



**Thaynã da Silva França**

**Engenharia de recursos para lidar com dados  
ruidosos na identificação esparsa sob as  
perspectivas de classificação e regressão**

**Tese de Doutorado**

Tese apresentada como requisito parcial para obtenção do grau de Doutor pelo Programa de Pós-graduação em Engenharia Mecânica, do Departamento de Engenharia Mecânica da PUC-Rio.

Orientador : Prof. Arthur Martins Barbosa Braga  
Coorientador: Prof. Helon Vicente Hultmann Ayala

Rio de Janeiro  
Maio de 2021



**Thaynã da Silva França**

**Engenharia de recursos para lidar com dados  
ruidosos na identificação esparsa sob as  
perspectivas de classificação e regressão**

Tese apresentada como requisito parcial para obtenção do grau de Doutor pelo Programa de Pós-graduação em Engenharia Mecânica da PUC-Rio. Aprovada pela Comissão Examinadora:

**Prof. Arthur Martins Barbosa Braga**

Orientador

Departamento de Engenharia Mecânica – PUC-Rio

**Prof. Helon Vicente Hultmann Ayala**

Coorientador

Departamento de Engenharia Mecânica – PUC-Rio

**Prof. Igor Braga de Paula**

Pontifícia Universidade Católica do Rio de Janeiro – PUC-Rio

**Prof. Ivan Fabio Mota de Menezes**

Pontifícia Universidade Católica do Rio de Janeiro – PUC-Rio

**Prof. Gideon Villar Leandro**

Universidade Federal do Paraná – UFPR

**Prof. Leandro dos Santos Coelho**

Pontifícia Universidade Católica do Paraná – PUCPR

Rio de Janeiro, 13 de Maio de 2021

Todos os direitos reservados. A reprodução, total ou parcial do trabalho, é proibida sem a autorização da universidade, do autor e do orientador.

### **Thaynã da Silva França**

Técnico em Mecânica pela Escola Técnica do Arsenal de Marinha do Rio de Janeiro. Bacharel em Matemática com ênfase em computação gráfica pela Universidade Federal Fluminense. Bacharel em Engenharia Mecânica pela Universidade Federal Fluminense. Mestre em Engenharia Mecânica pelo Instituto Militar de Engenharia.

#### Ficha Catalográfica

França, Thaynã

Engenharia de recursos para lidar com dados ruidosos na identificação esparsa sob as perspectivas de classificação e regressão / Thaynã da Silva França; orientador: Arthur Martins Barbosa Braga; coorientador: Helon Vicente Hultmann Ayala. – 2021.

118 f: il. color. ; 30 cm

Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Mecânica, 2021.

Inclui bibliografia

1. Engenharia Mecânica – Teses. 2. Robustez ao Ruído. 3. Identificação Esparsa. 4. Corrupção Ruidosa. 5. Engenharia de Recursos. 6. Pesquisa em Grade. I. Braga, Arthur M. B. II. Ayala, Helon Vicente H. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Mecânica. IV. Título.

CDD: 621

## Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Em adição, enfatiza-se que esta tese também foi realizada com apoio da Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro - Brasil (FAPERJ)

Por fim, gostaria de explicitar toda minha gratidão às amigas e aos amigos de longa data que, de alguma forma, tornaram minha trajetória como um todo mais leve. Sendo assim, é com muito apreço que exalto os nomes de Alci Jorge Ferreira Peres, Douglas Ferraz de Azevedo, Felipe Santos Gil, João Magnus Bailuni de Bragança, Lucas Carvalho Santos dos Reis, Luiz Felipe Barra Gomes, Luís Sérgio de Carvalho Roque, Mariana Silva Martins Alves, Mona Alves de Carvalho, Raiana Maria Arruda Teixeira, Raissa Tatiana Arruda Teixeira e Thiago Lopes Verbicário dos Santos.



## Resumo

França, Thaynã; Braga, Arthur M. B; Ayala, Helon Vicente H. **Engenharia de recursos para lidar com dados ruidosos na identificação esparsa sob as perspectivas de classificação e regressão.** Rio de Janeiro, 2021. 118p. Tese de Doutorado – Departamento de Engenharia Mecânica, Pontifícia Universidade Católica do Rio de Janeiro.

Os sistemas dinâmicos desempenham um papel crucial no que diz respeito à compreensão de fenômenos inerentes a diversos campos da ciência. Desde a última década, todo aporte tecnológico alcançado ao longo de anos de investigação deram origem a uma estratégia orientada a dados, permitindo a inferência de modelos capazes de representar sistemas dinâmicos. Além disso, independentemente dos tipos de sensores adotados a fim de realizar o procedimento de aquisição de dados, é natural verificar a existência de uma certa corrupção ruidosa nos referidos dados. Genericamente, a tarefa de identificação é diretamente afetada pelo cenário ruidoso previamente descrito, implicando na falsa descoberta de um modelo generalizável. Em outras palavras, a corrupção ao ruído pode ser responsável pela geração de uma representação matemática infiel de um determinado sistema. Nesta tese, no que diz respeito à tarefa de identificação, é demonstrado como a robustez ao ruído pode ser melhorada a partir da hibridização de técnicas de aprendizado de máquina, como aumento de dados, regressão esparsa, seleção de características, extração de características, critério de informação, pesquisa em grade e validação cruzada. Especificamente, sob as perspectivas de classificação e regressão, o sucesso da estratégia proposta é apresentado a partir de exemplos numéricos, como o crescimento logístico, oscilador Duffing, modelo FitzHugh-Nagumo, atrator de Lorenz e uma modelagem Suscetível-Infeccioso-Recuperado (SIR) do *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2).

## Palavras-chave

Robustez ao Ruído; Identificação Esparsa; Corrupção Ruidosa; Engenharia de Recursos; Pesquisa em Grade.

## Abstract

França, Thaynã; Braga, Arthur M. B (Advisor); Ayala, Helon Vicente H (Co-Advisor). **Feature engineering to deal with noisy data in sparse identification through classification and regression perspectives**. Rio de Janeiro, 2021. 118p. Tese de Doutorado – Departamento de Engenharia Mecânica, Pontifícia Universidade Católica do Rio de Janeiro.

The dynamical systems play a fundamental role related to the understanding of the phenomena inherent to several fields of science. Since the last decade, all technological advances achieved throughout years of research have given rise to a data oriented strategy, enabling the inference of dynamical systems. Moreover, regardless the sensor types adopted to perform the data acquisition procedure, it is natural to verify the existence of a certain noise corruption in such data. Generically, the identification task is directly affected by the noisy scenario previously described, which entails in the false discovery of a generalizable model. In other words, the noise corruption might be responsible to give rise to a worthless mathematical representation of a given system. In this thesis, with respect to the identification assignment, it is demonstrated how the robustness to noise may be improved from the hybridization of machine learning techniques, such as data augmentation, sparse regression, feature selection, feature extraction, information criteria, grid search and cross validation. Specifically, through classification and regression perspectives, the success of the proposed strategy is presented from numerical examples, such as the logistic growth, Duffing oscillator, FitzHugh–Nagumo model, Lorenz attractor and a Susceptible-Infectious-Recovered (SIR) modeling of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2).

## Keywords

Noise Robustness; Sparse Identification; Noisy Corruption; Feature Engineering; Grid Search.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>14</b>
1.1	Motivação	15
1.2	Objetivos	16
1.3	Contribuições Originais do Trabalho	17
1.4	Organização do Trabalho	17
<b>2</b>	<b>Revisão da Literatura</b>	<b>18</b>
<b>3</b>	<b>Identificação Esparsa</b>	<b>24</b>
3.1	Estrutura SINDy	24
3.2	Derivadas Temporais	31
3.3	Corrupção Ruidosa	34
3.4	Pré-processamento	36
3.5	Normas e Regressões	38
3.6	Engenharia de Recursos	45
3.7	Pesquisa em Grade	53
3.8	Critério de Informação	57
3.9	Métricas de Avaliação	61
3.10	Metodologia SINDyFE	63
<b>4</b>	<b>Estudos de Caso</b>	<b>68</b>
4.1	Crescimento Logístico	68
4.2	Oscilador Duffing	69
4.3	FitzHugh–Nagumo	71
4.4	Atrator de Lorenz	72
4.5	Suscetível-Infeccioso-Recuperado	73
<b>5</b>	<b>Resultados</b>	<b>75</b>
<b>6</b>	<b>Conclusão e Trabalhos Futuros</b>	<b>95</b>
	<b>Referências bibliográficas</b>	<b>97</b>
<b>7</b>	<b>Apêndice A</b>	<b>108</b>
7.1	Bases de Funções	108
7.2	Esquemas de Diferenças Finitas	110
7.3	Estimativas de Parâmetros Estatísticos	113
7.4	Otimização	115
7.5	Produto Interno e Externo	117

## Lista de figuras

Figura 3.1	Construção da matriz dos estados.	25
Figura 3.2	Disposição das características existentes no que concerne à construção da matriz de dados aumentada.	27
Figura 3.3	Superfície aproximada a fim de descrever como o número de características existentes na matriz de dados aumentada se relaciona com o número de estados atrelados ao sistema admitido e com o grau máximo do polinômio adotado na construção da referida matriz.	29
Figura 3.4	Construção da matriz dos estados suspensa.	30
Figura 3.5	Detalhamento da implementação do DFR, DFC e DFP.	32
Figura 3.6	Diferentes técnicas de pré-processamento de dados sob a perspectiva da identificação esparsa.	38
Figura 3.7	Diferentes famílias de pseudonormas e normas: a) Representação gráfica das pseudonormas unitárias $L_0$ e $L_{0.5}$ e da norma unitária $L_1$ ; b) Representação gráfica das normas unitárias $L_2$ , $L_5$ e $L_\infty$ .	40
Figura 3.8	Exclusão de características.	46
Figura 3.9	Direção de máxima covariância (reta sólida) associada a um conjunto de dados aumentados, onde este pode ser representado por apenas três características.	47
Figura 3.10	Dados correlacionados e descorrelacionados: a) Direções de máxima covariância para um dado problema detentor de somente duas características; b) Transformação capaz de descorrelacionar os dados aumentados nas direções de máxima covariância.	48
Figura 3.11	Descarte de dois termos de recuperação dentre os sete existentes.	52
Figura 3.12	Descrição da reconstrução completa e parcial (primeiro termo de recuperação) da matriz de dados aumentada centralizada.	53
Figura 3.13	Tipos de espaçamentos: a) Uniforme; b) Logarítmico.	54
Figura 3.14	Representação gráfica da superfície interpolada capaz de descrever como a cardinalidade do conjunto que contém as possíveis combinações dos termos de recuperação adotados com o objetivo de reconstruir a matriz de dados aumentada centralizada varia de acordo com a quantidade de características existentes e com o número de termos de recuperação descartados.	56
Figura 3.15	Seleção do modelo detentor do melhor suporte (rosa) dentre uma ampla coleção de modelos candidatos menos representativos (ciano).	60
Figura 3.16	Esquema detalhado dos passos realizados pela estratégia proposta.	67
Figura 4.1	Primeiro estudo de caso: a) Versão discreta do crescimento logístico com forçamento estocástico; b) Excitações de treino (ciano) e teste (rosa) impostas ao crescimento logístico.	68

Figura 4.2 Segundo estudo de caso: a) Retrato de fase do oscilador Duffing ao admitir uma excitação do tipo multisenso; b) Excitações de treino (ciano) e teste (rosa) aplicadas ao oscilador Duffing.	70
Figura 4.3 Terceiro estudo de caso: a) Retrato de fase obtido a partir da admissão de um sinal de entrada do tipo multisenso; b) Excitações de treino (ciano) e teste (rosa) aplicadas ao modelo de FitzHugh–Nagumo.	71
Figura 4.4 Quarto estudo de caso: a) Comportamento caótico apresentado pela versão não forçada do atrator de Lorenz; b) Excitações de treino (ciano) e teste (rosa) impostas ao sistema de Lorenz.	73
Figura 4.5 Quinto estudo de caso: a) Evolução histórica de uma modelagem Suscetível-Infeccioso-Recuperado não forçada do SARS-CoV-2; b) Excitações de treino (ciano) e teste (rosa) impostas à modelagem SIR.	74
Figura 5.1 Dinâmica do sistema real e dos modelos gerados por meio das estruturas SINDy e SINDyFE para o crescimento logístico ( $\eta = 20\%$ ).	78
Figura 5.2 Discrepâncias entre a dinâmica do sistema e dos modelos produzidos pelas estruturas SINDy e SINDyFE para o crescimento logístico ( $\eta = 20\%$ ).	79
Figura 5.3 Evolução ao longo do tempo do comportamento dinâmico do sistema real e dos modelos fornecidos por intermédio das abordagens SINDy e SINDyFE para o oscilador Duffing ( $\eta = 7\%$ ).	81
Figura 5.4 Diferenças entre a evolução histórica do sistema e dos modelos obtidos pelas estratégias SINDy e SINDyFE para o oscilador Duffing ( $\eta = 7\%$ ).	82
Figura 5.5 Evolução histórica do comportamento dinâmico do sistema real e dos modelos providos pelas estruturas SINDy e SINDyFE para o modelo de FitzHugh–Nagumo ( $\eta = 5\%$ ).	84
Figura 5.6 Discrepâncias entre o comportamento dinâmico do sistema real e dos modelos produzidos por intermédio das abordagens SINDy e SINDyFE para o modelo de FitzHugh–Nagumo ( $\eta = 5\%$ ).	85
Figura 5.7 Comportamento dinâmico do sistema e dos modelos produzidos por meio das metodologias SINDy e SINDyFE para o atrator de Lorenz ( $\eta = 5\%$ ).	88
Figura 5.8 Diferenças entre a evolução do sistema e dos modelos gerados a partir das estruturas SINDy e SINDyFE para o atrator de Lorenz ( $\eta = 5\%$ ).	88
Figura 5.9 Comportamento dinâmico do sistema e dos modelos fornecidos por meio das abordagens SINDy e SINDyFE para a modelagem Suscetível-Infeccioso-Recuperado do SARS-CoV-2 ( $\eta = 5\%$ ).	91

Figura 5.10 Discrepâncias entre a evolução temporal do sistema e dos modelos produzidos por intermédio das estruturas SINDy e SINDyFE para a modelagem Suscetível-Infeccioso-Recuperado do SARS-CoV-2 ( $\eta = 5\%$ ).

91

## Lista de tabelas

Tabela 5.1	Crescimento logístico (métricas de classificação).	79
Tabela 5.2	Crescimento logístico (métricas de regressão).	80
Tabela 5.3	Crescimento logístico (parâmetros e hiperparâmetros).	80
Tabela 5.4	Oscilador Duffing (métricas de classificação).	82
Tabela 5.5	Oscilador Duffing (métricas de regressão).	83
Tabela 5.6	Oscilador Duffing (parâmetros e hiperparâmetros).	83
Tabela 5.7	Modelo de FitzHugh–Nagumo (métricas de classificação).	85
Tabela 5.8	Modelo de FitzHugh–Nagumo (métricas de regressão).	86
Tabela 5.9	Modelo de FitzHugh–Nagumo (parâmetros e hiperparâmetros).	86
Tabela 5.10	Atrator de Lorenz (métricas de classificação).	89
Tabela 5.11	Atrator de Lorenz (métricas de regressão).	89
Tabela 5.12	Atrator de Lorenz (parâmetros e hiperparâmetros).	90
Tabela 5.13	Modelagem SIR do SARS-CoV-2 (métricas de classificação).	92
Tabela 5.14	Modelagem SIR do SARS-CoV-2 (métricas de regressão).	92
Tabela 5.15	Modelagem SIR do SARS-CoV-2 (parâmetros e hiperparâmetros).	93
Tabela 7.1	Esquemas com erros de segunda ordem.	110
Tabela 7.2	Esquemas com erros de sexta ordem.	111

## Lista de Abreviaturas

ACP – Análise de Componentes Principais  
CDM – Coeficiente de Determinação Múltipla  
CIA – Critério de Informação de Akaike  
CIB – Critério de Informação Bayesiano  
DFC – Diferenças Finitas Centrais  
DFP – Diferenças Finitas Progressivas  
DFR – Diferenças Finitas Regressivas  
DVS – Decomposição em Valores Singulares  
EAM – Erro Absoluto Médio  
F – Coeficiente de Similaridade Dice  
FN – Falso Negativo  
FP – Falso Positivo  
MQO – Mínimos Quadrados Ordinários  
OSEMA – Operador de Seleção e Encolhimento Mínimo Absoluto  
P – Precisão  
REQM – Raiz do Erro Quadrático Médio  
S – Sensibilidade  
SARS-CoV-2 – *Severe Acute Respiratory Syndrome Coronavirus 2*  
SINDy – *Sparse Identification of Nonlinear Dynamics*  
SIR – Suscetível-Infeccioso-Recuperado  
SQR – Soma Quadrática Residual  
SQT – Soma Quadrática Total  
VN – Verdadeiro Negativo  
VP – Verdadeiro Positivo



*I am a traveler of both time and space.*

**Robert Anthony Plant**, *Physical Graffiti*.

# 1

## Introdução

Na última década, houve um notável aporte tecnológico conquistado com a finalidade de beneficiar a sociedade como um todo. Dessa maneira, salienta-se que tal avanço nos permite gozar hoje de privilégios, como a diminuição do distanciamento interpessoal viabilizado por meio de chamadas de vídeo ao longo do globo [1], sistemas inteligentes de transporte sendo introduzidos a fim de aumentar a segurança nas estradas [2], a utilização de robôs para uso tático militar a fim de preservar a vida humana [3], estimativa de parâmetros econométricos tornando possível a análise de diferentes cenários econômicos [4], desenvolvimento de novas estratégias visando a utilização adequada de fármacos no tratamento de doenças [5], reconhecimento facial por meio de imagens [6] e previsões do tempo almejando mitigar os danos causados por catástrofes naturais [7]. Sendo assim, vivemos em uma era potencialmente digital, onde o principal elemento é o dado [8]. De um ponto de vista técnico, os dados por si só não são valiosos, porém o conjunto de informações que é possível extrair de tais dados pode ser algo inestimável. Dessa forma, pesquisadores contribuíram e contribuem desenvolvendo ferramentas, técnicas e metodologias a fim de extrair o máximo de informação possível dos dados.

Enfatiza-se ainda que tal desenvolvimento culminou no surgimento de novas áreas de pesquisa. Primeiramente foi possível verificar o surgimento de uma área intitulada identificação de sistemas, onde esta, por sua vez, tem como finalidade obter as equações de um dado sistema com o auxílio de dados adquiridos [9]. Conceitualmente, a identificação de sistemas pode ser caracterizada pelo nível de conhecimento prévio admitido na modelagem. Dessa maneira, as famílias caixa branca e caixa preta apresentam dois extremos no que diz respeito ao nível citado [10]. Mais especificamente, a primeira família dispõe somente de leis naturais, como as encontradas na biologia, na física e na química. Sendo assim, o sistema pode ser descoberto a partir da aplicação e do desenvolvimento de tais leis. Por outro lado, expõe-se que a segunda família é estritamente orientada a dados, ou seja, nenhum tipo de lei da natureza é incorporada à metodologia. Por fim, explica-se que a modelagem denominada caixa cinza pode ser descrita como uma hibridização das modelagens caixa branca e caixa preta [11, 12].

Dito isso, à medida que mais avanços tecnológicos eram conquistados, pôde-se evidenciar o nascimento de uma outra área de pesquisa. Detalhadamente, suportado pelo aumento significativo na capacidade tanto de armazenamento quanto de processamento de dados, se estabeleceu o aprendizado de máquina, onde este pode ser definido como o campo da ciência que estuda a capacidade dos computadores aprenderem padrões a partir de dados sem a necessidade de programação explícita. Em outras palavras, o aprendizado de máquina visa extrair informações de um conjunto de dados por meio da busca por padrões existentes em tal conjunto [13]. Formalmente, alicerçado em um teorema conhecido como sem almoço grátis, cita-se que em problemas de aprendizado de máquina é comum implementar diversas técnicas a fim de obter a melhor solução possível para um dado problema [14, 15].

Categoricamente, existem diversas metodologias avançadas orientadas a dados, onde estas, por sua vez, podem ser utilizadas a fim de gerar modelos candidatos para um dado sistema. Dessa maneira, almejando obter descrições matemáticas de um certo sistema, pode-se implementar metodologias, tais como métodos de aprendizado [16], algoritmos evolutivos [17], redes neurais artificiais [18], modelos não lineares auto regressivos de médias móveis [19], florestas aleatórias [20] e máquinas de vetores de suporte [21]. Pontua-se ainda que, uma vez que a admissão de modelos não espúrios é capaz de aumentar substancialmente o desempenho no que diz respeito a aplicações de controle, os avanços alcançados tanto pela identificação de sistemas quanto pelo aprendizado de máquina favorecem quase que diretamente uma importante área da ciência conhecida como sistemas de controle [22, 23, 24].

## 1.1 Motivação

Deter a capacidade de inferir modelos fidedignos de um dado sistema somente a partir de dados adquiridos é uma habilidade significativamente relevante. Não obstante, explana-se que a existência de uma entidade amplamente conhecida como ruído, é capaz de dificultar ou até mesmo inviabilizar a tarefa de identificação. Explicita-se ainda que a presença da entidade citada está intrinsecamente vinculada ao processo de aquisição de dados. Destaca-se também que, por intermédio de simulações computacionais, é tangível utilizar um modelo previamente inferido a fim de realizar investigações visando compreender melhor os fenômenos associados ao sistema analisado. Em paralelo, ressalta-se que os avanços tecnológicos conquistados nos últimos anos possibilitaram um aumento expressivo no que concerne à adoção do aprendizado de máquina em diversos campos da ciência, como biologia, física e química.

A surpreendente capacidade dos algoritmos de aprendizado de máquina conseguirem captar padrões nos mais variados tipos de dados, torna a utilização dessa poderosa ferramenta algo naturalmente justificável. Desse modo, a partir da adesão de tal aprendizado, pode-se explorar um certo conjunto de dados a fim de encontrar os padrões escondidos no mesmo. Sendo assim, sob a ótica da identificação de sistemas, é importante utilizar uma estratégia robusta o suficiente com a finalidade de descobrir o conjunto de equações que regem um dado sistema. Além disso, mesmo com a extensiva exploração do aprendizado de máquina apresentada na literatura, ainda é possível melhorar as metodologias existentes com o auxílio de ferramentas suportadas pelo referido aprendizado, viabilizando o avanço de tais metodologias com o propósito de atender finalidades específicas.

## 1.2 Objetivos

O objetivo principal desta tese é avançar a metodologia intitulada *Sparse Identification of Nonlinear Dynamics* (SINDy) [25]. De maneira mais precisa, o referido avanço disserta sobre uma estratégia que visa melhorar a robustez ao ruído no que diz respeito à tarefa de identificação de sistemas dinâmicos não lineares tanto sob uma perspectiva de classificação quanto sob uma perspectiva de regressão. Desse modo, a fim de realizar tal avanço, a estrutura denominada *Feature Engineering to Deal with Noisy Data in Sparse Identification* (SINDyFE) foi desenvolvida, onde esta é baseada na hibridização de uma coleção de técnicas de aprendizado de máquina, tais como aumento de dados [25, 26], regressão esparsa [27, 28], seleção de características [20, 29], extração de características [13, 30], critério de informação [31, 32], pesquisa em grade [20, 33] e validação cruzada [20, 34].

Expõe-se também que, a fim de acessar a metodologia mais apropriada no que diz respeito à inferência de modelos não espúrios a partir de dados ruidosos, métricas de avaliação, como precisão, sensibilidade, coeficiente de similaridade, erro absoluto médio, raiz do erro quadrático médio e coeficiente de determinação são consideradas. Além disso, são traçados comparativos entre as evoluções históricas dos estados do sistema real e dos estados dos modelos produzidos por intermédio das metodologias abordadas nesta tese. Em adição, como objetivo secundário, este material almeja tanto prover uma visão geral do aprendizado de máquina quanto promover a disseminação do mesmo. Explica-se ainda que, a fim de explorar tal objetivo, conceitos inerentes ao aprendizado de máquina são explorados em diferentes contextos com o propósito de mostrar sua elevada versatilidade.

### 1.3

#### Contribuições Originais do Trabalho

Em poucas palavras, a originalidade intrínseca ao conteúdo apresentado nesta tese pode ser observada sob duas perspectivas. Primeiramente, esta é a primeira vez que a estratégia pioneira é explorada a partir de uma ótica classificatória. Em outras palavras, no que diz respeito à compatibilidade dos termos não lineares admitidos nas descrições matemáticas tanto do sistema analisado quanto dos modelos inferidos, a estrutura proposta provê uma abordagem capaz de avaliar as similaridades existentes entre o sistema real e os modelos identificados por tal estrutura. Além disso, em caráter igualmente inédito, salienta-se que o principal avanço inerente à metodologia proposta disserta sobre o aumento da robustez ao ruído da abordagem pioneira, onde este viabiliza a identificação apropriada de sistemas dinâmicos mesmo quando condições mais severas de ruído são verificadas. De modo detalhado, enfatiza-se que o referido avanço é alicerçado na incorporação de uma ampla coleção de técnicas de aprendizado de máquina.

### 1.4

#### Organização do Trabalho

O conteúdo remanescente inerente a esta tese está dividido da seguinte forma. No Capítulo 2 é apresentada a revisão da literatura, objetivando familiarizar o leitor com diversos conceitos inerentes à identificação de sistemas e ao aprendizado de máquina. No Capítulo 3, explicita-se a teoria necessária a fim de compreender a metodologia original e a estratégia proposta. Em detalhes, tem-se que conceitos, como aumento de dados, regressão esparsa, seleção de características, extração de características, critério de informação e pesquisa em grade são explicados. Adicionalmente, o Capítulo 4 expõe os diferentes estudos de caso admitidos neste trabalho. Mais precisamente, estudos de caso, como o crescimento logístico [25, 35], o oscilador Duffing [36, 37], o modelo FitzHugh–Nagumo [38, 39], o atrator de Lorenz [25, 40] e uma modelagem Suscetível-Infeccioso-Recuperado (SIR) do *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2) [41, 42] são adotados. Fundamentado nas metodologias abordadas no presente trabalho, o Capítulo 5 reporta os resultados obtidos para os estudos de caso supracitados. Especificamente, são adotadas diferentes métricas de classificação e regressão a fim de atestar a superioridade da metodologia proposta em comparação com a original. No Capítulo 6 tanto as conclusões obtidas quanto os trabalhos futuros propostos são explicitados. Por fim, almejando garantir a compreensão completa da estratégia proposta, o Apêndice A é adicionado a este material.

A capacidade do aprendizado de máquina ser aplicado aos mais diferentes tipos de problemas, torna o referido aprendizado uma ferramenta extensamente adotada. Além disso, abordar um problema sob a perspectiva do aprendizado de máquina implica na necessidade de efetuar uma coleção de reflexões, visando enquadrar o problema corretamente e, conseqüentemente, acessar o conjunto de procedimentos e técnicas apropriadas para tal finalidade. As reflexões citadas ocorrem desde a definição do problema até o momento em que os resultados são interpretados. Ressalta-se ainda que esse trajeto como um todo não possui uma única sequência de passos lógicos, porém existem alguns pontos relevantes que podem servir de guia na abordagem de problemas sob a ótica do aprendizado de máquina. Por exemplo, é necessário estipular se o problema atacado se trata de uma inferência ou uma predição [13]. Deve-se especificar também se a abordagem utilizada é fundamentada na obtenção de um modelo generalizável ou se a mesma é alicerçada na comparação entre as instâncias de treino e as instâncias de teste [43]. Em paralelo, é necessário definir se a estratégia adotada é suportada pelo aprendizado do tipo em lote ou em tempo real [43].

Em adição, também é relevante estabelecer se o problema proposto pode ser abordado como um problema de classificação ou de regressão [34]. Os problemas de aprendizado de máquina também podem ser separados de acordo com o tipo de supervisão adotada. Em outras palavras, deve-se determinar se o problema é não supervisionado, semi-supervisionado, supervisionado ou supervisionado por reforço [44]. Dessa forma, tais reflexões são fundamentais para situar um dado problema e, conseqüentemente, restringir de forma significativa as abordagens disponíveis para o mesmo [20]. Além disso, existem alguns fatores que podem ser responsáveis tanto por gerar soluções não satisfatórias para um certo problema quanto aumentar substancialmente o desempenho de uma dada metodologia. Sendo assim, antes de implementar um algoritmo de aprendizado de máquina, é recomendável verificar se há necessidade de submeter o conjunto de dados a pré-processamentos específicos, como centralização, reescala, normalização e padronização [20]. Por outro lado, enfatiza-se que a seleção de variáveis plausíveis durante a confecção de um dado modelo pode evitar o vazamento de dados [12].

Algoritmos de aprendizado de máquina costumam obter resultados menos expressivos quando alimentados por variáveis altamente correlacionadas. Daí, é possível efetuar transformações no conjunto de dados visando mapear os dados inicialmente correlacionados em um novo conjunto não correlacionado [13]. Adicionalmente, por intermédio da adoção de técnicas de redução de dimensionalidade, pode-se evitar um fenômeno conhecido como maldição da dimensionalidade [45]. Explica-se também que cada algoritmo de aprendizado de máquina possui um conjunto de hiperparâmetros a ser definido. Sendo assim, visando tanto ajustar os hiperparâmetros inerentes a um dado modelo quanto avaliar a qualidade do mesmo, deve-se dividir o conjunto de dados em dois subconjuntos disjuntos intitulados treino e teste. Assim, baseado em um algoritmo de aprendizado de máquina previamente estabelecido, os dados de treino são utilizados com a finalidade de criar um modelo capaz de relacionar as variáveis dependentes com as independentes.

Objetivando acessar a performance do modelo gerado, submete-se o mesmo a dados nunca vistos, isto é, aplica-se o modelo gerado por meio dos dados de treino no conjunto de teste [46]. Em algumas aplicações, pode-se verificar a admissão do conjunto de validação, onde este é obtido a partir da extração de uma parcela dos dados de treino. Dessa maneira, a fim de obter o melhor conjunto de hiperparâmetros, pode-se submeter o conjunto de validação a uma técnica conhecida como validação cruzada [20]. Explana-se ainda que, suportada pela simplicidade implementacional, uma estratégia de validação cruzada conhecida como *K-fold* é rotineiramente empregada [20]. Paralelamente, caso o algoritmo de aprendizado de máquina admitido possua mais de um hiperparâmetro a ser definido, pode-se utilizar uma técnica denominada pesquisa em grade a fim de obter tais hiperparâmetros [43].

Expõe-se ainda que, quanto mais parâmetros um modelo possui, mais facilmente o mesmo poderá se ajustar a um certo conjunto de dados. Mais especificamente, um modelo detentor de um número elevado de parâmetros é geralmente traduzido como um modelo de baixo desempenho, estabelecendo, por sua vez, um cenário onde o modelo citado é dito não generalizável [45]. Por outro lado, um modelo que desfruta de um número insuficiente de graus de liberdade pode não possuir a flexibilidade mínima requerida, desfavorecendo um ajuste apropriado ao conjunto de dados [43]. Simplificadamente, modelos vinculados a um certo déficit paramétrico contém subajuste. De forma complementar, modelos associados a um dado superávit paramétrico possuem sobreajuste [20]. Idealmente, um modelo fidedigno, isto é, um modelo capaz de representar de forma genuína um dado sistema, deve equilibrar adequadamente subajuste e sobreajuste [43].

A fim de verificar se o modelo produzido por uma metodologia específica possui subajuste ou sobreajuste, pode-se aplicar uma estratégia intitulada curvas de aprendizado ou adotar uma abordagem via validação cruzada [20]. Adicionalmente, recomenda-se implementar diferentes algoritmos de aprendizado de máquina, pois cada um deles possui uma forma peculiar de buscar padrões nos dados, mostrando assim a inexistência de um algoritmo predominante [14, 15]. Dessa forma, a partir dos dados de teste, tem-se que os modelos obtidos por um grupo diversificado de metodologias podem ser elencados segundo uma métrica de avaliação adotada. É possível verificar a existência de uma coleção de métricas que podem ser adotadas a fim de mensurar o desempenho de um dado modelo. Todavia, cada tipo de tarefa está vinculada a um conjunto restrito de métricas que podem ser implementadas. Em detalhes, as métricas acurácia, precisão, sensibilidade e especificidade podem ser admitidas em problemas de classificação [43, 47, 48].

De maneira alternativa, métricas conhecidas, como erro absoluto médio, erro quadrático médio, raiz do erro quadrático médio e coeficiente de determinação múltipla podem ser assumidas em problemas de regressão [20, 43, 49]. Além disso, existem métricas capazes de penalizar modelos que apresentam elevado grau de liberdade, isto é, modelos complexos em demasia, acarretando na obtenção de modelos parcimoniosos e, por conseguinte, promovendo um equilíbrio adequado entre subajuste e sobreajuste [50, 51, 52]. Certos problemas apresentam uma linha tênue no que diz respeito ao enquadramento dos mesmos como um problema de classificação ou regressão. Um exemplo clássico de tal dualidade é personificado pelo problema da identificação de dígitos escritos à mão, onde identificar a intensidade de um dado pixel pode ser interpretado tanto como um problema de classificação quanto de regressão [43].

Pontua-se também que, independentemente do tipo de problema apresentado, os dados de teste só devem ser visualizados e manipulados na etapa de teste, evitando assim um problema denominado viés de espionagem. Além disso, se o conjunto de dados associado a um certo problema não possuir alguns valores no que diz respeito às variáveis dependentes, configurando a presença de valores faltantes, pode-se submeter tais variáveis à técnicas capazes de lidar com a anomalia descrita [20]. Ressalta-se ainda que, a fim de lidar com problemas de classificação não binários, é aconselhável adotar um procedimento conhecido como codificação *one-hot* [43]. Em adição, almejando aplicar algoritmos de aprendizado de máquina em problemas categóricos que possuem discrepâncias significativas em relação à distribuição das amostras nas classes existentes, é recomendável que sejam adotadas técnicas capazes de promover um certo balanceamento amostral [43].



É natural observar, principalmente em problemas de classificação, procedimentos que utilizam um conceito nomeado embaralhamento, onde este tem como finalidade embaralhar de maneira sistemática os dados de treino [18]. No entanto, tais procedimentos se tornam inviáveis em contextos onde uma certa estrutura e dependência temporal são verificadas [26]. Destaca-se ainda que, de forma similar aos problemas de regressão, pode-se utilizar o conceito apresentado como aumento de dados em problemas de classificação. Uma aplicação habitual do referido conceito se torna presente em problemas de classificação de imagens, onde a partir deste é possível aumentar o volume de dados por meio de transformações como translação e rotação [43]. Desse modo, após realizar as transformações citadas nos dados de treino, as imagens provenientes dos pré-processamentos adotados são incorporadas ao conjunto de treino. Explicita-se também que o aumento de dados foi amplamente explorado em diferentes contextos do aprendizado de máquina [53, 54, 55].

Os sistemas dinâmicos possuem uma relevância fundamental no que diz respeito à compreensão de fenômenos existentes na biologia, economia, engenharia, física e química [1, 2, 3, 4, 5, 6, 7]. Em paralelo, a identificação de sistemas disserta sobre a inferência de descrições matemáticas não espúrias de sistemas dinâmicos [26]. Salienta-se ainda que o tipo de sinal de entrada adotado para um dado sistema pode provocar ou o sucesso ou o fracasso no que diz respeito à identificação correta do sistema analisado. Adicionalmente, em algumas situações específicas, o tipo de sinal utilizado pode ser imposto de acordo com um certo conjunto de características que deseja-se explorar no sistema considerado [56]. Dessa maneira, objetivando identificar o sistema corretamente, tanto o tipo de sinal adotado quanto os parâmetros existentes no mesmo são quesitos importantes a serem avaliados [9].

Os sinais de entrada conhecidos como sinal de varredura, excitação multifásica, forçamento multiseno, sequência pseudo binária, impacto de pulso, explosão aleatória, ruído aleatório e varredura senoidal desempenham um papel crucial na tarefa de identificação [12, 45, 56]. Outro aspecto relevante na identificação de sistemas é a adoção da simulação do tipo livre em detrimento da simulação um passo à frente. Formalmente, a primeira família de simulação mencionada disserta sobre a realização de estimativas a partir somente de medições passadas. Em contraste, objetivando a mesma finalidade, as simulações vinculadas à família um passo à frente utilizam tanto as medições passadas quanto as medições atuais [57]. Sendo assim, a abordagem via simulação livre é capaz de gerar modelos mais generalizáveis, isto é, modelos mais fidedignos em comparação com o sistema real, favorecendo, por sua vez, a identificação adequada do sistema analisado [9].

Em particular, suportada pelo aumento de dados, a estratégia recém desenvolvida conhecida como *Sparse Identification of Nonlinear Dynamics* (SINDy) explora transformações não lineares [25]. Dessa maneira, a metodologia pioneira considera a produção de uma biblioteca de dados, onde esta é gerada a partir de uma dada combinação de termos não lineares. Daí, por meio da regressão esparsa, tal estratégia é capaz de produzir modelos representativos de um dado sistema [27]. Precisamente, a regressão esparsa impõe uma regularização capaz de encolher e selecionar os termos mais relevantes no que diz respeito à dinâmica intrínseca aos dados, viabilizando a identificação de sistemas dinâmicos não lineares [58]. Além disso, a elevada capacidade interpretativa associada à referida estrutura torna a adoção da mesma uma grande vantagem. Expõe-se também que, nos últimos anos, a metodologia original sofreu diversas modificações com o propósito de contemplar um conjunto expressivo de necessidades específicas [26].

Cronologicamente, a primeira modificação proposta para a estrutura SINDy viabilizou a identificação de sistemas descritos a partir do quociente entre dois polinômios e, posteriormente, uma abordagem sob a ótica do controle foi apresentada [59, 60]. Em seguida, uma extensão teórica da estrutura original foi realizada permitindo o tratamento de sistemas dinâmicos não autônomos e restrições físicas previamente conhecidas [61]. Subsequentemente, adaptações capazes de viabilizar a identificação de sistemas descritos por meio de equações diferenciais parciais foram introduzidas [62, 63]. Em adição, tanto uma análise no que diz respeito ao impacto gerado por dados altamente corrompidos quanto uma abordagem baseada no critério de informação foram efetuadas [64, 65]. Além disso, pode-se salientar que a teoria foi expandida para normas de esparsidade de grupos com o auxílio da admissão de um tipo de regressão intitulada regressão de crista [66].

Em sequência, um estudo foi realizado a fim de viabilizar o tratamento de sistemas detentores de coeficientes variáveis temporalmente e espacialmente [67]. Posteriormente, aplicações almejando analisar o processo de fratura hidráulica, comparar diferentes técnicas de controle em diversos sistemas dinâmicos, implementar a metodologia SINDy experimentalmente e possibilitar a identificação de sistemas híbridos também foram realizadas [68, 69, 70, 71]. Por fim, alicerçada na assunção de diferentes percentuais de separação do conjunto de dados, a robustez ao ruído da estratégia pioneira SINDy foi investigada a fim de reportar a capacidade da mesma inferir modelos não espúrios sob a perspectiva da engenharia [26]. Notoriamente, é tangível constatar que a magnitude do ruído vinculado aos dados adquiridos é capaz de impossibilitar o sucesso das metodologias supracitadas [72].

Até este ponto, um conjunto significativo de considerações, estratégias e ferramentas foram explicitadas visando prover uma visão geral do aprendizado de máquina. Todavia, é crucial manter em mente que ao abordar um problema sob a perspectiva do aprendizado de máquina é necessário enquadrar tal problema de acordo com os quesitos descritos a seguir.

- ▷ Classificação | Regressão
- ▷ Inferência | Predição
- ▷ Instância | Modelo
- ▷ Lote | Tempo real
- ▷ Reforço | Semi-supervisionado | Supervisionado | Não supervisionado

Motivados pela crítica robustez ao ruído apresentada pela estrutura original SINDy, uma generalização da referida estrutura denominada *Feature Engineering to Deal with Noisy Data in Sparse Identification* (SINDyFE) é desenvolvida, onde esta viabiliza a identificação de sistemas dinâmicos não lineares sob condições ruidosas. Pontua-se ainda que o tipo de aprendizado associado à abordagem proposta se enquadra, como regressão, inferência, modelo, lote e supervisionado. Além disso, embora a metodologia proposta seja desenvolvida a partir do ponto de vista da regressão, é traçado um comparativo entre os desempenhos da estratégia pioneira e da estrutura proposta sob uma perspectiva classificatória, mostrando assim a apreciável robustez ao ruído da metodologia proposta no que concerne à tarefa de identificação.

## 3

### Identificação Esparsa

No presente capítulo, são introduzidos conceitos como identificação esparsa, diferenças finitas, corrupção ruidosa, pré-processamento de dados, normas vetoriais, famílias de regressão, engenharia de recursos, seleção de características, extração de características, pesquisa em grade, validação cruzada, critérios de informação e métricas de avaliação.

#### 3.1

##### Estrutura SINDy

Notoriamente, suportado pela literatura disponível, pode-se constatar que o aumento de dados é um dos conceitos utilizados pela estrutura intitulada *Sparse Identification of Nonlinear Dynamics* (SINDy). Dessa maneira, a partir dos estados previamente mensurados, é possível gerar variáveis de forma artificial, possibilitando a criação de uma biblioteca de termos não lineares. Além disso, por intermédio da regressão esparsa, enfatiza-se que a estrutura citada viabiliza a obtenção de um modelo detentor de apenas alguns termos ativos, ou seja, um modelo esparso [25]. Detalhadamente, pode-se pontuar que a referida regressão possui um termo adicional de penalização, onde este permite o encolhimento de alguns termos até zero ao longo do procedimento de regressão, viabilizando, por sua vez, a descoberta de modelos matemáticos capazes de representar o conjunto de dados com o menor número possível de termos não lineares [27]. Assim, a metodologia SINDy torna exequível a identificação de sistemas dinâmicos, onde estes podem ser genericamente descritos por meio das Equações 3-1 e 3-2.

$$\dot{\mathbf{x}} = f[\mathbf{x}(t)] \quad ; \quad t > t_0 \quad (3-1)$$

$$x_0 = x(t) \quad ; \quad t = t_0 \quad (3-2)$$

Um objeto matemático conhecido como matriz é vastamente utilizado devido a sua capacidade de armazenar dados de diversas naturezas. Destaca-se ainda que, em problemas de regressão aplicados a sistemas dinâmicos, é comum organizar os dados em matrizes, como expõe a Figura 3.1.

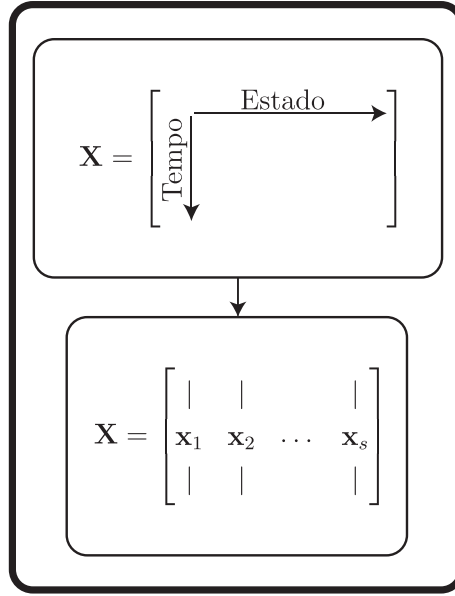


Figura 3.1: Construção da matriz dos estados.

Especificamente, tais matrizes são organizadas levando em conta a evolução temporal de cada tipo de medição realizada. De forma genérica, admite-se um sistema detentor de um certo número de estados  $s$  ( $s \in \mathbb{N}^*$ ), onde estes são descritos a partir de uma determinada quantidade de instâncias temporais  $n$  ( $n \in \mathbb{N}^*$ ). Formalmente, constrói-se a matriz dos estados  $\mathbf{X}$  ( $\mathbf{X} \in \mathbb{R}^{n \times s}$ ) a partir de vetores coluna  $\mathbf{x}_i$  ( $i = 1, 2, \dots, s$ ), onde estes, descrevem a evolução ao longo do tempo de cada um dos estados que compõem o sistema, como explicitado na Equação 3-3.

$$\mathbf{X} = \begin{bmatrix} x_1(t_1) & x_2(t_1) & \cdots & x_s(t_1) \\ x_1(t_2) & x_2(t_2) & \cdots & x_s(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(t_n) & x_2(t_n) & \cdots & x_s(t_n) \end{bmatrix} \quad (3-3)$$

A fim de realizar a identificação esparsa de um dado sistema dinâmico, é necessário acessar as derivadas de cada estado  $\dot{\mathbf{x}}_i$  ( $i = 1, 2, \dots, s$ ), onde estas podem ser obtidas tanto por meio de medições diretas quanto por intermédio de técnicas de aproximação. Desse modo, independentemente da maneira como tais derivadas são adquiridas, é possível dispô-las em uma matriz denominada matriz das derivadas dos estados  $\dot{\mathbf{X}}$  ( $\dot{\mathbf{X}} \in \mathbb{R}^{n \times s}$ ), onde cada linha da matriz citada está associada à evolução ao longo do tempo das derivadas dos estados. Matematicamente, com a finalidade de prover uma descrição mais formal da matriz previamente descrita, a Equação 3-4 é apresentada.

$$\dot{\mathbf{X}} = \begin{bmatrix} \dot{x}_1(t_1) & \dot{x}_2(t_1) & \cdots & \dot{x}_s(t_1) \\ \dot{x}_1(t_2) & \dot{x}_2(t_2) & \cdots & \dot{x}_s(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \dot{x}_1(t_n) & \dot{x}_2(t_n) & \cdots & \dot{x}_s(t_n) \end{bmatrix} \quad (3-4)$$

Note que a Equação 3-4 representa a matriz das derivadas dos estados de um sistema contendo uma dada quantidade de instâncias temporais  $n$  ( $n \in \mathbb{N}^*$ ) e um certo número de estados  $s$  ( $s \in \mathbb{N}^*$ ). Enfatiza-se ainda que, visando submeter um dado conjunto de dados à identificação esparsa, é necessário gerar a matriz de dados aumentada  $\Theta$  ( $\Theta \in \mathbb{R}^{n \times f}$ ). Assim, por intermédio de uma base de funções previamente definida, os elementos presentes na matriz de dados aumentada são arranjados segundo a evolução temporal de cada estado e de acordo com um determinado conjunto de combinações não lineares destes [25]. Dessa forma, a matriz de dados aumentada é composta por um determinado número de colunas  $f$  ( $f \in \mathbb{N}^*$ ), onde tal valor representa a quantidade de características admitidas [20]. Em paralelo, é válido ressaltar que a metodologia SINDy é fundamentada na hipótese de que o sistema dinâmico a ser identificado é detentor de uma representação esparsa, ou seja, o referido sistema pode ser descrito a partir de alguns poucos termos não nulos [27]. Matematicamente, um sistema dinâmico esparsa pode ser caracterizado pelo produto entre a matriz de dados aumentada  $\Theta$  ( $\Theta \in \mathbb{R}^{n \times f}$ ) e a matriz de esparsidade  $\Xi$  ( $\Xi \in \mathbb{R}^{f \times s}$ ), como descreve a Equação 3-5.

$$\dot{\mathbf{X}} = \Theta \Xi \quad (3-5)$$

Especificamente, cada coluna da matriz de dados aumentada é gerada a partir de dois conjuntos, onde o primeiro é dado pelas variáveis inicialmente mensuradas e o segundo é gerado, artificialmente, a partir de certas combinações das variáveis pertencentes ao primeiro conjunto. Note que a matriz de dados aumentada, por construção, deve possuir a mesma quantidade de linhas que a matriz dos estados, como explicita a Equação 3-6.

$$\Theta = \begin{bmatrix} \theta_1(t_1) & \theta_2(t_1) & \cdots & \theta_f(t_1) \\ \theta_1(t_2) & \theta_2(t_2) & \cdots & \theta_f(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \theta_1(t_n) & \theta_2(t_n) & \cdots & \theta_f(t_n) \end{bmatrix} \quad (3-6)$$

A viabilidade no que diz respeito à obtenção de uma representação esparsa de um dado sistema está intimamente ligada à forma com que a matriz de dados aumentada é gerada. Sendo assim, objetivando gerar tal matriz, uma das primeiras etapas da modelagem SINDy é a definição da base de funções utilizada. Detalhadamente, é possível verificar a adoção de bases de funções descritas a partir de polinômios racionais, polinômios reais e expansões da série de Fourier, para citar algumas [61]. Em particular, explica-se que a base polinomial real é comumente adotada devido a sua elevada capacidade de ajuste aos dados. De maneira formal, a Equação 3-7 apresenta a matriz de dados aumentada  $\Theta_d^{\mathbb{R}}$  ( $\Theta_d^{\mathbb{R}} \in \mathbb{R}^{n \times f}$ ), onde esta é construída a partir da assunção de uma base polinomial real de grau máximo  $d$  ( $d \in \mathbb{N}^*$ ).

$$\Theta_d^{\mathbb{R}} = \begin{bmatrix} | & | & | & | & | & \cdots & | \\ \mathbf{P}_0^{\mathbb{R}} & \mathbf{P}_1^{\mathbb{R}} & \mathbf{P}_2^{\mathbb{R}} & \mathbf{P}_3^{\mathbb{R}} & \mathbf{P}_4^{\mathbb{R}} & \cdots & \mathbf{P}_d^{\mathbb{R}} \\ | & | & | & | & | & \cdots & | \end{bmatrix} \quad (3-7)$$

Em geral, as características inerentes à matriz de dados aumentada são dispostas de acordo com os graus dos polinômios adotados na construção da mesma. Especificamente, a partir da Figura 3.2 é possível acessar como a construção da referida matriz é realizada.

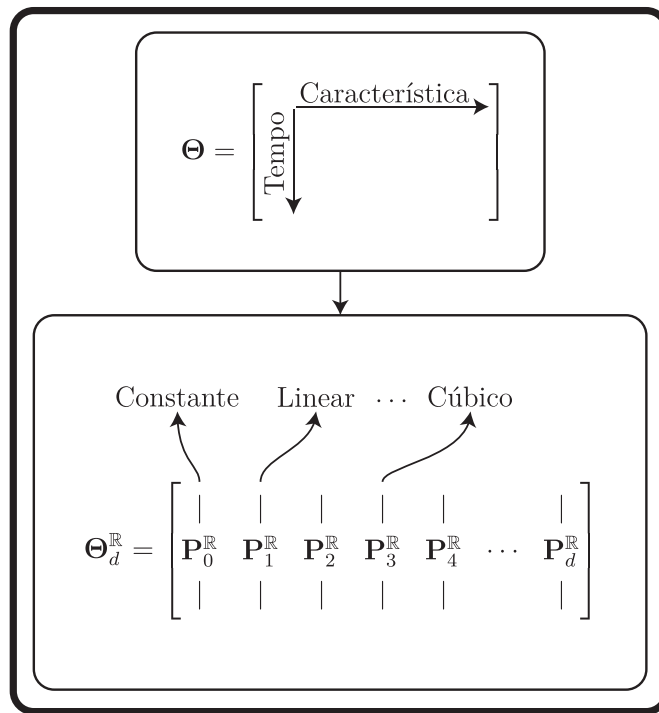


Figura 3.2: Disposição das características existentes no que concerne à construção da matriz de dados aumentada.

Note que as características presentes na matriz de dados aumentada são produzidas por intermédio das variáveis inicialmente mensuradas, ou seja, dos estados do sistema. Adicionalmente, desejando estender o conceito de bases de funções, expõe-se que diferentes famílias de bases são discutidas no Apêndice A do presente trabalho. Todavia, ao assumir, por hipótese, um sistema detentor da variável de estado  $\mathbf{x}_1$  ( $\mathbf{x}_1 \in \mathbb{R}^n$ ) e da variável de estado  $\mathbf{x}_2$  ( $\mathbf{x}_2 \in \mathbb{R}^n$ ), é possível concluir que o termo  $\mathbf{P}_3^{\mathbb{R}}$  ( $\mathbf{P}_3^{\mathbb{R}} \in \mathbb{R}^{n \times 4}$ ) presente na matriz de dados aumentada previamente descrita por intermédio da Equação 3-7 pode ser acessado a partir da Equação 3-8.

$$\mathbf{P}_3^{\mathbb{R}} = \begin{bmatrix} | & | & | & | \\ \mathbf{x}_1^3 & \mathbf{x}_1^2 \mathbf{x}_2 & \mathbf{x}_1 \mathbf{x}_2^2 & \mathbf{x}_2^3 \\ | & | & | & | \end{bmatrix} \quad (3-8)$$

Além disso, independentemente da quantidade de estados do sistema, a primeira coluna da matriz de dados aumentada  $\Theta_d^{\mathbb{R}}$  ( $\Theta_d^{\mathbb{R}} \in \mathbb{R}^{n \times f}$ ) é representada por um vetor contendo entradas unitárias cuja dimensão é numericamente igual ao número de instâncias temporais admitidas. Por outro lado, o termo  $\mathbf{P}_1^{\mathbb{R}}$  ( $\mathbf{P}_1^{\mathbb{R}} \in \mathbb{R}^{n \times s}$ ) é dado por uma submatriz, onde esta é composta unicamente pelos vetores de estado do sistema abordado. Complementarmente, explica-se que os demais termos presentes na matriz de dados aumentada dada pela Equação 3-7 são obtidos por meio do produto elemento a elemento entre os estados do sistema. Em particular, para um sistema contendo somente as duas variáveis de estado previamente descritas, tem-se que a segunda coluna da submatriz  $\mathbf{P}_3^{\mathbb{R}}$  ( $\mathbf{P}_3^{\mathbb{R}} \in \mathbb{R}^{n \times 4}$ ) pode ser obtida pela Equação 3-9.

$$x_1^2 x_2(t_i) = \left[ x_1(t_i) \right]^2 x_2(t_i) \quad (i = 1, 2, \dots, n) \quad (3-9)$$

Similarmente, para um sistema detentor de apenas duas variáveis de estado, tem-se que a quarta e última coluna da submatriz  $\mathbf{P}_3^{\mathbb{R}}$  ( $\mathbf{P}_3^{\mathbb{R}} \in \mathbb{R}^{n \times 4}$ ) pode ser gerada por meio da Equação 3-10.

$$x_2^3(t_i) = \left[ x_2(t_i) \right]^3 \quad (i = 1, 2, \dots, n) \quad (3-10)$$

Pontua-se ainda que, ao considerar sistemas forçados, é necessário alocar o sinal de entrada do sistema  $\mathbf{u}$  ( $\mathbf{u} \in \mathbb{R}^n$ ) na matriz de dados aumentada. Formalmente, a Equação 3-11 explicita a matriz de dados aumentada  $\Theta_{u,d}^{\mathbb{R}}$  ( $\Theta_{u,d}^{\mathbb{R}} \in \mathbb{R}^{n \times (f+1)}$ ) para a condição apresentada.



$$\Theta_{u,d}^{\mathbb{R}} = \begin{bmatrix} | & | & | & | & | & | & \cdots & | \\ \mathbf{P}_0^{\mathbb{R}} & \mathbf{P}_1^{\mathbb{R}} & \mathbf{u} & \mathbf{P}_2^{\mathbb{R}} & \mathbf{P}_3^{\mathbb{R}} & \mathbf{P}_4^{\mathbb{R}} & \cdots & \mathbf{P}_d^{\mathbb{R}} \\ | & | & | & | & | & | & \cdots & | \end{bmatrix} \quad (3-11)$$

Em adição, dependendo da dinâmica vinculada ao sistema analisado, pode ser interessante explorar o problema de identificação esparsa utilizando também o sinal de excitação  $\mathbf{u}$  ( $\mathbf{u} \in \mathbb{R}^n$ ) ao gerar as combinações não lineares impostas pela base de funções previamente adotada [26]. Salienta-se também que, ao admitir uma base polinomial real de grau máximo  $d$  ( $d \in \mathbb{N}^*$ ) e ao assumir um sistema detentor de  $s$  ( $s \in \mathbb{N}^*$ ) variáveis de estado, é estabelecido um problema contendo  $f$  ( $f \in \mathbb{N}^*$ ) características, onde o referido número de características pode ser calculado por intermédio da Equação 3-12.

$$f = \frac{(s+d)!}{s!d!} \quad (3-12)$$

Observe que a Figura 3.3 tem como finalidade evidenciar o aumento significativo do número de características presentes na matriz de dados aumentada, onde o aumento citado é verificado caso sejam admitidos tanto um sistema detentor de mais estados quanto um grau máximo mais elevado no que diz respeito à geração de tal matriz.

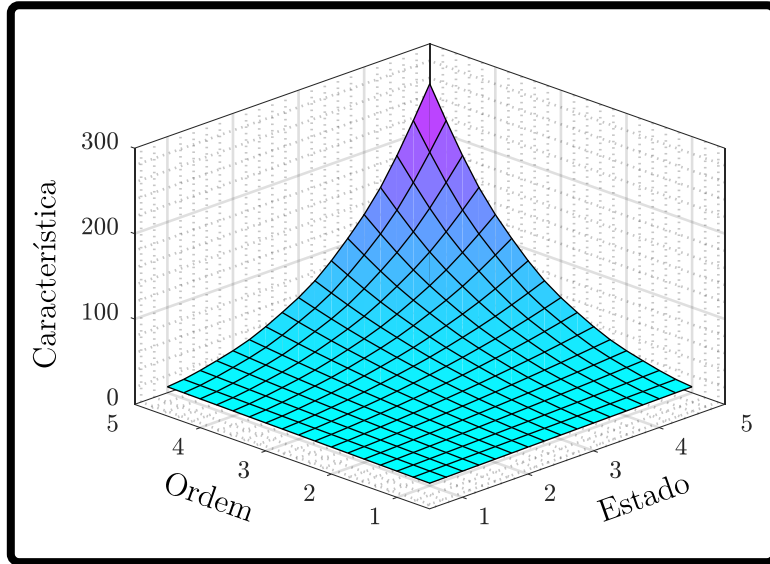


Figura 3.3: Superfície aproximada a fim de descrever como o número de características existentes na matriz de dados aumentada se relaciona com o número de estados atrelados ao sistema admitido e com o grau máximo do polinômio adotado na construção da referida matriz.

Explica-se ainda que a Equação 3-12 não contabiliza a característica referente ao sinal de entrada aplicado ao sistema. Desse modo, caso o forçamento seja considerado na construção da matriz de dados aumentada, deve-se acrescentar uma unidade ao número total de características descritas pela Equação 3-12. Por outro lado, se o sinal de excitação for utilizado nas combinações não lineares presentes na matriz de dados aumentada, é necessário adicionar uma unidade ao número de estados antes de utilizar a Equação 3-12, pois nesse caso o forçamento pode ser interpretado como um estado extra do sistema. Pontua-se também que os elementos nulos e não nulos presentes na matriz de esparsidade  $\Xi$  ( $\Xi \in \mathbb{R}^{f \times s}$ ) representam, respectivamente, os termos inativos e ativos do sistema analisado. Em detalhes, a Equação 3-13 descreve como os vetores de esparsidade  $\xi_i$  ( $i = 1, 2, \dots, s$ ) compõem a matriz de esparsidade, que, por hipótese, é esparsa [25].

$$\Xi = \begin{bmatrix} | & | & \cdots & | \\ \xi_1 & \xi_2 & \cdots & \xi_s \\ | & | & \cdots & | \end{bmatrix} \quad (3-13)$$

Explana-se também que, ao considerar sistemas de segunda ordem, é possível gerar a matriz dos estados suspensa  $\mathbf{X}_s$  ( $\mathbf{X}_s \in \mathbb{R}^{n \times (2s)}$ ) a partir dos estados do sistema  $\mathbf{x}_i$  ( $i = 1, 2, \dots, s$ ) e das derivadas de tais estados  $\dot{\mathbf{x}}_i$  ( $i = 1, 2, \dots, s$ ). Dito isso, a Figura 3.4 objetiva descrever como a matriz dos estados suspensa pode ser obtida.

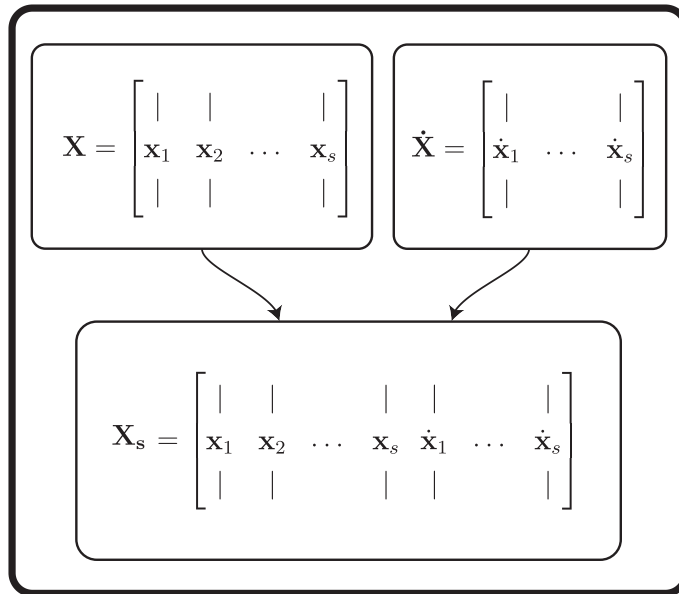


Figura 3.4: Construção da matriz dos estados suspensa.

Note que ao admitir sistemas de ordens superiores, tem-se que a matriz dos estados suspensa exercerá o mesmo papel que a matriz dos estados. Salienta-se ainda que a matriz dos estados suspensa descrita pela Figura 3.4 é obtida por meio de uma técnica conhecida como suspensão de variáveis. Desse modo, é possível estender a metodologia SINDy a fim de lidar com sistemas que sejam intrinsecamente descritos por equações diferenciais ordinárias de ordens superiores [70]. Por fim, com o auxílio das Equações 3-1 e 3-5, é tangível concluir que os vetores de esparsidade  $\xi_i$  ( $i = 1, 2, \dots, s$ ) podem ser interpretados como pesos, que quando multiplicados pela matriz de dados aumentada  $\Theta$  ( $\Theta \in \mathbb{R}^{n \times f}$ ) são capazes de reconstruir de maneira aproximada a dinâmica do sistema.

### 3.2

#### Derivadas Temporais

Em geral, as derivadas temporais dos estados de um dado sistema podem ser obtidas diretamente ou adquiridas por intermédio de estimativas. Mais especificamente, a obtenção direta de tais derivadas pode ser realizada por meio de sensores. Por outro lado, é possível acessar as derivadas dos estados de um dado sistema a partir de técnicas de aproximação [26]. Ressalta-se ainda que a realização de medições diretas de tais derivadas podem se tornar inviáveis em certos cenários. Precisamente, a obtenção das medições citadas pode ser dificultada ou inviabilizada caso sejam verificados problemas como elevado custo atrelado à aquisição dos sensores requeridos, inexistência de tecnologia disponível para construção de tais sensores, admissão de sensores defeituosos, dificuldade no acesso físico aos pontos de medição, assim como a instalação incorreta dos sensores necessários para tal finalidade.

Almejando evitar tais problemas, assume-se nesta tese que as referidas derivadas são adquiridas a partir de aproximações. Além disso, embora a metodologia SINDy explore de forma mais expressiva circunstâncias onde são admitidas a medição completa dos estados e de suas derivadas, a estrutura salientada também disserta sobre a estimação das derivadas dos estados a partir de uma abordagem via diferenças centrais [25]. Pontua-se também que uma certa perda de informação é verificada quando somente a abordagem citada é considerada. De maneira detalhada, a referida perda pode ser interpretada como uma diminuição no volume de dados utilizados na alimentação do algoritmo de identificação adotado. Daí, almejando contornar o problema proveniente da perda de informação, são implementados três esquemas distintos de diferenças finitas com o propósito de aproximar as derivadas dos estados, como expõe a Figura 3.5.

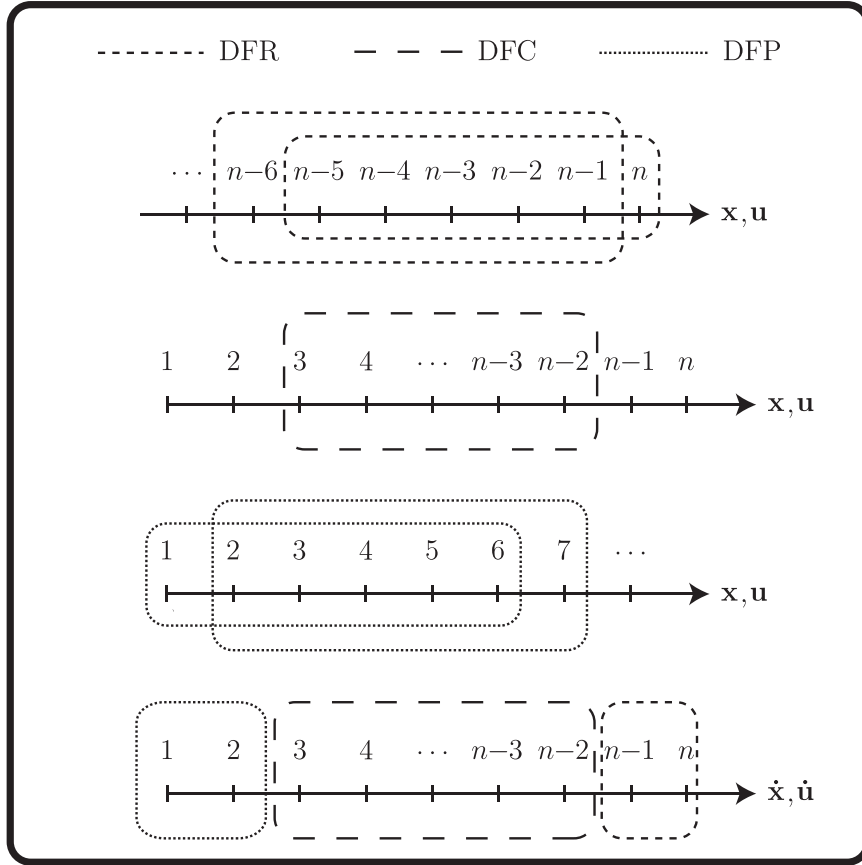


Figura 3.5: Detalhamento da implementação do DFR, DFC e DFP.

Especificamente, a Figura 3.5 tem como finalidade descrever quais amostras dos estados e do sinal de entrada devem ser utilizadas em cada um dos esquemas previamente salientados, assim como quais amostras das derivadas tanto dos estados quanto do referido sinal podem ser obtidas a partir de tais esquemas. Explica-se ainda que a metodologia proposta utiliza versões dos esquemas de Diferenças Finitas Regressivas (DFR), Diferenças Finitas Centrais (DFC) e Diferenças Finitas Progressivas (DFP) com erros da ordem de  $\mathcal{O}\{\Delta t^4\}$  [73]. Formalmente, expõe-se que a Equação 3-14 apresenta uma versão do esquema DFP com erro de quarta ordem, onde esta viabiliza a obtenção das derivadas dos estados para as duas primeiras amostras.

$$\left. \frac{d\mathbf{x}_i(t)}{dt} \right|_{\text{DFP}} \approx \frac{1}{12} \frac{1}{\Delta t} \left[ -25\mathbf{x}_i(t) + 48\mathbf{x}_i(t + \Delta t) - 36\mathbf{x}_i(t + 2\Delta t) + \right. \quad (3-14) \\ \left. + 16\mathbf{x}_i(t + 3\Delta t) - 3\mathbf{x}_i(t + 4\Delta t) \right]$$

De modo similar, fundamentado no esquema DFP, pode-se utilizar a Equação 3-15 com o propósito de obter a derivada do sinal de excitação para as duas primeiras amostras.

$$\left. \frac{d\mathbf{u}(t)}{dt} \right|_{\text{DFP}} \approx \frac{1}{12} \frac{1}{\Delta t} \left[ -25\mathbf{u}(t) + 48\mathbf{u}(t + \Delta t) - 36\mathbf{u}(t + 2\Delta t) + \right. \\ \left. + 16\mathbf{u}(t + 3\Delta t) - 3\mathbf{u}(t + 4\Delta t) \right] \quad (3-15)$$

Por outro lado, suportado pelo esquema DFC, é possível acessar as derivadas de cada um dos estados para as amostras centrais, isto é, as amostras ensandui-chadas pelas duas primeiras e duas últimas amostras. Assim, a Equação 3-16 descreve uma maneira de estimar as derivadas dos estados para tais amostras com erros de quarta ordem.

$$\left. \frac{d\mathbf{x}_i(t)}{dt} \right|_{\text{DFC}} \approx \frac{1}{12} \frac{1}{\Delta t} \left[ \mathbf{x}_i(t - 2\Delta t) - 8\mathbf{x}_i(t - \Delta t) + 8\mathbf{x}_i(t + \Delta t) + \right. \\ \left. - \mathbf{x}_i(t + 2\Delta t) \right] \quad (3-16)$$

De maneira análoga, ao levar em conta as amostras centrais, a derivada do sinal de entrada também pode ser estimada a partir do esquema DFC, como exprime a Equação 3-17.

$$\left. \frac{d\mathbf{u}(t)}{dt} \right|_{\text{DFC}} \approx \frac{1}{12} \frac{1}{\Delta t} \left[ \mathbf{u}(t - 2\Delta t) - 8\mathbf{u}(t - \Delta t) + 8\mathbf{u}(t + \Delta t) + \right. \\ \left. - \mathbf{u}(t + 2\Delta t) \right] \quad (3-17)$$

Não obstante, para as duas últimas amostras, a Equação 3-18 propõe uma versão do esquema DFR com erro de quarta ordem, viabilizando, por sua vez, a obtenção das derivadas dos estados do sistema.

$$\left. \frac{d\mathbf{x}_i(t)}{dt} \right|_{\text{DFR}} \approx \frac{1}{12} \frac{1}{\Delta t} \left[ 3\mathbf{x}_i(t - 4\Delta t) - 16\mathbf{x}_i(t - 3\Delta t) + \right. \\ \left. + 36\mathbf{x}_i(t - 2\Delta t) - 48\mathbf{x}_i(t - \Delta t) + 25\mathbf{x}_i(t) \right] \quad (3-18)$$

Explicita-se ainda que o esquema DFR também pode ser empregado com o objetivo de acessar a derivada do sinal de entrada para as duas últimas amostras, como descreve a Equação 3-19.

$$\left. \frac{d\mathbf{u}(t)}{dt} \right|_{\text{DFR}} \approx \frac{1}{12} \frac{1}{\Delta t} \left[ 3\mathbf{u}(t - 4\Delta t) - 16\mathbf{u}(t - 3\Delta t) + \right. \\ \left. + 36\mathbf{u}(t - 2\Delta t) - 48\mathbf{u}(t - \Delta t) + 25\mathbf{u}(t) \right] \quad (3-19)$$

Note que a perda de informação inerente às duas primeiras e às duas últimas amostras pode ser facilmente constatada a partir da Equação 3-16 ou da Equação 3-17. Analogamente, uma certa perda de informação também é verificada caso seja adotado somente o esquema DFP. De forma equivalente, a implementação exclusiva do esquema DFR também implica na perda de informação. Assim, visando evitar quaisquer perdas de informações, deve-se implementar os três esquemas apresentados de forma conjunta, possibilitando a estimativa integral das derivadas dos estados e do sinal de excitação. Observe que a adoção de técnicas que visam estimar as derivadas dos estados a partir dos estados previamente mensurados viabilizam o emprego da metodologia proposta. Em contraste, se uma dada aplicação disponibilizar as derivadas dos estados, é altamente recomendável utilizá-las, pois espera-se que a estratégia proposta obtenha um melhor desempenho quando submetida a um cenário onde não haja necessidade de aproximar tais grandezas. Finalmente, destaca-se que o Apêndice A deste trabalho apresenta esquemas de diferenças finitas com erros de segunda ordem e sexta ordem.

### 3.3

#### Corrupção Ruidosa

Excetuando condições ideais, tem-se que o ruído é uma entidade que está intrinsecamente ligada ao processo de aquisição de dados. Na prática, a corrupção ruidosa de tais dados pode ser gerada por problemas como sensores defeituosos, sensores instalados incorretamente ou, até mesmo, interferências. Precisamente, a aquisição de sinais ruidosos pode ser responsável por dificultar ou inviabilizar a tarefa de identificação. Dessa maneira, é importante descobrir sob que condições uma dada metodologia consegue operar sem ter sua performance significativamente prejudicada por cenários ruidosos, ou seja, é relevante efetuar uma análise de robustez ao ruído da metodologia admitida. Além disso, baseado na baixa robustez ao ruído da abordagem SINDy, expõe-se que avanços visando melhorar tal robustez podem alavancar a adoção da referida abordagem [26]. Sendo assim, é necessário ter acesso a diversos conjuntos de dados, onde estes contemplam cenários ruidosos distintos. Salienta-se ainda que, em problemas numéricos, é possível gerar os referidos cenários ruidosos de forma sintética, onde estes são introduzidos artificialmente por meio de uma operação matemática conhecida como produto de Hadamard [74]. De maneira formal, o produto de Hadamard entre a matriz  $\mathbf{A}$  ( $\mathbf{A} \in \mathbb{R}^{a \times b}$ ) e a matriz  $\mathbf{B}$  ( $\mathbf{B} \in \mathbb{R}^{a \times b}$ ), onde estas possuem um certo número de linhas  $a$  ( $a \in \mathbb{N}^*$ ) e uma determinada quantidade de colunas  $b$  ( $b \in \mathbb{N}^*$ ), é designado pela matriz  $\mathbf{H}$  ( $\mathbf{H} \in \mathbb{R}^{a \times b}$ ), como explicita a Equação 3-20.

$$\mathbf{H} = h_{ij} = a_{ij}b_{ij} = \mathbf{A} \odot \mathbf{B} \quad (i, j = [1, 2, \dots, a] \times [1, 2, \dots, b]) \quad (3-20)$$

Matematicamente, a partir da matriz dos estados não ruidosa  $\mathbf{X}$  ( $\mathbf{X} \in \mathbb{R}^{n \times s}$ ), de um percentual de ruído  $\eta$  ( $\eta \in \mathbb{R}_+$ ), da matriz contendo o desvio padrão de cada estado  $\mathbf{Z}$  ( $\mathbf{Z} \in \mathbb{R}^{n \times s}$ ) e de uma matriz possuindo entradas aleatórias segundo uma distribuição normal com média nula e variância unitária  $\mathbf{G}$  ( $\mathbf{G} \in \mathbb{R}^{n \times s}$ ) é tangível obter a matriz dos estados ruidosa  $\mathbf{X}_\eta$  ( $\mathbf{X}_\eta \in \mathbb{R}^{n \times s}$ ), como explica a Equação 3-21. Desse modo, pode-se adicionar ruído artificial a um conjunto de dados previamente adquirido [25].

$$\mathbf{X}_\eta = \mathbf{X} + \eta(\mathbf{Z} \odot \mathbf{G}) \quad (3-21)$$

Assim, cada coluna da matriz  $\mathbf{Z}$  ( $\mathbf{Z} \in \mathbb{R}^{n \times s}$ ) é composta por termos idênticos. Por exemplo, cada elemento da segunda coluna de tal matriz é dado pelo desvio padrão do segundo estado, como descreve a Equação 3-22.

$$\mathbf{Z} = \begin{bmatrix} | & | & | & \cdots & | \\ \mathcal{Z}\{\mathbf{x}_1\} & \mathcal{Z}\{\mathbf{x}_2\} & \mathcal{Z}\{\mathbf{x}_3\} & \cdots & \mathcal{Z}\{\mathbf{x}_s\} \\ | & | & | & \cdots & | \end{bmatrix} \quad (3-22)$$

Formalmente, baseado na Equação 3-23, pode-se computar facilmente o desvio padrão de cada estado a partir das evoluções históricas dos mesmos  $\mathbf{x}_i$  ( $i = 1, 2, \dots, s$ ) e de suas respectivas médias  $\bar{x}_i$  ( $i = 1, 2, \dots, s$ ) [75].

$$\mathcal{Z}\{\mathbf{x}_i\} = \sqrt{\frac{1}{n-1} \sum_{k=1}^n [x_i(k) - \bar{x}_i]^2} \quad (i = 1, 2, \dots, s) \quad (3-23)$$

Note que os erros obtidos pelos métodos apresentados a fim de estimar as derivadas dos estados e do sinal de entrada são amplificados pelo cenário ruidoso [76]. Em adição, explana-se que a evolução das variáveis de estado dos estudos de caso abordados nesta tese são inicialmente obtidos de forma numérica a partir da versão forçada e não ruidosa dos mesmos. Em seguida, suportado pela solução numérica previamente adquirida, pode-se gerar artificialmente diversos conjuntos de dados corrompidos pelo ruído aditivo, viabilizando a análise de robustez ao ruído tanto da estratégia pioneira quanto da estrutura proposta. Por fim, destaca-se que diferentes tipos de estatísticas são abordadas no Apêndice A deste material.

### 3.4

#### Pré-processamento

O pré-processamento de dados pode ser descrito como o conjunto de transformações que visam preparar os dados antes dos mesmos serem utilizados em um dado algoritmo de aprendizado de máquina. Precisamente, enfatiza-se que pré-processamentos como centralização, mudança de escala, normalização e padronização são vastamente adotados [20]. Expõe-se ainda que a adoção de tais processamentos prévios tem a capacidade de melhorar o desempenho de algumas técnicas e estratégias de aprendizado de máquina. Em detalhes, pode-se observar que certos tipos de pré-processamentos são imperativos quando métodos específicos de aprendizado de máquina são implementados [77]. Por exemplo, pode-se destacar que uma técnica de redução de dimensionalidade denominada Análise de Componentes Principais (ACP) requer a centralização prévia dos dados [13].

Em contraste, ressalta-se que nenhuma das extensões da estrutura pioneira SINDy presentes na literatura exigem qualquer tipo de pré-processamento. Salienta-se também que, com a finalidade de reconstruir os dados aumentados, o algoritmo de aprendizado de máquina apresentado nesta tese utiliza o método de redução de dimensionalidade previamente destacado. Dessa forma, ao considerar a implementação da metodologia proposta SINDyFE, é possível concluir que a centralização prévia da matriz de dados aumentada  $\Theta$  ( $\Theta \in \mathbb{R}^{n \times f}$ ) é mandatória. Sendo assim, como mostra a Equação 3-24, a matriz de dados aumentada centralizada  $\Theta^C$  ( $\Theta^C \in \mathbb{R}^{n \times f}$ ) pode ser obtida a partir da diferença entre as colunas da matriz de dados aumentada  $\theta_i$  ( $i = 1, 2, \dots, f$ ) e a média dos elementos presentes em cada uma das colunas citadas  $\bar{\theta}_i$  ( $i = 1, 2, \dots, f$ ).

$$\Theta^C = \theta_i^C = \theta_i - \bar{\theta}_i \quad (i = 1, 2, \dots, f) \quad (3-24)$$

Por outro lado, baseado na mudança de escala, salienta-se um pré-processamento intitulado reescala. Especificamente, a fim de gerar a matriz de dados aumentada reescalada  $\Theta^R$  ( $\Theta^R \in \mathbb{R}^{n \times f}$ ), deve-se transformar cada coluna da matriz de dados aumentada, produzindo, por sua vez, colunas detentoras de um valor ínfimo  $\theta^{\inf}$  ( $\theta^{\inf} \in \mathbb{R}$ ) e de um valor supremo  $\theta^{\sup}$  ( $\theta^{\sup} \in \mathbb{R}$ ), como propõe a Equação 3-25.

$$\Theta^R = \theta_i^R = \frac{\theta^{\sup} - \theta^{\inf}}{\theta_i^{\max} - \theta_i^{\min}} (\theta_i - \theta_i^{\min}) + \theta^{\inf} \quad (i = 1, 2, \dots, f) \quad (3-25)$$



Ressalta-se ainda que tal valor ínfimo deve ser estritamente menor do que o valor supremo citado. De modo complementar, explicita-se que os valores  $\theta_i^{\min} (i = 1, 2, \dots, f)$  e  $\theta_i^{\max} (i = 1, 2, \dots, f)$  presentes na Equação 3-25, representam, respectivamente, os valores mínimos e máximos existentes ao longo das colunas da matriz de dados aumentada. Explica-se ainda que, diferentemente do valor ínfimo e do valor supremo, o valor mínimo pode ser menor ou igual ao valor máximo. Em adição, pontua-se que tanto o simétrico da unidade quanto zero são valores comumente adotados como ínfimo, enquanto que a unidade é amplamente utilizada como valor supremo. Além disso, uma transformação denominada normalização também pode ser admitida como técnica de pré-processamento. Detalhadamente, a fim de acessar a matriz de dados aumentada normalizada  $\Theta^N (\Theta^N \in \mathbb{R}^{n \times f})$ , deve-se dividir os elementos de cada coluna da matriz de dados aumentada pela norma euclidiana de sua respectiva coluna, como reitera a Equação 3-26.

$$\Theta^N = \theta_i^N = \frac{\theta_i}{\|\theta_i\|} \quad (i = 1, 2, \dots, f) \quad (3-26)$$

Por fim, destaca-se que a técnica de pré-processamento intitulada padronização é capaz de produzir dados com média nula e variância unitária. Sendo assim, por intermédio do desvio padrão de cada característica existente na matriz de dados aumentada, é exequível acessar a matriz de dados aumentada padronizada  $\Theta^S (\Theta^S \in \mathbb{R}^{n \times f})$ , como apresenta a Equação 3-27.

$$\Theta^S = \theta_i^S = \frac{\theta_i^C}{\mathcal{Z}\{\theta_i\}} \quad (i = 1, 2, \dots, f) \quad (3-27)$$

Observe que todas as operações de pré-processamento apresentadas nesta tese são efetuadas ao longo das colunas da matriz de dados aumentada. Pontua-se também que as transformações de reescala e normalização não necessitam que a matriz de dados aumentada seja previamente centralizada e que, caso um dado algoritmo de aprendizado de máquina necessite ser alimentado por dados normalmente distribuídos, deve-se submeter o conjunto de dados inicialmente adquirido à padronização. Ressalta-se ainda que a Figura 3.6 visa prover uma melhor compreensão no que diz respeito à adoção de técnicas de pré-processamento sob a perspectiva da identificação esparsa, fornecendo, por sua vez, uma sequência de passos que contemplam procedimentos, como aquisição de dados, corrupção ruidosa, aproximação das derivadas, construção da matriz de dados aumentada e pré-processamento.

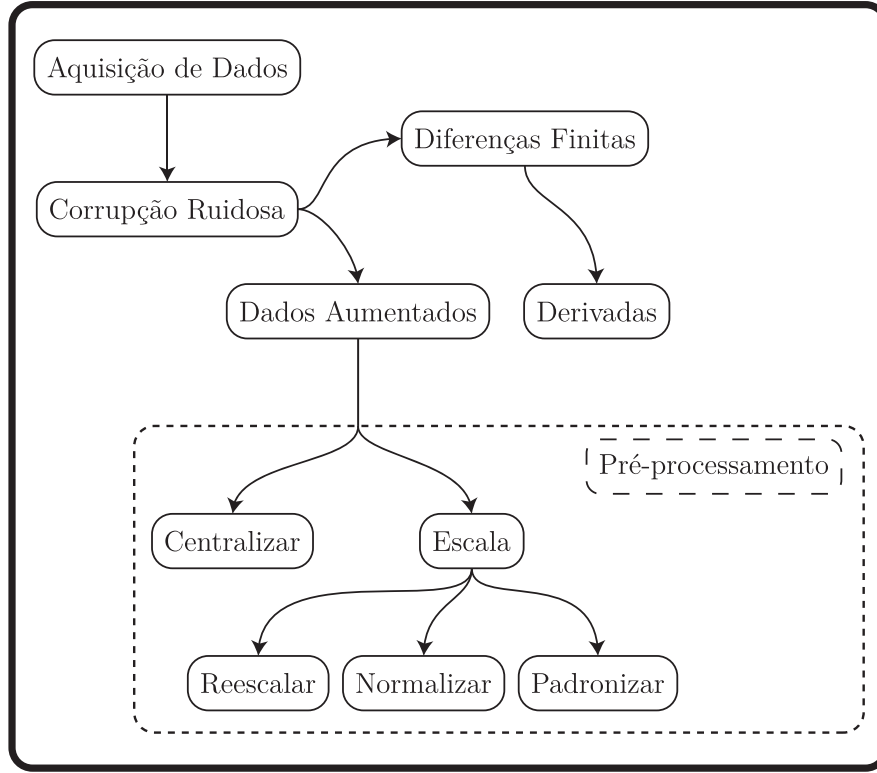


Figura 3.6: Diferentes técnicas de pré-processamento de dados sob a perspectiva da identificação esparsa.

### 3.5

#### Normas e Regressões

A fim de realizar a tarefa de identificação é necessário admitir uma dada família de regressão. Notavelmente, destaca-se que diferentes tipos de regressões são suportadas por normas vetoriais distintas, onde estas descrevem como a magnitude de um vetor pode ser obtida [78]. Formalmente, a norma- $p$  de um vetor  $\mathbf{w}$  ( $\mathbf{w} \in \mathbb{R}^n$ ) pode ser definida a partir da Equação 3-28.

$$\|\mathbf{w}\|_p = \left[ \sum_{k=1}^n |w(k)|^p \right]^{\frac{1}{p}} \quad (p \in \mathbb{Q}_+) \quad (3-28)$$

Observe ainda que a Equação 3-28 é capaz de gerar diferentes maneiras de mensurar distâncias. Em particular, explicita-se que a Equação 3-29 descreve a métrica de Manhattan [79].

$$\|\mathbf{w}\|_1 = \sum_{k=1}^n |w(k)| \quad (3-29)$$

Por outro lado, enfatiza-se que a Equação 3-28 é capaz de representar a norma euclidiana, onde esta é dada pela Equação 3-30.

$$\|\mathbf{w}\|_2 = \|\mathbf{w}\| = \left[ \sum_{k=1}^n |w(k)|^2 \right]^{\frac{1}{2}} \quad (3-30)$$

Matematicamente, geometrias distintas estão associadas a diferentes tipos de métricas. Especificamente, a geometria do táxi é baseada na métrica de Manhattan, ao passo que a geometria de Euclides é fundamentada na norma euclidiana. Observe ainda que a Equação 3-29 utiliza o conceito de módulo em sua formulação. Tecnicamente, pontua-se que o conceito de módulo está atrelado a grandezas escalares, enquanto que o conceito de norma está vinculado a grandezas vetoriais e matriciais. Ressalta-se ainda que a Equação 3-31 descreve a métrica de Chebyshev [80].

$$\|\mathbf{w}\|_\infty = \max \left\{ |w(k)| \right\} \quad (k = 1, 2, \dots, n) \quad (3-31)$$

Explica-se ainda que a Equação 3-28 é capaz de gerar infinitas relações. No entanto, algumas propriedades devem ser satisfeitas para que tais relações sejam consideradas efetivamente uma norma [10]. Almejando explorar tais propriedades, admite-se tanto um escalar  $\alpha$  ( $\alpha \in \mathbb{R}$ ) quanto os vetores  $\mathbf{w}_1$  ( $\mathbf{w}_1 \in \mathbb{R}^n$ ) e  $\mathbf{w}_2$  ( $\mathbf{w}_2 \in \mathbb{R}^n$ ). Sendo assim, expõe-se que uma norma efetiva deve satisfazer a propriedade da não negatividade.

$$\|\mathbf{w}_1\|_p \geq 0 \quad \left( \|\mathbf{w}_1\|_p = 0 \iff \mathbf{w}_1 = 0 \right) \quad (3-32)$$

Adicionalmente, a desigualdade triangular estabelece a segunda propriedade que uma norma deve verificar.

$$\|\mathbf{w}_1 + \mathbf{w}_2\|_p \leq \|\mathbf{w}_1\|_p + \|\mathbf{w}_2\|_p \quad (3-33)$$

Finalmente, destaca-se que a Equação 3-34 detalha a terceira e última propriedade que uma norma efetiva deve gozar.

$$\|\alpha \mathbf{w}_1\|_p = |\alpha| \|\mathbf{w}_1\|_p \quad (3-34)$$

Dessa maneira, a Equação 3-28 é capaz de gerar uma norma efetiva quando as três propriedades descritas anteriormente são verificadas, caso contrário uma pseudonorma é obtida. Em particular, explica-se que a Equação 3-35 designa uma pseudonorma, onde esta é dada a partir da quantidade de elementos não nulos presentes no vetor considerado.

$$\|\mathbf{w}\|_0 = \mathcal{N}\{\mathbf{w}\} \quad (3-35)$$

Pontua-se ainda que a métrica de Manhattan é uma escolha natural a fim de lidar com a regressão esparsa, pois a mesma é capaz de encolher uma coleção de variáveis até valores identicamente nulos [27]. Detalhadamente, a Figura 3.7 ilustra diferentes representações gráficas inerentes a diversas famílias de pseudonormas e normas.

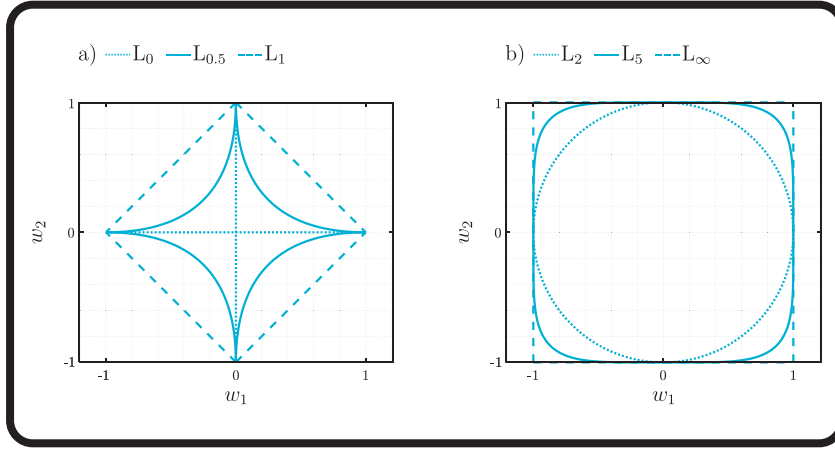


Figura 3.7: Diferentes famílias de pseudonormas e normas: a) Representação gráfica das pseudonormas unitárias  $L_0$  e  $L_{0.5}$  e da norma unitária  $L_1$ ; b) Representação gráfica das normas unitárias  $L_2$ ,  $L_5$  e  $L_\infty$ .

Desse modo, suportado pela regressão esparsa, é possível encolher e selecionar variáveis, visando produzir um modelo esparsa de um dado sistema [25]. Em detalhes, deseja-se obter os vetores de esparsidade  $\xi_i$  ( $i = 1, 2, \dots, s$ ), onde estes são responsáveis pela geração de certas estimativas  $\hat{y}_i$  ( $i = 1, 2, \dots, s$ ), que, por sua vez, representam aproximações de uma coleção de variáveis reais  $y_i$  ( $i = 1, 2, \dots, s$ ). Almejando tal objetivo, pode-se exprimir a diferença entre os valores reais do sistema e as estimativas destes a partir dos vetores de erro  $\epsilon_i$  ( $i = 1, 2, \dots, s$ ), como apresenta a Equação 3-36.

$$\epsilon_i = y_i - \hat{y}_i \quad (i = 1, 2, \dots, s) \quad (3-36)$$

Em adição, sob a perspectiva da identificação de sistemas dinâmicos, as variáveis  $\mathbf{y}_i$  ( $i = 1, 2, \dots, s$ ) e  $\hat{\mathbf{y}}_i$  ( $i = 1, 2, \dots, s$ ) apresentadas previamente designam as derivadas dos estados e suas estimativas, respectivamente. Assim, baseado na penalização  $L_1$ , pode-se admitir uma estratégia conhecida como Operador de Seleção e Encolhimento Mínimo Absoluto (OSEMA). Salienta-se ainda que a partir da estratégia citada é tangível definir a função de perda  $\mathcal{V}_1$  ( $\mathcal{V}_1 \in \mathbb{R}$ ), como mostra a Equação 3-37.

$$\mathcal{V}_1 = \underset{\boldsymbol{\xi}}{\operatorname{argmin}} \left\{ \sum_{k=1}^n [\boldsymbol{\varepsilon}_i(k)]^2 + \lambda \sum_{l=1}^s \|\boldsymbol{\xi}_l\|_1 \right\} \quad (3-37)$$

O hiperparâmetro de esparsidade  $\lambda$  ( $\lambda \in \mathbb{R}_+^*$ ) presente na Equação 3-37 é o termo incumbido de promover as penalizações vinculadas à família de regressão admitida. Além disso, a promoção da esparsidade inerente à regressão OSEMA determina que alguns termos sejam ajustados para valores identicamente nulos. Sendo assim, tal regressão opera como uma ferramenta seletora, fazendo com que as variáveis menos relevantes se tornem iguais a zero ao longo do processo de regressão [81]. Ressalta-se ainda que, embora a ocorrência de falsas descobertas por meio da regressão OSEMA tenha sido evidenciada, o referido hiperparâmetro de esparsidade é capaz de equilibrar a fidedignidade de um dado modelo e a esparsidade vinculada ao mesmo [82]. Contrastivamente, a partir do hiperparâmetro regularizador  $\lambda_R$  ( $\lambda_R \in \mathbb{R}_+^*$ ), pode-se explorar outras famílias de regressão, onde estas são baseadas em outros tipos de normas. Dessa forma, como apresenta a Equação 3-38, é possível acessar a função de perda  $\mathcal{V}_0$  ( $\mathcal{V}_0 \in \mathbb{R}$ ) por meio da penalização  $L_0$  [83].

$$\mathcal{V}_0 = \underset{\boldsymbol{\xi}}{\operatorname{argmin}} \left\{ \sum_{k=1}^n [\boldsymbol{\varepsilon}_i(k)]^2 + \lambda_R \sum_{l=1}^s \|\boldsymbol{\xi}_l\|_0 \right\} \quad (3-38)$$

Destaca-se ainda que o hiperparâmetro regularizador apresentado previamente é apenas uma generalização do hiperparâmetro de esparsidade. Adicionalmente, por intermédio da penalização  $L_2$ , uma classe de regressão conhecida como regressão de crista é estabelecida [84]. Dessa maneira, com o auxílio da Equação 3-39, é tangível descrever a função de perda  $\mathcal{V}_2$  ( $\mathcal{V}_2 \in \mathbb{R}$ ) associada à regressão citada anteriormente.

$$\mathcal{V}_2 = \underset{\boldsymbol{\xi}}{\operatorname{argmin}} \left\{ \sum_{k=1}^n [\boldsymbol{\varepsilon}_i(k)]^2 + \lambda_R \sum_{l=1}^s \|\boldsymbol{\xi}_l\|_2 \right\} \quad (3-39)$$

Por outro lado, enfatiza-se que uma família de regressão conhecida como rede elástica pode ser obtida a partir da função de perda  $\mathcal{V}_{1-2}$  ( $\mathcal{V}_{1-2} \in \mathbb{R}$ ). Mais formalmente, como reitera a Equação 3-40, a referida função de perda é suportada pela hibridização das penalizações  $L_1$  e  $L_2$  [85].

$$\mathcal{V}_{1-2} = \underset{\xi}{\operatorname{argmin}} \left\{ \sum_{k=1}^n [\varepsilon_i(k)]^2 + \lambda_R \sum_{l=1}^s \|\xi_l\|_1 + \lambda_R \sum_{l=1}^s \|\xi_l\|_2 \right\} \quad (3-40)$$

Em geral, independentemente do tipo de regressão admitida, é recomendável efetuar uma busca unidimensional com a finalidade de obter um hiperparâmetro que viabilize a geração de um modelo representativo do sistema abordado [86]. Em outras palavras, a partir de uma coleção de hiperparâmetros, procura-se um hiperparâmetro ótimo, onde este é capaz de entregar um modelo fiel no que diz respeito à dinâmica do sistema [25]. De maneira genérica, sob a perspectiva das regressões penalizadas, destaca-se que é necessário encontrar o hiperparâmetro responsável pela minimização das funções de perda correspondentes, pois tal hiperparâmetro é responsável pela obtenção de um modelo não espúrio do sistema. Pontua-se também que nenhuma abordagem com a finalidade de realizar a busca hiperparamétrica citada é descrita no artigo original da metodologia SINDy [25].

Todavia, salienta-se que uma técnica de validação cruzada intitulada *K-fold* é uma estratégia rotineiramente utilizada a fim de realizar a referida busca [20, 34]. Dito isso, é recomendável admitir a validação cruzada com o propósito de acessar um hiperparâmetro capaz de prover uma descrição matemática fidedigna do sistema considerado. Por inspeção, é exequível constatar que a assunção de um hiperparâmetro numericamente igual a zero implica na degeneração das funções de perda apresentadas anteriormente. Em particular, ao admitir um hiperparâmetro nulo, pode-se concluir que as funções de perda previamente descritas degeneram para uma versão não penalizada da regressão. Desse modo, como propõe a Equação 3-41, a função de perda não penalizada  $\mathcal{V}$  ( $\mathcal{V} \in \mathbb{R}$ ) pode ser descrita por intermédio de uma abordagem amplamente conhecida como Mínimos Quadrados Ordinários (MQO).

$$\mathcal{V} = \underset{\xi}{\operatorname{argmin}} \left\{ \sum_{k=1}^n [\varepsilon_i(k)]^2 \right\} \quad (3-41)$$

Precisamente, pode-se implementar o Algoritmo 1 a fim de encontrar o conjunto de parâmetros intrínsecos a uma dada metodologia que são capazes de minimizar a função de perda não penalizada.

**Algoritmo 1** Mínimos Quadrados Ordinários

**Entrada:** Derivadas dos estados  $\dot{\mathbf{X}}$  ( $\dot{\mathbf{X}} \in \mathbb{R}^{n \times s}$ ), matriz de dados aumentada  $\Theta$  ( $\Theta \in \mathbb{R}^{n \times f}$ ), palpite inicial da solução  $\Xi^\bullet$  ( $\Xi^\bullet \in \mathbb{R}^{f \times s}$ ), taxa de aprendizado  $\mu$  ( $\mu \in \mathbb{R}_+^*$ ), número máximo de iterações  $I$  ( $I \in \mathbb{N}^*$ ) e tolerância mínima admitida  $T$  ( $T \in \mathbb{R}_+^*$ ).

**Saída:** Modelo obtido por meio da regressão não penalizada.

---

```

1: procedure OLS( $\dot{\mathbf{X}}, \Theta, \Xi^\bullet, \mu, I, T$ )
2:    $\hat{\Xi} \leftarrow \Xi^\bullet$  ▷ Palpite Inicial
3:    $\hat{\mathbf{X}} \leftarrow \text{calculate\_derivatives}(\Theta, \hat{\Xi})$  ▷ Calcular Derivadas
4:    $\epsilon \leftarrow \dot{\mathbf{X}} - \hat{\mathbf{X}}$  ▷ Calcular Resíduos
5:    $\text{itn} \leftarrow 1$  ▷ Primeira Iteração
6:   while ( $\text{itn} \leq I$ ) or ( $\epsilon > T$ ) do ▷ Critérios de Parada
7:      $\nabla \leftarrow \text{gradient\_norm}(\epsilon, \Theta)$  ▷ Calcular Gradiente Unitário
8:      $\nabla^\mu \leftarrow \mu \nabla$  ▷ Atualizar Gradiente
9:      $\hat{\Xi} \leftarrow \hat{\Xi} + \nabla^\mu$  ▷ Atualizar Estimativa
10:     $\hat{\mathbf{X}} \leftarrow \text{calculate\_derivatives}(\Theta, \hat{\Xi})$  ▷ Atualizar Derivadas
11:     $\epsilon \leftarrow \dot{\mathbf{X}} - \hat{\mathbf{X}}$  ▷ Calcular Resíduos
12:     $\text{itn} \leftarrow \text{itn} + 1$  ▷ Atualizar Iteração
13:  end while
14:  return  $\hat{\Xi}$  ▷ Solução Não Penalizada
15: end procedure

```

---

Especificamente, por meio de uma estratégia denominada descida de encosta, realça-se que o algoritmo previamente explicitado é descrito de forma mais detalhada no Apêndice A deste trabalho [87]. Em adição, explica-se que os termos atrelados à regressão penalizada  $L_2$  podem ser encolhidos durante o processo de regressão para valores próximos de zero. Alternativamente, a fim de que alguns termos sejam ajustados para valores exatamente iguais a zero, deve-se adotar a regressão penalizada  $L_0$  ou admitir a regressão penalizada  $L_1$ . Entretanto, a complexidade computacional vinculada à implementação da regressão penalizada  $L_0$ , favorece a utilização da regressão penalizada  $L_1$ . Ressalta-se ainda que, assim como na estrutura SINDy, a adoção da regressão OSEMA na metodologia SINDyFE é justificada pela facilidade de implementação e promoção da esparsidade. Explicita-se ainda que, embora a estratégia pioneira e a estrutura proposta sejam alicerçadas na regressão penalizada  $L_1$ , é crucial mencionar que o procedimento de otimização adotado pelas abordagens consideradas nesta tese possuem diferenças substanciais. Além disso, destaca-se que a partir do Algoritmo 2 é possível implementar a regressão OSEMA, onde esta estabelece a penalização de quaisquer termos presentes nos vetores de esparsidade  $\xi_i$  ( $i = 1, 2, \dots, s$ ) cujo módulo é menor do que um dado hiperparâmetro de esparsidade  $\lambda$  ( $\lambda \in \mathbb{R}_+$ ) [27].

**Algoritmo 2** Operador de Seleção e Encolhimento Mínimo Absoluto

**Entrada:** Derivadas dos estados  $\dot{\mathbf{X}}$  ( $\dot{\mathbf{X}} \in \mathbb{R}^{n \times s}$ ), matriz de dados aumentada  $\Theta$  ( $\Theta \in \mathbb{R}^{n \times f}$ ), palpite inicial da solução  $\Xi^\bullet$  ( $\Xi^\bullet \in \mathbb{R}^{f \times s}$ ), taxa de aprendizado  $\mu$  ( $\mu \in \mathbb{R}_+^*$ ), hiperparâmetro de esparsidade  $\lambda$  ( $\lambda \in \mathbb{R}_+^*$ ), número máximo de iterações  $I$  ( $I \in \mathbb{N}^*$ ) e tolerância mínima admitida  $T$  ( $T \in \mathbb{R}_+^*$ ).

**Saída:** Modelo obtido a partir da regressão penalizada.

---

```

1: procedure LASSO( $\dot{\mathbf{X}}, \Theta, \Xi^\bullet, \lambda, \mu, I, T$ )
2:    $\hat{\Xi} \leftarrow \text{OLS}(\dot{\mathbf{X}}, \Theta, \Xi^\bullet, \mu, I, T)$  ▷ Palpite Inicial MQO
3:    $s \leftarrow \text{size}(\dot{\mathbf{X}}, 2)$  ▷ Número de Estados
4:    $c \leftarrow 10$  ▷ Número de Ciclos
5:   for  $\text{itn} \in \{1, 2, \dots, c\}$  do ▷ Encolhimento Sucessivo
6:      $I_{\text{small}} \leftarrow \text{abs}(\hat{\Xi}) < \lambda$  ▷ Encontrar Índices
7:      $\hat{\Xi}(I_{\text{small}}) \leftarrow 0$  ▷ Impor Esparsidade
8:     for  $i \in \{1, 2, \dots, s\}$  do ▷ Cada Estado
9:        $I_{\text{big}} \leftarrow \sim I_{\text{small}}(:, i)$  ▷ Cada Índice Ativo
10:       $\dot{\mathbf{X}} \leftarrow \dot{\mathbf{X}}(:, i)$  ▷ Cada Derivada
11:       $\Theta \leftarrow \Theta(:, I_{\text{big}})$  ▷ Cada Característica
12:       $\hat{\Xi} \leftarrow \hat{\Xi}(:, i)$  ▷ Cada Estimativa
13:       $\hat{\Xi}(I_{\text{big}}, i) \leftarrow \text{OLS}(\dot{\mathbf{X}}, \Theta, \hat{\Xi}, \mu, I, T)$  ▷ Regressão MQO
14:    end for
15:  end for
16:  return  $\hat{\Xi}$  ▷ Solução Penalizada
17: end procedure

```

---

Note que o Algoritmo 2 é baseado na admissão de dez ciclos, isto é, o procedimento de encolhimento e seleção de variáveis ocorre uma dezena de vezes. Em linhas gerais, almejando implementar a regressão OSEMA, é necessário fornecer um palpite inicial da solução, onde este é iterativamente incrementado nas direções opostas ao máximo crescimento, como descreve a abordagem intitulada descida de encosta. Dito isso, ao acessar a solução que minimiza a função de perda não penalizada, verifica-se quais termos presentes na referida solução são menores em módulo do que um dado hiperparâmetro de esparsidade e, posteriormente, substitui-se tais termos por zero, configurando, por sua vez, o encolhimento e a seleção de variáveis. Dessa maneira, após as dez iterações serem realizadas, a solução penalizada é obtida. Por fim, explana-se que a metodologia SINDy pode ser reduzida a uma estratégia conhecida como Decomposição em Modos Dinâmicos [88]. Dessa forma, visando obter o caso particular previamente citado, é necessário tanto admitir uma formulação temporalmente discreta para a versão não penalizada da metodologia SINDy quanto assumir que a matriz de dados aumentada produzida pela metodologia salientada é construída somente por termos lineares [89].



### 3.6

#### Engenharia de Recursos

A capacidade de um dado algoritmo de aprendizado de máquina descobrir padrões a partir de um conjunto de dados está diretamente conectada com a qualidade dos dados utilizados na alimentação do algoritmo admitido. Dessa forma, é importante ter acesso a dados representativos e, por conseguinte, dispor de características expressivas o suficiente, viabilizando, por sua vez, o desempenho adequado do algoritmo adotado [43]. Além disso, a admissão de uma grande quantidade de características com baixa representatividade, em detrimento de algumas poucas características relevantes, pode acarretar em um desempenho inapropriado do algoritmo considerado. Em adição, explicita-se que a engenharia de recursos disserta sobre estratégias, técnicas e métodos capazes de introduzir, selecionar e produzir características representativas para um dado problema [20, 87].

Em particular, a seleção de características tem como finalidade identificar e destacar as características mais úteis no que diz respeito à descrição de um sistema dinâmico [20]. Por outro lado, a extração de características visa produzir características mais significativas a partir das características existentes [43]. Ressalta-se ainda que a aquisição de novas características pode ser realizada por intermédio da coleta de novos dados. Todavia, almejando desenvolver uma metodologia versátil, enfatiza-se que abordagens suportadas pela aquisição de novos dados são descartadas. Desse modo, realça-se que a estratégia proposta explora a engenharia de recursos, objetivando automatizar os procedimentos de seleção e extração de características. De maneira precisa, a partir de uma investigação inicial, é tangível acessar quais características dentre as características existentes não são capazes de contribuir de forma significativa no que concerne à dinâmica do sistema analisado [90].

Dito isso, a fim de simplificar a matriz de dados aumentada centralizada, pode-se descartar as características menos expressivas do sistema analisado. Observe também que as características inicialmente obtidas são geradas a partir de uma base de funções previamente definida. Sendo assim, alicerçado na seleção de características, pode-se impor a exclusão de um conjunto de características presentes na matriz de dados aumentada centralizada [91]. Salienta-se ainda que o descarte de uma dada característica por meio da seleção de características, implica na perda total da informação associada à característica excluída. Adicionalmente, explica-se que a Figura 3.8 tem como objetivo personificar um cenário onde uma única característica é descartada dentre as características existentes, revelando a simplicidade inerente ao procedimento de seleção de características.

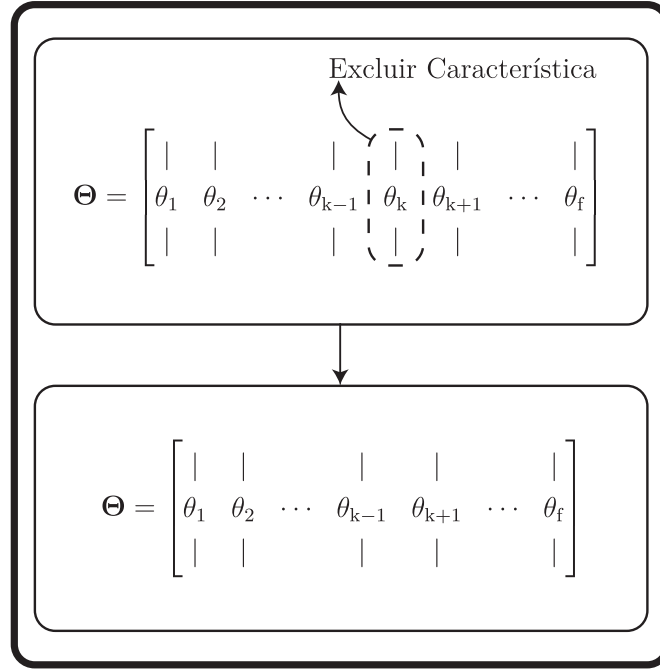


Figura 3.8: Exclusão de características.

Complementarmente, explica-se que a metodologia proposta também explora a engenharia de recursos sob a perspectiva da extração de características. Em específico, uma técnica de redução de dimensionalidade conhecida como Análise de Componentes Principais (ACP) é utilizada [13]. Salienta-se também que, caso seja admitido um grau máximo relativamente elevado para a base polinomial adotada, pode-se observar um aumento substancial no número de características presentes na matriz de dados aumentada, como expõe a Figura 3.3. Dessa forma, ao utilizar a técnica de redução de dimensionalidade previamente apresentada, é possível evitar um problema conhecido como maldição da dimensionalidade e fornecer dados aumentados mais representativos para o algoritmo de aprendizado de máquina empregado [45, 43]. Pontualmente, destaca-se que a técnica de redução de dimensionalidade citada é baseada na criação de novas características, onde estas são geradas a partir das direções de covariância mais significativas no que diz respeito aos dados aumentados. Destaca-se também que, diferentemente da seleção de características, a eliminação de uma característica produzida pelo ACP está vinculada a perda parcial de informação. Detalhadamente, a perda parcial salientada ocorre devido ao fato das novas características serem construídas a partir de certas combinações das características iniciais, onde as características atreladas a uma baixa capacidade de explicar os dados são comumente eliminadas. Expõe-se ainda que a Figura 3.9 tem como finalidade apresentar a direção de máxima covariância associada a um certo conjunto de dados.

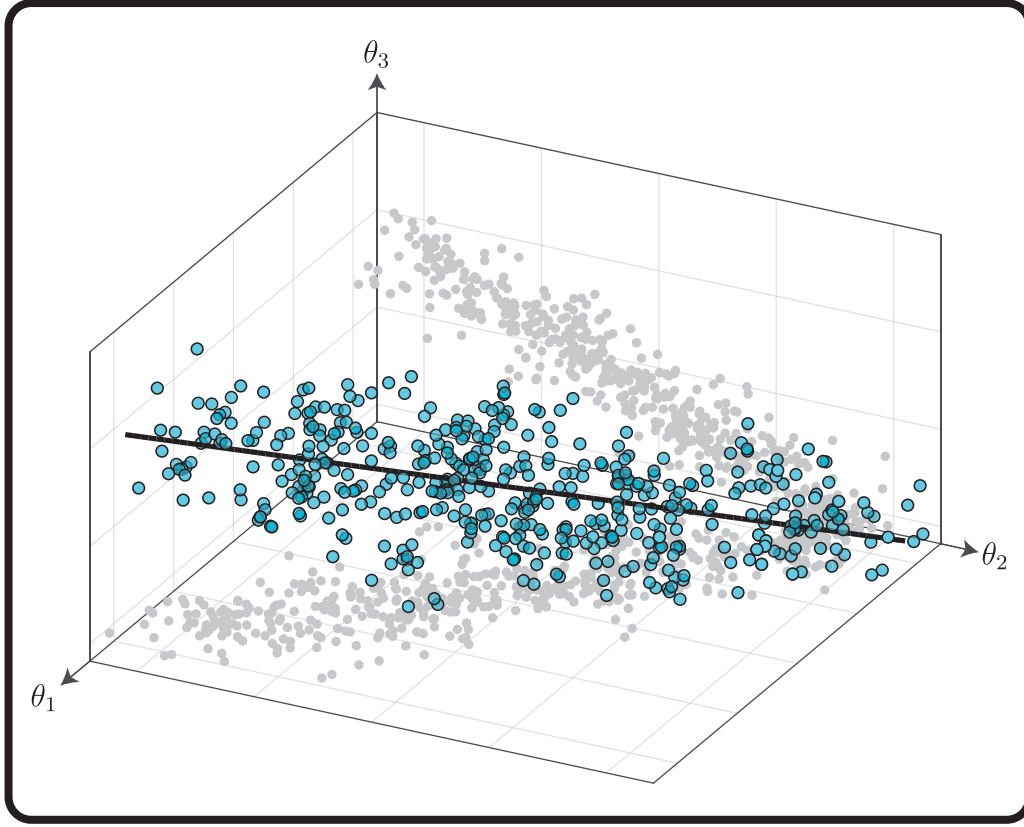


Figura 3.9: Direção de máxima covariância (reta sólida) associada a um conjunto de dados aumentados, onde este pode ser representado por apenas três características.

Pontua-se também que a técnica de redução de dimensionalidade previamente salientada é empregada neste trabalho com o objetivo de reconstruir a matriz de dados aumentada centralizada, onde para implementar tal técnica é necessário realizar a centralização prévia da matriz de dados aumentada. Paralelamente, caso sejam verificadas discrepâncias significativas entre as amplitudes das características existentes, é recomendável realizar algum tipo de pré-processamento na matriz de dados aumentada centralizada, objetivando aumentar o desempenho da metodologia proposta [43]. Em particular, o ACP é capaz de transformar as características inicialmente correlacionadas, onde estas se encontram na matriz de dados aumentada centralizada  $\Theta^C$  ( $\Theta^C \in \mathbb{R}^{n \times f}$ ) em um novo sistema descorrelacionado caracterizado pela matriz de pontuação  $\mathbf{S}$  ( $\mathbf{S} \in \mathbb{R}^{n \times f}$ ). Explana-se ainda que tal transformação é garantida pela centralização prévia da matriz de dados aumentada [92]. Adicionalmente, visando evitar o carregamento notacional, assume-se a supressão do superíndice existente na representação da matriz de dados aumentada centralizada. Desse modo, como mostra a Equação 3-42, a transformação salientada é designada pela matriz de carregamento  $\mathbf{L}$  ( $\mathbf{L} \in \mathbb{R}^{f \times f}$ ).

$$\mathbf{S} = \mathbf{\Theta}\mathbf{L} \quad (3-42)$$

Formalmente, as colunas da matriz de carregamento proveem as direções ortogonais do novo sistema descorrelacionado, onde estas, por sua vez, são dispostas em ordem decrescente de covariância. Geometricamente, a transformação  $\mathbf{L}$  ( $\mathbf{L} \in \mathbb{R}^{f \times f}$ ) projeta os dados aumentados, obtidos por meio da matriz de dados aumentada previamente centralizada, nas direções de máxima covariância. Dito isso, fundamentado na Equação 3-42, pode-se utilizar a Equação 3-43 com o intuito de transformar individualmente as características por intermédio das componentes principais  $l_i$  ( $i = 1, 2, \dots, f$ ).

$$s_i = \mathbf{\Theta}l_i \quad (i = 1, 2, \dots, f) \quad (3-43)$$

Realça-se ainda que, visando facilitar o entendimento do conteúdo abordado, a Figura 3.10 indica as direções de máxima covariância atreladas a um dado conjunto bidimensional de características e ilustra tanto a obtenção da característica  $s^*$  ( $s^* \in \mathbb{R}^2$ ) a partir da projeção da característica  $\theta^*$  ( $\theta^* \in \mathbb{R}^2$ ) na direção de máxima covariância  $l_1$  ( $l_1 \in \mathbb{R}^2$ ) quanto a localização no espaço de características vinculada ao centro dos dados aumentados, que, por sua vez, é designado pela variável  $\theta^\bullet$  ( $\theta^\bullet \in \mathbb{R}^2$ ).

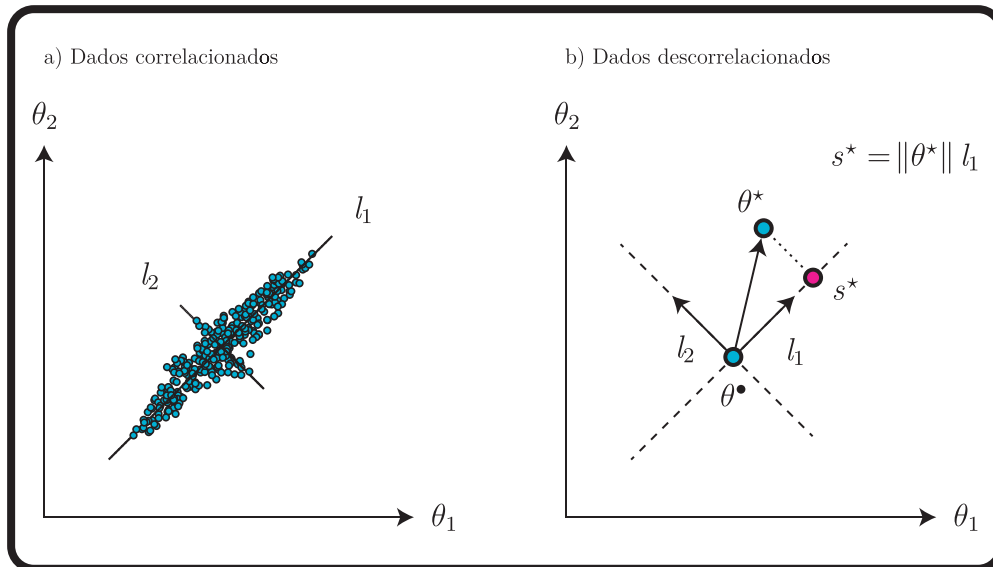


Figura 3.10: Dados correlacionados e descorrelacionados: a) Direções de máxima covariância para um dado problema detentor de somente duas características; b) Transformação capaz de descorrelacionar os dados aumentados nas direções de máxima covariância.

Não obstante, com a finalidade de dar sequência ao desenvolvimento do conteúdo teórico previamente iniciado, salienta-se que a matriz de carregamento  $\mathbf{L}$  ( $\mathbf{L} \in \mathbb{R}^{f \times f}$ ) é ortonormal e, por conseguinte, pode-se obter a Equação 3-44 por intermédio da Equação 3-42.

$$\mathbf{\Theta} = \mathbf{S}\mathbf{L}^T \quad (3-44)$$

Assim, a fim de obter tanto a matriz de pontuação quanto a matriz de carregamento, pode-se utilizar um método de decomposição matricial denominado Decomposição em Valores Singulares (DVS). Destaca-se ainda que o método supracitado é extremamente conveniente no que diz respeito à obtenção das referidas matrizes, pois o mesmo pode ser aplicado a matrizes retangulares [93]. Adicionalmente, a partir de uma única linha de código em linguagens *script* de alto nível, é tangível realizar a implementação do DVS com o objetivo de decompor a matriz de dados aumentada centralizada em um produto matricial contendo a matriz ortonormal  $\mathbf{U}$  ( $\mathbf{U} \in \mathbb{R}^{n \times f}$ ), a matriz diagonal  $\mathbf{\Sigma}$  ( $\mathbf{\Sigma} \in \mathbb{R}^{f \times f}$ ) e a matriz ortonormal  $\mathbf{V}$  ( $\mathbf{V} \in \mathbb{R}^{f \times f}$ ), como propõe a Equação 3-45.

$$\mathbf{\Theta} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \quad (3-45)$$

Salienta-se ainda que, com o propósito de relacionar as decomposições matriciais estabelecidas a partir das abordagens ACP e DVS, pode-se obter as Equações 3-46 e 3-47.

$$\mathbf{S} = \mathbf{U}\mathbf{\Sigma} \quad (3-46)$$

$$\mathbf{L} = \mathbf{V} \quad (3-47)$$

Além disso, a partir da decomposição matricial proposta pelo ACP, é possível encontrar as matrizes  $\mathcal{J}_\theta$  ( $\mathcal{J}_\theta \in \mathbb{R}^{f \times f}$ ) e  $\mathcal{O}_\theta$  ( $\mathcal{O}_\theta \in \mathbb{R}^{n \times n}$ ), onde estas são dadas pelas Equações 3-48 e 3-49, respectivamente.

$$\mathcal{J}_\theta = \mathbf{\Theta}^T \mathbf{\Theta} = \mathbf{L} \mathcal{J}_s \mathbf{L}^T \quad (3-48)$$

$$\mathcal{O}_\theta = \mathbf{\Theta} \mathbf{\Theta}^T = \mathbf{S} \mathbf{S}^T = \mathcal{O}_s \quad (3-49)$$

De maneira análoga, tais matrizes podem ser expressas por intermédio da decomposição DVS, como explicitam as Equações 3-50 e 3-51.

$$\mathcal{J}_\theta = \Theta^T \Theta = \mathbf{V} \Sigma^2 \mathbf{V}^T \quad (3-50)$$

$$\mathcal{O}_\theta = \Theta \Theta^T = \mathbf{U} \Sigma^2 \mathbf{U}^T \quad (3-51)$$

Ressalta-se que a obtenção das matrizes  $\mathcal{J}_\theta$  ( $\mathcal{J}_\theta \in \mathbb{R}^{f \times f}$ ) e  $\mathcal{O}_\theta$  ( $\mathcal{O}_\theta \in \mathbb{R}^{n \times n}$ ) são suportadas, respectivamente, por conceitos denominados produto interno matricial e produto externo matricial. Expõe-se ainda que tais conceitos são explicados de forma detalhada no Apêndice A deste material. Todavia, aspirando dar sequência à explanação do conteúdo remanescente, salienta-se que é possível acessar a matriz  $\mathcal{J}_s$  ( $\mathcal{J}_s \in \mathbb{R}^{f \times f}$ ), onde esta é gerada por meio do produto interno matricial associado à matriz de dados aumentada centralizada descrita a partir do sistema descorrelacionado. De maneira formal, pode-se obter a Equação 3-52 com o auxílio das Equações 3-47, 3-48 e 3-50.

$$\mathcal{J}_s = \Sigma^2 \quad (3-52)$$

Sendo assim, como propõe a Equação 3-53, ao dividir cada elemento da matriz supracitada por um certo escalar é tangível acessar a matriz de covariância amostral do sistema descorrelacionado  $\mathcal{C}_s$  ( $\mathcal{C}_s \in \mathbb{R}^{f \times f}$ ) [13].

$$\mathcal{C}_s = \frac{\mathcal{J}_s}{n-1} = \frac{\Sigma^2}{n-1} \quad (3-53)$$

Em paralelo, pode-se notar que a diagonalidade inerente à matriz de valores singulares  $\Sigma$  ( $\Sigma \in \mathbb{R}^{f \times f}$ ) acarreta na diagonalidade da matriz de covariância amostral do sistema descorrelacionado  $\mathcal{C}_s$  ( $\mathcal{C}_s \in \mathbb{R}^{f \times f}$ ). Categoricamente, diversas relações adicionais podem ser extraídas ao manipular as Equações 3-48, 3-49, 3-50 e 3-51, como exemplifica a Equação 3-54.

$$\mathcal{O}_s \mathbf{U} = \mathbf{U} \Sigma^2 \quad (3-54)$$

Em particular, após efetuar um conjunto de operações algébricas, é exequível obter à Equação 3-55 por intermédio da Equação 3-48, acarretando, por sua vez, na constatação de que tanto os termos presentes na diagonal principal da matriz de covariância amostral do sistema descorrelacionado quanto as colunas da matriz de carregamento representam, respectivamente, os autovaleores e os autovetores associados à matriz de covariância amostral do sistema correlacionado, como reitera a Equação 3-55.

$$\mathcal{C}_\theta \mathbf{L} = \mathbf{L} \mathcal{C}_s \quad (3-55)$$

A verificação da diagonalidade da matriz de covariância amostral do sistema descorrelacionado é esperada, pois os dados associados à matriz salientada pertencem a um sistema descorrelacionado. Em adição, destaca-se que é comum existir algum tipo de ruído associado a um dado valor singular [94]. De modo equivalente, pode-se verificar a presença de um certo ruído vinculado aos termos da diagonal principal da matriz de covariância amostral do sistema descorrelacionado  $c_i$  ( $i = 1, 2, \dots, f$ ). Dessa forma, visando acessar características mais relevantes, é possível investigar quais termos devem ser admitidos na reconstrução da matriz de dados aumentada centralizada. Precisamente, com o auxílio da Equação 3-56, é exequível computar os percentuais de recuperação  $P_i$  ( $i = 1, 2, \dots, f$ ) atrelados à reconstrução dos dados aumentados [95].

$$P_i = \frac{c_i}{\sum_{k=1}^f c_k} \quad (i = 1, 2, \dots, f) \quad (3-56)$$

Por construção, explicita-se que os termos presentes ao longo da diagonal principal da matriz de covariância amostral do sistema descorrelacionado são dispostos em ordem decrescente de magnitude. Adicionalmente, enfatiza-se que o percentual de recuperação cumulativo  $C_i$  ( $i = 1, 2, \dots, f$ ) pode ser calculado por meio da Equação 3-57.

$$C_i = \sum_{k=1}^i P_k \quad (i = 1, 2, \dots, f) \quad (3-57)$$

Salienta-se que a dificuldade em realizar a tarefa de identificação a partir de dados corrompidos pelo ruído pode ser amenizada ao admitir a exclusão de certos termos de recuperação na reconstrução da matriz de dados aumentada centralizada. Sendo assim, objetivando efetuar o procedimento de reconstrução da referida matriz, deve-se definir quais termos de recuperação serão descartados. Expõe-se também que o descarte de uma coleção de termos de recuperação está associado à utilização parcial tanto da matriz de pontuação quanto da matriz de carregamento. Dessa maneira, almejando facilitar a compreensão da abordagem proposta nesta tese, explica-se que a Figura 3.11 tem como finalidade elucidar um cenário onde são descartados apenas dois termos de recuperação dentre os sete disponíveis.

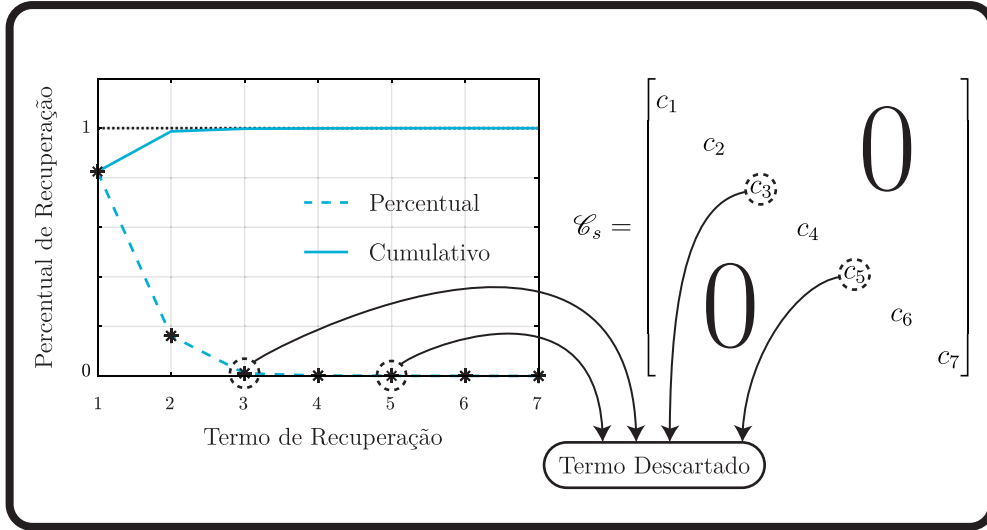


Figura 3.11: Descarte de dois termos de recuperação dentre os sete existentes.

De maneira precisa, ao levar em conta uma versão truncada da matriz de pontuação  $\tilde{\mathbf{S}}$  ( $\tilde{\mathbf{S}} \in \mathbb{R}^{n \times h}$ ), assim como uma versão truncada da matriz de carregamento  $\tilde{\mathbf{L}}$  ( $\tilde{\mathbf{L}} \in \mathbb{R}^{h \times h}$ ) é possível aproximar a matriz de dados aumentada centralizada  $\tilde{\mathbf{\Theta}}$  ( $\tilde{\mathbf{\Theta}} \in \mathbb{R}^{n \times h}$ ), como descreve a Equação 3-58.

$$\tilde{\mathbf{\Theta}} = \tilde{\mathbf{S}}\tilde{\mathbf{L}}^T \quad (3-58)$$

Mais especificamente, as versões truncadas das matrizes de pontuação e carregamento são produzidas a partir da supressão de certas colunas das referidas matrizes, onde as colunas excluídas correspondem aos termos de recuperação previamente assumidos como descartados. De maneira genérica, o número de termos não descartados  $h$  ( $h \in \mathbb{N}^*$ ) é menor do que o número de características existentes  $f$  ( $f \in \mathbb{N}^*$ ). Todavia, se o número de termos mantidos for numericamente igual ao número de características existentes, verifica-se a reconstrução completa da matriz de dados aumentada centralizada, estabelecendo, por sua vez, um cenário onde a técnica de redução de dimensionalidade intitulada ACP não se faz necessária. Desse modo, objetivando acessar dados mais representativos do sistema analisado, é possível explorar quais termos de recuperação devem ser adotados na reconstrução da matriz de dados aumentada centralizada. Por fim, ressalta-se que a Figura 3.12 tem o propósito de descrever como a reconstrução tanto completa quanto parcial da matriz de dados aumentada centralizada podem ser realizadas, onde a reconstrução parcial citada é baseada no descarte de todos os termos de recuperação existentes, excetuando o termo vinculado ao maior percentual de recuperação.



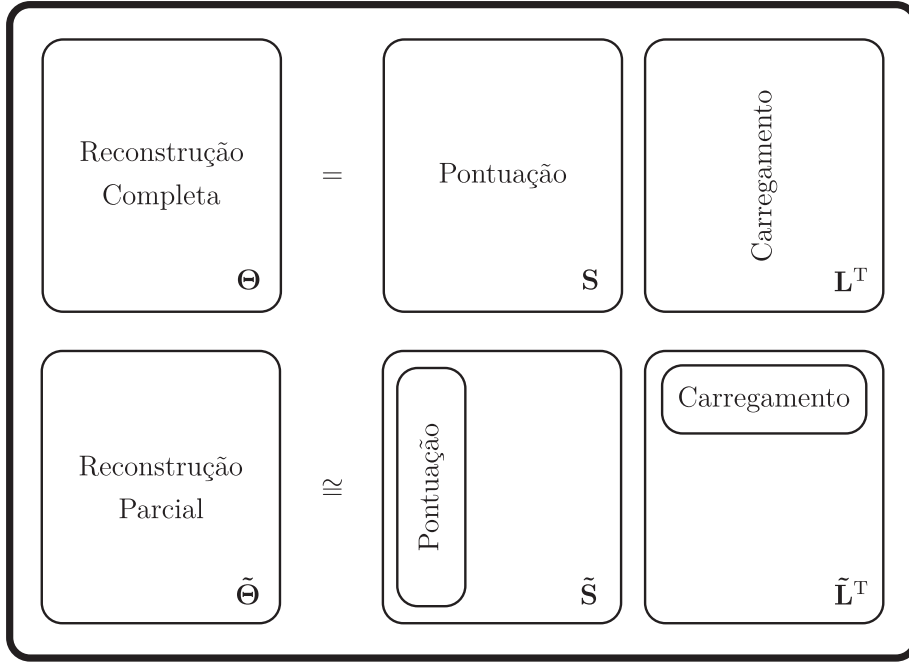


Figura 3.12: Descrição da reconstrução completa e parcial (primeiro termo de recuperação) da matriz de dados aumentada centralizada.

Note que a reconstrução parcial dos dados aumentados personificada pela Figura 3.12 é alicerçada no primeiro termo de recuperação. Além disso, a fim de realizar tal reconstrução, deve-se considerar tanto a primeira coluna da matriz de pontuação quanto a primeira coluna da matriz de carregamento. Em detalhes, a reconstrução parcial citada pode ser acessada a partir do produto entre uma matriz coluna e uma matriz linha, nesta ordem.

### 3.7 Pesquisa em Grade

Categoricamente, suportado pela literatura, pode-se verificar que apenas uma estratégia derivada da metodologia SINDy propõe algum tipo de busca hiperparamétrica. De forma mais precisa, a partir de uma busca unidimensional por um hiperparâmetro de esparsidade adequado, enfatiza-se que tal abordagem objetiva gerar descrições parcimoniosas de sistemas dinâmicos [65]. Todavia, ao assumir cenários ruidosos, a abordagem citada não é robusta o suficiente a fim de promover a identificação apropriada de tais sistemas. Sendo assim, a incorporação de uma técnica conhecida como pesquisa em grade é considerada [20]. Em detalhes, a partir de tal técnica pode-se efetuar uma busca bidimensional no espaço hiperparamétrico com a finalidade de encontrar tanto um hiperparâmetro de esparsidade adequado quanto um conjunto apropriado de termos de recuperação.

Desse modo, a busca bidimensional citada pode viabilizar a produção de modelos não espúrios mesmo sob condições mais severas de ruído. Mais formalmente, os conjuntos  $\Lambda$  ( $\Lambda \in \mathbb{R}^e$ ) e  $\Phi$  ( $\Phi \in \{0,1\}^f - \{0\}^f$ ) representam, respectivamente, um acervo contendo diversos hiperparâmetros de esparsidade e uma coleção possuindo as possíveis combinações dos termos de recuperação admitidos a fim de reconstruir a matriz de dados aumentada centralizada. Mais especificamente, a referida coleção é gerada por intermédio da quantidade máxima de termos de recuperação que são assumidos como mantidos  $h$  ( $h \in \mathbb{N}^*$ ). Pontua-se ainda que existem duas maneiras comumente utilizadas com a finalidade de gerar o conjunto de hiperparâmetros de esparsidade. De maneira precisa, pode-se definir o conjunto citado a partir de um espaçamento uniforme ou logarítmico, como mostra a Figura 3.13.

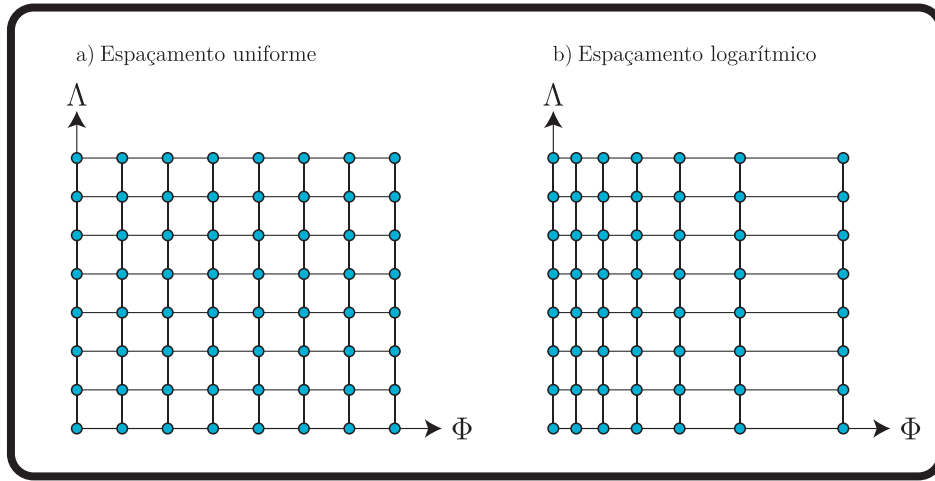


Figura 3.13: Tipos de espaçamentos: a) Uniforme; b) Logarítmico.

Observe também que, sob condições similares, a abordagem logarítmica é capaz de prover um conjunto concentrado de hiperparâmetros de esparsidade na vizinhança de zero, impulsionando, por sua vez, a produção de modelos menos influenciados pela magnitude do referido hiperparâmetro. Sendo assim, a assunção de um espaçamento logarítmico pode viabilizar a inferência de modelos mais representativos quando comparada com a adoção de um espaçamento do tipo uniforme. De forma genérica, independentemente do tipo de espaçamento admitido, deve-se definir a cardinalidade do conjunto de hiperparâmetros de esparsidade  $e$  ( $e \in \mathbb{N}^*$ ). Em particular, por meio de um hiperparâmetro de esparsidade mínimo  $\lambda_{\min}$  ( $\lambda_{\min} \in \mathbb{R}_+^*$ ) e de um hiperparâmetro de esparsidade máximo  $\lambda_{\max}$  ( $\lambda_{\max} \in \mathbb{R}_+^*$ ), é possível produzir um conjunto contendo uma coleção de hiperparâmetros de esparsidade igualmente espaçados, como propõe a Equação 3-59.

$$\lambda_i^E = \frac{i-1}{e-1} (\lambda_{\max} - \lambda_{\min}) + \lambda_{\min} \quad (i = 1, 2, \dots, e) \quad (3-59)$$

Por outro lado, almejando obter um conjunto logaritmicamente espaçado de hiperparâmetros de esparsidade, define-se o hiperparâmetro  $\alpha_{\min}$  ( $\alpha_{\min} \in \mathbb{R}$ ) e o hiperparâmetro  $\alpha_{\max}$  ( $\alpha_{\max} \in \mathbb{R}$ ). Daí, por meio da Equação 3-60, pode-se acessar um hiperparâmetro intermediário  $\alpha_i$  ( $i = 1, 2, \dots, e$ ).

$$\alpha_i = \frac{i-1}{e-1} (\alpha_{\max} - \alpha_{\min}) + \alpha_{\min} \quad (i = 1, 2, \dots, e) \quad (3-60)$$

Dessa forma, baseado no conceito de décadas, é possível produzir um conjunto logaritmicamente espaçado de hiperparâmetros de esparsidade, onde esse pode ser produzido a partir da Equação 3-61.

$$\lambda_i^L = 10^{\alpha_i} \quad (i = 1, 2, \dots, e) \quad (3-61)$$

Os conjuntos suportados tanto pela admissão de um espaçamento uniforme quanto pela assunção de um espaçamento logarítmico degeneram para um conjunto unitário se forem admitidos hiperparâmetros geradores iguais. Além disso, baseado nos princípios da análise combinatória, é exequível concluir que a cardinalidade do conjunto que contém as possíveis combinações dos termos de recuperação adotados com a finalidade de reconstruir a matriz de dados aumentada centralizada  $p$  ( $p \in \mathbb{N}^*$ ) depende do número de características disponíveis  $f$  ( $f \in \mathbb{N}^*$ ) e da quantidade de termos de recuperação não descartados  $h$  ( $h \in \mathbb{N}^*$ ), como descreve a Equação 3-62.

$$p = \sum_{k=0}^h \binom{f}{k} \quad (3-62)$$

Note que a quantidade de termos de recuperação descartados deve ser menor do que o número de características existentes, assim como a soma entre o número de termos de recuperação descartados e a quantidade de termos de recuperação mantidos deve ser numericamente igual ao número de características disponíveis. Adicionalmente, a Figura 3.14 almeja mostrar como a cardinalidade do conjunto que contém as possíveis combinações dos termos de recuperação adotados a fim de reconstruir a matriz de dados aumentada centralizada cresce à medida que a quantidade de características admitidas aumenta e o número de termos de recuperação descartados diminui.

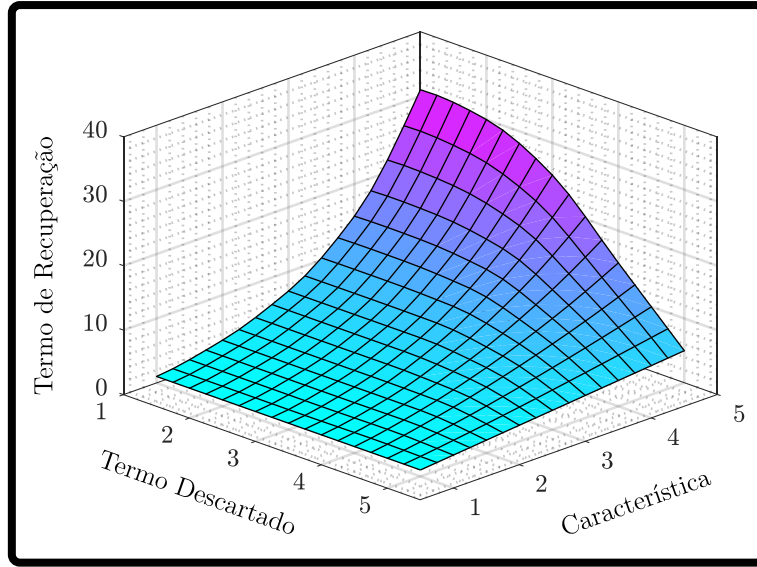


Figura 3.14: Representação gráfica da superfície interpolada capaz de descrever como a cardinalidade do conjunto que contém as possíveis combinações dos termos de recuperação adotados com o objetivo de reconstruir a matriz de dados aumentada centralizada varia de acordo com a quantidade de características existentes e com o número de termos de recuperação descartados.

Detalhadamente, ao admitir um problema detentor de quatro características, onde o número máximo de termos de recuperação mantidos é igual a dois, verifica-se que o conjunto que contém as possíveis combinações no que concerne aos termos de recuperação descartados e mantidos na reconstrução da matriz de dados aumentada centralizada é explicitado pela Equação 3-63.

$$\Phi = \left\{ (1, 1, 0, 0); (1, 0, 1, 0); (1, 0, 0, 1); (0, 1, 1, 0); (0, 1, 0, 1); (0, 0, 1, 1); (1, 0, 0, 0); (0, 1, 0, 0); (0, 0, 1, 0); (0, 0, 0, 1) \right\} \quad (3-63)$$

Explica-se ainda que os termos de recuperação descartados e mantidos no que diz respeito à reconstrução da matriz de dados aumentada centralizada são representados, respectivamente, pelos termos nulos e unitários presentes em cada elemento do conjunto dado pela Equação 3-63. Além disso, fundamentado na Figura 3.13, pode-se destacar que cada nó existente nas malhas previamente apresentadas está atrelado a um certo hiperparâmetro de esparsidade e a uma dada coleção de termos de recuperação que são mantidos a fim de reconstruir a matriz de dados aumentada centralizada, culminando na geração de um modelo candidato para cada nó das referidas malhas. Em seguida, por intermédio da adoção de métricas de avaliação, é tangível acessar o modelo mais fidedigno presente em um dado conjunto de modelos candidatos [43].

### 3.8

#### **Critério de Informação**

É notório que diversos princípios físicos, químicos e matemáticos são suportados pela simplicidade de como a natureza lida com alguns problemas impostos a ela. Particularmente, pode-se citar problemas inerentes a tópicos, tais como configuração eletrônica [96], estática [97], óptica [98] e termodinâmica [99]. Em adição, personificada pela navalha de Occam, é possível pontuar que a lei da parcimônia afirma que as explicações mais simples geralmente são as corretas [100]. Dessa maneira, sob uma ótica energética, é possível interpretar o modelo mais parcimonioso, dentre um conjunto de modelos candidatos, como o modelo menos energético capaz de explicar os dados vinculados a um dado sistema. Salienta-se ainda que a implementação da pesquisa em grade é capaz de viabilizar a geração de uma coleção de modelos candidatos, onde estes propõem certas relações entre as variáveis preditoras e as variáveis preditas do sistema analisado [77].

A fim de encontrar o modelo mais representativo presente em um conjunto de modelos candidatos, deve-se admitir uma métrica de avaliação. Paralelamente, destaca-se que a literatura disponibiliza um diversificado acervo de métricas, onde estas foram desenvolvidas com o propósito de atender finalidades específicas [43]. Em detalhes, expõe-se que as referidas finalidades são fundamentadas na capacidade das métricas citadas avaliarem problemas sob perspectivas, como classificação, regressão e seleção, para citar algumas. Precisamente, alicerçadas nos critérios de informação, enfatiza-se que as métricas de seleção são as mais indicadas em cenários onde a seleção de modelos parcimoniosos é requerida [65]. De maneira natural, é minimamente esperado que um modelo promissor de um dado sistema possua exatamente os mesmos termos não ativos que o mesmo [11].

É comum esperar que uma parte dos modelos candidatos apresentem um certo déficit paramétrico em relação ao sistema real, implicando na verificação de modelos que subajustam os dados. Em contraste, é provável que uma outra parcela dos referidos candidatos possuam demasiados parâmetros em comparação com o sistema real, acarretando na constatação de modelos que sobreajustam os dados. Adicionalmente, modelos que subajustam e sobreajustam os dados são modelos que apresentam, respectivamente, um alto viés e uma elevada variância. Salienta-se ainda que é possível verificar o nível de variância de um dado modelo a partir da submissão do mesmo a diferentes conjuntos de dados, por exemplo, os dados de teste. Sendo assim, quanto maior for a discrepância comportamental entre um modelo candidato e o sistema real, maior é a variância do modelo analisado [11].

Em especial, a verificação de um ajuste paramétrico adequado para um certo modelo candidato pode ser traduzida como a habilidade que o referido candidato possui em equilibrar de forma apropriada o subajuste e o sobreajuste. Dito isso, com o propósito de exemplificar tal ponderação, o sistema forçado unidimensional descrito a partir da Equação 3-64 é considerado, onde este é representado por meio de um polinômio real de quarta ordem.

$$\dot{x}_1(t) = -3 + 7x_1(t) + 2x_1^4(t) + u(t) \quad (3-64)$$

Dessa forma, para o exemplo descrito anteriormente, é esperado que a descrição matemática inerente a um certo candidato promissor possua exatamente os mesmos dois termos inativos que o sistema considerado. Além disso, no que concerne aos termos ativos, quanto menor for a discrepância entre os coeficientes do sistema real e os coeficientes de um dado modelo candidato, maior a capacidade do referido modelo explicar os dados. Destaca-se também que a principal vantagem das métricas de seleção pode ser descrita pela capacidade destas minimizarem os erros ao mesmo tempo que penalizam as complexidades desnecessárias existentes em um dado modelo, onde tais complexidades surgem a partir da adição inapropriada de parâmetros livres [101]. Desse modo, os critérios de informação fornecem uma abordagem que viabiliza a automatização da tarefa de seleção, promovendo, por sua vez, a identificação de modelos não espúrios [102, 103].

Historicamente, nos meados do século passado, foi proposto um método capaz de quantificar a perda de informação entre os valores reais do sistema e os valores estimados por um dado modelo [31]. Posteriormente, foi calculada a perda relativa de informação entre dois modelos, possibilitando a conexão entre a teoria apresentada pelo método previamente citado e a teoria estatística. Assim, a partir da maximização do logaritmo natural da função de verossimilhança o Critério de Informação de Akaike (CIA) foi desenvolvido [50, 104]. Matematicamente, ao assumir erros normalmente distribuídos com variância constante, pode-se obter a pontuação CIA a partir da Equação 3-65, onde esta depende de variáveis, como a quantidade de amostras existentes  $n$  ( $n \in \mathbb{N}^*$ ), o número de parâmetros livres presentes em um dado modelo candidato  $\rho$  ( $\rho \in \mathbb{N}^*$ ) e a soma quadrática dos resíduos entre o sistema real e o modelo admitido  $\Omega$  ( $\Omega \in \mathbb{R}$ ) [105].

$$\text{CIA} = n \ln \left( \frac{\Omega}{n} \right) + 2\rho \quad (3-65)$$

Salienta-se também que, caso o intercepto seja diferente de zero, é necessário levá-lo em conta no cálculo da quantidade de parâmetros livres existentes em um certo modelo candidato. Mais especificamente, o intercepto pode ser descrito como os termos constantes e diferentes de zero presentes em uma dada modelagem matemática. Além disso, por intermédio dos estados mensurados  $\mathbf{y}_i$  ( $i = 1, 2, \dots, s$ ) e de suas respectivas estimativas  $\hat{\mathbf{y}}_i$  ( $i = 1, 2, \dots, s$ ), pode-se acessar a soma quadrática a partir da Equação 3-66.

$$\Omega = \sum_{i=1}^s \sum_{k=1}^n [\mathbf{y}_i(k) - \hat{\mathbf{y}}_i(k)]^2 \quad (3-66)$$

Por outro lado, a penalização atrelada ao Critério de Informação Bayesiano (CIB) é obtida a partir do produto entre o dobro do número de parâmetros livres presentes na descrição matemática do modelo considerado  $\rho$  ( $\rho \in \mathbb{N}^*$ ) e o logaritmo natural da quantidade de amostras existentes  $n$  ( $n \in \mathbb{N}^*$ ). Dessa maneira, independentemente do critério de informação adotado, pode-se concluir que os modelos detentores de um maior sobreajuste estão associados a penalizações mais acentuadas. Em paralelo, a popularidade do critério bayesiano é justificada pelo fato da pontuação CIB ser formalmente consistente [32]. Detalhadamente, ao assumir erros normalmente distribuídos com variância constante, é possível utilizar a Equação 3-67 com o objetivo de calcular a pontuação CIB vinculada a cada modelo candidato.

$$\text{CIB} = n \ln \left( \frac{\Omega}{n} \right) + 2\rho \ln(n) \quad (3-67)$$

A partir do produto entre o número de hiperparâmetros de esparsidade admitidos  $e$  ( $e \in \mathbb{N}^*$ ) e a cardinalidade do conjunto que contém as possíveis combinações dos termos de recuperação adotados a fim de reconstruir a matriz de dados aumentada centralizada  $p$  ( $p \in \mathbb{N}^*$ ), pode-se acessar a quantidade total de modelos candidatos  $m$  ( $m \in \mathbb{N}^*$ ). Pontualmente, sob a perspectiva dos critérios de informação, objetiva-se encontrar o modelo menos penalizado capaz de minimizar a soma quadrática dos resíduos. Explana-se também que, com a finalidade de ranquear os modelos candidatos, é possível utilizar o conceito de pontuação relativa bayesiana, onde esta viabiliza a obtenção de um vetor que contém a pontuação bayesiana de cada modelo candidato  $\text{CIB}$  ( $\text{CIB} \in \mathbb{R}^m$ ). Formalmente, como mostra a Equação 3-68, pode-se acessar a pontuação relativa bayesiana  $\Delta\text{CIB}$  ( $\Delta\text{CIB} \in \mathbb{R}_+^m$ ) a partir de um vetor composto por termos iguais a menor pontuação bayesiana  $\text{CIB}_{\min}$  ( $\text{CIB}_{\min} \in \mathbb{R}^m$ ).

$$\Delta\text{CIB} = \text{CIB} - \text{CIB}_{\min} \quad (3-68)$$

Em adição, salienta-se a existência de diversas classificações que visam separar os modelos candidatos de acordo com a pontuação relativa dos mesmos. Especificamente, explana-se que os modelos candidatos são comumente classificados em categorias, como sem suporte, suporte fraco, suporte forte, pouco suporte, algum suporte e suporte substancial [65, 105]. Explana-se ainda que, além dos critérios de informação mencionados, foram desenvolvidos o Critério de Informação de Desvio [106] e a Descrição de Comprimento Mínimo [107]. Ressalta-se também que, fundamentado na consistência da pontuação bayesiana, admite-se nesta tese o emprego da pontuação relativa bayesiana [32]. Sendo assim, a fim de realizar a tarefa de seleção, busca-se o modelo candidato associado a uma pontuação relativa bayesiana mínima, onde esta pode ser traduzida como uma pontuação relativa nula, ou seja, procura-se o modelo candidato detentor do melhor suporte. Desse modo, suportada por um espaçamento do tipo logarítmico, a Figura 3.15 almeja ilustrar as pontuações relativas vinculadas a um certo conjunto de modelos candidatos, onde estes são gerados por intermédio da implementação de uma estratégia previamente apresentada como pesquisa em grade.

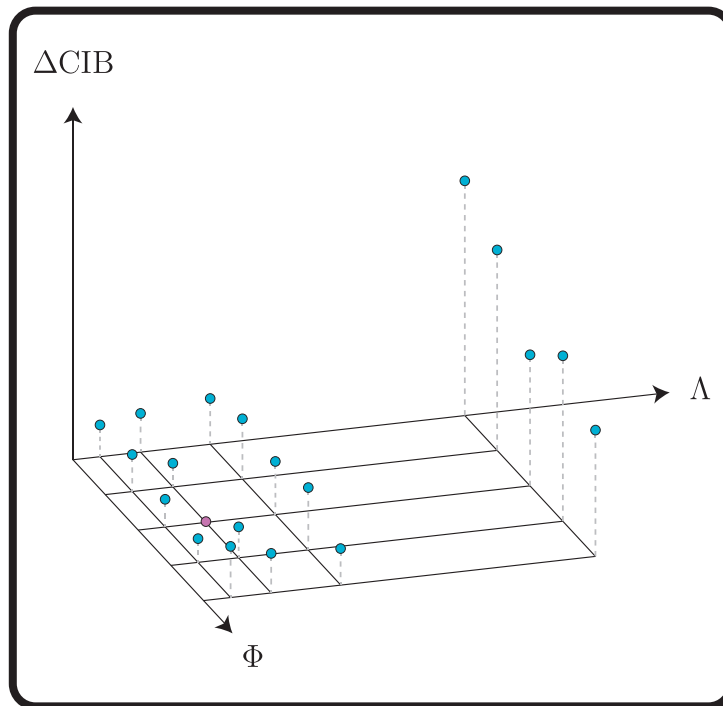


Figura 3.15: Seleção do modelo detentor do melhor suporte (rosa) dentre uma ampla coleção de modelos candidatos menos representativos (ciano).



### 3.9

#### Métricas de Avaliação

Naturalmente, uma vez que a presente tese visa demonstrar a capacidade da metodologia proposta superar a abordagem original no que diz respeito à inferência de sistemas dinâmicos a partir da admissão de dados ruidosos sob as perspectivas de classificação e regressão, é necessário adotar métricas de avaliação específicas. Em detalhes, sob a ótica da classificação, explicita-se que métricas, como Precisão (P), Sensibilidade (S) e Coeficiente de Similaridade Dice (F) são admitidas [48]. Ressalta-se também que, do ponto de vista classificatório, quanto mais próximo da unidade estiverem as três métricas de classificação previamente salientadas, mais similaridades existirão entre o modelo proposto e o sistema real. Observe que tais similaridades podem ser traduzidas como a compatibilidade entre o modelo considerado e o sistema real no que diz respeito à combinação de termos não lineares admitidos nas descrições matemáticas dos mesmos.

Mais precisamente, visando abordar a tarefa de identificação sob uma perspectiva classificatória, definem-se os termos nulos e não nulos presentes na descrição matemática de modelos e sistemas como termos inativos e ativos, respectivamente. A fim de calcular o desempenho dos modelos obtidos por intermédio das estruturas admitidas neste trabalho, é fundamental acessar variáveis, tais como o número de termos erroneamente classificados como inativos Falso Negativo (FN), a quantidade de termos corretamente rotulados como ativos Verdadeiro Positivo (VP), o número de termos equivocadamente categorizados como ativos Falso Positivo (FP) e a quantidade de termos assertivamente identificados como inativos Verdadeiro Negativo (VN) [108]. Assim, por meio das Equações 3-69 e 3-70, as duas primeiras métricas de classificação citadas anteriormente podem ser obtidas [87].

$$P = \frac{VP}{FP + VP} \quad (3-69)$$

$$S = \frac{VP}{FN + VP} \quad (3-70)$$

Além disso, enfatiza-se que a métrica de classificação previamente apresentada como coeficiente de similaridade foi desenvolvida com o propósito de combinar as métricas precisão e sensibilidade em um único escalar. Especificamente, como propõe a Equação 3-71, tal coeficiente é gerado a partir da média harmônica entre precisão e sensibilidade [43].

$$F = \frac{2}{\frac{1}{\bar{P}} + \frac{1}{\bar{S}}} \quad (3-71)$$

A média harmônica, diferentemente da média aritmética, impõe uma ponderação mais significativa quando valores mais baixos de precisão e sensibilidade são evidenciados. Dessa maneira, um dado modelo proposto possuirá um coeficiente de similaridade apreciável se forem verificados elevados valores de precisão e sensibilidade. Complementarmente, sob a ótica da regressão, a fidelidade dos modelos inferidos pode ser avaliada a partir de comparações entre o comportamento dinâmico de tais modelos e do sistema real. Dessa forma, com a finalidade de avaliar a qualidade dos modelos descobertos por uma dada metodologia, pode-se utilizar métricas, como Erro Absoluto Médio (EAM), Raiz do Erro Quadrático Médio (REQM) e Coeficiente de Determinação Múltipla (CDM) [20]. Salienta-se ainda que, quanto mais próximas de zero as métricas de regressão EAM e REQM estiverem, maior a similaridade dinâmica entre o modelo proposto e o sistema. Matematicamente, por intermédio dos estados do sistema  $\mathbf{y}_i$  ( $i = 1, 2, \dots, s$ ) e de suas estimativas  $\hat{\mathbf{y}}_i$  ( $i = 1, 2, \dots, s$ ), pode-se admitir as Equações 3-72 e 3-73 com a finalidade de computar as duas primeiras métricas de regressão previamente explicitadas.

$$\text{EAM} = \frac{1}{s} \sum_{i=1}^s \left[ \frac{1}{n} \sum_{k=1}^n \left| \mathbf{y}_i(k) - \hat{\mathbf{y}}_i(k) \right| \right] \quad (3-72)$$

$$\text{REQM} = \frac{1}{s} \sum_{i=1}^s \sqrt{\frac{1}{n} \sum_{k=1}^n \left[ \mathbf{y}_i(k) - \hat{\mathbf{y}}_i(k) \right]^2} \quad (3-73)$$

De maneira alternativa, a fim de acessar o grau de fidelidade entre o modelo inferido e o sistema propriamente dito, é possível considerar o coeficiente de determinação [49, 57]. Assim, almejando calcular tal coeficiente, é necessário obter um conjunto de variáveis intermediárias, onde estas são conhecidas como Soma Quadrática Residual (SQR) e Soma Quadrática Total (SQT). Formalmente, com o auxílio da Equação 3-74, pode-se acessar a primeira variável intermediária supracitada [109].

$$\text{SQR}_i = \sum_{k=1}^n \left[ \mathbf{y}_i(k) - \hat{\mathbf{y}}_i(k) \right]^2 \quad (i = 1, 2, \dots, s) \quad (3-74)$$

De modo complementar, suportado pela Equação 3-75, é possível obter a segunda variável intermediária previamente citada [109].

$$\text{SQT}_i = \sum_{k=1}^n [\mathbf{y}_i(k) - \bar{\mathbf{y}}_i]^2 \quad (i = 1, 2, \dots, s) \quad (3-75)$$

Expõe-se também que, a fim de utilizar a Equação 3-75, é necessário calcular o valor médio de cada estado do sistema  $\bar{\mathbf{y}}_i$  ( $i = 1, 2, \dots, s$ ). Por fim, como explicita a Equação 3-76, pode-se acessar a métrica de regressão denominada coeficiente de determinação [57].

$$\text{CDM} = \frac{1}{s} \sum_{i=1}^s \left( 1 - \frac{\text{SQR}_i}{\text{SQT}_i} \right) \quad (3-76)$$

Explica-se ainda que, quanto mais próximo da unidade o coeficiente de determinação estiver, maior a fidelidade do modelo admitido. Em geral, valores negativos do referido coeficiente podem ser traduzidos como uma baixa capacidade do modelo proposto representar o sistema. Rigorosamente, é preferível considerar um modelo ingênuo, onde este é baseado na média de cada estado do sistema do que assumir um modelo detentor de um coeficiente de determinação negativo. Finalmente, sob uma perspectiva de engenharia, destaca-se que o desempenho de um dado modelo é aceitável se for evidenciado um coeficiente de determinação maior ou igual a nove décimos [110].

### 3.10

#### Metodologia SINDyFE

Em linhas gerais, no que concerne às perspectivas de classificação e regressão, a metodologia intitulada *Feature Engineering to Deal with Noisy Data in Sparse Identification* (SINDyFE) avança a abordagem original denominada *Sparse Identification of Nonlinear Dynamics* (SINDy) no que diz respeito à inferência de sistemas dinâmicos a partir de dados corrompidos pelo ruído. Precisamente, a estratégia proposta consiste em uma hibridização de técnicas de aprendizado de máquina, tais como aumento de dados [25], regressão esparsa [28], seleção de características [29], extração de características [30], critério de informação [32], pesquisa em grade [33] e validação cruzada [87]. Pontua-se ainda que os dados associados aos estados do sistema a ser identificado são coletados por intermédio de sensores. Sendo assim, após a aquisição de tais dados, deve-se realizar a separação destes, produzindo, por sua vez, os conjuntos disjuntos de treino e teste [43].

Além disso, destaca-se que os dados de treino são utilizados na fase de treino do modelo. Em outros termos, o conjunto de treino é manipulado única e exclusivamente com o propósito de produzir um modelo representativo do sistema. Por outro lado, almejando evitar o viés de espionagem, recomenda-se que os dados de teste só sejam visualizados e manuseados na etapa de teste do modelo [105]. Adicionalmente, devido à causalidade temporal intrínseca aos dados admitidos no presente trabalho, realça-se que nenhuma estratégia baseada no embaralhamento de dados é incorporada ao algoritmo proposto [26]. Posteriormente, visando dar sequência aos procedimentos necessários com a finalidade de implementar a estrutura proposta nesta tese, deve-se construir a matriz de dados aumentada a partir da admissão de uma base de funções previamente definida [25].

Ressalta-se ainda que, com o auxílio da implementação de uma técnica conhecida como diferenças finitas, pode-se aproximar as derivadas temporais dos estados com erro de quarta ordem, evitando a perda de informação e descartando a necessidade de instalar mais sensores no sistema analisado [73]. Em sequência, deve-se realizar a centralização da matriz de dados aumentada com a finalidade de utilizar uma técnica de extração de características denominada Análise de Componentes Principais (ACP) [13]. Subsequentemente, admite-se uma coleção de hiperparâmetros de esparsidade e define-se o número máximo de termos de recuperação que podem ser mantidos com o objetivo de reconstruir a matriz de dados aumentada centralizada. Em detalhes, por intermédio de procedimentos, tais como regressão esparsa, pontuação relativa bayesiana, pesquisa em grade e validação cruzada é tangível automatizar o processo de seleção de modelos [20, 27, 32, 103].

Sendo assim, suportado pela descrição matemática de um dado modelo promissor, executa-se o procedimento de seleção de características, onde este tem como finalidade descartar as características menos representativas do sistema considerado [43]. Finalmente, após efetuar a seleção de características, submete-se novamente os dados a uma nova pesquisa em grade. Em particular, explana-se que tal pesquisa é alicerçada em uma versão simplificada do conjunto de características inicialmente admitidas, pois somente os dados aumentados que não foram excluídos no processo de seleção de características são mantidos na matriz de dados aumentada centralizada. Pontualmente, por meio do Algoritmo 3, é exequível obter tanto um hiperparâmetro de esparsidade adequado quanto o conjunto ótimo de termos de recuperação utilizados na reconstrução da matriz de dados aumentada centralizada, ou seja, é possível selecionar um modelo promissor do sistema analisado a partir de uma coleção de modelos candidatos [105, 111].

**Algoritmo 3** Pesquisa em Grade

**Entrada:** Estados do sistema  $\mathbf{X}$  ( $\mathbf{X} \in \mathbb{R}^{n \times s}$ ), matriz de dados aumentada centralizada  $\mathbf{\Theta}$  ( $\mathbf{\Theta} \in \mathbb{R}^{n \times f}$ ), palpite inicial da solução  $\mathbf{\Xi}^\bullet$  ( $\mathbf{\Xi}^\bullet \in \mathbb{R}^{f \times s}$ ), taxa de aprendizado  $\mu$  ( $\mu \in \mathbb{R}_+^*$ ), percentual de separação dos dados  $\varphi$  ( $\varphi \in (0, 1) \subset \mathbb{R}$ ), número máximo de termos de recuperação que podem ser descartados  $\delta$  ( $\delta \in \mathbb{N}$ ), conjunto de hiperparâmetros de esparsidade  $\Lambda$  ( $\Lambda \in \mathbb{R}^e$ ), número máximo de iterações  $I$  ( $I \in \mathbb{N}^*$ ) e tolerância mínima admitida  $T$  ( $T \in \mathbb{R}_+^*$ ).

**Saída:** Busca pelo modelo com o melhor suporte.

---

```

1: procedure GS( $\mathbf{X}, \mathbf{\Theta}, \mathbf{\Xi}^\bullet, \delta, \mu, \varphi, \Lambda, I, T$ )
2:    $\mathbf{X}_\tau, \mathbf{\Theta}_\tau, \mathbf{X}_\nu, \mathbf{\Theta}_\nu \leftarrow \text{split\_data}(\mathbf{X}, \mathbf{\Theta}, \varphi)$  ▷ Separar Dados
3:    $\dot{\mathbf{X}}_\tau, \dot{\mathbf{X}}_\nu \leftarrow \text{approx\_derivatives}(\mathbf{X}_\tau, \mathbf{X}_\nu)$  ▷ Aproximar Derivadas
4:    $\Phi \leftarrow \text{recovery\_set}(\mathbf{\Theta}_\tau, \delta)$  ▷ Conjunto de Recuperação
5:    $\text{ctr} \leftarrow 1$  ▷ Inicializar Contador
6:   for  $i \in \{1, 2, \dots, \text{length}(\Lambda)\}$  do ▷ Pesquisar Sobre  $\Lambda$ 
7:     for  $j \in \{1, 2, \dots, \text{length}(\Phi)\}$  do ▷ Pesquisar Sobre  $\Phi$ 
8:        $\dot{\mathbf{X}}, \lambda \leftarrow \dot{\mathbf{X}}_\tau, \Lambda^i$  ▷ Substituir Variáveis
9:        $\mathbf{\Theta} \leftarrow \text{feature\_extraction}(\mathbf{\Theta}_\tau, \Phi^j)$  ▷ Extrair Características
10:       $\mathbf{\Xi}_\tau^{ij} \leftarrow \text{LASSO}(\dot{\mathbf{X}}, \mathbf{\Theta}, \mathbf{\Xi}^\bullet, \lambda, \mu, I, T)$  ▷ Regressão OSEMA
11:       $\mathbf{\Xi}_\tau^{\text{ctr}} \leftarrow \mathbf{\Xi}_\tau^{ij}$  ▷ Rearranjar Modelo
12:       $\text{ctr} \leftarrow \text{ctr} + 1$  ▷ Atualizar Contador
13:    end for
14:  end for
15:   $m \leftarrow \text{length}(\Lambda) \times \text{length}(\Phi)$  ▷ Número de Candidatos
16:  for  $i \in \{1, 2, \dots, m\}$  do ▷ Cada Candidato
17:     $\hat{\mathbf{X}}_\nu \leftarrow \text{test\_model}(\mathbf{\Xi}_\tau^i, \mathbf{\Theta}_\nu)$  ▷ Testar Modelo
18:     $\text{BIC}^i \leftarrow \text{bic}(\mathbf{X}_\nu, \hat{\mathbf{X}}_\nu, \mathbf{\Xi}_\tau^i)$  ▷ Calcular CIB
19:  end for
20:   $\Delta\text{BIC} \leftarrow \text{BIC} - \min(\text{BIC})$  ▷ Calcular  $\Delta\text{CIB}$ 
21:   $\text{idx} \leftarrow \text{find}(\Delta\text{BIC} = \min(\Delta\text{BIC}))$  ▷ Menor Índice  $\Delta\text{CIB}$ 
22:   $\mathbf{\Xi}_\tau^* \leftarrow \mathbf{\Xi}_\tau^{\text{idx}}$  ▷ Modelo Promissor
23:  return  $\mathbf{\Xi}_\tau^*$  ▷ Melhor Suporte
24: end procedure

```

---

Detalhadamente, uma vez que deseja-se traçar um comparativo entre os desempenhos da estratégia proposta e da estrutura original sob as perspectivas de classificação e regressão, é fundamental que métricas de classificação e regressão sejam adotadas [48, 112]. Expõe-se ainda que o Algoritmo 4 visa prover uma descrição completa da metodologia proposta.

**Algoritmo 4** SINDyFE

**Entrada:** Estados do sistema  $\mathbf{X}$  ( $\mathbf{X} \in \mathbb{R}^{n \times s}$ ), matriz de dados aumentada centralizada  $\mathbf{\Theta}$  ( $\mathbf{\Theta} \in \mathbb{R}^{n \times f}$ ), palpite inicial da solução  $\mathbf{\Xi}^\bullet$  ( $\mathbf{\Xi}^\bullet \in \mathbb{R}^{f \times s}$ ), taxa de aprendizado  $\mu$  ( $\mu \in \mathbb{R}_+^*$ ), percentual de separação dos dados  $\varphi$  ( $\varphi \in (0, 1) \subset \mathbb{R}$ ), número máximo de termos de recuperação que podem ser descartados  $\delta$  ( $\delta \in \mathbb{N}$ ), conjunto de hiperparâmetros de esparsidade  $\Lambda$  ( $\Lambda \in \mathbb{R}^e$ ), número máximo de iterações  $I$  ( $I \in \mathbb{N}^*$ ) e tolerância mínima admitida  $T$  ( $T \in \mathbb{R}_+^*$ ).

**Saída:** Modelo parcimonioso.

---

```

1: procedure SINDyFE( $\mathbf{X}, \mathbf{\Theta}, \mathbf{\Xi}^\bullet, \delta, \mu, \varphi, \Lambda, I, T$ )
2:    $\mathbf{X}_\tau, \mathbf{\Theta}_\tau, \mathbf{X}_\nu, \mathbf{\Theta}_\nu \leftarrow \text{split\_data}(\mathbf{X}, \mathbf{\Theta}, \varphi)$            ▷ Separar Dados
3:    $\mathbf{\Xi}_\tau^* \leftarrow \text{GS}(\mathbf{X}, \mathbf{\Theta}, \mathbf{\Xi}^\bullet, \delta, \mu, \varphi, \Lambda, I, T)$            ▷ Pesquisa em Grade
4:    $\mathbf{\Theta} \leftarrow \text{feature\_selection}(\mathbf{\Xi}_\tau^*)$                              ▷ Selecionar Características
5:    $\mathbf{\Xi}_\nu^* \leftarrow \text{GS}(\mathbf{X}, \mathbf{\Theta}, \mathbf{\Xi}^\bullet, \delta, \mu, \varphi, \Lambda, I, T)$            ▷ Pesquisa em Grade
6:    $\hat{\mathbf{X}}_\nu^* \leftarrow \text{test\_model}(\mathbf{\Xi}_\tau^*, \mathbf{\Theta}_\nu)$                          ▷ Testar Modelo
7:    $\text{scores} \leftarrow \text{evaluation\_metrics}(\mathbf{X}_\nu, \hat{\mathbf{X}}_\nu^*)$                  ▷ Avaliar Modelo
8:   return  $\mathbf{\Xi}_\tau^*$                                                          ▷ Modelo Parcimonioso
9: end procedure

```

---

Dessa maneira, por intermédio da metodologia proposta, é possível avançar a abordagem original no que diz respeito à produção de modelos não espúrios de sistemas dinâmicos. Observe também que a adoção de técnicas que visam selecionar e extrair características estabelecem um cenário onde uma abordagem via engenharia de recursos é verificada [87]. Explica-se ainda que, como descreve o Algoritmo 4, o último estágio da tarefa de identificação como um todo disserta sobre a avaliação do modelo inferido. Desse modo, almejando acessar os desempenhos dos modelos produzidos tanto pela estrutura original quanto pela estratégia proposta, deve-se submeter tais modelos aos dados de teste. Notavelmente, caso seja possível adquirir as derivadas dos estados por meio de medições diretas, recomenda-se a utilização das mesmas, pois a adoção de técnicas de aproximação capazes de obter tais derivadas pode ser responsável por um aumento expressivo nos erros destas, culminando, por sua vez, na obtenção de modelos não generalizáveis. Salienta-se ainda que esta é a primeira vez que a hibridização de técnicas de aprendizado de máquina, tais como aumento de dados, regressão esparsa, seleção de características, extração de características, critério de informação, pesquisa em grade e validação cruzada são implementadas com a finalidade de avançar a abordagem original no que concerne à tarefa de identificação de sistemas dinâmicos a partir de dados ruidosos. Por fim, destaca-se que a Figura 3.16 tem o propósito de ilustrar a sequência de passos requeridos no que concerne à implementação da metodologia proposta nesta tese.

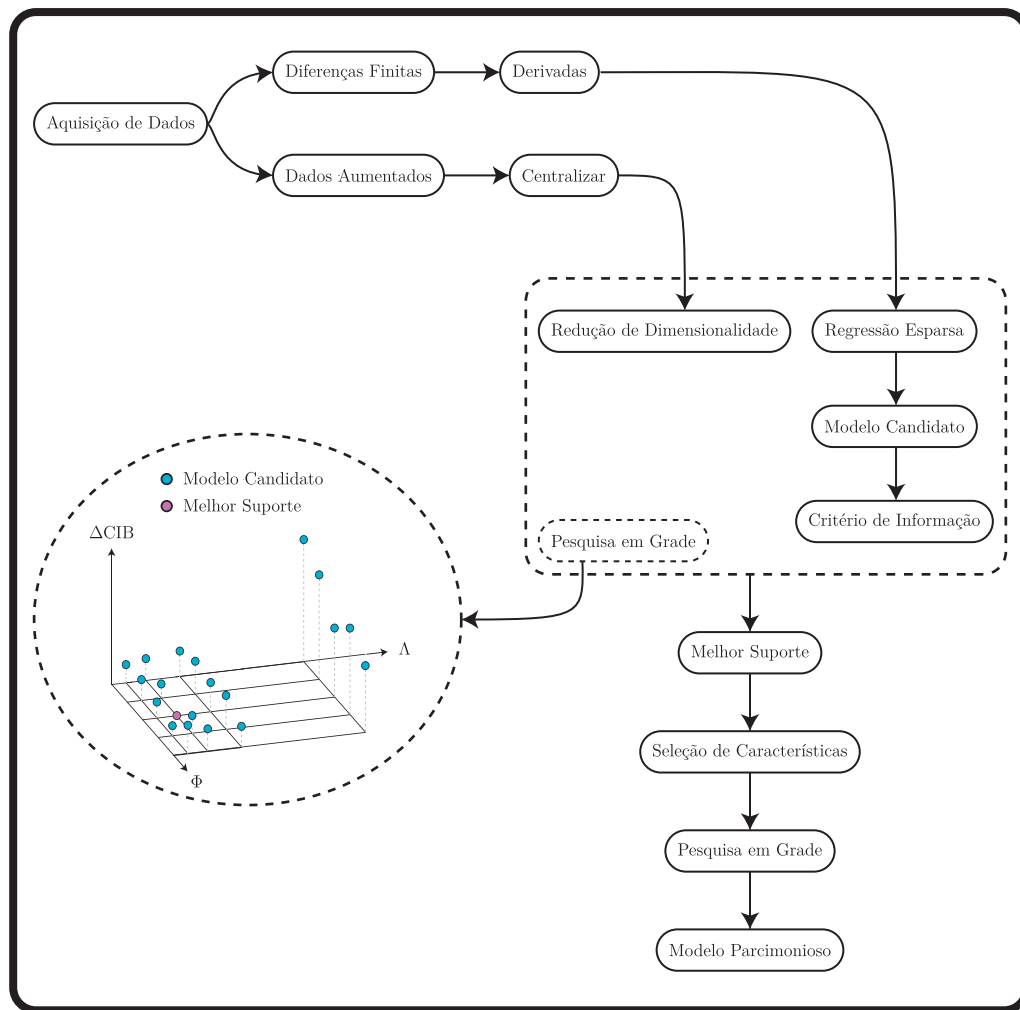


Figura 3.16: Esquema detalhado dos passos realizados pela estratégia proposta.

## 4

### Estudos de Caso

No presente capítulo, almejando comparar os desempenhos da estratégia pioneira e da metodologia proposta sob as perspectivas de classificação e regressão, cinco estudos de caso são apresentados. Mais especificamente, com o propósito de traçar tais comparativos, admite-se estudos de caso, tais como crescimento logístico [25, 35], oscilador Duffing [36, 37], modelo de FitzHugh–Nagumo [38, 39], atrator de Lorenz [25, 40] e uma modelagem Suscetível-Infeccioso-Recuperado [42, 113].

#### 4.1

##### Crescimento Logístico

É possível verificar na literatura a existência de sistemas capazes de mudar a dinâmica vinculada aos mesmos. Além disso, ressalta-se que certos sistemas apresentam um fenômeno conhecido como bifurcação [25]. Em particular, destaca-se um sistema dinâmico detentor de um hiperparâmetro de bifurcação, onde este é conhecido como crescimento logístico [35]. Dessa forma, por intermédio da Figura 4.1, é possível acessar a relação existente entre o hiperparâmetro de bifurcação supracitado e os estados estacionários associados à dinâmica do sistema considerado.

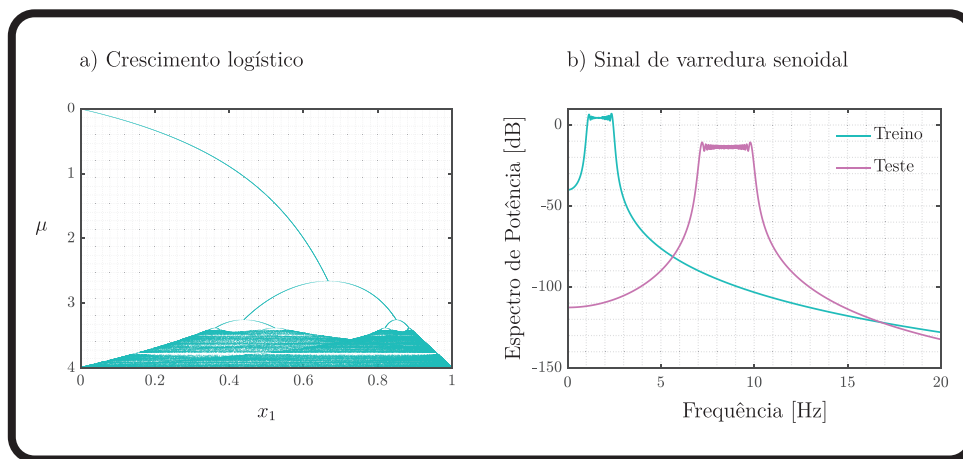


Figura 4.1: Primeiro estudo de caso: a) Versão discreta do crescimento logístico com forçamento estocástico; b) Excitações de treino (ciano) e teste (rosa) impostas ao crescimento logístico.



Adicionalmente, pode-se observar por meio da Figura 4.1 que um aumento no hiperparâmetro de bifurcação acarreta na elevação do número de estados estacionários atrelados ao sistema apresentado [35]. Salienta-se ainda que a versão discreta do crescimento logístico é amplamente conhecida por representar um modelo clássico de dinâmica populacional. Paralelamente, como explica a Equação 4-1, pode-se apresentar uma versão contínua do crescimento logístico forçado por meio da equação de Richard [114].

$$\dot{x}_1(t) = x_1(t) - x_1^2(t) + u(t) \quad (4-1)$$

Realça-se também que a aplicação do forçamento  $u$  no referido sistema pode causar tanto a diminuição quanto o aumento do tamanho da população  $x_1$ . Desse modo, como apresenta a Equação 4-2, admite-se uma excitação do tipo varredura senoidal a fim de realizar a tarefa de identificação [12].

$$u(t) = A \sin \left[ (\varphi_1 t + \varphi_2) t + \vartheta \right] \quad (4-2)$$

Observe que a referida excitação pode ser descrita a partir da amplitude  $A$  ( $A \in \mathbb{R}$ ), do ângulo de fase  $\vartheta$  ( $\vartheta \in \mathbb{R}$ ), do parâmetro  $\varphi_1$  ( $\varphi_1 \in \mathbb{R}$ ) e do parâmetro  $\varphi_2$  ( $\varphi_2 \in \mathbb{R}$ ), onde os dois últimos parâmetros apresentados dependem das frequências de excitação mínima e máxima aplicadas ao sistema. Dessa maneira, a Figura 4.1 ilustra os forçamentos de treino e teste impostos a tal sistema. Explica-se ainda que uma dada condição inicial  $x^\bullet$  ( $x^\bullet = [0.5]^\top$ ) é admitida com o propósito de gerar o conjunto de dados na íntegra.

## 4.2

### Oscilador Duffing

Desde o início do século passado, pôde-se observar tanto a descrição quanto a exploração de um sistema conhecido como oscilador Duffing [36]. Destaca-se ainda que a exploração citada é naturalmente justificada pela vasta riqueza dinâmica apresentada pelo referido sistema. Em detalhes, tal oscilador disserta sobre um problema de vibração forçada, onde este é descrito a partir de uma massa suspensa por meio de um amortecedor e de uma mola de rigidez cúbica. Ressalta-se também que o amortecedor e a mola presentes em tal sistema são dispostos paralelamente [115]. Matematicamente, suportado pela técnica de suspensão de variáveis, pode-se obter as Equações 4-3 e 4-4, onde estas caracterizam um oscilador detentor de uma massa unitária e não linearidade cúbica [37].

$$\dot{x}_1(t) = x_2(t) \quad (4-3)$$

$$\dot{x}_2(t) = -x_1(t) - 0,05x_2(t) - 0,02x_1^3(t) + u(t) \quad (4-4)$$

Expõe-se também que a descrição matemática apresentada anteriormente é dada a partir do sinal de entrada  $u$ , da coordenada medida em relação ao solo  $x_1$  e de sua derivada  $x_2$ . Detalhadamente, a coordenada medida em relação ao solo previamente destacada pode ser descrita por meio da subtração entre o deslocamento da massa e o deslocamento da excitação de base. Dessa maneira, a Figura 4.2 objetiva descrever o comportamento dinâmico apresentado pelo oscilador Duffing forçado.

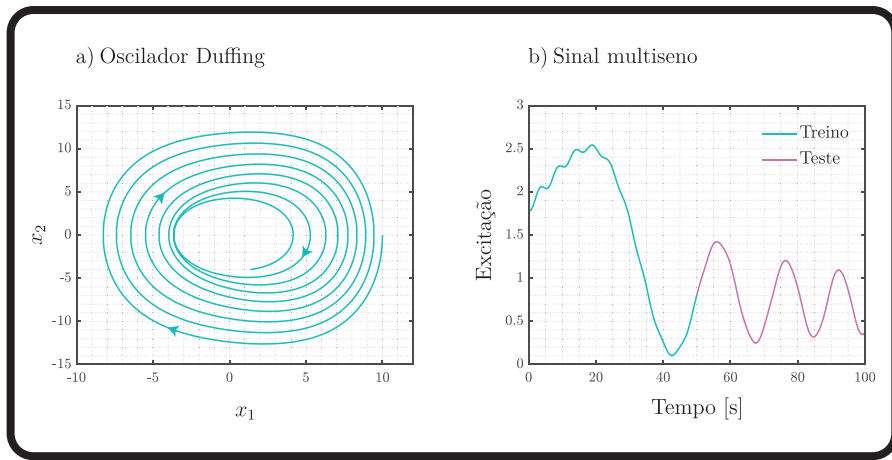


Figura 4.2: Segundo estudo de caso: a) Retrato de fase do oscilador Duffing ao admitir uma excitação do tipo multiseno; b) Excitações de treino (ciano) e teste (rosa) aplicadas ao oscilador Duffing.

Destaca-se também que, visando identificar corretamente o referido sistema, um forçamento conhecido como multiseno é aplicado a tal oscilador, como explicita a Figura 4.2. Especificamente, o forçamento previamente salientado é gerado a partir de uma coleção de amplitudes  $A_k$  ( $A_k \in \mathbb{R}$ ), frequências  $\varphi_k$  ( $\varphi_k \in \mathbb{R}$ ) e ângulos de fase  $\vartheta_k$  ( $\vartheta_k \in \mathbb{R}$ ), como explana a Equação 4-5.

$$u(t) = \sum_{k=1}^{\varsigma} A_k \cos(2\pi\varphi_k t + \vartheta_k) \quad (4-5)$$

Enfatiza-se ainda que, a partir de uma determinada quantidade de senoides superpostas  $\varsigma$  ( $\varsigma \in \mathbb{N}^*$ ), pode-se construir o sinal de excitação considerado [12]. Por fim, uma determinada condição inicial  $x^\bullet$  ( $x^\bullet = [10 \ 0]^T$ ) é assumida com o objetivo de produzir o conjunto de dados como um todo.

## 4.3

**FitzHugh–Nagumo**

O modelo de FitzHugh-Nagumo propõe uma representação bidimensional da atividade neural, onde a partir de tal representação é possível descrever os padrões inerentes tanto ao disparo neural quanto à propagação de sinais em neurônios [38]. Além disso, explana-se que a referida representação bidimensional admite, por hipótese, que um neurônio pode ser descrito por meio de um par de osciladores. Em outras palavras, a dinâmica neural proposta pela representação salientada pode ser interpretada como um processo oscilatório [39]. Formalmente, por intermédio do potencial da membrana  $x_1$ , de uma variável de recuperação  $x_2$  e da magnitude do estímulo imposto ao neurônio  $u$ , é tangível descrever matematicamente o modelo de FitzHugh-Nagumo, como reiteram as Equações 4-6 e 4-7 [114].

$$\dot{x}_1(t) = 0,1 + x_1(t) - x_2(t) - 0,33x_1^3(t) + u(t) \quad (4-6)$$

$$\dot{x}_2(t) = 0,1[x_1(t) - x_2(t)] \quad (4-7)$$

Pontua-se também que a riqueza dinâmica intrínseca ao modelo previamente apresentado pode ser evidenciada pela ampla coleção de fenômenos que o referido modelo é capaz de explicar. Detalhadamente, é possível destacar fenômenos, como acomodação de pico, bloco de excitação e excitação de ruptura anodal, para citar algumas. Explica-se ainda que, ao assumir um sinal de entrada do tipo multisenso, pode-se verificar a ocorrência do primeiro fenômeno citado, como ilustra a Figura 4.3.

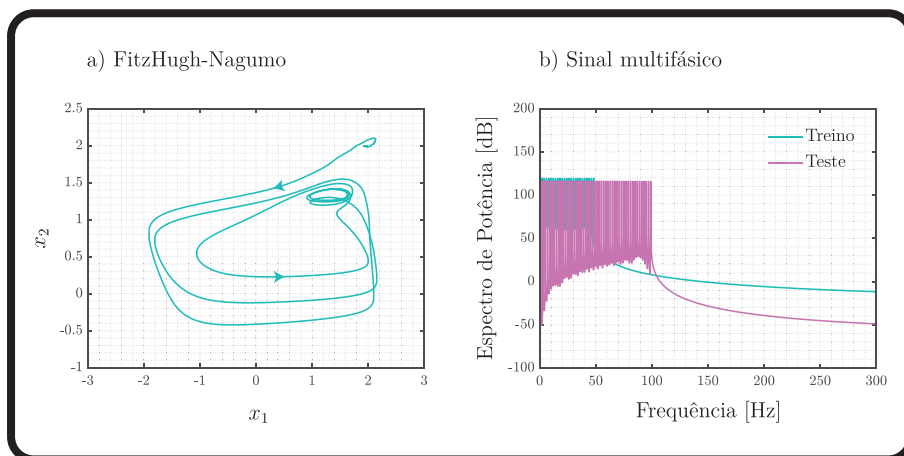


Figura 4.3: Terceiro estudo de caso: a) Retrato de fase obtido a partir da admissão de um sinal de entrada do tipo multisenso; b) Excitações de treino (ciano) e teste (rosa) aplicadas ao modelo de FitzHugh–Nagumo.

Salienta-se ainda que, visando obter um modelo parcimonioso no que diz respeito à tarefa de identificação, um sinal de entrada multifásico é admitido para o terceiro estudo de caso abordado nesta tese. Assim, por intermédio de parâmetros como amplitude  $A$  ( $A \in \mathbb{R}$ ), frequência de base  $\varphi_0$  ( $\varphi_0 \in \mathbb{R}$ ) e um conjunto de ângulos de fase  $\vartheta_k$  ( $\vartheta_k \in \mathbb{R}$ ), é exequível obter o sinal de entrada supracitado, como exprime a Equação 4-8.

$$u(t) = \sum_{k=1}^{\varsigma} A \cos(2\pi k \varphi_0 t + \vartheta_k) \quad (4-8)$$

Realça-se ainda que, a partir da superposição de uma determinada quantidade de senoides  $\varsigma$  ( $\varsigma \in \mathbb{N}^*$ ), pode-se produzir tal forçamento [116]. Por inspeção, é possível verificar que a excitação fornecida pela Equação 4-5 difere do forçamento dado pela Equação 4-8, basicamente, pelo fato do primeiro possuir amplitudes variáveis. Em paralelo, pontua-se que a Figura 4.3 descreve as excitações de treino e teste aplicadas ao modelo de FitzHugh–Nagumo. Finalmente, com a finalidade de gerar o conjunto de dados inerente ao presente estudo de caso, uma dada condição inicial  $x^\bullet$  ( $x^\bullet = [1 \ 2]^T$ ) é assumida.

#### 4.4

##### Atrator de Lorenz

O atrator de Lorenz é o sistema mais famoso no que diz respeito à capacidade de produzir soluções não periódicas, ou seja, devido ao fato das soluções inerentes a tal atrator não repetirem exatamente seus históricos à medida que o tempo transcorre. Adicionalmente, pode-se destacar que o referido atrator fornece uma descrição matemática de um fenômeno físico conhecido como convecção atmosférica [25]. De maneira detalhada, expõe-se que a variável  $x_1$  presente na descrição matemática de tal sistema é proporcional à intensidade do movimento de convecção e a variável  $x_2$  é proporcional à diferença de temperatura entre as correntes ascendentes e descendentes. Por fim, explica-se que a variável  $x_3$  é proporcional à distorção do perfil de temperatura vertical em relação à linearidade observada [40]. Desse modo, por meio das Equações 4-9, 4-10 e 4-11, é exequível acessar a descrição matemática do atrator previamente apresentado [25, 40].

$$\dot{x}_1(t) = -10x_1(t) + 10x_2(t) + u(t) \quad (4-9)$$

$$\dot{x}_2(t) = 28x_1(t) - x_2(t) - x_1(t)x_3(t) \quad (4-10)$$

$$\dot{x}_3(t) = -2,66x_3(t) + x_1(t)x_2(t) \quad (4-11)$$

Enfatiza-se também que a variável  $u$  presente no conjunto de equações fornecido anteriormente caracteriza o forçamento imposto ao sistema de Lorenz e que as variáveis de estado inerentes ao sistema citado representam a localização espacial de uma dada partícula. Além disso, como propõe a Figura 4.4, pode-se verificar o caos intrínseco ao atrator de Lorenz a partir das evoluções históricas de uma dada coleção de partículas.

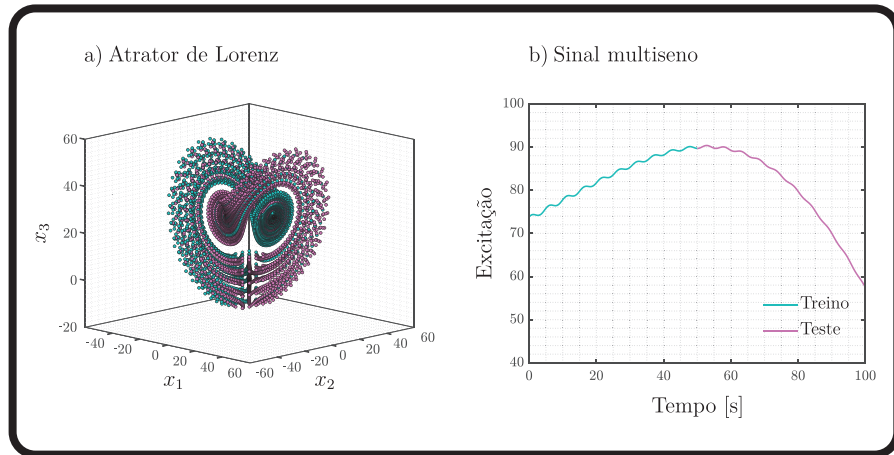


Figura 4.4: Quarto estudo de caso: a) Comportamento caótico apresentado pela versão não forçada do atrator de Lorenz; b) Excitações de treino (ciano) e teste (rosa) impostas ao sistema de Lorenz.

Em adição, almejando obter um modelo generalizável do sistema de Lorenz, pode-se assumir um sinal de entrada do tipo multisenso, como explicita a Figura 4.4. Finalmente, destaca-se que uma determinada condição inicial  $x^\bullet$  ( $x^\bullet = [-8 \ 8 \ 27]^T$ ) é admitida com o propósito de produzir o conjunto de dados em sua totalidade.

## 4.5

### Suscetível-Infecioso-Recuperado

Motivado pela natureza catastrófica de diversos eventos epidemiológicos, pesquisadores do mundo inteiro uniram esforços tentando explicar o comportamento dinâmico atrelado a tais eventos. Nas últimas décadas, pôde-se constatar o surgimento de diversas epidemias, como peste negra, varíola, gripe espanhola e febre amarela, para citar algumas. Atualmente, a humanidade enfrenta um novo cenário pandêmico evidenciado pela expressiva disseminação do *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2) [42, 113]. Formalmente, explica-se que as Equações 4-12, 4-13 e 4-14 tem como finalidade fornecer uma modelagem matemática capaz de descrever a evolução dinâmica inerente à disseminação do SARS-CoV-2 em uma população senegalesa contendo dez mil indivíduos.

$$\dot{x}_1(t) = -0,25x_1(t)x_2(t) + u(t) \quad (4-12)$$

$$\dot{x}_2(t) = 0,25x_1(t)x_2(t) - 0,03x_2(t) \quad (4-13)$$

$$\dot{x}_3(t) = 0,03x_2(t) \quad (4-14)$$

Ressalta-se também que a descrição matemática previamente apresentada é dada por meio de uma modelagem amplamente conhecida como Suscetível-Infeccioso-Recuperado (SIR), onde esta visa caracterizar a evolução dinâmica no que concerne ao cenário pandêmico salientado. Mais precisamente, a modelagem SIR assume que os indivíduos pertencentes a uma dada população podem ser separados em três grupos. De maneira detalhada, tais grupos podem ser caracterizados pelo percentual da população suscetível à doença  $x_1$ , pela taxa populacional de indivíduos doentes  $x_2$  e pela porcentagem de indivíduos recuperados da infecção  $x_3$  [117]. Assim, como retrata a Figura 4.5, no estágio zero da dinâmica epidemiológica considerada existem quatro indivíduos expostos ao vírus e nenhum indivíduo recuperado [41].

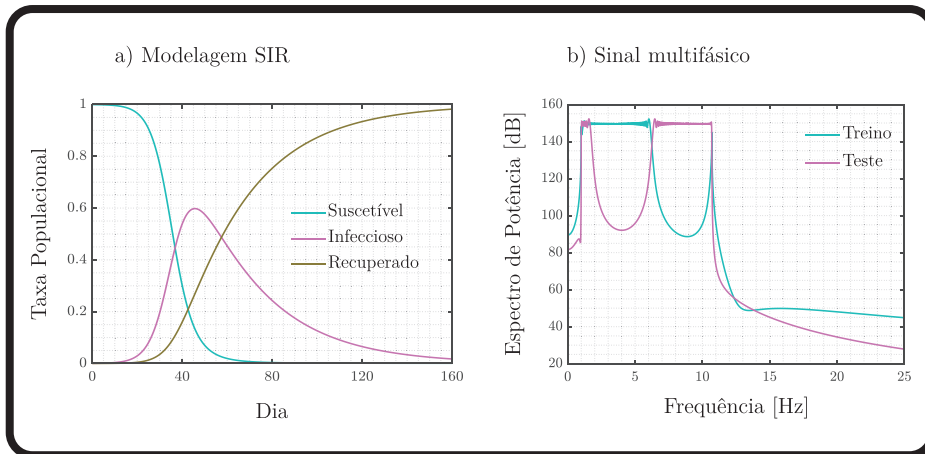


Figura 4.5: Quinto estudo de caso: a) Evolução histórica de uma modelagem Suscetível-Infeccioso-Recuperado não forçada do SARS-CoV-2; b) Excitações de treino (ciano) e teste (rosa) impostas à modelagem SIR.

Enfatiza-se também que, para cada instante de tempo, um determinado indivíduo pertencente à população só pode estar associado a um único grupo dentre os três existentes [41]. Além disso, analogamente ao modelo de FitzHugh-Nagumo, ressalta-se que uma excitação multifásica é imposta à modelagem SIR admitida, como descreve a Figura 4.5. Destaca-se ainda que a referida excitação pode ser produzida por intermédio da Equação 4-8 e que uma dada condição inicial  $x^\bullet$  ( $x^\bullet = [9996 \ 4 \ 0]^T$ ) é assumida com o objetivo de obter integralmente o conjunto de dados.

## 5

## Resultados

No presente capítulo, sob as perspectivas da classificação e da regressão, é realizado um detalhamento dos resultados obtidos por meio da metodologia original *Sparse Identification of Nonlinear Dynamics* (SINDy) e da estratégia proposta *Feature Engineering to Deal with Noisy Data in Sparse Identification* (SINDyFE) com o objetivo de acessar a estrutura mais indicada para lidar com a tarefa de identificação de sistemas dinâmicos em cenários ruidosos. Destaca-se também que, devido à causalidade temporal intrínseca aos sistemas dinâmicos admitidos nesta tese, nenhuma estratégia baseada no embaralhamento de dados é adotada [26]. Adicionalmente, uma vez que o presente estudo disserta sobre a identificação de sistemas forçados, é fundamental realizar uma investigação a fim de encontrar excitações que sejam capazes de prover a identificação adequada dos sistemas analisados. Mais especificamente, expõe-se que a investigação salientada foi realizada com o auxílio de sinais de entrada, tais como sinal de varredura, varredura senoidal, excitação multiseno, sequência pseudo binária, impacto de pulso, ruído aleatório, explosão aleatória, caminhada aleatória e forçamento multifásico [12, 45, 56, 116, 118].

Sendo assim, além de produzir os sinais de excitação, deve-se definir as condições iniciais impostas aos sistemas admitidos. Posteriormente, por meio de simulação computacional, é possível adquirir a evolução temporal dos estados de cada sistema. De forma precisa, suportado pelas expansões de Taylor de quarta e quinta ordem, explica-se que a evolução supracitada pode ser produzida a partir de um solucionador de equações diferenciais ordinárias do MATLAB amplamente conhecido como ode45. Salienta-se ainda que o solucionador previamente apresentado é desenvolvido com base em uma versão híbrida de passo variável do método de Runge-Kutta [119]. Além disso, com o propósito de gerar o conjunto de dados como um todo, foram realizadas simulações computacionais de cada estudo de caso considerando tanto um passo numericamente igual a um milésimo quanto cem mil instâncias temporais. Em seguida, alicerçado nos conceitos estabelecidos pela distribuição normal e aspirando analisar a robustez ao ruído das abordagens SINDy e SINDyFE, corrompe-se os dados artificialmente levando em conta uma diversificada coleção de percentuais de ruído [75].

Daí, munido das evoluções históricas dos estados e almejando acessar as derivadas dos mesmos sem nenhum tipo de perda de informação, pode-se aproximar tais derivadas com erros da ordem de  $\mathcal{O}\{\Delta t^4\}$  a partir de técnicas, como Diferenças Finitas Regressivas, Diferenças Finitas Centrais e Diferenças Finitas Progressivas [73]. Mais formalmente, sob a ótica da inferência de sistemas dinâmicos, os sinais de entrada, os estados e as derivadas dos estados podem ser interpretados como um sinônimo de conjunto de dados, caracterizando, por sua vez, a matéria prima básica capaz de viabilizar o processo de identificação. Explana-se ainda que um percentual de separação deve ser definido com o propósito de dividir os dados como um todo nos subconjuntos disjuntos de treino e teste. Ressalta-se também que, independentemente da metodologia adotada, os dados de treino são utilizados somente na etapa de treino do modelo, possibilitando o aprendizado e a descoberta de uma descrição matemática do sistema considerado.

Em contraste, após efetuar a partição de tais dados, deve-se reservar o conjunto de teste de forma que este só seja visualizado e manipulado no estágio final do processo de identificação, ou seja, na etapa de avaliação do modelo inferido. Dessa forma, enfatiza-se que a adesão de tais considerações são capazes de evitar a ocorrência do viés de espionagem, assim como facilitar o acesso à capacidade efetiva de um dado modelo equilibrar subajuste e sobreajuste [34, 105]. Explica-se também que todos os resultados apresentados no presente capítulo são fundamentados única e exclusivamente nos dados de teste [57]. Em detalhes, são adotadas métricas de avaliação, tais como Precisão (P), Sensibilidade (S), Coeficiente de Similaridade Dice (F), Erro Absoluto Médio (EAM), Raiz do Erro Quadrático Médio (REQM) e Coeficiente de Determinação Múltipla (CDM) [48, 20].

Esclarece-se ainda que, quanto mais próximas da unidade estiverem as métricas de classificação citadas, maior é a capacidade do modelo inferido identificar corretamente tanto os termos inativos quanto os termos ativos presentes no sistema. Em adição, quanto mais afastadas de zero estiverem as duas primeiras métricas de regressão explicitadas anteriormente, menores as similaridades observadas entre o comportamento dinâmico do modelo inferido e do sistema real. Finalmente, ao considerar a última métrica de regressão salientada, destaca-se que quanto mais próxima da unidade estiver a referida métrica, maior a fidedignidade do modelo identificado. Pontua-se também que, para ambas metodologias abordadas nesta tese, um percentual de separação numericamente igual a cinquenta por cento é assumido [25]. Em outros termos, admite-se uma partição aproximadamente igualitária entre o conjunto de treino e o conjunto de teste.



Expõe-se ainda que, almejando obter um desempenho satisfatório a partir da estratégia original, deve-se realizar uma investigação não automatizada visando encontrar um hiperparâmetro de esparsidade adequado [65]. Por outro lado, a estrutura proposta é alicerçada na automatização de procedimentos, como seleção de características representativas, aquisição de um hiperparâmetro de esparsidade apropriado e busca pelo conjunto ótimo de termos de recuperação admitidos na reconstrução da matriz de dados aumentada centralizada. Detalhadamente, a fim de implementar a metodologia proposta, foi definida uma quantidade máxima numericamente igual a cinco no que diz respeito aos termos de recuperação que podem ser descartados, assim como foi admitido um conjunto contendo cem hiperparâmetros de esparsidade logaritmicamente espaçados, onde os valores de mínimo e máximo definidos para tal hiperparâmetro são baseados, respectivamente, em zero e na unidade.

Realça-se ainda que, embora a estratégia pioneira não exija nenhum tipo de pré-processamento, é crucial pontuar a necessidade de realizar a centralização prévia da matriz de dados aumentada ao implementar a metodologia proposta [13]. Destaca-se também que os resultados apresentados neste material são exibidos nos formatos algébrico, tabular e gráfico. Em geral, ao levar em conta uma coleção de percentuais de ruído para ambas estruturas abordadas no presente trabalho, são fornecidas tabelas objetivando tanto prover as métricas de avaliação obtidas para cada um dos estudos de caso previamente descritos quanto reportar a quantidade de características selecionadas  $\sigma$  ( $\sigma \in \mathbb{N}^*$ ), o hiperparâmetro de esparsidade admitido  $\lambda$  ( $\lambda \in \mathbb{R}_+^*$ ) e o número de termos de recuperação descartados  $\delta$  ( $\delta \in \mathbb{N}$ ). Observe ainda que a soma entre a quantidade de termos de recuperação que são mantidos e descartados deve ser igual ao número de características existentes.

Dito isso, independentemente do percentual de ruído admitido para um dado estudo de caso, assume-se que a estratégia não automatizada original admite sempre o mesmo hiperparâmetro de esparsidade, onde este é obtido a partir da submissão da referida estratégia aos dados não ruidosos. Paralelamente, uma vez que a estrutura pioneira não é suportada pela engenharia de recursos, é natural definir que tal estrutura seleciona todas as características disponíveis, assim como não descarta nenhum termo de recuperação. Salienta-se também que, para percentuais de ruído específicos, são providas as descrições matemáticas e as representações gráficas dos modelos identificados a partir das abordagens SINDy e SINDyFE. Suplementarmente, aspirando fornecer resultados mais detalhados no que concerne à dinâmica dos modelos produzidos por tais abordagens, declara-se que as diferenças entre a evolução histórica do sistema e dos referidos modelos são ilustradas graficamente.

De maneira genérica, a fim de acessar as métricas de avaliação obtidas para os estudos de caso previamente estabelecidos, diferentes percentuais de ruído são assumidos. Em outras palavras, uma vez que a produção de modelos úteis do ponto de vista da engenharia é desejada, a robustez ao ruído de ambas metodologias são analisadas até serem verificados coeficientes de determinação menores ou iguais a noventa por cento [110]. Finalmente, a fim de explicitar os resultados, admite-se a representação destes a partir do truncamento da segunda casa decimal. Em particular, ao assumir um percentual de ruído de vinte por cento para o crescimento logístico, pode-se obter por meio da estrutura pioneira o modelo dado pela Equação 5-1.

$$\dot{x}_1(t) = 1,00u(t) \quad (5-1)$$

Em contraste, sob as mesmas condições, a estratégia proposta é capaz de inferir um modelo generalizável do sistema, como explica a Equação 5-2.

$$\dot{x}_1(t) = 0,99x_1(t) - 0,99x_1^2(t) + 1,00u(t) \quad (5-2)$$

Além disso, ao admitir percentuais de ruído superiores a vinte por cento, é possível verificar os equívocos sistemáticos cometidos pela estratégia original no que diz respeito à tarefa de identificação. Explica-se ainda que, ao considerar um percentual de ruído de vinte por cento, pode-se acessar a evolução histórica do sistema real e dos modelos produzidos a partir das metodologias SINDy e SINDyFE, como ilustra a Figura 5.1.

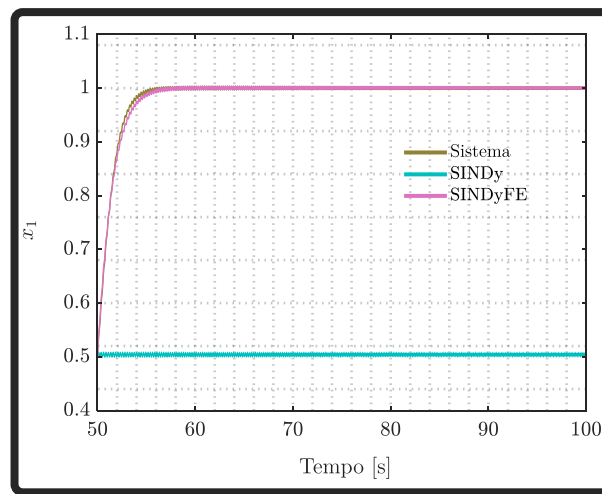


Figura 5.1: Dinâmica do sistema real e dos modelos gerados por meio das estruturas SINDy e SINDyFE para o crescimento logístico ( $\eta = 20\%$ ).

Precisamente, almejando disponibilizar uma comparação mais relevante no que diz respeito à dinâmica dos modelos identificados, é possível considerar as discrepâncias entre a evolução temporal do sistema real e dos referidos modelos, como retrata a Figura 5.2.

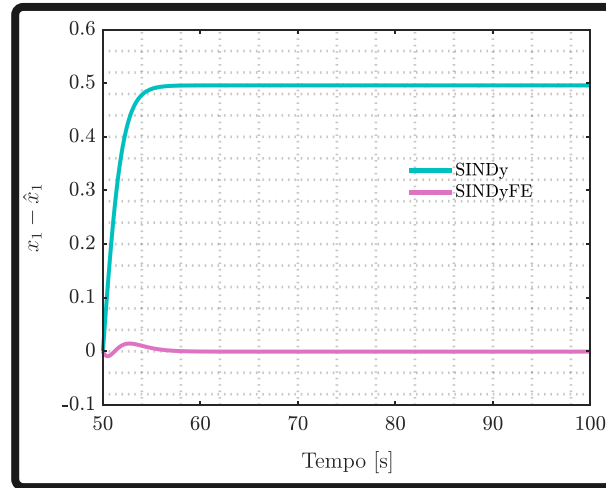


Figura 5.2: Discrepâncias entre a dinâmica do sistema e dos modelos produzidos pelas estruturas SINDy e SINDyFE para o crescimento logístico ( $\eta = 20\%$ ).

Note que ao comparar as descrições matemáticas inferidas por meio das abordagens admitidas nesta tese com o sistema real, pode-se concluir que, diferentemente da estratégia SINDy, o modelo gerado pela estrutura SINDyFE é ligeiramente diferente do sistema considerado. Enfatiza-se também que, para o percentual de ruído considerado, a estrutura original captura somente um termo classificado como verdadeiro positivo dentre os três termos existentes na descrição matemática do sistema analisado. Destaca-se também que, fundamentada em uma ampla coleção de cenários ruidosos, a Tabela 5.1 tem o propósito de fornecer os desempenhos das estratégias abordadas nesta tese sob uma perspectiva classificatória.

Tabela 5.1: Crescimento logístico (métricas de classificação).

$\eta$ [%]	SINDy			SINDyFE		
	P	S	F	P	S	F
0	1,00	1,00	1,00	1,00	1,00	1,00
10	1,00	1,00	1,00	1,00	1,00	1,00
20	0,60	1,00	0,75	1,00	1,00	1,00
30	0,60	1,00	0,75	0,75	0,75	0,75
40	0,60	1,00	0,75	0,75	0,75	0,75
50	0,60	1,00	0,75	0,75	0,75	0,75

Ressalta-se ainda que a admissão de um hiperparâmetro de bifurcação unitário acarreta na obtenção de um sistema sem bifurcações, ou seja, na constatação de um sistema detentor de um único estado estacionário. Dessa maneira, devido à simplicidade dinâmica inerente ao sistema abordado, é possível verificar que a metodologia proposta é capaz de obter modelos altamente fidedignos a tal sistema mesmo quando elevados percentuais de ruído são impostos. De maneira complementar, a Tabela 5.2 expõe as métricas de regressão verificadas para as estruturas SINDy e SINDyFE.

Tabela 5.2: Crescimento logístico (métricas de regressão).

$\eta$ [%]	SINDy			SINDyFE		
	EAM	REQM	CDM	EAM	REQM	CDM
0	$6,17 \times 10^{-8}$	$1,18 \times 10^{-6}$	1,00	$6,17 \times 10^{-8}$	$1,18 \times 10^{-6}$	1,00
10	$1,20 \times 10^{-6}$	$1,21 \times 10^{-5}$	1,00	$1,20 \times 10^{-6}$	$1,21 \times 10^{-5}$	1,00
20	$4,82 \times 10^{-3}$	$4,86 \times 10^{-2}$	-64,55	$2,19 \times 10^{-6}$	$4,48 \times 10^{-5}$	0,99
30	$4,82 \times 10^{-3}$	$4,86 \times 10^{-2}$	-64,55	$1,50 \times 10^{-5}$	$3,30 \times 10^{-4}$	0,99
40	$4,82 \times 10^{-3}$	$4,86 \times 10^{-2}$	-64,55	$1,98 \times 10^{-5}$	$4,26 \times 10^{-4}$	0,99
50	$4,82 \times 10^{-3}$	$4,86 \times 10^{-2}$	-64,55	$2,70 \times 10^{-5}$	$5,67 \times 10^{-4}$	0,99

Genericamente, no que concerne às métricas de classificação, é tangível verificar desempenhos similares entre a abordagem original e a estrutura proposta. De maneira alternativa, ao considerar as métricas de regressão, pode-se concluir que a metodologia proposta é substancialmente superior a estratégia original. Adicionalmente, no que diz respeito ao primeiro percentual de ruído admitido para o primeiro estudo de caso, pode-se afirmar que a incorporação da engenharia de recursos existente na estrutura proposta é desnecessária, pois todas as características existentes são utilizadas e nenhum termo de recuperação é descartado, como revela a Tabela 5.3.

Tabela 5.3: Crescimento logístico (parâmetros e hiperparâmetros).

$\eta$ [%]	SINDy			SINDyFE		
	$\sigma$	$\lambda$	$\delta$	$\sigma$	$\lambda$	$\delta$
0	4	$5,00 \times 10^{-1}$	0	4	$4,97 \times 10^{-1}$	0
10	4	$5,00 \times 10^{-1}$	0	3	$1,12 \times 10^{-1}$	0
20	4	$5,00 \times 10^{-1}$	0	3	$1,00 \times 10^{-2}$	1
30	4	$5,00 \times 10^{-1}$	0	3	$1,00 \times 10^{-2}$	0
40	4	$5,00 \times 10^{-1}$	0	3	$1,70 \times 10^{-1}$	0
50	4	$5,00 \times 10^{-1}$	0	3	$1,00 \times 10^{-2}$	0

Observe que a Tabela 5.3 viabiliza o acesso aos parâmetros e hiperparâmetros utilizados pelas estruturas admitidas nesta tese. Em adição, ao impor um percentual de ruído de sete por cento ao segundo estudo de caso, é exequível obter a partir da adoção da estrutura pioneira o modelo descrito por intermédio das Equações 5-3 e 5-4.

$$\dot{x}_1(t) = -0,02 + 1,00x_2(t) + 0,01u(t) \quad (5-3)$$

$$\dot{x}_2(t) = -1,03x_1(t) - 0,04x_2(t) - 0,01x_1^3(t) + 0,99u(t) \quad (5-4)$$

Por outro lado, ao levar em conta o mesmo conjunto de dados, explica-se que o modelo produzido por meio da implementação da metodologia proposta é representado pelas Equações 5-5 e 5-6.

$$\dot{x}_1(t) = 1,00x_2(t) \quad (5-5)$$

$$\dot{x}_2(t) = -1,03x_1(t) - 0,04x_2(t) - 0,01x_1^3(t) + 0,99u(t) \quad (5-6)$$

Além disso, ao comparar o sistema real com os modelos obtidos pelas estratégias admitidas neste material, pode-se constatar sob o ponto de vista classificatório que o modelo fornecido pela abordagem original dispõe de dois termos falsos positivos. Desse modo, suportada pelas estruturas SINDy e SINDyFE, a Figura 5.3 almeja descrever a dinâmica associada tanto ao sistema quanto aos modelos inferidos por tais estruturas.

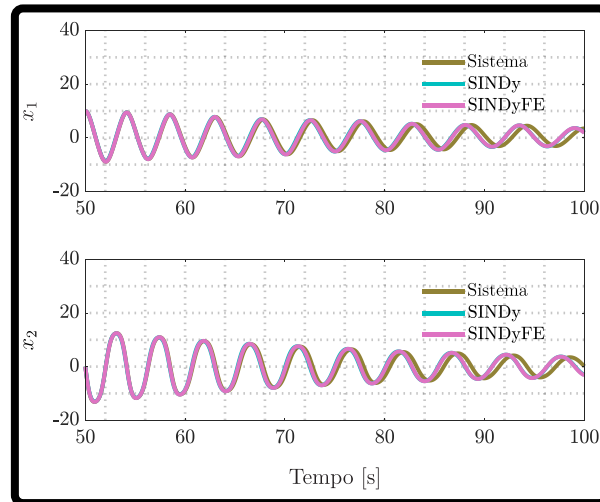


Figura 5.3: Evolução ao longo do tempo do comportamento dinâmico do sistema real e dos modelos fornecidos por intermédio das abordagens SINDy e SINDyFE para o oscilador Duffing ( $\eta = 7\%$ ).

Mais detalhadamente, como ilustra a Figura 5.4, as diferenças entre as evoluções temporais dos estados do sistema e dos estados do modelo inferido pela metodologia original envelopam as discrepâncias entre as evoluções históricas dos estados do sistema e dos estados do modelo identificado pela estrutura proposta, corroborando a superioridade da abordagem proposta.

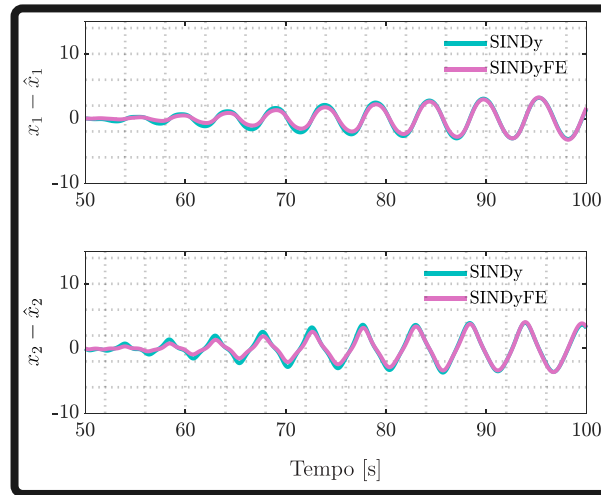


Figura 5.4: Diferenças entre a evolução histórica do sistema e dos modelos obtidos pelas estratégias SINDy e SINDyFE para o oscilador Duffing ( $\eta = 7\%$ ).

Destaca-se ainda que, ao considerar a Tabela 5.4, é viável acessar a superioridade da abordagem proposta em relação à estrutura pioneira sob uma perspectiva classificatória.

Tabela 5.4: Oscilador Duffing (métricas de classificação).

$\eta$ [%]	SINDy			SINDyFE		
	P	S	F	P	S	F
0	1,00	1,00	1,00	1,00	1,00	1,00
1	1,00	1,00	1,00	1,00	1,00	1,00
3	1,00	1,00	1,00	1,00	1,00	1,00
5	1,00	1,00	1,00	1,00	1,00	1,00
7	1,00	0,71	0,83	1,00	1,00	1,00
9	1,00	0,55	0,71	1,00	1,00	1,00

Genericamente, sob o ponto de vista da regressão, é possível constatar que os desempenhos obtidos por intermédio da adoção da metodologia proposta são ligeiramente melhores do que os desempenhos verificados pela estratégia pioneira, como detalha a Tabela 5.5.

Tabela 5.5: Oscilador Duffing (métricas de regressão).

$\eta$ [%]	SINDy			SINDyFE		
	EAM	REQM	CDM	EAM	REQM	CDM
0	<b><math>4,28 \times 10^{-7}</math></b>	<b><math>3,79 \times 10^{-6}</math></b>	<b>1,00</b>	<b><math>4,28 \times 10^{-7}</math></b>	<b><math>3,79 \times 10^{-6}</math></b>	<b>1,00</b>
1	$1,12 \times 10^{-3}$	$9,49 \times 10^{-3}$	<b>0,99</b>	<b><math>1,11 \times 10^{-3}</math></b>	<b><math>9,38 \times 10^{-3}</math></b>	<b>0,99</b>
3	$4,66 \times 10^{-3}$	$4,02 \times 10^{-2}$	<b>0,99</b>	<b><math>4,63 \times 10^{-3}</math></b>	<b><math>4,00 \times 10^{-2}</math></b>	<b>0,99</b>
5	$9,86 \times 10^{-3}$	$8,79 \times 10^{-2}$	0,98	<b><math>9,84 \times 10^{-3}</math></b>	<b><math>8,76 \times 10^{-2}</math></b>	<b>0,99</b>
7	<b><math>1,66 \times 10^{-2}</math></b>	<b><math>1,52 \times 10^{-1}</math></b>	<b>0,95</b>	<b><math>1,66 \times 10^{-2}</math></b>	<b><math>1,52 \times 10^{-1}</math></b>	<b>0,95</b>
9	$2,87 \times 10^{-2}$	$2,54 \times 10^{-1}$	0,87	<b><math>2,50 \times 10^{-2}</math></b>	<b><math>2,32 \times 10^{-1}</math></b>	<b>0,89</b>

Categoricamente, a identificação precipitada de um termo não genuíno intitulado força de reação do perfil sobre o rolamento foi reportada em uma aplicação experimental da estratégia pioneira [70]. Dessa maneira, pontua-se que a identificação equivocada de tal termo poderia ter sido evitada se a metodologia proposta fosse implementada. De modo complementar, a Tabela 5.6 almeja descrever os parâmetros e hiperparâmetros empregados com o propósito de implementar as estruturas abordadas neste material.

Tabela 5.6: Oscilador Duffing (parâmetros e hiperparâmetros).

$\eta$ [%]	SINDy			SINDyFE		
	$\sigma$	$\lambda$	$\delta$	$\sigma$	$\lambda$	$\delta$
0	11	$1,10 \times 10^{-2}$	0	4	$1,00 \times 10^{-2}$	1
1	11	$1,10 \times 10^{-2}$	0	4	$1,00 \times 10^{-2}$	1
3	11	$1,10 \times 10^{-2}$	0	4	$1,00 \times 10^{-2}$	1
5	11	$1,10 \times 10^{-2}$	0	4	$1,00 \times 10^{-2}$	1
7	11	$1,10 \times 10^{-2}$	0	4	$1,00 \times 10^{-2}$	1
9	11	$1,10 \times 10^{-2}$	0	4	$1,52 \times 10^{-2}$	0

Explicita-se também que, ao adotar um percentual de ruído de nove por cento para o segundo estudo de caso, tanto a seleção de características representativas quanto a admissão de um novo hiperparâmetro de esparsidade viabilizam a produção de um modelo ligeiramente mais parcimonioso do que o modelo identificado pela abordagem pioneira. Ressalta-se ainda que, visando apresentar os resultados obtidos para o terceiro estudo de caso abordado nesta tese, um percentual de ruído de cinco por cento é imposto ao modelo de FitzHugh-Nagumo. Dito isso, ao considerar o estudo de caso supracitado, é possível inferir a partir da implementação da metodologia original o modelo descrito pelas Equações 5-7 e 5-8.

$$\begin{aligned}
\dot{x}_1(t) = & 0,11 + 1,01x_1(t) - 1,15x_2(t) + 0,01x_1^2(t) + \\
& - 0,04x_1(t)x_2(t) - 0,20x_2^2(t) - 0,32x_1^3(t) + \\
& - 0,02x_1^2(t)x_2(t) - 0,03x_1(t)x_2^2(t) + 0,17x_2^3(t) + \\
& + 0,99u(t)
\end{aligned} \tag{5-7}$$

$$\dot{x}_2(t) = 0,11x_1(t) - 0,09x_2(t) - 0,04x_2^2(t) - 0,03x_1(t)x_2^2(t) \tag{5-8}$$

Note que a metodologia original fornece uma descrição matemática que contém uma quantidade considerável de termos equivocadamente identificados. Em contrapartida, sob o mesmo cenário ruidoso, pode-se obter por meio da estrutura proposta o modelo expresso pelas Equações 5-9 e 5-10.

$$\dot{x}_1(t) = 0,10 + 0,99x_1(t) - 1,00x_2(t) - 0,32x_1^3(t) + 0,99u(t) \tag{5-9}$$

$$\dot{x}_2(t) = 0,10x_1(t) - 0,11x_2(t) \tag{5-10}$$

Dessa forma, a Figura 5.5 tem como principal objetivo apresentar a dinâmica vinculada tanto ao sistema real quanto aos modelos inferidos a partir das estratégias SINDy e SINDyFE.

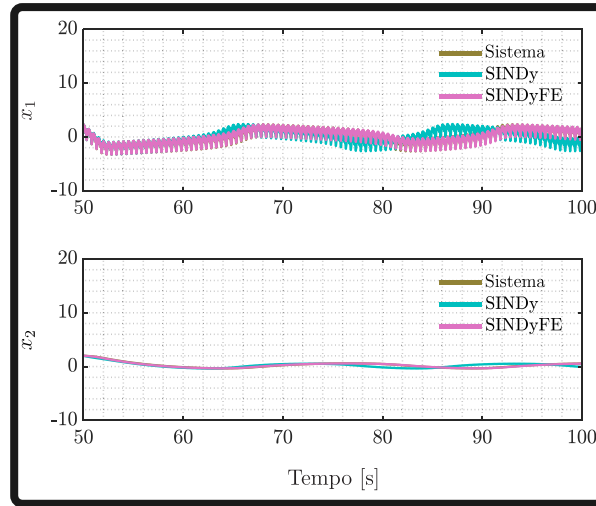


Figura 5.5: Evolução histórica do comportamento dinâmico do sistema real e dos modelos providos pelas estruturas SINDy e SINDyFE para o modelo de FitzHugh-Nagumo ( $\eta = 5\%$ ).

Observe que a Figura 5.5 não é capaz de revelar a metodologia mais apropriada no que diz respeito às similaridades dinâmicas entre o sistema e os modelos inferidos, configurando, por sua vez, um cenário onde a análise gráfica das discrepâncias comportamentais é relevante, como reitera a Figura 5.6.



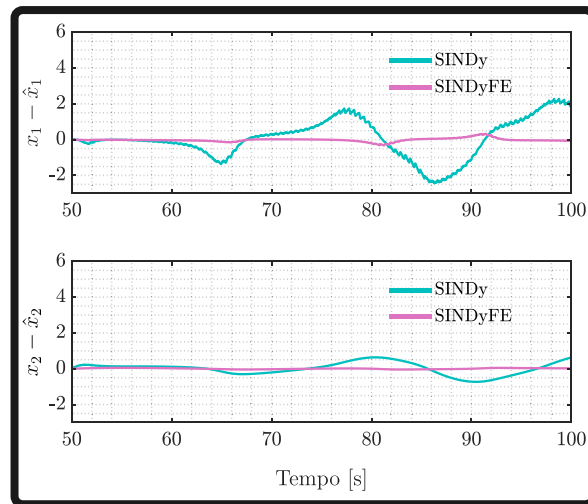


Figura 5.6: Discrepâncias entre o comportamento dinâmico do sistema real e dos modelos produzidos por intermédio das abordagens SINDy e SINDyFE para o modelo de FitzHugh-Nagumo ( $\eta = 5\%$ ).

Desse modo, alicerçado na Figura 5.6, pode-se concluir que as discrepâncias evidenciadas entre a evolução temporal do sistema e do modelo suportado pela estratégia proposta se encontram substancialmente mais próximas de zero. Além disso, como descreve a Tabela 5.7, os modelos gerados por meio das metodologias SINDy e SINDyFE apresentam termos espúrios do ponto de vista classificatório quando percentuais de ruído maiores ou iguais a dois por cento e a seis por cento são admitidos, respectivamente.

Tabela 5.7: Modelo de FitzHugh–Nagumo (métricas de classificação).

$\eta$ [%]	SINDy			SINDyFE		
	P	S	F	P	S	F
0	1,00	1,00	1,00	1,00	1,00	1,00
1	1,00	1,00	1,00	1,00	1,00	1,00
2	1,00	0,77	0,87	1,00	1,00	1,00
3	1,00	0,53	0,70	1,00	1,00	1,00
4	1,00	0,53	0,70	1,00	1,00	1,00
5	1,00	0,46	0,63	1,00	1,00	1,00
6	1,00	0,46	0,63	1,00	0,77	0,87

Enfatiza-se também que, independentemente do alto desempenho verificado para a estrutura proposta no que concerne à ótica classificatória, é fundamental avaliar os modelos produzidos pelas abordagens consideradas neste material sob a perspectiva da regressão, como expõe a Tabela 5.8.

Tabela 5.8: Modelo de FitzHugh–Nagumo (métricas de regressão).

$\eta$ [%]	SINDy			SINDyFE		
	EAM	REQM	CDM	EAM	REQM	CDM
0	$3,37 \times 10^{-7}$	$3,78 \times 10^{-6}$	<b>1,00</b>	$2,25 \times 10^{-7}$	$2,47 \times 10^{-6}$	<b>1,00</b>
1	<b><math>3,77 \times 10^{-4}</math></b>	<b><math>5,00 \times 10^{-3}</math></b>	<b>0,99</b>	<b><math>3,77 \times 10^{-4}</math></b>	<b><math>5,00 \times 10^{-3}</math></b>	<b>0,99</b>
2	<b><math>5,72 \times 10^{-4}</math></b>	<b><math>6,10 \times 10^{-3}</math></b>	<b>0,99</b>	<b><math>5,72 \times 10^{-4}</math></b>	<b><math>6,10 \times 10^{-3}</math></b>	<b>0,99</b>
3	$4,40 \times 10^{-3}$	$4,84 \times 10^{-2}$	0,87	<b><math>8,16 \times 10^{-4}</math></b>	<b><math>1,05 \times 10^{-2}</math></b>	<b>0,99</b>
4	$7,20 \times 10^{-3}$	$7,66 \times 10^{-2}$	0,69	<b><math>8,98 \times 10^{-4}</math></b>	<b><math>1,09 \times 10^{-2}</math></b>	<b>0,99</b>
5	$1,10 \times 10^{-2}$	$1,13 \times 10^{-1}$	0,31	<b><math>8,89 \times 10^{-4}</math></b>	<b><math>9,90 \times 10^{-3}</math></b>	<b>0,99</b>
6	$1,41 \times 10^{-2}$	$1,40 \times 10^{-1}$	-0,04	<b><math>9,50 \times 10^{-3}</math></b>	<b><math>1,07 \times 10^{-1}</math></b>	<b>0,38</b>

Assim, ao explorar o terceiro estudo de caso proposto, pode-se concluir que as estratégias SINDy e SINDyFE são efetivas para fins de engenharia quando percentuais de ruído menores ou iguais a dois por cento e a cinco por cento são admitidos, nesta ordem. Por fim, alicerçado na Tabela 5.9, pode-se acessar os parâmetros e hiperparâmetros admitidos por tais estratégias.

Tabela 5.9: Modelo de FitzHugh–Nagumo (parâmetros e hiperparâmetros).

$\eta$ [%]	SINDy			SINDyFE		
	$\sigma$	$\lambda$	$\delta$	$\sigma$	$\lambda$	$\delta$
0	11	$1,00 \times 10^{-2}$	0	5	$1,00 \times 10^{-2}$	0
1	11	$1,00 \times 10^{-2}$	0	5	$1,00 \times 10^{-2}$	0
2	11	$1,00 \times 10^{-2}$	0	5	$1,67 \times 10^{-2}$	0
3	11	$1,00 \times 10^{-2}$	0	5	$3,35 \times 10^{-2}$	0
4	11	$1,00 \times 10^{-2}$	0	5	$5,59 \times 10^{-2}$	0
5	11	$1,00 \times 10^{-2}$	0	5	$8,50 \times 10^{-2}$	0
6	11	$1,00 \times 10^{-2}$	0	6	$1,00 \times 10^{-2}$	0

Realça-se ainda que, nessa etapa do trabalho, as metodologias SINDy e SINDyFE são aplicadas em um sistema caótico conhecido como atrator de Lorenz. Em particular, ao assumir um percentual de ruído de cinco por cento, expõe-se que a estrutura pioneira produz um modelo altamente espúrio do sistema citado, reforçando a necessidade de avançar a estrutura previamente salientada com o propósito de desenvolver uma estratégia capaz de realizar a tarefa de identificação de forma adequada mesmo sob condições mais severas de ruído. Especificamente, suportado pelas Equações 5-11, 5-12 e 5-13, pode-se constatar que o modelo descoberto pela abordagem original admite em sua formulação uma quantidade expressiva de termos não genuínos.

$$\begin{aligned}
\dot{x}_1(t) = & 47,73 - 16,81x_1(t) + 9,21x_2(t) + 3,51x_3(t) + \\
& + 0,22x_1^2(t) - 0,88x_1(t)x_2(t) + 0,46x_1(t)x_3(t) + \\
& + 0,28x_2^2(t) + 0,24x_2(t)x_3(t) - 0,25x_3^2(t) + \\
& + 0,39u(t)
\end{aligned} \tag{5-11}$$

$$\begin{aligned}
\dot{x}_2(t) = & -31,38 + 30,10x_1(t) + 0,34x_2(t) - 1,07x_3(t) + \\
& - 0,12x_1^2(t) + 0,60x_1(t)x_2(t) - 1,22x_1(t)x_3(t) + \\
& - 0,19x_2^2(t) - 0,20x_2(t)x_3(t) + 0,12x_3^2(t) + \\
& + 0,34u(t)
\end{aligned} \tag{5-12}$$

$$\begin{aligned}
\dot{x}_3(t) = & -17,60 + 3,35x_1(t) - 0,05x_2(t) - 4,33x_3(t) + \\
& - 0,08x_1^2(t) + 1,27x_1(t)x_2(t) - 0,18x_1(t)x_3(t) + \\
& - 0,09x_2^2(t) - 0,05x_2(t)x_3(t) + 0,10x_3^2(t) + \\
& + 0,22u(t)
\end{aligned} \tag{5-13}$$

Alternativamente, ao assumir o mesmo cenário ruidoso, é possível acessar a partir da implementação da estratégia proposta um modelo altamente fidedigno ao atrator de Lorenz. De maneira mais formal, diferentemente da estratégia original, é possível verificar uma divergência diminuta no que concerne aos valores obtidos para os termos ativos presentes na descrição matemática do modelo identificado pela estrutura proposta, como explicitam as Equações 5-14, 5-15 e 5-16.

$$\dot{x}_1(t) = -9,98x_1(t) + 9,99x_2(t) + 0,99u(t) \tag{5-14}$$

$$\dot{x}_2(t) = 28,00x_1(t) - 1,00x_2(t) - 1,00x_1(t)x_3(t) \tag{5-15}$$

$$\dot{x}_3(t) = -2,66x_3(t) + 0,99x_1(t)x_2(t) \tag{5-16}$$

Por inspeção, sob uma perspectiva classificatória, é tangível ressaltar que o modelo inferido a partir da abordagem pioneira identifica erroneamente vinte e cinco termos como ativos e corretamente todos os termos inativos, enquanto que a metodologia proposta não identifica nenhum termo de maneira equivocada. Observe que a superioridade do algoritmo de aprendizado de máquina proposto nesta tese se torna cada vez mais evidente ao passo que estudos de caso mais complexos são considerados. Pontua-se ainda que a Figura 5.7 tem o propósito de ilustrar as evoluções históricas dos estados do sistema e as evoluções temporais dos estados dos modelos gerados por intermédio das abordagens SINDy e SINDyFE.

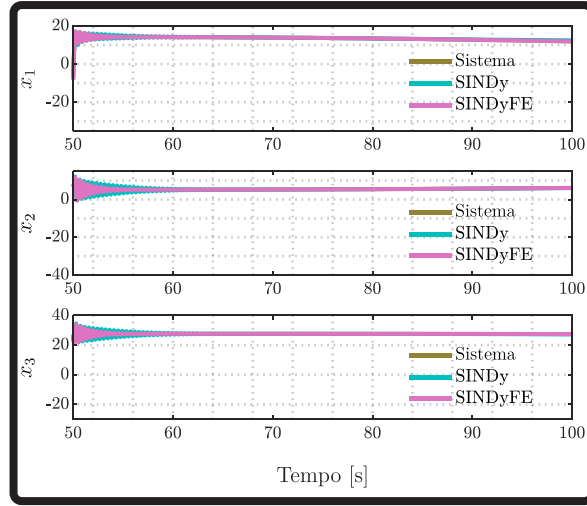


Figura 5.7: Comportamento dinâmico do sistema e dos modelos produzidos por meio das metodologias SINDy e SINDyFE para o atrator de Lorenz ( $\eta = 5\%$ ).

Mais precisamente, desejando prover uma análise dinâmica mais relevante dos resultados reportados anteriormente, pontua-se que a Figura 5.8 visa ilustrar as discrepâncias entre as evoluções temporais dos estados do sistema e as evoluções ao longo do tempo dos estados dos modelos obtidos por meio da implementação das estratégias SINDy e SINDyFE.

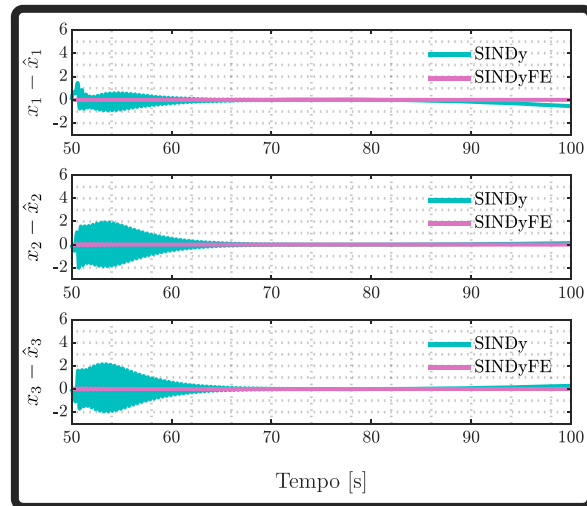


Figura 5.8: Diferenças entre a evolução do sistema e dos modelos gerados a partir das estruturas SINDy e SINDyFE para o atrator de Lorenz ( $\eta = 5\%$ ).

De maneira complementar, ao levar em conta um diversificado acervo de cenários ruidosos, pode-se salientar que a Tabela 5.10 tem como finalidade apresentar as métricas de classificação obtidas para o quarto estudo de caso abordado no presente trabalho.

Tabela 5.10: Atrator de Lorenz (métricas de classificação).

$\eta$ [%]	SINDy			SINDyFE		
	P	S	F	P	S	F
0	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>
1	<b>1,00</b>	0,47	0,64	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>
3	<b>1,00</b>	0,24	0,39	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>
5	<b>1,00</b>	0,24	0,39	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>
7	<b>1,00</b>	0,25	0,40	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>
9	<b>1,00</b>	0,25	0,40	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>
11	<b>1,00</b>	0,25	0,40	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>
13	<b>1,00</b>	0,25	0,40	0,88	<b>0,80</b>	<b>0,84</b>

Assim, é exequível concluir a partir da Tabela 5.10 que a estrutura pioneira identifica de maneira equivocada os termos inativos presentes no sistema de Lorenz quando percentuais de ruído maiores ou iguais a um por cento são impostos. Pontua-se também que a Tabela 5.11 viabiliza o acesso às métricas de regressão obtidas para o referido atrator.

Tabela 5.11: Atrator de Lorenz (métricas de regressão).

$\eta$ [%]	SINDy			SINDyFE		
	EAM	REQM	CDM	EAM	REQM	CDM
0	$1,94 \times 10^{-8}$	$3,16 \times 10^{-7}$	<b>1,00</b>	$6,92 \times 10^{-9}$	$1,04 \times 10^{-7}$	<b>1,00</b>
1	$3,80 \times 10^{-4}$	$5,94 \times 10^{-3}$	0,99	$1,26 \times 10^{-5}$	$1,89 \times 10^{-4}$	<b>1,00</b>
3	$2,81 \times 10^{-3}$	$3,60 \times 10^{-2}$	0,93	$5,36 \times 10^{-5}$	$8,70 \times 10^{-4}$	<b>1,00</b>
5	$6,18 \times 10^{-3}$	$7,09 \times 10^{-2}$	0,73	$8,61 \times 10^{-5}$	$1,41 \times 10^{-3}$	<b>0,99</b>
7	$1,06 \times 10^{-2}$	$1,09 \times 10^{-1}$	0,37	$2,28 \times 10^{-4}$	$3,87 \times 10^{-3}$	<b>0,99</b>
9	$4,56 \times 10^{-1}$	$1,54 \times 10^2$	$-1,24 \times 10^6$	$4,73 \times 10^{-4}$	$8,19 \times 10^{-3}$	<b>0,99</b>
11	$4,01 \times 10^{-1}$	$1,28 \times 10^2$	$-8,63 \times 10^5$	$6,98 \times 10^{-4}$	$1,20 \times 10^{-2}$	<b>0,99</b>
13	$4,05 \times 10^{-1}$	$1,36 \times 10^2$	$-9,67 \times 10^5$	$1,35 \times 10^{-2}$	$1,55 \times 10^{-1}$	<b>-0,24</b>

Desse modo, sob a perspectiva da engenharia, é tangível concluir por meio da Tabela 5.11 que a abordagem original pode ser adotada somente quando percentuais de ruído menores do que um por cento forem verificados, enquanto que a metodologia proposta pode ser admitida até mesmo quando percentuais de ruído iguais a onze por cento forem evidenciados. Em paralelo, no que diz respeito ao quarto e penúltimo estudo de caso abordado neste material, enfatiza-se que a Tabela 5.12 almeja explicitar os parâmetros e hiperparâmetros utilizados pelas estruturas SINDy e SINDyFE.

Tabela 5.12: Atrator de Lorenz (parâmetros e hiperparâmetros).

SINDy				SINDyFE		
$\eta$ [%]	$\sigma$	$\lambda$	$\delta$	$\sigma$	$\lambda$	$\delta$
0	11	$3,00 \times 10^{-2}$	0	6	$1,00 \times 10^{-2}$	1
1	11	$3,00 \times 10^{-2}$	0	6	$1,00 \times 10^{-2}$	0
3	11	$3,00 \times 10^{-2}$	0	7	$1,96 \times 10^{-1}$	0
5	11	$3,00 \times 10^{-2}$	0	6	$4,43 \times 10^{-2}$	1
7	11	$3,00 \times 10^{-2}$	0	6	$6,42 \times 10^{-2}$	1
9	11	$3,00 \times 10^{-2}$	0	7	$3,43 \times 10^{-1}$	0
11	11	$3,00 \times 10^{-2}$	0	6	$1,62 \times 10^{-1}$	1
13	11	$3,00 \times 10^{-2}$	0	11	$9,11 \times 10^{-1}$	0

Finalmente, ao impor um percentual de ruído de cinco por cento ao último estudo de caso abordado nesta tese, é possível obter um modelo da modelagem Suscetível-Infeccioso-Recuperado (SIR) do *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2) por intermédio da estratégia original, como retratam as Equações 5-17, 5-18 e 5-19.

$$\begin{aligned} \dot{x}_1(t) = & -2,04 + 0,96x_3(t) - 0,24x_1(t)x_2(t) + \\ & -0,01x_2(t)x_3(t) - 0,02x_3^2(t) + 0,99u(t) \end{aligned} \quad (5-17)$$

$$\dot{x}_2(t) = -0,86 + 0,01x_1(t) + 0,01x_3(t) + 0,24x_1(t)x_2(t) \quad (5-18)$$

$$\dot{x}_3(t) = 0,03x_2(t) \quad (5-19)$$

Observe que o modelo fornecido por meio da metodologia pioneira possui oito termos identificados de maneira equivocada, a saber, sete termos falsos positivos e um termo falso negativo. Realça-se também que, ao submeter a metodologia proposta ao mesmo conjunto de dados, é possível acessar um modelo expressivamente fidedigno à modelagem SIR considerada. De maneira detalhada, suportado pela implementação do algoritmo de aprendizado de máquina proposto nesta tese, é exequível obter um modelo generalizável da modelagem Suscetível-Infeccioso-Recuperado do SARS-CoV-2, como reiteram as Equações 5-20, 5-21 e 5-22.

$$\dot{x}_1(t) = -0,24x_1(t)x_2(t) + 0,99u(t) \quad (5-20)$$

$$\dot{x}_2(t) = -0,03x_2(t) + 0,24x_1(t)x_2(t) \quad (5-21)$$

$$\dot{x}_3(t) = 0,03x_2(t) \quad (5-22)$$

Dessa forma, por intermédio das estruturas SINDy e SINDyFE, a Figura 5.9 descreve o comportamento dinâmico tanto do sistema considerado quanto dos modelos inferidos pelas estruturas previamente explicitadas.

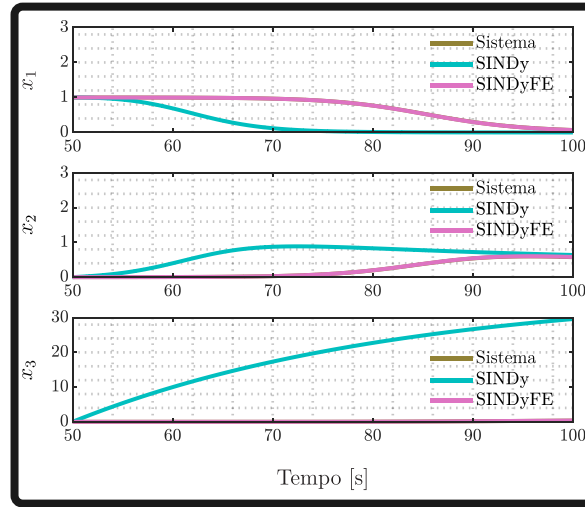


Figura 5.9: Comportamento dinâmico do sistema e dos modelos fornecidos por meio das abordagens SINDy e SINDyFE para a modelagem Suscetível-Infeccioso-Recuperado do SARS-CoV-2 ( $\eta = 5\%$ ).

Detalha-se ainda que, diferentemente do modelo produzido pela estratégia proposta, é possível constatar divergências comportamentais apreciáveis ao explorar as diferenças entre as evoluções ao longo do tempo dos estados do sistema e as evoluções históricas dos estados do modelo inferido por meio da metodologia original, como comprova a Figura 5.10.

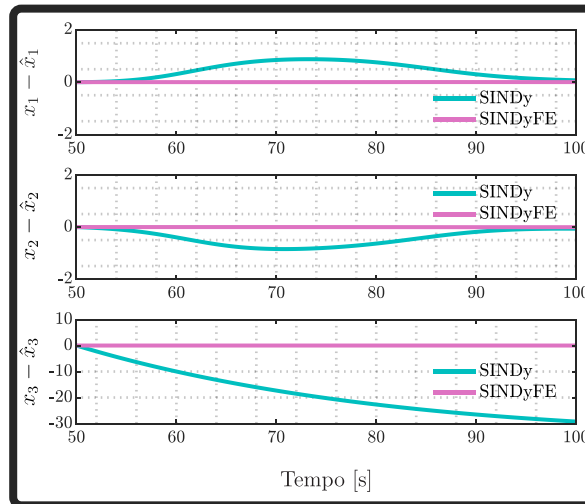


Figura 5.10: Discrepâncias entre a evolução temporal do sistema e dos modelos produzidos por intermédio das estruturas SINDy e SINDyFE para a modelagem Suscetível-Infeccioso-Recuperado do SARS-CoV-2 ( $\eta = 5\%$ ).

Adicionalmente, baseado na Tabela 5.13, pode-se acessar as métricas de classificação obtidas tanto para abordagem pioneira quanto para o algoritmo de aprendizado de máquina proposto neste material.

Tabela 5.13: Modelagem SIR do SARS-CoV-2 (métricas de classificação).

$\eta$ [%]	SINDy			SINDyFE		
	P	S	F	P	S	F
0	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>
1	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>
2	<b>1,00</b>	0,83	0,90	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>
3	0,83	0,62	0,71	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>
4	<b>1,00</b>	0,45	0,62	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>
5	0,83	0,41	0,55	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>
6	<b>1,00</b>	0,41	0,58	0,55	<b>0,71</b>	<b>0,62</b>

De modo complementar, realça-se também que a Tabela 5.14 tem como objetivo apresentar as métricas de regressão obtidas pela metodologia original e pela estratégia proposta.

Tabela 5.14: Modelagem SIR do SARS-CoV-2 (métricas de regressão).

$\eta$ [%]	SINDy			SINDyFE		
	EAM	REQM	CDM	EAM	REQM	CDM
0	$3,91 \times 10^{-8}$	$3,46 \times 10^{-7}$	<b>1,00</b>	$1,14 \times 10^{-10}$	$1,08 \times 10^{-9}$	<b>1,00</b>
1	$1,48 \times 10^{-5}$	$1,48 \times 10^{-4}$	<b>1,00</b>	$1,06 \times 10^{-5}$	$1,00 \times 10^{-4}$	<b>1,00</b>
2	$2,52 \times 10^{-3}$	$2,88 \times 10^{-2}$	0,52	$2,42 \times 10^{-5}$	$2,31 \times 10^{-4}$	<b>1,00</b>
3	$1,92 \times 10^{-1}$	$2,01 \times 10^0$	$-2,31 \times 10^3$	$2,80 \times 10^{-5}$	$2,88 \times 10^{-4}$	<b>1,00</b>
4	$2,89 \times 10^{-1}$	$3,24 \times 10^0$	$-5,95 \times 10^3$	$2,77 \times 10^{-5}$	$3,01 \times 10^{-4}$	<b>0,99</b>
5	$9,26 \times 10^{-1}$	$1,09 \times 10^1$	$-6,77 \times 10^4$	$2,56 \times 10^{-5}$	$3,17 \times 10^{-4}$	<b>0,99</b>
6	$7,62 \times 10^{-1}$	$9,04 \times 10^0$	$-4,63 \times 10^4$	$2,08 \times 10^{-1}$	$1,68 \times 10^0$	$-1,60 \times 10^3$

Sendo assim, suportado pela Tabela 5.14, é possível constatar mais uma vez a superioridade da estrutura proposta no que diz respeito à tarefa de identificação de sistemas dinâmicos não lineares em condições ruidosas. De maneira específica, ao considerar a perspectiva da engenharia, é exequível afirmar que a abordagem original e a metodologia proposta são capazes de produzir modelos fidedignos à modelagem Suscetível-Infeccioso-Recuperado do SARS-CoV-2, caso sejam admitidos percentuais de ruído menores ou iguais a um por cento e a cinco por cento, nesta ordem. Ressalta-se ainda que os parâmetros e hiperparâmetros utilizados pela estratégia pioneira e pela estrutura proposta podem ser acessados a partir da Tabela 5.15.



Tabela 5.15: Modelagem SIR do SARS-CoV-2 (parâmetros e hiperparâmetros).

SINDy				SINDyFE			
$\eta$ [%]	$\sigma$	$\lambda$	$\delta$	$\sigma$	$\lambda$	$\delta$	
0	11	$1,00 \times 10^{-2}$	0	3	$1,00 \times 10^{-2}$	0	
1	11	$1,00 \times 10^{-2}$	0	3	$1,00 \times 10^{-2}$	1	
2	11	$1,00 \times 10^{-2}$	0	3	$1,00 \times 10^{-2}$	1	
3	11	$1,00 \times 10^{-2}$	0	5	$1,38 \times 10^{-2}$	0	
4	11	$1,00 \times 10^{-2}$	0	3	$1,00 \times 10^{-2}$	0	
5	11	$1,00 \times 10^{-2}$	0	3	$1,00 \times 10^{-2}$	0	
6	11	$1,00 \times 10^{-2}$	0	7	$2,59 \times 10^{-1}$	0	

De maneira geral, ao levar em conta todos estudos de caso abordados e todos cenários ruidosos investigados, é possível notar que tanto a seleção de características quanto a busca por um hiperparâmetro de esparsidade adequado foram estratégias exaustivamente utilizadas pela abordagem proposta. Em outras palavras, a seleção de características e a referida busca hiperparamétrica foram, sozinhas, responsáveis pela produção de boa parte dos modelos não espúrios produzidos pela metodologia proposta. Entretanto, ao explorar certos conjuntos de dados considerados nesta tese, pôde-se observar que a adoção da extração de características também se mostrou imprescindível no que concerne à geração de modelos generalizáveis de sistemas dinâmicos. De modo preciso, alicerçado nos três primeiros estudos de caso admitidos no presente trabalho, pôde-se verificar a exclusão do termo de recuperação associado ao menor percentual de recuperação.

Paralelamente, no que diz respeito aos demais estudos de caso abordados neste material, pôde-se detectar o descarte do termo de recuperação vinculado ou ao maior percentual de recuperação ou ao segundo maior percentual de recuperação. De maneira natural, sob a perspectiva da identificação de sistemas dinâmicos não lineares, ao passo que a quantidade de características requeridas com a finalidade de descrever de maneira adequada um certo sistema ruidoso aumenta, mais a admissão da extração de características se torna justificável, como corrobora o fenômeno da maldição da dimensionalidade. Categoricamente, enfatiza-se que as considerações previamente salientadas devem ser encaradas como uma tendência genérica, pois é possível que cenários específicos distintos sejam revelados quando outros conjuntos de dados sejam admitidos na alimentação do algoritmo de aprendizado de máquina proposto nesta tese, como reitera o teorema sem almoço grátis.

De forma detalhada, com o propósito de identificar sistemas dinâmicos, foram utilizadas técnicas de aprendizado de máquina, tais como aumento de dados, regressão esparsa, seleção de características, extração de características, critério de informação, pesquisa em grade e validação cruzada. Genericamente, sob as perspectivas da classificação e da regressão, pôde-se concluir que a metodologia proposta supera a estratégia pioneira no que diz respeito à tarefa de identificação de sistemas dinâmicos exclusivamente a partir de dados, demonstrando, por sua vez, sua apreciável robustez ao ruído. Por fim, realça-se que essa é a primeira demonstração explícita de como a hibridização das técnicas de aprendizado de máquina previamente salientadas pode ser explorada com o objetivo de inferir modelos não espúrios de sistemas dinâmicos não lineares sob condições mais severas de ruído.

Esta tese apresenta, pela primeira vez, uma hibridização de técnicas de aprendizado de máquina, tais como aumento de dados, regressão esparsa, seleção de características, extração de características, critério de informação, pesquisa em grade e validação cruzada. Desse modo, sob as perspectivas da classificação e da regressão, a estratégia proposta superou a estrutura original em sistemas como crescimento logístico, oscilador Duffing, modelo de FitzHugh–Nagumo, atrator de Lorenz e uma modelagem Suscetível-Infecioso-Recuperado. Em detalhes, enfatiza-se que a engenharia de recursos, o critério de informação, a busca em grade e a validação cruzada, viabilizaram a automatização de procedimentos, tais como a seleção de características relevantes, a obtenção do conjunto ótimo de termos de recuperação e a aquisição de um hiperparâmetro de esparsidade adequado. Portanto, alicerçado em uma comparação extensiva, demonstrou-se que o algoritmo de aprendizado de máquina proposto neste material é mais indicado para lidar com a tarefa de identificação de sistemas dinâmicos sob condições mais severas de ruído.

Os experimentos conduzidos no presente trabalho revelaram a robustez ao ruído apreciável da metodologia proposta, favorecendo a inferência de modelos parcimoniosos. Sendo assim, uma vez acessada uma descrição matemática adequada do sistema analisado, é possível admitir tal descrição a fim de realizar tarefas futuras, como predições com horizontes de previsão mais amplos e aplicações de controle. Em geral, o presente estudo revelou que a extração de características foi responsável pela adoção do termo de recuperação associado ou ao maior ou ao segundo maior percentual de recuperação, implicando em um pequeno aumento do custo computacional. Em contraste, ao passo que aumentam tanto o número de estados adotados com o propósito de descrever um determinado sistema quanto o grau máximo definido para a base de funções com o objetivo de gerar os dados aumentados, nota-se um crescimento fatorial no número de características inicialmente admitidas, culminando em uma elevação significativa do custo computacional. Enfatiza-se também que a etapa de seleção de características visa simplificar ao máximo o número de características inicialmente consideradas, evitando o aumento do custo computacional e promovendo a descoberta de modelos generalizáveis.

Em adição, pesquisas futuras serão devotadas a aplicações empíricas da metodologia original e da estrutura proposta, assim como ao desenvolvimento de uma interface de programação de aplicação. Particularmente, ao considerar a admissão de filtros nos sensores de um dado sistema, serão investigados os desempenhos tanto da abordagem original quanto da estratégia proposta. Por outro lado, alicerçado no carregamento do conjunto de dados e na configuração de poucos parâmetros, o desenvolvimento da interface de programação de aplicação viabilizará o provimento do melhor modelo descoberto a partir de uma coleção de algoritmos de aprendizado de máquina. Além disso, a interface citada fornecerá, prontamente, uma comparação adequada entre os modelos produzidos a partir de cada uma das metodologias admitidas, acelerando, por sua vez, a tarefa de identificação de sistemas dinâmicos.

## Referências bibliográficas

- [1] GUNKEL, S. N. B.; STOKKING, H. M.; HINDRIKS, R.; KONINCK, T. **VR conferencing: communicating and collaborating in photo-realistic social immersive environments.** International Journal of Virtual Reality, p. 45–46, 2019.
- [2] YANG, S.; LIU, W.; SUN, D.; AND LI, C. **A new extended multiple car-following model considering the backward-looking effect on traffic flow.** Journal of Computational and Nonlinear Dynamics, 8(1), 2013.
- [3] WHETTEN, J. M.; GOODRICH, M. A.; YISONG, G. **Beyond robot fan-out: Towards multi-operator supervisory control.** In: IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN AND CYBERNETICS, p. 2008–2015, 2010.
- [4] MULLAINATHAN, S.; SPIESS, J. **Machine learning: an applied econometric approach.** Journal of Economic Perspectives, 31(2):87–106, May 2017.
- [5] MENDEN, M. P.; IORIO, F.; GARNETT, M.; MCDERMOTT, U.; BENES, C. H.; BALLESTER, P. J.; RODRIGUEZ, J. S. **Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties.** PLOS ONE, 8(4):1–7, 2013.
- [6] YAMAN, M. A.; SUBASI, A.; RATTAY, F. **Comparison of random subspace and voting ensemble machine learning methods for face recognition.** Symmetry, 10(11), 2018.
- [7] OUYANG, S.; YI, L.; Y. WU. **Prediction of natural disasters.** International Journal of General Systems, 29(6):897–912, 2000.
- [8] COHEN, P. R.. **DARPA’s big mechanism program.** Physical Biology, 12(4), 2015.
- [9] BILLINGS, S. A. **Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains.** John Wiley & Sons, 2013.

- [10] KEESMAN, K. J. **System identification: an introduction**. Springer, 2011.
- [11] LJUNG, L. **System identification: theory for the user**. Prentice Hall, 1999.
- [12] PINTELON, R.; SCHOUKENS, J. **System identification: a frequency domain approach**. John Wiley & Sons, 2012.
- [13] TAYLOR, J. S.; CRISTIANINI, N. **Kernel methods for pattern analysis**. Cambridge University Press, 2004.
- [14] WOLPERT, D. H.; MACREADY, W. G. **No free lunch theorems for search**, 1995.
- [15] WOLPERT, D. H.; MACREADY, W. G. **No free lunch theorems for optimization**. Transactions on Evolutionary Computation, 1(1):67–82, 1997.
- [16] YANGMING, X.; HOLLERBACH, J. M. **A robust ensemble data method for identification of human joint mechanical properties during movement**. IEEE Transactions on Biomedical Engineering, 46(4):409–419, 1999.
- [17] SCHMIDT, M.; LIPSON, H. **Distilling free-form natural laws from experimental data**. Science, 324(5923):81–85, 2009.
- [18] LEWIS, N. D. **Deep time series forecasting with Python: an intuitive introduction to deep learning for applied time series modeling**. CreateSpace Independent Publishing Platform, 2016.
- [19] BOX, G. E.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. **Time series analysis: forecasting and control**. John Wiley & Sons, 2015.
- [20] HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning: data mining, inference and prediction**. Springer, New York, 2009.
- [21] MÜLLER, K. R.; SMOLA, A. J.; RÄTSCH, G.; SCHÖLKOPF, B.; KOHLMORGEN, J.; VAPNIK, V. **Predicting time series with support vector machines**. Artificial Neural Networks: ICANN, p. 999–1004, 1997.
- [22] WIGLEY, P. B.; EVERITT, P. J.; HENGEL, V.; BASTIAN, J. W.; SOORIYABANDARA, M. A.; MCDONALD, G. D.; HARDMAN, K. S.; QUINLIVAN, C. D.; MANJU, P.; KUHN, C. C. N.; PETERSEN, I. R.; LUITEN, A.; HOPE,

- J. J.; ROBINS, N. P.; HUSH, M. R. **Fast machine-learning online optimization of ultra-cold-atom experiments**. *Scientific Reports*, 6(1), 2015.
- [23] JIN, L.; LI, S.; YU, J.; HE, J. **Robot manipulator control using neural networks: a survey**. *Neurocomputing*, 285:23–34, 2018.
- [24] FRANÇA, T.; GRUZMAN, M. **Structural control by means of electromagnetic actuators**. *Sodebras*, 14(165), 2019.
- [25] BRUNTON, S. L.; PROCTOR, J. L.; KUTZ, J. N. **Discovering governing equations from data by sparse identification of nonlinear dynamical systems**. *Proceedings of the National Academy of Sciences of the United States of America*, 113(15):3932–3937, 2016.
- [26] FRANÇA, T. **Identificação esparsa de dinâmicas não lineares: um estudo de caso detalhado**. Trabalho de conclusão de curso, Departamento de Engenharia Mecânica, Universidade Federal Fluminense, Rio de Janeiro, 2020.
- [27] TIBSHIRANI, R. **Regression shrinkage and selection via the LASSO: a retrospective**. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3):273–282, 2011.
- [28] FAN, C.; HUANG, Y.; WANG, Q. **Sparsity-promoting polynomial response surface: a new surrogate model for response prediction**. *Advances in Engineering Software*, 77:48–65, 2014.
- [29] CHANDRASHEKAR, G.; SAHIN, F. **A survey on feature selection methods**. *Computers & Electrical Engineering*, 40(1):16–28, 2014.
- [30] DU, X.; ZHU, F. **A novel principal components analysis (pca) method for energy absorbing structural design enhanced by data mining**. *Advances in Engineering Software*, 127:17–27, 2019.
- [31] KULLBACK, S.; LEIBLER, R. A. **On information and sufficiency**. *Annals of Mathematical Statistics*, 22(1):79–86, 1951.
- [32] SCHWARZ, G. **Estimating the dimension of a model**. *Annals of Statistics*, 6(2):461–464, 1978.
- [33] LI, S.; FANG, H.; LIU, X. **Parameter optimization of support vector regression based on sine cosine algorithm**. *Expert Systems with Applications*, 91:63–77, 2018.

- [34] NELLO, C.; TAYLOR, J. S. **An introduction to support vector machines: and other kernel-based learning methods.** Cambridge University Press, New York, 1999.
- [35] ISMAIL, S. M.; SAID, L. A.; REZK, A. A.; RADWAN, A. G.; MADIAN, A. H.; ABU-ELYAZEED, M. F.; SOLIMAN, A. M. **Generalized fractional logistic map encryption system based on FPGA.** *AEU - International Journal of Electronics and Communications*, 80:114–126, 2017.
- [36] DUFFING, G. **Ingenieur: erzwungene schwingungen bei veränderlicher eigenfrequenz und ihre technische bedeutung.** *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, 1(1):72–73, 1921.
- [37] QU, H.; LI, T.; CHEN, G. **Multiple analytical mode decompositions (M-AMD) for high accuracy parameter identification of nonlinear oscillators from free vibration.** *Mechanical Systems and Signal Processing*, 117:483–497, 2019.
- [38] FITZHUGH, R. **Impulses and physiological states in theoretical models of nerve membrane.** *Biophysical Journal*, 1(6):445–466, 1961.
- [39] SUBASHINI, M. M.; SAHOO, S. K. **Pulse coupled neural networks and its applications.** *Expert Systems with Applications*, 41(8):3965–3974, 2014.
- [40] LORENZ, E. N. **Deterministic nonperiodic flow.** *Journal of the Atmospheric Sciences*, 20(2):130–141, 1963.
- [41] NDIAYE, B. M.; TENDENG, L.; SECK, D. **Analysis of the covid-19 pandemic by sir model and machine learning technics for forecasting.** *Arxiv*, 2020.
- [42] KATRIS, C. **A time series-based statistical approach for outbreak spread forecasting: Application of covid-19 in greece.** *Expert Systems with Applications*, 166, 2021.
- [43] RASCHKA, S.; MIRJALILI, V. **Python machine learning: machine learning and deep learning with Python, Scikit-Learn, and TensorFlow.** Packt Publishing, 2019.
- [44] SARANGAPANI, J. **Neural network control of nonlinear discrete-time systems.** Taylor & Francis, 2006.



- [45] NORGAARD, M.; RAVN, O.; POULSEN, N. K.; HANSEN, L. K. **Neural networks for modelling and control of dynamic systems: a practitioner's handbook**. Springer, 2000.
- [46] VAPNIK, V. N. **Statistical learning theory**. John Wiley & Sons, 1998.
- [47] BROWN, J. B. **Classifiers and their metrics quantified**. *Molecular Informatics*, 37, 2018.
- [48] REDDY, G. T.; REDDY, M. P. K.; LAKSHMANNA, K.; KALURI, R.; RAJPUT, D. S.; SRIVASTAVA, G.; BAKER, T. **Analysis of dimensionality reduction techniques on big data**. *IEEE Access*, 8:54776–54788, 2020.
- [49] HABER, R.; UNBEHAUEN, H. **Structure identification of nonlinear dynamic systems - a survey on input/output approaches**. *Automatica*, 26(4):651–677, 1990.
- [50] AKAIKE, H. **Information theory and an extension of the maximum likelihood principle**. 2nd International Symposium on Information Theory, p. 267–281, 1973.
- [51] NAKAMURA, T.; JUDD, K.; MEES, A. I.; SMALL, M. **A comparative study of information criteria for model selection**. *International Journal of Bifurcation and Chaos*, 16:2153–2175, 2006.
- [52] PENNY, W. D. **Comparing dynamic causal models using aic, bic and free energy**. *NeuroImage*, 59:319–330, 2012.
- [53] CRUTCHFIELD, J. P.; MCNAMARA, B. S. **Equations of motion from a data series**. *Complex Systems*, 1:417–452, 1987.
- [54] MEZIĆ, I. **Spectral properties of dynamical systems, model reduction and decompositions**. 2005.
- [55] COIFMAN, R. R.; LAFON, S. **Diffusion maps**. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006.
- [56] SODERSTROM, T.; STOICA, P. **System identification**. Prentice Hall, 1989.
- [57] AYALA, H. V. H.; HABINEZA, D.; RAKOTONDRABE, M.; KLEIN, C. E.; COELHO, L. S. **Nonlinear black-box system identification through neural networks of a hysteretic piezoelectric robotic micromanipulator**. *IFAC-PapersOnLine*, 48(28):409–414, 2015.

- [58] WANG, W-X.; LAI, Y-C.; GREBOGI, C. **Data based identification and prediction of nonlinear and complex dynamical systems**. *Physics Reports*, 644:1–76, 2016.
- [59] MANGAN, N. M.; BRUNTON, S. L.; PROCTOR, J. L.; KUTZ, J. N. **Inferring biological networks by sparse identification of nonlinear dynamics**. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, 2(1):52–63, 2016.
- [60] BRUNTON, S. L.; PROCTOR, J. L.; KUTZ, J. N. **Sparse identification of nonlinear dynamics with control (SINDYc)**. *IFAC*, 49(18):710–715, 2016.
- [61] LOISEAU, J-C.; BRUNTON, S. L. **Constrained sparse Galerkin regression**. *Journal of Fluid Mechanics*, 838:42–67, 2018.
- [62] SCHAEFFER, H. **Learning partial differential equations via data discovery and sparse optimization**. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2197), 2017.
- [63] RUDY, S. H.; BRUNTON, S. L.; PROCTOR, J. L.; KUTZ, J. N. **Data-driven discovery of partial differential equations**. *Science Advances*, 3(4), 2017.
- [64] TRAN, G.; WARD, R. **Exact recovery of chaotic systems from highly corrupted data**. *Multiscale Modeling & Simulation*, 15(3):1108–1129, 2017.
- [65] MANGAN, N. M.; KUTZ, J. N.; BRUNTON, S. L.; PROCTOR, J. L. **Model selection for dynamical systems via sparse regression and information criteria**. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2204), 2017.
- [66] SCHAEFFER, H.; TRAN, G.; WARD, R. **Learning dynamical systems and bifurcation via group sparsity**. *Arxiv*, 2017.
- [67] RUDY, S.; ALLA, A.; BRUNTON, S. L.; KUTZ, J. N. **Data-driven identification of parametric partial differential equations**. *SIAM Journal on Applied Dynamical Systems*, 18(2):643–660, 2019.
- [68] NARASINGAM, A.; KWON, J. S-I. **Data-driven identification of interpretable reduced-order models using sparse regression**. *Computers & Chemical Engineering*, 119:101–111, 2018.

- [69] KAISER, E.; KUTZ, J. N.; BRUNTON, S. L. **Sparse identification of nonlinear dynamics for model predictive control in the low-data limit.** *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 474(2219), 2018.
- [70] GOHAROODI, S. K.; DEKEMELE, K.; DUPRE, L.; LOCCUFIER, M.; CREVECOEUR, G. **Sparse identification of nonlinear duffing oscillator from measurement data.** *IFAC*, 51(33):162–167, 2018.
- [71] MANGAN, N. M.; ASKHAM, T.; BRUNTON, S. L.; KUTZ, J. N.; PROCTOR, J. L. **Model selection for hybrid dynamical systems via sparse regression.** *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 475(2223):2018–0534, 2019.
- [72] JINKI, K.; HARNE, R. L.; WANG, K. W. **Predicting non-stationary and stochastic activation of saddle-node bifurcation.** *Journal of Computational and Nonlinear Dynamics*, 12(1), 2016.
- [73] SADIKU, M. N. O. **Numerical techniques in electromagnetics.** CRC Press, 2009.
- [74] ARFKEN, G. B.; WEBER, H. J. **Mathematical methods for physicists.** Academic Press, 2005.
- [75] NAVIDI, W. **Statistics for engineers and scientists.** McGraw-Hill, 2006.
- [76] CHARTRAND, R. **Differentiation of noisy, nonsmooth data.** International Scholarly Research Network, 2011.
- [77] XING, L.; PHAM, D. T. **Neural networks for identification, prediction and control.** Springer-Verlag, Berlin, 1995.
- [78] BARTLE, R. G. **The elements of integration and Lebesgue measure.** John Wiley & Sons, 1995.
- [79] SALEM, S. B.; NAOUALI, S.; SALLAMI, M. **A computational cost-effective clustering algorithm in multidimensional space using the manhattan metric: application to the global terrorism database.** *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 11:702–707, 2017.

- [80] QURESHI, C.; COSTA, S. I. R. **On perfect q-ary codes in the maximum metric.** In: INFORMATION THEORY AND APPLICATIONS WORKSHOP (ITA), p. 1–4, 2016.
- [81] REID, S.; TIBSHIRANI, R.; FRIEDMAN, J. **A study of error variance estimation in lasso regression.** *Statistica Sinica*, 26(1):35–67, 2016.
- [82] SU, W.; BOGDAN, M.; CANDÈS, E. **False discoveries occur early on the lasso path.** *Annals of Statistics*, 45(5):2133–2150, 2017.
- [83] HUANG, J.; JIAO, Y.; LIU, Y.; LU, X. **A constructive approach to  $l_0$  penalized regression.** *Journal of Machine Learning Research*, 19(10):1–37, 2018.
- [84] NIKOLAENKO, V.; WEINSBERG, U.; IOANNIDIS, S.; JOYE, M.; BONEH, D.; TAFT, N. **Privacy-preserving ridge regression on hundreds of millions of records.** In: IEEE SYMPOSIUM ON SECURITY AND PRIVACY, p. 334–348, 2013.
- [85] LENTERS, V.; PORTENGEN, L.; RIGNELL-HYDBOM, A.; JÖNSSON, B. A. G.; LINDH, C. H.; PIERSMA, A. H.; TOFT, G.; BONDE, J. P.; HEEDERIK, D.; RYLANDER, L.; VERMEULEN, R. **Prenatal phthalate, perfluoroalkyl acid, and organochlorine exposures and term birth weight in three birth cohorts: multi-pollutant models based on elastic net regression.** *Environmental Health Perspectives*, 124(3):365–372, 2016.
- [86] VAZ, L. E.; PEREIRA, A.; MENEZES, I. F. M. **Programação matemática: teoria, algoritmos e aplicações na engenharia.** Rio de Janeiro, 2012.
- [87] CHOLLET, F. **Deep learning with Python.** Manning Publications Co., 2017.
- [88] SCHMID, P. J. **Dynamic mode decomposition of numerical and experimental data.** *Journal of Fluid Mechanics*, 656:5–28, 2010.
- [89] MEZIĆ, I. **Analysis of fluid flows via spectral properties of the koopman operator.** *Annual Review of Fluid Mechanics*, 45(1):357–378, 2013.
- [90] ABEEL, T.; HELLEPUTTE, T.; VAN DE PEER, Y.; DUPONT, P.; SAEYS, Y. **Robust biomarker identification for cancer diagnosis with**

- ensemble feature selection methods. *Bioinformatics*, 26(3):392–398, 2010.
- [91] FENG, L.; MIAO, D.; PEDRYCZ, W. **Granular multi-label feature selection based on mutual information**. *Pattern Recognition*, 67:410–423, 2017.
- [92] AL-SHUDEIFAT, M. A.; MEHAIRI, A. A.; SAEED, A. S.; BALAWI, S. **Application of the proper orthogonal decomposition method for cracked rotors**. *Journal of Computational and Nonlinear Dynamics*, 13(11), 2018.
- [93] JAIN, S.; TISO, P. **Simulation-free hyper-reduction for geometrically nonlinear structural dynamics: a quadratic manifold lifting approach**. *Journal of Computational and Nonlinear Dynamics*, 13(7), 2018.
- [94] GAVISH, M.; DONOHO, D. L. **The optimal hard threshold for singular values is  $4/\sqrt{3}$** . *IEEE Transactions on Information Theory*, 60(8):5040–5053, 2014.
- [95] FERNANDEZ, D.; GONZALEZ, C.; MOZOS, D.; LOPEZ, S. **FPGA implementation of the principal component analysis algorithm for dimensionality reduction of hyperspectral images**. *Journal of Real-Time Image Processing*, 16(5), 2019.
- [96] ZHOU, G-D. **Fundamentals of structural chemistry**. World Scientific Publishing, 1993.
- [97] DOW, J. O. **A unified approach to the finite element method and error analysis procedures**. Academic Press, 1999.
- [98] FARKAS, H.; KÁLY-KULLAI, K.; SIENIUTYCZ, S. **Variational and extremum principles in macroscopic systems**. Elsevier, 2005.
- [99] GUIASU, S.; SHENITZER, A. **The principle of maximum entropy**. *The Mathematical Intelligencer*, 7:42–48, 1985.
- [100] SUNTOLA, T. **The short history of science - or the long path to the union of metaphysics and empiricism**. Physics Foundations Society, 2018.
- [101] GKIIOULEKAS, I.; PAPAGEORGIOU, L. G. **Piecewise regression analysis through information criteria using mathematical programming**. *Expert Systems with Applications*, 121:362–372, 2019.

- [102] SCHABER, J.; MAX, F.; LI, J.; TIGER, C.; HOHMANN, S.; KLIPP, E. **Automated ensemble modeling with modelmage: analyzing feedback mechanisms in the sho1 branch of the HOG pathway.** Plos One, 6(3):1–7, 2011.
- [103] BLAKE, I. M.; MARTIN, R.; GOEL, A.; KHETSURIANI, N.; EVERTS, J.; WOLFF, C.; WASSILAK, S.; AYLWARD, R. B.; GRASSLY, N. C. **The role of older children and adults in wild poliovirus transmission.** Proceedings of the National Academy of Sciences, 111(29):10604–10609, 2014.
- [104] AKAIKE, H. **A new look at the statistical model identification.** IEEE Transactions on Automatic Control, 19(6):716–723, 1974.
- [105] BURNHAM, K. P.; ANDERSON, D. R. **Model selection and multimodel inference: a practical information-theoretic approach.** Springer-Verlag, 2002.
- [106] LINDE, A. **DIC in variable selection.** Statistica Neerlandica, 59(1):45–56, 2005.
- [107] BISHOP, C. M. **Pattern recognition and machine learning.** Springer, 2006.
- [108] MASKEY, A.; CHO, M. **Cubesatnet: ultralight convolutional neural network designed for on-orbit binary image classification on a 1u cubesat.** Engineering Applications of Artificial Intelligence, 96, 2020.
- [109] HABER, R.; UNBEHAUEN, H. **Structure identification of nonlinear dynamic systems - a survey on input/output approaches.** Automatica, 26(4):651–677, 1990.
- [110] SCHAIBLE, B.; XIE, H.; LEE, Y-C. **Fuzzy logic models for ranking process effects.** IEEE Transactions on Fuzzy Systems, 5(4):545–556, 1997.
- [111] NILASHI, M.; IBRAHIM, O.; BAGHERIFARD, K. **A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques.** Expert Systems with Applications, 92:507–520, 2018.
- [112] LIU, Y.; TANG, S.; LOZANO, C. F.; MUNTEANU, C. R.; PAZOS, A.; YU, Y-Z.; TAN, Z.; DÍAZ, H. G. **Experimental study and random forest**

- prediction model of microbiome cell surface hydrophobicity.** Expert Systems with Applications, 72:306–316, 2017.
- [113] LI, R.; PEI, S.; CHEN, B.; SONG, Y.; ZHANG, T.; YANG, W.; SHAMAN, J. **Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (sars-cov-2).** Science, 368(6490):489–493, 2020.
- [114] SCHAEFFER, H.; MCCALLA, S. G. **Sparse model selection via integral terms.** PHYSICAL REVIEW E, 96(2), 2017.
- [115] CATTANI, C.; GANDINO, E.; MARCHESIELLO, S. **Identification of a duffing oscillator under different types of excitation.** Mathematical Problems in Engineering, 2010.
- [116] HOAGG, J. B.; LACY, S. L.; BABUSKA, V.; BERNSTEIN, D. S. **Sequential multisine excitation signals for system identification of large space structures.** In: AMERICAN CONTROL CONFERENCE, 2006.
- [117] HOLMES, M. H. **Introduction to the foundations of applied mathematics.** Springer, New York, 2009.
- [118] COWPERTWAIT, P. S. P.; METCALFE, A. V. **Introductory time series with R.** Springer Publishing Company, Incorporated, 2009.
- [119] MOLER, C. B. **Numerical Computing with MATLAB.** SIAM, Philadelphia, 2008.

## 7

### Apêndice A

O presente capítulo disponibiliza todo material complementar necessário a fim de prover uma maior compreensão do conteúdo abordado nesta tese. Detalhadamente, descreve-se de forma mais minuciosa tópicos, tais como bases de funções, esquemas de diferenças finitas, estimativas de parâmetros estatísticos, otimização, produto interno e produto externo.

#### 7.1

##### Bases de Funções

A escolha da base de funções utilizada com a finalidade de construir a matriz de dados aumentada é uma etapa fundamental no que diz respeito à tarefa de identificação de sistemas dinâmicos não lineares via regressão esparsa. De modo genérico, a menos que exista algum tipo de conhecimento prévio sobre o sistema analisado, é altamente recomendável aplicar diferentes tipos de bases com a finalidade de descobrir a base de funções mais adequada no que diz respeito à capacidade de capturar de maneira satisfatória os padrões escondidos no conjunto de dados. Precisamente, almejando descobrir a base de funções mais apropriada para um dado sistema, pode-se admitir bases suportadas por famílias como polinômios racionais, polinômios reais, polinômios de Chebyshev e expansões da série de Fourier, para citar algumas. De maneira detalhada, é possível ressaltar que os polinômios reais são amplamente conhecidos por sua elevada flexibilidade no que concerne ao ajuste de dados, justificando, por sua vez, a exaustiva adoção da referida família como base de funções. Por exemplo, com a finalidade de identificar um determinado sistema não forçado descrito a partir do estado  $\mathbf{x}_1$  ( $\mathbf{x}_1 \in \mathbb{R}^n$ ) e do estado  $\mathbf{x}_2$  ( $\mathbf{x}_2 \in \mathbb{R}^n$ ), pode-se assumir uma base polinomial real de grau máximo numericamente igual a três. Formalmente, como expõe a Equação 7-1, é tangível obter a matriz de dados aumentada associada às considerações previamente salientadas.

$$\Theta_3^{\mathbb{R}} = \begin{bmatrix} | & | & | & | & | & | & | & | & | & | \\ \mathbf{1} & \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_1^2 & \mathbf{x}_1\mathbf{x}_2 & \mathbf{x}_2^2 & \mathbf{x}_1^3 & \mathbf{x}_1^2\mathbf{x}_2 & \mathbf{x}_1\mathbf{x}_2^2 & \mathbf{x}_2^3 \\ | & | & | & | & | & | & | & | & | & | \end{bmatrix} \quad (7-1)$$



Note que a variável  $n$  ( $n \in \mathbb{N}^*$ ) representa a quantidade de instâncias temporais presentes nos vetores de estado descritos anteriormente e, por conseguinte, o número de linhas da matriz de dados aumentada previamente apresentada. Além disso, a fim de identificar corretamente certos sistemas dinâmicos, pode ser relevante admitir uma base de funções exclusivamente trigonométrica. Pontua-se ainda que a base de funções alicerçada na expansão da série de Fourier pode ser interpretada como um caso particular da base trigonométrica previamente citada. Mais especificamente, almejando obter a matriz de dados aumentada fundamentada em tal expansão, deve-se definir um número natural não nulo conhecido como fator multiplicativo de frequência máximo e utilizar somente as funções seno e cosseno na confecção da referida matriz. Sendo assim, ao considerar um sistema livre de forçamento descrito por meio do estado  $\mathbf{x}_1$  ( $\mathbf{x}_1 \in \mathbb{R}^n$ ) e ao admitir um fator multiplicativo de frequência máximo igual a dez, pode-se gerar a matriz de dados aumentada baseada na expansão de Fourier, como revela a Equação 7-2.

$$\Theta_{10}^{\mathbb{R}} = \begin{bmatrix} | & | & | & | & | & | \\ \text{sen}(\mathbf{x}_1) & \cos(\mathbf{x}_1) & \text{sen}(2\mathbf{x}_1) & \cos(2\mathbf{x}_1) & \cdots & \cos(10\mathbf{x}_1) \\ | & | & | & | & | & | \end{bmatrix} \quad (7-2)$$

Realça-se também que a base de funções suportada pela expansão da série de Fourier utiliza majoritariamente números naturais não nulos como fatores multiplicativos de frequência. Adicionalmente, com a finalidade de gerar modelos não espúrios, pode-se considerar tanto a omissão de algumas características quanto a produção de dados aumentados fundamentados na hibridização de diferentes tipos de bases de funções. Expõe-se ainda que, com o propósito de realizar a tarefa de identificação de sistemas dinâmicos forçados, é necessário admitir o sinal de excitação  $\mathbf{u}$  ( $\mathbf{u} \in \mathbb{R}^n$ ) na matriz de dados aumentada. Assim, ao considerar um sistema forçado descrito por intermédio do estado  $\mathbf{x}_1$  ( $\mathbf{x}_1 \in \mathbb{R}^n$ ) e do estado  $\mathbf{x}_2$  ( $\mathbf{x}_2 \in \mathbb{R}^n$ ), é exequível assumir uma base de funções híbrida, contemplando, por exemplo, uma base polinomial real de grau máximo igual a dois e uma base alicerçada na expansão da série de Fourier com fator multiplicativo de frequência máximo numericamente igual a cinco, como descreve a Equação 7-3.

$$\Theta_{2,5}^{\mathbb{R},\mathbb{F}} = \begin{bmatrix} | & | & | & | & | & | & | & | & | & | \\ \mathbf{1} & \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{u} & \mathbf{x}_1^2 & \mathbf{x}_1\mathbf{x}_2 & \mathbf{x}_2^2 & \text{sen}(\mathbf{x}_1) & \cdots & \cos(5\mathbf{x}_2) \\ | & | & | & | & | & | & | & | & | & | \end{bmatrix} \quad (7-3)$$

## 7.2

**Esquemas de Diferenças Finitas**

É possível constatar que tanto a estrutura original quanto os avanços propostos para referida estrutura assumem abordagens que consistem ou na aquisição das derivadas dos estados a partir da instalação de mais sensores no sistema ou na obtenção de tais derivadas por intermédio de técnicas de aproximação. Entretanto, ao levar em conta quesitos financeiros e técnicos, é usual descartar abordagens alicerçadas na aquisição de novos dados, acarretando na assunção de que tais derivadas são obtidas a partir de aproximações. Detalhadamente, no que diz respeito tanto à metodologia pioneira quanto as suas extensões, destaca-se que as estratégias admitidas visando estimar as derivadas citadas ou são baseadas em um método intitulado Regularização da Variação Total ou são fundamentadas em um esquema denominado Diferenças Centrais. Todavia, objetivando tornar o algoritmo proposto mais rápido e almejando evitar a perda de informação, pode-se admitir três esquemas de diferenças finitas a fim de aproximar as derivadas salientadas.

Nesta tese é considerada a utilização de três esquemas de diferenças finitas com erros de quarta ordem, onde estes são conhecidos como Diferenças Finitas Regressivas (DFR), Diferenças Finitas Centrais (DFC) e Diferenças Finitas Progressivas (DFP). Precisamente, é possível implementar os esquemas DFP, DFC e DFR com o propósito de aproximar as derivadas dos estados para as duas primeiras amostras, para as amostras centrais e, finalmente, para as duas últimas amostras, respectivamente. No entanto, ao considerar aplicações específicas, é natural desejar que tais derivadas sejam disponibilizadas ou em intervalos de tempo menores ou com maiores precisões. Sendo assim, almejando acessar as referidas derivadas mais rapidamente, pode-se empregar os três esquemas supracitados com erros de segunda ordem. De maneira mais formal, como destaca a Tabela 7.1, é possível acessar os coeficientes utilizados com a finalidade de obter integralmente as aproximações das derivadas dos estados com erros de segunda ordem.

Tabela 7.1: Esquemas com erros de segunda ordem.

Defasagem	DFP	DFC	DFR
$x(t - 2\Delta t)$	0	0	$1/2$
$x(t - \Delta t)$	0	$-1/2$	$-2$
$x(t)$	$-3/2$	0	$3/2$
$x(t + \Delta t)$	2	$1/2$	0
$x(t + 2\Delta t)$	$-1/2$	0	0

Assim, ao admitir um sistema detentor de um dado número de estados  $s$  ( $s \in \mathbb{N}^*$ ), onde tais estados  $\mathbf{x}_i$  ( $i = 1, 2, \dots, s$ ) possuem uma certa quantidade de amostras  $n$  ( $n \in \mathbb{N}^*$ ), pode-se obter as derivadas dos estados para a primeira amostra a partir do esquema DFP, como explica a Equação 7-4.

$$\left. \frac{d\mathbf{x}_i(t)}{dt} \right|_{\text{DFP}} \approx \frac{1}{2} \frac{1}{\Delta t} \left[ -3\mathbf{x}_i(t) + 4\mathbf{x}_i(t + \Delta t) - \mathbf{x}_i(t + 2\Delta t) \right] \quad (7-4)$$

Por outro lado, suportada pelo esquema DFC, a Equação 7-5 objetiva prover as estimativas das derivadas dos estados para as amostras centrais.

$$\left. \frac{d\mathbf{x}_i(t)}{dt} \right|_{\text{DFC}} \approx \frac{1}{2} \frac{1}{\Delta t} \left[ -\mathbf{x}_i(t - \Delta t) + \mathbf{x}_i(t + \Delta t) \right] \quad (7-5)$$

Finalmente, por intermédio do esquema DFR, pode-se aproximar as derivadas dos estados para a última amostra com o auxílio da Equação 7-6.

$$\left. \frac{d\mathbf{x}_i(t)}{dt} \right|_{\text{DFR}} \approx \frac{1}{2} \frac{1}{\Delta t} \left[ \mathbf{x}_i(t - 2\Delta t) - 4\mathbf{x}_i(t - \Delta t) + 3\mathbf{x}_i(t) \right] \quad (7-6)$$

Em contraste, ao admitir erros de sexta ordem, pode-se acessar os coeficientes utilizados com o propósito de estimar as derivadas dos estados na íntegra, como explicita a Tabela 7.2.

Tabela 7.2: Esquemas com erros de sexta ordem.

Defasagem	DFP	DFC	DFR
$x(t - 6\Delta t)$	0	0	1/6
$x(t - 5\Delta t)$	0	0	-6/5
$x(t - 4\Delta t)$	0	0	15/4
$x(t - 3\Delta t)$	0	-1/60	-20/3
$x(t - 2\Delta t)$	0	3/20	15/2
$x(t - \Delta t)$	0	-3/4	-6
$x(t)$	-49/20	0	49/20
$x(t + \Delta t)$	6	3/4	0
$x(t + 2\Delta t)$	-15/2	-3/20	0
$x(t + 3\Delta t)$	20/3	1/60	0
$x(t + 4\Delta t)$	-15/4	0	0
$x(t + 5\Delta t)$	6/5	0	0
$x(t + 6\Delta t)$	-1/6	0	0

Enfatiza-se também que os esquemas fornecidos por meio da Tabela 7.2 podem ser explorados em cenários onde uma elevada precisão é requerida e, paralelamente, nenhuma restrição é imposta no que diz respeito ao intervalo de tempo despendido a fim de obter as derivadas dos estados. Especificamente, alicerçado no esquema DFP, é tangível aproximar as derivadas dos estados para as três primeiras amostras, como explicita a Equação 7-7.

$$\begin{aligned} \left. \frac{d\mathbf{x}_i(t)}{dt} \right|_{\text{DFP}} \approx \frac{1}{60} \frac{1}{\Delta t} \Big[ & -147\mathbf{x}_i(t) + 360\mathbf{x}_i(t + \Delta t) + \\ & -450\mathbf{x}_i(t + 2\Delta t) + 400\mathbf{x}_i(t + 3\Delta t) + \\ & -225\mathbf{x}_i(t + 4\Delta t) + 72\mathbf{x}_i(t + 5\Delta t) + \\ & -10\mathbf{x}_i(t + 6\Delta t) \Big] \end{aligned} \quad (7-7)$$

De modo similar, ao considerar o esquema DFC, pode-se estimar as derivadas dos estados para as amostras centrais por intermédio da Equação 7-8.

$$\begin{aligned} \left. \frac{d\mathbf{x}_i(t)}{dt} \right|_{\text{DFC}} \approx \frac{1}{60} \frac{1}{\Delta t} \Big[ & -\mathbf{x}_i(t - 3\Delta t) + 9\mathbf{x}_i(t - 2\Delta t) + \\ & -45\mathbf{x}_i(t - \Delta t) + 45\mathbf{x}_i(t + \Delta t) + \\ & -9\mathbf{x}_i(t + 2\Delta t) + \mathbf{x}_i(t + 3\Delta t) \Big] \end{aligned} \quad (7-8)$$

Por fim, baseado no esquema DFR, é possível aproximar as derivadas dos estados para as três últimas amostras, como reitera a Equação 7-9.

$$\begin{aligned} \left. \frac{d\mathbf{x}_i(t)}{dt} \right|_{\text{DFR}} \approx \frac{1}{60} \frac{1}{\Delta t} \Big[ & 10\mathbf{x}_i(t - 6\Delta t) - 72\mathbf{x}_i(t - 5\Delta t) + \\ & + 225\mathbf{x}_i(t - 4\Delta t) - 400\mathbf{x}_i(t - 3\Delta t) + \\ & + 450\mathbf{x}_i(t - 2\Delta t) - 360\mathbf{x}_i(t - \Delta t) + \\ & + 147\mathbf{x}_i(t) \Big] \end{aligned} \quad (7-9)$$

Dessa forma, ao empregar as Equações 7-4, 7-5 e 7-6, é viável estimar as derivadas dos estados com erros de segunda ordem. De modo alternativo, ao utilizar as Equações 7-7, 7-8 e 7-9, pode-se aproximar as derivadas dos estados com erros de sexta ordem. Genericamente, por meio de combinações lineares de certas defasagens amostrais de um dado sinal, é exequível estimar integralmente a derivada do mesmo. Em particular, igualmente aos estados do sistema, a derivada do sinal de excitação pode ser acessada por intermédio dos esquemas de diferenças finitas previamente descritos.

### 7.3

#### Estimativas de Parâmetros Estatísticos

Categoricamente, alicerçado na estatística, destaca-se que conceitos como parâmetro e estatística estão, respectivamente, ligados à noção de população e amostra. Em particular, pode-se definir parâmetro e estatística como qualquer valor capaz de resumir algum tipo de informação de uma determinada população e de uma dada amostra, nesta ordem. Dito isso, uma vez que os dados utilizados na inferência de sistemas dinâmicos é um pequeno subconjunto do conjunto de dados capaz de contemplar os possíveis cenários empíricos, é tangível extrair estatísticas de tal subconjunto. De modo geral, no que concerne à tarefa de identificação, verifica-se que a exploração de estatísticas é comumente posta em segundo plano, porém, para alguns problemas, é recomendável realizar a exploração citada. Especificamente, a partir de tais estatísticas, pode-se, por exemplo, acessar a real necessidade de efetuar algum tipo de pré-processamento nos dados, assim como fornecer um ponto de partida no que diz respeito à especulação de quais abordagens podem ser adotadas com a finalidade de capturar os padrões escondidos nos dados.

Realça-se ainda que, sob a perspectiva do aprendizado de máquina, a abordagem estatística como um todo deve ser baseada somente nos dados de treino, impedindo, por sua vez, a ocorrência de um problema intitulado viés de espionagem. Adicionalmente, fundamentado na identificação esparsa, pontua-se que a coleção de características admitidas em um dado problema é a matéria-prima utilizada com o propósito de obter estimativas de parâmetros estatísticos, tais como média, desvio padrão, variância, covariância e correlação. Assim, ao considerar um problema detentor de um certo número de características  $f$  ( $f \in \mathbb{N}^*$ ), onde estas são representadas pela variável  $\theta_i$  ( $i = 1, 2, \dots, f$ ) e descritas a partir de uma determinada quantidade de amostras  $n$  ( $n \in \mathbb{N}^*$ ), pode-se calcular a média amostral associada a cada característica existente, como pontua a Equação 7-10.

$$\bar{\theta}_i = \frac{1}{n} \sum_{k=1}^n \theta_i(k) \quad (i = 1, 2, \dots, f) \quad (7-10)$$

Notavelmente, conforme foi explicitado no Capítulo 3, as características compõem as colunas da matriz de dados aumentada  $\Theta$  ( $\Theta \in \mathbb{R}^{n \times f}$ ). Explana-se também que, a fim de acessar a matriz de dados aumentada centralizada  $\Theta^C$  ( $\Theta^C \in \mathbb{R}^{n \times f}$ ), é necessário obter os valores médios das características inicialmente admitidas. Além disso, ressalta-se que a variância amostral de cada característica pode ser computada pela Equação 7-11.

$$\mathcal{V}\{\theta_i\} = \frac{1}{n-1} \sum_{k=1}^n \left[ \theta_i(k) - \bar{\theta}_i \right]^2 \quad (i = 1, 2, \dots, f) \quad (7-11)$$

De modo complementar, pode-se acessar o desvio padrão amostral de cada característica por intermédio da Equação 7-12.

$$\mathcal{Z}\{\theta_i\} = \sqrt{\frac{1}{n-1} \sum_{k=1}^n \left[ \theta_i(k) - \bar{\theta}_i \right]^2} \quad (i = 1, 2, \dots, f) \quad (7-12)$$

Observe que o desvio padrão amostral é numericamente igual a raiz quadrada da variância amostral. De maneira alternativa, suportado pelo produto interno matricial, pode-se obter facilmente a variância amostral e o desvio padrão amostral de uma dada característica ao considerar a centralização da mesma a partir da média. Note que a variância amostral e o desvio padrão amostral são capazes de exprimir somente como uma característica varia individualmente. Por outro lado, almejando acessar como duas características variam em conjunto, pode-se calcular a covariância amostral associada a tais características, como mostra a Equação 7-13.

$$\mathcal{C}\{\theta_i, \theta_j\} = \frac{1}{n-1} \sum_{k=1}^n \left[ \theta_i(k) - \bar{\theta}_i \right] \left[ \theta_j(k) - \bar{\theta}_j \right] \quad (i, j = 1, 2, \dots, f) \quad (7-13)$$

Por inspeção, é possível constatar que a variância amostral é um caso particular da covariância amostral. Em adição, ao considerar o produto interno matricial atrelado à matriz de dados aumentada centralizada  $\mathcal{J}_\theta$  ( $\mathcal{J}_\theta \in \mathbb{R}^{f \times f}$ ), pode-se obter a matriz de covariância amostral vinculada aos dados aumentados  $\mathcal{C}_\theta$  ( $\mathcal{C}_\theta \in \mathbb{R}^{f \times f}$ ). Matematicamente, alicerçado na Equação 7-14, explica-se que os elementos presentes na matriz de covariância amostral representam a covariância amostral associada a cada par de características.

$$\mathcal{C}_\theta = \frac{\mathcal{J}_\theta}{n-1} \quad (7-14)$$

De modo formal, pode-se verificar que a covariância amostral é sensível à mudança de escala, ou seja, é possível demonstrar que a referida estatística não é capaz de acessar de maneira apropriada a força da relação linear efetiva entre duas características. Sendo assim, com a finalidade de contornar tal problema, pode-se admitir uma versão padronizada da covariância amostral denominada correlação ou coeficiente de Pearson, como propõe a Equação 7-15.

$$\mathcal{P}\{\theta_i, \theta_j\} = \frac{\mathcal{C}\{\theta_i, \theta_j\}}{\mathcal{Z}\{\theta_i\}\mathcal{Z}\{\theta_j\}} \quad (i, j = 1, 2, \dots, f) \quad (7-15)$$

Em detalhes, diferentemente da covariância amostral, que não possui limite teórico e, conseqüentemente, pode variar de menos infinito a mais infinito, enfatiza-se que a correlação é limitada inferiormente pelo simétrico da unidade e superiormente pela unidade.

## 7.4

### Otimização

Em linhas gerais, é natural constatar que os algoritmos costumam possuir certos procedimentos de otimização incorporados aos mesmos. Pontualmente, devido à adoção de linguagens *script* de alto nível, pode-se ressaltar que os procedimentos salientados são usualmente tratados como uma caixa preta. Sendo assim, é fundamental investigar de modo mais profundo como os dados são manuseados ao longo de tais procedimentos. Adicionalmente, uma vez que tanto a metodologia original quanto a estratégia proposta lidam recorrentemente com um problema de mínimos quadrados, ressalta-se que uma abordagem conhecida como descida de encosta é adotada com a finalidade de prover a solução para o referido problema. Em particular, enfatiza-se que a abordagem citada é alicerçada no cálculo multivariável e que o sucesso das metodologias consideradas neste material dependem da minimização da função de perda explicitada por meio da Equação 7-16.

$$\mathcal{V} = \underset{\xi}{\operatorname{argmin}} \left\{ \sum_{k=1}^n [\mathbf{y}_i(k) - \hat{\mathbf{y}}_i(k)]^2 \right\} \quad (7-16)$$

Realça-se ainda que a função de perda supracitada é suportada pela diferença entre as derivadas dos estados  $\mathbf{y}_i$  ( $i = 1, 2, \dots, s$ ) e suas estimativas  $\hat{\mathbf{y}}_i$  ( $i = 1, 2, \dots, s$ ). Além disso, alicerçado na regressão esparsa, destaca-se que as estimativas das derivadas dos estados podem ser obtidas por intermédio dos vetores de esparsidade  $\xi_i$  ( $i = 1, 2, \dots, s$ ), onde tais vetores possuem uma quantidade de termos numericamente igual ao número de características existentes  $f$  ( $f \in \mathbb{N}^*$ ). Mais formalmente, é tangível obter as estimativas das derivadas dos estados por meio da Equação 7-17.

$$\hat{\mathbf{y}}_i = \Theta \xi_i \quad (i = 1, 2, \dots, s) \quad (7-17)$$

Desse modo, almejando acessar a solução que minimiza a função de perda, deve-se calcular as derivadas parciais de tal função em relação aos vetores de esparsidade. Em adição, baseado em uma conveniência matemática, admite-se a multiplicação da função de perda por um meio antes de iniciar o cálculo das derivadas citadas, como apresenta a Equação 7-18.

$$\frac{\partial \mathcal{V}}{\partial \boldsymbol{\xi}_i} = \frac{\partial}{\partial \boldsymbol{\xi}_i} \left\{ \frac{1}{2} \sum_{k=1}^n [\boldsymbol{\varepsilon}_i(k)]^2 \right\} = \frac{2}{2} \sum_{k=1}^n \left\{ \boldsymbol{\varepsilon}_i(k) \left[ \frac{\partial}{\partial \boldsymbol{\xi}_i} \boldsymbol{\varepsilon}_i(k) \right] \right\} \quad (7-18)$$

Note que a solução capaz de minimizar a função de perda não é afetada pela multiplicação previamente realizada e que a Equação 7-19 pode ser facilmente obtida a partir das Equações 7-17 e 7-18.

$$\frac{\partial \mathcal{V}}{\partial \boldsymbol{\xi}_i} = \sum_{k=1}^n \left\{ \boldsymbol{\varepsilon}_i(k) \frac{\partial}{\partial \boldsymbol{\xi}_i} \left[ \mathbf{y}_i(k) - (\boldsymbol{\Theta} \boldsymbol{\xi}_i) \Big|_k \right] \right\} \quad (7-19)$$

Precisamente, uma vez que as derivadas dos estados não dependem dos vetores de esparsidade, é possível concluir que as derivadas parciais das derivadas dos estados em relação aos vetores de esparsidade são numericamente iguais a zero, viabilizando, por sua vez, a obtenção da Equação 7-20.

$$\frac{\partial \mathcal{V}}{\partial \boldsymbol{\xi}_i} = \sum_{k=1}^n \left\{ \boldsymbol{\varepsilon}_i(k) \frac{\partial}{\partial \boldsymbol{\xi}_i} \left[ \left( -\boldsymbol{\Theta} \boldsymbol{\xi}_i \right) \Big|_k \right] \right\} \quad (7-20)$$

De maneira lógica, fundamentado no cálculo multivariável, é exequível constatar que a Equação 7-21 pode ser obtida por meio da Equação 7-20.

$$\nabla_i = \frac{\partial \mathcal{V}}{\partial \boldsymbol{\xi}_i} = - \sum_{k=1}^n \left[ \boldsymbol{\varepsilon}_i(k) \boldsymbol{\Theta} \right] \quad (7-21)$$

Observe que as derivadas parciais descritas por intermédio da Equação 7-21 representam uma entidade matemática amplamente conhecida como gradiente  $\nabla_i$  ( $i = 1, 2, \dots, s$ ). Mais especificamente, explica-se que a entidade citada é uma grandeza vetorial geometricamente caracterizada pela sua capacidade de fornecer a direção de máximo aclave. Por outro lado, uma vez que a minimização da função de perda é desejada, pontua-se que a direção de máximo declive deve ser obtida. De maneira formal, ao considerar o vetor oposto ao gradiente, pode-se acessar a direção de máximo declive  $\blacktriangledown_i$  ( $i = 1, 2, \dots, s$ ), como reitera a Equação 7-22.



$$\nabla_i = -\nabla_i = \sum_{k=1}^n [\varepsilon_i(k) \Theta] \quad (7-22)$$

Além disso, com a finalidade de favorecer a convergência da solução, é comum utilizar a versão normalizada do vetor responsável por prover a direção de máximo declive, isto é, recomenda-se a admissão do versor associado ao vetor oposto ao gradiente, como explana a Equação 7-23.

$$\nabla_i = \frac{\nabla_i}{\|\nabla_i\|} \quad (7-23)$$

Assim, ao definir tanto uma determinada taxa de aprendizado  $\mu$  ( $\mu \in \mathbb{R}_+^*$ ) quanto um dado palpite inicial da solução  $\Xi^\bullet$  ( $\Xi^\bullet \in \mathbb{R}^{f \times s}$ ), é possível obter uma solução aproximada  $\Xi^*$  ( $\Xi^* \in \mathbb{R}^{f \times s}$ ), onde esta, por sua vez, minimiza a função de perda. Em detalhes, por intermédio de uma versão escalada do vetor oposto ao gradiente  $\nabla_i^\mu$  ( $i = 1, 2, \dots, s$ ), pode-se avançar de maneira iterativa na direção de mínimo e, eventualmente, obter a solução capaz de minimizar a função de perda, como reitera a Equação 7-24.

$$\nabla_i^\mu = \mu \nabla_i \quad (7-24)$$

Finalmente, ao levar em conta os passos previamente detalhados como um todo, é exequível implementar todos os algoritmos apresentados ao longo do Capítulo 3 desta tese.

## 7.5

### Produto Interno e Externo

Notavelmente, como foi explorado no Capítulo 3, existe uma vasta riqueza teórica e prática atrelada tanto ao produto interno matricial quanto ao produto externo matricial. Em particular, pôde-se apreciar a abundância conceitual vinculada ao conceito de norma vetorial, assim como relacionar decomposições matriciais com a estatística. Desse modo, é relevante investigar de forma detalhada os referidos produtos objetivando obter uma melhor compreensão dos mesmos. Mais especificamente, visa-se explorar tais produtos quando matrizes detentoras de uma única coluna são admitidas. Assim, ao assumir um vetor genérico  $\mathbf{v}$  ( $\mathbf{v} \in \mathbb{R}^n$ ), onde este contém uma certa quantidade de amostras  $n$  ( $n \in \mathbb{N}^*$ ), é possível obter o produto interno matricial associado à matriz coluna considerada, como apresenta a Equação 7-25.

$$\mathcal{J}_{\mathbf{v}} = \mathbf{v}^T \mathbf{v} = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \sum_{k=1}^n v_k^2 \quad (7-25)$$

Observe que o produto interno matricial dado pela Equação 7-25 degenera no habitual produto interno vetorial, mostrando, por sua vez, a relação íntima entre o produto interno e a norma euclidiana. Complementarmente, baseado na análise tensorial, pode-se acessar o produto externo matricial atrelado à matriz coluna previamente descrita, como expõe a Equação 7-26.

$$\mathcal{O}_{\mathbf{v}} = \mathbf{v} \mathbf{v}^T = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} = \begin{bmatrix} v_1 v_1 & v_1 v_2 & \cdots & v_1 v_n \\ v_2 v_1 & v_2 v_2 & \cdots & v_2 v_n \\ \vdots & \vdots & \ddots & \vdots \\ v_n v_1 & v_n v_2 & \cdots & v_n v_n \end{bmatrix} \quad (7-26)$$

Por inspeção, uma vez que os números reais gozam da propriedade comutativa da multiplicação, pode-se constatar que a matriz dada pela Equação 7-26 é simétrica. Salienta-se também que, ao manipular adequadamente vetores e matrizes, é possível extrair a partir do produto interno matricial e do produto externo matricial estimativas de parâmetros estatísticos, tais como média, desvio padrão e variância. Finalmente, contanto que os produtos explorados sejam bem definidos, explica-se que a admissão de matrizes retangulares distintas pode ser considerada.