

# Identificação não linear no espaço de estados de motor-gerador via regressão esparsa

Klarissa Carvalho\* Leilane dos Anjos\* Raissa Paiva\*  
Raphael Teixeira\*

\* *Universidade Federal do Pará,  
Campus Universitário de Tucuruí,  
Faculdade de Engenharia Elétrica*

---

## Resumo:

Este trabalho estuda a aplicação da técnica SINDy na identificação de modelos lineares e não lineares, no espaço de estados, de um sistema motor-gerador. A partir de dados coletados de sinais de entrada e saída do sistema, são gerados termos polinomiais que concorrem na composição da estrutura das equações de estado. O problema de identificação é formulado a partir de uma estratégia de regressão esparsa que busca selecionar os termos mais relevantes para cada equação e os seus parâmetros. A implementação mostra bom desempenho local, em termos da métrica da normalização do erro médio quadrático (NRMSE), dos modelos lineares. São identificados ainda modelos não lineares a partir de sinais que excursionam o sistema em uma ampla faixa de operação. Estes modelos demonstram um desempenho crescente com o grau da não linearidade e o limiar de seleção dos termos polinomiais, atingindo ajuste de até 98,7% para não linearidades de quinta ordem, demonstrando a relevância da estratégia.

---

## 1. INTRODUÇÃO

A identificação de sistemas compreende técnicas baseadas em dados medidos, destacando-se por gerar boas representações dinâmicas lineares e não lineares, para diferentes classes de sistemas, cujas leis físicas são desconhecidas *a priori* (Aguirre, 2015). Destacam modelos baseados em séries de Volterra e com termos regressivos como AR, ARX, NARX e NARMAX (Billings, 2013).

Este trabalho apresenta um estudo de identificação de sistemas para representações no espaço de estados, a identificação esparsa de dinâmicas não lineares (SINDy, do acrônimo em inglês)<sup>1</sup> (Brunton et al., 2016). A técnica parte das medições dos estados e das entradas, estabelece com estes sinais uma biblioteca de termos candidatos para compor as equações de estado e faz uso de uma estratégia de regressão esparsa que busca concomitantemente selecionar os termos mais relevantes e determinar o valor de seus parâmetros.

A importância da identificação SINDy é a sua capacidade de eliminar parte dos termos do modelo candidato, selecionando aqueles de maior representatividade dinâmica, devido à esparsidade. Isto gera um modelo mais compacto, típico de estruturas em espaço de estados. Este resultado advém da composição da função custo que minimiza o erro médio quadrático do modelo penalizando a magnitude dos parâmetros. Isto é vantajoso comparativamente ao método dos mínimos quadrados clássico que identifica todos os parâmetros do modelo candidato, gerando por consequência um modelo com maior complexidade estrutural.

O método ganhou relevância nos últimos anos, tendo recebido importantes reformulações, como a SINDy-PI

(Kaheman et al., 2020), sendo aplicado na modelagem de sistemas relevantes em amplos cenários de engenharia (Abdullah et al., 2022),-(Foster et al., 2022). Um pacote python chamado PySINDy (de Silva et al., 2020) foi desenvolvido visando sistematizar a aplicação computacional do método, sendo aqui utilizado.

Neste trabalho, a SINDy é aplicada à identificação de modelos lineares e não lineares de um sistema experimental de bancada do tipo motor-gerador. Ensaio preliminares explicitam uma não linearidade do tipo ganho variável neste sistema. Dados de entrada e saída deste sistema foram coletados em 3 pontos de operação sendo utilizados na identificação de modelos lineares locais em espaço de estados. A validação mostra que estes modelos são consistentes apenas nas vizinhanças de cada ponto de operação.

Buscando capturar o comportamento não linear, foi realizado um ensaio de identificação para dados de entrada e saída em uma ampla faixa de operação do ganho do sistema. Modelos com estrutura polinomial de grau 2 até 5 apresentaram capacidade crescente de ajuste aos dados de validação, que atingiu até 98,74% medido pela métrica NRMSE. Este resultado mostrou-se dependente também do limiar, um hiperparâmetro do SINDy associado à seleção de termos dos modelos, que estabelece um compromisso entre a complexidade estrutural e sua precisão.

O trabalho é organizado a partir da seção 2, onde o método SINDy é desenvolvido, seção 3 que apresenta a bancada motor-gerador e a metodologia de coleta de dados, a seção 4 que apresenta a identificação de modelos lineares e análise dos resultados, a seção 5 com os modelos não lineares e a seção 6 que apresenta uma conclusão do trabalho.

---

<sup>1</sup> *Sparse Identification of Nonlinear Dynamics*

## 2. ALGORITMO SINDY

Seja um sistema dinâmico de tempo discreto<sup>2</sup> definido por

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k), \quad (1)$$

com  $k$  denotando o múltiplo inteiro do período de amostragem  $T_s$ , para  $m$  entradas  $\mathbf{u}_k = [u_1, u_2, \dots, u_m]^T$  e  $n$  variáveis de estado  $\mathbf{x}_k = [x_1, x_2, \dots, x_n]^T$ , cujas dinâmicas são governadas pelas funções de estado  $\mathbf{f}(\cdot)$ . A estrutura deste sistema de equações é bastante geral, permitindo acomodar a forma canônica linear e uma grande diversidade de comportamentos não lineares.

Visando formular a identificação destas classes de modelos a partir de dados medidos de  $\mathbf{x}_k$  e  $\mathbf{u}_k$ , considera-se que cada função  $\mathbf{f}(\cdot)$  possa ser aproximada pela combinação de  $p$  termos  $\boldsymbol{\theta}(\mathbf{x}_k, \mathbf{u}_k)$  parametrizados pelos coeficientes  $\boldsymbol{\xi} = [\xi_1, \xi_2, \dots, \xi_p]^T$  correspondentes, na forma (2):

$$\mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) \approx \sum_{i=1}^p \xi_i \boldsymbol{\theta}_i(\mathbf{x}_k, \mathbf{u}_k) = \boldsymbol{\Theta}(\mathbf{x}_k, \mathbf{u}_k) \boldsymbol{\xi}, \quad (2)$$

sendo  $\boldsymbol{\Theta}(\mathbf{x}_k, \mathbf{u}_k)$  o conjunto de termos candidatos a compor a estrutura inicial do modelo. Esta biblioteca pode conter uma grande quantidade de termos dadas as possibilidades algébricas e o desconhecimento *a priori* de quais destes devem compor cada função de estado ao final da identificação. Nestes termos, o problema proposto possui o duplo objetivo de determinar valores apropriados e não nulos para o menor número possível de coeficientes em  $\boldsymbol{\xi}$ , em cada equação de estado, de modo a conferir esparsidade, uma propriedade típica desta classe de modelos.

Um modelo de estado pode exibir formas algébricas polinomiais, trigonométricos, exponenciais, racionais, entre vários outros, de modo que a composição de  $\boldsymbol{\Theta}(\mathbf{x}_k, \mathbf{u}_k)$  é bastante diversa. Este trabalho utiliza exclusivamente termos polinomiais, os quais são aproximadores universais de funções que exibem grande versatilidade e acomodam tanto as representações lineares quanto as não lineares. Assim, para compor uma biblioteca de termos polinomiais parte-se de um conjunto de medições de  $\mathbf{X}_k$  e de  $\mathbf{U}_k$  na forma:

$$\mathbf{X}_k = \begin{bmatrix} | & | & \dots & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_n \\ | & | & \dots & | \end{bmatrix} \quad \text{e} \quad \mathbf{U}_k = \begin{bmatrix} | & | & \dots & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_m \\ | & | & \dots & | \end{bmatrix},$$

para então gerar a biblioteca de termos polinomiais

$$\boldsymbol{\Theta}(\mathbf{X}, \mathbf{U}) = [\mathbf{1} \ \mathbf{X} \ \mathbf{U} \ \mathbf{X}\mathbf{U} \ \mathbf{X}^2\mathbf{U} \ \dots \ \mathbf{X}^{d_x}\mathbf{U}^{d_u}]. \quad (3)$$

Nesta estrutura, se se desejar identificar um modelo linear são suficientes termos expressos em  $\mathbf{X}$  e  $\mathbf{U}$ . Já para modelos não lineares deve-se incluir também termos com  $2 \leq d_x, d_u \leq d$ , sendo  $d = d_x + d_u$  o grau máximo do termo polinomial considerado. Note-se que os produtos, como  $\mathbf{X}\mathbf{U}$ , constituem cruzamentos das colunas de  $\mathbf{X}$  por  $\mathbf{U}$ , o que é comumente denotado por  $\mathbf{X} \otimes \mathbf{U}$ , não devendo ser confundido com o produto matricial ordinário.

Em posse de  $\boldsymbol{\Theta}(\mathbf{X}, \mathbf{U})$  pode-se então estabelecer um sistema de equações candidatas a representar as funções de estado. Todas são inicializadas com os mesmos termos, expressos na biblioteca, diferenciando uma das outras

pelos seus coeficientes  $\boldsymbol{\xi} = [\xi_1, \xi_2, \dots, \xi_p]$ . É pertinente agregar então a estrutura do modelo na forma

$$\mathbf{X}_{k+1} = \boldsymbol{\Theta}(\mathbf{X}_k, \mathbf{U}_k) \boldsymbol{\Xi}, \quad (4)$$

onde

$$\boldsymbol{\Xi} = \begin{bmatrix} | & | & \dots & | \\ \boldsymbol{\xi}_1 & \boldsymbol{\xi}_2 & \dots & \boldsymbol{\xi}_n \\ | & | & \dots & | \end{bmatrix} \quad (5)$$

é uma matriz cujas colunas correspondem aos vetores de parâmetros de cada uma das funções buscadas e  $\mathbf{X}_{k+1}$  uma matriz de medição dos estados um passo a frente, visando preservar a relação dinâmica expressa em (1).

A depender dos hiperparâmetros  $n$ ,  $m$  e  $d$ , a quantidade total de termos em  $\boldsymbol{\Xi}$  pode ser bastante numerosa. Neste caso, a busca de realizar a identificação por meio da solução do problema de regressão (4) por minimização quadrática, por exemplo, pode ser contraproducente, na medida em que manterá todos os termos propostos na estrutura inicial do modelo. Como se sabe, modelos de estado costumam exibir representação compacta, de modo que embutir no problema uma penalização de  $\boldsymbol{\Xi}$  buscando uma estrutura final esparsa e mais representativa do sistema pode ser bastante proveitosa.

A estratégia SINDy formula então o seguinte problema de otimização para a determinação de  $\boldsymbol{\Xi}$ :

$$\boldsymbol{\Xi} = \underset{\boldsymbol{\Xi}'}{\operatorname{argmin}} \|\mathbf{X}_{k+1} - \boldsymbol{\Theta}(\mathbf{X}, \mathbf{U}) \boldsymbol{\Xi}'\|_2 + \lambda \|\boldsymbol{\Xi}'\|_1 \quad (6)$$

O problema estabelece uma relação de compromisso entre a minimização do erro médio quadrático, aqui medido pela norma quadrática  $\|\cdot\|_2$  e a magnitude dos parâmetros, que é penalizada pela norma unitária  $\|\cdot\|_1$ . O hiperparâmetro  $\lambda$  regulariza estas duas parcelas.

Busca-se então determinar  $\boldsymbol{\Xi}$  por estratégias de regressão que levem a modelos com poucos termos e que conduzam a um bom ajuste entre os dados. Esta função custo é bem conhecida no contexto de estatística, compondo um problema de regressão conhecido como LASSO<sup>3</sup>.

Uma variação deste problema de otimização para o caso proposto, é o de substituir a determinação de  $\boldsymbol{\Xi}$  pelas suas colunas individualmente, o que é conduzida pelo algoritmo de mínimos quadrados com limite sequencial - STLSQ<sup>4</sup>, para o qual a função custo para cada coluna pode ser então definida como:

$$\boldsymbol{\xi}_k = \underset{\boldsymbol{\xi}_k'}{\operatorname{argmin}} \|\mathbf{x}_k - \boldsymbol{\Theta}(\mathbf{X}, \mathbf{U}) \boldsymbol{\xi}_k'\|_2 + \lambda \|\boldsymbol{\xi}_k'\|_1 \quad (7)$$

Esta estratégia permite resolver o problema sequencialmente incluindo um limiar  $\tau$  como restrição para os termos identificados, de modo que os parâmetros  $\boldsymbol{\xi}_k < \tau$  são eliminados do modelo, sendo esta a estratégia que confere esparsidade ao método. O interesse no seu uso é a melhoria da robustez na determinação dos parâmetros mesmo para dados ruidosos Brunton et al. (2016).

<sup>2</sup> Brunton et al. (2016) apresenta a formulação contínua e autônoma.

<sup>3</sup> *Least Absolute Shrinkage and Selection Operator*

<sup>4</sup> *Sequential thresholded Least-Squares*

---

**Algorithm 1** Algoritmo SINDy

---

```
1: procedure SINDY( $\mathbf{X}_k, \mathbf{U}$ )           ▷ Chama SINDy.
2:    $\Theta \leftarrow (\mathbf{X}_k, \mathbf{U})$          ▷ Cria biblioteca  $\Theta$ .
3:    $\Xi_o \leftarrow \|\mathbf{X}_{k+1} - \Theta(\mathbf{X}, \mathbf{U})\Xi'\|_2$   ▷ Inicializa  $\Xi$  por
   mínimos quadrados.
4:   for  $k \in (1, 2, \dots, n)$  do       ▷ Para cada coluna:
5:      $\xi_{ko} = |\Xi_o[:, k]|$            ▷ Seleciona coluna  $k$ .
6:      $\xi_k = \xi_{ko} > \tau$              ▷ Elimina termos  $< \tau$ .
7:      $\Xi[:, k] = \xi_k$ 
8:   end for
9:   return  $\Xi$                        ▷ Retorna parâmetros.
10: end procedure
```

---

O algoritmo 1 apresenta a estratégia STLSQ para a solução do problema SINDy, sendo esta a abordagem adotada neste trabalho.

### 3. METODOLOGIA

#### 3.1 O sistema motor-gerador

O sistema experimental do trabalho é uma bancada motor-gerador, utilizada em laboratório de ensino de engenharia, a qual possui duas máquinas de corrente contínua acopladas pelo eixo de rotação. A entrada é a tensão de armadura do primeiro motor e a saída a tensão induzida na armadura do segundo motor (gerador). A Figura 1 apresenta a estrutura física da bancada, com os dois motores DC, um drive de potência, uma placa Arduino UNO e um filtro RC que apresenta uma frequência de corte de  $15.91\text{Hz}$  dado um resistor de  $1\text{k}\Omega$  e um capacitor de  $10\mu\text{F}$ .

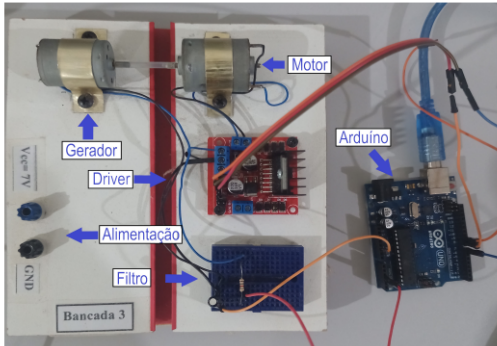


Figura 1. Bancada motor-gerador, com destaque para os subsistemas: driver L298N, filtro RC e placa Arduino UNO.

A bancada é acionada a partir de um programa de computador escrito em linguagem *python*. O programa estabelece uma comunicação serial entre o computador e uma placa Arduino UNO. O *loop* de comunicação consiste em: a placa Arduino recebe o valor do sinal de tensão a ser aplicado nos terminais do Motor (codificado por PWM) e, então ler a tensão gerada nos terminais do gerador. A bancada é energizada por uma fonte DC assimétrica com  $15\text{V}$ . Um driver de potência (L298N) realiza a interface entre a placa arduino e o motor. O sinal de tensão lido nos terminais do gerador são filtrados (Filtro RC), visando atenuar ruídos de medição e então entregues ao conversor AD<sup>5</sup> da placa Arduino UNO.

<sup>5</sup> Conversor analógico-digital

A caracterização do ganho em regime permanente é realizada ao aplicar diferentes valores de tensão DC ( $V_a$ ), para a faixa de operação entre 0 a  $14\text{V}$ . Os valores correspondentes no gerador, em regime permanente, ( $V_\omega$ ), são coletados. A Figura 2 apresenta uma curva  $V_a \times V_\omega$ , para todos os pontos coletados. Nota-se a presença de uma zona morta, típica de motores DC, e um ganho variável do sistema até a região de saturação, características que expressam não linearidades. Com base no comportamento não linear observado, duas frentes para a modelagem do sistema são realizadas, uma baseada em modelos lineares e outra em modelos não lineares, visando avaliar o desempenho da técnica SINDy nestes diferentes contextos.

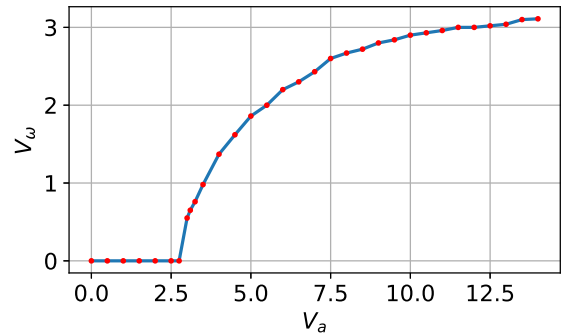


Figura 2. Ensaio aplicado à Bancada motor-gerador de tensão de entrada ( $V_a$ ) versus tensão de saída ( $V_\omega$ ), em Volts.

#### 3.2 Modelagem Linear

Esta etapa busca determinar modelos matemáticos lineares, no espaço de estados, por identificação SINDy nas vizinhanças de diferentes pontos de operação do sistema. Estes pontos de operação referem-se a valores de tensão fixos de entrada e saída na região de ganho variável da bancada motor gerador. Foram utilizados três pontos com tensões de 5, 0, 8, 0 e 12, 5V na entrada. Um sinal do tipo Sequência Binária Pseudoaleatória (*Pseudo-Random Binary Sequence* - PRBS), gerado em Python, foi adicionado a estes valores DC dos pontos de operação visando perturbar o sistema garantindo uma persistência de excitação (Aguirre, 2015). Assim, dados de entrada e saída foram coletados, com um período de amostragem  $T_s = 0,03\text{s}$ , para serem utilizados na identificação dos modelos lineares.

Em posse dos dados, foram realizados procedimentos de pré-processamento visando adequá-los a identificação linear na SINDy. Os procedimentos foram: (i) exclusão de uma parcela das amostras coletadas referentes ao transitório; (ii) retirada do valor médio dos dados de entrada e saída visando eliminar tendência nos modelos; (iii) divisão dos dados residuais em duas parcelas, sendo 60% das amostras destinadas para a identificação e 40% para a validação dos modelos, e (iv) composição das estruturas de dados matriciais de identificação e validação a serem utilizadas na SINDy. Presume-se que os modelos identificados para cada um dos três conjuntos de dados coletados apresentem um ajuste satisfatório apenas nas proximidades dos respectivos pontos de operação.

### 3.3 Modelagem Não Linear

Para esta etapa buscou-se determinar modelos matemáticos não lineares no espaço de estados através da identificação SINDy. As não linearidades avaliadas foram do tipo polinomial, as quais exibem uma grande versatilidade na aproximação de funções. Com isso pressupõe-se que estes modelos não lineares demonstrem maior capacidade de representação da dinâmica do sistema por uma ampla faixa de operação. Para isso houve a necessidade de um sinal de entrada do tipo Sequência Multinível Pseudoaleatória (Pseudo Random Multilevel Sequence-PRMLS), que excita o sistema em diferentes níveis de ganho.

Neste caso, o pré-processamento manteve os dados íntegros particionando-os em 70% para identificação e 30% para a validação do modelos. O ajuste dos hiperparâmetros destes modelos não lineares constitui uma etapa importante da identificação, uma vez que o desempenho está diretamente ligado ao grau das não linearidades e ao limiar responsável por reduzir a complexidade estrutural e promover a esparsidade. De modo geral, modelos de grau e limiar elevados exibem melhor ajuste e estrutura compacta.

### 3.4 A biblioteca pySINDy

Para a implementação computacional da identificação SINDy este trabalho faz uso do pacote python pySINDy (de Silva et al., 2020), que comporta funções de pré-processamento de dados, definição dos modelos, composição da base matricial de termos candidatos, solução do problema de otimização da SINDy e simulação de modelos. O Código (1) apresenta um código python de chamada das funções pysindy com estas funcionalidades.

```

1 import pysindy as ps
2 model = ps.SINDy(t_default = Ts,
3                 discrete_time = True,
4                 optimizer = ps.STLSQ(tau),
5                 feature_library = ps.Polynomial(d))
6 model.fit(X, U, t = Ts)
7 model.print()
8 Xs = model.simulate(Xv0, uv)

```

Código 1. Código Python de identificação PySINDy.

Note-se que o pacote permite declarar ainda o tipo de otimizador, a natureza das não linearidades e o grau (no caso das polinomiais), além do valor do limiar.

## 4. RESULTADOS

### 4.1 Identificação via SINDy de Modelos lineares.

Após a realização dos ensaios para a aquisição de dados, em cada ponto de operação definido para a modelagem linear, estes foram pré-processados e carregados no algoritmo de identificação do pySINDy visando a determinação dos modelos. Nestes procedimentos foram adotados para todos os casos termos de grau 1 e limiar de 0,01, além da utilização do otimizador STLSQ. Os resultados são apresentados nas Equações 8, 9 e 10 que serão denotadas, respectivamente, por  $\mathcal{M}_1$ ,

$$\begin{aligned} x_1[k+1] &= x_2[k] \\ x_2[k+1] &= -0,258x_1[k] + 1,114x_2[k] + 0,064u[k], \end{aligned} \quad (8)$$

$\mathcal{M}_2$

$$\begin{aligned} x_1[k+1] &= x_2[k] \\ x_2[k+1] &= -0,103x_1[k] + 0,847x_2[k] + 0,042u[k] \end{aligned} \quad (9)$$

e  $\mathcal{M}_3$

$$\begin{aligned} x_1[k+1] &= x_2[k] \\ x_2[k+1] &= 0,016x_1[k] + 0,558x_2[k] + 0,024u[k]. \end{aligned} \quad (10)$$

Cada modelo resultou em um sistema de duas equações, uma vez que foram consideradas duas variáveis de estado, com a primeira equação comum, indicando sempre a relação pré-estabelecida entre os estados. A variável  $x_1$  é a própria saída do sistema  $y(k)$  e a variável  $x_2$  a saída atrasada  $y(k-1)$ . Os três modelos apresentaram coeficientes diferentes em cada uma das variáveis, o que implica que cada ponto de operação do sistema possui sua própria representação dinâmica linear.

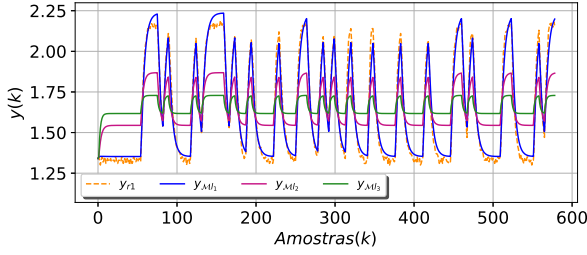
Para avaliar a qualidade destes modelos foram realizadas simulações com os dados de validação. Os resultados gráficos obtidos deste procedimento são apresentados na Figura 3. Ao realizar um processo de validação cruzada, aplicou-se os sinais de entrada de cada ponto de operação, associados às respectivas saídas do sistema real, denotadas por  $y_{r1}$ ,  $y_{r2}$  e  $y_{r3}$ , aos três modelos identificados, obtendo os respectivos sinais de saída, caracterizados no gráfico por  $y_{m1}$ ,  $y_{m2}$  e  $y_{m3}$ . Nota-se que na Figura 3(a), o ajuste de  $y_{r1}$  aos três modelos apresentados é adequado apenas para  $y_{m1}$ , sendo bastante discrepante para  $y_{m2}$  e  $y_{m3}$ . O mesmo ocorre com os outros dois sinais de saída, cujos resultados são apresentados nas Figuras 3(b) e 3(c). Este fato confirma a hipótese de que cada modelo linear será apropriado apenas nas vizinhanças do seu ponto de operação.

Para medir a qualidade destes ajustes expressos nos resultados gráficos foi utilizada a métrica da normalização da raiz do erro médio quadrático (NRMSE), dada por

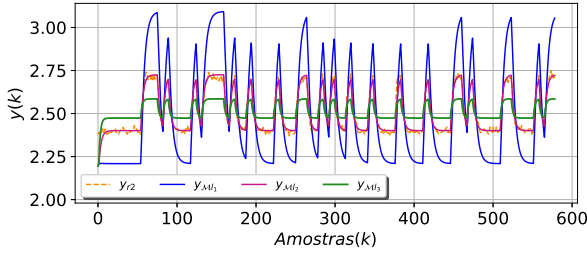
$$J_i = \left( 1 - \sqrt{\frac{\sum (y_{r_i} - \hat{y}_{m_i})^2}{\sum (y_{r_i} - \mu_{r_i})^2}} \right)^{\frac{1}{2}} 100\%, \quad (11)$$

sendo  $J_i$  a medida do ajuste e  $\mu_{r_i}$  a média da saída real do sistema. Esta métrica irá medir 100% quando o ajuste for perfeito e valores tendendo a 0 quanto pior for o ajuste.

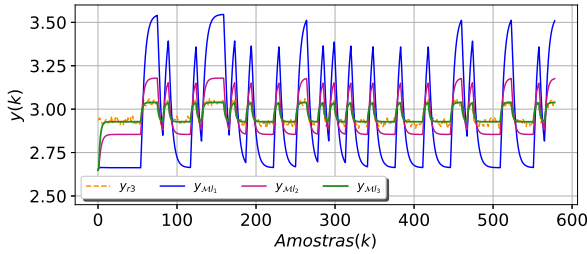
A Figura 4 apresenta o ajuste para cada modelo na forma de uma matriz de confusão, onde se verifica quantitativamente que cada modelo apresenta os melhores ajustes para o seu respectivo conjunto de dados. O ajuste para  $\mathcal{M}_1$ , por exemplo, é de 97,6% para o ponto 1 e de 87,9% para o ponto 3.



(a) Validação de  $Ml_1$ .



(b) Validação de  $Ml_2$ .



(c) Validação de  $Ml_3$ .

Figura 3. Dados de simulação dos modelos lineares visando a validação cruzada.

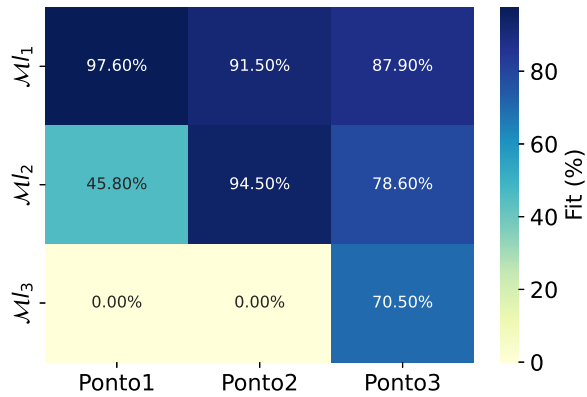


Figura 4. Ajuste cruzado por NRMSE dos dados *versus* modelos.

#### 4.2 Identificação SINDy: Modelos não lineares.

Realizado o pré-processamento dos dados obtidos do ensaio não linear, os mesmos foram então aplicados ao algoritmo de identificação do pySINDy, buscando determinar a modelagem do comportamento não linear do sistema dinâmico. Foram identificados modelos com termos polinomiais de 2°, 3°, 4° e 5° grau, para três diferentes valores de

limiares, 0,0005, 0,005 e 0,02. A ideia é verificar se estes modelos exibem melhores resultados que os lineares, progressivamente em relação tanto ao grau quanto ao limiar.

As Equações 12 e 13, exibem os modelos obtidos com melhor ajuste aos dados de validação, com limiar 0,0005 para grau 2 e 0,02 para grau 3. Estes serão referidos, respectivamente por,  $Mnl_2$  e  $Mnl_3$ .

$$\begin{aligned}
 x_1[k+1] &= x_2[k] \\
 x_2[k+1] &= -0,1791 - 0,802x_1[k] + 1,812x_2[k] \\
 &\quad + 0,079u[k] - 0,041x_1[k] + 0,323x_1[k]x_2[k] \\
 &\quad - 0,015x_1[k]u[k] - 0,317x_2[k]^2 + 0,008x_2[k]u[k] \\
 &\quad - 0,002u[k]^2
 \end{aligned} \quad (12)$$

$$\begin{aligned}
 x_1[k+1] &= x_2[k] \\
 x_2[k+1] &= -0,0381 + 0,390x_1[k] + 0,397x_2[k] \\
 &\quad + 0,038u[k] - 0,430x_1[k]^2 + 0,433x_1[k]x_2[k] \\
 &\quad - 0,155x_1[k]u[k] + 0,056x_2[k]^2 + 0,182x_2[k]u[k] \\
 &\quad + 0,223x_1[k]^3 - 0,418x_1[k]^2x_2[k] + 0,188x_1[k]x_2[k]^2 \\
 &\quad + 0,072x_1[k]x_2[k]u[k] - 0,082x_2[k]^2u[k]
 \end{aligned} \quad (13)$$

Para os dois modelos matemáticos apresentados, observa-se um sistema de duas equações onde a primeira indica a relação pré-estabelecida entre os estados. A presença de termos constantes e de produtos entre as variáveis caracterizam estas equações como não lineares, analisando os dois modelos é notável afirmar que ambos tem uma quantidade de termos diferentes que é influenciado tanto pelo grau quanto o limiar.

As equações dinâmicas de 4° e 5° grau foram omitidas por limitação de espaço. Contudo, o ajuste para  $Mnl_2$  foi de 98,49%,  $Mnl_3$  de 98,59%,  $Mnl_4$  e 98,64%  $Mnl_5$  de 98,74%. É notável que o aumento do grau da não linearidade resulta em melhoria do ajuste, ainda que sutil. Destaca-se ainda, que estes resultados são todos superiores aos melhores modelos lineares, embora aqui o sistema esteja sendo excitado em uma ampla faixa de operação.

A Figura 5 apresenta o resultado de simulação com os dados de validação para  $Mnl_2$  e  $Mnl_3$ , cujas saídas simuladas  $y_{m2}$  e  $y_{m3}$ , foram confrontados com os dados do sistema real  $y_r$ . Note-se ainda que foi identificado um modelo linear para estes dados, cuja saída simulada  $y_{m1}$  é utilizada como base de comparação, tendo contudo um ajuste limitado a 72,85%. As curvas para  $y_{m2}$  e  $y_{m3}$  demonstram a melhoria do ajuste com o aumento do grau da não linearidade, o que é ressaltado visualmente pela porção em destaque do gráfico.

## 5. CONCLUSÃO

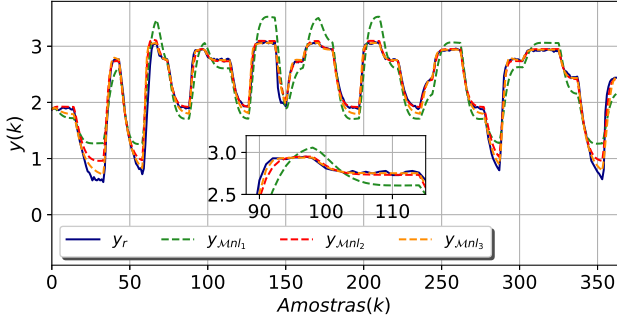


Figura 5. Validação de  $Mnl_2$  e  $Mnl_3$  quanto o grau da não linearidade.

A Figura 6 apresenta o efeito do limiar sobre os modelos identificados. São apresentados dados referentes a três modelos de 5° grau, cujas saídas simuladas são denotadas por  $y_{m51}$ ,  $y_{m52}$  e  $y_{m53}$ , os quais foram identificados com limiares crescentes de 0,0005, 0,005 e 0,02. Os resultados gráficos demonstram uma melhoria no ajuste com o aumento do limiar, o que é expresso no destaque da Figura. Note-se que modelos de limiares maiores implicam em um maior número de termos no modelo identificado. Estes resultados são similares aqueles observados para a variação do limiar na identificação de modelos de graus diferentes.

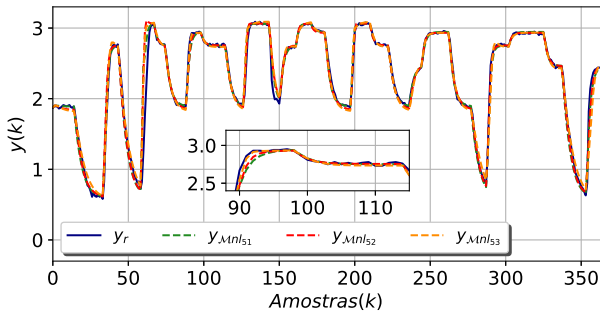


Figura 6. Validação de  $Mnl_5$  quanto ao limiar.

Este fato é apresentado na Figura 7 que confronta o ajuste dos modelos a diferentes valores dos limiares. Modelos de grau 2 mostraram-se bastante sensíveis ao aumento do limiar, enquanto modelos de grau 3, 4 e 5, tiveram resultados similares para os diferentes valores adotados. A diferença de desempenho para estes últimos deve-se mais ao grau da não linearidade.

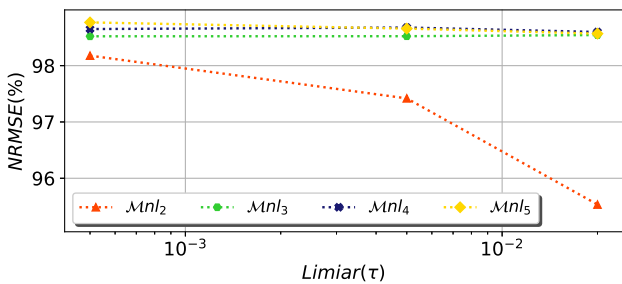


Figura 7. Ajuste por NRMSE (%) de modelos não lineares em função do limiar  $\tau$ .

Este trabalho apresentou o método SINDy para identificação de modelos lineares e não lineares no espaço de estados para um sistema dinâmico real, uma bancada de laboratório de ensino tipo motor-gerador. Ao realizar ensaios que denotaram características não lineares do sistema, dando ênfase para a presença de uma região de ganho variável, foram coletados dados de entrada e saída em determinados pontos de operação desta região visando a identificação de modelos lineares. Os resultados obtidos demonstraram que, estes modelos apresentam um bom desempenho apenas nas vizinhanças dos pontos de operação escolhidos.

Deste modo, buscou-se determinar modelos capazes de capturar o comportamento do sistema de forma mais abrangente, ao identificar modelos não lineares, exitando o sistema em uma ampla faixa de operação. Avaliando os resultados apresentados, ao delimitarmos o SINDy a encontrar modelos de 2° à 5° ordem para diferentes limiares, observou-se que o método demanda de uma relação de compromisso entre a precisão do modelo com os dois hiperparâmetros avaliados. Tipicamente, modelos mais precisos são obtidos a medida em que se aumenta a ordem da não linearidade e o limiar, permitindo que uma quantidade maior de termos possam compor a estrutura final do modelo, fornecendo boas representações que conseguem explicar de forma espaçar sistemas complexos como a bancada motor-gerador.

## REFERÊNCIAS

- Abdullah, F., Alhajeri, M.S., and Christofides, P.D. (2022). Modeling and control of nonlinear processes using sparse identification: Using dropout to handle noisy data. *Industrial & Engineering Chemistry Research*, 61(49), 17976–17992.
- Aguirre, L.A. (2015). *Introdução à Identificação de Sistemas-Técnicas Lineares e Não-Lineares Aplicadas a Sistemas Reais*. Editora UFMG.
- Billings, S.A. (2013). *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. John Wiley & Sons.
- Brunton, S.L., Proctor, J.L., and Kutz, J.N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, 113(15), 3932–3937.
- de Silva, B.M., Champion, K., Quade, M., Loiseau, J.C., Kutz, J.N., and Brunton, S.L. (2020). PySINDy: a python package for the sparse identification of nonlinear dynamics from data. *arXiv preprint arXiv:2004.08424*.
- Foster, J.A., Decuyper, J.R., De Troyer, T., and Runacres, M. (2022). Estimating a sparse nonlinear dynamical model of the flow around an oscillating cylinder: Using sparse identification of nonlinear dynamics (SINDy). In *2022 ISMA International Conference on Noise and Vibration Engineering*.
- Kaheman, K., Kutz, J.N., and Brunton, S.L. (2020). SINDy-PI: a robust algorithm for parallel implicit sparse identification of nonlinear dynamics. *Proceedings of the Royal Society A*, 476(2242), 20200279.