ao se aumentar o número de funções, aumenta-se a quantidade de parâmetros a serem estimados. Nos experimentos realizados, ao aumentar até 20 funções na base, diversos experimentos conseguiram uma modelagem com resultado de MSE na ordem de 10^{-4} , como são os casos dos testes 2, 4, 7 e 12 na Tabela 3.

No caso do algoritmo da OB, não foi necessário especificar os pontos iniciais por se tratar de um método que usa o processo gaussiano para gerar os pontos a serem observados. Neste caso, foi preciso definir uma função de aquisição e a EI foi a função escolhida, pois considera a magnitude da melhoria que o ponto pode proporcionar na busca pelo mínimo. Assim como para o caso de LM, o aumento das funções na base fizeram com que o resultado do MSE melhorasse, conseguindo atingir valores na ordem de 10^{-4} . Em relação ao LM, esse método apresentou um melhor resultado se comparado o número de iterações necessárias para a otimização atingir um resultado de MSE inferior a 10^{-3} e, conseguiu modelar o mesmo sistema com um MSE 22% menor. Contudo, algumas melhorias podem ser feitas para evitar que o algoritmo fique preso em mínimos locais rapidamente, ficando como uma sugestão para trabalhos futuros.

Os resultados apresentados neste documento apontam que as estratégias propostas são alternativas viáveis para tratar o problema da estimação dos polos ótimos de Kautz, pois, possibilitaram a modelagem do *benchmark* analisado, em termo de MSE, na ordem de 10^{-1} a 10^{-2} menor do que outros dois trabalhos da literatura que modelaram o mesmo conjunto de dados.

Como trabalhos futuros espera-se explorar algoritmos de otimização metaheurísticos, como *Ant Colony Optimization* e *Particle Swarm Optimization*, que podem trazer melhorias na forma de realização da busca dos parâmetros b e c . Outra estratégia a ser estudada é a aplicação de diferentes tipos de função de covariância na otimização bayesiana, como a função de Matérn, ou ainda, explorar diferentes funções de aquisição, como *Entropy Search* e *Knowledge Gradient*. Para o algoritmo de LM pode ser sugerido como trabalho futuro o estudo de diferentes formas de atualização do parâmetro λ , além da forma clássica tratada neste trabalho, e analisar o impacto sobre a busca dos parâmetros b e c das funções de Kautz.

6.1 PUBLICAÇÕES

Durante o processo de elaboração deste trabalho, o seguinte artigo foi publicado em periódico:

- SERAFIN, Higor S; OROSKI, Elder; LAZZARETTI, André E. Estimation of Kautz Poles in Wiener-Volterra Models Using Levenberg-Marquardt Algorithm. **Learning & NonLinear Models journal, Special Issue - Applications of Computational Intelligence in Emerging Electrical Systems, 2022.**

REFERÊNCIAS

AGUIRRE, L. A. **Introdução à Identificação de Sistemas**. 3. ed. [S.l.]: UFMG, 2007.

AKÇAY, Hüseyin; HEUBERGER, Peter. A frequency-domain iterative identification algorithm using general orthonormal basis functions. **Automatica**, Elsevier, v. 37, n. 5, p. 663–674, 2001.

ARONSZAJN, Nachman. Theory of reproducing kernels. **Transactions of the American mathematical society**, JSTOR, v. 68, n. 3, p. 337–404, 1950.

AYALA, Helon Vicente Hultmann; CRUZ, Luciano F da; FREIRE, Roberto Z; COELHO, Leandro dos Santos. Cascaded free search differential evolution applied to nonlinear system identification based on correlation functions and neural networks. *In: IEEE. 2014 IEEE Symposium on Computational Intelligence in Control and Automation (CICA)*. [S.l.], 2014. p. 1–7.

BELT, Harm J.W. **Orthonormal Bases for Adaptive Filtering**. [S.l.: s.n.], 1997.

BODIN, Per; WAHLBERG, Bo. Thresholding in high order transfer function estimation. *In: IEEE. Proceedings of 1994 33rd IEEE Conference on Decision and Control*. [S.l.], 1994. v. 4, p. 3400–3405.

BRAGA, Márcio Feliciano *et al.* Modelos de volterra: identificação não paramétrica e robusta utilizando funções ortonormais de kautz e generalizadas. [sn], 2011.

BRINKER, Albertus C den. Laguerre-domain adaptive filters. **IEEE transactions on signal processing**, IEEE, v. 42, n. 4, p. 953–956, 1994.

BROCHU, Eric; CORA, Vlad M; FREITAS, Nando De. A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. **arXiv preprint arXiv:1012.2599**, 2010.

BROOME, Paul W. Discrete orthonormal sequences. **Journal of the ACM (JACM)**, ACM New York, NY, USA, v. 12, n. 2, p. 151–168, 1965.

CALVIN, James; ŽILINSKAS, A. On the convergence of the p-algorithm for one-dimensional global optimization of smooth functions. **Journal of Optimization Theory and Applications**, Springer, v. 102, n. 3, p. 479–495, 1999.

CALVIN, JAMES M; ZILINSKAS, ANTANAS. A one-dimensional p-algorithm with convergence rate $o(n^{-3})$ for smooth functions. Citeseer, 2003.

CAMPELLO, Ricardo JGB; OLIVEIRA, Gustavo HC; AMARAL, Wagner C. Identificação e controle de processos via desenvolvimentos em séries ortonormais. parte a: Identificação. **Sba: Controle & Automação Sociedade Brasileira de Automatica**, SciELO Brasil, v. 18, n. 3, p. 301–321, 2007.

CAMPELLO, R. J. G. B.; OLIVEIRA, G. H. C. Modelos não lineares. *In: _____. [S.l.]*: Edgard Blücher, 2007. cap. 4.

CHAN, Lai-Wan; SZETO, Chi-Cheong. Training recurrent network with block-diagonal approximated levenberg-marquardt algorithm. *In: IEEE. IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No. 99CH36339). [S.l.]*, 1999. v. 3, p. 1521–1526.

CHATTERJEE, Animesh. Identification and parameter estimation of a bilinear oscillator using volterra series with harmonic probing. **International Journal of Non-Linear Mechanics**, Elsevier, v. 45, n. 1, p. 12–20, 2010.

CHATTERJEE, Animesh; VYAS, Nalinaksh S. Non-linear parameter estimation with volterra series using the method of recursive iteration through harmonic probing. **Journal of Sound and Vibration**, Elsevier, v. 268, n. 4, p. 657–678, 2003.

CHATTERJEE, Animesh; VYAS, Nalinaksh S. Nonlinear parameter estimation in rotor-bearing system using volterra series and method of harmonic probing. **J. Vib. Acoust.**, v. 125, n. 3, p. 299–306, 2003.

CHATTERJEE, Animesh; VYAS, Nalinaksh S. Non-linear parameter estimation in multi-degree-of-freedom systems using multi-input volterra series. **Mechanical Systems and Signal Processing**, Elsevier, v. 18, n. 3, p. 457–489, 2004.

CHEN, Y; LIU, ZY; ZHANG, HJ. Research and simulation of volterra series kernel identification on wiener model. *In: IEEE. 2010 The 2nd Conference on Environmental Science and Information Application Technology. [S.l.]*, 2010. v. 4, p. 95–98.

CHENG, CM; PENG, ZK; ZHANG, WM; MENG, Guang. Volterra-series-based nonlinear system modeling and its engineering applications: A state-of-the-art review. **Mechanical Systems and Signal Processing**, Elsevier, v. 87, p. 340–364, 2017.

CHOU, Chun Tung; VERHAEGEN, Michel; JOHANSSON, Rolf. Continuous-time identification of siso systems using laguerre functions. **IEEE Transactions on Signal Processing**, IEEE, v. 47, n. 2, p. 349–362, 1999.

COLEMAN, Thomas F; LI, Yuying. An interior trust region approach for nonlinear minimization subject to bounds. **SIAM Journal on optimization**, SIAM, v. 6, n. 2, p. 418–445, 1996.

CORA, Vlad M. **Model-based active learning in hierarchical policies**. 2008. Tese (Doutorado) — University of British Columbia, 2008.

CUNHA, Joao Carlos Vieira da. O método dos mínimos quadrados: uma proposta ao ensino médio para o ajuste por retas. 2014.

DAN, Hiroshige; YAMASHITA, Nobuo; FUKUSHIMA, Masao. Convergence properties of the inexact levenberg-marquardt method under local error bound conditions. **Optimization methods and software**, Taylor & Francis, v. 17, n. 4, p. 605–626, 2002.

EYKHOFF, P. **System Identification: Parameter and State Estimation**. Wiley, 1979. (Wiley-Interscience Publication). Disponível em: <https://books.google.com.br/books?id=nVI7ngEACAAJ>.

FAN, Jin-yan. A modified levenberg-marquardt algorithm for singular system of nonlinear equations. **Journal of Computational Mathematics**, JSTOR, p. 625–636, 2003.

FAN, Jin-yan; YUAN, Ya-xiang. On the quadratic convergence of the levenberg-marquardt method without nonsingularity assumption. **Computing**, Springer, v. 74, n. 1, p. 23–39, 2005.

FETTWEIS, A. On the factorization of transfer matrices of lossless two-ports. **IEEE Transactions on Circuit Theory**, IEEE, v. 17, n. 1, p. 86–94, 1970.

FETTWEIS, Alfred. Digital filter structures related to classical filter networks. **Arch. Elektr. Übertragung**, v. 25, n. 2, p. 79–89, 1971.

FU, Y.; DUMONT, G.A. An optimum time scale for discrete laguerre network. *In: IEEE TRANSACTIONS AUTOMATIC CONTROL*. [S.l.: s.n.], 1993.

GINSBOURGER, David; RICHE, Rodolphe Le; CARRARO, Laurent. A multi-points criterion for deterministic parallel global optimization based on kriging. *In: NCP07*. [S.l.: s.n.], 2007.

GRUBER, JK; BORDONS, Carlos; OLIVA, A. Nonlinear mpc for the airflow in a pem fuel cell using a volterra series model. **Control Engineering Practice**, Elsevier, v. 20, n. 2, p. 205–217, 2012.

GUNNARSSON, Svante; WAHLBERG, Bo. Some asymptotic results in recursive identification using laguerre models. **International journal of adaptive control and signal processing**, Wiley Online Library, v. 5, n. 5, p. 313–333, 1991.

HEUBERGER, Peter SC; HOF, Paul MJ van den; WAHLBERG, Bo. **Modelling and identification with rational orthogonal basis functions**. [S.l.]: Springer Science & Business Media, 2005.

IKEHARA, Shikao. A method of wiener in a nonlinear circuit. Research Laboratory of Electronics, Massachusetts Institute of Technology, 1951.

ISERMANN, Rolf; MÜNCHHOF, Marco. **Identification of Dynamic Systems**. [S.l.]: Springer, 2011.

JING, Xingjian; LANG, Ziqiang; BILLINGS, Stephen A. Determination of the analytical parametric relationship for output spectrum of volterra systems based on its parametric characteristics. **Journal of mathematical analysis and applications**, Elsevier, v. 351, n. 2, p. 694–706, 2009.

JONES, Donald R. A taxonomy of global optimization methods based on response surfaces. **Journal of global optimization**, Springer, v. 21, n. 4, p. 345–383, 2001.

JONES, Donald R; SCHONLAU, Matthias; WELCH, William J. Efficient global optimization of expensive black-box functions. **Journal of Global optimization**, Springer, v. 13, n. 4, p. 455–492, 1998.

KAUTZ, W. Transient synthesis in the time domain. **IRE Transactions on Circuit Theory**, 1954.

KAUTZ, William H. Network synthesis for specified transient response. Massachusetts Institute of Technology, Research Laboratory of Electronics, 1952.

KEANE, Andy J. Statistical improvement criteria for use in multiobjective design optimization. **AIAA journal**, v. 44, n. 4, p. 879–891, 2006.

KING, RE; PARASKEVOPOULOS, PN. Digital laguerre filters. **International Journal of Circuit Theory and Applications**, Wiley Online Library, v. 5, n. 1, p. 81–91, 1977.

KNOWLES, Joshua. Parego: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. **IEEE Transactions on Evolutionary Computation**, IEEE, v. 10, n. 1, p. 50–66, 2006.

KORENBERG, Michael J; HUNTER, Ian W. The identification of nonlinear biological systems: Volterra kernel approaches. **Annals of biomedical engineering**, Springer, v. 24, n. 2, p. 250–268, 1996.

KUSHNER, Harold J. A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. 1964.

LEE, Yuk-Wing. Synthesis of electric networks by means of the fourier transforms of laguerre's functions. **Journal of Mathematics and Physics**, Wiley Online Library, v. 11, n. 1-4, p. 83–113, 1932.

LEMMA, DT; RAMASAMY, M; SHUHAIMI, M. System identification using orthonormal basis filters. **Journal of Applied Sciences(Faisalabad)**, Asian Network for Scientific Information, 308-Lasani Town Sargodha Rd Faisalabad 38090 Pakistan, v. 10, n. 21, p. 2516–2522, 2010.

LEVENBERG, K. A method for the solution of certain problems in least squares. **Quart. Appl. Math. Vol 2**, 1944.

LIZOTTE, Daniel James. Practical bayesian optimization. 2008.

MA, Changfeng; JIANG, Lihua. Some research on levenberg–marquardt method for the nonlinear equations. **Applied mathematics and Computation**, Elsevier, v. 184, n. 2, p. 1032–1040, 2007.

MA, Changfeng; TANG, Jia. The quadratic convergence of a smoothing levenberg–marquardt method for nonlinear complementarity problem. **Applied mathematics and computation**, Elsevier, v. 197, n. 2, p. 566–581, 2008.

MA, Changfeng; TANG, Jia; CHEN, Xiaohong. A globally convergent levenberg–marquardt method for solving nonlinear complementarity problem. **Applied mathematics and computation**, Elsevier, v. 192, n. 2, p. 370–381, 2007.

MACHADO, Jeremias Barbosa *et al.* Modelagem de sistemas não-lineares por base de funções ortonormais generalizadas com funções internas. [sn], 2011.

MACHADO, R. CAMPELLO J.; AMARAL, W. Modelos lineares gobf com funções internas. *In: VIII Congresso Brasileiro de Automática. [S.l.: s.n.]*, 2010.

MALMQUIST, F. Sur la determination d'une classe de fonctions analytiques par leurs valeurs dans un ensemble donné de points. **Comptes Rendus du Sixième Congrès des Mathématiciens Scandinaves**, 1925.

MARINELLI, Maura Ferreira *et al.* Métodos de quadrados mínimos. 2002.

MARMARELIS, Vasilis Z. Identification of nonlinear biological systems using laguerre expansions of kernels. **Annals of biomedical engineering**, Springer, v. 21, n. 6, p. 573–589, 1993.

MARQUARDT, D. An algorithm for least-squares estimation of nonlinear parameters. **SIAM J. Appl. Math Vol 1**, 1963.

MASNADI-SHIRAZI, Mohammad A; AHMED, Nasir. Optimum laguerre networks for a class of discrete-time systems. **IEEE Transactions on Signal Processing**, IEEE, v. 39, n. 9, p. 2104–2108, 1991.

MASTROMAURO, RA; LISERRE, M; DELL'AQUILA, A. Frequency domain analysis of inductor saturation in current controlled grid converters. *In: IEEE. IECON 2007-33rd Annual Conference of the IEEE Industrial Electronics Society. [S.l.]*, 2007. p. 1396–1401.

MEDEIROS, Luiz Antônio da Silva *et al.* **Acelerando o método de Levenberg-Marquardt para a minimização da soma de quadrados de funções com restrições em caixa**. 2008. Tese (Doutorado) — UNIVERSIDADE ESTADUAL DE CAMPINAS, 2008.

MENDEL, Jerry M. A unified approach to the synthesis of orthonormal exponential functions useful in systems analysis. **IEEE Transactions on Systems Science and Cybernetics**, IEEE, v. 2, n. 1, p. 54–62, 1966.

MIRRI, D; IUCULANO, G; TRAVERSO, PA; PASINI, G; FILICORI, F. Non-linear dynamic system modelling based on modified volterra series approaches. **Measurement**, Elsevier, v. 33, n. 1, p. 9–21, 2003.

MOČKUS, Jonas. On bayesian methods for seeking the extremum. *In: SPRINGER. Optimization techniques IFIP technical conference. [S.l.]*, 1975. p. 400–404.

MOCKUS, Jonas. The bayesian approach to global optimization. *In: System Modeling and Optimization. [S.l.]*: Springer, 1982. p. 473–481.

MOCKUS, Jonas; TIESIS, Vytautas; ZILINSKAS, Antanas. The application of bayesian methods for seeking the extremum. **Towards global optimization**, v. 2, n. 117-129, p. 2, 1978.

MORÉ, Jorge J. The levenberg-marquardt algorithm: implementation and theory. *In: Numerical analysis. [S.l.]*: Springer, 1978. p. 105–116.

MORÉ, Jorge J. Recent developments in algorithms and software for trust region methods. *In: Mathematical programming The state of the art. [S.l.]*: Springer, 1983. p. 258–287.

MORTENSEN, RE. **Nonlinear System Theory: The Volterra/Wiener Approach. [S.l.]**: JS-TOR, 1983.

NALBANTOĞLU, Volkan; BOKOR, Jozsef; BALAS, Gary; GASPAR, Peter. System identification with generalized orthonormal basis functions: an application to flexible structures. **Control Engineering Practice**, Elsevier, v. 11, n. 3, p. 245–259, 2003.

NINNESS, Brett; GOMEZ, J-C; WELLER, Steven. MIMO system identification using orthonormal basis functions. *In: IEEE. Decision and Control, 1995., Proceedings of the 34th IEEE Conference on. [S.l.]*, 1995. v. 1, p. 703–708.

NINNESS, Brett; GUSTAFSSON, Fredrik. A unifying construction of orthonormal bases for system identification. **IEEE Transactions on automatic control**, IEEE, v. 42, n. 4, p. 515–521, 1997.

NURGES, U. Laguerre models in problems of approximation and identification of discrete-systems. **Automation and Remote Control**, PLENUM PUBL CORP CONSULTANTS BUREAU 233 SPRING ST, NEW YORK, NY 10013, v. 48, n. 3, p. 346–352, 1987.

OROSKI, Elder. Identificação de sistemas não lineares utilizando modelos narx, funções ortonormais e otimização heurística. 2015.

OROSKI, João Roberto Deroco Martins Elder; BAUCHSPIESS, Adolfo. Identificação de um levitador magnético não linear a partir de funções de kautz e algoritmos genéticos. *In: Congresso Nacional de Matemática Aplicada à Indústria. [S.l.: s.n.]*, 2014.

PARTINGTON, Jonathan R. Approximation of delay systems by fourier-laguerre series. **Automatica**, Elsevier, v. 27, n. 3, p. 569–572, 1991.

PEREZ, Hector; TSUJII, Shigeo. A system identification algorithm using orthogonal functions. **IEEE Transactions on signal processing**, IEEE, v. 39, n. 3, p. 752–755, 1991.

RANGANATHAN, Ananth. The levenberg-marquardt algorithm. 2004.

ROSA, Alex da. Desenvolvimento de modelos discretos de volterra usando funções de kautz. [sn], 2005.

RUGH, Wilson J.. **Nonlinear System Theory: The Volterra/Wiener Approach. [S.l.]**: Johns Hopkins University Press, 1981.

SANSONE, Giovanni. Orthogonal functions. Interscience, 1959.

SASENA, Michael James. **Flexibility and efficiency enhancements for constrained global design optimization with kriging approximations. [S.l.]**: University of Michigan, 2002.

SCHETZEN, Martin. The volterra and wiener theories of nonlinear systems. Wiley, 1980.

SNOEK, Jasper; LAROCHELLE, Hugo; ADAMS, Ryan P. Practical bayesian optimization of machine learning algorithms. **Advances in neural information processing systems**, v. 25, 2012.

STRANG, G. **Álgebra Linear e suas Aplicações**. [S.l.]: Cengage Learning, 2010.

TAKENAKA, S. On the orthogonal functions and a new formula of interpolation,. **J. Math.**, 1925.

VAIDYANATHAN, P. A unified approach to orthogonal digital filters and wave digital filters, based on lbr two-pair extraction. **IEEE transactions on circuits and systems**, IEEE, v. 32, n. 7, p. 673–686, 1985.

VOLTERRA, Vito. Theory of functionals and of integral and integro-differential equations. **Bull. Amer. Math. Soc**, v. 38, n. 1, p. 623, 1932.

WAHLBERG, B. System identification using laguerre models. **IEEE TRANSACTIONS ON AUTOMATIC CONTROL**, 1991.

WAHLBERG, Bo. System identification using kautz model. *In: IEEE TRANSACTIONS AUTOMATIC CONTROL*. [S.l.: s.n.], 1994.

WAHLBERG, Bo. Orthonormal basis function models: A transformation analysis. *In: 14th Triennial World Congress*. [S.l.: s.n.], 1999.

WAHLBERG, Bo; HANNAN, Edward J. Parametric signal modelling using laguerre filters. **The Annals of Applied Probability**, JSTOR, p. 467–496, 1993.

WAHLBERG, Bo; MÄKILÄ, P.M. On approximation of stable linear dynamical systems using laguerre and kautz functions. *In: Automatica*. [S.l.: s.n.], 1996.

WALSH, J. L. **Interpolation and Approximation by Rational Functions in the Complex Domain**. [S.l.]: American Mathematical Society, 1935. XX.

WIENER, Norbert. **The Fourier integral and certain of its applications**. [S.l.]: CUP Archive, 1933.

WIENER, N. Response of a nonlinear device to noise, radiation lab. **Massachusetts Inst. Technology, Rept**, v. 129, p. 78, 1942.

WIENER, Norbert. Extrapolation, interpolation and smoothing of stationary. **Time Series, with Engineering Applications**, MIT press, 1949.

WIENER, Norbert. Nonlinear problems in random theory. **Nonlinear Problems in Random Theory**, by Norbert Wiener, pp. 142. ISBN 0-262-73012-X. Cambridge, Massachusetts, USA: The MIT Press, August 1966.(Paper), p. 142, 1966.

WIGREN, Torbjörn; SCHOUKENS, Maarten. **Coupled electric drives data set and reference models**. [S.l.]: Department of Information Technology, Uppsala Universitet, 2017.

WILAMOWSKI, Bogdan M; CHEN, Yixin; MALINOWSKI, Aleksander. Efficient algorithm for training neural networks with one hidden layer. In: IEEE. **IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No. 99CH36339)**. [S.l.], 1999. v. 3, p. 1725–1728.

WRIGHT, J. Nocedal S. **Numerical Optimization**. [S.l.]: New York:Springer, 1999.

WU, Jian; FRAZIER, Peter. The parallel knowledge gradient method for batch bayesian optimization. **Advances in neural information processing systems**, v. 29, 2016.

YAMASHITA, Nobuo; FUKUSHIMA, Masao. On the rate of convergence of the levenberg-marquardt method. In: **Topics in numerical analysis**. [S.l.]: Springer, 2001. p. 239–249.

YOON, Yongsoon; SUN, Zongxuan. Nonlinear identification and robust tracking control of a camless engine valve actuator based on a volterra series representation. In: IEEE. **2014 American Control Conference**. [S.l.], 2014. p. 1535–1540.

YUAN, Ya-xiang. **Trust region algorithms for nonlinear equations**. [S.l.]: Hong Kong Baptist University, Department of Mathematics, 1994.

ZAKARIA, Mohd Zakimi; MANSOR, Zakwan; NOR, Azuwir Mohd; BAHARUDIN, Mohammad Ezral; SAAD, Mohd Sazli. Modeling coupled electric drives systems using a modified narmax model. In: AIP PUBLISHING LLC. **AIP Conference Proceedings**. [S.l.], 2021. v. 2339, n. 1, p. 020015.

ZERVOS, CCPR; BÉLANGER, PR; DUMONT, Guy Albert. On pid controller tuning using orthonormal series identification. **Automatica**, Elsevier, v. 24, n. 2, p. 165–175, 1988.