

# Anais do I Encontro Regional de Matemática Aplicada e Computacional do Vale do Itajaí

(Organizadores)

Francis Félix Córdova Puma  
Hugo José Lara Urdaneta

1ª Edição

Blumenau

**UFSC**

2019

**Organização e Publicação:**

Universidade Federal de Santa Catarina

Departamento de Engenharia de Controle, Automação e Computação

Departamento de Matemática

Centro de Blumenau

<http://ermac.blumenau.ufsc.br>

**ISBN: 978-65-80460-31-1**

**Créditos:**

Editor em  $\text{\LaTeX}$ : Luiz-Rafael Santos

usando pacote `confproc` do  $\text{\LaTeX}$ , versão 0.8 (by V. Verfaillie)

# SUMÁRIO – ERMAC-2019

---

## Resumos

- 1 *Gean C. Araujo, Rosângela T. Guedes*  
1-ET-1 Funções Hiperbólicas, Funções Hiperbólicas Inversas e Aplicações
- 3 *Paula C. R. Ertel, Renan G. Romano*  
2-ET-1 Resolução de problema isoperimétrico via séries de Fourier
- 5 *Eduardo O. Belinelli, Paulo Laerte Natti, Eliandro R. Cirilo, Neyva M. L. Romeiro*  
3-ET-2 Geração de Malha para Descrever a Dispersão da Ferrugem da Soja no Paraná
- 7 *Elaine Bandura, Marlon Soares*  
5-ET-2 Um estudo sobre códigos corretores de erros aplicados à codificação genética
- 9 *Fernanda B. Damasceno, Orlando S. Pereira, Cristiane S. S. Pereira*  
6-ET-2 Modelagem matemática para estudo da evolução da Tuberculose no Estado do Rio de Janeiro nos últimos cinco anos
- 11 *Leonardo Veit, Delair Bavaresco*  
7-ET-3 O problema da roda quadrada: uma abordagem a partir da Geometria Dinâmica e impressora 3D
- 13 *João Vitor Pamplona, Luiz-Rafael Santos*  
8-ET-3 Quadrados mágicos: propriedades matemáticas e computacionais
- 15 *Mateus S. Oliveira, Leonardo K. Sacht*  
9-ET-4 Transformações Conformes da esfera para o plano
- 17 *João V. R. de Jesus, Alex L. R. Rese, Rodrigo Lyra*  
11-ET-4 Identificação e classificação de estrelas usando máquinas de vetores de suporte
- 19 *Anderson J. de Oliveira, Alice N. de Oliveira*  
13-ET-7 Aplicação de Códigos Corretores de Erros no processo de geração de sequências de DNA
- 21 *Gabriel Dinse, Leonardo Mejía*  
14-ET-7 Sistema robotizado usando tecnologias de atuação por cabos
- 23 *Felipe P. Gomes, Marciano Pereira*  
16-ET-9 Números de Catalan: apresentação, propriedades e aplicações
- 25 *Leonardo F. Silva, Felipe Fidalgo*  
17-ET-9 Determinante de Cayley-Menger
- 27 *Luis F. Campognara, Felipe D. Costa*  
18-ET-10 Tese de malha para uma simulação em LES
- 29 *Arlindo D. C. Júnior, Rafael Maroneze, Luis F. Campognara, Luiz E. Medeiros, Felipe D. Costa*  
19-ET-10 Estrutura da camada limite estável na região do Pampa Gaúcho e sua representação por modelos numéricos
- 31 *Natália R. Almeida, Naiara C. Ap. dos Santos, Eraldo S. Silva, Vivian C. De castro, Lígia Carreira*  
20-ET-11 Propriedades psicométricas da adaptação transcultural do instrumento de avaliação de atitudes relacionadas ao lazer de Ragheb e Beard
- 33 *Renato Poli Mari, Amanda Schuch, Edna Evaristo, Gabrielle C. Kratz, José Alexandre B. Valle, Rita T. S. Curto Valle, Jully S. P. Filippi*  
22-ET-11 Uso de delineamento composto central para otimização de processo de transferência de calor
- 35 *Thiago F. De Oliveira, Eliandro R. Cirilo, Neyva, M. L. Romeiro, Paulo L. Natti*  
24-ET-12 Análise numérica do método quase linear que modela a equação telegráfica reativa difusiva
- 37 *Márcio R. Fernandes*  
25-ET-12 Simulações numéricas com um Problema Não-Linear de Advecção-Difusão

- 39 *Aldair Forster, Antonio R. R. Bendô, Juliana Costi, Eduardo de P. Kirinus, William C. Marques, Thaísa B. Trombetta*  
26-ET-12 Circulação dirigida pelo vento em uma lagoa de águas rasas: a Lagoa Mirim, Rio Grande do Sul, Brasil
- 41 *Bryan A. Siqueira, Louise Reips*  
27-ET-12 Discretização de Equações Diferenciais Parabólicas
- 43 *Jamile S. dos Santos, Daniel G. A. Vigo*  
28-ET-13 Implementação em Python do modelo SEIR estocástico aplicado à propagação de informação
- 45 *Luiz A. S. França, Louise Reips*  
31-ET-13 Estudo da evolução de tumores sólidos através de Equações Diferenciais
- 47 *Lisiane Reips, Roger L. H. Lavarda*  
34-ET-14 Demonstração do Problema do Caixeiro Viajante utilizando Robótica e Algoritmos Genéticos
- 49 *Guilherme R. Wehmuth, Hugo J. L. Urdaneta, Tiago D. C. Busarello*  
35-ET-14 Um modelo de Fluxo de Potência Ótimo para agrupamento virtual de geradores distribuídos
- 51 *Guilherme Philippi, Felipe Fidalgo*  
36-ET-14 DMDGP: um problema real

---

## Trabalhos Completos

- 53 *Marcos Pinheiro de Lima, Leslie D. Pérez-Fernández, Júlio C. R. Claeyssen*  
01-ET-1 Homogeneização da Equação do Calor para meios continuamente microperiódicos via Método de Funções Teste Oscilantes.
- 61 *Flávia T. Giordani*  
02-ET-2 Existência de acoplamentos espaciais favorecendo sincronização em um modelo metapopulacional.
- 69 *Edgar Lacerda de Aguiar, Gustavo H. M. Mendonça, Sandro R. Dias, Thiago S. Rodrigues*  
03-ET-2 Criação de bactérias mutantes in silico de *S. agalactiae* híbridas entre *Oreochromis niloticus* e *Homo Sapiens* por meio de Algoritmos Genéticos e Lógica Fuzzy.
- 75 *Ana Carla F. N. Gomes, Adriano de Cezaro*  
04-ET-2 Modelo SIRC Fracionário: comparação com dados reais.
- 83 *João Inácio M. Bezerra, Alexandre Molter*  
05-ET-3 Implementação de um programa computacional para o diagrama de bifurcações de sistemas caóticos.
- 91 *André Tragancin, Gilmário Santos*  
06-ET-4 Uma proposta baseada em Visão Computacional na Determinação de Pontos de Referência no perímetro de vôo para drones.
- 97 *Michele Valentino, Luiz Fernando C. Alberto, Yuri Cândido da S. Ribeiro*  
07-ET-5 Uma extensão do princípio de invariância para a classe de sistemas positivos com parâmetros incertos.
- 103 *Jean Carlo Moraes, Cristina Zaniol*  
09-ET-6 Previsão da Inflação: uma análise a partir de Redes Neurais.
- 109 *Janine Garcia, Jayne Garcia, Elisa Henning, Ana M. Hackenberg*  
10-ET-7 Definição de um modelo representativo das habitações de interesse social de Joinville/SC para simulação computacional.
- 117 *Janine Garcia, Elisa Henning, Andreza Kalbusch, Jayne Garcia*  
11-ET-7 Verificação da sazonalidade no consumo de água em residências unifamiliares de Joinville.
- 125 *Thiago M. Fernandes, João E. Abdalla Filho*  
13-ET-7 Uma análise dos métodos semi-empíricos de cálculo de pressão de falha de dutos corroídos.
- 133 *Paul Krause, Joseph Tribbia*  
15-ET-7 Controle do caos em sistemas dinâmicos: aplicação à inicialização e previsão de nuvens em um modelo climático.

- 139 *Julius M. de Barros Filho, Fernanda L. Sá Ferreira, Moisés A. S. M. de Araújo*  
16-ET-8 Uma proposta matemática e computacional para o desenvolvimento da Metodologia da Matriz Densidade tight-binding ordem-N.
- 147 *Fernanda L. Sá Ferreira, Julius M. de Barros Filho, Moisés A. S. M. de Araújo*  
17-ET-8 Sobre funções reais de varável matricial: tratamento no cálculo de estrutura eletrônica e algoritmos de busca.
- 155 *Cláudia J. F. Gonçalves, Emerson L. M. Carmelo, Irene N. Nakaoka*  
18-ET-9 Números de Ramsey em grafos multipartidos para 4-ciclos e estrelas.
- 163 *Tiago A. S. Boza, Hercules A. Feitosa*  
19-ET-9 Semântica relacional para a lógica proposicional do plausível.
- 171 *Felipe O. Teixeira, Solange R. dos Santos*  
20-ET-11 Estatística não-paramétrica aplicada na análise dos fitoterápicos Meratrim e Morosil.
- 177 *Henrique Favre, Anselmo Chaves Neto*  
21-ET-11 Aplicação do delineamento experimental DSD (Definitive Screening Design) em Otimização de produto industrial e propostas para a seleção dos termos do modelo.
- 185 *Sheila Regina Oro, Luana Engeler Felippi*  
22-ET-11 Modelagem do IMCRB utilizando duas metodologias para extração de fatores.
- 193 *Rita T. S. Curto Valle, Luis A. Lourenço, Lucile C. Peruzzo, Cintia K. Imme, José Alexandre B. Valle*  
23-ET-11 Otimização de parâmetros de processo de extração de antocianinas utilizando metodologia de superfície de resposta.
- 201 *William Soares Girão, Luiz Satoru Ochi*  
24-ET-11 Algoritmo Formação Social de Opinião.
- 209 *Arthur Scolari Fagundes, Rodrigo Rossi, Eduardo Varriale da Silva*  
26-ET-12 Método Implícito de Diferenças Finitas aplicado ao Processo de Conformação Eletromagnética.
- 217 *Aldair Forster, Juliana Costi, William Correa Marques, Antonio R. R. Bendô*  
27-ET-12 Modelagem Hidrodinâmica Bidimensional aplicada a inundações: mapeamento e estudo de padrões no escoamento.
- 225 *Luciane Inês A. Schuh, Igor Mozolevski, Márcio A. Murad*  
28-ET-12 Método de Galerkin descontínuo aplicado a problemas de escoamento em meios porosos com fraturas.
- 233 *Everton Boos, Fermín S. V. Bazán*  
29-ET-12 Método iterativo baseado na iteração de Schultz para problemas discretos mal postos.
- 241 *Edgar L. Aguiar, Giselle C. Falcão*  
30-ET-13 Previsibilidade pluviométrica: um estudo matemático analítico e estatístico sobre o volume das precipitações de São Paulo.
- 247 *Stefânia S. Glaeser, Adriano de Cezaro*  
31-ET-13 Um estudo sobre a sincronização do ritmo circadiano: Modelo PIM.
- 255 *Jaime Conte Zandona, Claudio Z. Petersen, Marcelo Schramm*  
32-ET-13 Solução das Equações da Cinética Pontual de nêutrons pelo Método de Fator de Integração Implícito.
- 263 *Leticia M. Kaufman, Guilherme J. Weymar, Régis S. de Quadros*  
33-ET-13 Modelagem das concentrações das espécies químicas existentes na combustão do Metanol via mecanismo reduzido.
- 271 *Bruno Cunha Bartz, Marcelo Schramm, Cláudio Zen Petersen*  
35-ET-13 Aplicação do método de Diferenças Finitas na equação da difusão de nêutrons unidimensional.
- 277 *Marcelo S. T. A. Lourenço, Lucas S. Rocha, Marcelo Schramm, Cláudio Z. Petersen*  
36-ET-13 Solução da Cinética Pontual de Nêutrons via Equação In-Hour Particularizada.
- 283 *Roberto M. S. Décio Júnior, Leslie D. Pérez-Fernández, Julian Bravo-Castillero*  
37-ET-13 Avaliação do comportamento efetivo de compostos microperiódicos não-lineares, com contato imperfeito entre as fases, através do Método de Homogeneização Assintótica: uma introdução com o caso unidimensional.

- 291 *Antonio Carlos Valdiero, Roberta Goergen, Marcia R. M. H. Porsch, Luiz Antonio Rasia, Maurício Oberdörfer, Jocarly P. de Souza*  
38-ET-13 Modelagem Matemática de uma bancada robotizada com acionamento pneumático para reabilitação física.
- 299 *Rita de Cassia S. C. Valle, Regilene S. Silva, Miriam Albara, Cintia Marangoni*  
39-ET-13 Modelagem estatística do processo de destilação de membranas aplicado à recuperação de águas residuais da Indústria Têxtil.
- 307 *José Luiz F. Alves, Valdemar F. S. Filho, Ricardo A. F. Machado, Cintia Marangoni*  
40-ET-13 Modelagem empírica de uma unidade piloto de destilação de película descendente assistida por termossifão: determinação de funções de transferência.
- 315 *Anderson E. Schvertner, Emerson V. Castelani, Francisco N. C. Sobral*  
41-ET-14 O Método de Levenberg-Marquardt para Problemas de Otimização de Menor Valor Ordenado.
- 323 *Rafaela Filippozzi, Douglas S. Gonçalves, Luiz-Rafael Santos*  
43-ET-14 Um algoritmo geométrico para o Problema de Inclusão no Envoltório Convexo.

## Funções Hiperbólicas, Funções Hiperbólicas Inversas e Aplicações

**Gean C. Araujo**

Universidade Tecnológica Federal do Paraná  
86300-000, Cornélio Procópio, PR  
E-mail: geanaraujo@alunos.utfpr.edu.br

**Rosângela T. Guedes**

Universidade Tecnológica Federal do Paraná - Departamento de Matemática  
86300-000, Campus Cornélio Procópio, Cornélio Procópio, PR  
E-mail: rtguedes@utfpr.edu.br

**Resumo:** O objetivo desta pesquisa é aplicações de funções hiperbólicas. Para isso, os métodos utilizados nesse trabalho foram a utilização de software gráfico para estudo das funções mencionadas acima aliada aos conceitos teóricos matemáticos de funções, cálculo e equações diferenciais ordinárias. Uma das principais conclusões deste trabalho foram como essas funções podem ser aplicadas em situações reais os quais destacamos a equação da tractriz e a catenária.

**Palavras-chave:** Aplicações. Função Hiperbólica. Função Hiperbólica Inversa.

### Introdução

Neste trabalho, será desenvolvido alguns resultados que envolvem funções hiperbólicas e funções hiperbólicas inversas para resolver integrais, cujo integrando possui expressões do tipo  $\sqrt{a^2 - u^2}$ ,  $\sqrt{a^2 + u^2}$  e  $\sqrt{u^2 - a^2}$  em que  $a > 0$  e uma aplicação. Outra aplicação importante das funções hiperbólicas é a catenária que consiste em uma curva descrita pelo cosseno hiperbólico. Portanto, essas funções aumentaram o interesse da sociedade em construir e aprimorar construções e desenvolver telégrafos, ponte entre outros.

### Aplicações: Equação da Tractriz e Catenária

Dentre os diversos tipos de métodos de integração, o que merece mais atenção nesse trabalho é a integração por substituição trigonométrica. Ela consiste em substituir o integrando de uma função algébrica por uma função trigonométrica que pode facilitar as contas. Desta forma, o resultado dessa integral envolve essas funções trigonométricas e com as devidas substituições adequadas  $u = a\sin(\theta)$ ,  $u = a\cos(\theta)$  e  $u = a\sec(\theta)$ , respectivamente. Porém, apenas por diferenciação da função  $y = \sinh^{-1}\left(\frac{x}{a}\right)$  e integrando o resultado obtemos  $\int \frac{1}{\sqrt{x^2+a^2}} dx$ . Uma aplicação interessante desse tema é a solução da equação da tractriz, cuja curva é definida de um objeto arrastado ao longo de um plano horizontal por um fio de comprimento constante quando a outra extremidade do fio se move ao longo de uma reta no plano. Na figura 1, pode-se observar essa curva.

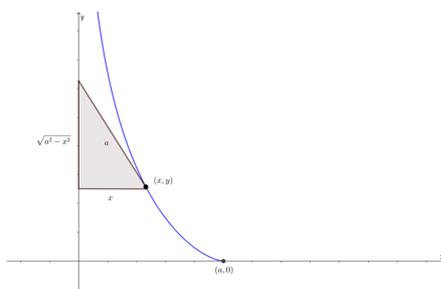


Figura 1: Equação da tractriz

Consideramos o seguinte problema:

- Em todo ponto P da tractriz, a reta  $\overline{AP}$  é tangente em P

- Distância de  $A = (x_a, 0)$  até  $P = (x, y)$  é constante igual a  $a$
- A cúspide  $P_0 = (0, a)$  está na tratriz

Seja  $f$  a função tratriz e  $y = f(x), \forall x \in \mathbb{R}$  e a reta que passa pelo os pontos  $A = (x_a, 0)$  até  $P = (x, y)$  é dada por  $f'(x)(x - x_0)$ . Assim, pelo Teorema de Pitágoras, podemos mostrar que a função  $f$  será solução do problema do valor inicial:

$$\begin{cases} y' = \pm \frac{y}{\sqrt{a^2 - y^2}} \\ y(0) = a \end{cases}$$

Cuja solução  $x \pm a[\varpi - \tanh(\varpi)]$  e  $y = a \operatorname{sech}(\varpi)$ . Além disso, outra aplicação importante das funções hiperbólicas é a catenária. Ela consiste em uma família de curvas planas semelhante às que seriam geradas por uma corda suspensa pela as suas extremidades e sujeitas à ação da gravidade. A curva procurada da catenária será obtida pela solução de uma equação diferencial ordinária de segunda ordem não linear dada por:

$$f'' = a\sqrt{1 + (f'(x))^2}$$

Com condições iniciais  $f(0) = a$  e  $f'(0) = 0$ , pois o elo suspenso foi posicionado na origem. Portanto, a equação da forma da catenária é dada pela função hiperbólica e a sua equivalente exponencial:

$$y = \frac{e^{ax} + e^{-ax}}{2} = \frac{1}{a}(\operatorname{cosh}(ax) - 1)$$

Além disso, uma força aplicada em um ponto qualquer da curva a divide igualmente por todo material. Por isso é usada para a fabricação de materiais como o fundo das latas de refrigerante, iglus e túneis. Outro aspecto importante que costuma causar confusão é a de que a catenária é uma parábola. De fato, ela não é, pois, a parábola é definida por uma função algébrica, enquanto que a catenária é transcendente, sua equação é obtida a partir da função cosseno hiperbólico. E também, a catenária é bem comum do ponto de vista da natureza, visto que pode ser encontrada nas extremidade do ovo, na rede elétrica de uma via férrea, nos fios de alta tensão, nas cordas suspensas de um anúncio de um supermercado, nos separadores de filas de bancos, nas barracas de camping e em pontes, neste caso, a catenária invertida estabelece equilíbrio e modernidade do ponto de vista arquitetônico (VASCONCELOS, 2013).

## Conclusão

- A função cosseno hiperbólico é conhecida pela aplicação da catenária onde é muito observada em pontes, linhas de transmissão, e também, no formato de ovos;
- Observamos na aplicação da equação da tratriz e da catenária a importância como solução das funções hiperbólicas

## Referências

- [1] LEITHOLD, Louis. **O Cálculo com Geometria Analítica**. 3. ed. São Paulo: HARBRA LTDA., 1994.
- [2] STEWART, James; ROMO, Jorge Humberto. **Cálculo**. Cengage Learning, 2008.
- [3] VASCONCELOS, Jerry Gleison Sagueiro Fidanza. **Funções Hiperbólicas: História, Conceito e Aplicação**. Dissertação, Universidade Federal do Amazonas, Manaus, 2013.

## Resolução do Problema Isoperimétrico via Séries de Fourier

**Paula C. R. Ertel**      **Renan G. Romano**

Departamento de Matemática, UFSC - Blumenau

89036-004, Blumenau, SC

E-mail: paulacristina.ertel@hotmail.com,

r.g.romano@ufsc.br

### RESUMO

28 de abril de 2019

O objetivo deste trabalho é apresentar uma demonstração da desigualdade isoperimétrica no plano usando a teoria das séries de Fourier, também conhecida como análise harmônica clássica. Essa teoria tem suas origens no trabalho *Theorie mathématique de la chaleur* desenvolvida pelo matemático e físico francês Jean-Baptiste Joseph Fourier em 1811 e sua aplicação ao problema isoperimétrico segue o trabalho de Hurwitz como proposto em [1].

Seja  $\gamma : [a, b] \rightarrow \mathbb{R}^2$  uma curva fechada e simples. Dizemos que a curva  $C = \{\gamma(t) : t \in [a, b]\}$  é *retificável* quando existe uma constante  $K > 0$  tal que  $\sum_{i=1}^n |\gamma(t_i) - \gamma(t_{i-1})| \leq K$  para toda partição  $a = t_0 < t_1 < \dots < t_n = b$  do intervalo  $[a, b]$ . Para curvas retificáveis, vale a chamada *desigualdade isoperimétrica*:

$$A \leq \frac{L^2}{4\pi}, \quad (1)$$

onde  $A$  é a área delimitada pela curva  $C$  e  $L$  seu perímetro. O *problema isoperimétrico*, ou *problema de Dido*, consiste em encontrar a curva fechada, simples e retificável que delimita a maior área possível com comprimento fixo. Em tal curva, no caso o círculo em  $\mathbb{R}^2$ , temos a igualdade em (1).

Para abordar o problema isoperimétrico usando as séries de Fourier iremos admitir algumas hipóteses sobre uma parametrização da curva  $C$ , as quais garantem a aplicabilidade da teoria de Fourier nesta demonstração. Neste caso, usando a identidade de Parseval, podemos concluir que

$$L^2 - 4\pi A = 2\pi^2 \sum_{n=1}^{\infty} [(na_n - \beta_n)^2 + (nb_n - \alpha_n)^2 + (n^2 - 1)(\alpha_n^2 + \beta_n^2)], \quad (2)$$

em que os coeficientes  $a_n$ ,  $b_n$ ,  $\alpha_n$  e  $\beta_n$  decorrem do Teorema de Fourier aplicado às funções  $x$  e  $y$  tais que  $\gamma(t) = (x(t), y(t))$ . A partir disso, basta notar que o termo do lado direito da igualdade (2) é não-negativo, o que demonstra a desigualdade isoperimétrica. Ainda, supondo a igualdade em (1), podemos obter informações sobre os coeficientes de Fourier  $a_n$ ,  $b_n$ ,  $\alpha_n$  e  $\beta_n$  e concluir que  $\gamma$  é a parametrização de um círculo com raio  $r = \sqrt{A/\pi}$ .

**Palavras-chave:** *Problema Isoperimétrico, Séries de Fourier, Identidade de Parseval.*

### Referências

- [1] D. G. Figueiredo, "Análise de Fourier e Equações Diferenciais Parciais", 4 ed. Projeto Euclides, IMPA, Rio de Janeiro, 2014.
- [2] R. Iório Jr, V. de M. Iório, "Equações Diferenciais Parciais: uma Introdução", 3 ed. Projeto Euclides, IMPA, Rio de Janeiro, 2013.



## Geração de Malha para Descrever a Dispersão da Ferrugem da Soja no Paraná

Paulo Laerte Natti      Eduardo Oliveira Belinelli \*

Departamento de Matemática, CCE, UEL,

86057-970, Londrina, PR

E-mail: plnatti@uel.br    edubelinelli@hotmail.com

Eliandro Rodrigues Cirilo      Neyva Maria Lopes Romeiro

Universidade Estadual de Londrina - Departamento de Matemática

86057-970, Londrina, PR

E-mail: ercirilo@uel.br    nromeiro@uel.br

### RESUMO

A ferrugem asiática é uma doença fungicida, dispersada pelo ar, que afeta todos os anos a produtividade da soja em escala mundial. Compreender como os fatores atmosféricos influenciam na dispersão da ferrugem asiática pelo ar, é de fundamental importância para avaliar os impactos causados nas plantações de soja e prever medidas de controle e manejo da doença [1,3]. Nesse contexto, este trabalho tem por objetivo gerar uma malha retangular para simular o transporte atmosférico da ferrugem asiática no estado do Paraná, realizado por meio de uma equação dispersiva-convectiva-reativa.

O modelo matemático utilizado para descrever o transporte atmosférico do fungo da ferrugem asiática, é uma equação diferencial parcial (EDP) bidimensional com termos difusivo, convectivo e reativo, ou seja,

$$D \left( \frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} \right) - \rho \left( \frac{\partial C \vec{v}}{\partial x} + \frac{\partial C \vec{v}}{\partial y} \right) - \frac{\partial C}{\partial t} = -\lambda C \quad (1)$$

$$C(b_1, 0) = C(b_2, 0) = C(T, 0) = 0 \quad (2)$$

$$C(b_1, t) = C_0 \quad (3)$$

$$\frac{\partial C}{\partial n}(b_2, t) = 0, \quad (4)$$

onde  $D$  denota o coeficiente de difusividade molecular do fungo da ferrugem suspenso no ar,  $C(x, y, t)$  é a concentração de fungos,  $\rho$  é a massa específica do fluido,  $\vec{v}(x, y, t)$  é o campo de velocidades do ar e  $\lambda$  é a taxa de mortalidade do fungo da ferrugem durante seu transporte atmosférico.

A EDP (1) descreve o transporte do fungo da ferrugem pelo ar. A equação (2) são as condições iniciais para a concentração de fungos em toda a malha (veja a Figura 1), isto é, quando  $t = 0$  não há ferrugem no estado do Paraná. Considerando que para  $t > 0$  existe uma injeção contínua  $C_0$  de fungos que entram pelas células  $b_1$ , fronteira Paraná-Paraguai, devido às correntes atmosféricas, a condição de contorno para essa fronteira é dada pela equação (3). Nas demais células  $b_2$  da fronteira aplicam-se condições de contorno tipo Neumann, equação (4).

---

\*Mestrando em Matemática Aplicada e Computacional - PGMAC/UUEL.

O processo de geração de malhas visa a decomposição de um determinado domínio geométrico em vários elementos, de modo a permitir o uso de um método numérico para aproximar a solução das equações governantes do problema em questão [2]. O domínio geométrico do problema em estudo (estado do Paraná) é ilustrado pela Figura 1. A malha em cinza divide o domínio em células. Os rótulos nas células são tais que:  $F$  são células vazias que não contém nenhuma informação, mas pertencem ao domínio do problema e nas simulações não são resolvidas qualquer equação sobre elas.  $T$  são células interiores (ao estado do Paraná), onde o transporte de fungos pelo ar é descrito pelas equações (1-4). Nas células  $b_1$  e  $b_2$  são aplicadas as condições de contorno sobre elas (fronteira do estado do Paraná).

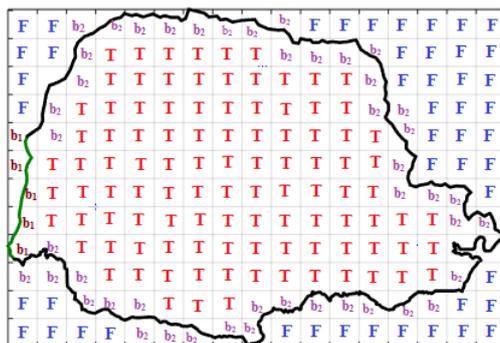


Figura 1: Domínio geométrico do problema em estudo.

Para construir a borda do estado do Paraná foi utilizada a interpolação polinomial spline cúbica. Em seguida, discretizando o modelo (1-4) por Diferenças Finitas, vamos simular a dispersão atmosférica do fungo da ferrugem pelo ar. Os resultados serão comparados com registros de ferrugem no Paraná.

**Palavras-chave:** *Ferrugem asiática, geração de malhas, diferenças finitas.*

### Agradecimentos

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

### Referências

- [1] S. A. Isard, S. H. Gage, P. Comtois and J. M. Russo, Principles of the atmospheric pathway for invasive species applied to soybean rust, *BioScience.*, 55 (2005) 851-861.
- [2] L. G. P. Machado, “Malhas adaptativas em domínios definidos por fronteiras curvas”, Dissertação de Mestrado, USP, São Carlos, 2008.
- [3] C. A. Minchio, M. G. Canteri, L. H. Fantin e M. A. de A. Silva, Epidemias de ferrugem asiática no Rio Grande do Sul explicadas pelo fenômeno ENOS e pela incidência da doença na entressafra, *Summa Phytopathologica.*, 42 (2016) 321-326.

## Um Estudo Sobre Códigos Corretores de Erros Aplicados à Codificação Genética

**Elaine Bandura**      **Marlon Soares**

Depto de Matemática, Universidade Estadual do Centro-Oeste (UNICENTRO)

85040-167, Guarapuava, PR

E-mail: elaine.bandura@gmail.com,      marlonsoares@unicentro.br

### **RESUMO**

Propostos inicialmente por R. W. Hamming, no final da década de 1940, os códigos corretores de erros são atualmente utilizados tanto em atividades rotineiras, como a comunicação via celular, quanto em atividades inusitadas, como o envio de imagens de outros planetas por sondas espaciais.

Embora a transmissão de informação seja, em geral, considerada na comunicação digital, que se dá no espaço, ou seja, de um lugar para outro, ela também pode ser considerada no contexto da genética. Neste contexto, a transmissão de informação genética se dá no tempo, ou seja, de um momento para outro, sendo o ácido desoxirribonucléico (DNA) o canal de comunicação utilizado. Dessa forma, é natural perguntar se a estrutura teórica dos códigos corretores de erros pode contribuir para a compreensão dos processos biológicos envolvidos na codificação genética. A resposta, segundo G. Battail, em [1], é afirmativa.

Iniciadas na década de 1970, as investigações envolvendo métodos da teoria da informação na genética não obtiveram sucesso. Passadas cerca de três décadas, com a conclusão do projeto genoma, que demandou cerca de 13 anos e a contribuição de mais de 5000 cientistas de 17 países, houve um novo interesse no assunto. Tal interesse continua até o presente momento, apesar da quantidade reduzida de pesquisadores tratando do assunto.

Segundo Márcio de Castro Silva Filho, do Departamento de Genética da Escola Superior de Agricultura Luiz de Queiroz (Esalq), da USP, em entrevista à Revista de Pesquisa FAPESP, em setembro de 2015: "O assunto é extremamente complexo, poucas pessoas conseguem transitar nas duas áreas, da genética e dos códigos corretores de erros...".

De fato, por si só, a teoria matemática envolvida nos códigos corretores de erros abrange, além de várias subáreas da Álgebra, como Teoria dos Números, Teoria dos Anéis, Teoria dos Corpos e Teoria dos Grupos, tópicos de Álgebra Linear e de Espaços Métricos. Além disso, o completo entendimento do assunto exige que sejam conhecidos conceitos da Teoria da Informação e de Genética.

Dentre as pesquisas investigando o uso de códigos corretores de erros na codificação genética, é possível encontrar estudos envolvendo códigos lineares auto duais sobre anéis em [4] e sobre corpos em [6]. Além disso, é possível encontrar estudos envolvendo códigos lineares cíclicos sobre anéis em [5] e [2]. Em particular, é possível encontrar estudos envolvendo códigos BCH tanto sobre anéis quanto sobre corpos em [3].

Apesar de fazer uso de uma teoria matemática bastante ampla e que inclui tópicos de nível mais avançado, os artigos que tratam do assunto não contemplam tal teoria. Diante disso sentiu-se a necessidade de realizar uma pesquisa exploratória, de natureza predominantemente teórica, para investigar os detalhes da teoria matemática envolvida nas aplicações dos códigos corretores de erros à codificação genética e algumas dessas aplicações.

O desenvolvimento dessa pesquisa, que encontra-se em andamento, inclui: estudar os princípios básicos da Teoria da Informação e da Genética, bem como os principais tópicos de Álgebra Linear, dos Espaços Métricos, da Teoria dos Números, da Teoria dos Anéis, da Teoria dos Corpos e da Teoria dos Grupos; realizar um levantamento sobre o estado da arte da publicação científica envolvendo o tema da pesquisa; a partir desse levantamento, identificar e validar a teoria matemática dos códigos corretores de erros necessária à sua aplicação na codificação genética; estudar aplicações dos códigos corretores de erros na codificação genética.

**Palavras-chave:** *Correção de Erros, DNA, Teoria da Informação*

## Referências

- [1] G. Battail, “An outline of informational genetics”. Morgan e Claypool Publishers, 2008.
- [2] N. Bennenni, K. Guenda, S. Mesnager, DNA cyclic codes over rings, *Advances in Mathematics of Communications*, vol. 11, pp. 83-98, 2017.
- [3] M. M. Brandão, L. Spoladore, L. C. B. Faria, A. S. L. Rocha, M. C. Silva-Filho, R. Palazzo Jr., Ancient DNA sequence revealed by error-correcting code, *Scientific Reports*, vol. 5, pp. 1-9, 2015.
- [4] B. Feng, S. Bai, B. Chen, X. Zhou, The constructions of DNA codes from linear self-dual codes over  $Z_4$ , “International Conference on Computer Information Systems and Industrial Applications”, 2015.
- [5] J. Liang, L. Wang, On cyclic DNA codes over  $F_2 + uF_2$ , *Journal of Applied Mathematics and Computing*, vol. 51, pp. 81-91, 2016.
- [6] Z. Varbanov, T. Todorov, M. Hristova, A method for constructing DNA codes from additive self-dual codes over  $GF(4)$ , *ROMAI Journal*, v. 10, pp. 208-211, 2014.

## Modelagem Matemática para Estudo da Evolução da Tuberculose no Estado do RJ nos últimos 5 anos

**Orlando S. Pereira**      **Fernanda B. Damasceno\***

Universidade Federal Rural do Rio de Janeiro, Departamento de Matemática  
23890-000, Seropédica, RJ

E-mail: fbdamasceno7@gmail.com, orlandopereira@ufrj.br,

**Cristiane S.S. Pereira**

Universidade de Vassouras- Mestrado Profissional em Ciências Ambientais  
27700-000, Vassouras, RJ

E-mail: crispereirauss@gmail.com

### RESUMO

A Tuberculose é uma doença infectocontagiosa cujo agente transmissor é chamado *Mycobacterium tuberculosis*, também conhecido como bacilo de Koch [2]. A apresentação da tuberculose se dá, principalmente, na forma pulmonar, no entanto pode afetar outros órgãos. Segundo o Ministério da Saúde, essa doença é a que mais mata no mundo, ultrapassando mortes causadas por HIV e malária juntos. Conforme a Organização Mundial da Saúde(OMS), a tuberculose é responsável por quase 2 milhões de óbitos a cada ano.

O presente trabalho consistiu em aplicar um modelo matemático clássico, criado em 1927 por Kermack e McKendrick, para o estudo da evolução temporal da Tuberculose no Estado do Rio de Janeiro, a partir de dados do Ministério da Saúde, no período de 2013 a 2018.

O modelo SIR( Suscetíveis-Infectedos-Recuperados) consiste de uma epidemia em que um indivíduo deve ser isolado, podendo ser curado e tornar-se imune a doença ou morrer. A dinâmica deste modelo consiste em analisar os indivíduos infectados e também os que se recuperam da doença, como ilustrado na Figura 1,



Figura 1: Modelo compartimental SIR sem dinâmica Vital.

onde  $S = S(t)$  é o número de indivíduos sadios e aptos a contraírem a doença por contatos infecciosos,  $I = I(t)$  é o número de indivíduos anteriormente suscetíveis que contraíram a doença de outros infecciosos e passam imediatamente a serem transmissores dela por contato num instante  $t$  e  $R = R(t)$  é o número de indivíduos que, de infecciosos passam a não mais participar da dinâmica epidêmica.

Após coleta de dados dos infectados e estudos preliminares, foi feita uma simulação numérica do modelo com Matlab(MathWorks Inc.). A justificativa em trabalhar o modelo SIR, além das questões acadêmicas, é analisar de maneira mais ampla os dados de infectados que se dispõe a cada ano e fazer projeções, de modo que novas medidas sejam tomadas, por exemplo, pelos governantes, caso a doença não esteja controlada. Este trabalho foi inspirado no trabalho de Almeida(2017)[1]. Estudos correlatos sobre este tema podem ser vistos na literatura, como em Sabino (2011)[3], Ribeiro Luis(2012)[2], dentre outros.

A formulação matemática do modelo SIR é representado por um sistema de equações diferenciais ordinárias, de primeira ordem e não- linear,

---

\* bolsista de Iniciação Científica FAPERJ/RJ

$$\text{Modelo SIR} \quad \begin{cases} \frac{dS}{dt} = -\beta SI \\ \frac{dI}{dt} = \beta SI - \mu I \\ \frac{dR}{dt} = \mu I \end{cases} \quad t \geq 0 \text{ e condições iniciais, } S(0) = S_0, I(0) = I_0, R(0) = 0.$$

Abaixo na tabela, são apresentados os dados para simulação

### Dados Tuberculose (Estado do Rio de Janeiro)

Ano	Nº de infectados	População	$\beta= I/N$
2013	13.347	16.616.344	0,000803
2014	13.106	16.726.184	0,000784
2015	13.364	16.841.138	0,000794
2016	13.578	16.947.738	0,000801
2017	13.871	17.051.465	0,000813
2018	13.875	17.159.760	0,000809
Média	13.524	16.890.438	<b>0,000801</b>

Fonte: SINAN: Sistema de Informação de Agravos de Notificação

Assim, simulou-se o problema considerando a média dos dados da população. Logo,  $N = 16.890.438$ ,  $S(0) = 16.876.914$  e  $I(0) = 13.524$ , com os parâmetros  $\beta = 0,000801$  e  $\mu = 0,124093$ , sendo este último determinado via regressão linear dos dados dos infectados em relação ao tempo, e neste caso, obteve-se a taxa de  $0,0015$ , que igualando a  $\beta SI - \mu I$  determinou-se  $\mu = 0,124093$ . Na Figura 2, apresenta-se o gráfico da simulação no Software Matlab.

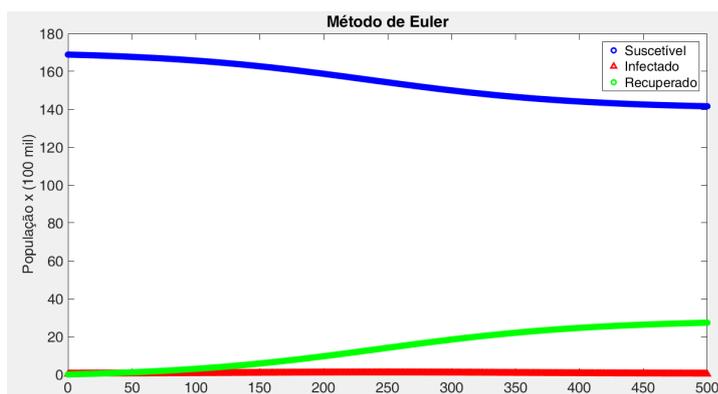


Figura 2: Simulação no Matlab via método de Euler

Por fim, calculou-se a taxa de reprodutibilidade basal  $R_0 = \frac{\beta S(0)}{\mu} \cong 1,09 > 1$  e como  $R_0 > 1$ , conclui-se que a Tuberculose não está controlada, com tendência epidêmica em um período curto de tempo, caso nada seja feito.

**Palavras-chave:** Modelagem Matemática, Epidemiologia, Tuberculose

## Referências

- [1] C.P. ALMEIDA, O.S. PEREIRA, C.S.S. PEREIRA , "Modelagem Matemática e Simulação Numérica de um Sistema de Equações Diferenciais: Estudo de Caso para Análise de Doença(s) Infecciosa(s) no Estado do RJ "; Anais do ERMAC UFRRJ, p. 140-143, Seropédica, 2017.
- [2] M.H. RIBEIRO LUIS, "Modelos Matemáticos em Epidemiologia"; Tese de Mestrado, IGCE/UNESP, Rio Claro, 2012 .
- [3] M.R. SABINO , "Efeitos da Vacinação e do Tratamento na Dinâmica da Transmissão da Tuberculose"; Tese de mestrado, IMECC/UNICAMP, 2011.

## O Problema da Roda Quadrada: Uma Abordagem a Partir da Geometria Dinâmica e Impressora 3D

**Leonardo Consorte Veit\***

**Delair Bavaresco**

Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande Do Sul- Campus Bento Gonçalves  
95700-206, Bento Gonçalves-RS

E-mail: leonardo.consorteweit@yahoo.com.br, delair.bavaresco@bento.ifrs.edu.br,

### RESUMO

O problema da bicicleta de rodas quadradas, apresentado por Hall e Wagon [1], propõe a discussão da existência de uma forma de andar de bicicleta com rodas de formas diferentes sem causar muitos solavancos e força para realizar o movimento de rolamento. No entanto, a abordagem dos autores traz ilustrações bastante precárias e com pouco detalhamento das construções geométricas por eles utilizadas. Nesse trabalho apresentamos a exploração dos entes geométricos com uso de software de geometria dinâmica e a confecção de um mecanismo com rodas quadradas por meio de impressora 3D após modelagem tridimensional em software específico. Nessas discussões utilizamos curvas paramétricas aplicadas ao planejamento e modelagem tridimensional da pista e da roda do sistema dinâmico do mecanismo.

A roda como conhecida atualmente garante que cada ponto da sua volta está a mesma distância do centro da roda, o conhecido eixo, e por essa distância ser constante o movimento em uma estrada plana é sem oscilações. Posto a roda circular em uma estrada que não é perfeitamente lisa, ocorre que o centro da roda passa a ser variante em sua trajetória. Mas qual seria a estrada apropriada para a roda na forma de um quadrado. Se um quadrado é posto em movimento sobre uma superfície plana, a distância do centro em relação a superfície não é constante. Sendo assim o problema consiste em construir uma pista que permite que o centro da roda viaje em uma linha constante e que as irregularidades da roda sejam corrigidas pela pista.

A parte inicial dos estudos constituiu da solução bidimensional do problema com simulação a partir de construções dinâmicas no *software* Geogebra. Para a realização das simulações foi definido e implementadas curvas paramétricas nas quais um ponto (vértice da roda) passou a dar o movimento de rolamento. Posteriormente, foi construído a roda no formato quadrado com dependência dinâmica desse ponto.

De acordo com os autores, a obtenção da curva que determina a pista, se dá pela solução de uma equação diferencial ordinária. Contudo há condições para que uma roda quadrada role sobre uma superfície ou uma curva. Primeiramente devemos garantir que o lado da roda tenha um ponto de contato em toda sua extensão, e além disso, a distância percorrida pela roda e pela pista deve ser a mesma em qualquer ponto de movimento. A segunda condição proposta é que o centro geométrico da roda se desloque sempre na horizontal sem variações.

A obtenção da curva característica da pista é obtida em consequência das características definidas para a roda. Para uma roda quadrada se dá por uma sequência de cossenos hiperbólicos, ou seja, se obtivermos a estrada para um lado do quadrado garantiremos que essa curva vai ser exatamente igual para os próximos lados, sendo periódica em toda sua extensão. Com a função que descreve a pista conseguimos incrementar a simulação permitindo realizar a generalização baseado em parâmetros de ampliação. Quando verificamos que a estrada para a roda era válida, seus múltiplos, consequentemente, foram válidos também.

A solução da roda quadrada permite generalização para rodas definidas por polígonos regulares. Consequentemente para polígonos regulares de muitos lados a pista se aproxima de uma linha reta. Esses resultados podem ser observados na generalização construída no Geogebra, mostrada na Figura 1, a seguir.

Visto que o projeto em duas dimensões fornece características para sua expansão no tridimensional, foi realizado a simulação de uma partição ondulada de uma pista circular. A superfície dessa partição é em formato de um arco de circunferência e construída de acordo com

a quantidade de partições desejada. Para essa implementação utilizamos 28 partições para compor a pista circular ondulada.

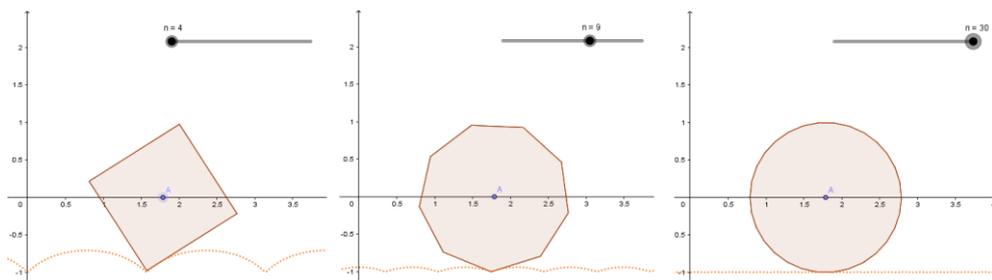


Figura 1: Simulação para generalização das trajetórias construídas no *GeoGebra*.

Para modelagem tridimensional de cada partição, foi utilizado o software Blender, o qual fornece ferramentas de modelagem e animação, além de possuir caixa de entrada para sistema de equações matemáticas. Após a modelagem tridimensional foi confeccionada uma partição com impressão 3D de modo que o projeto começou a tomar forma. Verificado as consistências da partição, foi realizado as impressões dos 27 lócus restantes da pista.

Já a roda para esta pista se torna um paralelepípedo com cortes em suas arestas de acordo com o ângulo que a partições se encontram, e dessa forma o eixo de simetria da roda se torna uma reta com ângulo da partição. Para adequar as rodas à pista circular completa, foram confeccionadas rodas de tamanhos diferentes, com dimensões de acordo com a distância do centro da pista a posição em que a roda se deslocaria.

Por fim, cada partição correspondente a um lado da roda foi confeccionada de uma cor para facilitar a visualização do projeto. Para implementação do movimento foi realizado a tração nos eixos com um motor movido a bateria. O resultado desse projeto é mostrado na Figura 2, a seguir.



Figura 2: Pista modelada no software Blender e sua representação física.

A realização da pesquisa que culminou com a produção desse trabalho nos mostrou que as condições analíticas e paramétricas para definir as simulações no software Geogebra demandam aprofundamentos de estudos e precisão de cálculos. As maiores dificuldades se evidenciaram na determinação dos parâmetros para modelagem tridimensional e para simulação. Vale ressaltar que a maioria de curvas com condições de fechamento não possui uma equação diferencial de simples resolução, sendo necessário o uso de aproximações numéricas para sua solução. No entanto, a utilização de recursos tecnológicos contemporâneos permite novas discussões para problemas clássicos e a obtenção de soluções de forma diferenciada e facilitada pelo processamento numérico e pela visualização e confecção tridimensional.

**Palavras-chave:** *Curvas paramétricas, Geogebra, Blender*

## Referências

[1] HALL, Leon; WAGON, Stan. Roads and Wheels. Mathematics Magazine, Washington D.C., v. 65, n. 5, p.283-301, Dec.,1992.

# A Álgebra Linear dos Quadrados Mágicos

**João Vitor Pamplona**

Universidade Federal de Santa Catarina, Departamento de Matemática  
89036-002, Campus Blumenau, Blumenau, SC  
E-mail: joaovitorpamplona@hotmail.com,

**Luiz Rafael dos Santos**

Universidade Federal de Santa Catarina, Departamento de Matemática  
89036-002, Campus Blumenau, Blumenau, SC  
E-mail: l.r.santos@ufsc.br.

## RESUMO

Uma lenda chinesa diz que certa vez uma tartaruga surgiu do Rio Lo durante uma inundação e em seu casco estava esculpido um padrão diferente: um *grid* três-por-três contendo números diferentes tais que a soma de cada uma das três linhas, das três colunas e das duas diagonais totalizavam 15, que na forma matricial pode ser apresentada como

$$\begin{bmatrix} 4 & 9 & 2 \\ 3 & 5 & 7 \\ 8 & 1 & 6 \end{bmatrix}.$$

Sua real origem não é conhecida, mas há registros de sua existência em épocas anteriores à nossa era, na China e na Índia. Por exemplo, o quadrado de 9 casas é encontrado em um manuscrito árabe no fim do Século VIII atribuído a Apolônio de Tiana.

Um *quadrado mágico* de ordem  $n$  é uma matriz quadrada  $A$  de ordem  $n$  com  $n^2$  números tais que a soma dos elementos de cada linha e de cada coluna, bem como da diagonal principal e da diagonal secundária é o mesmo número, chamado constante mágica.

Neste trabalho foi desenvolvido, em linguagem computacional, um método que gerasse quadrados mágicos de qualquer ordem. Tal implementação foi feita em **Julia** [1] com três algoritmos distintos:

1. Para quadrados mágicos de ordem *ímpar*[2];
2. Para quadrados mágicos de ordem *par divisível por 4*, chamado também de quadrado mágico duplo par;
3. Para quadrados mágicos de ordem *par não divisível por 4*.

Ressalta-se ainda, que os algoritmos 2 e 3 foram desenvolvido neste trabalho.

**Palavras-chave:** *Quadrado mágico, implementação, algoritmo, matemática computacional.*

## Referências

- [1] BEZANSON, J. et al. Julia: A Fresh Approach to Numerical Computing. *Siam Review*, Society for Industrial and Applied Mathematics, v. 59, n. 1, p. 65–98, fev. 2017. Disponível em: <<http://epubs.siam.org/doi/10.1137/141000671>>.
- [2] LEONHARD EULER. On magic squares. *ArXiv Mathematics e-prints*, ago. 2004.

## Transformações Conformes da Esfera para o Plano

**Mateus Souza Oliveira\***      **Leonardo Koller Sacht†**

Universidade Federal de Santa Catarina - Departamento de Matemática  
88040-900, Campus Trindade, Florianópolis, SC  
E-mail: matews1943@gmail.com, leonardo.sacht@ufsc.br.

### RESUMO

Imagens com grandes campos de visão, as ditas imagens panorâmicas, são uma maneira mais completa de representar o mundo. Com intuítos tanto práticos quanto artísticos, estas imagens apresentam distorções não naturais, as quais são o principal problema da área. Por exemplo, linhas retas do mundo aparecem curvadas na imagem, e objetos aparecem distorcidos. Consideramos estes como sendo os problemas principais nos panoramas, os quais denominamos, matematicamente, como falta de curvatura nula e de conformalidade, respectivamente. Como visto na principal referência [1], não existem funções que satisfazem ambas as características. Portanto, optamos por tentar caracterizar as transformações conformes, por acharmos que elas serão usadas de forma mais abrangente. Em suma, uma projeção é dita conforme quando preserva ângulos e proporcionalidade, e mapeia círculos em círculos.

Em nossos estudos, consideramos uma imagem panorâmica como sendo uma representação com campo visual (*field of view* em inglês, por isso usamos a sigla FOV) superior a  $180^\circ$  de longitude por  $90^\circ$  de latitude. Muitos autores não se preocupam em enfatizar isto por trabalharem com imagens que abrangem o campo de visão esférico total. Estas são as chamadas imagens equirretangulares. Os algoritmos que propomos até o momento aceitam FOVs variados, desde que respeitem os domínios das projeções em questão.

Uma outra forma de abordar as imagens equirretangulares é através da representação por longitudes e latitudes  $r : [-\pi, \pi) \times (-\frac{\pi}{2}, \frac{\pi}{2}) \rightarrow \mathbb{S}^2$  (em que  $\mathbb{S}^2$  é a esfera unitária centrada na origem do  $\mathbb{R}^3$ ), dada por  $(\lambda, \phi) \mapsto (\cos(\lambda) \cdot \cos(\phi), \sin(\lambda) \cdot \cos(\phi), \sin(\phi))$ . Como a aplicação  $r$  é uma bijeção, é comum tratamos o domínio equirretangular  $([-\pi, \pi) \times (-\frac{\pi}{2}, \frac{\pi}{2}))$  e a esfera como o mesmo objeto.

Por ser um assunto que está sendo abordado pela primeira vez, ainda não sabemos se seremos capazes de caracterizar o espaço das transformações conformes. Então, nosso objetivo atual tem sido o aperfeiçoamento de um algoritmo que usaremos para averiguar se uma projeção é ou não conforme.

Explicaremos os passos de execução desse código, além de expor alguns testes. Primeiro, o usuário indica quantos pontos irá querer na malha, a qual é uma discretização regular do domínio equirretangular (quanto mais pontos, mais próximo estamos da teoria). Depois, é escolhida a transformação, com a qual é montado o vetor da projeção, com os valores da transformação em cada ponto da malha discretizada. Nessas avaliações, escolhemos projeções clássicas da área, além da aplicação identidade. Expomos a seguir suas formas algébricas:

- i) a projeção de Mercator  $M : [-\pi, \pi) \times (-\frac{\pi}{2}, \frac{\pi}{2}) \rightarrow \mathbb{R}^2$ , é dada por  $(\lambda, \phi) \mapsto (u(\lambda, \phi), v(\lambda, \phi)) = (\lambda, \log(\sec(\phi) + \tan(\phi)))$ .
- ii) a projeção Estereográfica  $E : (-\pi, \pi) \times (-\frac{\pi}{2}, \frac{\pi}{2}) \rightarrow \mathbb{R}^2$ , é dada por  $(\lambda, \phi) \mapsto (u(\lambda, \phi), v(\lambda, \phi)) = \left( \frac{2 \cdot \sin(\lambda) \cdot \cos(\phi)}{\cos(\lambda) \cdot \cos(\phi) + 1}, \frac{2 \cdot \sin(\phi)}{\cos(\lambda) \cdot \cos(\phi) + 1} \right)$ .

---

\*Graduando do curso de Matemática - Bacharelado na Universidade Federal de Santa Catarina

†Professor do Departamento de Matemática da Universidade Federal de Santa Catarina

- iii) a projeção Identidade  $I : [-\pi, \pi) \times (-\frac{\pi}{2}, \frac{\pi}{2}) \rightarrow \mathbb{R}^2$ , é dada por  $(\lambda, \phi) \mapsto (u(\lambda, \phi), v(\lambda, \phi)) = (\lambda, \phi)$ .
- iv) a projeção Perspectiva  $P : (-\frac{\pi}{2}, \frac{\pi}{2}) \times (-\frac{\pi}{2}, \frac{\pi}{2}) \rightarrow \mathbb{R}^2$ , é dada por  $(\lambda, \phi) \mapsto (u(\lambda, \phi), v(\lambda, \phi)) = (\tan(\lambda), \frac{\tan(\phi)}{\cos(\lambda)})$ .

Na teoria (e no algoritmo), usamos o teorema 1 para avaliar a conformalidade de funções.

**Teorema 1.** Sejam uma projeção  $P : \Omega \subseteq \mathbb{S}^2 \rightarrow \mathbb{R}^2$ ,  $P(\lambda, \phi) = (u(\lambda, \phi), v(\lambda, \phi))$  e um ponto  $p = (\lambda, \phi)$  no interior de  $\Omega$ . A projeção  $P$  é conforme (a menos de reflexões) se, e somente se, satisfaz as equações de Cauchy-Riemann na esfera:  $\frac{\partial u}{\partial \phi}(p) = -\frac{1}{\cos(\phi)} \cdot \frac{\partial v}{\partial \lambda}(p)$  e  $\frac{\partial v}{\partial \phi}(p) = \frac{1}{\cos(\phi)} \cdot \frac{\partial u}{\partial \lambda}(p)$ .

*Demonstração.* Encontra-se em [1], teorema 2.1. ■

Ressaltamos que, entre as quatro projeções apresentadas anteriormente, apenas Mercator e Estereográfica são conformes, pois são as únicas que satisfazem o teorema 1.

Por fim, é calculado o produto da matriz da discretização das equações de Cauchy-Riemann, através de diferenças finitas, com o vetor da projeção. O erro é dado pela norma do vetor resultante, que se aproxima de zero no caso das funções conformes. Expomos alguns dos testes realizados com as quatro transformações na figura 1.

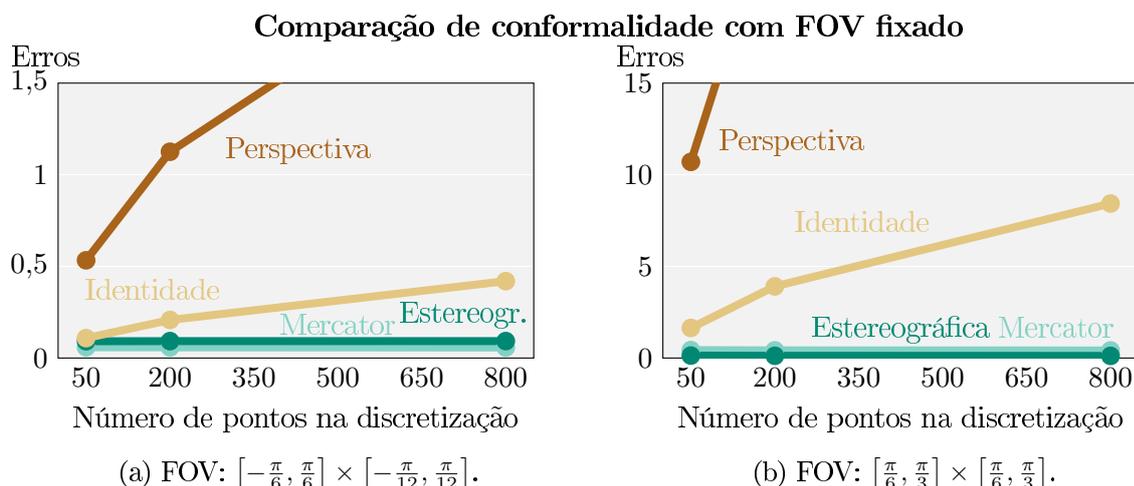


Figura 1: Mesmo com um FOV baixo (menos de 3% da área do domínio equiretangular) (a), há uma clara divergência nos valores de erro das projeções não conformes, e uma sutil convergência das conformes. Ao se deslocar o campo de visão do centro do domínio equiretangular (b), o erro produzido pelas projeções não conformes é ainda mais elevado, mesmo reduzindo a área do campo de visão pela metade.

Ao se comparar funções discretizadas com projeções conformes discretizadas, conseguimos ver se elas satisfazem ou não as equações de Cauchy-Riemann, mesmo usando uma quantidade relativamente baixa de pontos na malha (testes com 800 pontos apresentados na figura 1, levaram em média 0,03 segundos, para cada projeção).

**Palavras-chave:** *Imagens Panorâmicas, Projeções Conformes*

## Referências

- [1] L. K. Sacht, “Content-Based Projections for Panoramic Images and Videos”, Dissertação de Mestrado, IMPA, 2010.  
URL: [http://w3.impa.br/~leo-ks/publications/thesis\\_leonardo\\_sacht\\_2010.pdf](http://w3.impa.br/~leo-ks/publications/thesis_leonardo_sacht_2010.pdf)

## Identificação e Classificação de Estrelas Utilizando Máquinas de Vetores de Supervisão

**João Victor Ribeiro de Jesus**

**Alex Luciano Roesler Rese**

**Rodrigo Lyra**

Universidade do Vale do Itajaí (UNIVALI) – Laboratório de Inteligência Aplicada

88302-901, Campus Itajaí, Itajaí, SC

E-mail: joao.v.ribeiro@outlook.com, alexrese@univali.br, rlyra@univali.br

### RESUMO

O presente trabalho aplicou métodos de processamento e classificação de imagens para a área da astronomia, nestes processos foram extraídos dados para realizar o treinamento e posteriormente sua classificação. Na abordagem adotada, foram utilizadas técnicas que visam reduzir o ruído em imagens para ressaltar as informações mais relevantes, e Máquinas de Vetores de Supervisão foram empregadas para realizar o treinamento e posteriormente a classificação. Nos experimentos realizados a estratégia adotada permitiu a identificação de modo eficiente para duas estrelas, Polaris (alfa Ursa Menor) e Mimosa (beta Crucis).

### **Introdução**

Desde os primórdios da civilização humana, estão presentes os fatores astronômicos para sua orientação e deslocamento em terra, e posteriormente muito utilizado na navegação [5]. As estrelas Polaris e Mimosa são exemplos de referência utilizadas para este fim.

Este resumo reporta a realização de um estudo sobre estas estrelas, aplicando processamento e classificação de imagens com o objetivo de identificar, computacionalmente, estes astros.

### **Desenvolvimento**

Foi realizado a obtenção de um conjunto de imagens da esfera celeste. O conjunto utilizado como amostra foi extraído do repositório “STScI Digitized Sky Survey” [1].

As imagens foram submetidas primeiramente a um processamento, com o objetivo de facilitar sua classificação no SVM. Este processamento é dividido em quatro etapas (Figura 1): (i) Na primeira etapa, as cores da imagem são alteradas para escala de cinza. (ii) Após esta etapa, a imagem é submetida ao processo de limiarização, que consiste em dividir a imagem em duas classes, fundo e objetos, sendo uma delas representada pela cor preta e a outra pela cor branca [3,9]. Os pixels com valores maiores que o limiar especificado são transformados em branco e os com valor menor são transformados em preto. (iii) A terceira etapa aplica uma erosão na imagem, retirando a borda de elementos maiores e removendo elementos menores. Isto faz com que as estrelas menores desapareçam por apresentarem pouca densidade de pixels, o que diz respeito a baixa relevância para a identificação de objetos. (iv) A última etapa consiste na dilatação dos objetos da imagem, evidenciando elementos relevantes [9].

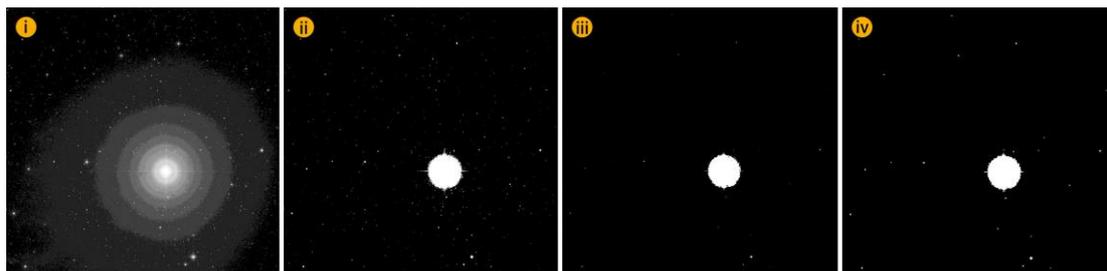


Figura 1. Etapas de pré-processamento: (i) escala de cinza, (ii) limiarização, (iii) erosão e (iv) dilatação

Para realizar a classificação foi utilizado o método de Máquinas de Vetores de Supervisão (SVM), que consiste em um conjunto de aprendizado supervisionado para reconhecimento de padrões. Inicialmente, o SVM recebe um conjunto de informações base para treinamento e entrega como resultado um modelo de classificação que é utilizado para categorizar os dados, como parâmetro para o SVM foi utilizada a distância relativa (Figura 2) entre a estrela a ser classificada em relação as suas vizinhas, além destas, a magnitude (tamanho) também foi utilizada.

Para a implementação dos tratamentos da imagem foi utilizado a linguagem de programação C++ e a biblioteca OpenCV. Esta biblioteca foi escolhida por possuir licença livre para o uso acadêmico e possuir técnicas necessárias para o processamento de imagens utilizados neste trabalho [2].

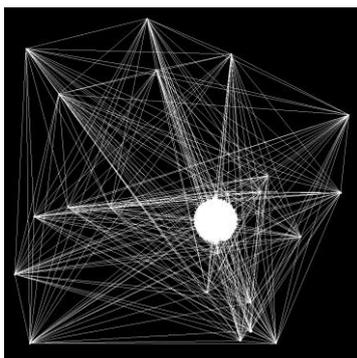


Figura 2. Distância relativa e Magnitude das estrelas

## Considerações Finais

Neste projeto foi realizada a aplicação de quatro técnicas de processamento de imagem sendo estas: (i) Escala de Cinza, (ii) Limiarização, (iii) erosão e (iv) dilatação. Estas foram aplicadas como um pré-processamento para o treinamento de SVM. Com isto foram extraídas as informações de maior relevância para a classificação dos conjuntos.

Para o treinamento do SVM foram utilizadas 44 imagens para o primeiro treinamento e 544 para o segundo e terceiro, analisando 10 imagens para cada experimento.

O trabalho apresentou resultados interessantes para a classificação de estrelas utilizando SVM. Sendo possível identificar conjuntos distintos de imagens mesmo quando ambos eram exemplos negativos da estrela analisada. Portanto, é possível concluir que foram identificados subconjuntos dentro de conjuntos de imagens.

**Palavras-chave:** *Processamento de Imagem, Visão Computacional, Estrelas*

## Referências

- [1] Association of Universities for Research in Astronomy, (Org.). The STScI Digitized Sky Survey. Disponível em: <[http://stdu.stsci.edu/cgi-bin/dss\\_form](http://stdu.stsci.edu/cgi-bin/dss_form)>. Acesso em: 10 set. 2015.
- [2] G. Bradski, and A. Kaehler, Learning OpenCV: Computer Vision with the OpenCV Library. vol. 1, (2008).
- [3] W. Burger and M. J. Burge, Digital Image Processing. London: Springer London, (2008).
- [4] P. Soille, Morphological Image Analysis. Berlin, Heidelberg: Springer Berlin Heidelberg. vol. 24, (1999)
- [5] K. S. Oliveira Filho and M. F.O. Saraiva. Astronomia e Astrofísica. vol. 2, pp. 760, (2003)

## Aplicações de Códigos Corretores de Erros no Processo de Geração de Sequências de DNA

**Alice Noronha de Oliveira\***

Programa de Pós-Graduação em Estatística Aplicada e Biometria Unifal-MG,  
37130-001, Alfenas, MG  
E-mail: alicenoronhaa@hotmail.com,

**Anderson José de Oliveira**

Universidade Federal de Alfenas - Departamento de Matemática  
37130-001, Alfenas, MG  
E-mail: ajoliveira01@gmail.com.

### RESUMO

A necessidade de se garantir que uma mensagem seja armazenada ou transmitida de forma confiável pelos diversos meios existentes, exige o uso da teoria dos códigos corretores de erros. Durante a transmissão ou armazenamento de informações, podem ocorrer algumas interferências ou perturbações, que fazem com que a mensagem recebida seja diferente da original. Porém, através de técnicas adequadas de codificação e decodificação é possível garantir que a mensagem enviada chegará com uma quantidade de erros pequena ou sem erros ao seu destino. Nos últimos anos, com o avanço tecnológico, a procura por serviços de alta confiabilidade para a transmissão de dados digitais impulsionou o estudo da teoria dos códigos corretores de erros, tornando-o um campo de pesquisa ativo e com aplicação em diversas áreas do conhecimento. Dentre os diversos códigos corretores de erros existentes, os códigos BCH tem se destacado devido às características simples e sua aplicação em diversos contextos, podendo ser utilizado na construção de métodos práticos em hardware e software, em aplicações espaciais e até mesmo no âmbito biológico. A relação dos códigos corretores de erros com a biologia foi apontada por [5] e [2], que evidenciaram a existência de uma estrutura matemática nas sequências de DNA associadas aos códigos corretores de erros, também identificaram, reproduziram e classificaram matematicamente diferentes sequências de DNA por meio de modelos de sistema de comunicação de informação genética análogos ao modelo de sistema de comunicação digital. A reprodução dessas sequências ocasiona uma diminuição de tempo e de gastos laboratoriais para análises mutacionais. Na mesma linha de pesquisa, [1] apresentaram algumas aplicações de estruturas matemáticas no código genético, tornando-se ferramenta eficaz na análise de diversos fenômenos biológicos. Além disso, mostraram uma modelagem do código genético possibilitando a identificação de características, propriedades e implicações do modelo. [4] mostrou os benefícios de associar as tecnologias computacionais com o uso de análises da biotecnologia, através da apresentação de um algoritmo que faz a predição mutacional em sequências de DNA por meio do processo de decodificação. [3] propuseram uma conexão entre a constelação de 64 sinais presentes no modulador genético de um sistema de comunicação genético com um sistema de comunicação digital. Com isso, foi possível contribuir em uma área em franca expansão, fazendo com que a biologia se torne uma ciência fundamentada em conceitos matemáticos. Com base nesses estudos, pressupõe-se a relação entre engenharia (códigos corretores de erros), em particular os códigos BCH, biologia e estruturas matemáticas. Perante isso, o objetivo deste trabalho é apresentar a conexão

---

\*Mestranda do Programa de Pós-Graduação em Estatística Aplicada e Biometria

existente entre essas três áreas do conhecimento, por meio do detalhamento teórico dos códigos BCH e as estruturas algébricas utilizadas, além da relação das mesmas com sequências de DNA, e que através do algoritmo de geração de proteínas proposto por [5] e [2] possamos reproduzir sequências de DNA obtidas por meio do banco de dados NCBI (em inglês, National Center for Biotechnology Information), para análise de ocorrências de mutações genéticas.

**Palavras-chave:** *Códigos BCH, Comunicação Genética, Estruturas Algébricas.*

## Referências

- [1] A. J. de Oliveira, R. Palazzo Júnior. Análise Algébrica dos Rotulamentos Associados ao Mapeamento do Código Genético. Dissertação (Mestrado), Faculdade de Engenharia Elétrica, Campinas, 2012.
- [2] A. S. L. Rocha, R. Palazzo Júnior. Modelo de sistema de comunicação digital para o mecanismo de importação de proteínas mitocondriais através de códigos corretores de erros. Tese (Doutorado), Faculdade de Engenharia Elétrica e de Computação, Unicamp, 2010.
- [3] B. P. Pires, A. J. de Oliveira. Aplicações de estruturas algébricas na conexão entre sistemas de comunicação padrão e genética, *Sigmae*, vol.6, pp. 1-14, 2017.
- [4] D. G. Pereira, R. Palazzo Júnior, Uma abordagem computacional para análise de sequências de DNA por meio dos códigos corretores de erros. Dissertação (Mestrado), Faculdade de Engenharia Elétrica, Campinas, 2014.
- [5] L. C. B. Faria, R. Palazzo Júnior. Existências de Códigos Corretores de Erros e Protocolos de Comunicação em Sequências de DNA. Tese (Doutorado), Faculdade de Engenharia Elétrica e de Computação, Unicamp, 2011.

## Sistema robotizado usando tecnologias de atuação por cabos

**Gabriel Dinse\***

**Leonardo Mejia**

Departamento de Engenharia de Controle, Automação e Computação  
Universidade Federal de Santa Catarina - Campus Blumenau  
E-mail: gabriel\_dinse@hotmail.com, leonardo.mejia.rincon@ufsc.br,

### RESUMO

O estudo de manipuladores e robôs atuados por cabos data principalmente da década dos 80, quando Landsberg e Sheridan propuseram a substituição de elos rígidos de uma plataforma de Stewart por cabos flexíveis [1]. Desde então, alguns trabalhos já foram desenvolvidos nessa área, visando a sua utilização em diferentes aplicações e em escalas diversas. Manipuladores e robôs atuados por cabos oferecem grandes vantagens em relação a sistemas tradicionais que utilizam apenas corpos rígidos devido ao seu grande espaço de trabalho alcançável, facilidade de reconfiguração, alta capacidade de carga e alto desempenho dinâmico [2]. Os graus de liberdade e a forma do movimento destes sistemas são dados pela quantidade de cabos atuados e pela disposição deles com relação à plataforma atuada [3].

Segundo Ming e Higuchi [4], Manipuladores e robôs atuados por cabos podem ser classificados em: **CRPM** (Completely Restrained Parallel Manipulator) e **IRPM** (Incompletely Restrained Parallel Manipulator). Verhoeven [5] aponta também a existência de uma terceira categoria chamada **RRPM** (Redundantly Restrained Parallel Manipulator). As classificações foram realizadas em função da diferença entre o número de cabos ( $n$ ) e o número de graus de liberdade do manipulador ( $m$ ). As categorias anteriormente citadas respeitam as relações mostradas na sequência: 1). **CRPM**:  $n = m + 1$ ; 2). **IRPM**:  $n \leq m$ ; 3). **RRPM**:  $n \geq m + 1$ .

Uma classificação adicional apresentada por Verhoeven [5], divide os manipuladores em seis grupos de acordo com o número de graus de liberdade controláveis, e com o tipo de movimento do efetuador final como mostrado a seguir: **classe 1: "1T"**, manipulador que permite apenas movimento linear de qualquer ponto de interesse no efetuador final do manipulador; **classe 2: "2T"**, manipulador que permite o movimento planar de qualquer ponto de interesse no efetuador final do manipulador; **classe 3: "1R2T"**, manipulador que permite o movimento planar completo de qualquer ponto de interesse no efetuador final do manipulador; **classe 4: "3T"**, manipulador que permite o movimento espacial de qualquer ponto de interesse no efetuador final do manipulador; **classe 5: "2R3T"**, manipulador que permite o movimento espacial de uma barra usada como efetuador final do manipulador; **classe 6: "3R3T"**, manipulador que permite o movimento espacial completo de um corpo usado como efetuador final do manipulador.

O presente projeto visa a concepção, construção, controle e programação de um manipulador atuado por cabos das classes CRPM-1R2T. Nesse sentido, e respeitando as considerações previamente apresentadas, podemos descrever este equipamento como sendo uma plataforma fixa unida a uma plataforma móvel através de quatro cabos tensionados, sendo estes: **i)** dois cabos atuados por motores presos aos vértices inferiores da plataforma fixa e **ii)** outros dois cabos unidos a pesos suspensos em polias nos vértices superiores também da plataforma fixa. Estes pesos são colocados com o objetivo de manter os cabos sempre sob tensão sem precisar de um sistema de tração ativo (*i.e. motores*). Neste sistema, a tensão nos cabos é gerada pela combinação linear das componentes das forças geradas pelos motores e pelos pesos, o que se diferencia de sistemas onde o peso do manipulador gera a tensão nos cabos [6].

Como os cabos estarão sempre tensionados a análise da cinemática de posição do sistema pode ser simplificada interpretando os cabos como elos de comprimento variável. Através desta análise, é possível a obtenção das equações da cinemática inversa que descrevem o movimento do robô usando uma abordagem puramente vetorial onde a cinemática inversa é dada pela obtenção do comprimento dos cabos atuados pelos motores a partir da posição da plataforma móvel, de forma que apenas o movimento do robô será analisado mas sem considerar as forças envolvidas.

Devido que o manipulador é categorizado como CRPM-1R2T, este pode ter o seu efetuador final rotacionado e transladado sobre um plano, portanto este manipulador apresenta três graus de liberdade. Obtidos os cálculos da cinemática e geometria do manipulador é possível desenvolver uma grande variedade de aplicações interessantes envolvendo seu movimento. Para realizar o controle do robô, são gerados comandos no computador baseados na escolha do usuário através de uma interface gráfica criada no software Qt Creator e que permitirá o cálculo de trajetórias otimizadas que serão posteriormente comunicadas a um sistema embarcado responsável pelo controle dos motores de passo.

Futura aplicações desta plataforma de pesquisa, poderão envolver uma análise das forças envolvidas no sistema, de modo que análises complementares da capacidade de força e manipulabilidade dentre outras poderão ser estudadas.

**Palavras-chave:** *Robô por Cabos, Cinemática, Geração de trajetórias.*

## Referências

- [1] LANSDBERGER, S. E.; SHERIDAN, T. B. A New Design for Parallel Link Manipulators. In: Proc. Sys. Man. and Cybernetics Conf., Tucson, AZ. 1985. p. 812-814.
- [2] LESSANIBAHRI, Saman; CARDOU, Philippe; CARO, Stéphane. Parasitic Inclinations in Cable-Driven Parallel Robots using Cable Loops. In: CIRP DESIGN CONFERENCE, 28., 2018, Nantes. Proceedings. Nantes: Elsevier, 2018. v. 1, p. 296 - 301.
- [3] MURARO, Thaís. Análise das condições algébricas de tracionamento dos cabos em mecanismos do tipo Cable-Suspended usando matroides. 2017. 100 f. Tese (Doutorado) - Curso de Engenharia Mecânica, Departamento de Engenharia Mecânica, Universidade Federal de Santa Catarina, Florianópolis, 2019.
- [4] MING, A. and HIGUCHI, T. Study on multiple degree-offreedom positioning mechanism using wires (part1)- concept, design and control. Int. Journal of the Jap. Soc. for Precision Engeneering 28 (2): pp 131-138. 1994.
- [5] VERHOEVEN, R. Analysis of the workspace of tendon-based Stewart platforms. Ph.D. dissertation. University DuisburgEssen. Duisburg, Germany, 2004.
- [6] ZI, Bin et al. Comparative study of cable parallel manipulators with and without hybrid-driven planar five-bar mechanism. Applied Mathematical Modelling. [s. L.], p. 5994-6017. 15 dez. 2014.

## Números de Catalan: Apresentação, Propriedades e Aplicações

Marciano Pereira      Felipe Pereira Gomes\*

Universidade Estadual de Ponta Grossa,

Ponta Grossa, PR

E-mail: marciano@uepg.br,    felipegomes1501@gmail.com,

### RESUMO

Os números de Catalan são assim chamados em homenagem ao matemático belga Eugène Charles Catalan (1814 - 1894), responsável pela divulgação e demonstração de propriedades dessa sequência numérica. Contudo, atribui-se a descoberta de tais números ao notável matemático suíço Leonard Euler, por volta do ano 1751.

Presentes em problemas de combinatória, álgebra, ciência da computação, teoria dos grafos e geometria, os Números de Catalan são pouco conhecidos no Brasil e possuem pouca literatura na língua portuguesa. O objetivo deste trabalho é fazer um estudo sobre os Números de Catalan, apresentando seu histórico, principais propriedades e algumas aplicações, que podem ser aplicadas em sala de aula. Além disso, difundi-los e torná-los mais conhecidos.

**Definição 1.** Os números de Catalan, representados por  $C_n$ , são definidos por

$$C_n = \frac{1}{n+1} \binom{2n}{n} = \frac{(2n)!}{(n+1)!n!},$$

para todo inteiro  $n \geq 0$ . [?]

Os seis primeiros números da sequência numérica de Catalan são 1, 1, 2, 5, 14 e 42.

Euler propôs um problema ao matemático alemão Christian Goldbach, onde procurava descobrir qual o número de possibilidades de desenhar as diagonais de um polígono convexo sem que haja interseção entre elas. Das trocas de correspondências entre os dois matemáticos surgiu a primeira forma de encontrar os números de Catalan, através de uma fórmula recursiva. [?]

**Fórmula de recorrência de Euler.** Os números de Catalan podem ser obtidos pela seguinte relação de recorrência

$$C_n = \frac{4n-2}{n+1} \cdot C_{n-1},$$

para  $n \geq 1$ , onde define-se  $C_0 = 1$ .

Já no ano de 1756, o matemático alemão Johann Von Segner, através dos estudos de Euler sobre a triangulação de polígonos, publicou uma segunda relação de recorrência para a obtenção dos números de Catalan.

---

\*Mestrando PROFMAT

### Relação de recorrência de Segner.

$$C_n = C_0 \cdot C_{n-1} + C_1 \cdot C_{n-2} + \dots + C_{n-1} \cdot C_0,$$

para  $n \geq 1$ , onde  $C_0 = 1$ .

**Propriedades.** Os números de Catalan gozam de diversas propriedades, mas entre elas destacamos algumas que tratam de paridade, primalidade e divisibilidade.

**Teorema 1.** Para  $n > 0$ , o número de Catalan  $C_n$  é ímpar se, e somente se,  $n$  é um número de Mersenne, ou seja, da forma  $2^m - 1$ , onde  $m \geq 1$ . [?]

**Teorema 2.** Os únicos números de Catalan que são primos são os números  $C_2$  e  $C_3$ . [?]

**Teorema 3.** Para todo  $n \geq 0$ ,  $C_n | n + 2$  se, e somente se,  $n \leq 3$ . [?]

**Aplicações.** Os números de Catalan são utilizados na resolução de problemas envolvendo análise combinatória, dos quais alguns exemplos são descritos a seguir:

**Exemplo 1: Problema da triangulação de polígonos convexos.** De quantas maneiras podemos dividir um polígono convexo em triângulos, desenhando suas diagonais, de modo que não haja interseção entre elas?

**Exemplo 2: Caminhos reticulados de  $(0, 0)$  até  $(n, n)$ .** Em uma malha quadriculada de tamanho  $n \times n$ , de quantas maneiras partindo do ponto  $(0, 0)$  podemos chegar ao ponto  $(n, n)$ , sendo possível apenas passos de comprimento 1 e nas direções leste ou norte, desde que o caminho não cruze a diagonal da malha que une os pontos  $(0, 0)$  e  $(n, n)$ ?

**Exemplo 3: Problema das sequências de parênteses.** De quantas maneiras podemos multiplicar  $n + 1$  números, multiplicando sempre dois a dois, de modo que a ordem da multiplicação seja respeitada?

Para ilustrar, se  $n = 3$ , temos as seguintes possibilidades:

$$(x_1(x_2(x_3x_4))), ((x_1x_2)(x_3x_4)), (x_1((x_2x_3)x_4)), (((x_1x_2)x_3)x_4) e ((x_1(x_2x_3))x_4).$$

Ou seja, 5 possibilidades, que corresponde ao número de Catalan  $C_3$ .

Assim, por exemplo, para  $n = 5$ , não é preciso listar as possibilidades, basta substituímos  $n = 5$  na fórmula combinatória para o cálculo de  $C_n$ . Então, temos que  $C_5 = \frac{1}{5+1} \binom{10}{5} = \frac{(10)!}{(6)!5!} = 42$  possibilidades.

**Palavras-chave:** Catalan, Análise Combinatória, Matemática Discreta

## Referências

- [1] N.G.B. Junior, “Bijeções Envolvendo os Números de Catalan”, Dissertação de Mestrado, IMECC-Unicamp, 2014.
- [2] T. Koshy, M. Salmassi, “Parity and Primality of Catalan Numbers”, *The College Mathematics Journal*, 37 (2006) 52-53.
- [3] T. Koshy, “Catalan Numbers with Applications”, Oxford University Press, New York, 2008.

## Determinante de Cayley-Menger e o volume de poliedros convexos

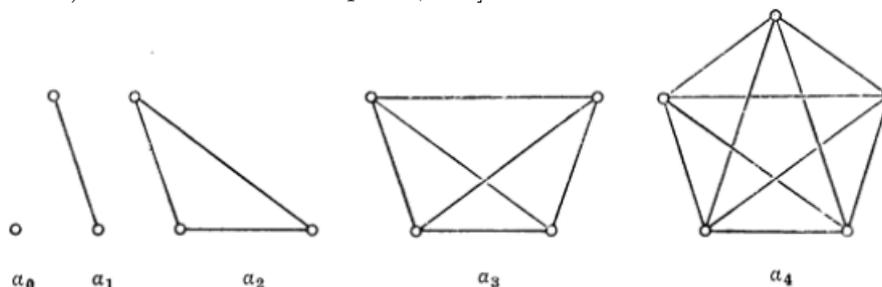
**Leonardo Felipe Silva**      **Felipe Fidalgo**

Universidade Federal de Santa Catarina  
 Departamento de Matemática, Centro de Blumenau  
 89036-004, Blumenau, Santa Catarina, Brasil  
 E-mail: l.f.silva@grad.ufsc.br, felipe.fidalgo@ufsc.br

### RESUMO

Um dos problemas primordiais da matemática de distâncias estava relacionado com objetos de  $n$  dimensões mergulhados em espaços de dimensões menores. Em 1841, o jovem Arthur Cayley publicou um *paper* sobre a relação entre as distâncias de cinco pontos no espaço. Ele mostrou, utilizando argumentos matriciais, que um tetraedro de quatro dimensões somente pode existir em um espaço tridimensional se o seu volume for zero. Anos mais tarde, Karl Menger estabeleceu uma nova caracterização para espaços métricos à partir do conceito de *sistema de congruência*. O determinante de Cayley harmonizava com a métrica de Menger, fato observado por Leonard Blumenthal[5], que uniu ambos e, assim, formalizou o que hoje conhecemos como *determinante de Cayley-Menger*. Uma das ferramentas de Blumenthal foi a deveras conhecida Fórmula de Heron. Em seu livro *Theory and applications of Distance Geometry*, o seguinte resultado é apresentado: “*O volume quadridimensional de um simplex quadridimensional mergulhado no espaço tridimensional é zero.*”.

Figura 1: Aqui representados,  $n$ -simplices, com  $n = 1, \dots, 5$ . [Coxeter, H.S.M. (1973). *Regular Polytopes* (3rd ed.). New York: Dover. p 120,121.]



Segundo Coxeter [6], “quaisquer  $n + 1$  pontos não pertencentes à um espaço  $n - 1$  são vértices de um *simplex*  $n$ -dimensional.”. Ou seja, reconhecemos o *simplex* como a região do espaço fechada pelo menor número de vértices. Note que na Figura 1 acima o objeto formado para o caso onde  $n = 3$  é o triângulo e os casos subsequentes são generalizações de um triângulo para dimensões maiores que 3. Podemos então facilmente perceber a relação deste problema com a fórmula de Heron, pois esta calcula a área de um triângulo à partir das medidas de seus lados. De fato,

**Teorema 1** (Fórmula de Heron). [2] *Seja um triângulo ABC, cujos lados medem a, b e c, e, seja  $s = \frac{1}{2}(a + b + c)$  o seu semiperímetro. Então, temos que  $A = \sqrt{s(s - a)(s - b)(s - c)}$  é a área de ABC.*

O seguinte teorema é o determinante que Cayley demonstrou em [3].

**Teorema 2.** *Dados os pontos  $x_1, \dots, x_5 \in \mathbb{R}^4$ , todos pertencentes à um subespaço 3D afim de  $\mathbb{R}^4$ , seja  $d_{ij} = \|x_i - x_j\|^2$  para cada  $i, j \leq 5$ . Então,*

$$\begin{vmatrix} 0 & d_{12}^2 & d_{13}^2 & d_{14}^2 & d_{15}^2 & 1 \\ d_{21}^2 & 0 & d_{23}^2 & d_{24}^2 & d_{25}^2 & 1 \\ d_{31}^2 & d_{32}^2 & 0 & d_{34}^2 & d_{35}^2 & 1 \\ d_{41}^2 & d_{42}^2 & d_{43}^2 & 0 & d_{45}^2 & 1 \\ d_{51}^2 & d_{52}^2 & d_{53}^2 & d_{54}^2 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{vmatrix} = 0.$$

Note que o Teorema 2 acima é apresentado para o caso em que  $n = 5$ . Em seu livro, Blumenthal formula o seguinte resultado: o volume  $n$ -dimensional  $V_n$  de um  $n$ -simplex em  $\mathbb{R}^n$  com uma matriz de comprimento lateral  $d = (d_{ij}|i, j \leq n + 1)$  é:

$$V_n^2 = \frac{(-1)^{n-1}}{2^n(n!)^2} \Delta.$$

**Teorema 3** (Menger). *Um espaço semi-métrico não vazio  $S$  é congruentemente integrável em  $\mathbb{R}^n$  (mas não em qualquer  $\mathbb{R}^r$  para  $r < n$ ) se, e somente se: (a)  $S$  contiver um subconjunto de  $(n + 1)$  pontos  $S'$  que é congruente com um subconjunto de  $(n + 1)$  pontos independente de  $\mathbb{R}^n$ ; e (b) cada subconjunto de  $(n + 3)$  pontos  $U$  de  $S$  contendo  $S'$  é congruente com um subconjunto de  $(n + 3)$  pontos de  $\mathbb{R}^n$ .*

Menger havia sido fortemente influenciado pelo movimento de formalização do método matemático, encabeçado por David Hilbert na década de 20, no desenvolvimento dos seus axiomas de espaços métricos. Ele organizou uma série de seminários entre 1929 e 1937, e em meio aos seus estudantes, um nome recebe destaque. Kurt Gödel, que no curso de sua vida exilou-se em Princeton, nos EUA, tornando-se “amigo das caminhadas matutinas de Albert Einstein”, foi responsável pelo desenvolvimento de resultados importantíssimos neste campo. Entretanto, os 3 teoremas acima apresentados são de notável relevância pois carregam em si o conceito central da Geometria de Distâncias.

**Palavras-chave:** *volume de um  $n$ -simplex, Determinante de Cayley-Menger, espaços métricos, geometria de distâncias.*

## Referências

- [1] L. Liberti; C. Lavor, Six mathematical gems from the history of Distance Geometry, *International Transactions in Operational Research*, (2015), 1-22.
- [2] Heath, Thomas L. (1921). A History of Greek Mathematics (Vol II). Oxford University Press. pp. 321–323.
- [3] A. Cayley. A theorem in the geometry of position. *Cambridge Mathematical Journal*, II:267–271, 1841.
- [4] K. Menger. Untersuchungen uber allgemeine Metrik. *Mathematische Annalen*, 100:75–163, 1928.
- [5] L. Blumenthal. Theory and Applications of Distance Geometry. Oxford University Press, Oxford, 1953.
- [6] Coxeter, H.S.M. (1973). Regular Polytopes (3rd ed.). New York: Dover. p120,121.

## Teste da malha para uma simulação em LES

**Luis Fernando Camponogara\***      **Felipe Denardin Costa**

Fundação Universidade Federal do Pampa - Campus Alegrete

97546-550, Alegrete, RS

E-mail: lfcamponogara@gmail.com,      E-mail: fdenardin@unipampa.edu.br

### RESUMO

Uma simulação numérica, é composta por três etapas o pré-processamento, a escolha do solver e o pós-processamento. Durante o pré-processamento é determinado as condições iniciais do sistema (condições de contorno), a determinação da geometria, a criação da malha para o sistema em análise e a determinação das propriedades do material. Já na escolha do solver é determinado as equações que serão resolvidas, os modelos físicos a serem adotados e com isso dado início a simulação numérica. No pós-processamento, é realizado os gráficos de análise dos dados [3]. Segundo [4] um escoamento turbulento é caracterizado em três regiões: Sub-camada viscosa, onde as forças viscosas predominam no escoamento; camada externa, onde as forças cisalhantes predominam o escoamento; e a camada intermediária ou superposição, onde os dois tipos de cisalhamento são importantes. O presente trabalho tem como objetivo principal realizar a análise de validação de uma malha para uma simulação de grandes turbilhões através do uso da condição de parede. A condição de parede é expressa pela lei interna de parede (LoW, do inglês *Law of the Wall*, Eq.(1)) e pela lei logarítmica da camada intermediária Eq.(2).

$$u^+ = \frac{u}{u^*} = \frac{Y u^*}{\nu} = y^+, \quad (1)$$

$$\frac{u}{u^*} = \frac{1}{\kappa_*} \ln \frac{Y u^*}{\nu} + B, \quad (2)$$

onde  $u^+$  representa a velocidade adimensionalizada,  $u$  a velocidade do escoamento [m s<sup>-1</sup>],  $u^*$  representa a velocidade de atrito [m s<sup>-1</sup>],  $Y$  representa a espessura da camada externa [m],  $y^+$  representa a espessura da camada externa adimensionalizada,  $\nu$  representa a viscosidade cinemática [m<sup>2</sup> s<sup>-1</sup>],  $\kappa_*$  representa a constante de Von Kármán, a qual  $\kappa_* \approx 0,41$  e  $B$  representa uma constante adimensional com valor  $\approx 5$ .

O domínio computacional adotado para a realização da simulação possui as mesmas dimensões da seção de teste II do túnel de vento “Professor Joaquim Blessmann”, o qual se encontra no Laboratório de Aerodinâmica das Construções (LAC) da Universidade Federal do Rio Grande do Sul (UFRGS). As configurações do túnel de vento, assim como um modelo esquemático do mesmo podem ser encontrado em [1]. As dimensões da seção de teste são 2 m, 0,9 m, 1,32 m. A escolha desse domínio se justifica pela possibilidade de se realizar experimentos práticos comprovando validando a simulação. O software de dinâmica dos fluidos computacional (CFD, do inglês *Computational Fluid Dynamics*) utilizado para realizar as simulações foi o OpenFOAM 2.4.0, ele possui um conjunto de bibliotecas ("solvers"), os quais podem ser editados devido ao mesmo possuir seu código ser aberto.

---

\*Mestrandos do programa de Pós Graduação em Engenharia Mecânica (PPEng)

Para esse mesmo domínio foi testado 2 condições de malha, sendo a primeira com 70 divisões em  $x$ , 90 em  $y$  e 70 em  $z$ , para esse primeiro caso não foi adicionado um fator de refinamento junto a superfície. Enquanto que, para o segundo caso a malha apresenta 70 divisões em  $x$ , 180 em  $y$  e 70 em  $z$ , para esse segundo caso foi adicionado um fator de refinamento junto a superfície inferior de 20 vezes. Desta forma a altura no ponto mais próximo a superfície inferior é de 0,000785472 m e as maiores células apresentam uma altura de 0,0157094 m.

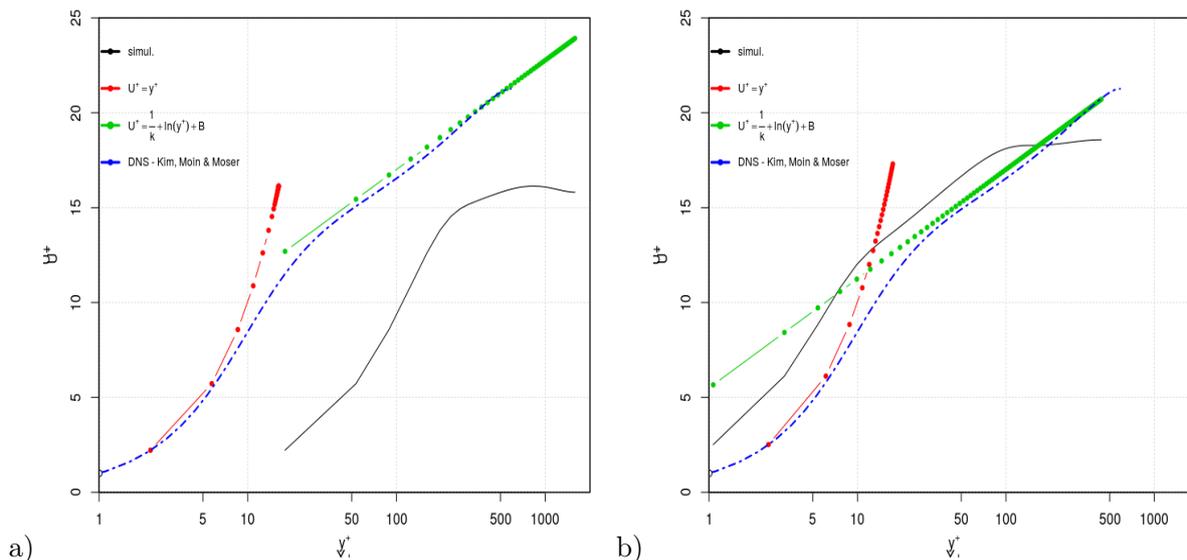


Figura 1: Condição de parede para as 2 malhas testadas. a) Malha sem refinamento. b) Com refinamento nas proximidades da parede.

Analisando os gráfico para lei de parede, observa-se no primeiro caso que a linha preta, que representa a simulação realizada, se encontra bem afastada da curva teórica (linha vermelha e verde) e da em DNS (do inglês *Direct Numerical Simulation*) (linha azul). Os dados da simulação em DNS foram apresentada por [2], enquanto para o segundo caso observa-se que o comportamento é muito similar ao da curva teórica porém ela esta super estimando um pouco a curva teórica. A influencia da diferença do refinamento nas bordas pode ser observada também nos perfis de velocidade do vento dentro da camada limite (figura não mostrada).

**Palavras-chave:** CFD, LoW, Simulação Numérica

## Referências

- [1] UFRGS, LAC. Disponível em: < <http://www.ufrgs.br/lac/TV%20zoom.htm>>. Acessado em: 17 abril de 2019.
- [2] KIM, J.; MOIN, P.; MOSER, R. Turbulence statistics in fully developed channel flow at low reynolds number. *Journal of Fluid Mechanics*, Cambridge University Press, v. 177, p. 133–166, 1987.
- [3] TU, J.; YEOH, G.-H.; LIU, C. *Computational fluid dynamics: a practical approach*. [S.l.]: Butterworth-Heinemann, 2011.
- [4] WHITE, F. M. *Mecânica dos Fluidos-6*. [S.l.]: AMGH Editora, 2010.

# ESTRUTURA DA CAMADA LIMITE ESTÁVEL NA REGIÃO DO PAMPA GAÚCHO E SUA REPRESENTAÇÃO POR MODELOS NUMÉRICOS

**Arlindo D. Carvalho Jr**

**Rafael Maroneze**

Universidade Federal de Santa Maria, Programa de Pós Graduação em Meteorologia, UFSM  
97105-900, Santa Maria, RS

E-mail: arlindocarvalhojr@gmail.com, rafaelmaroneze@gmail.com,

**Luis Fernando Camponogara**

Universidade Federal do Pampa - Programa de Pós-Graduação em Engenharias UNIPAMPA  
97546-550, Campus Alegrete, Alegrete, RS

E-mail: lfcamponogara@gmail.com

**Luiz Eduardo Medeiros**

**Felipe Denardin Costa**

Universidade Federal do Pampa,- Docentes do Programa de Pós-Graduação em Engenharias  
97546-550, Campus Alegrete, Alegrete, RS

E-mail: luizmedeiros@unipampa.edu.br, felipecosta@unipampa.edu.br,

## RESUMO

O presente trabalho visa fazer uma comparação entre os modelos de coluna atmosférica (MCA) E-1 e Boulac com dados adquiridos que foram coletados no sítio experimental estava localizado em uma área de 24 ha pertencente ao Bioma Pampa em Santa Maria, RS, Brasil (29.7241° S, 53.759°). Foram medidos os fluxos de superfície e lançadas radiossondas a cada 3h permitindo combinar essas informações para analisar o comportamento das variáveis turbulentas nos regimes de camada limite convectiva (CLC) e camada limite estável (CLE). Os modelos numéricos têm grande dificuldade em modelar o regime fortemente estável devido aos ventos muito fracos, fazendo com que a velocidade de fricção chegue a valores muito baixos não realísticos. Para isso usaremos as funções de estabilidade que procuram contribuir para que se mantenha mais turbulência nos modelos.

Nas equações prognósticas dos modelos se assume que a componente vertical média do vento é nula, ou seja,  $\bar{w} = 0$ , considera-se a atmosfera seca, ( $q = 0$ ), onde  $q$  é a umidade, e a atmosfera sendo horizontalmente homogênea, para isso consideram-se nulas as derivadas horizontais, chega-se no seguinte sistema de equações para as variáveis médias do escoamento :

$$\frac{\partial \bar{u}}{\partial t} = f(\bar{v} - v_g) - \frac{\partial(\overline{u'w'})}{\partial z} \quad (1)$$

$$\frac{\partial \bar{v}}{\partial t} = f(u_g - \bar{u}) - \frac{\partial(\overline{v'w'})}{\partial z} \quad (2)$$

$$\frac{\partial \bar{\theta}}{\partial t} = - \frac{\partial(\overline{w'\theta'})}{\partial z}. \quad (3)$$

Onde  $\bar{u}$ ,  $\bar{v}$ ,  $\bar{\theta}$  são os componentes de velocidade zonal, meridional e a temperatura potencial, respectivamente, e  $u_g$ ,  $v_g$ , e  $f$  são as componentes do vento geostrófico e o parâmetro de coriolis, respectivamente. Os fluxos de momentum  $\overline{u'w'}$ ,  $\overline{v'w'}$ , e o fluxo de calor  $\overline{w'\theta'}$  são momentos estatístico de 2ª ordem de valores desconhecidos, assim como se desconhece a variação das

componentes médias do vento e da temperatura no tempo. Então é feita uma parametrização dos fluxos de momentum e de calor como sendo proporcional ao gradiente das variações médias, que são componentes já conhecidas, assim consegue-se resolver o sistema de equações.

Esse processo denominado teoria-K. Onde  $K_m$  é o coeficiente de difusão turbulenta, determinado nos nossos modelos por  $K_m = (\alpha e)^{1/2} l_m$ , onde  $\alpha$  é uma constante,  $e$  é a energia cinética turbulenta, e  $l_m$  é o comprimento de mistura, geralmente a maior dificuldade dos modelos é a boa estimativa do coeficiente de difusão turbulenta.

A diferença entre os modelos E-1 e Boulac está na determinação do comprimento de mistura. Para o caso e-1 tem-se

$$l_m = \frac{kz}{1 + kz/\lambda_0} f_m \quad (4)$$

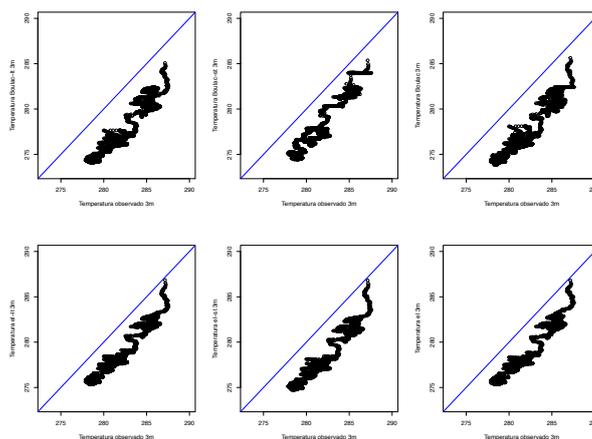
onde  $z$  é a altura da camada,  $\lambda_0$  é dado por  $\lambda_0 = 0,0004\sqrt{u_G^2 + v_G^2} f^{-1}$ , e  $f_m$  é a função de estabilidade, que é definida de duas formas: cauda curta e cauda longa. Para o caso Boulac, tem-se:

$$l_{up} = \int_Z^{Z+l_{up}} \beta(\theta(Z) - \theta(Z')) dZ' \quad (5)$$

$$l_{down} = \int_{Z-l_{down}}^Z \beta(\theta(Z') - \theta(Z)) dZ' \quad (6)$$

onde  $\beta$  é o coeficiente de empuxo,  $Z$  e  $Z'$  são níveis de altura.

Figura 1: Erro de temperatura a 3m.



A figura mostra o erro de temperatura comparando os modelos e suas respectivas funções de estabilidade com os dados observados a 3m, mostrando que os modelos subestimam as temperaturas, pois os gráficos aparecem inteiramente pro lado observado. Geralmente o Boulac é mais turbulento na camada limite planetária, o que é bom para casos de forte estabilidade atmosférica.

**Palavras-chave:** *Camada limite estável, Modelo de coluna atmosférica, observações*

## Referências

- [1] G.A. Degrazia, et al, "An overview of the micrometeorological field campaign at Santa Maria, Southern Brazil: the Pampa?2016 experiment." *Meteorological Applications* 25.3 (2018): 435-444.
- [2] R.T. McNider, et al, "Predictability of the stable atmospheric boundary layer." *Journal of the atmospheric sciences* 52.10 (1995): 1602-1614.
- [3] R.B. Stull, *An introduction to boundary layer meteorology*. Vol. 13. Springer Science & Business Media, 2012.

## Análise das propriedades psicométricas da adaptação transcultural da Escala de Atitude do Lazer em idosos

**Natália R. Almeida\***

**Naiara C. Ap. dos Santos**

**Eraldo S. Silva**

Universidade Estadual de Maringá, Departamento de Estatística, DES/UEM,

87020-900, Maringá, PR

E-mail: nataliarochaalmeida@hotmail.com

naicaroline2@gmail.com

eraldoschunk@gmail.com

**Vivian C. de Castro**

**Lígia Carreira**

Universidade Estadual de Maringá, Dpto de Enfermagem

CEP: 87020-900, Maringá, PR

E-mail: vivian.carla5@hotmail.com,

ligiacarreira.uem@gmail.com.

### RESUMO

Os primeiros estudos científicos a respeito do lazer são da década de 60 no campo da sociologia. Em que, entende-se como lazer um conjunto de atividades escolhidas livremente, nas quais proporcionam prazer, bem estar, condições de recuperação psicossomática e de desenvolvimento pessoal e social. Por outro lado, no campo da psicologia, o estudo do lazer está pautado no processo de desenvolvimento pessoal como uma continuidade de etapas: a vida adulta, maturidade e velhice [3].

O lazer, além de direito constitucional, é considerado uma necessidade humana básica. No entanto, as atividades produtivas e/ou remuneradas são as mais valorizadas pela sociedade e o fato dos idosos não exercerem tais atividades pode fazê-los sentir uma necessidade maior de lazer e interação social. Um dos fatores determinantes do envolvimento em atividades de lazer é a atitude face ao lazer, em que uma atitude positiva tende a estar relacionada com maior envolvimento nestas atividades. Assim, a mensuração da atitude face ao lazer torna-se essencial por estruturar intervenções para promoção de atitude positiva e, portanto, bem-estar biopsicossocial [2].

Desta forma, o presente estudo tem por objetivo verificar as atitudes dos idosos, em relação ao lazer por meio do instrumento *Leisure Attitude Scale* de Ragheb e Beard (1982), submetido a adaptação transcultural [1, 6, 8] e composto por 27 itens com respostas na escala *Likert* variando de 1 a 5, apropriando-se da metodologia de modelagem de equações estruturais. Para verificar o grau com que os fatores encontrados satisfazem a estrutura esperada realizou-se análise fatorial confirmatória (AFC). Para tanto, utilizou-se como índices de qualidade do ajustamento o  $\chi^2/gl$  (razão qui-quadrado e graus de liberdade), CFI (*comparative fit index*), GFI (*goodness of fit index*) e RMSEA (*root mean square error of approximation*) [5]. As análises foram realizadas no *software* R (versão 3.5.0) [7] e os diagramas de caminhos no *software* AMOS (versão 24) [4]. Participaram do estudo 384 idosos atendidos e cadastrados nas unidades básicas de saúde (UBS) da cidade de Maringá-PR até o ano de 2017, sendo 241(62.8%) mulheres e 143(37.2%) homens, com média de idade de  $70.9 \pm 7.64$  anos. Em sua maioria, os participantes eram casados ou em união estável (61, 2%), residiam com familiares (86, 2%), possuíam entre 1 e 4 anos de estudo (50, 5%) e renda pessoal de até um salário mínimo (48, 9%).

A amostra apresentou tamanho adequado para a análise fatorial e adequabilidade do ajuste do modelo, avaliada por meio do índice de *Kaiser-Meyer-Olkin* ( $KMO = 0.92$ ), foi excelente.

\*Agradecimento à CAPES pelo apoio financeiro.

Assim, por meio da Análise fatorial exploratória (AFE), com rotação Varimax, reteve-se três fatores, os quais explicam ao menos 5% da estrutura de variância e covariância dos dados e são bastante próximos aos retidos por Ragheb e Beard (1982). O primeiro fator foi operacionalizado pelos itens 13 a 24, explicando 34.60% da variabilidade total dos dados e foi interpretado como sendo o domínio “Sentimento em relação ao lazer (sentimento)”. O segundo fator explica 7,25% da variabilidade total dos dados sendo este operacionalizado pelos 12 primeiros itens do instrumento e interpretado como o domínio “Pontos de vista e crenças sobre o lazer (crenças)”. Por fim, o terceiro fator foi operacionalizado pelos itens 26, 28, 29, 34 e 35 explicando 6.10% da variabilidade total, o qual foi interpretado como sendo o domínio “Intenções de realizar atividades de lazer (intenções)”.

O modelo que busca explicar a relação entre os fatores sentimento, crenças e intenções apresentou correlações positivas entre os três fatores. Destes, sentimento e crença possui uma maior correlação quando comparada com as demais, indicando que quanto mais o indivíduo acredita nos benefícios das atividades de lazer, melhor ele se sentirá ao praticá-la. Os índices  $\chi^2/gl$ , CFI, GFI, RMSEA e  $p - value$  ( $H_0 : rmsea \leq 0.05$ ), mostram que o modelo proposto tem ajuste aceitável [5], sendo um entre alguns possíveis modelos que podem explicar as inter-relações entre Sentimento em relação ao lazer, Pontos de vista e crenças sobre o lazer e as Intenções de realizar atividades de lazer.

**Palavras-chave:** *Modelagem de Equações Estruturais, Análise Fatorial Confirmatória, Validação de Construto, Lazer de Idosos.*

## Referências

- [1] CASTRO V. C de; Adaptação transcultural e avaliação das propriedades psicométricas da leisure attitude scale para a pessoa idosa. 2019. 127 f. Tese (Programa de Pós-graduação em Enfermagem - PSE), Universidade Estadual de Maringá, Maringá, 2019.
- [2] FREIRE T., FONTE C. Escala de atitudes face ao lazer em adolescentes e jovens adultos. *Paidéia* (Ribeirão Preto). 2007;17(36):79-87.
- [3] GASPARI, J. C. de; SCHWARTZ, G. M. O idoso e a ressignificação emocional do lazer. *Psicologia: teoria e pesquisa*, SciELO Brasil, p. 069–076, 2005.
- [4] IBM®SPSS®. AMOS 24.0. Amos Development Corporation: Meadville-PA USA, 1983-2013. URL <http://amosdevelopment.com/>.
- [5] MARÔCO, J. Análise de equações estruturais: Fundamentos teóricos, software e aplicações. [S.l.]: ReportNumber, Lda, 2010.
- [6] RAGHEB, M. G.; BEARD, J. G. Measuring leisure attitude. *Journal of Leisure Research*, 14(2), p. 155-167, 1982.
- [7] R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [8] TEIXEIRA, A.; FREIRE, T. The Leisure Attitude Scale: psychometrics properties of a short version for adolescents and young adults. *Leisure/Loisir*, 37, 1, 2013.

## Uso de delineamento composto central para otimização de processo de transferência de calor

**Renato P. Mari    Amanda Schuch    Edna Evaristo    Gabrielle C. Kratz**

Universidade Federal de Santa Catarina - Depto de Engenharia Têxtil, DET

89036-256, Campus Blumenau, SC

E-mail: [renato.poli.mari@gmail.com](mailto:renato.poli.mari@gmail.com),

**José Alexandre B. Valle    Rita C. S. C. Valle    July S. P. Filippi**

Universidade Federal de Santa Catarina – Depto. de Engenharia Têxtil, DET

Programa de Pós-Graduação em Engenharia Têxtil - PGETEX

89036-256, Campus Blumenau, SC

### RESUMO

Sistemas de transferência de calor são utilizados em muitas indústrias, quer seja para aquecer ou resfriar fluidos que serão utilizados nos mais diferentes processos. Seu funcionamento baseia-se no princípio da existência de um gradiente de temperatura que age como força motriz desta transferência térmica [2]. O custo associado a um processo de transferência térmica geralmente é elevado, fato que justifica a importância do controle preciso de cada variável a fim de obter-se a maior eficiência do processo [7].

O planejamento experimental é o subcampo da estatística preocupado com a otimização da informação em investigações [4]. A metodologia do planejamento composto central rotacional (DCCR) consiste em um grupo de procedimentos, estatísticos e matemáticos, que podem ser usados no estudo das inter-relações entre uma ou mais respostas (variáveis dependentes) com fatores (variáveis independentes) e seus respectivos níveis [1]. A aplicação deste conjunto de técnicas permite a realização de um número reduzido de ensaios com informações que permitem chegar à otimização do processo, além dos cálculos de erro e determinação dos resíduos envolvidos nos testes[5].

Neste trabalho utilizou-se um trocador de calor de placas paralelas com sentido dos fluidos contracorrente. Foram admitidas como variáveis independentes temperatura de entrada do fluido quente (Tqent), vazão do fluido quente (Qq) e vazão do fluido frio (Qf) e como resposta a quantidade de calor média trocada (qm) e eficiência (Ef%) utilizando as relações apresentadas no livro Transferência de Calor: Um texto básico - Ozisik [6], sendo medido nos ensaios a temperatura de entrada do fluido frio (Tfent), as temperaturas de saída do fluido quente (Tqsai) e do fluido frio (Tfsai). O DCCR totalizou 23 ensaios com realização de triplicata no ponto central e valor de  $\alpha = \pm 1,68$ . Os níveis estudados de cada variável -1 e +1 foram 40 e 60 K; 2,4 e 6,4 L/min, e 2,4 e 6,4 L/min para Tqent, Qq e Qf, respectivamente.

Foram determinados os coeficientes de regressão para a resposta de interesse do processo e realizou-se a análise de variância (ANOVA) a um nível de significância de 5%. Os dados experimentais foram ajustados num modelo polinomial (Eq. 1), considerando os componentes lineares e de interação que apresentam efeito significativo sobre a resposta [3].

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_i \sum_j \beta_{ij} x_i x_j \quad (\text{Eq1})$$

Os modelos estatísticos de regressão obtidos para qm e Ef são apresentados nas equações 2 e 3, respectivamente.

$$qm = 3710 + 1048Qq - 251Qq^2 + 665Qf - 194Qf^2 + 1393Tqent + 376QqQf + 443QqTqent + 205 QfTqent \quad (Eq2)$$

$$Ef\% = 0,624 + 0,111Qq - 0,031Qq^2 - 0,149 Qf + 0,017Qf^2 + 0,010Tq + 0,011 QqQf \quad (Eq3)$$

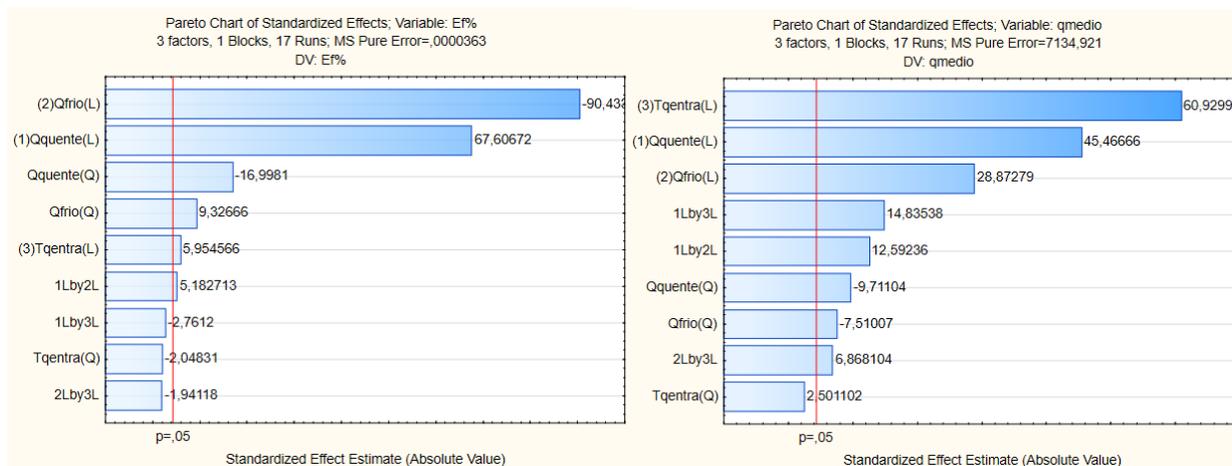


Figura 1: diagrama de Pareto dos efeitos das variáveis temperatura do fluido quente (Tqent), vazão do fluido quente (Qq) e vazão do fluido frio (Qf) sobre qm (a) e Ef% (b).

O uso do planejamento experimental composto central rotacional permitiu a obtenção de informações relevantes sobre o sistema de troca térmica usando trocador a placas paralelas. Com um número reduzido de ensaios foi possível à obtenção de modelos estatísticos quadráticos que podem ser usados para prever respostas com valores diferentes daqueles testados experimentalmente. A análise de dados por meio de um software estatístico também foi de extrema importância para o tratamento dos dados, trazendo confiabilidade para o estudo além de permitir conclusões rápidas e eficientes.

**Palavras-chave:** *Transferência de Calor, Planejamento Experimental, Software Estatístico*

## Referências

[1] B. Barros Neto, R. E. Bruns, I. S. Scarminio, Como fazer experimentos: pesquisa e desenvolvimento na ciência e na indústria. [s.l.]: Porto Alegre: Bookman, 2010.

[2] A. C Ferreira, Estudo comparativo da troca térmica em trocadores de calor dos tipos casco/tubos e de placas Unoesc & Ciência - ACET, p. 57-66, Edição Especial 201

[3] D.C. Montgomery, G. C. RUNGER, “Estatística aplicada e probabilidade para engenheiros”, 2. ed. Rio de Janeiro: LTC, 2003, 463 p.

[4] J. Morgan, X. Deng, Experimental design. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, v.2, n. 2, p. 164-172. 2012.

[5] C. A. Nunes, et al. The use of statistical software in food science and technology: Advantages, limitations and misuses Food Research International n. 75 p. 270–280, 2015.

[6] M. N. Ozisik, “Transferência de Calor: Um texto básico”. Guanabara Koogan, 1990. 661p.

[7] Textília Máquina têxtil, disponível em <[http://www.textilia.net/materias/ler/textil/maquina-textil/uso\\_racional\\_da\\_energia\\_na\\_industria\\_textil](http://www.textilia.net/materias/ler/textil/maquina-textil/uso_racional_da_energia_na_industria_textil)>, acessado em 12/04/2019

## Análise numérica do método quase linear que modela a equação telegráfica reativa difusiva

Eliandro Rodrigues Cirilo

Thiago Ferreira de Oliveira\*

Neyva Maria Lopes Romeiro

Paulo Laerte Natti

Centro de Ciências Exatas, Departamento de Matemática, UEL, Londrina, PR

E-mail: ercirilo@uel.br, thiagoxp41@gmail.com, nromeiro@uel.br, plnatti@uel.br.

### RESUMO

A equação telegráfica reativa difusiva, que pode modelar alguns processos de invasão biológica, eventualmente é dada sob a forma [2]:

$$\tau \frac{\partial^2 S}{\partial t^2} + \left(1 - \tau \frac{dF(S)}{dS}\right) \frac{\partial S}{\partial t} = D \frac{\partial^2 S}{\partial x^2} + F(S) \quad (1)$$

onde  $S = S(x, t)$  é a densidade populacional ao longo do espaço unidimensional  $x$  no tempo  $t$ . O termo  $D$  é chamado de coeficiente de difusão e  $\tau$  é uma constante de retardo temporal. Neste trabalho considera-se o modelo exponencial, escrito como  $F(S) = kS$ , cuja constante  $k$  exprime o crescimento de densidade.

Ainda que a equação (1) tenha solução analítica para o modelo exponencial, submetida a condições iniciais e de contorno, a abordagem numérica computacional para se obter as soluções se faz importante, pois quando utilizados os modelos Logístico ou Allee a equação não tem ainda solução analítica.

O código computacional desenvolvido parte da hipótese de que a discretização do domínio unidimensional  $x$  é como o mostrado na Figura 1. Nesta figura os pontos discretos são rotulados

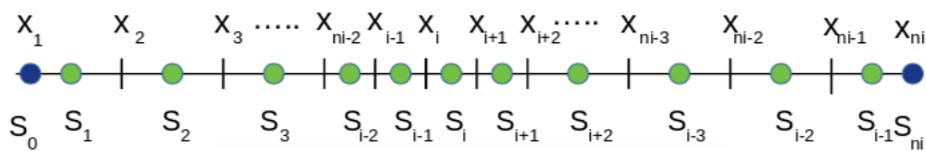


Figura 1: Domínio computacional

e enumerados por  $x_i$  com  $i = 1, \dots, ni$ , o termo  $S_i$  com  $i = 0, \dots, ni$  é centrado no intervalo discreto. Nos pontos  $S_0$  e  $S_{ni}$  são aplicadas as condições de contorno, e para os demais pontos ( $i = 1, \dots, ni - 1$ ) é calculada a equação diferencial parcial para cada lapso temporal  $\Delta t$ , a partir do estado inicial  $t = 0$ .

Para a discretização da equação (1), aplicando-se o método das diferenças finitas [1], resulta que

$$S_i^{k+1} = \left(\frac{1}{C_i + \tau \tilde{C}_i}\right) \left[C_{i-1} S_{i-1}^{k+1} + C_{i+1} S_{i+1}^{k+1} + \tilde{b} + \tau \bar{b}\right] \quad (2)$$

\*bolsista de mestrado CAPES

onde os índices superiores  $k, k + 1$  designam momentos temporais, e

$$\begin{aligned} C_i &= \frac{1}{\Delta t} + \frac{2D}{\Delta x^2}, & \tilde{C}_i &= \frac{1}{\Delta t^2} - \frac{1}{\Delta t} \left. \frac{dF}{dS} \right|_i^{k+1}, \\ C_{i-1} &= D/\Delta x^2, & C_{i+1} &= D/\Delta x^2, \\ \tilde{b} &= F_i^{k+1} + \frac{1}{\Delta t} S_i^k, & \bar{b} &= \left[ \frac{2}{\Delta t^2} - \frac{1}{\Delta t} \left. \frac{dF}{dS} \right|_i^{k+1} \right] S_i^k - \frac{1}{\Delta t^2} S_i^{k-1}. \end{aligned}$$

Ressalta-se que o conjunto de equações algébricas não lineares (2) é localmente linear no nível iterativo  $k + 1$ . É justamente deste fato que surge a nomenclatura quase linear. Desta forma a análise numérica, com o modelo exponencial, permite a validação do método. A validação corrobora com o desenvolvimento de estudos futuros, em especial para os modelos Logístico e efeito Allee.

**Palavras-chave:** *Equação telegráfica, método quase linear, invasão biológica*

## Agradecimentos

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

## Referências

- [1] Cuminato, J.A.; Meneguette J., M. “Discretização de Equações Diferenciais Parciais: Técnicas de Diferenças Finitas”. Rio de Janeiro: SBM, 2013
- [2] E. E. Holmes. Are Diffusion Models too Simple? A Comparison with Telegraph Models of Invasion, *The American Naturalist*, 1993. DOI:10.1086/285572.

## Simulações Numéricas com um Problema Não-Linear de Advecção-Difusão

**Márcio Rodolfo Fernandes**

Universidade Federal de Santa Catarina - Departamento de Matemática  
88040-900, Campus Trindade, Florianópolis, SC  
E-mail: marcio.fernandes@ufsc.br

**RESUMO:** Esse trabalho descreve uma combinação de métodos de discretização e linearização para a realização de simulações numéricas com um problema não-linear de advecção-difusão.

**Palavras-chave:** *Métodos Numéricos, Elementos Finitos, Petrov-Galerkin*

### 1. Introdução

O transporte e o armazenamento de petróleo têm um risco significativo de vazamentos acidentais com grandes consequências. Tal risco torna-se maior em zonas costeiras e especialmente indesejável em regiões onde indústrias petroquímicas coexistam com atividades pesqueiras e/ou turísticas.

Acidentes deste tipo têm ocorrido ao longo do tempo com grandes consequências. Assim, o planejamento de políticas de prevenção e minimização dos efeitos de tais vazamentos constitui um dos desafios da comunidade científica, sendo o conhecimento do movimento e da área atingida por essas manchas de petróleo após um vazamento, parte das informações necessárias para a decisão das ações a serem tomadas.

Segundo Fay [2], a evolução do processo se divide em três fases. Inicialmente, o regime dominante é o gravitacional-inercial que dura por volta de uma hora ou mais, dependendo do óleo, e se caracteriza pelo fato dele ainda não ter se misturado à água. Em seguida, temos o regime gravitacional-viscoso no qual o deslocamento da mancha se deve principalmente aos efeitos combinados dos ventos e correntes marítimas, durando por volta de duas semanas, dependendo do volume do derrame. Finalmente, ocorre o regime dominado pela viscosidade e pela tensão superficial, em que a mancha já não pode ser vista a olho nu, estando o petróleo totalmente misturado com a água e podendo durar meses.

Segundo Benqué [1], a fase representada pelo regime gravitacional-viscoso pode ser descrita por uma equação não-linear de advecção-difusão, juntamente com condições iniciais e de contorno adequadas. Este modelo será apresentado a seguir.

### 2. Modelagem

$$\frac{\partial u}{\partial t} - c\Delta(u^3) + \vec{\beta}(x,t) \cdot \nabla u = f(x,t), \quad \Omega \times I \quad (1)$$

$$u(x,t) = g(x), \quad x \in \partial\Omega_-, \quad t \in I \quad (2)$$

$$\frac{\partial u(x,t)}{\partial \vec{\eta}} = 0, \quad x \in \partial\Omega_+, \quad t \in I \quad (3)$$

$$u(x,0) = u_0(x), \quad x \in \Omega \quad (4)$$

onde  $I = (0, T] \subset \mathfrak{R}$ ,  $\Omega \subset \mathfrak{R}^2$  tem sua fronteira dividida em duas partes:  $\partial\Omega = \partial\Omega_+ \cup \partial\Omega_-$ ,  $\partial\Omega_+ = \{x \in \partial\Omega / \vec{\eta}(x) \cdot \vec{\beta}(x,t) \geq 0\}$ ,  $\partial\Omega_- = \{x \in \partial\Omega / \vec{\eta}(x) \cdot \vec{\beta}(x,t) < 0\}$ ;  $\vec{\eta}(x)$  é o vetor unitário normal à fronteira do domínio;  $u$  representa a altura da mancha medida em relação ao nível da água;  $f$  é uma fonte de poluentes;  $\vec{\beta}$  é um vetor que combina os efeitos de arrasto dos ventos, correntes e marés, enquanto  $c$  é um coeficiente utilizado para modificar a intensidade da difusão e para mudar o caráter da equação.

### 3. Metodologia

Existem várias possibilidades de combinação entre métodos de linearização e discretização, no espaço e no tempo, para o problema apresentado. Optamos por discretizar o tempo por diferenças finitas e o espaço via elementos finitos. A linearização foi feita de forma simples, aproveitando a discretização da variável temporal e o fato de que  $\Delta(u^3) = \text{div}(3u^2 \nabla u)$ :

$$\frac{u^n - u^{n-1}}{k_n} - \text{div}(3c(u^{n-1})^2 \nabla u^n) + \vec{\beta}(x, t_n) \cdot \nabla u^n = f(x, t_n), u^0(x) = u_0(x). \quad (5)$$

$u^n(x)$ : aproximação para  $u(x, t_n)$ ,  $k_n = t_n - t_{n-1}$  é passo local no tempo.

Do ponto de vista teórico, a equação (1) apresenta caráter parabólico ou hiperbólico, dependendo de sua própria solução. Isso se deve à não-linearidade do termo de difusão que pode variar de um valor estritamente positivo para zero e vice-versa, dependendo do valor da solução. Em tais regiões, não conhecidas *a priori*, transições rápidas da solução podem ocorrer, levando ao aparecimento de oscilações espúrias na aproximação numérica. Para contornar esse problema, a discretização espacial de (5) foi feita através de um método do tipo Petrov-Galerkin, isto é, um método de elementos finitos que usa funções-teste do tipo  $v + \delta \vec{\beta} \cdot \nabla v$ , com  $\delta$  escolhida localmente e dependente da relação entre a difusão e a advecção em cada elemento.

### 4. Resultados e Conclusões

Simulamos, inicialmente, num cenário com solução exata conhecida:  $\Omega = [-3,5;3,5] \times [-1;1]$ ,  $c = 1$ ,  $\vec{\beta} = (1,0)$ ,  $f \equiv 0$ ,  $g \equiv 0$ ,  $k_n = k = 5 \times 10^{-2}$  e condição inicial dada por

$$u_0(x) = \sqrt{\max\left\{(0,25)^2 - \frac{1}{18}(x_1^2 + x_2^2), 0\right\}}.$$

Tempo	t=0,5	t=1	t=1,5	t=2	t=2,5	t=3
Erro Absoluto	0,00266	0,00278	0,00282	0,00287	0,00349	0,00455

Tabela 1: Erro absoluto entre as soluções exata e aproximada

Ressaltamos que algoritmo não apresentou as oscilações indesejáveis presentes em soluções via métodos de elementos finitos na forma padrão. Além disso, a solução computada teve boa concordância com a solução exata do problema, nesse cenário de simulação, como podemos observar pelos dados da Tabela 1.

Após realizar várias outras simulações, em diversos cenários possíveis, pretendemos passar a situações mais reais, levando em consideração o cálculo do escoamento incompressível para a obtenção do campo de velocidades  $\vec{\beta}$ , por intermédio das equações de Navier-Stokes. A partir disso, o escoamento da água deixará de ser dado, passando a ser uma das incógnitas do problema, que agora passará a ser um sistema contendo as equações de Navier-Stokes juntamente com a equação não-linear de advecção-difusão, que governa o espalhamento propriamente.

### Referências

- [1] J. P. Benqué, A. Hauguel & P. L. Viollet, Engineering Application of Computational Hydraulics, Pitman Advanced Publishing Program, London, 1982.
- [2] J. A. Fay, The Spread of Oil Slicks on a Calm Sea, Oil on the Sea, D. P. Hoult Plenum Press, pp 53-63, 1969.

## **Circulação dirigida pelo vento em uma lagoa de águas rasas: a Lagoa Mirim, Rio Grande do Sul, Brasil**

**Antonio R. R. Bendô\***      **Eduardo de P. Kirinus**

Universidade Federal do Rio Grande, FURG – Escola de Engenharia, PPGE0  
96201-900, Rio Grande, RS

E-mail: rayltonbendo@gmail.com      ekirinus@gmail.com

**Aldair Forster**      **Juliana Costi**      **Wiliam C. Marques**

Universidade Federal do Rio Grande, FURG – Departamento de Matemática, Física e Estatística, IMEF  
96201-900, Rio Grande, RS

E-mail: aldairforster@gmail.com      ju.costi@gmail.com      wilianmarques@furg.br

**Thaísa Beloti Trombetta**

Universidade Federal do Rio Grande, FURG – Departamento de Oceanologia.  
thaisa\_bt@hotmail.com

### **RESUMO**

O presente estudo teve por objetivo investigar a oscilação no nível de água da Lagoa Mirim (LM) durante eventos extremos de vento, visando dar suporte a tomada de decisões frente a futuras operações de dragagens na região. Para isso, foi utilizado o modelo hidrodinâmico TELEMAC-2D. A simulação percorreu um período de cinco anos (2000-2004), e o domínio computacional foi forçado pelas descargas dos principais rios e pela ação dos ventos.

O modelo TELEMAC-2D resolve as equações de Saint-Venant para escoamentos em superfície livre aceitando a hipótese de equilíbrio hidrostático. Além disso, o mesmo considera a evolução da superfície livre em função do tempo e utiliza as equações de advecção-difusão para o transporte das concentrações dos traçadores (salinidade, temperatura ou sedimentos em suspensão) [1]. A turbulência foi parametrizada com o modelo k- $\epsilon$ , que resolve as equações de transporte e dissipação da energia cinemática turbulenta. Esse modelo é amplamente usado em lagoas costeiras e na plataforma continental sul do Brasil [2] [3] [4].

Para investigar a variabilidade do padrão de circulação da LM causada por ventos, então, foram identificados cinco eventos extremos de ventos (I, II, III, IV e V). Essa escolha foi feita, de acordo com a magnitude e direção do vento que gerou impacto na elevação da superfície de água da LM. Os cenários foram identificados em diferentes estações do ano durante o período simulado, demonstrando que a região possui alta variabilidade sazonal do vento. A validação do modelo hidrodinâmico, foi feita através da comparação entre séries temporais de níveis medido e o resultado da modelagem. Essa comparação resultou em uma boa concordância (coeficiente de correlação de Pearson  $> 0.83$ ), desse modo, os resultados mostram que o modelo hidrodinâmico é bem sucedido em representar a variabilidade temporal do nível da água.

Os resultados mostram que em média o nível de água é cerca de 0.10 m maior na região sul da LM em comparação com o norte, esta configuração é uma consequência da predominância de ventos do quadrante nordeste, que represa a água proveniente dos rios, criando uma altura de água superior na parte sul da lagoa. Por outro lado, durante eventos extremos esse padrão pode se inverter em consequência da configuração adotada pelo vento, como observado nos eventos I, II, III e IV, onde os ventos eram predominantemente de sudoeste, sul e sudeste, desse modo provocaram variações na superfície da água da LM de -1 m na parte sul e até +1 m na parte norte. Já durante o evento V, o vento era de nordeste e causou uma elevação de 0.5 m da parte sul em relação ao norte.

A utilização de modelagem numérica se mostrou eficaz para fornecer uma melhor compreensão da circulação em lagoas costeiras de águas rasas, bem como fornecer importantes compreensões ambientais sobre eventos extremos para implementação de futuras operações de dragagem na LM.

**Palavras-chave:** *Modelagem Hidrodinâmica, Circulação, Eventos extremos de vento*

## Referências

- [1] Hervouet, J.M. “Hydrodynamics of free surface Flows. Modelling with the finite element method”. John Wiley & Sons, Ltda, West Sussex, 2007.
- [2] Kirinus, E.P., Marques, W. C. Viability of the application of marine current power generators in the south Brazilian shelf. Applied Energy, vol. 155, pp. 23-34, (2015).
- [3] Marques, W. C., Stringari, C. E., Kirinus, E. P., Möller Jr, O. O., Toldo Jr, E. E., Andrade, M. M. Numerical modeling of the Tramandaí beach oil spill, Brazil—Case study for January 2012 event. Applied Ocean Research, vol. 65, pp. 178-191, (2017).
- [4] Costi, J., Marques, W. C., de Paula Kirinus, E., de Freitas Duarte, R., Arigony-Neto, J. Water level variability of the Mirim-São Gonçalo system, a large, subtropical, semi-enclosed coastal complex. Advances in Water Resources, 117, 75-86, (2018).

## Discretização de Equações Diferenciais Parabólicas

**Bryan Aoliabe Siqueira\***

Departamento de Matemática, UFSC

89036-002 Blumenau - SC

E-mail: bryan.siqueira@grad.ufsc.br

**Louise Reips**

Universidade Federal de Santa Catarina - Departamento de Matemática

88036-002, Campus Blumenau, Blumenau, SC

E-mail: l.reips@ufsc.br

### RESUMO

Com aplicações em diversas áreas, as equações diferenciais parciais são objeto de estudo de inúmeras pesquisas, pois permitem, através de seu uso, a criação de modelos matemáticos hábeis para descrever de forma realista, muitos fenômenos biológicos, físicos, econômicos, etc. Tais modelos matemáticos são testados através da simulação computacional, que permite analisar, de forma quantitativa, os dados obtidos através da comparação entre a simulação e o sistema o qual desejou-se modelar. Para tal, o domínio em estudo deve ser discretizado através da criação de uma malha refinada de pontos, de modo satisfatório a garantir uma boa aproximação dos resultados simulados, obtendo um comportamento semelhante ou igual ao esperado.

Neste trabalho temos como objetivo o estudo de equações parabólicas, que são equações de evolução e modelam fenômenos que evoluem de modo irreversível, com o tempo, e tendo assim, uma das variáveis com um caráter temporal que a distingue das demais. Esse aspecto serve de motivação para o estudo do caso bidimensional, onde usaremos a variável  $x$  com sentido espacial e  $t$  como variável evolutiva. Buscamos aqui obter soluções utilizando o método de diferenças finitas (MDF), discretizando o domínio e substituindo as derivadas presentes na equação diferencial por aproximações que envolvem somente valores numéricos da função.

O modelo fundamental das equações parabólicas é a equação dada por

$$u_t - a(x, t)u_{xx} = r(x, t), \quad (1)$$

Para esse trabalho, consideraremos a condição inicial:

$$u(x, t) = \psi(x),$$

sujeito às condições de fronteira:

$$\begin{aligned} u(0, t) &= f(t), \\ u(L, t) &= g(t), \end{aligned}$$

onde  $a(x, t) > 0$  e constante,  $x \in [0, L]$  e  $t \in [0, T]$  e  $r(x, t) = 0$ . Para as aproximações numéricas utilizaremos os métodos Explícito e Implícito fazendo um comparativo entre ambos. Além disso, apresentaremos a respectiva análise de Estabilidade, Erro de Truncamento Local e Erro Global para o problema em questão.

**Palavras-chave:** *Equações Diferenciais Parciais, Equações Parabólicas, Diferenças Finitas*

---

\*Licenciando em Matemática pela Universidade Federal de Santa Catarina - Campus Blumenau

## **Referências**

- [1] J. A. Cuminato and M. M. Junior, Discretização de equações diferenciais parciais: técnicas de diferenças finitas - Rio de Janeiro: SBM, 2013.
- [2] M. C. C. Cunha, Métodos numéricos - 2ª edição - Campinas, SP: Editora da Unicamp, 2000.

# IMPLEMENTAÇÃO EM PYTHON DO MODELO SEIR ESTOCÁSTICO APLICADO À PROPAGAÇÃO DE INFORMAÇÃO

Jamile S. dos Santos      Daniel G. Alfaro Vigo

Departamento de Ciência da Computação, IM-UFRJ,  
Programa de Pós-Graduação em Informática (PPGI),  
Universidade Federal do Rio de Janeiro (UFRJ),  
21941-590 Rio de Janeiro, RJ, Brasil

E-mail: jamilesantosdossantos@gmail.com/ dgalfaro@dcc.ufrj.br

## RESUMO

O objetivo desse trabalho é apresentar o modelo *SEIR estocástico* e realizar sua implementação em Python, esse modelo consiste em uma adaptação do modelo *SEIR* apresentado em [1]. Em trabalhos futuros, o mesmo será usado para realizar inferências sobre a propagação de informação em diversas redes sociais e fazer comparações com outros modelos conhecidos.

O processo de transmissão de doenças numa epidemia tem sido modelado por ferramentas matemáticas determinísticas e probabilísticas. Em 1927, Kermack e McKendrick criaram o modelo *SIR* para o estudo de doenças infecciosas, que descreve a população total ( $N$ ) dividida em suscetíveis ( $S$ ), infectados ( $I$ ) e recuperados ( $R$ ).

A analogia existente entre o processo de disseminação de uma doença contagiosa e o de propagação de uma informação oferece a possibilidade de descrever-las por modelos matemáticos semelhantes. Nesse sentido, em [3] foi usada uma versão simplificada do modelo *SIR*, consistindo em um sistema com duas EDOs, para descrever a propagação de informação em mídias sociais.

Seguindo a mesma ideia, em [1] foi introduzido o modelo *SEIR* (determinístico), também baseado no modelo *SIR*, para descrever a propagação boca-a-boca de informação no Flickr (site de compartilhamento de imagens). O comportamento dinâmico do processo de transmissão se dá com uma população total ( $N$ ), dividida em suscetíveis ( $S$ ), expostos ( $E$ ), infectados ( $I$ ) e recuperados ( $R$ ) com taxas de contato:  $\alpha$  (probabilidade de um usuário suscetível, com pelo menos um amigo infectado, ser exposto ao conteúdo),  $\beta$  (probabilidade de que um usuário exposto marcar uma foto e tornar-se infectado. Os usuários expostos têm probabilidade  $1 - \beta$  para passar para o estado recuperado) e  $\sigma$  (probabilidade de um usuário remover uma foto infectada de sua lista de favoritos, ficando recuperado). O modelo consiste em um sistema de quatro equações de diferenças para as grandezas envolvidas.

No modelo *SEIR estocástico*, desenvolvido nesse trabalho aplicando as técnicas descritas em [2], a dinâmica desses grupos da população total é descrita como uma cadeia de Markov de ordem 1. As probabilidades das transições entre esses grupos é dada por:

$$\begin{cases} P((S_{t+\Delta t}, E_{t+\Delta t}, I_{t+\Delta t}) - (S_t, E_t, I_t) = (-1, 1, 0)) = \alpha \frac{S_t I_t}{N} \Delta t + o(\Delta t), \\ P((S_{t+\Delta t}, E_{t+\Delta t}, I_{t+\Delta t}) - (S_t, E_t, I_t) = (0, -1, 1)) = \beta \gamma E_t \Delta t + o(\Delta t), \\ P((S_{t+\Delta t}, E_{t+\Delta t}, I_{t+\Delta t}) - (S_t, E_t, I_t) = (0, -1, 0)) = (1 - \beta) \gamma E_t \Delta t + o(\Delta t), \\ P((S_{t+\Delta t}, E_{t+\Delta t}, I_{t+\Delta t}) - (S_t, E_t, I_t) = (0, 0, -1)) = \sigma I_t \Delta t + o(\Delta t), \end{cases} \quad (1)$$

onde  $S_t$ ,  $E_t$ ,  $I_t$  representam a quantidade de indivíduos suscetíveis, expostos e infectados no tempo  $t$ . Esse modelo coincide com o *SEIR* determinístico apresentado em [1] quando fazemos  $\Delta t = 1$ ,  $\gamma = 1$  e desprezamos as flutuações aleatórias.

Para ilustrar a dinâmica do modelo SEIR estocástico foi desenvolvido um código em linguagem Python usando algoritmos clássicos para a simulação de processos aleatórios [4]. A figura 1 mostra o resultado de 5 simulações da propagação de uma foto, considerando uma população total ( $N$ ) igual a 5.000 usuários e os parâmetros do modelo como  $\alpha = \beta = 0,2$  (representa 20% de chance de um usuário marcar uma foto como favorita),  $\sigma = 0$  e  $\gamma = 1$ . O resultado da nossa simulação nos mostra que a população de suscetíveis e infectados, não sofrem grandes alterações depois de aproximadamente 200 dias.

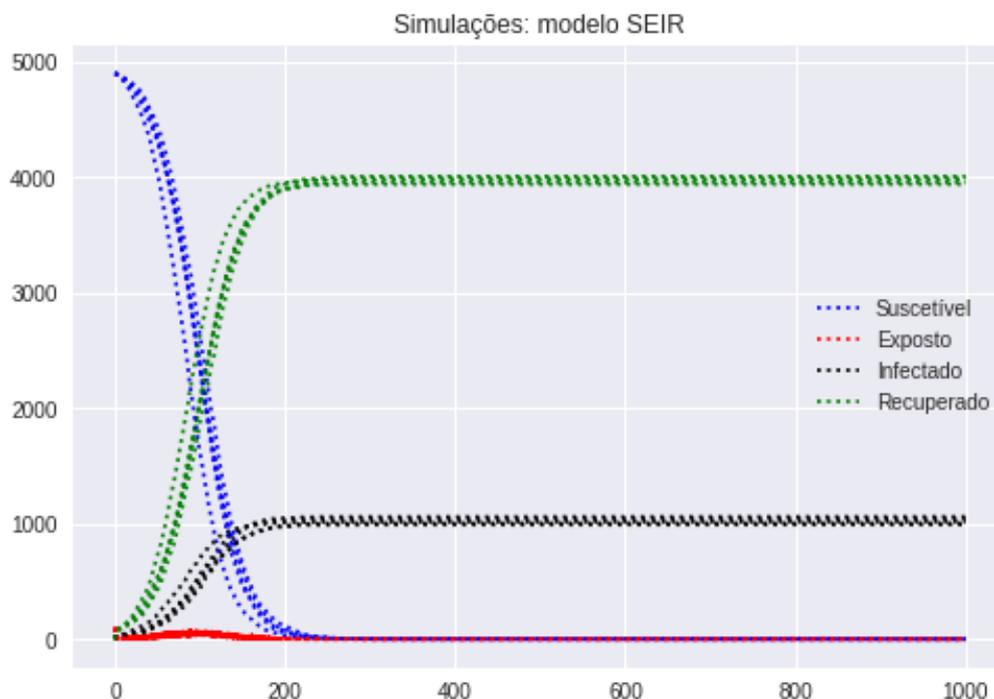


Figura 1: Resultados das simulações usando o modelo SEIR estocástico

**Palavras-chave:** *Propagação de informação, Modelo SEIR, Modelagem estocástica*

## Referências

- [1] M. Cha, F. Benevenuto, Y-Y Ahn, K.P. Gummadi, Delayed information cascades in Flickr: Measurement, analysis, and modeling, *Computer Networks*, 56 (2012), 1066-1076.
- [2] P.E. Greenwood and L.F. Gordillo, Stochastic epidemic modeling, em “Mathematical and statistical estimation approaches in epidemiology” (G. Chowell, J.M. Hyman, L.M.A. Bettencourt e C. Castillo-Chavez, eds.) pp. 31–52. Springer, Dordrecht, 2009..
- [3] A.S. Quadros, J.V. Valério, and D. Alfaro Vigo, Adaptação do modelo SI de epidemiologia para propagação de informação, em “Anais do XXXIII Congresso da Sociedade Brasileira de Computação”, v.1, pp. 1495–1500, 2013.
- [4] Sheldon M. Ross, “Simulation”, Academic Press, 2013.

## Estudo da evolução de tumores sólidos através de Equações Diferenciais

**Luiz Augusto S. França\***

Departamento de Engenharias (ENG), Campus Blumenau, UFSC  
89036-002, Blumenau, SC  
E-mail: scheuermannfranca@gmail.com

**Louise Reips**

Universidade Federal de Santa Catarina - Departamento de Matemática  
89036-002, Campus Blumenau, Blumenau, SC  
E-mail: l.reips@ufsc.br

### **RESUMO**

O câncer é uma doença caracterizada pelo crescimento desordenado de células que invadem os órgãos e tecidos. Tal proliferação anormal das células provoca a formação de uma massa celular, que é denominada tumor e que pode evoluir de modo rápido e agressivo (neoplasia). Um diagnóstico precoce permite a detecção e acompanhamento da atividade de tumores malignos, possibilitando, assim, um melhor tratamento e possivelmente uma maior taxa de sobrevivência dos pacientes. Baseado nisso, o objetivo desta pesquisa é o estudo de equações diferenciais aplicadas à modelagem do crescimento de tumores sólidos, se direcionando mais especificamente, à Equação de Gompertz [1]. A metodologia constou na simulação computacional da Equação de Gompertz utilizando-se do software MATLAB. Para tal, os parâmetros iniciais utilizados foram retirados da bibliografia consultada [2] [3], o que possibilitou a comparação entre os dados obtidos e os já publicados na literatura. Após a simulação dos sistemas tumorais, houve o arcabouço para propor a inserção de fatores na equação, os quais representam o modelo matemático, de modo a descrever os tratamentos quimioterápico [4] e radioterápico aplicados à equação, bem como a análise dos efeitos desses. Nesse sentido, simulou-se a aplicação destes fatores na equação e fez-se a comparação gráfica da evolução tumoral sem os mesmos. Mediante a análise dos resultados obtidos, conclui-se que a Equação de Gompertz oferece uma base sólida para a modelagem de sistemas tumorais, além de que a inserção de fatores de tratamento pode, de fato, auxiliar no retardamento da evolução patológica.

**Palavras-chave:** *Câncer, Equação de Gompertz, Simulação computacional, Tratamento.*

### **Referências**

- [1] E. W. Boyce, R. C. DiPrima, “Equações diferenciais elementares e problemas de valores de contorno”, Rio de Janeiro, LTC, 2006.
- [2] J. S. Domingues, Análise do Modelo de Gompertz no crescimento de tumores sólidos e inserção de um fator de tratamento, *Biomatemática*, 21 (2011) 103-112.
- [3] G. F. C. Filho, N. A. de Aguiar Neto, C. M. da Silva, Solução da equação diferencial de Gompertz na análise de crescimento de tumores malignos da próstata, *Rev. de Eng. e Pesquisa Aplicada*, 2 (2016) 551-555.
- [4] D. B. Rodrigues, “Terapia Antiangiogênica de tumores utilizando células produtoras de endostatina encapsuladas em dispositivos de imunisolamento”, Dissertação de Mestrado, Instituto de Pesquisas Energéticas e Nucleares, São Paulo, 2008.



## Demonstração do Problema do Caixeiro Viajante utilizando Robótica e Algoritmos Genéticos

**Roger Luis Hoff Lavarda**      **Lisiane Reips\***

Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul - Campus Ibirubá  
Bacharelado em Ciência da Computação  
98200-000, Ibirubá, RS

E-mail: roger.lavarda@ibiruba.ifrs.edu.br,      lisiane.reips@ibiruba.ifrs.edu.br

### **RESUMO**

A necessidade de realizar tarefas de forma ótima e no menor tempo possível vem acompanhando a evolução dos processos tecnológicos, tornando, cada vez mais, necessário trabalhar com Problemas de Otimização Combinatória (POC). Um exemplo de POC é o Problema do Caixeiro Viajante (PCV), o qual foi o tema escolhido para a fundamentação do presente trabalho.

O PCV é considerado um Problema de Otimização Combinatória[1] e, devido à sua importância em implicações práticas, os estudos sobre os algoritmos, para resolver tal problema, iniciaram-se muito cedo. Algoritmos Genéticos (AGs) foram utilizados, pela primeira vez, por Goldberg e Grefenstette, em 1985, para resolver o PCV[3, 6].

O objetivo em resolver o PCV é encontrar o caminho mais curto entre uma cidade e outra, dentre uma determinada lista de cidades, onde o caixeiro viajante visita as cidades apenas uma vez, e depois retorna para a cidade inicial. Esse problema, computacionalmente, é considerado NP-hard[7, 2], também conhecido como NP-complexo.

O PCV é muito utilizado como modelo em encaminhamento de veículos[9], problemas de programação[4], problemas de mapeamento físico[5], entre outros.

Para resolver o PCV, utiliza-se AGs, que possuem um diferencial com relação aos demais algoritmos.

AGs fazem parte da classe dos Algoritmos Evolutivos (AEs), que geram soluções para problemas de otimização, através de técnicas inspiradas na evolução natural, como herança, mutação, seleção[8].

Como metodologia, utilizou-se a análise de literaturas e bibliografias, com o intuito de definir relações entre o tema escolhido e possíveis soluções já encontradas; o desenvolvimento de um robô, que demonstra na prática a solução para o PCV; a utilização de AGs para que o robô percorra caminhos escolhidos pelo usuário e a análise dos resultados obtidos, a fim de concluir o quão satisfatória fora a pesquisa. O desenvolvimento do trabalho encontra-se na fase de estudo da literatura e planejamento.

Para solucionar o PCV, o robô será programado com, no mínimo, três algoritmos que percorrerão as rotas, identificando o menor custo possível, de tempo e distância. Para que isso ocorra, o robô identificará, através de marcos visuais, pontos coloridos, em uma base quadrangular de madeira. O robô calculará a distância entre um ponto e outro e guardará a informação de cada distância, para posterior cálculo de tempo e distância mínimos.

Como o robô estará em uma área desconhecida, ele será informado sobre cada marco visual, através de um sensor de cor. Então o robô armazenará as cores para que consiga identificar a distância de uma cor à outra.

Após o aprendizado das cores e distâncias, o robô irá partir do seu ponto de saída, perfazendo o menor trajeto possível, na menor distância, percorrendo todos os pontos coloridos, sem passar duas vezes pelo mesmo ponto e então retornará ao seu ponto de saída.

A utilização do robô para demonstrar os percursos a serem seguidos, com soluções ótimas ou quase ótimas, foi uma opção devido à eficácia determinada pela precisão dos seus sensores, bem como a inteligência nas estratégias de planejamento de caminhos.

**Palavras-chave:** *Problema do Caixeiro Viajante, Robótica, Algoritmos Genéticos.*

## Referências

- [1] A. Philip, A. A. Taofiki, O. Kehinde, A genetic algorithm for solving travelling salesman problem, *Int. J. Adv. Comput. Sci. Appl.*, vol. 2, pp. 26-29, (2011).
- [2] C. H. Papadimitriou, K. Steiglitz, "Combinatorial Optimization: Algorithms and Complexity", Prentice Hall of India Private Limited, India, 1997.
- [3] D.E. Goldberg, J.R. Lingle, Alleles, loci and the traveling salesman problem, em "Proceedings of the First International Conference on Genetic Algorithms and Their Applications" pp. 154-159, Pittsburgh PACarnegie Mellon University, Pittsburgh, 1985.
- [4] D. Whitley, T. Starkweather, F.D. Ann, Scheduling Problems and Traveling Salesmen: The Genetic Edge Recombination Operator, em "Proceedings of the 3rd International Conference on Genetic Algorithms" pp. 133-140, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 1989.
- [5] F. Alizadeh, R.M. Karp, R.M. Newberg, et al., Physical mapping of chromosomes: a combinatorial problem in molecular biology, em "Proceedings of the fourth annual ACM-SIAM Symposium on Discrete algorithms " pp. 371-381, Society for Industrial and Applied Mathematics Philadelphia, PA, USA, 1993.
- [6] J. J. Grefenstette, R. Gopal, B. Rosmaita, et al., Genetic algorithm for TSP, em "Proceedings of the First International Conference on Genetic Algorithms and Their Applications" pp. 160-168, Pittsburgh PACarnegie Mellon University, Pittsburgh, 1985.
- [7] M. R. Garey, D. S. Johnson, "Computers and Intractability: A Guide to the Theory of NP-Completeness", Freeman W.H., San Francisco, 1979.
- [8] Shen, Y.-C., Hsu, C.-H., & Hsieh, S.-H., Integrated genetic algorithms and cloud technology to solve travelling salesman problem on Hadoop, em " Proceedings of the 4th IEEE International Conference on Cloud Computing Technology and Science" vol. 1, pp. 566-569, IEEE Computer Society, USA, 2012.
- [9] S. Lin, Computer solutions of the traveling salesman problem, *Bell Syst. Tech. J.*, vol. 10, pp. 2245-2269, (1965).

## Um Modelo de Fluxo de Potência Ótimo para Agrupamento Virtual de Geradores Distribuídos

**Guilherme Renkel Wehmuth\***      **Hugo Jose Lara Urdaneta**  
**Tiago D. Curi Busarello**

Universidade Federal de Santa Catarina - Departamento de Engenharias  
89036-002, Campus Blumenau, SC

E-mail: guilherme.renkel@grad.ufsc.br, hugo.lara.urdaneta@ufsc.br, tiago.busarello@ufsc.br

### RESUMO

O constante aumento na demanda por energia tem impulsionado a busca por melhorias na transmissão da rede elétrica, garantindo assim o desenvolvimento de ferramentas que façam uso otimizado dos recursos energéticos e financeiros disponíveis. A fim de selecionar de maneira ótima os diversos parâmetros que compõem a rede elétrica, um modelo amplamente utilizado é o de fluxo de potência ótimo (FPO). Nesse trabalho, temos como objetivo estudar problemas de otimização oriundos do planejamento de despacho de potência em um operador de sistema de distribuição (Distribution System Operator(DSO)) [4,5]. Sendo tais problemas modelados em termos de equações de fluxo de potência, podemos descrever o sistema elétrico como um conjunto de  $n$  barras interligadas, por onde uma potência complexa, descrita na forma  $S = P + jQ$ , é transmitida. Entre cada uma das barras há uma admitância  $y(i, k)$  associada, às quais, em conjunto, geram uma matriz de admitância  $Y(n, n)$ . A matriz de admitância, para fins de cálculo, é normalmente separada em uma matriz  $G(n, n)$ , que contém a parte real, e uma matriz  $B(n, n)$ , a qual possui a parte complexa de  $Y$ .

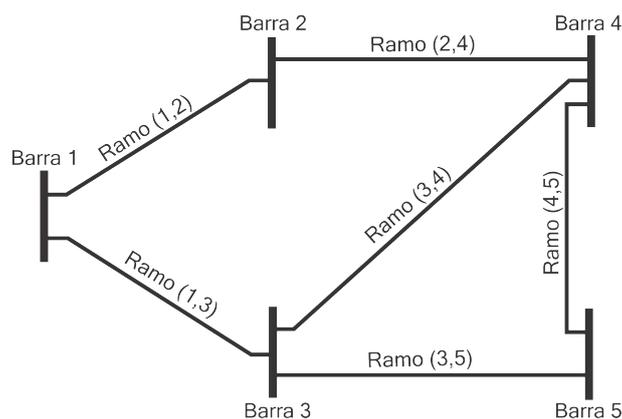


Figura 1: Sistema elétrico composto por 5 barras

Considerando o estado operativo da rede em regime permanente senoidal, a modelagem matemática é feita a fim de contabilizar a potência complexa em cada um dos elementos que compõem o sistema, além do fasor de tensão em cada uma das  $n$  barras. Conforme apresentado

\*bolsista de Iniciação Científica PIBIC/CNPq

na referência [4], o seguinte modelo não linear é gerado:

$$P_i(V, \delta) = V_i \sum_{k=1}^N V_k (G_{ik} \cos(\delta_i - \delta_k) + B_{ik} \sin(\delta_i - \delta_k)) \quad (1)$$

$$Q_i(V, \delta) = V_i \sum_{k=1}^N V_k (G_{ik} \sin(\delta_i - \delta_k) - B_{ik} \cos(\delta_i - \delta_k)) \quad (2)$$

As equações de potência fazem parte das restrições do modelo de programação não linear a ser resolvido, que calcula valores de injeção de potência, em barras predeterminadas, e cuja função objetivo buscar minimização das perdas e a redução de custos. O modelo foi implementado utilizando a linguagem de programação JULIA e a biblioteca JuMP [2]. Ela foi escolhida por ser uma linguagem open-source de alto nível, o que garante uma participação muito efetiva da comunidade científica, constantes melhorias e alta performance computacional.

O trabalho é a primeira etapa de um projeto que visa integrar as ferramentas de FPO no contexto de Smart Grids[1,3].

**Palavras-chave:** *Fluxo Ótimo de Potência, Otimização, Sistema Elétrico.*

## Referências

- [1] H. M. Ayres D. Salles W. Freitas "A Practical Second-Order Based Method for Power Losses Estimation in Distribution Systems With Distributed Generation" IEEE Trans. Power Syst vol. 29 no. 2 pp. 666-674 Mar. 2014.
- [2] I. Dunning, J. Huchette, and M. Lubin, "Jump: A modeling language for mathematical optimization", SIAM Review, vol. 59, no. 2, pp. 295-320, 2017.
- [3] J. Yu Y. Jiao X. Wang M. Ni "Bi-level distributed optimal dispatch of micro grid clusters based on mutual communication" 2015 5th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT) pp. 2480-2485 2015.
- [4] S. Frank, S. Rebennack, "A Primer on Optimal Power Flow: Theory, Formulation, and Practical Examples", Working Papers 2012-14, Colorado School of Mines, Division of Economics and Business.
- [5] T. D. C. Busarello, H. J. L. Urdaneta, A. Péres, J. A. Pomilio and M. G. Simões, "Cluster Formation Strategy as Ancillary Service of a Battery Storage System in a Secondary Distribution System," 2018 13th IEEE International Conference on Industry Applications (INDUSCON), São Paulo, Brazil, 2018, pp. 467-473.

## DMDGP: Um Problema Real

Guilherme Philippi\*      Felipe Fidalgo

Universidade Federal de Santa Catarina  
89036-004, Campus Blumenau, Blumenau, SC  
E-mail: guilherme.philippi@grad.ufsc.br,    felipe.fidalgo@ufsc.br

### RESUMO

Existe uma relação muito forte com a forma geométrica das moléculas orgânicas e suas funções em organismos vivos [1]. Sabe-se que uma das estruturas principais da vida é construída com os aminoácidos que formam as proteínas. Logo, para conhecer a estrutura geométrica dessas moléculas, é preciso estudar a sua geometria [2]. Existe uma ferramenta muito importante para quem trabalha nesta área: um repositório online contendo dados de todas as proteínas catalogadas chamado *Worldwide Protein Data Bank* (ou wwPDB).

O problema de encontrar as posições dos átomos de uma molécula, tendo como entradas algumas distâncias entre átomos próximos (obtidas através de experimentos de Ressonância Magnética Nuclear [3]), é conhecido na literatura como *Molecular Distance Geometry Problem* (MDGP), que é uma particularização do *Distance Geometry Problem* (DGP) [4]. Tal problema, munido de uma ordem conveniente para percorrer seus átomos (dada pelo *Discretization Vertex Order Problem*, ou simplesmente, DVOP), pode ser discretizado, gerando o *Discretizable MDGP* (DMDGP), como segue formalmente definido [5].

**Discretizable Molecular Distance Geometry Problem (DMDGP):** Dados um grafo ponderado e não-direcionado  $G = (V, E, d)$ , onde  $d : E \rightarrow \mathbb{R}_+$ , o subconjunto de vértices  $U_0 = \{v_1, v_2, v_3\}$  e uma relação de ordem total em  $V$  que satisfaz a seguinte relação de axiomas:

1.  $G[U_0]$  é um clique em três vértices (iniciando a configuração);
2. para todo vértice  $v_i$  com posto  $i = \rho(v_i) > 3$  nesta ordem,  $G[U_i]$  é uma clique com quatro vértices (ordem de discretização, dada anteriormente) e
3. para cada vértice  $v_i$ , com posto  $i = \rho(v_i) > 3$ , juntamente com  $\{v_{i-3}, v_{i-2}, v_{i-1}\}$ , vale a desigualdade

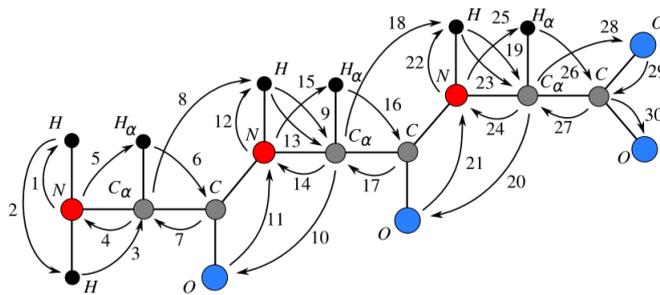
$$d_{i-3, i-1} < d_{i-3, i-2} + d_{i-2, i-1}, \quad (\text{Desigualdade Triangular Estrita})$$

encontre uma imersão  $x : V \rightarrow \mathbb{R}^3$  tal que valha  $\|x(v_i) - x(v_j)\| = d_{i,j}$ ,  $\forall \{v_i, v_j\} \in E$ .

Note que, para ser possível estudar o problema acima, precisamos encontrar uma ordem conveniente nos vértices a partir de algumas informações sobre a geometria da molécula, dadas a priori. Nos interessa aqui, de fato, devido sua periodicidade, a existência de uma subestrutura nas proteínas chamada *Cadeia Principal*, que possui uma geometria rica e bem conhecida [1]. Através de dados experimentais de cristalografia, sabe-se sobre a geometria média dessa subestrutura [6], onde os comprimentos e ângulos entre as ligações dos átomos que a formam são fixas, na média, a menos de erros de medida. Tendo posse de tais informações, pode-se percorrer os átomos da molécula através da Cadeia Principal, repetindo os que possuem propriedades conhecidas, afim de fazer valer os três axiomas do DMDGP. Isto foi feito em [5] propondo o *hand-crafted vertex order*, conforme esboça a Figura 1 (extraída do texto original).

\*bolsista de Iniciação Científica PIBIC/CNPq

Isto é, neste trabalho, buscou-se estudar a geometria de distâncias aplicada a geometria molecular, visando produzir realizações válidas das posições geométricas dos átomos em uma molécula de proteína. Para o projeto, criou-se um programa para facilitar as simulações do problema, que aceita como entrada instâncias de proteínas do wwPDB e retorna um arquivo com a instância reordenada (ordenação hc) obtendo a discretização do problema.



# Homogeneização da Equação do Calor para Meios Continuamente Microperiódicos via Método das Funções Teste Oscilantes

Júlio C. R. Claeysen, Marcos P. de Lima,

Instituto de Matemática e Estatística, UFRGS,  
91509-900, Porto Alegre, RS

E-mail: julio@mat.ufrgs.br, marcos.lima@ufrgs.br,

**Leslie D. Pérez-Fernández**

Instituto de Física e Matemática, UFPel,  
96001-970, CP: 354, Pelotas, RS

E-mail: leslie.fernandez@ufpel.edu.br.

**Resumo:** Neste trabalho apresentamos uma aplicação do método de Tartar das funções teste oscilantes, o qual propõe homogeneizar empregando funções teste que variam rapidamente na formulação variacional do problema estudado, de maneira que ao tender o parâmetro geométrico  $\varepsilon$  para zero, toda a formulação contenha termos que convergiram para um limite num certo sentido (forte ou fraco). A aplicação do método de Tartar é descrita para um problema da condução do calor linear, do qual se deduzem expressões para o fluxo de calor efetivo, a condutividade térmica efetiva e a capacidade calorífica efetiva.

**Palavras-chave:** Homogeneização, Método das Funções Teste Oscilantes, Convergência de Sequências de Operadores

## 1 Introdução

Geralmente, os fenômenos físicos associados às propriedades de meios micro-heterogêneos ocorrem na “microescala”. Assume-se que o comprimento característico  $l$  desta escala “microscópica” (aquela em que ocorre a heterogeneidade) é muito maior que aquele da escala molecular, mas muito menor que o comprimento  $L$  característico da escala macroscópica,  $l \ll L$ . Em tais situações, dizemos que o meio heterogêneo apresenta separação das escalas estruturais (caracterizada pelo parâmetro geométrico  $\varepsilon = l/L$ ,  $0 < \varepsilon \ll 1$ ), suas propriedades apresentam variação rápida com relação a posição e que cumpre a hipótese do contínuo, ou seja, ele pode ser visto como um contínuo na escala microscópica e, portanto propriedades efetivas podem lhe ser atribuídas [9]. Mais precisamente, a hipótese de homogeneidade equivalente estabelece que, na macroescala, o meio heterogêneo é fisicamente equivalente a certo meio homogêneo ideal, de maneira que as propriedades efetivas do primeiro são as propriedades constantes do segundo. Assim, ao processo de obtenção do comportamento efetivo do meio micro-heterogêneo dá-se o nome de “homogeneização”. Sob as hipóteses de separação de escalas e do contínuo, podemos especificar dois usos da homogeneização: a obtenção de uma boa aproximação da solução do problema original e a obtenção do comportamento efetivo do meio heterogêneo. Neste trabalho enfoca-se o segundo uso, ou seja, obter grandezas que caracterizam o comportamento macroscópico do meio heterogêneo: as propriedades físicas efetivas. Para entender este processo num sentido matemático, precisamos entender o que é convergência de operadores. Para isso, considere uma sequência de equações da forma

$$A^\varepsilon u^\varepsilon = f \tag{1}$$

com uma única solução  $u^\varepsilon$  pertencente à algum espaço de função  $X$ . Para uma sequência de operadores  $A^\varepsilon$ , obtemos uma sequência correspondente de soluções  $u^\varepsilon$  para estas equações. Se olharmos de um ângulo diferente, isto significa que temos uma sequência de operadores  $A^\varepsilon$  que determina uma sequência de funções  $u^\varepsilon$  tal que a mesma imagem  $f$  é obtida. Se a sequência  $\{u^\varepsilon\}$  tende para um limite  $u^0 \in X$ , ou seja, se

$$u^\varepsilon \rightarrow u^0 \quad (2)$$

em algum sentido razoável, pode-se questionar se existe um operador  $A^0$  que produza a mesma imagem  $f$  desta função limite  $u^0$ , ou seja, se para algum  $A^0$

$$A^0 u^0 = f. \quad (3)$$

Se este é o caso com o mesmo  $A^0$  para qualquer  $f$  definida,  $A^0$  poderia ser visto como o limite de  $\{A^\varepsilon\}$  em relação a capacidade de transformar funções em  $X$ , isto é, no sentido de operadores. Ao pesquisar a equação do limite resolvida pela função limite  $u^0$ , temos que lidar com dois problemas. A primeira é encontrar suposições sob as quais existe um limite  $A^0$ , com propriedades tais que (3) tenha uma solução única. A segunda dificuldade é a da determinação de  $A^0$ , e tem sido estudada principalmente para problemas periódicos. Entende-se por meio periódico, aquele no qual a heterogeneidade pode ser reproduzida mediante a replicação periódica de um elemento recorrente, chamado de célula básica. Aplicaremos o método de funções teste oscilantes, o qual foi desenvolvido por Luc Tartar [7]. Em suma, a partir da formulação variacional de (1) complementada por condições de contorno e iniciais

$$a^\varepsilon(u^\varepsilon, v) = F(v), \quad \forall v \in D(\Omega), \quad (4)$$

obtemos uma forma variacional limite, quando  $\varepsilon \rightarrow 0$ ,

$$a^0(u^0, v) = F(v), \quad \forall v \in D(\Omega). \quad (5)$$

Porém o estudo do limite (4) só é possível graças a um problema auxiliar, o qual tem como solução funções oscilantes. Uma vez que as soluções oscilantes compõem a função teste  $v$  em (4) o limite (5) é obtido.

## 2 Formulação do problema e descrição do método

Para ilustrar o método, iremos homogeneizar um problema para uma equação parabólica que modela a condução do calor em um meio microperiódico caracterizado pelo parâmetro geométrico  $\varepsilon$ :

$$\begin{cases} c\left(\frac{x}{\varepsilon}\right) \frac{\partial u^\varepsilon}{\partial t} - \nabla_x \cdot \left( A\left(\frac{x}{\varepsilon}\right) \nabla_x u^\varepsilon \right) = f(t, x), & t \in (0, T), \quad x \in \Omega \subset \mathbb{R}^N, \\ u^\varepsilon(x, 0) = u_0(x), & x \in \Omega, \\ u^\varepsilon(x, t) = 0, & t \in (0, T), \quad x \in \partial\Omega, \end{cases} \quad (6)$$

onde  $\Omega$  é um aberto conexo,  $Y = [0, 1]^N$  é a célula básica na variável local  $y = x/\varepsilon$ ,  $u_0 \in L^2(\Omega)$ ,  $c(y) \in L^2_\#(Y)$ ,  $|Y| = 1$ , é a capacidade calorífica,  $x$  é a posição,  $t$  é o tempo,  $u^\varepsilon$  é a temperatura,  $A(y) = (a_{ij}(y))_{1 \leq i, j \leq N}$  ( $a_{ij} \in L^2_\#(Y)$ ) é a condutividade térmica e  $f \in L^2((0, T) \times \Omega)$  é a fonte de calor.  $A$  é um componente condutor  $Y$ -periódico, não necessariamente isotrópico e que é limitado e definido positivo, isto é, existem duas constantes positivas  $0 < \alpha \leq \beta < +\infty$  tais que, para qualquer vetor  $\xi \in \mathbb{R}^N$  e em qualquer ponto  $y \in Y$ ,  $\alpha|\xi|^2 \leq A(y)\xi \cdot \xi \leq \beta|\xi|^2$ , e  $c(y)$  é uma função  $Y$ -periódica, positiva e limitada,  $0 < c^- \leq c(y) \leq c^+ < +\infty \forall y \in Y$ . Observe que o problema (6) admite uma única solução  $u^\varepsilon$  no espaço  $L^\infty(0, T; L^2(\Omega)) \cap \mathcal{W}$ , onde

$$\mathcal{W} = \left\{ v \mid v \in L^2(0, T; H_0^1(\Omega)), \frac{\partial v}{\partial t} \in L^2(0, T; H^{-1}(\Omega)) \right\} \quad (7)$$

é um espaço de Banach com respeito à norma

$$\|v\|_{\mathcal{W}} = \|v\|_{L^2(0,T;H_0^1(\Omega))} + \left\| \frac{\partial v}{\partial t} \right\|_{L^2(0,T;H^{-1}(\Omega))}, \quad (8)$$

e que satisfaz a estimativa [2]

$$\|u^\varepsilon\|_{L^\infty(0,T;L^2(\Omega))} + \|u^\varepsilon\|_{\mathcal{W}} \leq C [\|u_0\|_{L^2(\Omega)} + \|f\|_{L^2((0,T)\times\Omega)}]. \quad (9)$$

Isto significa que a sequência  $\{u^\varepsilon\}$ , indexada por uma sequência  $\varepsilon$  convergente a 0, é limitada, na norma do espaço  $L^2(0,T;H_0^1(\Omega))$ . Portanto, existe uma subsequência fracamente convergente para algum  $u^0(t,x) \in L^2(0,T;H_0^1(\Omega))$ , que não depende do parâmetro geométrico  $\varepsilon$ . Assim nosso objetivo é encontrar a equação homogeneizada que  $u^0$  satisfaz.

Em suma, o principal objetivo do método é conhecer o limite, quando  $\varepsilon \rightarrow 0$ , da formulação variacional do problema (6), isto é, de

$$- \int_0^T \int_\Omega \left[ c\left(\frac{x}{\varepsilon}\right) u^\varepsilon(t,x) \frac{\partial \phi}{\partial t} - A\left(\frac{x}{\varepsilon}\right) \nabla_x u^\varepsilon \cdot \nabla_x \phi + f(x,t)\phi(t,x) \right] dxdt = 0 \quad (10)$$

para toda  $\phi(t,x) \in D((0,T) \times \Omega)$ . Infelizmente, o lado esquerdo de (10) envolve o produto de duas sequências que convergem fracamente, o qual não converge para o produto dos limites. Portanto, não podemos passar ao limite em (10) sem argumentos adicionais. Para contornar isto, se constrói certas funções teste que permitem estudar o limite de (10).

Uma vez que a sequência  $\{u^\varepsilon\}$  é limitada na norma de  $L^2(0,T;H_0^1(\Omega))$ , pela aplicação do teorema de Rellich-Kondrachov [3], existem uma subsequência (denotada como  $\varepsilon_k$ ), um  $u^0(t,x) \in L^2(0,T;H_0^1(\Omega))$  e um  $\sigma^0(t,x) \in L^2((0,T) \times \Omega)$ , tais que

$$\begin{cases} u^{\varepsilon_k} \rightharpoonup u^0 \text{ fracamente em } L^2(0,T;H_0^1(\Omega)), \\ u^{\varepsilon_k} \rightarrow u^0 \text{ fortemente em } L^2((0,T) \times \Omega), \\ \sigma^\varepsilon(t,x) \equiv A^\varepsilon(x) \nabla_x u^\varepsilon(t,x) \rightharpoonup \sigma^0(t,x) \text{ fracamente em } L^2((0,T) \times \Omega). \end{cases} \quad (11)$$

Defina

$$\omega_j^\varepsilon(x) = \varepsilon \omega_j\left(\frac{x}{\varepsilon}\right) = x_j - \varepsilon \chi_j\left(\frac{x}{\varepsilon}\right) \quad (12)$$

onde  $\omega_j = y_j - \chi_j(y)$  e  $\chi_j(y) \in W_\#(Y)$  é solução do problema local adjunto

$$\begin{cases} -\nabla_y \cdot (A^t(y) (e_j + \nabla_y \chi_j(y))) = 0, & y \in Y, \\ \chi_j(y), & Y \text{ - periódica,} \end{cases} \quad (13)$$

em que

$$W_\#(Y) = \left\{ v \in H_\#^1(Y) : \mathcal{M}_Y(v) = \frac{1}{|Y|} \int_Y v dy = 0 \right\}. \quad (14)$$

**Teorema 1.** *Seja  $f^\varepsilon(x) = f(x/\varepsilon)$  uma função  $Y$ -periódica em  $L^2(Y)$ . Então quando  $\varepsilon \rightarrow 0$   $f^\varepsilon(x) \rightharpoonup \frac{1}{|Y|} \int_Y f(y) dy$ , fracamente em  $L^2(\Omega)$ .*

Pelo Teorema 1 sabemos que

$$\omega_j^\varepsilon \rightharpoonup x_j \text{ fracamente em } L^2(\Omega) \quad (15)$$

e por (12)

$$\nabla_x \omega_j^\varepsilon = e_j - \nabla_y \chi_j\left(\frac{x}{\varepsilon}\right) \quad (16)$$

em que  $\nabla_y \omega_j$  é  $Y$ -periódica, uma vez que  $\chi_j$  é  $Y$ -periódica. Logo, pelo Teorema 1

$$\nabla_y \omega_j\left(\frac{x}{\varepsilon}\right) \rightharpoonup \mathcal{M}(e_j - \nabla_y \chi_j(y)) = e_j - \mathcal{M}(\nabla_y \chi_j(y)) = e_j \text{ fracamente em } L^2(\Omega). \quad (17)$$

Consequentemente, temos as seguintes convergências:

$$\begin{cases} \omega_j^\varepsilon \rightharpoonup x_j \text{ fracamente em } H^1(\Omega), \\ \omega_j^\varepsilon \rightarrow x_j \text{ fortemente em } L^2(\Omega), \end{cases} \quad (18)$$

onde usamos o Teorema de Rellich-Kondrachov novamente. Agora introduzimos o vetor

$$\eta_j^\varepsilon(x) = A^t \left( \frac{x}{\varepsilon} \right) \nabla_x \omega_j^\varepsilon(x) = A^t \left( \frac{x}{\varepsilon} \right) \nabla_y \omega_j \left( \frac{x}{\varepsilon} \right). \quad (19)$$

Uma vez que  $A^t$  é  $Y$ -periódica,  $A^t \left( \frac{x}{\varepsilon} \right) \nabla_y \omega_j \left( \frac{x}{\varepsilon} \right)$  é  $Y$ -periódica também. Logo, pelo Teorema 1 segue

$$\eta_j^\varepsilon(x) \rightharpoonup \int_Y A^t(y) \nabla_y \omega_j(y) dy = \int_Y A^t(y) (e_j - \nabla_y \chi_j) dy \text{ fracamente em } (L^2(\Omega))^N. \quad (20)$$

Provemos agora que  $\eta_j^\varepsilon$  satisfaz

$$\int_{\Omega} \eta_j^\varepsilon \cdot \nabla_x v dx = 0, \quad v \in H_0^1(\Omega). \quad (21)$$

Para fazer isso, considere  $\varphi \in D(\Omega)$  e defina  $\varphi^\varepsilon(y) = \varphi(\varepsilon y)$ , em quase todo  $\mathbb{R}^N$ . Obviamente  $\varphi^\varepsilon$  pertence a  $D(\mathbb{R}^N)$ . Assim, tem-se

$$\int_{\mathbb{R}^N} A^t(y) \nabla_y \omega_j(y) \nabla_y \varphi^\varepsilon(y) dy = 0. \quad (22)$$

Fazendo uma mudança de variável  $x = \varepsilon y$  segue que

$$\int_{\Omega} A^t \left( \frac{x}{\varepsilon} \right) \nabla_y \omega \left( \frac{x}{\varepsilon} \right) \nabla_y \varphi \left( \frac{x}{\varepsilon} \right) dy = 0, \quad \forall \varphi \in D(\Omega), \quad (23)$$

desde que  $\text{supp } \varphi \subset \Omega$ . Uma vez que  $H_0^1(\Omega)$  é o fecho de  $D(\Omega)$ , segue (21). Considere  $\varphi_1 \in D(\Omega)$ ,  $\varphi_2 \in D(0, T)$ ,  $\varphi_2(0) = 0$ ,  $\varphi_2(T) = 1$  e escolha  $\phi = \varphi_1(x) \varphi_2(t) \omega_j^\varepsilon$  como função teste em (10) e  $\varphi_1 \varphi_2 u^\varepsilon$  como função teste em (21). Temos respectivamente,

$$\begin{aligned} & \int_0^T \int_{\Omega} \sigma^\varepsilon(t, x) \cdot \nabla_x \omega_j^\varepsilon \varphi_1 \varphi_2 dx dt + \int_0^T \int_{\Omega} \sigma^\varepsilon(t, x) \cdot \omega_j^\varepsilon \nabla_x \varphi_1 \varphi_2 dx dt \quad (24) \\ &= \int_0^T \int_{\Omega} \left[ c \left( \frac{x}{\varepsilon} \right) u^\varepsilon \varphi_1 \frac{\partial \varphi_2}{\partial t} \omega_j^\varepsilon + f \varphi_1 \varphi_2 \omega_j^\varepsilon \right] dx dt + \int_{\Omega} u_0 \varphi_1 \omega_j^\varepsilon dx, \quad \forall \varphi_1 \in D(\Omega), \varphi_2 \in D(0, T), \end{aligned}$$

$$\int_0^T \int_{\Omega} \eta_j^\varepsilon \cdot \varphi_2 \nabla_x \varphi_1 u^\varepsilon dx dt + \int_0^T \int_{\Omega} \eta_j^\varepsilon \cdot \varphi_1 \varphi_2 \nabla_x u^\varepsilon dx dt = 0, \quad \forall \varphi_1 \in D(\Omega), \varphi_2 \in D(0, T). \quad (25)$$

Por definição, tem-se

$$\sigma^\varepsilon \cdot \nabla_x \omega_j^\varepsilon = A^\varepsilon \nabla_x u^\varepsilon \cdot \nabla_x \omega_j^\varepsilon = A^{t\varepsilon} \nabla \omega_j^\varepsilon \cdot \nabla_x u^\varepsilon = \eta_j^\varepsilon \cdot \nabla_x u^\varepsilon. \quad (26)$$

Portanto, ao subtrair (25) de (24) obtemos

$$\begin{aligned} & \int_0^T \int_{\Omega} \sigma^\varepsilon(t, x) \cdot \omega_j^\varepsilon \nabla_x \varphi_1 \varphi_2 dx dt - \int_0^T \int_{\Omega} \eta_j^\varepsilon \cdot \varphi_2 \nabla_x \varphi_1 u^\varepsilon dx dt - \int_{\Omega} u_0 \varphi_1 \omega_j^\varepsilon dx \quad (27) \\ &= \int_0^T \int_{\Omega} \left[ c \left( \frac{x}{\varepsilon} \right) u^\varepsilon \varphi_1 \frac{\partial \varphi_2}{\partial t} \omega_j^\varepsilon + f \varphi_1 \varphi_2 \omega_j^\varepsilon \right] dx dt, \quad \forall \varphi_1 \in D(\Omega), \varphi_2 \in D(0, T). \end{aligned}$$

Passando  $\varepsilon \rightarrow 0$ , obtém-se em cada termo de (27)

$$\lim_{\varepsilon \rightarrow 0} \int_0^T \int_{\Omega} \sigma^\varepsilon(t, x) \cdot \omega_j^\varepsilon \nabla_x \varphi_1 \varphi_2 dx dt = \int_0^T \int_{\Omega} \sigma^0(t, x) \cdot x_j \nabla_x \varphi_1 \varphi_2 dx dt, \quad (28)$$

$$\lim_{\varepsilon \rightarrow 0} \int_0^T \int_{\Omega} \eta_j^\varepsilon \cdot \varphi_2 \nabla_x \varphi_1 u^\varepsilon dx dt = \int_0^T \int_{\Omega} \int_Y A^t(y) (e_j - \nabla_y \chi_j) \cdot \varphi_2 \nabla_x \varphi_1 u^0 dy dx dt, \quad (29)$$

$$\lim_{\varepsilon \rightarrow 0} \int_0^T \int_{\Omega} c\left(\frac{x}{\varepsilon}\right) u^\varepsilon \varphi_1 \frac{\partial \varphi_2}{\partial t} \omega_j^\varepsilon dx dt = \int_0^T \int_{\Omega} \int_Y c(y) u^0 \varphi_1 \frac{\partial \varphi_2}{\partial t} x_j dy dx dt, \quad (30)$$

$$\lim_{\varepsilon \rightarrow 0} \int_0^T \int_{\Omega} f \varphi_1 \varphi_2 \omega_j^\varepsilon dx dt = \int_0^T \int_{\Omega} f \varphi_1 \varphi_2 x_j dx dt, \quad (31)$$

$$\lim_{\varepsilon \rightarrow 0} \int_{\Omega} u_0 \varphi_1 \omega^\varepsilon dx = \int_{\Omega} u_0 \varphi_1 x_j dx. \quad (32)$$

Isto é,

$$\begin{aligned} \int_0^T \int_{\Omega} [\sigma^0(t, x) \cdot x_j \nabla_x \varphi_1 \varphi_2 - A_j^* \cdot \varphi_2 \nabla_x \varphi_1 u^0] dx dt &= \int_0^T \int_{\Omega} c^* u^0 \varphi_1 \frac{d\varphi_2}{dt} x_j + f \varphi_1 \varphi_2 x_j dx dt \\ &+ \int_{\Omega} u_0 \varphi_1 x_j dx, \end{aligned} \quad (33)$$

que pode ser reescrita na forma

$$\begin{aligned} &\int_0^T \int_{\Omega} \sigma^0(t, x) \cdot \nabla_x (x_j \varphi_1 \varphi_2) dx dt - \int_0^T \int_{\Omega} \sigma^0(t, x) \cdot e_j \varphi_1 \varphi_2 dx dt \\ &- \int_0^T \int_{\Omega} A_j^* \cdot \varphi_2 \nabla_x \varphi_1 u^0 dx dt = \int_0^T \int_{\Omega} c^* u^0 \varphi_1 \frac{d\varphi_2}{dt} x_j + f \varphi_1 \varphi_2 x_j dx dt + \int_{\Omega} u_0 \varphi_1 x_j dx, \end{aligned} \quad (34)$$

$\forall \varphi_1 \in D(\Omega), \varphi_2 \in D(0, T)$  onde

$$A^* = (A_j^*)_{1 \leq j \leq N} = \left( \int_Y A^t(y) (e_j + \nabla_y \chi_j) dy \right)_{1 \leq j \leq N} \quad \text{e} \quad c^* = \int_Y c(y) dy. \quad (35)$$

Para a função teste  $\phi = x_j \varphi_1(x) \varphi_2(t)$  temos

$$\int_0^T \int_{\Omega} \sigma^0(t, x) \cdot \nabla_x (x_j \varphi_1 \varphi_2) dx dt = \int_0^T \int_{\Omega} c^* u^0 \varphi_1 \frac{d\varphi_2}{dt} x_j + f \varphi_1 \varphi_2 x_j dx dt + \int_{\Omega} u_0 \varphi_1 x_j dx. \quad (36)$$

Substituindo (36) em (34) segue que

$$\begin{aligned} &\int_0^T \int_{\Omega} c^* u^0 \varphi_1 \frac{d\varphi_2}{dt} x_j + f \varphi_1 \varphi_2 x_j dx dt - \int_0^T \int_{\Omega} \sigma^0(t, x) \cdot e_j \varphi_1 \varphi_2 dx dt \\ &- \int_0^T \int_{\Omega} A_j^* \cdot \varphi_2 \nabla_x \varphi_1 u^0 dx dt = \int_0^T \int_{\Omega} c^* u^0 \varphi_1 \frac{d\varphi_2}{dt} x_j + f \varphi_1 \varphi_2 x_j dx dt, \end{aligned} \quad (37)$$

e, portanto,

$$\int_0^T \int_{\Omega} \sigma^0(t, x) \cdot e_j \varphi_1 \varphi_2 dx dt = - \int_0^T \int_{\Omega} A_j^* \cdot \varphi_2 \nabla_x \varphi_1 u^0 dx dt = \int_0^T \int_{\Omega} A_j^* \cdot \nabla_x u^0 \varphi_2 \varphi_1 dx dt. \quad (38)$$

Disso, segue

$$\sigma^0(t, x) \cdot e_j = A_j^* \cdot \nabla_x u^0 = A^* e_j \cdot \nabla_x u^0 = A^{t*} \nabla_x u^0 \cdot e_j \quad (39)$$

e portanto

$$\sigma^0(t, x) = A^{t*} \nabla_x u^0. \quad (40)$$

Pela densidade das funções suaves de  $H_0^1(\Omega)$  e  $H^1(0, T)$ , a equação (33) é válida para toda função teste  $\phi_1 \in D(\Omega)$  e  $\phi_2 \in D(0, T)$ . Uma vez que  $A^{t*}$  e  $c^*$  são limitadas e definidas positivas [2],  $\alpha |\xi|^2 \leq A^{t*}(y) \xi \cdot \xi \leq \frac{\beta^2}{\alpha} |\xi|^2$ ,  $c^- \leq c^* \leq c^+$ , garante-se que (33) admite uma única solução em  $L^\infty(0, T; L^2(\Omega)) \cap \mathcal{W}$  e isso prova que qualquer subsequência de  $\{u^\varepsilon\}$ , e não somente uma

subseqüência, converge para a solução homogeneizada  $u^0 \in L^\infty(0, T; L^2(\Omega)) \cap \mathcal{W}$ . Portanto o problema limite quando  $\varepsilon \rightarrow 0$  de (6) é

$$c^* \frac{\partial u^0}{\partial t} - \nabla_x \cdot (A^{t*} \nabla_x u^0) = f(t, x) \quad (41)$$

e sua única solução  $u^0$  é o limite de toda subseqüência de  $u^\varepsilon$ , ou seja,  $u^\varepsilon \rightarrow u^0$ , quando  $\varepsilon \rightarrow 0$ .

Observe que nesta abordagem, se utiliza da equação (25) para anular os termos envolvendo funções fracamente convergentes e assim fornecer o fluxo efetivo. Uma abordagem alternativa utilizada por Allaire [1], para obter a equação homogeneizada diretamente, foi utilizando a função teste  $\varphi(t, x) + \varepsilon \nabla_x \varphi(t, x) \cdot \chi\left(\frac{x}{\varepsilon}\right)$  em que  $\chi$  é solução de (13). Este método também pode ser usado para possíveis problemas monótonos não-lineares [6]. No entanto, neste caso, o procedimento de subtrair os dois problemas com a escolha transversal das funções de teste, a fim de poder passar para o limite, não é aplicável. Isso significa que é necessário ter uma maneira alternativa de lidar com o processo de limite na configuração mais geral. A chave para isso é a chamada compacidade compensada, que foi introduzida pela primeira vez por Murat [5] e Tartar [8]. A compacidade compensada permite, em certo sentido, obter a convergência dos produtos de duas seqüências  $u^\varepsilon$  e  $v^\varepsilon$  ao produto dos respectivos limites, sem que nenhuma das seqüências seja fortemente convergente. Em vez disso, propriedades de convergência adequadas podem ser impostas às derivadas em ambas as seqüências. Um protótipo de compactação compensada é o resultado, conhecido como o lema divergência-rotacional (*div-curl lemma*) [7].

### 3 Considerações Finais

Neste trabalho, mostramos como homogenizar a equação do calor, através do método das funções teste oscilantes, considerando uma capacidade calorífica e um tensor de condutividade, não necessariamente isotrópico, que variam rapidamente com relação à posição. Apontamos uma abordagem alternativa, mais direta, para obtenção do comportamento efetivo usando uma função teste específica. Graças à preservação das propriedades de  $A^{t*}$  (limitada e definida positiva), se garante a existência de uma única solução para o problema homogeneizado e assim que toda a seqüência  $\{u^\varepsilon\}$  converge para única solução  $u^0$ . Uma vez que o problema homogeneizado apresenta coeficientes constantes, obter uma solução para ele torna-se mais simples, comparado ao problema original. Além disso, é possível mostrar que  $\|u^\varepsilon - u^0\|_{L^2(0, T; H_0^1(\Omega))} = \mathcal{O}(\varepsilon)$ , ou seja, que a solução homogeneizada é uma boa aproximação da solução exata [4].

### Referências

- [1] G. Allaire, Homogenization and Two-Scale Convergence, SIAM Journal on Mathematical Analysis, v.23, n.6 (1992) 1482–1518.
- [2] D. Cioranescu, P. Donato, “An Introduction to Homogenization”, University Press, Oxford, 1999.
- [3] L.C. Evans, “Partial Differential Equations”, Graduate Studies in Mathematics, v. 19, American Mathematical Society, 2010.
- [4] M.P. de Lima, “Homogeneização matemática de meios micro-heterogêneos com estrutura periódica”, Dissertação de Mestrado, IFM-UFPel, 2016.
- [5] F. Murat, Compacité par compensation, Annali della Scuola Normale Superiore di Pisa, Classe di Scienze, 4a série, tome 5, n.3 (1978) 489–507.
- [6] M. Olsson, “G-convergence and Homogenization of some Monotone Operators”, Doctoral Thesis, Mid Sweden University, 2008.

- [7] L. Tartar, “The General Theory of Homogenization: A Personalized Introduction”, Berlin, Springer, 2010.
- [8] L. Tartar, Compensated compactness and applications to partial differential equations, In Nonlinear analysis and mechanics: Heriot–Watt Symposium, Research Notes in Mathematics, 39, Pitman, London, v. 4 (1979) 136–212.
- [9] S. Torquato, “Random Heterogeneous Materials: Microstructure and Macroscopic Properties”, New York, Springer, 2002.



# Existência de acoplamentos espaciais favorecendo sincronização em um modelo metapopulacional

Flávia Tereza Giordani  
UFSC - Departamento de Matemática  
Campus Trindade  
88040-900, Florianópolis, SC  
E-mail: flavia.giordani@ufsc.br.

## Resumo:

No presente trabalho, investigamos a estabilidade assintótica de atratores caóticos sincronizados relacionados a um modelo metapopulacional discreto de uma única espécie. Refinando os resultados em [1], estabelecemos critérios para a solução do problema inverso de obter uma matriz de conectividade entre os sítios de forma que a estabilidade do atrator no espaço de fase do sistema seja prescrita. Tais resultados podem ser relacionados ao estudo da viabilidade da construção de corredores de conservação.

**Palavras-chave:** *Sincronização, metapopulação, números de Lyapunov*

## 1 Introdução

O interesse crescente da comunidade científica em torno da extinção das espécies e perda de biodiversidade, tem levado a diversos estudos sobre os efeitos da migração em redes de populações espacialmente estruturadas [2, 11, 16]. De fato, estudos recentes [9, 12] indicam que a migração de indivíduos entre habitats pode contribuir favoravelmente para a persistência populacional e para a diversidade genética e biológica das espécies. Dessa forma, a viabilidade da construção de rotas migratórias apropriadas (corredores de conservação) conectando habitats outrora isolados é de interesse fundamental para conservacionistas. No entanto, conforme destacado em [5, 12], a arquitetura desses corredores deve ser pensada com cautela, pois uma maior conectividade entre os sítios pode levar à ocorrência de densidades populacionais locais de baixa densidade flutuando próximas ao estado sincronizado, o que por sua vez, aumenta as chances de extinção global. Isso ocorre porque oscilações síncronas reduzem a possibilidade de ocorrência do denominado “efeito resgate”, que é a recolonização de um sítio por imigrantes de outros sítios [8, 12].

Neste trabalho, abordamos um modelo metapopulacional discreto de uma única espécie e  $N$  sítios conectados por processos migratórios [7, 15]. Considerando o caso de migração independente da densidade, analisamos a viabilidade de estabelecer conexões entre os sítios de forma que órbitas caóticas sincronizadas tenham um comportamento prescrito de estabilidade. Estabelecemos critérios, baseados nos números de Lyapunov, para a existência de uma matriz de conectividade apropriada, de forma a favorecer a sincronia ou a assincronia das populações locais. Para isso, um problema inverso de autovalores para matrizes simétricas e duplamente estocásticas deve ser analisado, nas linhas da investigação em [10, 1]. Conforme discussão prévia, essa análise é importante para o design de corredores de conservação, uma vez que a estabilidade de atratores caóticos sincronizados no espaço de fase do sistema torna a metapopulação mais suscetível à sincronização das populações locais, aumentando o risco de extinção [13, 16, 15].

Ao longo do trabalho,  $\mathcal{M}_N$  denota o conjunto das matrizes duplamente estocásticas, simétricas, irredutíveis, não-negativas, com diagonal nula e de ordem  $N$ .

## 2 Modelo

Consideremos uma coleção de sítios enumerados por  $1, 2, \dots, N$ , os quais são fragmentos de habitat que possuem recursos necessários à reprodução e à sobrevivência das populações que o habitam. Denotamos por  $x_j^t$  a densidade populacional no sítio  $j$ , no tempo  $t$ . Na ausência de migração entre os sítios admitimos que a dinâmica local é descrita por  $x_j^{t+1} = f(x_j^t)$ ,  $j = 1, 2, \dots, N$ ,  $t \in \mathbb{Z}_+$ , sendo  $f : \mathbb{R} \rightarrow \mathbb{R}$  de classe  $C^{1+\varsigma}$ ,  $0 < \varsigma < 1$ , uma função que incorpora processos de reprodução e sobrevivência em cada sítio. Os sítios estão conectados por processos migratórios e supomos que a dispersão de indivíduos de um determinado sítio para outros sítios ocorre em cada geração após o processo de dinâmica local. Mais precisamente, a cada geração, uma fração  $\mu_{ji}$  de indivíduos do sítio  $i$  migra para o sítio  $j$ , em um processo de curta duração, de forma que podemos supor que não há morte de indivíduos quando em migração. Neste trabalho vamos admitir que  $\mu_{ji} = c_{ji}\bar{\mu}$ , sendo  $0 < \bar{\mu} \leq 1$  a fração de indivíduos que migra em cada sítio e  $0 \leq c_{ji} \leq 1$ ,  $c_{ii} = 0$ , a proporção dos indivíduos que migram do sítio  $i$  e chegam no sítio  $j$ . Com isso, o sistema que descreve a dinâmica dessa rede de populações é dado por

$$x_j^{t+1} = (1 - \bar{\mu})f(x_j^t) + \sum_{i=1}^N c_{ji}\bar{\mu}f(x_i^t), \quad t \in \mathbb{Z}_+, \quad j \in \{1, 2, \dots, N\}. \quad (1)$$

Note que, no modelo acima,  $c_{ji}\bar{\mu}x_i^t$  representa a quantidade de indivíduos que deixam o sítio  $i$  e movem-se para o sítio  $j$  no tempo  $t$ . Assim, os dois primeiros termos à direita da equação (1) representam a quantidade de indivíduos que permanece no sítio  $j$  após o processo de migração, e o segundo termo representa a quantidade de indivíduos provenientes dos sítios vizinhos ao sítio  $j$ . Nesse trabalho vamos considerar a hipótese técnica de que  $C = [c_{ij}] \in \mathcal{M}_N$ . Note que essa hipótese é consistente com o fato de que o processo de migração é 100% bem sucedido, uma vez que nesse caso devemos ter  $\sum_{j=1}^N c_{ji} = 1$ . Vamos denotar por  $\mathbf{F} : \mathbb{R}^N \rightarrow \mathbb{R}^N$  o mapa global do sistema (1).

Dizemos que uma solução  $(x_i^t)_{1 \leq i \leq N, t \in \mathbb{Z}_+}$  do sistema (1) é *sincronizada* se  $x_i^t \equiv x_s^t$  verifica-se, para todo  $i = 1, 2, \dots, N$  e cada  $t \in \mathbb{Z}_+$ . Soluções sincronizadas assumem valores no subespaço unidimensional

$$\mathbf{S} = \text{span}\{\mathbf{v}\}, \quad \mathbf{v} = (1, 1, \dots, 1)_N^T, \quad (2)$$

do espaço de fase do sistema. É importante observar que a hipótese de que  $C \in \mathcal{M}_N$  é uma condição suficiente para a invariância do subespaço  $\mathbf{S}$  com relação a  $\mathbf{F}$  e, em particular, para a existência de soluções sincronizadas [14, 7].

## 3 Análise de estabilidade

Nessa seção nos referimos a vários conceitos de sistemas dinâmicos, como noções de atrator assintoticamente estável, sela caótica e estabilidade no sentido de Lyapunov; referimos o leitor a [3, 6] para detalhes.

Vamos admitir que  $\mathbf{F}|_{\mathbf{S}}$  tem um atrator assintoticamente estável  $\Omega$ . Queremos analisar o seguinte problema: sob que condições  $\Omega$  é um atrator assintoticamente estável para  $\mathbf{F}$ ? Isso significa que queremos avaliar se a base de atração de  $\Omega$  com relação a  $\mathbf{F}$  contém um subconjunto aberto  $U$  do  $\mathbb{R}^N$ , de forma que  $\Omega \subset U$ . Em outras palavras, se para condições iniciais suficientemente próximas a  $\Omega$ , mas não necessariamente em  $\mathbf{S}$ , a órbita correspondente é atraída para  $\Omega$ . Do ponto de vista biológico, essa questão é fundamental para a quantificação do que Earn et al. [5] denominaram de *coerência espacial* da metapopulação: uma medida sintética do quão suscetível o sistema é à sincronização. Mais precisamente, a coerência espacial quantifica a propensão de órbitas que iniciam próximas ao estado sincronizado tem de se aproximarem de  $\mathbf{S}$ , ou seja, de se tornarem coerentes (sincronizadas). Quanto menor a coerência espacial o

sistema é menos suscetível à extinção global, e o design de corredores de conservação deve levar em consideração esse ingrediente [5, 12].

Para fazer essa análise vamos usar a teoria da estabilidade linear a qual consiste em linearizar  $\mathbf{F}$  em torno de um ponto  $\mathbf{x}^0 \in \Omega$ , isto é, dada uma perturbação  $\tilde{\mathbf{x}}^0$  de  $\mathbf{x}^0$ , escrever

$$\tilde{\mathbf{x}}^{t+1} - \mathbf{x}^{t+1} \approx D\mathbf{F}(\mathbf{x}^t)D\mathbf{F}(\mathbf{x}^{t-1}) \dots D\mathbf{F}(\mathbf{x}^0)\Delta_0 = D\mathbf{F}^t(\mathbf{x}^0)\Delta_0, \quad t = 0, 1, 2, \dots \quad (3)$$

sendo  $\Delta_0 = \tilde{\mathbf{x}}_0 - \mathbf{x}_0$ . É bem conhecido o fato de que, pelo menos formalmente, pode-se avaliar o quanto a órbita perturbada  $\{\tilde{\mathbf{x}}^t\}_{t=0,1,2,\dots}$  se “afasta” ou se “aproxima” da órbita sincronizada  $\{\mathbf{x}^t\}_{t=0,1,2,\dots}$ , através da análise do fator  $D\mathbf{F}^t(\mathbf{x}^0)$  (ver [6]). Observamos que, sendo  $\mathbf{x} = xv$ , temos  $D\mathbf{F}(\mathbf{x}) = f'(x)[\mathbf{I}_N - \bar{\mu}B]$ , em que  $B = \mathbf{I}_N - C$ . e que  $\lambda_0 = 0$  é autovalor simples de  $B$  tendo autoespaço associado precisamente igual a  $\mathbf{S}$  [7]. Dessa forma, é natural [14, 3] analisarmos o comportamento das componentes transversas a soluções sincronizadas introduzindo a decomposição  $\mathbb{R}^N = \mathbf{S} \oplus \mathbf{S}^\perp$  e considerando uma base  $\beta$  do  $\mathbb{R}^N$  formada por  $v$  e  $N - 1$  vetores L.I. pertencentes a  $\mathbf{S}^\perp$ . Seja  $Q$  a matriz de mudança de base correspondente, então existe uma matriz  $A$  de ordem  $(N - 1) \times (N - 1)$  de forma que  $B = Q^{-1}\mathcal{B}Q$ , sendo

$$\mathcal{B} = \begin{bmatrix} 0 & \dots & 0 \\ \vdots & & \\ 0 & A & \end{bmatrix}. \quad (4)$$

Como consequência, podemos escrever  $D\mathbf{F}(\mathbf{x}) = Q^{-1}\mathbf{J}(\mathbf{x})Q$ , sendo

$$\mathbf{J}(\mathbf{x}) = f'(x) \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & \mathbf{I}_{N-1} - \bar{\mu}A & & \\ 0 & & & \end{bmatrix}, \quad t = 0, 1, 2, \dots \quad (5)$$

Isso significa que, se considerarmos a base  $\beta$ , perturbações paralelas à órbita  $\{\mathbf{x}^t\}_{t=0,1,2,\dots}$  são amplificadas/reduzidas através do fator  $\prod_{i=0}^t f'(x^i)$ , enquanto que o comportamento de perturbações transversais a essa órbita dependem do fator  $\prod_{i=0}^t f'(x^{t-i})(\mathbf{I}_{N-1} - \bar{\mu}A)$ . Isso motiva a seguinte definição [3].

**Definição 3.1.** Considere  $\mathbf{x} = xv \in \Omega$ , então o número paralelo de Lyapunov em  $\mathbf{x}$  é definido como

$$L_{//}(\mathbf{x}) = \lim_{t \rightarrow +\infty} \prod_{i=0}^t |f'(x^i)|^{1/t}, \quad (6)$$

sendo  $x^0 = x$  e  $x^{i+1} = f(x^i)$ ,  $i \in \mathbb{Z}_+$ . O número transversal de Lyapunov em  $\mathbf{x}$  na direção  $\mathbf{u} \in \mathbb{R}^N$  é dado por

$$L_{\perp}(\mathbf{x}, \mathbf{u}) = \lim_{t \rightarrow +\infty} \left\| \prod_{i=0}^t f'(x^{t-i})(\mathbf{I}_{N-1} - \bar{\mu}A)\mathbf{u}_{N-1}^\perp \right\|^{1/t}, \quad (7)$$

sendo  $[\mathbf{u}]_\beta = (0, \mathbf{u}_{N-1}^\perp)^T$ .

Vamos supor que o atrator  $\Omega$  suporta uma medida invariante de probabilidade  $\rho$  que é ergódica. Então, como consequência do Teorema de Oseledets [3], para  $\rho$ -quase todo  $x \in \mathcal{O}$ , o limite (6) existe e independe de  $x$ ; para  $\mathbf{u} \in \mathbb{R}^N$  fixado e  $\rho$ -quase todo  $x \in \mathcal{O}$ , o limite (7) existe e independe de  $x$ ; além disso (7) assume um dos valores

$$L_{\perp}^1(\rho) < L_{\perp}^2(\rho) < \dots < L_{\perp}^s(\rho),$$

sendo  $1 \leq s \leq N - 1$ . É importante observar que os limites acima, embora independam de  $\mathbf{x}$ , dependem da medida  $\rho$ , uma vez que atratores podem suportar mais de uma medida ergódica [6]. Vamos definir

$$L_{\perp}^{max} = \sup_{\rho} L_{\perp}^s(\rho), \quad L_{\perp}^{min} = \inf_{\rho} L_{\perp}^1(\rho),$$

sendo o sup e o inf tomados sobre todas as medidas invariantes de probabilidade e ergódicas com suporte em  $\Omega$ . Em [3] foi demonstrado que  $0 \leq L_{\perp}^{min} \leq L_{\perp}^{max} < \infty$  e também que:

**Teorema 3.1.** *Se  $L_{\perp}^{max} < 1$ ,  $\Omega$  é assintoticamente estável como um atrator em  $\mathbb{R}^N$ ; por outro lado, se  $L_{\perp}^{max} > 1$ ,  $\Omega$  é Lyapunov instável como atrator em  $\mathbb{R}^N$ . Caso tenhamos  $L_{\perp}^{min} > 1$ ,  $\Omega$  é uma sela caótica normalmente repelente.*

## 4 Resultados Principais

Para investigar a existência da matriz de conectividade  $C$  de forma que o atrator  $\Omega$  tem um comportamento prescrito como atrator em  $\mathbb{R}^N$ , vamos utilizar os seguintes resultados técnicos, os quais são adaptações daqueles demonstrados em [10].

**Teorema 4.1.** *Dados reais  $\eta_1 \geq \eta_2 \geq \dots \geq \eta_N$ , para que exista  $C \in \mathcal{M}_N$  de forma que  $\sigma(C) = \{\eta_1, \eta_2, \dots, \eta_N\}$  é necessário que  $\eta_1 = 1$ ,  $\eta_N \geq -1$  e  $\eta_1 + \eta_2 + \dots + \eta_N \geq 0$ . Por outro lado, uma condição suficiente para a existência de  $C$ , é que  $\eta_1 = 1$  e  $\{\eta_2, \dots, \eta_N\} \subset [-1/(N - 1), 1]$ .*

**Proposição 4.1.** *Dado  $\eta \in [-1, 1]$ , existe  $C \in \mathcal{M}_N$  tal que  $\eta \in \sigma(C)$ .*

Como estamos considerando o caso em que  $C$  é simétrica, então  $A$  dada em (4) é uma matriz diagonal  $A = \text{diag}(\lambda_2, \lambda_3, \dots, \lambda_N)$ , sendo  $0 \leq \lambda_2 \leq \lambda_3 \leq \dots \leq \lambda_N \leq 2$ . Então temos o seguinte resultado, cuja prova é imediata.

**Teorema 4.2.** *São válidas as igualdades  $L_{\perp}^s(\rho) = L_{\parallel}^s(\rho)\bar{\Lambda}$  e  $L_{\perp}^1(\rho) = L_{\parallel}^1(\rho)\underline{\Lambda}$ , sendo*

$$\bar{\Lambda} = \max\{|1 - \bar{\mu}\lambda_i|, \lambda_i \in \sigma(A)\}, \quad \underline{\Lambda} = \min\{|1 - \bar{\mu}\lambda_i|, \lambda_i \in \sigma(A)\}.$$

Note que  $\bar{\Lambda}$ ,  $\underline{\Lambda}$  dependem somente da topologia da rede. Como consequência do Teorema acima vemos facilmente que  $L_{\perp}^{max} = \bar{\Lambda}L_{\parallel}^{max}$ , e  $L_{\perp}^{min} = \underline{\Lambda}L_{\parallel}^{min}$  com  $L_{\parallel}^{max} = \sup_{\rho} L_{\parallel}(\rho)$ ,  $L_{\parallel}^{min} = \inf_{\rho} L_{\parallel}(\rho)$ . Queremos discutir a seguinte questão: conhecendo  $L_{\parallel}^{max}$  e  $L_{\parallel}^{min}$ , os quais não dependem da topologia da rede, existe uma matriz  $C \in \mathcal{M}_N$ , de forma que  $\Omega$  tenha um comportamento de estabilidade prescrito como um atrator em  $\mathbb{R}^N$ ? Essa questão é facilmente respondida no caso em que  $L_{\parallel}^{max} \leq 1$ . De fato, para toda matriz  $C \in \mathcal{M}_N$  temos  $\sigma(A) \subset ]0, 2]$  o que implica em  $\bar{\Lambda} < 1$  e assim  $L_{\perp}^{max} < 1$ . Concluímos então que, nessa situação, para qualquer matriz  $C \in \mathcal{M}_N$ ,  $\Omega$  é assintoticamente estável como um atrator em  $\mathbb{R}^N$ . Isso reflete o fato bem conhecido de que a migração independente da densidade não tem o efeito de desestabilizar globalmente órbitas que são localmente estáveis [14]. No entanto, a questão é mais sutil quando consideramos atratores locais com dependência sensível a condições iniciais. Para analisar esse caso, vamos supor que  $L_{\parallel}^{max} > 1$ , ou seja,  $\Omega$  é caótico como atrator em  $S$ . O seguinte Teorema é um refinamento dos resultados estabelecidos em [1].

**Teorema 4.3.** *Seja  $0 < \bar{\mu} \leq 1$  e suponha que  $L_{\parallel}^{max} > 1$ . Então:*

1. *Existe  $C \in \mathcal{M}_N$  de forma que  $\Omega$  é Lyapunov instável como um atrator em  $\mathbb{R}^N$ .*
2. *Se  $\frac{N}{N-1} > \frac{1}{\bar{\mu}} \left(1 - \frac{1}{L_{\parallel}^{max}}\right)$ , existe  $C \in \mathcal{M}_N$  de forma que  $\Omega$  é assintoticamente estável como um atrator em  $\mathbb{R}^N$ .*

3. Se  $\frac{N}{N-1} < \frac{1}{\bar{\mu}} \left(1 - \frac{1}{L_{\parallel}^{max}}\right)$ ,  $\Omega$  é Lyapunov instável como um atrator em  $\mathbb{R}^N$ , seja qual for  $C \in \mathcal{M}_N$ .

4. Se  $L_{\parallel}^{min} > 1$  e  $\frac{N}{N-1} < \frac{1}{\bar{\mu}} \left(1 - \frac{1}{L_{\parallel}^{min}}\right)$ , existe  $C \in \mathcal{M}_N$  de forma que  $\Omega$  é uma sela caótica normalmente repelente.

*Demonstração.* 1. Tome qualquer  $0 < \lambda < \min \left\{ \frac{1}{\bar{\mu}} \left(1 - \frac{1}{L_{\parallel}^{max}}\right), 2 \right\}$ . Então, definindo  $\eta = 1 - \lambda$ , vemos facilmente que  $-1 < \eta < 1$ . Logo, conforme Teorema 4.1, existe  $C \in \mathcal{M}_N$  tal que  $\eta \in \sigma(C)$ . Pelo fato que

$$0 < \lambda < \frac{1}{\bar{\mu}} \left(1 - \frac{1}{L_{\parallel}^{max}}\right) \implies \frac{1}{L_{\parallel}^{max}} < 1 - \bar{\mu}\lambda < 1,$$

o Teorema 4.2 implica que  $\bar{\Lambda} L_{\parallel}^{max} \geq |1 - \bar{\mu}\lambda| L_{\parallel}^{max} > 1$  e obtemos o resultado como consequência do Teorema 3.1.

2. Considere quaisquer  $\lambda_2 \leq \lambda_3 \leq \dots \leq \lambda_N$  de forma que

$$\frac{1}{\bar{\mu}} \left(1 - \frac{1}{L_{\parallel}^{max}}\right) < \lambda_k < \min \left\{ \frac{1}{\bar{\mu}} \left(1 + \frac{1}{L_{\parallel}^{max}}\right), \frac{N}{N-1} \right\}, \quad k = 2, 3, \dots, N.$$

Então, definindo  $\eta_k = 1 - \lambda_k$ , vemos que  $\eta_2 \geq \eta_3 \geq \dots \geq \eta_N$  e que

$$\frac{1}{\bar{\mu}} \left(1 - \frac{1}{L_{\parallel}^{max}}\right) < \lambda_k < \frac{N}{N-1} \implies \frac{-1}{N-1} < \eta_k < 1$$

para cada  $k = 2, 3, \dots, N$ . Consequentemente, em vista do Teorema de forma que  $\sigma(C) = \{1, \eta_2, \dots, \eta_N\}$ . Como, para cada  $k = 2, 3, \dots, N$ , temos

$$\frac{1}{\bar{\mu}} \left(1 - \frac{1}{L_{\parallel}^{max}}\right) < \lambda_k < \frac{1}{\bar{\mu}} \left(1 + \frac{1}{L_{\parallel}^{max}}\right) \implies \frac{-1}{L_{\parallel}^{max}} < 1 - \bar{\mu}\lambda_k < \frac{1}{L_{\parallel}^{max}},$$

então  $\bar{\Lambda} < \frac{1}{L_{\parallel}^{max}}$ . Assim, o Teorema 4.2 assegura que  $L_{\perp}^{max} < 1$  e devido ao Teorema 3.1 obtemos o resultado desejado.

3. De acordo com o Teorema 4.1, dados quaisquer  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$ , uma condição necessária para que uma matriz  $C \in \mathcal{M}_N$  seja tal que  $\sigma(C) = \{1 - \lambda_1, 1 - \lambda_2, \dots, 1 - \lambda_N\}$  é que  $\lambda_1 = 0$ ,  $\lambda_N \leq 2$  e que

$$1 + (1 - \lambda_1) + (1 - \lambda_2) + \dots + (1 - \lambda_N) \geq 0.$$

Isso significa que deve ser válida a desigualdade  $\lambda_2 + \dots + \lambda_N \leq N$ , e, como consequência, também  $(N-1)\lambda_2 \leq N$ . Então, supondo que  $\frac{N}{N-1} < \frac{1}{\bar{\mu}} \left(1 - \frac{1}{L_{\parallel}^{max}}\right)$ , para toda  $C \in$

$\mathcal{M}_N$ , temos  $\lambda_2 < \frac{1}{\bar{\mu}} \left(1 - \frac{1}{L_{\parallel}^{max}}\right)$ , ou seja,  $1 - \bar{\mu}\lambda_2 > \frac{1}{L_{\parallel}^{max}}$ . Isso implica que

$$\bar{\Lambda} \geq |1 - \bar{\mu}\lambda_2| > \frac{1}{L_{\parallel}^{max}}$$

o que, por sua vez, nos permite concluir que  $L_{\perp}^{max} > 1$  e o resultado segue novamente pelo Teorema 3.1.

4. Considere quaisquer  $\lambda_2 \leq \lambda_3 \leq \dots \leq \lambda_N$  de forma que  $0 < \lambda_k < \frac{N}{N-1}$ . Isso significa que, para todo  $k = 2, 3, \dots, N$ ,  $\eta_k = 1 - \lambda_k$  satisfaz  $\frac{-1}{N-1} < \eta_k < 1$ , ou seja, podemos utilizar o Teorema 4.1 para garantir a existência de uma matriz  $C \in \mathcal{M}_N$  tal que  $\sigma(C) = \{1, \eta_2, \dots, \eta_N\}$ . Como, para todo  $k = 2, 3, \dots, N$ ,

$$0 < \lambda_k < \frac{N}{N-1} < \frac{1}{\bar{\mu}} \left( 1 - \frac{1}{L_{\parallel}^{min}} \right) \implies \frac{1}{L_{\parallel}^{min}} < 1 - \bar{\mu}\lambda_k < 1,$$

pelo Teorema 4.2 temos  $L_{\perp}^{min} = \underline{\Lambda} L_{\parallel}^{min} > 1$ . Em vista do Teorema 3.1, concluímos que  $\Omega$  é uma sela caótica normalmente repelente. □

**Corolário 4.1.** *Suponha que  $L_{\parallel}^{max} > 1$  e que  $\left( 1 - \frac{1}{L_{\parallel}^{max}} \right) < \bar{\mu} \leq 1$ . Então, seja qual for  $N \in \{2, 3, 4, \dots\}$ , existe  $C \in \mathcal{M}_N$  de forma que  $\Omega$  é assintoticamente estável como um atrator em  $\mathbb{R}^N$ .*

*Demonstração.* De fato, nesse caso,  $\frac{1}{\bar{\mu}} \left( 1 - \frac{1}{L_{\parallel}^{max}} \right) < 1 < \frac{N}{N-1}$  e o resultado segue pelo Item 2 do Teorema 4.3. □

## 5 Conclusão

Conforme destacado em [16], a dispersão em metapopulações tem efeito estabilizador, dependendo de sua intensidade. Baixas taxas de migração entre populações que são localmente caóticas induzem uma maior propensão a dinâmicas assíncronas regulares do sistema acoplado [13], fato que é corroborado por evidências experimentais [4, 12]. Tal cenário é importante para a persistência das espécies envolvidas, uma vez que oscilações síncronas caóticas podem favorecer sua extinção enquanto que assincronia favorece a recolonização de populações extintas [5, 13]. Os resultados obtidos no Teorema 4.3 reforçam essa perspectiva, uma vez que, para  $L_{\parallel}^{max} > 1$  fixado, o fator  $\left( 1 - \frac{1}{L_{\parallel}^{max}} \right)$  é amplificado por  $\frac{1}{\bar{\mu}}$  nas desigualdades dos Itens 2 e 3. Assim, se a taxa de migração é baixa,  $\Omega$  é Lyapunov instável como atrator em  $\mathbb{R}^N$ , independentemente da forma como os sítios são conectados. Nessas circunstâncias, órbitas típicas que correspondem a densidades locais flutuando próximas ao estado sincronizado não ficam arbitrariamente próximas de  $S$ , em outras palavras, a metapopulação é menos suscetível às denominadas *oscilações coerentes* [5, 12]. Por outro lado, como ilustra o Corolário 4.1, no caso de altas taxas de migração, independentemente do tamanho da rede, o sistema é suscetível a oscilações caóticas síncronas estáveis, dependendo da forma como ocorre o acoplamento. Isso vem ao encontro dos resultados em [4, 5, 12], que predizem o aumento do risco de extinção em metapopulações com altas taxas de migração. Para minimizar os riscos de extinção de populações distribuídas em habitats fragmentados, biólogos tem concentrado esforços no design dos denominados corredores de conservação [12]. Para que estes sejam efetivos, promovendo a persistência das espécies, deve-se arquitetá-los de forma a se evitar ao máximo flutuações síncronas caóticas [5]. Nesse sentido, nossos resultados mostram que, no contexto do modelo apresentado, a construção de tais corredores pode ser idealizada. De fato, sejam quais forem  $\bar{\mu}$  e  $N$ , podemos acoplar os sítios de acordo com uma determinada matriz  $C \in \mathcal{M}_N$  que torna  $\Omega$  Lyapunov instável (ou mesmo uma sela caótica), ou seja, a assincronia é favorecida. Pretendemos futuramente analisar uma questão de optimalidade: qual o valor de  $\bar{\mu}$  moderadamente grande de forma que o sistema possa se beneficiar do processo migratório e, ao mesmo tempo, suficientemente pequeno de forma que o risco de extinção seja minimizado.

## Referências

- [1] J. A. Barrionuevo and J. A. L. Silva, Stability and synchronism of certain coupled dynamical systems, *SIAM J. Math. Anal.*, 40 (2006) 939–951.
- [2] Y. Ben-Zion, Y. Fried e N. M. Shnerb, Migration, coherence and persistence in a fragmented landscape, *Theor. Biol.*, 5 (2012) 481-493.
- [3] J. Buescu, “Exotic Attractors: From Lyapunov Stability to Riddled Basins”, Birkhauser-Verlag, Basel, 1997.
- [4] S. Dey e A. Joshi, Stability via asynchrony in Drosophila Metapopulations with low migration rates, *Science*, 312 (2006) 434-436.
- [5] D. J. D. Earn, S. A. Levin e P. Rohani, Coherence and Conservation, *Science*, 290 (2000) 1360-1364.
- [6] J. P. Eckman e D. Ruelle, Ergodic theory of chaos and strange attractors, *Reviews of Modern Physics*, 57 (1985) 617-656.
- [7] F. T. Giordani, ” Migração Dependente da Densidade em Modelos Metapopulacionais”, Tese de Doutorado, IM-UFRGS, 2008.
- [8] M. Heino, V. Kaitala, E. Ranta e J. Lindström, Synchronous dynamics and rates of extinction in spatially structured populations, *Proc. R. Soc. London B*, 264 (1997) 481–486.
- [9] M. Holyoak, Connectance and Connectivity, em “Encyclopedia of Ecology” (S. V. Jorgensen e B. Fath, eds.) pp. 737-743, Elsevier, New York, 2008.
- [10] S. Hwang e S. Pyo, The inverse eigenvalue problem for symmetric doubly stochastic matrices, *Linear Algebra and its Appl.*, 379 (2004) 77–83.
- [11] R. A. Ims e N. G. Yoccoz, Studying Transfer Processes in Metapopulations: Emigration, Migration and Colonization, em “Metapopulation Biology - Ecology, Genetics, and Evolution” (I.A. Hanski e M.E. Gilpin, eds.) pp. 247-264, Academic Press, San Diego, 1997.
- [12] J. Molofsky e J. B. Ferdy, Extinction dynamics in experimental metapopulations, *Proc. Natl. Academy Sci.*, 102 (2005) 3726-3731.
- [13] G. D. Ruxton, Low levels of immigration between chaotic populations can reduce system extinction by inducing asynchronous regular cycles, *Proc. R. Soc. London. B*, 256 (1994) 189-193.
- [14] J. A. L. Silva e F. T. Giordani, Density-dependent Migration and Synchronism in Metapopulations, *Bull. Math. Biol.*, **68** (2006) 451-465.
- [15] J. A. L. Silva, J. A, Barrionuevo e F. T. Giordani, Synchronism in population networks with non linear coupling, *Nonlinear Analysis: Real World Applications*, **11** (2010) 1005-1016.
- [16] E. Tromeur, L. Rudolf e T. Gross, Impact of dispersal on the stability of metapopulations, *J. Theor. Biol.*, 392 (2016) 1-11.



# Criação de bactérias mutantes *in silico* de *S. agalactiae* híbridas entre *Oreochromis niloticus* e *Homo sapiens* por meio de algoritmos genéticos e lógica *Fuzzy*

Edgar L. de Aguiar\*

Departamento de Modelagem Matemática e Computacional, CEFET-MG, Belo Horizonte, MG

Gustavo H. M. Mendonça

Departamento de Modelagem Matemática e Computacional, CEFET-MG, Belo Horizonte, MG

Sandro R. Dias

Departamento de Modelagem Matemática e Computacional, CEFET-MG, Belo Horizonte, MG

Thiago. S. Rodrigues

Departamento de Modelagem Matemática e Computacional, CEFET-MG, Belo Horizonte, MG

**Resumo:** Os microrganismos são encontrados nos mais diversos reinos biológicos, possuindo uma ampla distribuição nos locais mais distintos no planeta, sendo responsáveis por causar diversos impactos ambientais e em diversos outros microrganismos. Um dos microrganismos de elevado potencial biotecnológico e patogênico são as bactérias. O objetivo desse trabalho é gerar bactérias artificiais mutantes híbridas de humano com peixe da espécie *Streptococcus agalactiae* por meio de algoritmos evolutivos, combinando os genes de bactérias encontradas em peixes e humanos, e sua posterior classificação em uma máquina de inferência *Fuzzy*. Para auxiliar os estudos análises comparativas entre os possíveis perfis genômicos das bactérias, como classificá-las e tratá-las de forma preventiva, evitar possíveis surtos. A Inteligência Computacional em conjunto com modelo de lógica *Fuzzy* foram muito úteis para auxiliar nas tomadas de decisões e classificação dos diferentes perfis das bactérias mutantes.

**Palavras-chave:** Algoritmo Genético. Bactéria Híbrida. Inteligência Computacional. Lógica *Fuzzy*. *Streptococcus Agalactiae*.

## 1 Introdução

Os microrganismos podem ser classificados e observados em diferentes reinos biológicos, sendo amplamente distribuídos pela natureza. A alta diversidade de microrganismos contribui para uma grande quantidade de informações que impulsionam estudos em diversas áreas como: agrícola, medicinal e industrial [12]. Os microrganismos podem ser detectados nos lugares mais hostis, afetando diretamente o padrão ambiental e outros organismos habitantes do mesmo. Levando-se em consideração essas informações, e ao extrapolar para o convívio humano, é possível apontar que tais características possam refletir em impactos significativos na economia, saúde e ambiente de forma positiva ou negativa [8].

Dentre os microrganismos mais estudados estão as bactérias. Elas são de grande importância medicinal, biotecnológica, veterinária, ambiental e um dos mais antigos organismos da Terra, com amostras localizadas em rochas de 3,8 milhões de anos [8]. Existem vários fatores que tornam as bactérias bons modelos para diversos estudos genômicos como: conservação de funções celulares em comum com organismos mais complexos, a possibilidade de se obter rapidamente em meio de cultura adequado uma elevada quantidade de material biológico, e possui a vantagem de ter um custo de desenvolvimento significativamente menor quando comparado a organismos mais complexos [5], [8].

---

\*edgarlaguiar@gmail.com

Entre as inúmeras características das bactérias, uma das mais importantes é a reprodução, pois elas permite a geração de mutações, transferências de genes entre espécies, trocas e ganhos de fatores de virulência e aquisição de novas resistências [8]. Os novos fatores de virulência e resistência aumentam significativamente o risco para os hospedeiros, pois a cada nova resistência adquirida, aumenta-se a necessidade de se desenvolver novos antibacterianos e bactericidas para combatê-los, sendo que a cada dia, antigos fármacos se tornam menos eficazes e, assim, aumentam o custo de pesquisas para se combater as novas resistências obtidas [11].

Tendo em vista as informações que são categoricamente discutidas pela literatura, nota-se que os microrganismos, em especial as bactérias, despertam grande interesse em pesquisas direcionadas as áreas abordadas neste trabalho. O total de projetos de genoma depositados na base de dados (BD) de Gold desde da sua criação, em 2007, o número de bactérias com genomas depositados sempre foi maior que dos demais domínios. Em 2011, o BD armazenava aproximadamente 10 mil genomas bacterianos, e no ano de 2015, o valor chegou a aproximadamente 50 mil genomas, teve um salto significativo de valor em 2018 pra aproximadamente 140 mil [3].

Dentro do filo das *Firmicutes* encontra-se o gênero *Streptococcus*. Esse gênero atualmente tem 121 espécies, 23 subespécies identificadas e, aproximadamente, 100 mil artigos relacionados [7], [9]. Dentre as espécies mais estudadas, destaca-se a *Streptococcus Agalactiae*, uma bactéria patogênica que afeta diversos hospedeiros.

Atualmente existem apenas por volta de 40 genomas completos de *S. agalactiae* depositados no banco de dados públicos, tornando-se necessário um aumento significativo no número de linhagens, pois esse valor baixo é um grande dificultador das análises comparativas. Outros fatores complexos são: o custo elevado de sequenciamento com plataformas NGS (*Next-Generation Sequencing*) e a elevada dificuldade técnica da montagem de genoma completo. Por tais motivos é necessário desenvolvimento de novas abordagens *in silico* para simular novas linhagens.

Neste trabalho, foi implementado um algoritmo bioevolutivo, de modo a criar computacionalmente bactérias mutantes que contenham genes das *S. agalactiae* de isoladas de peixe e de humano para futuras análises de transferência horizontal de genes [10]. Estas bactérias serão geradas procurando serem mais fieis a realidade possível considerando o perfil genômico da espécie e sua probabilidade de sobrevivência e sua capacidade de adaptativa. Serão dados maiores pesos aos genes de peixes originários da Ásia e genes de Humanos da América, pois o maior objetivo é estudar e prevenir mutações de bactérias com características destes peixes e com capacidade de afetarem humano na América. Em seguida, será criada uma máquina de inferência *Fuzzy* onde as bactérias serão classificadas considerando as informações obtidas pelo AG, e resultará em uma nota para cada bactérias referente à sua probabilidade de sobrevivência e sua capacidade de adaptação. Uma das formas de realizar essa classificação é por meio da utilização de Inteligência Computacional (IC). Pois através de uma máquina IC desenvolvida em conjunto com as especialistas Biomedicina e Medicina pela Santa Casa BH Ensino e Pesquisa, e Microbiologia pela UFMG será possível avaliar a eficiência da máquina *Fuzzy*, tal com as bactérias, auxiliando na comparação e na análise dos resultados.

## 2 Algoritmo Genético

Observando a evolução tecnológica principalmente na inteligência artificial, fica claro que a natureza serve como fonte de inspiração para os cientistas. Um dos exemplos mais difundidos é a computação evolucionária, baseada no princípio da evolução das espécies e na genética. Neste contexto, existem algoritmos evolucionário, criados à partir das ideias de evolução dos indivíduos na natureza, inspiradas pelo Darwinismo e Lamarckismo. Os algoritmos de computação evolutiva trabalham com um repertório (população) de soluções candidatas que interagem entre si e competem pela permanência na população. A evolução é alcançada basicamente pelos processos de reprodução dos indivíduos com herança genética, variação em uma população de indivíduos (mutação, que ajuda a criar diversidade na população) e aplicação da "seleção natural" para produção da próxima geração, onde indivíduos mais bem adaptados tem maiores chances de

sobreviverem [6].

Segundo Goldberg e Holland(1988), um algoritmo genético (AG) é uma técnica de busca que localizar uma sequência ótima, através do processamento de uma população de sequências inicializadas aleatoriamente, usando técnicas inspiradas na biologia evolutiva, como hereditariedade, mutação, seleção natural e Recombinação (*Crossover*) [4].

O primeiro passo do algoritmo genético é iniciar uma população de indivíduos (que pode ser feita de forma aleatória ou determinista), representando um conjunto de soluções candidatas para o problema. Em seguida, a população é avaliada, de maneira que cada indivíduo recebe um valor de Aptidão (*Fitness*), indicando quão boa aquela solução é para o problema. Este valor é utilizado para comparação entre os indivíduos e auxilia na probabilidade dos indivíduos com maior aptidão para serem selecionados na próxima geração e gerar descendentes no cruzamento, porém não descartando a possibilidade de indivíduos menos adaptados também se reproduzirem ou serem selecionados. Esta característica torna o processo de seleção probabilístico, de acordo com o valor de aptidão de cada indivíduo [6].

### 3 Lógica *Fuzzy*

Ao descrevermos certos fenômenos ou características relacionados ao mundo é comum a utilização de graus que representam qualidades ou verdades parciais. Como exemplo, podemos considerar o grupo "pessoas altas", e uma abordagem para caracterizar este grupo é dada de maneira que os indivíduos sejam considerados altos com mais ou menos intensidade, ou seja, existem elementos que pertenceriam mais à classe dos altos que outros. Por consequência, quanto menor (ou maior) for a medida da altura do indivíduo, menor (ou maior) será seu grau de pertinência a esta classe. Desse modo, podemos dizer que os indivíduos pertencem à classe das pessoas altas, com mais ou menos intensidade. Partindo deste tipo de questões, onde a propriedade que define o conjunto é incerta, que surgiu a teoria dos conjuntos *Fuzzy*, que tem crescido consideravelmente em nossos dias, tanto do ponto de vista teórico como nas aplicações em diversas áreas de estudo como computação e matemática [2].

Na lógica *Fuzzy* a ambiguidade semântica da inteligência humana pode ser representada por meio de variáveis linguísticas e seus termos primários. Uma variável linguística é uma entidade utilizada para se representar de modo impreciso e, portanto, linguístico, um conceito ou variável de um dado problema. Ela admite como valores as expressões linguísticas (chamadas de termos primários), em contraste com uma variável numérica que assume apenas valores precisos. Os termos primários de uma variável linguística formam a sua estrutura de conhecimento. Por exemplo, a variável linguística "Altura" poderia admitir os termos primários "Muito alto", "razoavelmente baixo" [1].

A propriedade sintática adotada nos modelos *Fuzzy* de Mamdani define regras de produção que possuem relações *Fuzzy* tanto em seus antecedentes quanto em seus consequentes [1]. Neste modelo, o módulo de interface (entrada) recebe valores numéricos e os converte em conjuntos *Fuzzy* equivalentes, ocorrendo uma conversão escalar  $\rightarrow$  *Fuzzy* (Fuzzificação). A máquina de inferência então busca em seu banco de conhecimento e processa as regras disparadas pela entrada, fazendo uma composição das mesmas. Então, o módulo de interface de saída recebe um conjunto *Fuzzy* (processado anteriormente) para cada variável de saída e o converte em um valor escalar correspondente, gerando saídas compatíveis com os sistemas externos [1].

### 4 Análise Comparativa Genômica

A análise comparativa genômica consiste em comparar e selecionar quais os genes que são encontrados em determinadas linhagens do *Cluster* sobre uma determinada condição [13]. Essas análises inicialmente podem ser realizadas pelos identificadores de genes, porém o mesmo gene é anotado com diferentes termos em diversas linhagens tal fato dificulta uma análise comparativa

por termos e ontologias. Foi necessário realizar alinhamento de sequências locais entre os genes de cada linhagem do Cluster para uma melhor curadoria, foi utilizado BlastN com os parâmetros mínimo 90% identidade e 90% de cobertura.

Após obtenção das linhagens no formato gbff no Site do NCBI, as linhagens foram separadas em diversos clusters gênicos separados por espécie, localização, sendo:

1° - *Cluster Homo sapiens*: análises *Core*, *Parcial* e *Única*;

2° - *Cluster Homo sapiens América X Homo sapiens Ásia*: análises *Core*, *Parcial* e *Única*;

3° - *Cluster Oreochromis sp*: análises *Core*, *Parcial* e *Única*;

4° - *Cluster Oreochromis sp América X Oreochromis sp Ásia*: análises *Core*, *Parcial* e *Única*;

5° - *Cluster Homo sapiens X Oreochromis sp*: análises *Core*, *Parcial* e *Única*.

A análise comparativa do *Core* consiste em selecionar quais genes são encontradas em todas as linhagens do *cluster*. Já a *Parcial* são todos os genes que existem em um determinado grupo e não existem no outro e o *Única* são os genes que existem exclusivamente em um sub grupo refinado de um determinado grupo. As saídas das análises comparativas foram usadas para as etapas seguintes do trabalho.

## 5 Implementação do Algoritmo Genético

O algoritmo genético foi implementado em linguagem Java 6, na IDE NetBeans 8.0.2. Os dados dos referentes aos genes foram obtidas no NCBI, após análises comparativas.

Todos os genes do *Core* Genoma foram definidos com peso 1 (pois são essenciais), os genes do *Parcial Core* foram definidos com peso 2, os genes *Exclusiva* de peixe América foi dado peso 1,5. Já para genes *Exclusiva* de peixe da Ásia foi dado peso 3, pois são genes que tem maior importância para este trabalho, uma vez que os peixes da Ásia afetando humanos da América seriam o caso mais grave devido às diferentes possibilidades de mutações e o atual despreparo para esta situação hipotética. Pelo mesmo motivo, os genes *Exclusiva* de humano da América possuem peso 3. Finalmente, aos genes *Exclusiva* de humano da Ásia foi dado peso 2.

Para uma bactéria ser factível, e mais fiel a uma possível bactéria real, foi considerado que a mesma deve ter tamanho entre 2000kb e 2400kb, deve possuir todo o core genoma do hospedeiro (humano), além de possuir entre 40% e 60% do *Parcial Core* (peixe) e deve possuir entre 40% e 70% do *Única Core*.

Para a seleção e foi utilizada a roleta simples e ranking. O número de pareamentos para o cruzamento foi definido como a metade do número de indivíduos da população. Se a população possui 100 bactérias, ocorrerão 50 pareamentos.

O *crossover* foi implementado considerando os genes *Parcial* e *Única* das bactérias, uma vez que o core genoma do hospedeiro deve estar totalmente presente na bactéria (sendo inclusive critério de factibilidade), não fazendo sentido utilizar *crossover* neste grupo de genes. O *Crossover* é realizado com 1 ponto de corte definido aleatoriamente para o *Parcial Core* e outro para o *Única core* da bactéria, podendo ser pontos diferentes.

Os parâmetros utilizados nas execuções do AG:

Tamanho da População: 250;

N. de Gerações: 400;

Taxa de Mutação: 0,3;

Taxa de Cruzamento: 0,8;

Critério de parada: Número máximo de gerações atingido.

## 6 Máquina de Inferência *Fuzzy*

As bactérias resultantes das execuções do algoritmo completo de AG foram classificadas em uma máquina *Fuzzy* de Mamdani. Este sistema possui três variáveis linguísticas de entrada e duas de saída. As variáveis linguísticas de entrada são o tamanho da bactéria, sua pontuação

(valor) de *Partial Core* e sua pontuação (valor) de *Unique Core*. A entrada (tamanho) que admite os termos primários pequeno, ideal e grande. Os três termos foram representados por funções triangulares, sendo que o termo ideal abrange de 1800kb a 2600kb, possuindo pertinência 1 em 2200kb. As variáveis de entrada *Score Partial* e *Score Unique* admitem os termos primários baixo, médio e alto, todas sendo representadas por funções triangulares. Para estas variáveis, quanto maior seu valor, melhor a bactéria.

As variáveis de saída foram adaptabilidade e probabilidade de sobrevivência. Ambas admitem os termos primários muito baixa, baixa, média, alta e muito alta. Procuramos então a bactéria com a melhor adaptabilidade e melhor probabilidade de sobrevivência possível. Os termos muito baixa e muito alta foram representados por funções trapezoidais, enquanto os termos baixa, média e alta foram representados por funções triangulares. As partições *Fuzzy* de saída admitem valores entre 0 e 1, indicando qual bactéria que possui melhor valor de adaptabilidade e probabilidade de sobrevivência.

## 7 Resultados Obtidos

O AG foi executado 10 vezes em sequencia devido tempo de execução e a limitação computacional da máquina utilizada. Após fim de cada execução foi selecionado a melhor bactéria gerada(elitismo). Em seguida as características de cada uma foram inseridas na máquina de inferência *Fuzzy*, criada no Matlab, e as bactérias foram classificadas considerando sua adaptabilidade e probabilidade de sobrevivência.

Após a classificação das bactérias pela máquina *Fuzzy*, as especialistas Microbiologia e Medicina também classificaram as bactérias considerando os mesmos fatores. Os resultados da classificação pela máquina *Fuzzy* e pelo especialistas podem ser vistos na Tabela 1, onde é mostrado para cada uma das execuções o tamanho da bactéria, sua pontuação do *Partial Core*, pontuação do *Unique Core*, classificação de 0 a 1 referente à adaptabilidade da bactéria pela máquina *Fuzzy*, classificação 0 a 1 referente à adaptabilidade da bactéria pela máquina *Fuzzy*, classificação quanto adaptação e sobrevivência pelo especialista, respectivamente. Os resultados demonstram que as bactérias geradas pelo AG possuem bons atributos tanto em adaptabilidade quanto em probabilidade de sobrevivência, sendo bons padrões para um estudo aprofundado de suas cargas genéticas e características, além das consequências de sua existência. Isso fica evidenciado ao perceber que todas as 10 bactérias alcançaram pontuação de mais de 60% para ambos os atributos considerados tanto pela classificação da máquina *Fuzzy* quanto pelas especialistas.

Tabela 1: Resultado das classificações das bactérias pelas especialistas e Fuzzy

Exec.	Tamanho	Score P.	Score U.	Nota Fuzzy		Nota Especial 1		Nota Especial 2	
				Adapt.	Sobreviv.	Adapt.	Sobreviv.	Adapt.	Sobreviv.
1	2395	3809	16,47	0,656	0,661	0,611		0,67	
2	2397	3809	18,50	0,657	0,714	0,62		0,73	
3	2400	3829	13,52	0,671	0,653	0,70		0,66	
4	2397	3799	17,72	0,621	0,685	0,60		0,69	
5	2400	3842	15,96	0,814	0,655	0,76		0,665	
6	2399	3821	16,05	0,653	0,654	0,605		0,665	
7	2399	3817	18,10	0,655	0,696	0,61		0,71	
8	2398	3825	17,42	0,656	0,677	0,611		0,68	
9	2398	3812	15,12	0,651	0,651	0,609		0,66	
10	2400	3816	17,51	0,654	0,678	0,61		0,685	

## 8 Conclusões

Aplicação da Máquina *Fuzzy* se mostrou muito útil para auxiliar na tomada de decisões e na classificação do perfil das bactérias, pois através da mesma foi possível condensar diversas regras,

criar um classificador funcional e tornando replicável a outros microrganismos. Observamos que após execuções dos algoritmos foi possível gerar mutantes híbridos humano e peixe factíveis que tiveram bons resultados de adaptabilidade e sobrevivência.

As opiniões das especialistas na classificações das bactérias foram bem próxima da classificação realizada pela máquina *Fuzzy*, isso demonstra que a mesma está bem calibrada para esta função. Dependendo do perfil desejado da bactéria e do objetivo, podem ser selecionadas diferentes bactérias. Por exemplo no caso de melhor bactéria com maior capacidade de adaptativa seria a bactéria 5 que possui valor de 0,814 de acordo com a máquina *Fuzzy* e 0,760 na opinião do especialista. Já no caso de melhor bactéria com maior capacidade de sobrevivência seria a bactéria 2 com valor de 0,714 atribuído pela máquina *Fuzzy* e 0,730 de acordo com o especialista.

Seria uma opção interessante para trabalhos futuros o desenvolvimento de outros algoritmos e meta heurísticas para comparar os resultados e assim com uma obtenção de uma nova máquina mais robusta pra melhores testes.

## Referências

- [1] Almeida, Paulo E. M., and Alexandre G. E., **Sistemas fuzzy**, SO Sistemas Inteligentes: fundamentos e aplicações, Cap 7, Manole,(2003): 169-201.
- [2] Amendola, Mariangela, and Laécio Carvalho Barros. **Manual do uso da teoria dos conjuntos Fuzzy no MATLAB 6.5**, Feagri e Imecc/Unicamp (2005): 1-44.
- [3] Gold, **GOLD Statistics 2019.**, Disponível em:<<https://gold.jgi-psf.org/statistics>>. Acesso em: 13 de Março de 2019.
- [4] Goldberg, David E., and John H. Holland., **Genetic algorithms and machine learning**, Machine learning 3.2 (1988): 95-99.
- [5] JR, M. J. P.; Chan, E.; Kreig, N. R., **Microbiologia: conceitos e aplicações**, Vol. 1., São Paulo, Makron Books, 1996.
- [6] Linden, R., **Algoritmos genéticos**, 2a edição, Brasport, 2008.
- [7] Lpsn, **List of Prokaryotic names**, Disponível em: <http://www.bacterio.net/streptococcus.html>. Acesso em: 13 de Março de 2019.
- [8] Madigan, M. T. et al., **Microbiologia de Brock** , 14a Edição, Artmed Editora, 2016.
- [9] NCBI, **PubMed Search Streptococcus**. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/?term=Streptococcus> , Acesso em: 13 de Março de 2019.
- [10] Shi-Kunne X, van Kooten M, et al., **The Genome of the Fungal Pathogen Verticillium dahliae Reveals Extensive Bacterial to Fungal Gene Transfer** Genome Biol Evol. 2019;11(3):855–868. doi:10.1093/gbe/evz040
- [11] Quinn, P., et al., **Microbiologia veterinária e doenças infecciosas**, [S.l.], Artmed Editora, 2005.
- [12] Tortora, G. J.; Case, C. L.; Funke, B. R., **Microbiologia** 12a Edição, Artmed Editora, 2016
- [13] Venter, J.C., Adams, M.D., Myres, E.W. et al., **The Sequence of the Human Genome**, Science. 291(5507): 1304-1351.

## Modelo SIRC Fracionário: Comparação com dados reais

Ana Carla F. N. Gomes,      **Adriano de Cezaro**,

Instituto de Matemática, Estatística e Física, IMEF, FURG,  
96203-900, Rio Grande, RG

E-mail: anagomes.mat@gmail.com,    adrianocezaro@furg.br.

**Resumo:** Neste trabalho mostraremos que, para alguns parâmetros específicos e uma escolha da ordem das derivadas feitos pelo método da exaustão, o modelo SIRC fracionário representa mais fidedignamente dados reais de influenza A, obtidos do DATASUS-RS, quando comparado com o modelo SIRC com derivadas ordinárias proposto por [1]. Para além de comparações, provamos que o modelo SIRC fracionário proposto é bem posto no sentido de Hadamard e que a sua solução pode ser estendida a toda a semirreta positiva.

**Palavras-chave:** Modelo Compartimental SIRC Fracionário, Boa Colocação, Comparação com Dados Reais, Influenza A.

### Introdução

A influenza é uma doença respiratória infecciosa de origem viral em que seu agravamento pode levar ao óbito, especialmente nos indivíduos que apresentam fatores ou condições de risco para as complicações da infecção. Nas pessoas acima de 60 anos, por exemplo, destaca-se como uma das principais causas de mortalidade e importante causa de morbidade, frequentemente necessitando de hospitalização [5].

Tendo em vista a vastidão do grupo de risco da influenza, o fato de ser uma doença de fácil contaminação e de espalhamento rápido, a quantidade de pessoas afetadas é preocupante. No ano de 2011, por exemplo, foram registradas 750.006 internações por influenza e pneumonia [4].

Desta forma, as influenzas formam um panorama de riscos importante para a população que, na visão dos governantes, representam um quadro oneroso aos cofres públicos. Justificamos assim a necessidade de estudar e entender a dinâmica de propagação das mesmas.

Nesta contribuição daremos ênfase à modelagem matemática do problema que envolve o vírus da influenza A, que, geralmente, tem uma pequena mutação a cada ano, com relação à parte que não sofre mutação, chamada de cepa [1].

As influenzas são tradicionalmente descritas por modelos compartimentais do tipo SIR [2, 1]. Neste tipo de modelo, o total da população é segmentada em três compartimentos: suscetível (S), infeccioso (I), recuperado (R). No entanto, a capacidade de mutação dos vírus de influenza justifica a introdução de um novo compartimento ao modelo SIR, chamado de imunidade cruzada (C). Neste, são considerados a parcela da população que não são suscetíveis ao mesmo subtipo do vírus, mas a uma mutação deste. Tal modelo compartimental é conhecido na literatura por SIRC, introduzindo originalmente por Casagrandi et al. [1].

As epidemias virais como a influenza A só não se tornam pandemias graças à capacidade de nosso sistema imunológico de reconhecer, mesmo que parcialmente, os vírus que já foi exposto e reagir a esses de forma a dificultar o contágio novamente. A essa capacidade chamamos de memória imunológica [9].

O modelo SIRC da forma que foi proposto por [1] não possui a capacidade de reproduzir os efeitos de memória imunológica, haja vista que o operador diferencial ordinário só reflete condições locais [3].

**Novidades desta contribuição:** Uma das principais novidades de nossa proposta baseia-se na introdução de memória através da utilização de uma dinâmica fracionária no modelo SIRC, obtida a partir de operadores diferenciais não locais do tipo Caputo [3].

Do ponto de vista teórico, mostraremos que o modelo SIRC fracionário é bem-posto no sentido de Hadamard. Além disso, garantimos que a solução do modelo SIRC fracionário possui uma extensão para a semirreta positiva. Por outro lado, apresentaremos uma comparação entre o modelo SIRC fracionários e dados reais de influenza A, obtidos do DATASUS-RS. Verificamos que, a partir de uma escolha de parâmetros (não ótima e feita utilizando o método da exaustão) e da ordem fracionária da derivada para a dinâmica, o modelo SIRC fracionário representa melhor os dados reais, que o modelo SIRC com derivadas ordinárias.

**Estrutura do artigo:** Dividimos esta contribuição da seguinte forma: na Seção 1, apresentaremos o modelo SIRC com dinâmica fracionária. Na Seção 2, explicitaremos resultados de boa colocação e extensão da solução do problema de valor inicial (PVI) fracionário. Na Seção 3, mostraremos os resultados numéricos comparativos entre o modelo SIRC e SIRC fracionário com relação aos dados reais. Na Seção 4, apresentaremos nossas conclusões e trabalhos futuros.

## 1 Modelo SIRC Fracionário

Mudanças de comportamento em indivíduos como resposta às modificações ambientais são uma característica essencial à vida. Tais mudanças tornam-se mais evidentes nos casos de epidemias repetitivas, quando ocorre um efeito que chamamos de memória imunológica, e.g., [9] e referências. Em outras palavras, a memória imunológica é a capacidade que o sistema imune adaptativo tem de responder de forma mais rápida e eficiente quando ocorre exposição à antígenos encontrados previamente.

Nesta contribuição visamos a introdução de memória na dinâmica compartimental SIRC através de derivadas de ordem fracionária do tipo Caputo [3], com o objetivo de levar em conta a memória epidemiológica existente naturalmente na modelagem de influências. Desta forma, nos propomos a estudar o modelo fracionário

$$\begin{aligned} D_*^\theta(S(t)) &= \mu(N - S(t)) - \beta S(t)I(t) + \gamma C(t) \\ D_*^\theta(I(t)) &= \beta S(t)I(t) + \sigma\beta C(t)I(t) - (\mu + \alpha)I(t) \\ D_*^\theta(R(t)) &= (1 - \sigma)\beta C(t)I(t) + \alpha I(t) - (\mu + \delta)R(t) \\ D_*^\theta(C(t)) &= \delta R(t) - \beta C(t)I(t) - (\mu + \gamma)C(t), \end{aligned} \quad (1)$$

onde o parâmetro  $\mu > 0$  representa a taxa de mortalidade e natalidade. Os parâmetros  $\alpha > 0$ ,  $\delta > 0$  e  $\gamma > 0$  são os inversos do tempo em que os indivíduos ficam respectivamente nos compartimentos I, R e C. Já  $\sigma > 0$ , representa a probabilidade média de reinfeção do indivíduo que esteja no compartimento C e  $\beta > 0$  é a taxa de contato entre indivíduos dos compartimentos S e I. Estes parâmetros fazem com que as populações fluam entre os compartimentos, mudando seus estados perante a doença.  $D_*^\theta$  representa o operador diferencial de Caputo [3], para  $\theta \in ]0, 1]$ .

Nesta abordagem iremos considerar a dinâmica fracionária (1) com as seguintes hipóteses:

**H1)**  $\theta \in ]0, 1]$  é a ordem fracionária do operador diferencial de Caputo  $D_*^\theta$ .

**H2)** As condições iniciais para a dinâmica (1) são dadas pelas quantidades  $S(0) = S_0 > 0$ ,  $I(0) = I_0 \geq 0$ ,  $R(0) = R_0 \geq 0$  e  $C(0) = C_0 \geq 0$ , respectivamente.

**H3)** A taxa de mortalidade e natalidade  $\mu > 0$  são consideradas iguais.

**H4)** Os parâmetros do PVI fracionário (1) são todos constantes e não negativos.

Das hipóteses **H1)** - **H4)**, podemos observar os seguintes fatos, os quais afetam também a calibragem dos parâmetros do modelo.

**Observação 1.** Da hipótese **H3)**, segue-se que o total da população  $N = S + I + R + C$  permanece constante. Disto segue que o inverso da taxa de natalidade (mortalidade)  $\mu^{-1}$  pode

ser calibrado com a média do tempo de vida da população e, por isso, podemos supor que  $\mu$  é bem conhecida.

Da hipótese **H4**), segue que os demais parâmetros no modelo (1) podem ser estimados através de observações clínicas mais prolongadas. Nessa abordagem faremos a calibração dos parâmetros utilizando valores da literatura e um método de exaustão.

## 2 Boa Colocação para o modelo SIRC fracionário

Nessa seção, mostraremos a boa colocação do PVI fracionário (1), bem como a existência de solução estendida à semirreta positiva.

Por simplicidade, usaremos a seguinte notação na parte restante deste trabalho. Por  $u(t) = (S(t), I(t), R(t), C(t))^T$ , denotamos a função real a valores em  $\mathbb{R}^4$ , cujas componentes são as dinâmicas dos indivíduos suscetíveis  $S$ , infectados  $I$ , recuperados  $R$  e imunidade cruzada  $C$ , respectivamente. Em particular  $u(0) = (S(0), I(0), R(0), C(0))^T := u_0$  corresponde às condições iniciais do PVI fracionário (1) dadas pela hipótese **H2**). Além disso, denotaremos por  $f(t, u(t)) := (\mu(N - S(t)) - \beta S(t)I(t) + \gamma C(t), \beta S(t)I(t) + \sigma\beta C(t)I(t) - (\mu + \alpha)I(t), (1 - \sigma)\beta C(t)I(t) + \alpha I(t) - (\mu + \delta)R(t), \delta R(t) - \beta C(t)I(t) - (\mu + \gamma)C(t))^T$  a função vetorial correspondente ao lado direito do PVI fracionário (1).

Com essas notações, nós podemos reescrever o PVI fracionário (1) com os valores das condições iniciais dadas pela hipótese **H2**) e  $\theta \in ]0, 1]$ , como o seguinte problema de valor inicial fracionário

$$D_*^\theta u(t) = f(t, u(t)), \quad u(0) = u_0. \quad (2)$$

Primeiramente, provaremos o seguinte lema auxiliar.

**Lema 1.** *Assuma que as hipóteses **H1**) - **H4**) sejam verdadeiras. Então, a matriz Jacobiana  $J := J(f(t, u(t)))$  correspondente à função vetorial  $f(\cdot, \cdot)$  que representa ao lado direito do PVI fracionário (2) é uniformemente limitada por uma constante  $M$ , independente de  $S, T, I, R$ , no intervalo  $[0, T]$ .*

*Demonstração.* Um cálculo direto mostra que a matriz Jacobiana  $J$  é representada por:

$$J = \begin{bmatrix} -\mu - \beta I(t) - \rho & -\beta S(t) & 0 & \gamma \\ \beta I(t) & \beta S(t) + \sigma\beta C(t) - (\mu + \alpha) & 0 & \sigma\beta I(t) \\ \rho & (1 - \sigma)\beta C(t) + \alpha & -(\mu + \delta) & (1 - \sigma)\beta I(t) \\ 0 & -\beta C(t) & \delta & -\beta I(t) - (\mu + \gamma) \end{bmatrix}$$

Segue-se de [7, Teo. 9] que cada componente de  $J$  é contínua em  $[0, T]$ . Como  $[0, T]$  é compacto, o Teorema de Weierstrass implica na existência de uma constante  $M$ , independente de  $S, I, R, C$ , de tal modo que  $\|J\| \leq M$ .  $\square$

Como forma de completude do trabalho, segue abaixo a definição da condição de Lipschitz com relação à segunda variável, para uma função a valores vetoriais.

**Definição 1.** *Seja  $f : U \rightarrow E$  uma função definida em um conjunto aberto  $U \subset I \times E$ , onde  $I$  é um intervalo não degenerado da reta real,  $E$  é um espaço normado. Dizemos que  $f$  é Lipschitz contínua com relação à segunda variável, se a seguinte condição é satisfeita;*

$$\|f(t, u(t)) - f(t, v(t))\|_E \leq L |u(t) - v(t)|_U,$$

para alguma constante  $L > 0$  independente de  $(t, u(t)), (t, v(t)) \in U$ . Onde,  $\|\cdot\|_E$  e  $|\cdot|_U$  representam as normas nos espaços vetoriais  $E$  e  $U$ , respectivamente.

Os resultados a seguir são essenciais para garantir a boa colocação do PVI fracionário (1).

**Lema 2.** *Consideremos as hipóteses **H1**) - **H4**) verdadeiras. Então a função vetorial  $f(t, u(t))$  em (2) é contínua com relação a  $t \in [0, T]$  e Lipschitz contínua com respeito à segunda coordenada, como na Definição (1).*

*Demonstração.* O resultado de continuidade enunciado segue da continuidade das operações e funções envolvidas na definição de  $f(\cdot, \cdot)$  que representa o lado direito de (1). Além disso, o teorema do valor médio [8] implica que

$$\|f(t, u(t)) - f(t, v(t))\| \leq \|J(su(t) + (1-s)v(t))\| \|u(t) - v(t)\|.$$

Assim, a afirmativa do lema segue agora do resultado de limitação uniforme da matriz Jacobiana  $J$  dada pelo Lema (1).  $\square$

**Lema 3.** *Suponha que as hipóteses **H1** - **H4** sejam verdadeiras. Seja  $f(t, u(t))$  e  $u(t)$  como definidos no Lema 2. Então, existem constantes  $c_1, c_2$  e  $M$  (que depende apenas da população total  $N$  e dos parâmetros no lado direito do PVI fracionário (1)) e um parâmetro  $0 \leq s < 1$  de tal modo que*

$$\|f(t, u(t))\| \leq c_1 + c_2 \|u\|^s.$$

Além disso  $\|f(t, u(t))\| \leq M$  uniformemente em  $[0, +\infty[$ .

*Demonstração.* Primeiramente, é importante notar que a população total  $N$  é constante como consequência da hipótese **H3**. Podemos concluir assim que  $\|u\|$  é uniformemente limitada por  $N$ . Portanto, a prova da primeira hipótese deste Lema implica a segunda. Para provar a primeira afirmação, vamos observar que, para qualquer  $s \in ]0, 1]$ , temos  $\|u\| = \|u\|^{1-s} \|u\|^s$ . Haja visto que  $N$  é constante (da hipótese **H3**), podemos concluir a existência de  $\tilde{c}$  (que depende somente de  $N$ ) de tal modo que  $\|u\| = \tilde{c} \|u\|^s$ . Portanto, para concluir o Lema, é suficiente provar a primeira hipótese para  $s = 1$ . Para isso, detalhamos o cálculo da primeira coordenada de  $f(\cdot, \cdot)$ , já que a ideia pode ser replicada de maneira análoga para as demais componentes. Note que

$$\begin{aligned} \|\mu(N - S(t)) - \beta S(t)I(t) + \gamma C(t)\| &\leq \mu|N| + \mu|S(t)| + \beta|S(t)||I(t)| + \gamma|C(t)| + \leq \\ &\mu|N| + (\mu + \beta|N|)|S(t)| + \gamma|C(t)| \leq \mu|N| + \max\{1, (\mu + \beta|N|), \gamma\} \|u\|. \end{aligned}$$

No cálculo acima, fica clara a existência de constantes  $c_1$  e  $c_2$  (dependendo apenas dos parâmetros do sistema e de  $N$ ) satisfazendo  $\|f(t, u(t))\| \leq c_1 + c_2 \|u\|^s$ , concluindo a prova.  $\square$

Agora estamos prontos para provar a boa colocação do modelo SIRC fracionário (1).

**Teorema 1.** *Suponha que as hipóteses **H1** - **H4** sejam verdadeiras. Então, existe um  $0 < T_* < T$  tal que o PVI fracionário (1)-**H2** possui uma única solução contínua  $u(t) := (S(t), I(t), R(t), C(t))^T$  para  $t \in [0, T_*]$ . Além disso, a solução  $u$  depende continuamente dos dados iniciais, dos parâmetros e da ordem da derivada  $\theta \in ]0, 1]$ .*

*Demonstração.* Como consequência do Lema (2) obtemos que a função  $f(t, u(t))$ , representando o lado direito do PVI fracionário (1), é contínua e Lipschitz em  $[0, T]$ . Assim, segue de [3, Teo. 6.5], [3, Teo. 6.20], [3, Teo. 6.21] e [3, Teo. 6.22], a existência de  $T_* = T_*(M) \in ]0, T[$  de modo que o PVI fracionário (1), com condições iniciais como na hipótese **H2**, possui uma única solução contínua  $u(t)$  para  $t \in [0, T_*]$ , que depende continuamente dos dados iniciais, dos parâmetros e da ordem da derivada, respectivamente.  $\square$

Note que a magnitude de  $T_*$  dada pelo Teorema 1 pode ser (em geral é) muito pequena, inviabilizando qualquer aplicação na vida real. O próximo resultado garante uma extensão contínua da solução de (1) para a semirreta positiva.

**Teorema 2.** *Suponha as hipóteses do Teorema 1 verdadeiras. Então, a solução do PVI fracionário (1) -**H3** possui uma única extensão contínua ao intervalo  $[0, +\infty[$ .*

*Demonstração.* Segue do Lema 3 que  $\|f(t, u(t))\| \leq c_1 + c_2 \|u\|^s$  para algum  $s \in [0, 1[$ . Além disso, o Lema 1 garante que a matriz jacobiana  $J$  é uniformemente limitada. Por fim, do Lema 2 obtemos que  $f$  é Lipschitz contínua com relação à segunda variável, em  $R^4$ . Agora, o resultado enunciado segue diretamente de [3, Corolário 6.7].  $\square$

### 3 Resultados Numéricos

Nesta seção apresentaremos alguns testes numéricos comparando o comportamento do PVI fracionário (1)-**H3**) com relação a um conjunto de dados reais de indivíduos infectados pela Influenza A.

Os dados reais são do estado do Rio Grande do Sul do ano de 2010, obtidos pelo sistema DATASUS, sendo ao todo 34 e fornecidos por semana epidemiológica. Apesar de serem poucos em quantidade e esparsos são os dados que se tem conhecimento sobre o contágio da Influenza A.

Por serem registrados semanalmente, optamos por usar eles uniformemente distribuídos, assim nas simulações não faz sentido usar qualquer unidade de tempo no modelo que não reflita tal distribuição. Estes dados são representados na Figura 1 (a) e (b) como os pontos em azul.

Os parâmetros do modelo foram calibrados usando o método de exatidão e são apresentados na Tabela 1.

Parâmetros	Casagrandi et al.	Parâmetros Modificados
$\mu$	0,02	0,02
$\alpha$	1,355	0,27
$\delta$	0,75	0,75
$\gamma$	0,35	0,35
$\sigma$	0,12	0,12
$\beta$	52,14	10,43
$N$	100	100

Tabela 1: Parâmetros do modelo SIRC.

Para as simulações, utilizamos um método numérico baseado em adaptações do Método Trapezoidal para operadores fracionários. Detalhes de convergência e estabilidade do método numérico podem ser encontrados em [6]. O intervalo de tempo é usado de maneira a ter esta estabilidade e o passo utilizado foi  $10^{-4}$ .

Utilizamos como ordem da derivada  $\theta = 1$ , para mostrar o que ocorre com a dinâmica da doença conforme os resultados de [1]. Para analisarmos o comportamento do modelo com efeitos de memória epidemiológica, sabemos que  $\theta \in ]0, 1]$  e quanto menor, maior fica o efeito de memória.

Conforme melhor discutido em [7], escolhemos  $\theta = 0,9$  como ordem da derivada, para a memória epidemiológica estar presente, mas não ser muito efetiva, já que não estamos propondo nenhum tipo de interferência externa nesta simulação, como seria o caso um modelo com vacinação.

A Figura 1 (a) mostra a evolução do compartimento dos indivíduos infectados  $I(t)$  do modelo SIRC e SIRC fracionário (1), com  $\theta = 1$  e com  $\theta = 0,9$ , respectivamente. Para esta simulação utilizamos os parâmetros do modelo retirados de [1] e que podem ser vistos na coluna Casagrandi et al. da Tabela 1.

Ao compararmos a evolução das dinâmicas dos indivíduos infectados (utilizando o modelo SIRC e SIRC fracionário) percebemos que ambos os modelos não correspondem de maneira fidedigna os dados reais.

Tendo em vista este fato e levando em conta as características de localidade da dinâmica evolutiva da Influenza A, propusemos uma calibragem dos parâmetros mais sensíveis do modelo (1), a saber  $\alpha$  e  $\beta$ , como podem ser vistos na coluna Parâmetros modificados da Tabela 1.

A calibragem foi feita utilizando um método de exatidão, isto é, propomos valores para os parâmetros  $\alpha$  e  $\beta$  mantendo uma proporção com o correspondente na coluna Casagrandi et al. da Tabela 1 e comparamos os resultados simulados dos modelos SIRC e SIRC fracionário com os dados reais.

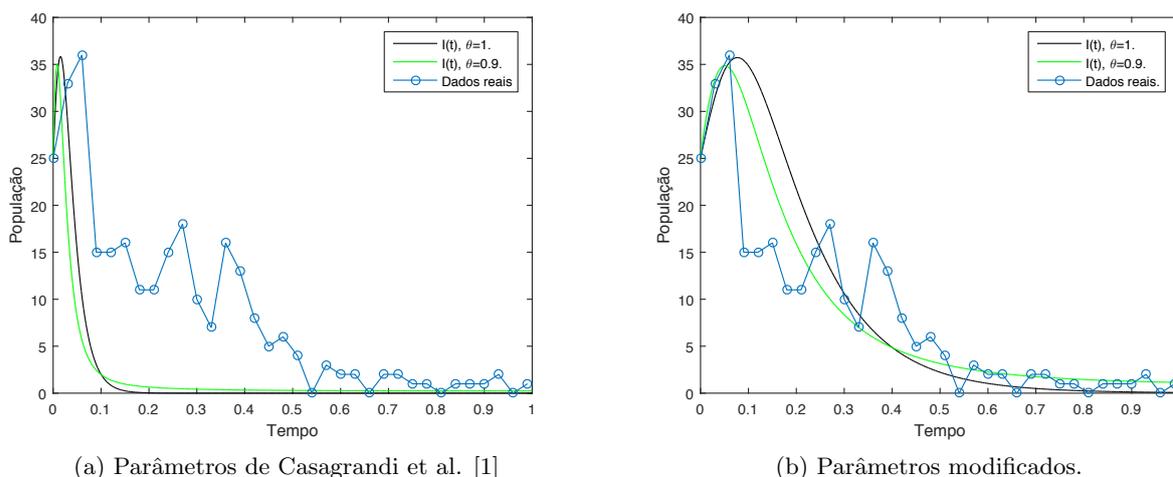


Figura 1: Comparação de  $I(t)$ , para diferentes valores de  $\theta$ , com os dados reais.

Percebemos que à medida que diminuíamos os valores de  $\alpha$  e  $\beta$ , as simulações para o compartimento dos infectados  $I$  aproximava-se mais dos dados reais, até chegarmos nos parâmetros dados pelos valores na coluna Parâmetros modificados da Tabela 1.

Não apresentaremos aqui todas as simulações, por questão de espaço. Optamos por expor a simulação com os parâmetros originais (coluna Casagrandi et al. na Tabela 1) e a melhor aproximação obtida (coluna Parâmetros modificados da Tabela 1).

E que percebemos no decorrer da aplicação do método é que para valores relativamente menores que os apresentados na coluna Parâmetros modificados da Tabela 1, a simulação voltava a se distanciar dos dados reais, fazendo com que considerássemos esse conjunto de parâmetros como o melhor para o modelo em comparação com esses dados reais.

O resultado da simulação para os modelos SIRC e SIRC fracionário (com  $\theta = 0,9$ ) da dinâmica do compartimento dos indivíduos infectados  $I$  em comparação aos dados reais podem ser vistos na Figura 1 (b). A Figura 3 mostra o erro relativo entre os modelos, comparando o compartimento dos infectados  $I$  do modelo SIRC e SIRC fracionário (com  $\theta = 0.9$ ) com os dados reais.

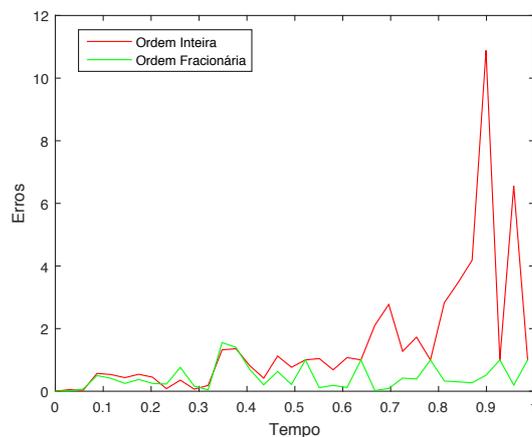


Figura 2: Gráfico de erros das simulações.

Haja vista que o erro relativo das simulações para o compartimento dos indivíduos infectados  $I$  para o modelo SIRC fracionário ( $\theta = 0.9$ ) é sempre menor que para o modelo SIRC, podemos comprovar nossa tese: o modelo SIRC fracionário é uma alternativa interessante para modelar

a dinâmica de epidemias como a Influenza A. Para além disso, podemos concluir que o erro relativo do modelo SIRC fracionário é ainda menor quando o tempo passa. Alegamos que esse fenômeno se deve aos efeitos de memória inerentes à epidemias de Influenza e que são capturados com mais intensidade pela dinâmica fracionária quando  $t \gg 0$ .

## 4 Conclusões

Neste trabalho estudamos o modelo SIRC fracionário como uma proposta de modelagem para a dinâmica da influenza A. Provamos que o modelo SIRC fracionário é bem posto no sentido de Hadamard e que a solução pode ser estendida de maneira única para a semirreta positiva. Além disso, fizemos comparações entre os modelos SIRC e SIRC fracionário com dados reais.

As simulações numéricas mostraram que, a partir de uma escolha dos parâmetros feito pelo método da exaustão, o modelo SIRC fracionário representa melhor os dados reais que a dinâmica de ordem inteira (modelo SIRC).

Os trabalhos futuros são a calibração dos parâmetros usando técnicas de minimização, e.g. quadrados mínimos, buscar outros estudos que tragam parâmetros melhores e a implementação de vacinação ao modelo SIRC.

Em particular, a vacinação para a influenza já é fornecida pelo SUS, principalmente para os grupos de risco. Portanto, é de interesse global entender o que acontece com a dinâmica da doença quando parte da população é vacinada [4]. A vacinação é mais uma forma de promover memória imunológica ao sistema, justificando assim a utilização de dinâmicas fracionárias para descrever o fenômeno.

## Referências

- [1] R. Casagrandi et al, The SIRC model and influenza A, *Mathematical biosciences*, 200, (2006), 152-169.
- [2] O. Diekmann, J. A. P. Heesterbeek e J. A. Metz, The legacy of Kermack and McKendrick, *Publications of the Newton Institute*, 5 (1995) 95-115.
- [3] K. Dirtheim, “The analysis of fractional differential equations: An application-oriented exposition using differential operators of Caputo type”, Springer Science and Business Media, 2010.
- [4] P. M. S. B. Francisco et al, Impacto da vacinação contra influenza na mortalidade por doenças respiratórias em idosos, *Revista de Saúde Pública*, 39 (2005) 75-81.
- [5] P. M. S. B. Francisco et al, Tendência da mortalidade por doenças respiratórias em idosos do Estado de São Paulo, 1980 a 1998, *Revista de Saúde Pública*, 37 (2003) 191-196.
- [6] R. Garrappa, Trapezoidal methods for fractional differential equations: Theoretical and computational aspects, *Mathematics and Computers in Simulation*, 110 (2015) 96-112.
- [7] A. C. F. N. Gomes e A. De Cezaro, Um estudo sobre a memória epidemiológica: modelo SIRC fracionário, *Revista Eletrônica Paulista de Matemática*, 10 (2017) 194-210.
- [8] E. L. Lima, “Análise real Vol. 2”, Impa, 2015.
- [9] A. Pimenov et al, Memory effects in population dynamics: spread of infectious disease as a case study, *Mathematical Modelling of Natural Phenomena*, 7 (2012) 204-226.



# Implementação de um programa computacional para o diagrama de bifurcações de sistemas caóticos

**João I. M. Bezerra,**

Centro de Desenvolvimento Tecnológico, CDTEC, UFPEL,  
Campus Porto, 96010-610, Pelotas, RS  
E-mail: jimbezerra@inf.ufpel.edu.br,

**Alexandre Molter**

Instituto de Física e Matemática, IFM, UFPEL,  
Campus Universitário, 96160-000, Capão do Leão, RS  
E-mail: alexandre.molter@ufpel.edu.br

**Resumo:** *O estudo do comportamento de sistemas dinâmicos que apresentam comportamento caótico vem atraindo a atenção de pesquisadores nas diferentes áreas. Nas últimas décadas, o desenvolvimento computacional proporcionou estudos mais detalhados do comportamento destes sistemas. No entanto, os programas em que são simuladas as características destes sistemas normalmente são ocultados, não dando acesso dos códigos aos usuários. Tendo em vista o acesso a um código explicado e detalhado, o objetivo deste trabalho é apresentar um programa computacional para o diagrama de bifurcações de sistemas contínuos que apresentam comportamento caótico.*

**Palavras-chave:** *Comportamento Caótico, Diagrama de Bifurcações, Programa Computacional*

## 1 INTRODUÇÃO

A análise de sistemas dinâmicos que apresentam comportamento caótico tem sido amplamente divulgada nos meios científicos. Dentre as características analisadas nestes sistemas estão o diagrama de bifurcações, as seções de Poincaré e os expoentes de Lyapunov, todas elas usadas para analisar se o sistema apresenta, ou não, comportamento caótico. Há diversos trabalhos que apresentam simulações computacionais destas análises, dentre os quais pode-se destacar [4], [7], [2] e [9].

Normalmente, as implementações destas ferramentas, que aparecem nestes trabalhos, são feitas por intermédio de programas fechados, os quais não dão ao usuário acesso aos códigos. [7], por exemplo, utilizou o pacote MATCONT2.5.1, do software MATLAB. O uso destes programas, embora forneça soluções precisas, pode dificultar a compreensão destas análises feitas nos sistemas aos acadêmicos, em especial para os iniciantes nesta área.

Tendo em vista a dificuldade de acesso a códigos computacionais abertos e explicativos, este trabalho tem como objetivo disponibilizar um programa aberto para a construção do diagrama de bifurcações de sistemas de tempo contínuo que apresentam comportamento caótico. O trabalho disponibiliza o passo a passo como este diagrama é construído e sua implementação no MATLAB.

São apresentados exemplos da implementação do diagrama de bifurcações dos sistemas de Rössler [6], Lorenz [5] e de um sistema presa-predador [4], [7], e discutidas as particularidades da implementação para cada um destes sistemas.

## 2 DIAGRAMA DE BIFURCAÇÕES

Antes de se apresentar o algoritmo para a construção do diagrama de bifurcações, será feita uma breve discussão sobre o aparecimento das bifurcações nos sistemas dinâmicos. Uma bifurcação está relacionada com a mudança qualitativa das soluções do sistema no plano de fase, à medida que o parâmetro analisado do sistema passa por um valor crítico, podendo ser classificada em local ou global. A bifurcação local está relacionada com a mudança na estabilidade do sistema de acordo com as raízes do polinômio característico, onde pode ocorrer mudança qualitativa no comportamento das soluções perto dos pontos de equilíbrio, podendo estes pontos serem calculados matematicamente, enquanto a bifurcação global prevê mudanças nas órbitas tais como a periodicidade ou não das soluções, não podendo ser calculadas matematicamente. O objetivo deste trabalho é justamente a disponibilização de um algoritmo para a construção do diagrama que permite a visualização destas bifurcações, que são importantes no estudo destes sistemas, especialmente quando apresentam comportamento caóticos.

No caso de instabilidade, os valores das variáveis do sistema irão apresentar comportamento oscilatório ao longo do tempo. Estas oscilações podem ser periódicas, ou não. No caso de ocorrer a periodicidade, significa que o sistema terá comportamento previsível, não sendo sensível às condições iniciais. Por outro lado, quando as oscilações forem aperiódicas, o sistema será imprevisível e sensível às condições iniciais. A figura abaixo mostra a diferença entre oscilações: uma oscilação periódica, na qual há sempre o mesmo ponto de máximo, e uma oscilação não periódica, em que estes pontos variam. A periodicidade, ou não, destas oscilações, pode ser observada nos pontos de máximo ou de mínimo ao longo da trajetória.

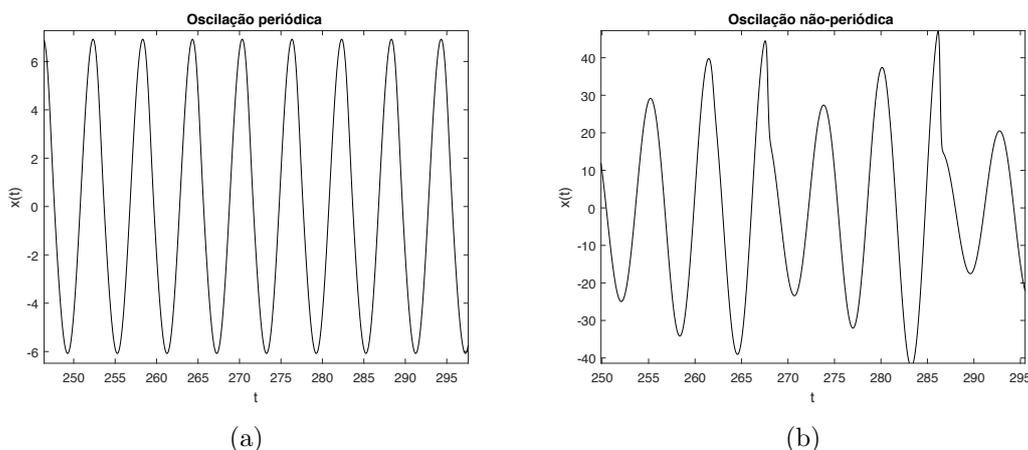


Figura 1: (a) Oscilação Periódica; (b) Oscilação não-periódica

A construção do diagrama começa com a escolha do parâmetro a ser variado, e seu intervalo de variação. Após, para cada valor dentro deste intervalo, são feitos os passos que seguem:

- O sistema é integrado.
- Os resultados da integração são analisados. Mesmo que a trajetória seja periódica, inicialmente haverá um período de oscilações não periódicas até que ocorra o estabelecimento da periodicidade. Este período é denominado transiente. Portanto, a análise dos resultados deve ser feita após o período transiente.
- O resultado da integração é uma matriz, onde cada linha representa um instante de tempo e cada coluna representa uma variável. No período que deve ser analisado, verificar se os valores da variável de interesse são valores de máximo [8]. Isso pode ser feito comparando o valor da coluna da variável (1 para  $x$ , 2 para  $y$ , 3 para  $z$ ) no instante de tempo corrente, ou seja, na linha analisada, é maior do que o valor da linha anterior como o da próxima

linha. Caso este valor seja um valor máximo, plote ele em função do valor do parâmetro no diagrama. Se desejado, também pode-se plotar o valor de mínimo.

Abaixo, segue o algoritmo.

<b>Algoritmo:</b> diagrama de bifurcações	
<b>Entrada:</b> sistema, intervalo de integração, duração do transiente, intervalo de variação do parâmetro, variável de interesse, condições iniciais	
1	<b>enquanto</b> <i>intervalo de variação do parâmetro</i> <b>faça</b>
2	Integre o sistema;
3	Selecione os valores após o transiente;
4	<b>enquanto</b> <i>transiente</i> <b>faça</b>
5	<b>se</b> <i>ponto de máximo</i> <b>então</b>
6	Plote no diagrama;
7	<b>fim</b>
8	<b>fim</b>
9	<b>fim</b>
<b>Saída:</b> diagrama de bifurcações	

Em anexo, no final do trabalho, está a implementação deste algoritmo em MATLAB. A seguir, exemplos da implementação do algoritmo nos sistemas de Lorenz, Rössler e presa-predador. As variáveis utilizadas no diagrama de bifurcações são as encontradas nas bibliografias([7], [1]), para fins de comparação.

### 3 APLICAÇÕES E DISCUSSÕES

#### 3.1 SISTEMA DE LORENZ

O sistema de Lorenz [5], que representa um modelo simples de convecção térmica de um fluido em uma caixa, é descrito pelas seguintes equações:

$$\begin{aligned} \dot{x} &= \sigma(y - x), \\ \dot{y} &= rx - y - xz, \\ \dot{z} &= xy - \beta z, \end{aligned} \tag{1}$$

A variável  $x$  em (1) está relacionada à intensidade do movimento do fluido, enquanto as variáveis  $y$  e  $z$  estão relacionadas às variações de temperatura nas direções horizontal e vertical do movimento. As equações de Lorenz envolvem, também, três parâmetros,  $\sigma$ ,  $r$ ,  $\beta$ , todos reais e positivos.  $\sigma$  e  $\beta$  dependem do material e das propriedades geométricas do fluido. Para a análise do sistema, usa-se  $\sigma = 10$  e  $\beta = \frac{8}{3}$ . O parâmetro  $r$  é proporcional a diferença de temperatura  $\Delta T$  entre as extremidades da caixa, e aqui será investigado o comportamento das soluções para diferentes valores, à medida que variam os valores de  $r$ .

Após a obtenção do polinômio característico do sistema no ponto crítico

$$(\sqrt{\beta(r-1)}, \sqrt{\beta(r-1)}, r-1)$$

e análise de sua estabilidade pelo método de Routh-Hurwitz, tem-se que o sistema será instável para  $r \geq 24,7$ . Portanto, o diagrama de bifurcações será construído para valores de  $r$  neste intervalo.

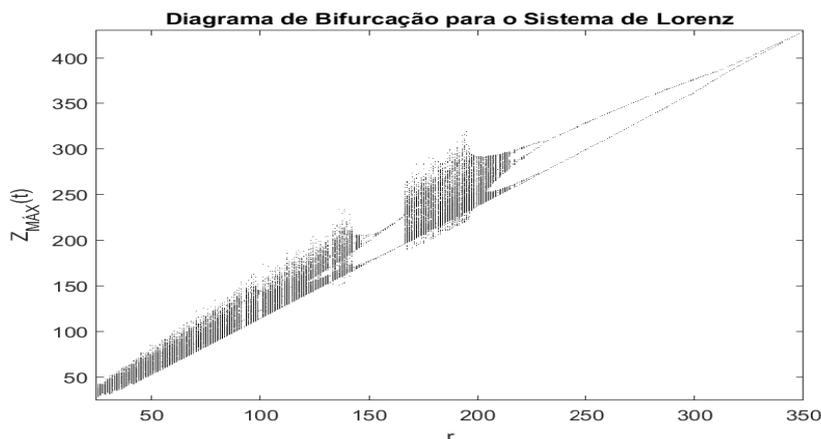


Figura 2: Diagrama de bifurcações para o sistema de Lorenz

Na figura acima, percebe-se que o sistema apresenta comportamento aperiódico até  $r \approx 160$ , a partir de quando os pontos de máximo oscilam entre dois valores. Por volta de  $r = 180$ , novamente o sistema se comporta de forma aperiódica, até aproximadamente  $r = 240$ , onde os pontos de máximo novamente oscilam entre dois valores, e posteriormente em  $r \approx 340$ , onde passa a ter apenas um valor de máximo.  $Z_{MÁX}(t)$  são os valores de máximo da variável de estado  $z(t)$  do sistema de Lorenz.

### 3.2 SISTEMA DE RÖSSLER

O sistema de Rössler [6] é um sistema não linear, autônomo, tridimensional, descrito pelas seguintes equações:

$$\begin{aligned} \dot{x} &= -y - z, \\ \dot{y} &= x + ay, \\ \dot{z} &= b + (x - c)z, \end{aligned} \tag{2}$$

Os parâmetros do sistema são  $a$ ,  $b$  e  $c$ , todos reais e positivos. Para a análise do sistema,  $a = b = 0,1$ , e o parâmetro  $c$  será variado, com o objetivo de se estudar como o sistema irá se comportar para diferentes valores deste parâmetro. Pela análise feita pelo critério de Routh-Hurwitz, para  $1 \leq c \leq 45$ , o sistema será instável, e o intuito é analisar para quais valores neste intervalo o sistema apresenta comportamento periódico. Isto é analisado de acordo com o diagrama de bifurcações que segue.

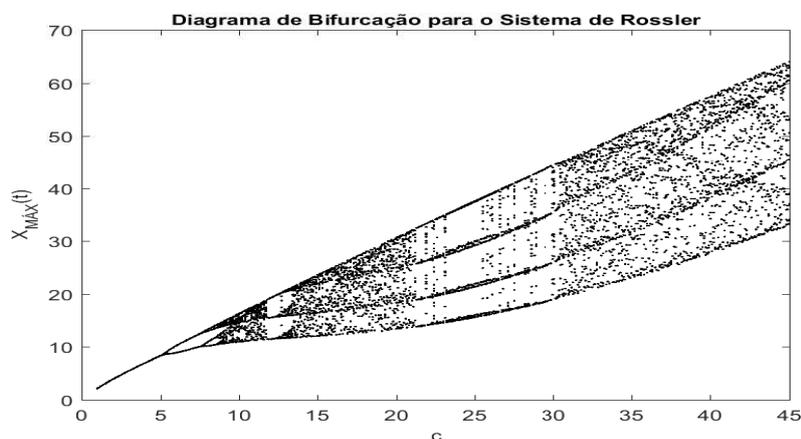


Figura 3: Diagrama de bifurcações para o sistema de Rössler

Deste diagrama, pode-se inferir que para valores de  $1 \leq c \leq 8,6$ , o sistema apresenta comportamento periódico. Após, embora venham a existir algumas janelas periódicas, para a maior parte destes valores o sistema será imprevisível.

### 3.3 MODELO PRESA-PREDADOR

O modelo presa-predador, apresentado no trabalho de [7], é dado por:

$$\begin{aligned} \dot{x} &= x(1-x) - \frac{a_1xy}{1+\alpha\xi+b_1x}, \\ \dot{y} &= \frac{\beta y(x+c\xi)}{1+\alpha\xi+b_1x} - \frac{a_2yz}{1+\alpha\eta+b_2y} - d_1y, \\ \dot{z} &= \frac{z\gamma(y+e\eta)}{1+\alpha\eta+b_2y} - d_2z, \end{aligned} \tag{3}$$

onde,  $x$  representa a população da presa e  $z$  representa a população do predador de ordem mais alta. Já  $y$  representa a população da espécie que é predadora de  $x$  e presa de  $z$ .

O sistema foi analisado com os seguintes valores de parâmetros:  $a_1 = 5, a_2 = 0,1, \xi = 0,1, \eta = 0, b_1 = 2,8, b_2 = 2, \beta = 4,8, \gamma = 0,08, d_1 = 0,4, d_2 = 0,01, c = 0,95, e = 0,85$ , valores definidos por [7]. O parâmetro  $\alpha$ , relacionado a quantidade de comida disponível, foi variado. O diagrama foi feito, tanto para a presa, como para o predador.

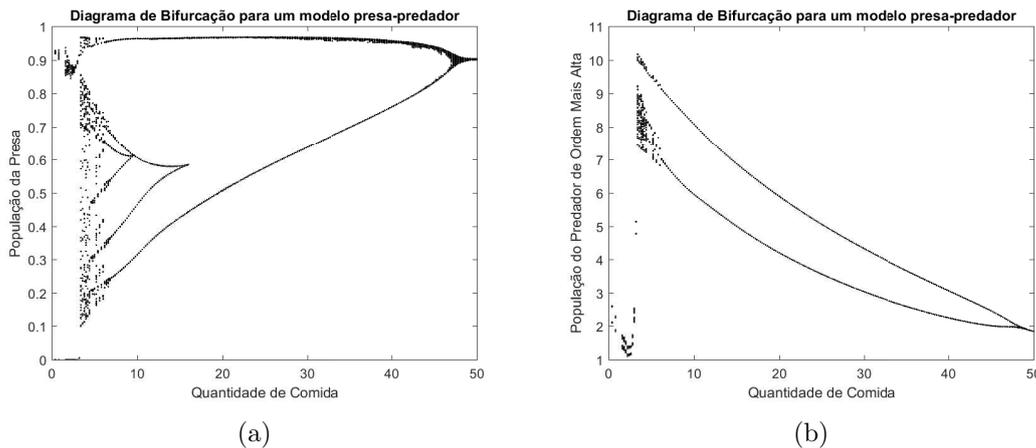


Figura 4: (a) População da presa (b) População do Predador

Este sistema apresenta algumas singularidades em relação aos estudados anteriormente. Por exemplo, para a construção do diagrama, foram usados os pontos de máximo e de mínimo. Para quantidades baixas de comida, tanto a população da presa como a do predador apresentam comportamento caótico. À medida que se aumenta a quantidade de comida, os valores máximos e mínimos das populações se repetem periodicamente, até se manterem estáveis com  $\alpha = 47,5$ . Destaca-se também que para a obtenção do diagrama neste sistema, foi necessário o intervalo de integração ser consideravelmente maior que nos sistemas anteriores.

## 4 CONCLUSÕES

Neste trabalho foi apresentado um programa computacional, com detalhamento do código, para implementação do diagrama de bifurcações de sistemas contínuos que apresentam comportamento caótico. Foram mostradas aplicações para os sistemas de Rössler, Lorenz e presa-predador. Os resultados obtidos foram comparados (visualmente) com resultados apresentados na bibliografia e houve uma similaridade dos mesmos. Como continuação deste estudo, o objetivo é implementar um algoritmo também para o mapa de Poincaré e o Expoente de Lyapunov,

incluindo as variações deste último. Estes programas também serão disponibilizados de forma aberta, para utilização acadêmica.

## 5 AGRADECIMENTOS

O primeiro autor agradece à FAPERGS(Fundação de Amparo à pesquisa do Estado do RS) pelo apoio recebido.

## Referências

- [1] G. Alvarez, S. Li, Cryptographic requirements for chaotic secure communications, *International Journal of Bifurcation and Chaos* 16, 8 (2006) 2129-2151.
- [2] R. Barrio, A.M. Martínez, S. Serrano, D.Wilczak, When chaos meets hyperchaos: 4D Rössler model, *Physics Letters A* 379, 38 (2015) 2300-2305.
- [3] F. M. Danca, M. Romera, G. Pastor, F. Vitini, Finding Attractors of Continuous-Time Systems by Parameter Switching, *Nonlinear Dynamics* 67, 4 (2011) 2317-2342.
- [4] A. Hastings, T. Powell, Chaos in a Three-Species Food Chain, *Ecology* 72, 3 (1991) 896-903.
- [5] E. N. Lorenz, Deterministic non-periodic flow, *Journal of Atmospheric Sciences* 20 (1963) 130-141.
- [6] O. Rössler, An equation for continuous chaos, *Physics Letters A* 57, 5 (1976) 397-398.
- [7] B. Sahoo, S. Poria, Chaos to Order: Role of Additional Food to Predator in a Food Chain Model, *Differential Equations and Dynamical Systems* 23, 2 (2015) 129-146.
- [8] S. H. Strogatz, "Nonlinear Dynamics and Chaos", Westview Press, Cambridge, 2000.
- [9] D. Wilczak, S. Serrano, R. Barrio, Coexistence and Dynamical Connections between Hyperchaos and Chaos in the 4D Rössler System: A Computer-Assisted Proof, *SIAM Journal on Applied Dynamical Systems* 15, 1 (2016) 356-390.

## 6 ANEXOS

Código no MATLAB para construir o diagrama de bifurcações de um sistema

```
function bifurc(sistema, t_inicio, t_fim, t_transiente, parametro
, variavel, y0)

global taxa

for i = 1:length(parametro)
    taxa = parametro(i); %Valor do parâmetro para a integração
    [t,y] = ode45(sistema, [t_inicio t_fim], y0); %Integra o
        sistema
    contador = find(t > t_transiente); %Seleciona os valores após
        o transiente
    y = y(contador, variavel); %Variável = 1: valores de x, 2
        para y e 3 para z
    for j = 2:(length(y)-1) %Analisa os valores da variável
        selecionada
```

```

        %if ((y(j-1,1)+eps) < y(j,1) && y(j,1) > (y(j+1,1)+eps))
            %Se for valor de máximo, marca no gráfico
            %Marcar se for máximo ou mínimo
            if ((y(j-1,1)+eps) < y(j,1) && y(j,1) > (y(j+1,1)+eps)
                || (y(j-1,1)+eps) > y(j,1) && y(j,1) < (y(j+1,1)+eps))
                plot(taxa, y(j,1), 'k.', 'MarkerSize', 1);
                hold on;
            end
        end
    end
end

title('Diagrama de Bifurcação para o Sistema de Chua');
xlabel('a');
ylabel('X_{MÁX}(t)');

```

Script que descreve o Sistema de Lorenz. Passado como o parâmetro *sistema* no script *bifurc*

```

function f = lorenz(t, X)

global taxa %Parâmetro que será variado, no sistema de Lorenz,
           taxa é r.

SIGMA = 10; BETA = 8/3; %Valores dos parâmetros

dx=SIGMA*(X(2)-X(1));           %Equação 1 do sistema
dy=-X(1)*X(3)+taxa*X(1)-X(2); %Equação 2 do sistema
dz=X(1)*X(2)-BETA*X(3);        %Equação 3 do sistema
f = [dx; dy; dz];               %Sistema na forma matricial

```

Formato de chamada do script para a construção do diagrama

```

sistema = @lorenz;           %Sistema a ser integrado
t_inicio = 0;                %Tempo de início da integração
t_fim = 300;                 %Tempo final de integração
t_transiente = 200;         %Duração do tempo transiente
parametro = 25:0.5:325;     %Intervalo de variação do parâmetro
variavel = 3;                %Variável a ser analisada: 1 = x, 2 = y,
                             3 = z;
y0 = [0 1 0];                %Condições iniciais para a integração
%Chamada do script
bifurc(sistema, t_inicio, t_fim, t_transiente, parametro,
       variavel, y0);

```



# Uma Proposta Baseada em Estimativa de Pose (P3P) na Determinação de Pontos de Referência no Perímetro de Voo para Drones

**André L.T. Filho**

Aluno do Bacharelado em Ciência da Computação, UDESC  
Depto de Ciências de Computação, UDESC  
Joinville, SC  
E-mail: altfxd@gmail.com.

**Gilmário B. dos Santos**

Professor do Dep. de Ciências de Computação, UDESC  
Joinville, SC  
E-mail: gilmario.santos@udesc.br

**Resumo:** *Drones trouxeram possibilidades de aplicação em diversas áreas, dentre elas estão o mapeamento aéreo e mais recentemente a utilização de nuvens de drones para efeitos luminosos em eventos artístico-culturais. Uma demanda importante desse tipo de aplicação é a delimitação de uma área no espaço cujas fronteiras limitam o voo do(s) drone(s). O problema é que muitas vezes se conhecem apenas as distâncias entre os pontos de referência que delimitam o perímetro dessa área de voo, sem conhecimento das coordenadas espaciais desses pontos. Esse trabalho discute uma estratégia de estimativa de posição (pose estimation) a qual pode ser adaptada à solução desse problema. Um experimento é realizado em um ambiente geometricamente simulado e os resultados são apresentados.*

**Palavras-chave:** *Estimativa de Pose, Modelo de câmera, Otimização*

## 1 Exposição do Problema

Veículos Aéreos Não Tripulados (VANTs) têm sido empregados ao longo da história, especialmente para uso militar. Há relatos de uso de balões na primeira guerra mundial para fins de bombardeio, bem como as bombas V1 e V2 utilizadas na segunda guerra [1]. No Brasil, os primeiros relatos de VANT ocorreram na década de 80 [2].

O termo “drone” tem origem em idiomas de regiões ao norte/noroeste da Europa, particularmente do inglês, significando um zumbido contínuo ou propriamente designando um zangão. A designação “drone” para descrever um VANT ganhou popularidade a partir da disponibilidade comercial de equipamentos compactos de fácil pilotagem, manobráveis e capazes de embarcar diversos dispositivos, tais como câmeras (RGB), sensores infravermelhos, sensores multiespectrais e outros tipos de sensores (Figura 1). Essa combinação de fatores levaram o drone a usos que se estendem desde o mero entretenimento até a utilização em atividades comerciais e científicas. No Brasil o uso de drones remonta a década de oitenta e vem ganhando terreno como plataforma para várias aplicações, tais como cadastro de propriedades, segurança, monitoramento de obras, agricultura de precisão, entre outras [2].

O emprego de drones muitas vezes demanda a sua localização espacial em tempo real, para tanto, alguns autores tais como Skoda [3] utilizam processamento de imagens e retroprojeção de pontos (2D para 3D) como uma forma de estimar a posição espacial do equipamento durante



Figura 1: Esquerda: drone do tipo multi-rotor “quadricóptero”. Direita: drone do tipo asa fixa. Fontes: commons.wikimedia.org e upload.wikimedia.org/wikipedia/commons

o voo. Por sua vez, Su et al. [4] também utilizam marcadores e sensores de infravermelho, associados a cameras para a estimativa de posição do drone.

Neste trabalho se propõe a utilização do método conhecido por P3P (*Perspective 3 Point*) descrito em [5] e [6] para estimativa de pose baseado em pontos de referência, trata-se de um problema de visão computacional com as seguintes considerações:

- O voo é delimitado pela área na forma de um polígono cujo perímetro é marcado por um conjunto de pontos de referência  $P = \{p_1, p_2, \dots, p_k\}$  e pode ser construído a partir de triângulos (Figura 2);
- A cena da área de voo planejada deve exibir os pontos de referência do polígono, sem oclusões (Figura 2);
- Dados  $p_m$  e  $p_n$  dois pontos consecutivos do perímetro  $m = i, n = i + 1$ , tem que ser previamente conhecida a distância  $d_{mn}$  (Figura 2);
- O drone embarca um sistema GPS.

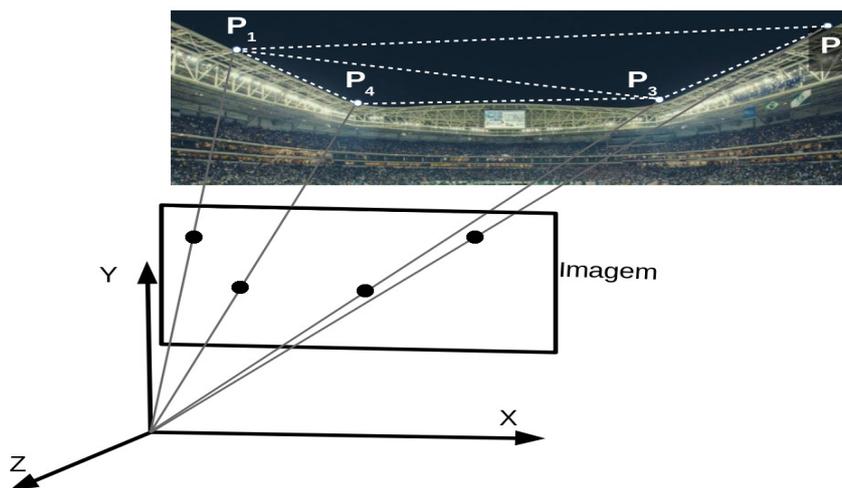


Figura 2: Pontos de referência delimitando um perímetro quadrilátero de voo  $P_1, P_2, P_3, P_4$  sobre o estádio e projeções desses pontos sobre uma imagem. O quadrilátero pode ser composto por dois triângulos de vértices  $P_1, P_2, P_3$  e  $P_1, P_3, P_4$

## 2 Método Proposto

No caso exibido na Figura 2, os dois triângulos  $\Delta_1$  de vértices  $P_1, P_2, P_3$  e  $\Delta_2$  de vértices  $P_1, P_3, P_4$  formam o quadrilátero  $P_1, P_2, P_3, P_4$ , a proposta neste trabalho consiste em repetir o método P3P [6] para cada triângulo, determinando todos os pontos de referência do contorno do quadrilátero.

Inicialmente é importante descrever a geometria da cena. Na Figura 2, os raios projetores lançam-se sobre o plano da imagem em direção ao foco na origem do sistema de coordenadas. A distância entre o centro e o plano de imagem corresponde à distância focal, portanto há versores (vetores de norma unitária) que servem de guia sobre cada raio projetor de maneira que seja possível encontrar os pontos projetores desejados  $P_1, P_2, P_3, P_4$ , ou a melhor estimativa possível dos valores das suas coordenadas. O princípio básico corresponde a  $P_i = a_i * q_i$ , onde  $q_i$  corresponde ao versor direcionado ao respectivo ponto de referência  $P_i$ . Por meio de  $q_i$  alcança-se o ponto projetor utilizando do adequado valor do escalar  $a_i$ :

A formulação fornece três equações, uma para cada ponto  $P_i$ . São desconhecidas as coordenadas de  $P_i$  mas há o conhecimento das distâncias entre pontos consecutivos, sabe-se que para  $p_m$  e  $p_n$  pontos consecutivos ( $m = i, n = i + 1$ ) é conhecida a distância  $d_{mn}$ . Temos então que  $d_{mn}^2 = \|P_m - P_n\|^2 = \|a_m * q_m - a_n * q_n\|^2$  e desenvolvendo esse termo chega-se a Equação 1.

$$d_{mn}^2 = a_m^2 - 2 * a_m * a_n (q_m \cdot q_n) + a_n^2 \quad (1)$$

Dessa forma, da Equação 1 deriva-se o sistema conforme a Equação 2, cujas incógnitas são os escalares  $a_1, a_2$  e  $a_3$ .

$$\begin{aligned} f(a_1, a_2, a_3) &= a_1^2 - 2a_1a_2(q_1 \cdot q_2) + a_2^2 - d_{12}^2 \\ g(a_1, a_2, a_3) &= a_2^2 - 2a_2a_3(q_2 \cdot q_3) + a_3^2 - d_{23}^2 \\ h(a_1, a_2, a_3) &= a_1^2 - 2a_1a_3(q_1 \cdot q_3) + a_3^2 - d_{31}^2 \end{aligned} \quad (2)$$

Utilizando o conceito de linearização de função a várias variáveis, obtém-se a formulação da Equação 3 [5][6], na qual  $J$  corresponde ao Jacobiano das expressões na Equação 2 e  $\delta_i$  corresponde ao ajuste a ser realizado ao valor inicial de  $A = (a_1, a_2, a_3)^T$  em um processo de otimização que conduzirá ao melhor valor dos escalares  $a_1, a_2, a_3$  dentro das restrições impostas:

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} f(a_1, a_2, a_3) \\ g(a_1, a_2, a_3) \\ h(a_1, a_2, a_3) \end{bmatrix} + J \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix} \quad (3)$$

A solução desse sistema consiste em iterar  $A$  conforme a Equação 4, abaixo, determinando as raízes das equações envolvidas nas variáveis  $a_1, a_2, a_3$ . De posse dessas variáveis calculam-se os vértices de  $\Delta_1$  utilizando  $P_i = a_i * q_i$  com  $i \in 1, 2, 3$ , ficando determinadas as coordenadas de  $P_1, P_2, P_3$ , restando desconhecida as coordenadas do ponto  $P_4$ . Para a sua determinação, basta aplicar a mesma solução utilizando o triângulo  $\Delta_2$  e o perímetro do contorno do quadrilátero está identificado por meio das coordenadas dos vértices  $P_1, P_2, P_3, P_4$ .

$$A^{k+1} = A^k - J^{-1}(A^k)f(A^k) \quad (4)$$

### 2.1 Experimento

A metodologia experimental adotada consiste em testar o potencial da proposta em um ambiente simulado. Para tanto um sistema foi modelado simulando a geometria de uma câmera e seu referencial de coordenadas, conforme ilustrado na Figura 3. Como não foi adotada um sistema de medidas específico, utiliza-se o termo *u.m.* como unidade métrica.

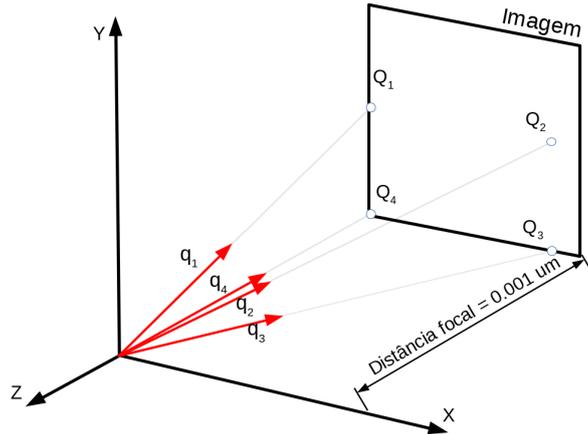


Figura 3: Projeções  $Q_1, Q_2, Q_3$  e  $Q_4$  dos pontos de referência e versores nas direções dos pontos  $P_1, P_2, P_3$  e  $P_4$ , são desconhecidas as coordenadas desses pontos.

Ainda na mesma figura, a câmera é modelada de maneira que o seu foco ocorra na origem do sistema de coordenadas  $(X, Y, Z)$ . O plano de imagem desse sistema ocorre a uma distância focal igual a  $10^{-3}u.m.$  (a qual pode ser determinada via método de calibração [6]). Esse sistema captura uma cena na imagem que exhibe as projeções  $Q_1, Q_2, Q_3$  e  $Q_4$  de quatro pontos de referência  $P_1, P_2, P_3$  e  $P_4$ .

Apesar do desconhecimento das coordenadas dos pontos  $P_1, P_2, P_3$  e  $P_4$ , são conhecidas as distâncias entre pares consecutivos desses pontos. Ou seja, são conhecidos ou determináveis os valores das distâncias entre  $P_m$  e  $P_n$  consecutivos, identificáveis e contidos no perímetro do quadrilátero delimitador da área de voo do drone.

Manualmente ou por meio de processamento de imagens (aplicando extrator de feições) são determinadas as coordenadas dos pontos  $Q_1, Q_2, Q_3, Q_4$  na imagem. Com essa informação e conhecendo as dimensões de pixel do sensor que captura a cena, são determinadas as coordenadas dos respectivos pontos  $Q_i$  [6].

As projeções  $Q_i$  com a distância focal  $10^{-3}u.m.$ , encontram-se na Equação 5, onde  $Q_1, Q_2, Q_3$  e  $Q_4$  estão nas colunas da matriz:

$$Q = \begin{bmatrix} 0.0 & 3 * 10^{-4} & 3 * 10^{-4} & 0.0 \\ 2 * 10^{-4} & 2 * 10^{-4} & 0.0 & 0.0 \\ -10^{-3} & -10^{-3} & -10^{-3} & -10^{-3} \end{bmatrix} \quad (5)$$

A partir dessas coordenadas são calculados os versores  $q_1, q_2, q_3$  e  $q_4$  (Figura 3) associados aos pontos  $Q_1, Q_2, Q_3, Q_4$ , onde  $q_i = \frac{Q_i}{\|Q_i\|}$ . São calculados dois conjuntos de versores considerando cada triângulo,  $\Delta_1$  e  $\Delta_2$ , conforme exibido na Figura 4.

Dos pontos  $P_1, P_2, P_3, P_4$  conhece-se apenas as distâncias interpontos:  $d_{12} = 30,0 u.m.$ ;  $d_{2,3} = 30,0 u.m.$ ;  $d_{34} = 36,05555 u.m.$ ;  $d_{41} = 20,0 u.m.$

De posse de  $q_i$  calcula-se a matriz do Jacobiano por meio da Equação 6. As distâncias interpontos são utilizadas para o cálculo das funções  $f(A^k)$  na Equação 4.

$$J = \begin{bmatrix} 2a_1 - 2(q_1 \cdot q_2)a_2 & 2a_2 - 2(q_1 \cdot q_2)a_1 & 0.0 \\ 0.0 & 2a_2 - 2(q_2 \cdot q_3)a_3 & 2a_3 - 2(q_2 \cdot q_3)a_2 \\ 2a_1 - 2(q_4 \cdot q_1)a_4 & 0.0 & 2a_3 - 2(q_3 \cdot q_1)a_1 \end{bmatrix} \quad (6)$$

Ao resolver o sistema exibido na Equação 4 para os pontos de vértices dos triângulos  $\Delta_1$  e depois para  $\Delta_2$ , obtém-se o conjunto  $\hat{P}$  exibido na Equação 7, na qual os pontos  $\hat{P}_1, \hat{P}_2, \hat{P}_3, \hat{P}_4$  estão em colunas. O sistema foi testado em um notebook Vaio, core I7, rodando no SO Linux distribuição Ubuntu 14, em uma máquina virtual Oracle em 8GB de memória RAM. As implementações em Matlab.

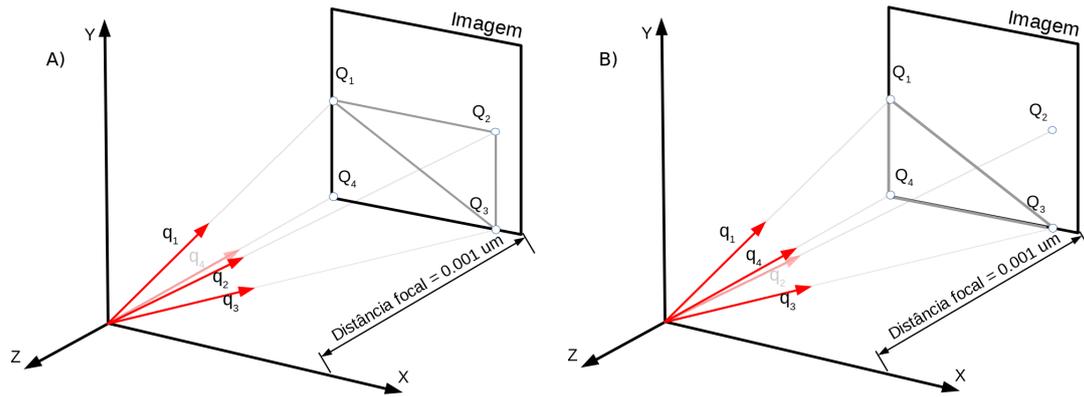


Figura 4: São calculados dois conjuntos de versores  $q_1, q_2, q_3$  e  $q_1, q_3, q_4$  em relação a cada triângulo.

$$\hat{P} = \begin{bmatrix} 0.0 & 30.0 & 30.0 & 0.0 \\ 20.0 & 20.0 & 0.0 & 0.0 \\ -100.0 & -100.0 & -100.0 & -100.0 \end{bmatrix} \quad (7)$$

## 2.2 Verificando Pontos Obtidos

Para verificar a acurácia dos resultados ( $\hat{P}$ ) basta utilizar a matriz de projeção da câmera e projetar os pontos obtidos em  $\hat{P}$  (exibidos na Equação 7). A matriz de projeção pode ser obtida via calibração de câmera [6], porém, como aqui se utiliza um ambiente virtual de experimento, a matriz de projeção pode ser deduzida da geometria da cena. Conforme a Figura 5, as relações entre triângulos apresentados na referida geometria permitem deduzir a matriz de projeção  $M$  exibida na Equação 8, onde  $DF$  corresponde à distância focal.

Ao aplicar a  $\hat{Q}^h = M * \hat{P}$  obtém-se as projeções  $\hat{Q}^h$  no espaço homogêneo, os quais são convertidos para o espaço cartesiano por meio da expressão  $\hat{Q} = \frac{\hat{Q}_i^h}{w_i}$ , onde  $w_i$  é a terceira coordenada de  $\hat{Q}_i^h$ .

Quanto menor a distância euclidiana entre  $\hat{Q}$  e o conjunto de pontos  $Q$  (Equação 5), menor será o erro dos resultados obtidos em  $\hat{P}$  em relação aos valores esperados para  $P = P_1, P_2, P_3, P_4$ . No presente experimento o erro foi de 0.007.

É importante mencionar que os resultados obtidos são referenciados no sistema de câmera, podendo ser mapeados para um sistema de coordenadas GPS sem grandes dificuldades utilizando a técnica de Gram-Schmidt, por exemplo.

$$M = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & \frac{1}{DF} \end{bmatrix} \quad (8)$$

## 3 Conclusão

Este trabalho relata experimento de visão computacional no campo da estimação de pose o qual pode ser útil a aplicações que utilizem drones e que demandem controle automático de navegação. Para trabalhos futuros nesse campo, pretende-se utilizar um drone para testes em condições reais.

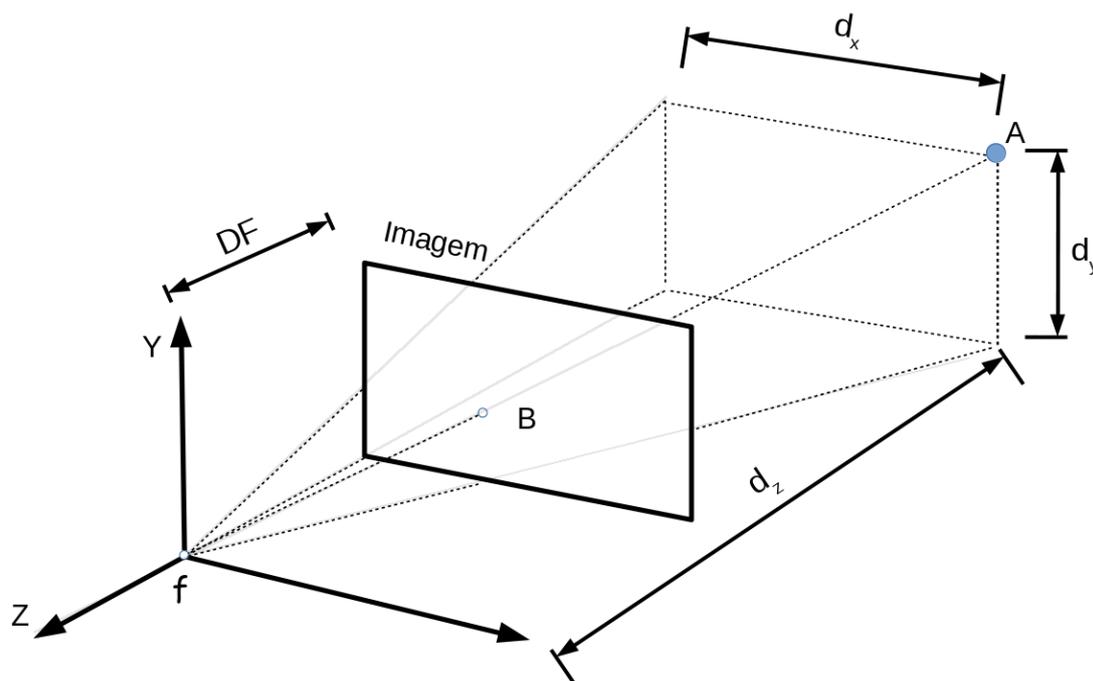


Figura 5: Geometria da cena.

## Referências

- [1] E. Ubiratan, A origem dos vant, *Aero Magazine*, n.248, jan/2015. Documento acessado em 15/04/2019 e disponível no URL <https://aeromagazine.uol.com.br/revista/248>
- [2] F. A. Medeiros et al., Utilização de um veículo aéreo não-tripulado em atividades de imageamento georeferenciado, *Ciência Rural*, v.38, n.8, p.2375-2378, nov/2008.
- [3] J. Skoda, R. Bartak, Camera-Based Localization and Stabilization of a Flying Drone, em "Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference", pp. 372-377, 2015. Documento acessado em 15/04/2019 e disponível no URL: <https://www.aaai.org/ocs/index.php/FLAIRS/FLAIRS15/paper/view/10398>
- [4] W. Su, A. Ravankar, A. A. Ravankar, Y. Kobayashi e T. Emaru, UAV Pose Estimation Using IR and RGB Cameras, em "Proceedings of the 2017 IEEE/SICE International Symposium on System Integration", pp. 151-156, Taipei, Taiwan, December/2017. Documento acessado em 15/04/2019 e disponível no URL: <https://ieeexplore.ieee.org/document/8279204>
- [5] M. Fischler e R. Boilles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, *Communications of ACM*, v. 24, n. 6,1981.
- [6] L. Shapiro e S. Stockman, "Computer Vision", Pearson, 2001.
- [7] D.G Lowe, Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, 60(2), 91-110, 2004.

# Uma Extensão do Princípio de Invariância Para a Classe de Sistemas Positivos com Parâmetros Incertos

Michele C. Valentino

Depto Acadêmico de Matemática, UTFPR,  
86300-000, Cornélio Procópio, PR  
E-mail: valentino.utfpr.edu.br,

**Luís F. C. Alberto,**      **Yuri C. S. Ribeiro,**

EESC-USP Departamento de Engenharia Elétrica  
13566-590, São Carlos, SP

E-mail: lfcaberto@usp.br,    yuricandido@gmail.com.

**Resumo:** Neste trabalho, consideramos a classe dos sistemas positivos com parâmetros incertos e apresentamos uma extensão do princípio de invariância, o qual é capaz de estimar um atrator para a solução que será invariante com respeito a não somente ao chaveamento, mas também com relação as incertezas dos parâmetros. Nesta extensão, o conjunto onde a derivada da função de Lyapunov ao longo da solução do sistema é positiva pode ser não vazio e por esta razão conseguimos aplicar o resultado na análise de uma classe maior de sistemas, principalmente aqueles que ainda não conhecemos, ou não existe, sua função de Lyapunov. Um sistema do tipo presa predador é considerado para mostrarmos a eficiência do resultado.

**Palavras-chave:** *Sistemas Positivos, Parâmetros Incertos, Atrator, Princípio de Invariância*

## 1 Introdução

Muitos problemas práticos são descritos por modelos, os quais têm a característica de manter suas trajetórias não negativas sempre que a condição inicial é não negativa. Modelos com tal característica, como por exemplo os modelo biológicos, são denominados **sistemas positivos** e têm sido muito desenvolvido, principalmente a parte da teoria que estuda a estabilidade de sua solução, nesta última década.

Uma revisão teórica sobre sistemas positivos e resultados de estabilidade para esta classe de sistema pode ser encontrada em [1, 7, 6]. Resultados que utilizam a teoria de Lyapunov para analisar a solução desses modelos [1, 5, 3], exploram a propriedade de que a derivada da função de Lyapunov somente deve ser negativa no conjunto  $\mathbb{R}_+^n = \{x \in \mathbb{R}^n : x_i \geq 0, 1 \leq i \leq n\}$ . No entanto, ainda é possível obter condições menos conservadoras que as apresentadas em [5], considerando a extensão do princípio de invariância de LaSalle. A extensão do princípio de invariância foi primeiramente obtida para equações diferenciais ordinárias [9] e depois foi desenvolvida para sistemas discretos [2], sistemas periódicos [4] e sistemas chaveados [11]. A principal vantagem dessas extensões é que a derivada da função auxiliar pode assumir valores positivos em conjuntos limitados, relaxando a suposição sobre a função de Lyapunov. Mais ainda, explorando essas extensões podemos obter estimativas do atrator e da área de atração, que são uniformes com respeito a lei de chaveamento, incluindo aqueles sistemas que não admitem função de Lyapunov.

Neste artigo, propomos uma extensão do princípio de invariância para uma classe geral de sistemas não lineares positivos. A principal característica desta extensão é que a função escalar

auxiliar não precisa ser positiva definida e o conjunto onde sua derivada é positiva deve ser limitado somente em  $\mathbb{R}_+^n$ .

No decorrer do texto consideramos  $int(\mathbb{R}_+^n) = \{x \in \mathbb{R}^n : x_i > 0, 1 \leq i \leq n\}$  e  $Bd(\mathbb{R}_+^n) = \mathbb{R}_+^n - int(\mathbb{R}_+^n)$ .

## 2 Preliminares

Considere o sistema dinâmico autônomo não linear:

$$\dot{x}(t) = f(x(t)) \tag{1}$$

onde  $x \in \mathbb{R}^n$  e  $f$  é uma função de classe  $C^1$ . A solução de (1) com condição inicial  $x(t_0) = x_0$  será denotada por  $\varphi(t, x_0)$ .

**Definição 1.** *Seja  $\varphi(t, x_0)$  a solução de (1) com condição inicial  $x(t_0) = x_0$  para todo  $t \geq 0$ . O conjunto  $\omega$ -limite de  $x_0$  é o conjunto:*

$$\omega^+(x_0) = \{p \in \mathbb{R}^n : \exists \{t_n\} \text{ com } t_n \rightarrow \infty \text{ quando } n \rightarrow \infty \text{ talque } \varphi(t_n, x_0) \rightarrow p\}.$$

Em [8] foi demonstrado o seguinte resultado a respeito das propriedades dos conjuntos limite. Se a solução  $\varphi(t, x_0)$  de (1) é limitada para todo  $t \geq 0$ , então  $\omega^+(x_0)$  é um conjunto não-vazio, compacto, conexo e invariante. Mais ainda,  $\varphi(t, x_0)$  é atraída para  $\omega^+(x_0)$ , isto é

$$\lim_{t \rightarrow \infty} dist(\varphi(t, x_0), \omega^+(x_0)) = 0.$$

Considere as funções de classe  $C^1$ ,  $V, c : \mathbb{R}^n \rightarrow \mathbb{R}$  tais que  $-\nabla V(x) \cdot f(x) \geq c(x)$ , para todo  $x \in \mathbb{R}^n$  e defina o conjunto  $A = \{x \in \mathbb{R}^n : c(x) < 0\}$ . Pelos resultados provados em [9], podemos dizer que, se existe um número real  $l := \sup_{x \in A} V(x)$  tal que  $\bar{\Omega}_l = \{x \in \mathbb{R}^n : V(x) \leq l\}$  é limitado, então toda solução limitada  $\varphi(t, x_0)$  de (1) é atraída para  $B$ , em que  $B$  é o maior conjunto invariante em  $\{x \in \mathbb{R}^n : \nabla V(x) \cdot f(x) = 0\} \cup \bar{\Omega}_l$ . Este resultado é a versão global da extensão do princípio de invariância para a classe de sistemas autônomos não lineares. Na versão local deste resultado, é suposto a existência de um número real  $L > 0$  tal que  $\Omega_L = \{x \in \mathbb{R}^n : V(x) < L\}$  é limitado. Então, se existe um número real  $\tilde{l}$  tal que  $\sup_{x \in \tilde{A}} V(x) = \tilde{l} < L$ , em que  $\tilde{A} = \{x \in \Omega_L : c(x) < 0\}$ , temos que toda solução iniciando em  $\Omega_L$  é atraída para o maior conjunto invariante em  $\{x \in \Omega_L : c(x) = 0\} \cup \{x \in \mathbb{R}^n : V(x) \leq \tilde{l}\}$ . Versões uniformes em relação aos parâmetros dos sistemas foram apresentadas em [10]. Na próxima seção será apresentada uma extensão do princípio de invariância para sistemas não lineares positivos com parâmetros incertos.

## 3 Uma extensão do princípio de invariância

Nesta seção, consideramos a seguinte classe de sistemas:

$$\dot{x} = f(x, \lambda), \tag{2}$$

em que  $\lambda$  é o vetor de parâmetros do sistema que pertence ao conjunto compacto  $\Lambda \subset \mathbb{R}^m$ ,  $x \in \mathbb{R}^n$ . O sistemas (2) é dito ser  $\Lambda$ -positivamente invariante se, para todo  $x_0 \in \mathbb{R}_+^n$  e para todo  $\lambda \in \Lambda$ , a solução  $\varphi(t, x_0, \lambda) \in \mathbb{R}_+^n$ ,  $\forall t \geq 0$ . Uma condição suficiente e necessária para isso é:

$$\forall x \in Bd(\mathbb{R}_+^n) : x_i = 0 \implies f_i(x, \lambda) \geq 0 \text{ para todo } \lambda \in \Lambda.$$

Para um  $\lambda \in \Lambda$  fixado, podemos definir o conjunto  $\omega_\lambda$ -limite como segue

$$\omega_\lambda^+(x_0) = \{p \in \mathbb{R}^n : \exists \{t_n\} \text{ com } t_n \rightarrow \infty \text{ quando } n \rightarrow \infty \text{ tal que } \varphi(t_n, x_0, \lambda) \rightarrow p\}.$$

Com o objetivo de apresentar uma extensão do princípio de invariância para os sistemas  $\Lambda$ -invariante positivos, consideramos uma função de classe  $C^1$ ,  $V : \mathbb{R}^n \times \Lambda \rightarrow \mathbb{R}$ , funções contínuas  $a, b, c : \mathbb{R}^n \rightarrow \mathbb{R}$  satisfazendo

$$\begin{aligned} a(x) \leq V(x, \lambda) \leq b(x), \quad \forall (x, \lambda) \in \mathbb{R}^n \times \Lambda & \quad (3) \\ -\nabla V(x, \lambda).f(x, \lambda) \geq c(x), \quad \forall (x, \lambda) \in \mathbb{R}^n \times \Lambda, & \quad (4) \end{aligned}$$

e os conjuntos

$$\begin{aligned} A_L &= \{x \in \mathbb{R}_+^n : a(x) \leq L\}, & B_L &= \{x \in \mathbb{R}_+^n : b(x) \leq L\}, \\ C_L &= \{x \in A_L : c(x) < 0\}, & E_L &= \{x \in A_L : c(x) = 0\}. \end{aligned} \quad (5)$$

**Teorema 1.** *Considere o sistema  $\Lambda$ -invariante positivo (2). Vamos supor a existência de um número real  $L$  tal que  $A_L$  é não vazio e limitado. Mais ainda, supomos que existe  $\ell \in \mathbb{R}$  tal que  $\sup_{x \in C_L} b(x) \leq \ell < L$ . Então, para qualquer  $\lambda \in \Lambda$  fixado:*

(i) *Se  $x_0 \in B_\ell$ , então  $\varphi(t, x_0, \lambda)$  permanece em  $A_\ell$  para todo  $t \geq 0$ . Portanto,  $\varphi(t, x_0, \lambda)$  é atraído para o maior conjunto invariante em  $A_\ell$  quando  $t \rightarrow +\infty$ .*

(ii) *Se  $x_0 \in \text{int}(B_L - B_\ell)$ , então  $\varphi(t, x_0, \lambda)$  é atraído para o maior conjunto invariante contido em  $A_\ell \cup E_L$ .*

*Demonstração.* Para um  $\lambda \in \Lambda$  fixado e  $x_0 \in \mathbb{R}_+^n$ , seja  $[0, t_+)$  o intervalo maximal de existência da solução  $\varphi(t, x_0, \lambda)$  do sistema  $\Lambda$ -invariante positivo (2). Então,  $\varphi(t, x_0, \lambda) \in \mathbb{R}_+^n, \forall t \in [0, t_+)$ .

(i) Primeiro, para um  $\lambda \in \Lambda$  fixado, consideramos que  $x_0 \in B_\ell \subset \mathbb{R}_+^n$  e supomos que existe  $\bar{t} \in [0, t_+)$  tal que  $\varphi(\bar{t}, x_0, \lambda) \notin A_\ell$ . Uma vez que  $V(\varphi(0, x_0, \lambda)) \leq b(\varphi(0, x_0, \lambda)) \leq \ell$  e  $V(\varphi(\bar{t}, x_0, \lambda)) \geq a(\varphi(\bar{t}, x_0, \lambda)) > \ell$ , devido a positividade da trajetória da solução e a continuidade de  $V$ , temos a existência de  $t^*$  tal que  $V(\varphi(t^*, x_0, \lambda)) = \ell$  e  $b(\varphi(t, x_0, \lambda)) \geq V(\varphi(t, x_0, \lambda)) > \ell$  para todo  $t \in (t^*, \bar{t})$ , mas isso é uma contradição, pois  $-\nabla V(\varphi(t, x_0, \lambda), \lambda).f(x, \lambda) \geq c(\varphi(t, x_0, \lambda)) \geq 0$  para todo  $t$  tal que  $\varphi(t, x_0, \lambda) \notin B_\ell \subseteq A_\ell$ . Assim,  $\varphi(t, x_0, \lambda) \in A_\ell$  para todo  $t \in [0, t_+)$  e  $t_+ = +\infty$ . Portanto,  $\varphi(t, x_0, \lambda)$  é atraída para o maior conjunto invariante em  $A_\ell$ .

(ii) Agora, considere  $x_0 \in \text{int}(B_L - B_\ell)$  e um  $\lambda \in \Lambda$  fixado. Se existe  $\bar{t} \in (0, t_+)$  tal que  $\varphi(\bar{t}, x_0, \lambda) \in B_\ell$ , então o resultado seguirá da primeira parte desta demonstração. Caso contrário, uma vez que  $C_L \subseteq B_\ell$ , temos  $a(\varphi(t, x_0, \lambda)) \leq V(\varphi(t, x_0, \lambda), \lambda) \leq V(\varphi(0, x_0, \lambda), \lambda) \leq b(\varphi(0, x_0, \lambda)) = b(x_0) < L$  para todo  $t \in [0, t_+)$  e então,  $\varphi(t, x_0, \lambda) \in B_L \subseteq A_L$ . Assim, concluímos que  $t_+ = +\infty$  e que existe  $r \in \mathbb{R}$  tal que  $V(\varphi(t, x_0, \lambda), \lambda) \rightarrow r$  quando  $t \rightarrow +\infty$ . Seja  $c \in \omega_\lambda^+(x_0)$ , então existe uma sequência  $t_n \rightarrow \infty$  quando  $n \rightarrow +\infty$  tal que  $\varphi(t_n, x_0, \lambda) \rightarrow c$  e, pela continuidade de  $V$ , temos que  $V(\varphi(t_n, x_0, \lambda), \lambda) \rightarrow V(c)$ . Assim, pela unicidade do limite  $V(c) = r$ , isto é,  $V(x) = r$  para todo  $x \in \omega_\lambda^+(x_0)$  e então  $\nabla V(x, \lambda).f(x, \lambda) = 0$  em  $\omega_\lambda^+(x_0)$ . Portanto, a solução é atraída para o maior conjunto invariante em  $E_L$ .  $\square$

O próximo resultado apresenta uma versão global do Teorema 1.

**Teorema 2. (Versão Global)** *Considere o sistema  $\Lambda$ -invariante positivo (2). Se para um  $\lambda \in \Lambda$  fixado, existir  $\ell \in \mathbb{R}$  tal que  $\sup_{x \in C} b(x) \leq \ell$  em que  $C = \{x \in \mathbb{R}_+^n : c(x) < 0\}$ , então:*

(i) *Se  $x_0 \in B_\ell$  então, toda solução limitada  $\varphi(t, x_0, \lambda)$  permanece em  $A_\ell$  para todo  $t \geq 0$ . Portanto,  $\varphi(t, x_0, \lambda)$  é atraída para o maior conjunto invariante em  $A_\ell$  quando  $t \rightarrow +\infty$ .*

(ii) *Se  $\varphi(t, x_0, \lambda)$  é uma solução limitada para todo  $t \geq 0$ , então é atraída para o maior conjunto invariante em  $A_\ell \cup E$ , em que  $E = \{x \in \mathbb{R}_+^n : c(x) = 0\}$ .*

*Demonstração.* Seja  $[0, +\infty)$  o intervalo maximal da solução limitada do sistema  $\Lambda$ -invariante positivo (2). Então, para todo  $\lambda \in \Lambda$ ,  $\varphi(t, x_0, \lambda) \in \mathbb{R}_+^n, \forall t \in [0, t_+)$  sempre que  $x_0 \in \mathbb{R}_+^n$ .

(i) A demonstração é análoga a demonstração do item (i) do Teorema 1.

(ii) Sabemos que  $C \subset B_\ell \subset A_\ell$ , então para todo  $(x, \lambda) \in (\mathbb{R}_+^n - B_\ell) \times \Lambda$  temos que  $\nabla V(x, \lambda).f(x, \lambda) \leq 0$ . Se existe  $\bar{t} \in \mathbb{R}$  tal que  $\varphi(\bar{t}, x_0, \lambda) \in B_\ell$ , então a demonstração segue de (i). Agora, supomos

que  $\varphi(t, x_0, \lambda)$  é uma solução limitada para todo  $t \geq 0$  e  $\varphi(t, x_0, \lambda) \notin B_\ell$  para todo  $t \geq 0$ . Assim,  $\nabla V(\varphi(t, x_0, \lambda), \lambda) \cdot f(\varphi(t, x_0, \lambda), \lambda) \leq 0$  e portanto  $t \mapsto V(\varphi(t, x_0, \lambda), \lambda)$  é uma função não crescente. Uma vez que  $V(\varphi(t, x_0, \lambda), \lambda)$  é limitada, existe  $r \in \mathbb{R}$  tal que  $V(\varphi(t, x_0, \lambda), \lambda) \rightarrow r$  quando  $t \rightarrow +\infty$ . Seja  $c \in \omega_\lambda^+(x_0)$  então, existe uma sequência  $t_n \rightarrow \infty$  quando  $n \rightarrow +\infty$  tal que  $\varphi(t_n, x_0, \lambda) \rightarrow c$  e pela continuidade de  $V$ , temos que  $V(\varphi(t_n, x_0, \lambda), \lambda) \rightarrow V(c)$ . Assim, pela unicidade do limite  $V(c, \lambda) = r$ , temos  $V(x, \lambda) = r$  para todo  $x \in \omega_\lambda^+(x_0)$  e  $c(x) \leq -\nabla V(x, \lambda) \cdot f(x, \lambda) = 0$  em  $\omega_\lambda^+(x_0)$ , isto é,  $c(x) = 0$  para todo  $x \in \omega_\lambda^+(x_0)$ . Uma vez que  $\omega_\lambda^+(x_0)$  é um conjunto invariante com respeito a (2), então a solução é atraída para o maior conjunto invariante em  $E_\ell$ .  $\square$

**Corolário 1.** *Se consideramos que  $a : \mathbb{R}^n \rightarrow \mathbb{R}$  satisfaz  $a(x) \rightarrow \infty$  quando  $\|x\| \rightarrow \infty$  no Teorema 2, então toda solução será atraída para o maior conjunto invariante em  $A_\ell \cup E$ .*

*Demonstração.* Por (i) da demonstração do Teorema 2, temos  $\varphi(t, x_0, \lambda) \in A_\ell$  para todo  $t \in [0, t_+)$  sempre que  $x_0 \in B_\ell$ , em que  $[0, t_+)$  é o intervalo maximal de existência da solução. Uma vez que  $a$  é radialmente ilimitada, todo conjunto de nível de  $a$  é um conjunto limitado. Assim, a solução é limitada e  $\varphi(t, x_0, \lambda) \in A_\ell$  para todo  $t \geq 0$ . Agora, considere  $x_0 \in \mathbb{R}_+^n$  tal que  $x_0 \notin B_\ell$ ,  $\varphi(t, x_0, \lambda) \notin B_\ell$  para todo  $t \in [0, t_+)$  e escolha  $L \geq b(\varphi(0, x_0, \lambda))$ . Vamos provar que a solução  $\varphi(t, x_0, \lambda)$  é limitada para todo  $t \geq 0$ . Na verdade, temos que  $V(\varphi(0, x_0, \lambda)) \in B_L$  e assim analogamente a demonstração (i) do Teorema 2 podemos dizer que  $\varphi(t, x_0, \lambda) \in A_L$  para todo  $t \in [0, t_+)$  (pois  $C \subset B_\ell \subseteq B_L$ ). Uma vez que  $A_L$  é um conjunto limitado, então a solução é limitada e  $\varphi(t, x_0, \lambda) \in A_L$  para todo  $t \geq 0$ . Portanto, o que falta da demonstração segue da segunda parte da demonstração do Teorema 2.  $\square$

**Observação 1.** *Se  $E \subseteq A_\ell$ , então toda solução  $\varphi(t, x_0)$  é atraída para o maior conjunto invariante em  $A_\ell$ . Neste caso,  $A_\ell$  é uma estimativa do atrator do sistema  $\Lambda$ -invariante positivo (2).*

**Corolário 2.** *Seja  $A_\ell = \bigcup_{i=1}^{m_1} A_{\ell_i}$ ,  $B_\ell = \bigcup_{i=1}^{m_2} B_{\ell_i}$ , e considere o conjunto  $D = \{x \in \mathbb{R}^n : \nabla V(x, \lambda) \cdot f(x, \lambda) > 0, \forall \lambda \in \Lambda\}$  tal que  $D = \bigcup_{i=1}^{m_3} D_i$ . Se existe  $j_1 \in \{1, \dots, m_1\}$ ,  $j_2 \in \{1, \dots, m_2\}$ ,  $j_3 \in \{1, \dots, m_3\}$  e  $\ell \in \mathbb{R}$  satisfazendo  $D_{j_3} \subseteq B_{\ell_{j_2}} \subseteq A_{\ell_{j_1}}$ , tal que  $A_{\ell_{j_1}}$  é limitado e  $A_{\ell_{j_1}} \cap (\bigcup_{i \neq j_3}^{m_3} D_i) = \emptyset$ , então a solução  $\varphi(t, x_0)$  com  $x_0 \in \text{int}(B_{\ell_{j_2}})$  é atraída para o maior conjunto invariante em  $A_{\ell_{j_1}}$ .*

*Demonstração.* Supomos a existência de  $j_1 \in \{1, \dots, m_1\}$ ,  $j_2 \in \{1, \dots, m_2\}$ ,  $j_3 \in \{1, \dots, m_3\}$  e  $\ell \in \mathbb{R}$  satisfazendo  $D_{j_3} \subseteq A_{\ell_{j_1}} \subseteq B_{\ell_{j_2}}$ . Seja a solução  $\varphi(t, x_0)$  com  $x_0 \in \text{int}(B_{\ell_{j_2}})$ , então a demonstração deste corolário é análoga a primeira parte da demonstração do Teorema 1.  $\square$

No próximo exemplo numérico, analisaremos o comportamento assintótico de (6) admitindo a existência de uma incerteza de  $\pm 3\%$  no parâmetro  $\alpha$ . Note que, embora consideramos este pequeno erro, os pontos de equilíbrio do sistema (6) mudam de acordo com o parâmetro  $\lambda \in \Lambda$  fixado. Mesmo assim, mostraremos que a estimativa do atrator não dependerá do valor de  $\alpha$ .

**Exemplo 1.** *Considere o seguinte sistema:*

$$\begin{cases} \dot{x} = \alpha\beta x + (\beta - \alpha)x^2 - \gamma xy - x^3, \\ \dot{y} = \zeta y + xy. \end{cases} \quad (6)$$

com  $\Lambda = \{\lambda = (\alpha, \beta, \gamma, \zeta) \in \mathbb{R}^4 : \beta = 3, \gamma = 0.5, \zeta = 2.1 \text{ e } 0.97 \leq \alpha \leq 1.03\}$ . Escolhendo a função  $V$  da seguinte forma:

$$\begin{aligned} V(x, y, \lambda) &= 1.35(18.93(x - 2.1)^4 - 36.34(x - 2.1)^3(y - d) - 31.73(x - 2.1)^3 \\ &+ 113.7(x - 2.1)^2(y - d)^2 + 363.4(x - 2.1)^2(y - d) + 270.2(x - 2.1)^2 \\ &+ 63.09(x - 2.1)(y - d)^2 + 102.7(x - 2.1)(y - d) - 0.9347(x - 2.1) \\ &+ 63.89(y - d)^3 + 138.5(y - d)^2 + 18.43(y - d)), \end{aligned} \quad (7)$$

em que  $d = 1.8(2.1 - \alpha) \in [1.926, 2.034]$ , temos que o conjunto  $D$  é compacto em  $\mathbb{R}_+^2$  (veja a região pontilhada na Figura 1. Assim, escolhendo  $\ell = 28$  e

$$a(x, y) = 0.86(18.93(x - 2.16)^4 - 36.34(x - 2.16)^3(y - 2.01) - 31.73(x - 2.16)^3 + 113.7(x - 2.16)^2(y - 2.01)^2 + 363.4(x - 2.16)^2(y - 2.01) + 270.2(x - 2.16)^2 + 63.09(x - 2.16)(y - 2.01)^2 + 102.7(x - 2.16)(y - 2.01) - 0.9347(x - 2.16) + 63.89(y - 2.01)^3 + 138.5(y - 2.01)^2 + 18.43(y - 2.01))$$

e

$$b(x, y) = 3.6(18.93(x - 2.19)^4 - 36.34(x - 2.19)^3(y - 1.98) - 31.73(x - 2.19)^3 + 113.7(x - 2.19)^2 * (y - 1.98)^2 + 363.4(x - 2.19)^2(y - 1.98) + 270.2(x - 2.19)^2 + 63.09(x - 2.19)(y - 1.98)^2 + 102.7(x - 2.19)(y - 1.98) - 0.9347(x - 2.19) + 63.89(y - 1.98)^3 + 138.5(y - 1.98)^2 + 18.43(y - 1.98)),$$

os quais asseguram a inclusão da componente conexa  $D \subseteq B_{28} \subseteq A_{28}$ , temos que, pelo Corolário 2, a solução  $\varphi(t, x_0)$  com  $x_0 \in B_{28}$  é atraída para um conjunto invariante em  $A_{28}$ , isto é, a solução é atraída para um ponto de equilíbrio contido no conjunto de pontos  $\{(2.1, 1.8(2.1 - \alpha)), 0.97 \leq \alpha \leq 1.03\}$ , como podemos ver na Figura 1.

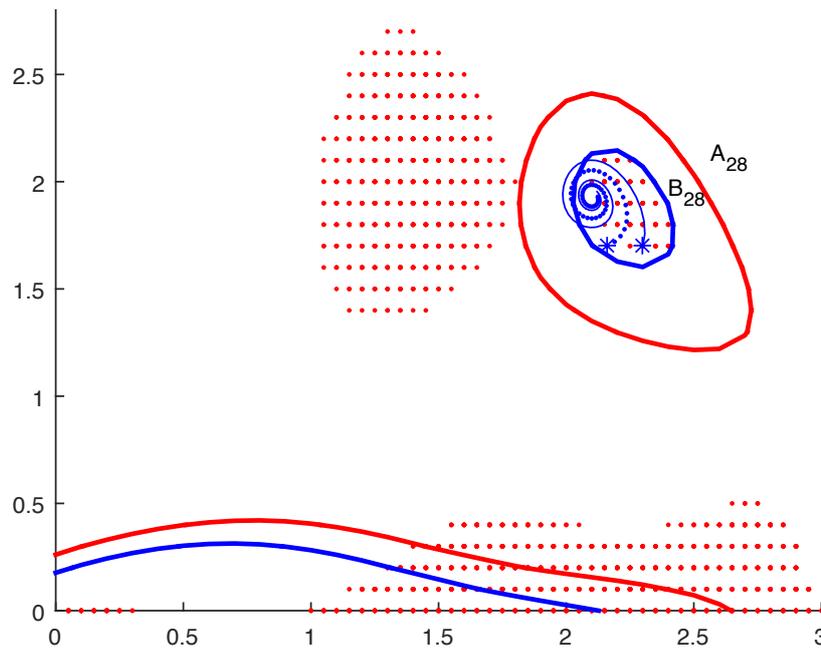


Figura 1: Componentes conexas dos conjunto de nível de  $A_{28}$ ,  $B_{28}$ , solução  $\varphi(t, x_0)$  com  $\alpha = 1.03$  e  $x_0 = (2.16, 1.7)$  (curva pontilhada), e  $\varphi(t, x_0)$  com  $\alpha = 1$  e  $x_0 = (2.3, 1.7)$  do Exemplo 1.

## 4 Conclusão

Neste trabalho foi apresentada uma extensão uniforme do princípio de invariância para a classe de sistemas positivos com parâmetros incertos, a qual é capaz de fornecer estimativas de conjuntos atratores e suas respectivas áreas de atração, mesmo quando a derivada da função auxiliar  $V$  assume valores positivos. O termo uniforme significa que as estimativas do atrator

e da área de atração não dependem dos parâmetros incertos do sistema. Trabalhos futuros irão propor condições de estabilidade menos conservadoras, por considerar as propriedades de algumas classes de sistemas positivos, como por exemplo, os sistemas cooperativos.

## Referências

- [1] P. U. Abara, “On the stability of positive nonlinear systems: Cooperative and concave system dynamics with applications to distributed networks”, Dissertação de Mestrado, University of Padova, 2014.
- [2] L. F. C. Alberto, T. R. Calliero e A. C. P. Martins, An invariance principle for nonlinear discrete autonomous dynamical systems, *IEEE Trans. Automat. Contr.*, 52 (2017) 692-697.
- [3] T. Chen e X. Liu,  $\mu$ -stability of nonlinear positive systems with unbounded time-varying delays, *IEEE Trans. Neural Netw. Learn. Syst.*, 1 (2016) 1-6.
- [4] W. R. Coimbra e L.F.C. Alberto, A uniform invariance principle for periodic systems with applications to synchronization, *Syst. Control Lett.*, 97 (2016) 48-54.
- [5] T. Kaczorek, Positivity and stability of discrete-time and continuous-time nonlinear systems, *Prz. Elektrotech.*, 1 (2015) 129-132.
- [6] T. Kaczorek, Stability of fractional positive nonlinear systems, *Arch. Control Sci.*, 25 (2015) 491-495.
- [7] P. De Leenheer and D. Aeyels, Stability properties of equilibria of classes of cooperative systems, *IEEE Trans. Automat. Contr.*, 46 (2001) 1996-2001.
- [8] Jacob Palis, Jr. and Welington de Melo, “Geometric Theory of Dynamical Systems: an Introduction”, Springer-Verlag, Berlin, 1982.
- [9] H. M. Rodrigues, L. F. C. Alberto e N. G. Bretas, On the invariance principle. generalizations and applications to synchronism, *IEEE Trans. Circuits Syst. I*, 47 (2000) 730-739.
- [10] H.M. Rodrigues, L.F.C. Alberto e N.G. Bretas, Uniform invariance principle and synchronization. robustness with respect to parameter variation, *J. Differ. Equ.*, 169 (2001) 228-254.
- [11] M.C. Valentino, V. A. Oliveira, L.F.C. Alberto e D. A. Sant Anna, An extension of the invariance principle for dwell-time switched nonlinear systems, *Syst. Control Lett.*, 61 (2012) 580-586.

## Previsão da Inflação: uma análise a partir de Redes Neurais

Cristina Zaniol, Jean Carlo Moraes,

Programa de Pós-Graduação em Matemática Aplicada, IME, UFRGS,  
91509-900, Porto Alegre, RS

E-mail: [criszaniol@gmail.com](mailto:criszaniol@gmail.com), [jean.moraes@ufrgs.br](mailto:jean.moraes@ufrgs.br),

**Resumo:** O controle inflacionário tem ganhado muita importância nos estudos econômicos nas últimas décadas, principalmente devido a adoção de metas para inflação a partir da década de 90, utilizada como política monetária por diversos países. Para fazer uma análise da inflação, uma medida que separa componentes persistentes e transitórios de um índice de inflação tem se mostrado importante. Na literatura, esta medida é conhecida como Núcleo de Inflação. Neste trabalho, são propostos Núcleos de Inflação baseados em wavelets das famílias Daubechies e Symlets, sendo utilizadas Redes Neurais para previsão da inflação para diversos horizontes de tempo. Para melhores resultados, são incorporados aos núcleos de inflação outros indicadores: a variação percentual mensal do Ibovespa, da confiança do consumidor e do índice de expectativas futuras. Os resultados de previsão utilizando wavelets são comparados com as previsões obtidas a partir dos Núcleos de Inflação utilizados pelo Bacen. Conclui-se que os núcleos baseados em wavelets apresentam um resultado superior quando comparado com os núcleos oficiais.

**Palavras-chave:** *Inflação, Núcleo de Inflação, previsão, Redes Neurais.*

### 1 Introdução

A inflação, embora um fenômeno bem retratado, suscitou discussões ao longo das últimas décadas devido o seu grande impacto em todas as esferas da sociedade, desde do planejamento familiar até a alocação de recursos de agentes financeiros e estatais. O controle inflacionário ganhou importância nos estudos econômicos, principalmente a partir dos anos 60, com vários casos de *hiperinflação*. No início da década de 90, vários países adotaram metas para a inflação, entre eles Austrália e Nova Zelândia [14]. Nestes países, aliado ao uso de estratégia de controle inflacionário, foi incorporado o uso do Núcleo de Inflação, uma ferramenta que possibilita o acompanhamento da inflação desconsiderando ruídos de curto prazo.

O Núcleo da Inflação representa tanto uma tendência, quanto um indicador da inflação futura. Assim, a inflação pode ser descrita através de uma componente persistente, o núcleo, e outra transitória, que representa um ruído [5, 7, 14]. Matematicamente, podemos descrever a inflação como

$$\pi_t = \pi_t^* + \epsilon_t, \quad (1)$$

onde  $\epsilon_t$  é a volatilidade presente no índice cheio, que representa uma transitoriedade do índice de inflação,  $\pi_t$ , que quando descartada resta apenas o Núcleo de Inflação,  $\pi_t^*$ . O problema do cálculo do Núcleo de Inflação está justamente em como separar estes dois componentes, o permanente e o transitório. Não há consenso sobre o que constitui um Núcleo de Inflação ideal, mas normalmente utiliza-se um conjunto de critérios para avaliá-los, em que deve satisfazer ao menos estes: redução de volatilidade e capacidade preditiva da inflação [3, 5, 12].

Neste trabalho, são propostos núcleos de inflação baseado em wavelets que, embora seja uma ferramenta importante para a decomposição da inflação e construção do núcleo, não possui grande capacidade preditiva. Desta forma, adota-se as Redes Neurais para geração de previsões, uma vez que possibilitam o tratamento de dados altamente complexos sem a descrição de um

modelo analítico específico [1,9]. Ademais, é possível tornar a previsão da inflação mais acurada com a incorporação no modelo de outros indicadores como, por exemplo, a expectativa e a confiança do mercado. Para verificar o desempenho do modelo proposto, os núcleos baseados em wavelets tem o resultado de predição comparados com os núcleos de inflação oficiais brasileiros: os núcleos por exclusão da série IPCA-EX, o de dupla ponderação (IPCA-DP) e o de médias aparadas (IPCA-MA) [14]. Ainda que haja estudos que utilizem wavelets para o cálculo do núcleo de inflação [2, 5, 8, 10], destaca-se que a utilização conjunta de wavelets e Redes Neurais não foi encontrada na literatura.

Este trabalho está dividido em: Seção 2 e 3 apresentam o arcabouço teórico para o desenvolvimento do modelo, cujos resultados são apresentados na Seção 4 e, em seguida, são tecidas algumas considerações finais na Seção 5.

## 2 Wavelets

Wavelets são famílias de bases ortonormais que podem ser usadas para representar funções. Desde o surgimento da técnica, wavelets têm sido usadas com êxito em várias aplicações para remover ruídos e desvelar o verdadeiro sinal. Em  $L^2(\mathbb{R})$ , bases de wavelets são obtidas através de translações e dilatações de uma função escala  $\psi$ , em que os elementos da base são:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right), \quad a > 0, \quad (2)$$

com  $a$  e  $b$  representando os parâmetros de escala e de translação, respectivamente. Na forma discreta, a wavelet é escrita como

$$\psi_{m,n}(t) = \frac{1}{a_0^{m/2}}\psi(a_0^{-m}t - nb_0), \quad (3)$$

onde  $m$  e  $n$  tem a mesma função que  $a$  e  $b$  na Equação 2, com  $a = a_0^m$ ,  $m \in \mathbb{Z}$  e  $a_0 \neq 1$ ; e  $b = nb_0a_0^m$ ,  $n \in \mathbb{Z}$ . Para reescrever a série temporal como uma representação multirresolução (decomposição das funções em componentes de diferentes volatidades), define-se os coeficientes de detalhe e de aproximação. O coeficiente de detalhe é dado por

$$T_{m,n} = \int_{-\infty}^{\infty} f(t)\psi_{m,n}(t)dt, \quad (4)$$

e o coeficiente de aproximação,

$$S_{m,n} = \int_{-\infty}^{\infty} f(t)\phi_{m,n}(t)dt. \quad (5)$$

A aproximação na escala  $m$  é definida por

$$f_m(t) = \sum_{n=-\infty}^{\infty} S_{m,n}\phi_{m,n}(t) \quad (6)$$

e os detalhes,

$$d_m = \sum_{n=-\infty}^{\infty} T_{m,n}\psi_{m,n}(t). \quad (7)$$

Desta forma, é possível reescrever uma série temporal como a soma da aproximação e dos detalhes, em que

$$f(t) = f_{m_0}(t) + \sum_{m=-\infty}^{m_0} d_m(t). \quad (8)$$

Os detalhes  $d_m$  são relacionados com os componentes de mais alta frequência da série temporal (maior volatidade) e  $f_m$ , de mais baixa frequência (menor volatidade) [6, 11, 13].

### 3 Redes Neurais

As Redes Neurais são algoritmos com habilidade de aprendizagem, generalizando padrões que podem ser utilizados na classificação e na predição [1,9]. Um Rede Neural pode ser representada por um conjunto de neurônios, onde cada neurônio  $k$  realiza uma ponderação de suas entradas e pode ser matematicamente representado por

$$u_k = \sum_{j=1}^m w_{kj}x_j, \quad (9)$$

onde  $w_{kj}$  é o peso utilizado pelo neurônio  $k$  para um dado de entrada  $x_j$ . A resposta da Rede Neural formada por um neurônio  $k$ ,  $y_k$ , é dada por

$$y_k = \sigma(v_k), \quad (10)$$

onde

$$v_k = u_k + b_k, \quad (11)$$

com  $b_k$  representando um parâmetro associado com limiar de ativação do neurônio; e  $\sigma(\cdot)$  é uma função de ativação, cuja finalidade é análoga às sinapses do neurônio: inibir ou excitar a transmissão da informação quando é ultrapassado determinado limiar.

### 4 Resultados

O cálculo do núcleo de inflação baseado em wavelets foi conduzido da seguinte forma: a partir do IPCA, foram calculadas decomposições em cinco níveis e, na reconstrução, foram mantidos apenas três níveis de detalhes (frequências referentes ao período até 16 meses). Desta forma, o sinal torna-se mais suave e tende a representar a tendência da inflação. As wavelets utilizadas são da família das Daubechies e Symlets, mais especificamente: *db2*, *db4*, *db6*, *db8*, *db10*, *sym3*, *sym4*, *sym6*, *sym8*, *sym10* [11]. Estas medidas de núcleo, então, foram utilizadas como dado de entrada para uma Rede Neural, juntamente com a variação percentual mensal do Ibovespa, do índice de confiança do consumidor e do índice de expectativas futuras. Tanto o IPCA, quanto as medidas de núcleo de inflação oficiais, a confiança do consumidor e o índice de expectativa futura utilizados neste trabalho foram obtidos no Sistema Gerenciador de Séries Temporais do Bacen [4], sendo delimitadas no período de julho de 2006 a janeiro de 2019.

A Rede Neural utilizada possui apenas uma camada escondida, conforme Figura 1, sendo classificada como recorrente ou recursiva. Enquanto é utilizado um *lag* 12 para os núcleos de inflação, o Ibovespa, a confiança e a expectativa, para o IPCA, é utilizado um *lag* de 4, isto é, utilizando a representação temporal,

$$y(t) = \sigma(y(t-1), y(t-2), y(t-3), y(t-4), x(t-1), x(t-2), \dots, x(t-11), x(t-12)). \quad (12)$$

O algoritmo de aprendizagem adotado foi o SCG (*Scaled Conjugate Gradient*) e o Erro Quadrático Médio (EQM) foi utilizado como critério de performance. Os dados foram divididos assim: 80% foram utilizados para o treinamento, 10% para validação e 10% para teste. Para a saída, é utilizado apenas um neurônio; na camada escondida, por sua vez, são utilizados  $n$  neurônios, com  $n$  variando de 1 a 100, e a *função linear saturada simétrica* é adotada como função de ativação, sendo esta representada por

$$\sigma(v_k) = \begin{cases} -1, & \text{se } v_k \leq -1 \\ v_k, & \text{se } -1 < v_k < 1 \\ 1, & \text{se } v_k \geq 1 \end{cases} . \quad (13)$$

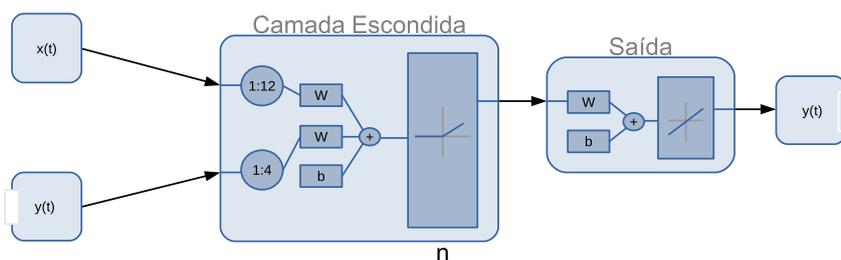


Figura 1: Arquitetura da Rede Neural com  $n$  neurônios na camada escondida.

Tabela 1: Erro Quadrático Médio (EQM) e Coeficiente de Determinação ( $R^2$ ) da previsão gerada pela Rede Neural de uma camada escondida.

	Neurônios	EQM da Rede Neural			EQM da Previsão					$R^2$
		Treino	Validação	Teste	k+1	k+3	k+6	k+9	k+12	
IPCA-EX0	33	0.0102	0.0735	0.0949	0.0251	0.0148	0.0082	0.0103	0.0084	0.6953
IPCA-EX1	66	0.0095	0.0784	0.0466	0.0202	0.0542	0.0304	0.0219	0.0163	0.7555
IPCA-EX2	56	0.0093	0.0633	0.0608	0.0199	0.0046	0.0111	0.0817	0.0612	0.7602
IPCA-EX3	56	0.0083	0.0760	0.0592	0.0203	0.0013	0.0174	0.0921	0.0704	0.7563
IPCA-DP	66	0.0045	0.0920	0.0686	0.0198	0.0051	0.0119	0.0999	0.0666	0.7622
IPCA-MA	66	0.0059	0.0962	0.0831	0.0228	0.0103	0.0065	0.0272	0.0175	0.7295
db2	15	0.0020	0.0243	0.0482	0.0089	0.0014	0.0012	0.0285	0.0258	0.8914
db4	61	0.0012	0.0368	0.0317	0.0079	0.0071	0.0193	0.0518	0.0471	0.9044
db6	69	0.0007	0.0202	0.0212	0.0047	0.0004	0.0126	0.0074	0.0246	0.9424
db8	45	0.0042	0.0310	0.0333	0.0099	0.0046	0.0015	0.0020	0.0014	0.8790
db10	69	0.0013	0.0192	0.0266	0.0057	0.0000	0.0045	0.0091	0.0195	0.9327
sym3	15	0.0020	0.0243	0.0482	0.0089	0.0014	0.0012	0.0285	0.0258	0.8914
sym4	61	0.0010	0.0333	0.0310	0.0073	0.0166	0.0196	0.0619	0.0571	0.9121
sym6	54	0.0012	0.0286	0.0204	0.0059	0.0002	0.0002	0.0002	0.0021	0.9290
sym8	56	0.0042	0.0208	0.0140	0.0068	0.0025	0.0023	0.0061	0.0078	0.9172
sym10	37	0.0024	0.0290	0.0141	0.0063	0.0007	0.0031	0.0027	0.0186	0.9228

A Tabela 1 apresenta os melhores resultados de performance, isto é, aqueles com menor EQM, assim como o Coeficiente de Determinação ( $R^2$ ), uma medida de explicação do modelo e seu ajustamento ao valor observado do IPCA. Para verificar a capacidade preditiva de cada um dos modelos gerados pela Rede Neural, foram realizados testes para diversos horizontes de tempo. Para cada um deles, a rede neural realiza testes em que os dados até o mês  $k$  são utilizados na preparação da rede neural e, posteriormente, são realizadas previsões para um  $(k + 1)$ , três  $(k + 3)$ , seis  $(k + 6)$ , nove  $(k + 9)$  e doze meses  $(k + 12)$ .

Primeiramente, em relação ao número de neurônios, é possível perceber que as redes neurais necessitaram mais de 30 neurônios para gerar o melhor resultado. Dentre as exceções, o núcleo *db2* e o *sym3* que necessitaram apenas 15 neurônios para uma capacidade de explicação do modelo de 89.14%. Nenhum dos núcleos ultrapassou 70 neurônios, sendo que a maioria dos melhores resultados estão concentrados entre 50 e 70 neurônios.

Quanto à construção da rede neural, é possível perceber que o EQM do treino foi inferior a 0.01 para todos os núcleos, com exceção do IPCA-EX0; da validação, inferior a 0.10, onde os maiores erros são observados nas previsões geradas a partir dos núcleos de inflação oficiais; o mesmo comportamento em relação à validação pode ser observado para o teste. Percebe-se que a construção da rede neural com o uso dos núcleos de inflação baseados em wavelets proporciona EQM menores que os núcleos oficiais.

Em relação à previsão, percebe-se mais uma vez que a previsão com a utilização dos núcleos de inflação baseados em wavelets proporciona, em geral, melhores resultados. Para a previsão de um mês, os três melhores resultados são *db6*, *db10* e *sym10*; para três meses, *db10*, *sym6* e

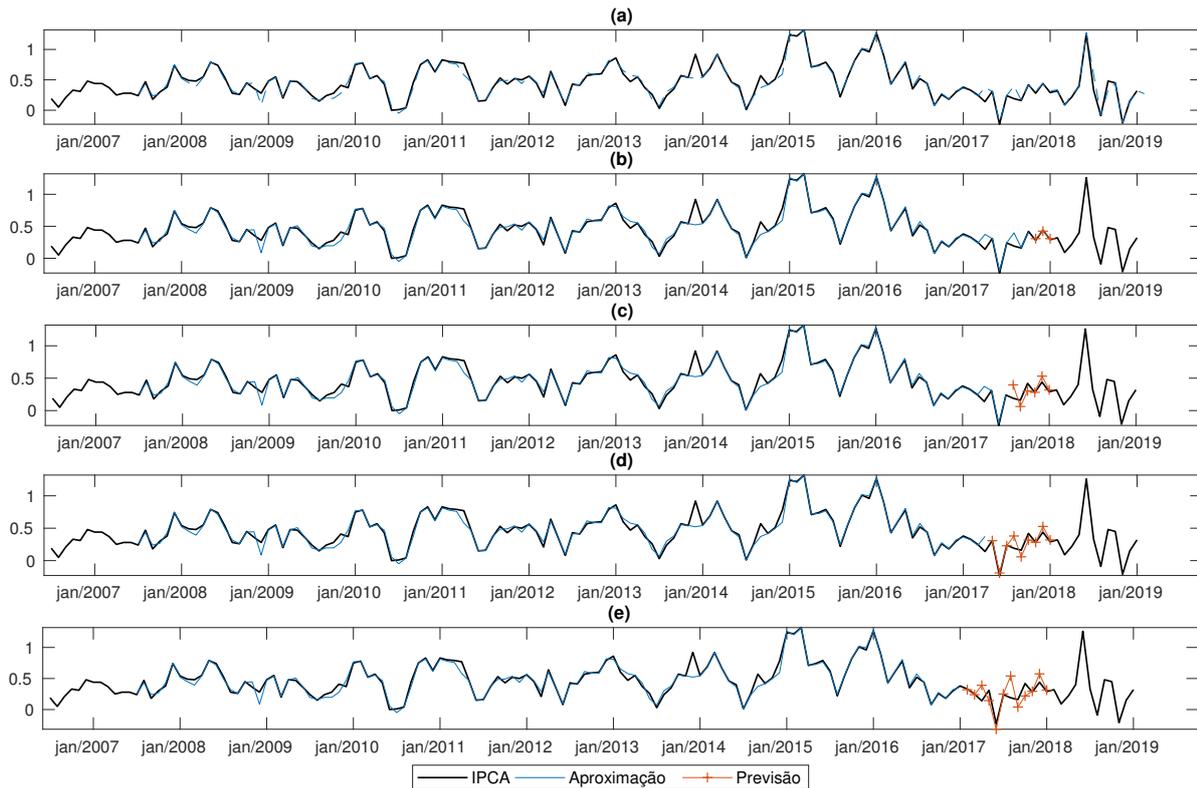


Figura 2: Resultado da rede neural usando o núcleo de inflação baseado na wavelet  $db6$ . Em preto, o IPCA e, em azul, a aproximação gerada pela rede neural: (a) aproximação da rede neural; em vermelho, previsão para (b) três meses, (c) seis meses, (d) nove meses e (e) doze meses.

o IPCA-EX3; para seis meses,  $sym6$ ,  $db2$  e  $sym3$ ; para nove meses,  $sym6$ ,  $db8$  e  $sym10$ ; e para doze meses,  $db8$ ,  $sym6$  e  $sym8$ . Percebe-se que há desempenho diferente para diversos horizontes de tempo dependendo do núcleo utilizado.

Por fim, em relação à capacidade de explicação do modelo, percebe-se que os núcleos oficiais possuíram um desempenho inferior aos dos núcleos baseados em wavelets: os primeiros não conseguem explicar 80% do modelo, enquanto que os núcleos baseados em wavelets explicam mais de 87%. O melhor resultado é o do núcleo  $db6$ , que consegue explicar 94.24% do período analisado (Figura 2). Percebe-se que, para períodos com singularidades geradas por eventos díspares, como o aumento do preço da energia elétrica no início de 2015 e a greve dos caminhoneiros em maio de 2018, puderam ser descritos pela rede neural; este fato deve-se à incorporação de indicadores diversos ao núcleo da inflação.

## 5 Conclusão

Os resultados obtidos neste trabalho indicam que várias medidas de núcleo apresentadas possuem um resultado mais satisfatório na predição da inflação que os núcleos de inflação oficiais; tanto na construção da rede neural e na previsão, quanto na capacidade de explicação da série temporal, é possível perceber que os núcleos baseados em wavelets apresentam um desempenho muito superior aos demais. Desta forma, há indícios que os núcleos de inflação baseados em wavelets sejam mais apropriadas para a previsão e, conseqüentemente, para o auxílio no delineamento das metas de inflação que as medidas atualmente utilizadas pelo Bacen. Estudos futuros incluem novas arquiteturas de redes, com mais camadas escondidas, e também a utilização de outros dados macroeconômicos para a predição da inflação.

## Agradecimentos

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

## Referências

- [1] A. K. Alexandridis, and A. D. Zapranis. *Wavelet neural networks: with applications in financial engineering, chaos, and classification*. John Wiley & Sons, 2014.
- [2] A. A. Denardin, A. Kozakevicius, A. A. Schmidt. Avaliação Da Medida De Núcleo De Inflação Baseada No Método Wavelet Para O Brasil. *Proceedings of the 44th Brazilian Economics Meeting*, 2018.
- [3] BANCO CENTRAL DO BRASIL. Relatório de inflação: Viés e Capacidade Preditiva das Medidas de Inflação Subjacentes - Dezembro 2018. *Comitê de Política Monetária*, 2018. Url: <https://bit.ly/2soy3Xd>.
- [4] BANCO CENTRAL DO BRASIL. SGS - Sistema Gerenciador de Séries Temporais - v2.1. Url:<https://bit.ly/1ivyjfd>
- [5] D. Baqaee. Using wavelets to measure core inflation: The case of New Zealand. *The North American Journal of Economics and Finance*, volume 21, número 3, p. 241-255, 2010.
- [6] C. Chui and Q. Jiang. *Applied Mathematics: Data Compression, Spectral Fourier Analysis, Wavelets, and Applications*. Mathematics Textbooks for and Engineering. Atlantis Press, 2013.
- [7] T. Cogley. A simple adaptive measure of core inflation. *Journal of Money, Credit and Banking*, 34(1), p. 94-113.
- [8] K. Dowd, J. Cotter, L. Loh. US core inflation: A wavelet analysis. *Macroeconomic Dynamics*, volume 15, número 4, p. 513-536, 2011.
- [9] S. Haykin. *Neural Networks and Learning Machines*, vol.10, Prentice Hall, 2009.
- [10] E. Lahura, M. Vega. Evaluation of Wavelet Based Core Inflation Measures: Evidence from Perú. *Documento de Trabajo*, Departamento de Economía - Pontificia Universidad Católica del Peru, número 320, 2011.
- [11] S. Mallat. *A Wavelet Tour of Signal Processing*. Third Edition: The Spa Academic Press, 3rd edition, 2008.
- [12] C. R. Marques, P. D. Neves, L. M. Sarmiento. Evaluating core inflation indicators. *Economic modelling*, volume 20, número 4, p. 765-775, 2003.
- [13] Y. Nievergelt. *Wavelets Made Easy*. Birkhauser, 1999.
- [14] T. N. T. Silva Filho, F. M. R. Figueiredo (2011). Has core inflation been doing a good job in Brazil?. *Revista Brasileira de Economia*, 65(2), p. 207-233, 2011.

## Definição de um modelo representativo das habitações de interesse social de Joinville/SC para simulação computacional

**Jayne Garcia<sup>a</sup>, Elisa Henning<sup>b</sup>, Ana Mirthes Hackenberg<sup>a</sup>, Janine Garcia<sup>a</sup>**

<sup>a</sup> Universidade do Estado de Santa Catarina - Departamento de Engenharia Civil

<sup>b</sup> Universidade do Estado de Santa Catarina - Departamento de Matemática  
89219-710, Centro de Ciências Tecnológicas, Joinville, SC

E-mail: jaynegarcia00@gmail.com, elisa.henning@udesc.br, ana.hackenberg@udesc.br,  
janinegarcia01@gmail.com

**Resumo:** *A simulação computacional é uma ferramenta que permite avaliar o desempenho e consumo energético da edificação, desde que se obtenha um modelo que reproduza as suas reais condições. Ao se analisar um grupo de edificações, é conveniente aplicar métodos estatísticos de amostragem, buscando determinar um modelo representativo do grupo analisado. Tendo isso em vista, o objetivo do estudo foi definir um modelo representativo das habitações multifamiliares de interesse social localizadas no município de Joinville/SC, para uso em simulação computacional. No estudo, foi aplicado o método de amostragem não probabilística por julgamento. Primeiramente, realizou-se um levantamento das características construtivas de cinco habitações sociais existentes em Joinville, por meio dos projetos e memorial descritivo dessas habitações. Após isso, foi determinado o intervalo de confiança de 95% para as variáveis quantitativas, utilizando a aproximação pela distribuição t de Student. Quanto às variáveis qualitativas, verificou-se a moda do conjunto de dados. Dessa forma, foi definido um modelo representativo para as habitações analisadas, que atendessem aos intervalos de confiança e as características qualitativas mais frequentes. Por fim, realizou-se a modelagem do projeto representativo no SketchUp, o qual será utilizado em simulações futuras.*

**Palavras-chave:** *habitação de interesse social; modelo representativo; simulação computacional.*

### 1. Introdução

O Programa Minha Casa Minha Vida foi uma iniciativa do governo federal, voltado para as famílias de baixa renda, buscando atender a demanda social por moradia [3]. No entanto, alguns autores afirmam que os projetos das habitações sociais construídas no âmbito do programa são padronizados, sem considerar as particularidades do clima local e o desempenho térmico da edificação [1, 8].

Nesse sentido, torna-se necessária a análise do comportamento térmico e energético dessas habitações, possibilitando a implementação de medidas que proporcionem condições mínimas de conforto aos usuários. Tal análise requer o conhecimento de inúmeras variáveis, tendo em vista a sua complexidade e de modo que reflita o real desempenho da edificação. Dentre as ferramentas disponíveis, destaca-se a simulação computacional, que permite avaliar a eficiência energética para diferentes alternativas de projeto, bem como estimar o consumo de energia [9].

Nos casos em que a simulação do desempenho termoenergético envolve um grupo de edificações, é útil aplicar técnicas específicas de amostragem com a finalidade de determinar um modelo computacional, cujas características o tornam representativo do grupo analisado [4]. Assim sendo, a estatística é um instrumento fundamental para desenvolver pesquisas em diversas áreas (incluindo a engenharia), sobretudo quando há restrições de tempo, dinheiro e outros recursos [5].

Com base no exposto, este artigo tem como objetivo definir um modelo representativo das habitações multifamiliares de interesse social localizadas no município de Joinville/SC, para uso em simulação computacional.

## 2. Método

Na presente pesquisa, adotou-se a técnica de amostragem não probabilística por julgamento, que compreende a seleção intencional de elementos para pertencer à amostra, com base no julgamento de que tais elementos sejam representativos da população [6, 12].

Sendo assim, foi definido um modelo representativo das cinco habitações multifamiliares de interesse social existentes em Joinville (Figura 1), mediante a caracterização dessas habitações em termos quantitativos e qualitativos, bem como análise estatística dos dados coletados.

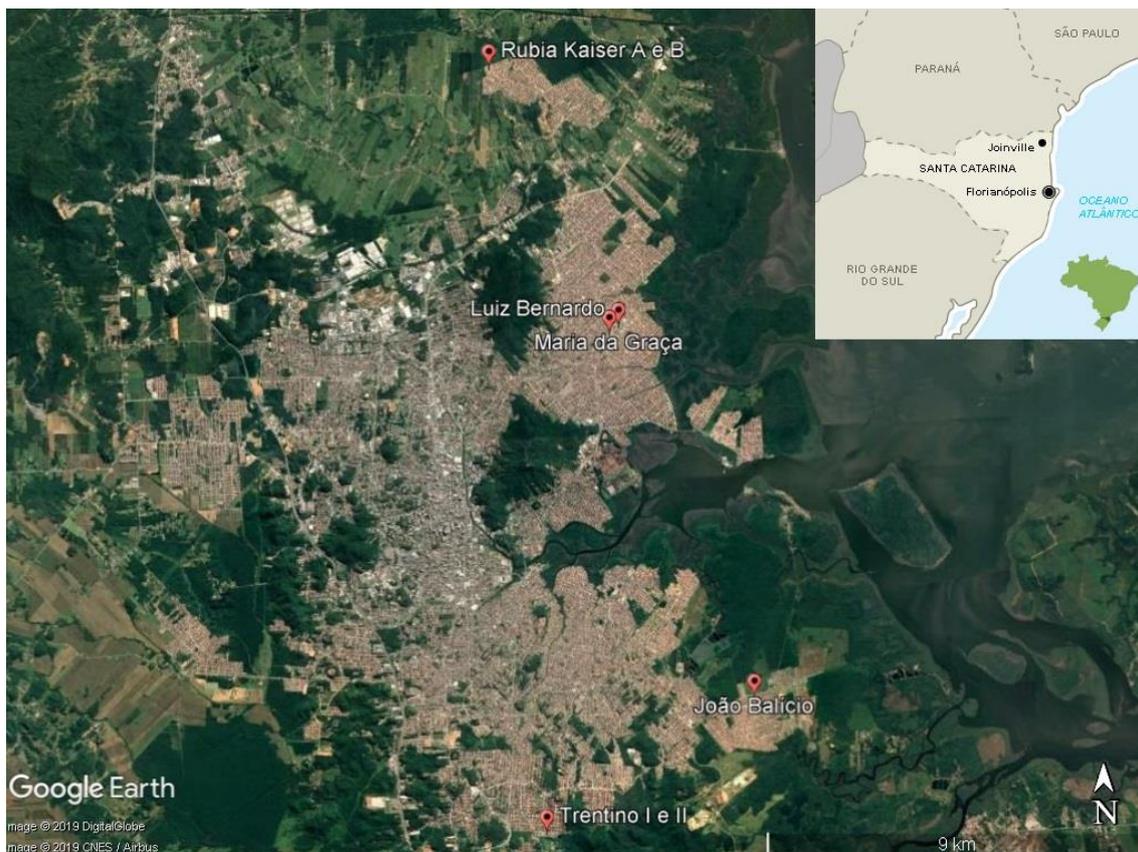


Figura 1: Habitações multifamiliares de interesse social de Joinville/SC

O método descrito a seguir foi aplicado em estudos similares para amostras pequenas [13, 14]. Salienta-se que são consideradas amostras pequenas aquelas em que o número de elementos é inferior a 30 [10]. Dessa forma, pode-se afirmar que a presente pesquisa atende as particularidades de aplicação do método.

A primeira etapa da pesquisa envolveu o levantamento das características geométricas e construtivas das habitações sociais analisadas, de acordo com seus respectivos projetos e memorial descritivo, os quais foram disponibilizados pela instituição financeira Caixa Econômica Federal.

Nessa etapa, foram levantadas características relacionadas à geometria da edificação, número total de blocos e unidades habitacionais (apartamentos), área dos ambientes e área útil da UH<sup>1</sup>, tipo e dimensões das aberturas, além dos materiais e componentes construtivos que compõem a envoltória (piso, parede e cobertura).

Em seguida, realizou-se a classificação e análise estatística dos dados levantados, com o auxílio do software R versão 3.5.1 e interface RStudio. Para as variáveis quantitativas, referentes à área dos ambientes e área útil da UH, verificou-se a aderência dos dados à

<sup>1</sup> Unidade Habitacional - UH

distribuição normal através do teste de Shapiro-Wilk. É importante ressaltar que, de modo geral, o teste de Shapiro-Wilk é o mais poderoso dentre outros testes de normalidade, tais como Kolmogorov-Smirnov, Lilliefors, Anderson-Darling e D'Agostino-Pearson [11, 7]. Por isso, optou-se por utilizar o teste em questão.

Além dessas variáveis, considerou-se também o percentual de abertura para ventilação nos ambientes de permanência prolongada (sala e dormitórios). Esse percentual é obtido pelo cálculo da área das aberturas em relação a área do ambiente.

Na sequência, foi realizada uma análise descritiva dos dados quantitativos, na qual foram calculados a média, desvio padrão e coeficiente de variação. Caso a suposição de normalidade fosse aceita, aplicou-se a aproximação pela distribuição t de Student, a fim de obter o intervalo de confiança de 95% para a média. Quanto às variáveis qualitativas, verificou-se a moda do conjunto de dados, ou seja, as propriedades das aberturas e envoltória que ocorriam com maior frequência.

Por fim, dentre as HIS<sup>2</sup> analisadas, averiguou-se aquela que atendia simultaneamente ao intervalo de confiança de 95% para as variáveis quantitativas e as características mais frequentes das variáveis qualitativas. Outro critério utilizado foi a NBR 9.050 [2], uma vez que o modelo representativo também deveria atender à norma de acessibilidade. Foi elaborado então um checklist com as especificações de projeto contidas na norma.

Ao final, realizou-se a modelagem da unidade habitacional representativa no programa SketchUp 2017 com o plug in Euclid 0.9.3, podendo ser utilizado posteriormente em simulações termoenergéticas no programa EnergyPlus.

### 3. Resultados e Discussões

A seguir, são apresentados os resultados do estudo no que se refere à definição do modelo representativo. Na etapa de coleta dos dados, foi possível levantar as principais características das HIS multifamiliares de Joinville/SC. Contudo, destaca-se que foram encontradas incongruências entre os projetos e o memorial descritivo das habitações. Nesses casos, foram consideradas as informações contidas nos projetos, visto que tais documentos foram utilizados durante a execução dessas habitações.

A Tabela 1 mostra as características das habitações sociais multifamiliares existentes em Joinville/SC. Percebe-se que as habitações sociais analisadas são conjuntos habitacionais de grande porte, com diversos blocos e centenas de unidades habitacionais. Além disso, observa-se que o número de pavimentos em cada bloco varia entre 4 e 5. Na maioria dos casos, foram constatadas 4 unidades habitacionais por pavimento. As unidades habitacionais contavam com sala, cozinha, banheiro, dois dormitórios e área de circulação entre os dormitórios, em sua maioria.

HIS	UH	Blocos	Pavimentos por bloco	UH por pavimento
João Balício	100	3	5	$\frac{8}{4}$
Luiz Bernardo	280	14	5	4
Maria da Graça	320	20	4	4
Rubia Kaiser A e B	640	40	4	4
Trentino I e II	784	49	4	4
Total	2124	-	-	-

Tabela 1: Características das habitações sociais de Joinville/SC

Por meio da Tabela 2, é possível analisar a estatística descritiva, probabilidade de significância do teste de Shapiro-Wilk e intervalo de confiança de 95% para as variáveis quantitativas. Nota-se que, em geral, não houve uma variabilidade alta nos dados quantitativos

<sup>2</sup> Habitação de Interesse Social - HIS

levantados. Com relação ao teste de Shapiro-Wilk, a probabilidade de significância (p-valor) foi superior ao nível de significância adotado ( $\alpha=0,05$ ) em todas as variáveis consideradas. Consequentemente, não há evidências para rejeitar a hipótese nula de que os dados são normalmente distribuídos e, portanto, a aproximação pela distribuição t de Student pode ser aplicada. Desse modo, foram determinados os limites inferior e superior do intervalo para cada variável quantitativa, com nível de confiança de 95%.

As características mais frequentes das variáveis qualitativas são apresentadas na Tabela 3 e Tabela 4. Convém ressaltar que as propriedades das aberturas nos ambientes e composições dos componentes construtivos da envoltória eram semelhantes nas HIS, diferenciando-se entre si em poucos aspectos.

Área (m <sup>2</sup> )	Mínimo	Máximo	Média	Desvio padrão	Coeficiente de variação (%)	P-valor ( $\alpha=0,05$ )	Intervalo de confiança (95%)		Valores adotados
							Inferior	Superior	
Sala	10,25	12,40	11,43	1,06	9,24	0,07	10,12	12,74	10,25
Cozinha	5,82	7,87	6,84	0,96	14,08	0,24	5,65	8,04	7,02
Dormitório 1	8,40	9,28	8,69	0,36	4,09	0,18	8,25	9,13	8,58
Dormitório 2	6,72	7,62	7,24	0,37	5,10	0,64	6,78	7,69	7,08
Banheiro	2,88	5,04	4,15	0,84	20,31	0,75	3,10	5,20	4,76
Circulação	1,23	2,04	1,51	0,46	30,41	0,06	0,37	2,65	2,04
Área útil UH	38,01	40,43	39,25	0,90	2,29	0,89	38,14	40,37	39,71

Percentual de abertura (%)	Mínimo	Máximo	Média	Desvio padrão	Coeficiente de variação (%)	P-valor ( $\alpha=0,05$ )	Intervalo de confiança (95%)		Valores adotados
							Inferior	Superior	
Sala	5,57	17,59	13,42	0,05	36,10	0,29	7,40	19,43	15,13
Dormitório 1	15,52	17,14	16,60	0,01	3,94	0,23	15,79	17,41	16,79
Dormitório 2	18,90	21,43	19,94	0,01	5,18	0,62	18,66	21,22	20,34

Tabela 2: Estatística descritiva e intervalo de confiança de 95%

Aberturas (janela)	Dimensões (cm)	Área (m <sup>2</sup> )	Tipo	Material
Sala	160x120	1,92	correr (2 folhas)	esquadria de alumínio e vidro simples (3mm)
Cozinha	100x100	1,00	correr (2 folhas)	esquadria de alumínio e vidro simples (3mm)
Dormitórios	120x120	1,44	correr (2 folhas)	esquadria de alumínio e vidro simples (3mm)
Banheiro	60x60	0,36	maxim-ar	esquadria de alumínio e vidro simples (3mm)

Tabela 3: Propriedades das aberturas

Envoltória	Composição
Piso	contato com o solo, laje e piso cerâmico
Parede	alvenaria estrutural de bloco cerâmico (14x19x29cm), com revestimento interno/externo (2cm) e pintura
Cobertura	laje maciça de concreto (10cm) e telha de fibrocimento (6mm)

Tabela 4: Composição da envoltória

Os valores adotados (Tabela 2) referem-se à habitação de interesse social definida como modelo representativo, cujas características atendiam aos limites do intervalo de confiança, propriedades das aberturas e composição da envoltória mais recorrentes, assim como às especificações da NBR 9.050.

O modelo representativo possui 39,71m<sup>2</sup> de área útil e 3 ambientes de permanência prolongada (sala e 2 dormitórios), sendo que as dimensões dos ambientes são indicadas na Tabela 2. A planta baixa do modelo representativo é apresentada na Figura 2, enquanto que a modelagem geométrica pode ser visualizada na Figura 3. Ressalta-se que a planta baixa do modelo representativo atende às especificações da norma de acessibilidade, sendo acessível a pessoas em cadeira de rodas.

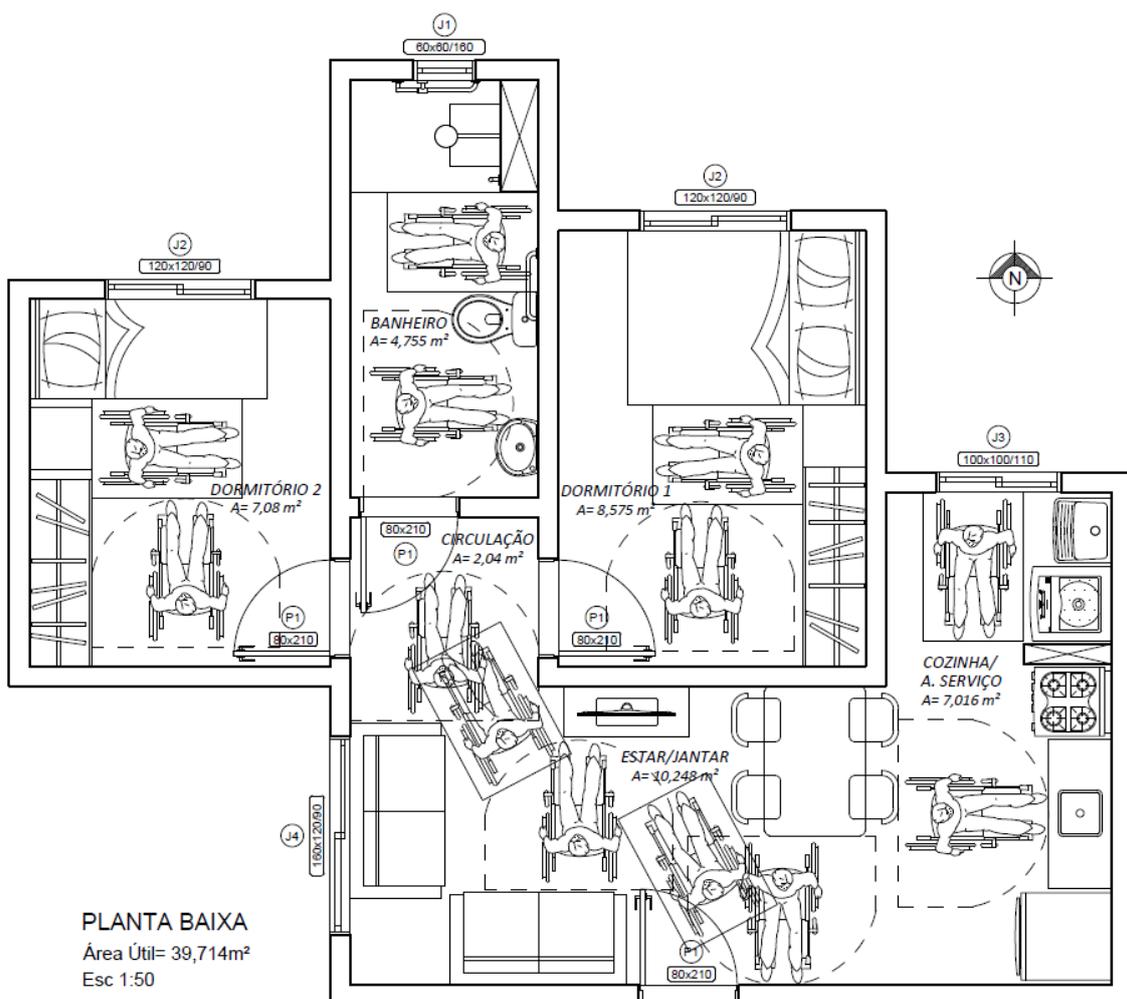


Figura 2: Planta baixa do modelo representativo

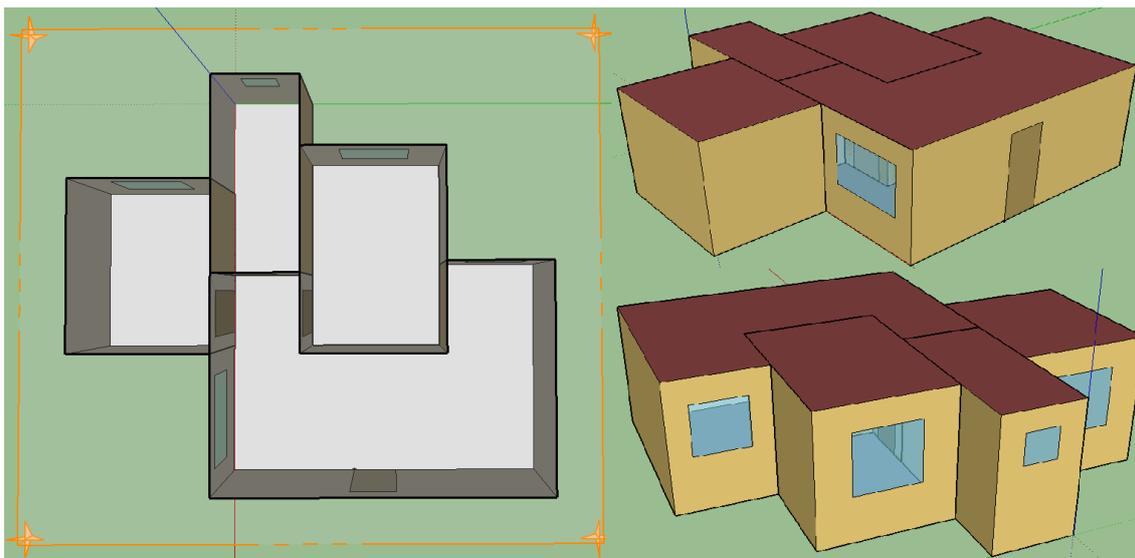


Figura 3: Modelagem geométrica no SketchUp

#### 4. Conclusão

O método adotado permitiu definir um modelo para uso em simulação computacional de forma a representar adequadamente a população, considerando um nível de significância de 5%. Entretanto, é necessário destacar que os resultados obtidos podem ter sido afetados devido à certas limitações, como o tamanho da amostra.

Na etapa de levantamento dos dados, verificou-se que os projetos das habitações sociais implementadas em Joinville seguem o mesmo padrão, apresentando poucas diferenças nas características construtivas e especificações dos materiais.

A partir do modelo representativo, será possível analisar o desempenho termoenergético dessas habitações em simulações futuras, utilizando o programa EnergyPlus. Além disso, pretende-se verificar os parâmetros mais influentes na eficiência energética do modelo, levando em consideração as características climáticas locais.

#### Referências

- [1] Almeida, L. S. S., “Influência de parâmetros físicos e geométricos no desempenho termoenergético de habitações unifamiliares”, Dissertação de Mestrado, Programa de Pós-Graduação em Engenharia Civil - Universidade Federal de Santa Catarina, 2014.
- [2] Associação Brasileira de Normas Técnicas – ABNT, “NBR 9050: acessibilidade a edificações, mobiliário, espaços e equipamentos urbanos”, Rio de Janeiro, 2015.
- [3] Câmara Brasileira da Indústria da Construção – CBIC, “Perenidade dos programas habitacionais”, CBIC, Brasília, 2016.
- [4] Carlo, J. C., “Desenvolvimento de metodologia de avaliação da eficiência energética do envoltório de edificações não-residenciais”, Tese de Doutorado, Programa de Pós-Graduação em Engenharia Civil - Universidade Federal de Santa Catarina, 2008.
- [5] Devore, J. L., “Probabilidade e estatística para engenharia e ciências”, Cengage Learning, São Paulo, 2016.
- [6] Kelley, K.; Clark, B.; Brown, V.; Sitzia, J., Good practice in the conduct and reporting of survey research, *Internacional Journal for Quality in Health Care*, vol. 15, no. 3, pp. 261-266, (2003).

- [7] Keskin, S., Comparison of several univariate normality tests regarding type I error rate and power of the test in simulation based small samples, *Journal of Applied Science Research*, vol. 2, no. 5, pp. 296-300, (2006).
- [8] Kowaltowski, D. C. C. K.; Granja, A. D., The concept of desired value as a stimulus for change in social housing in Brazil, *Habitat International*, vol. 35, no. 3, pp. 435-446, (2011).
- [9] Mendes, N.; Westphal, F. S.; Lamberts, R.; Cunha Neto, J. A. B., Uso de instrumentos computacionais para análise do desempenho térmico e energético de edificações no Brasil, *Ambiente Construído*, vol. 5, no. 4, pp. 47-68, (2005).
- [10] Montgomery, D.C.; Runger, G. C., “Applied Statistics and Probability for Engineers”, Wiley, USA, 2014.
- [11] Razali, N. M., Wah, Y. B., Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests, *Journal of Statistical Modeling and Analytics*, vol. 2, no.1, pp. 21-33, (2011).
- [12] Reis, M. M., “Estatística aplicada à administração”, UFSC, Florianópolis, 2008.
- [13] Schaefer, A.; Almeida, L. S. S.; Ghisi, E.; Freitas, M.N.; Duarte, E, Determinação de modelos de habitação de interesse social da grande Florianópolis para uso em simulação computacional, em “IV Workshop - Rede de Pesquisa: Uso racional de água e eficiência energética em habitações de interesse social”, pp. 259-280, Aracaju, 2012.
- [14] Triana, M. A.; Lamberts, R.; Sassi, P., Characterisation of representative building typologies for social housing projects in Brazil and its energy performance, *Energy Policy*, vol. 87, pp. 524-541, (2015).



## Verificação da sazonalidade no consumo de água em residências unifamiliares de Joinville

**Janine Garcia<sup>a</sup>, Elisa Henning<sup>b</sup>, Andreza Kalbusch<sup>a</sup>, Jayne Garcia<sup>a</sup>**

<sup>a</sup> Universidade do Estado de Santa Catarina - Departamento de Engenharia Civil

<sup>b</sup> Universidade do Estado de Santa Catarina - Departamento de Matemática  
89219-710, Centro de Ciências Tecnológicas, Joinville, SC

E-mail: janinegarcia01@gmail.com, elisa.henning@udesc.br, andreza.kalbusch@udesc.br,  
jaynegarcia00@gmail.com

**Resumo:** *O estudo do comportamento sazonal do consumo de água é fundamental para o planejamento da demanda de água e criação de programas voltados à conservação desse recurso. Nesse sentido, o objetivo deste estudo foi verificar a existência de sazonalidade no consumo de água em residências unifamiliares de Joinville. Para isso, foram utilizados dados de consumo de água de 43 residências unifamiliares da cidade, constituindo uma amostra não probabilística por conveniência. A verificação da existência de sazonalidade no consumo de água foi efetuada pela aplicação dos testes ANOVA e Kruskal-Wallis. Com o auxílio do software R, foram realizadas análises estatísticas e concluiu-se que não existe a componente sazonal tanto nas séries temporais de consumo médio de água como nas medianas de consumo. Dessa forma, o consumo de água nas residências unifamiliares avaliadas não é influenciado pelos diferentes meses, estações e épocas do ano.*

**Palavras-chave:** *Sazonalidade; Consumo de água; Joinville.*

### 1. Introdução

A água é um dos recursos vitais mais importantes da Terra e o desenvolvimento de pesquisas nessa área tem sido amplamente impulsionado pelo crescimento populacional, desenvolvimento econômico e mudanças nos padrões de consumo. Fatores como esses irão elevar a demanda mundial de água em 20 a 30% até 2050 [3]. Consequentemente, a sua gestão e uso racional tornaram-se assuntos de relevância na criação de políticas públicas [22].

O abastecimento urbano é responsável por 23,8% da quantidade total de água retirada no Brasil, ficando atrás apenas do setor agrícola [1]. Em razão disso, uma compreensão aprofundada dos mecanismos que modelam o consumo doméstico de água permite a criação e implementação de estratégias, programas e políticas melhor orientados e mais eficazes na conservação desse recurso vital [12].

Analisar o consumo de água em edificações residenciais é primordial na gestão do uso da água em cidades. Acrescenta-se também que, na área predial, o uso racional de água tem papel fundamental na concepção de edifícios sustentáveis [21]. A economia alcançada pelos programas de gestão da demanda de água pode ter um impacto significativo sobre o sistema de abastecimento [20].

Como é difícil modelar uma série temporal em sua totalidade, considerando suas diversas componentes, é útil dividi-la em categorias para melhor entendimento. Tradicionalmente, a série temporal é desmembrada em quatro componentes: tendência, ciclo, sazonalidade e flutuação irregular [14].

Em geral, uma série exige comportamento periódico com períodos quando similaridades na série ocorrem depois de  $s$  intervalos de tempo [2]. Assim, um padrão sazonal ocorre quando uma série temporal é afetada por fatores sazonais como a época do ano ou dias da semana [8].

A análise da componente sazonal em séries temporais já é implementada em muitos estudos relativos à água como previsão da demanda e qualidade da água [5],[4],[18],[23]. Desta forma, este artigo tem como objetivo verificar a existência de sazonalidade no consumo de água em residências unifamiliares de Joinville.

## 2. Método

No presente estudo, adotou-se a edificação unifamiliar (casa e geminado) como unidade de amostragem para a coleta de informações. O procedimento utilizado foi a amostragem não probabilística por conveniência. Amostragem não probabilística compreende a seleção intencional de indivíduos dentro da população a fim de estabelecer uma amostra representativa do universo populacional [9],[10]. A amostra de conveniência é formada por elementos que o pesquisador reuniu por serem acessíveis [7], sendo obtida sem aleatoriedade sistemática [6]. A pesquisa consistiu na coleta de informações de 43 residências unifamiliares distribuídas em 25 bairros da cidade de Joinville, de acordo com a Figura 1.

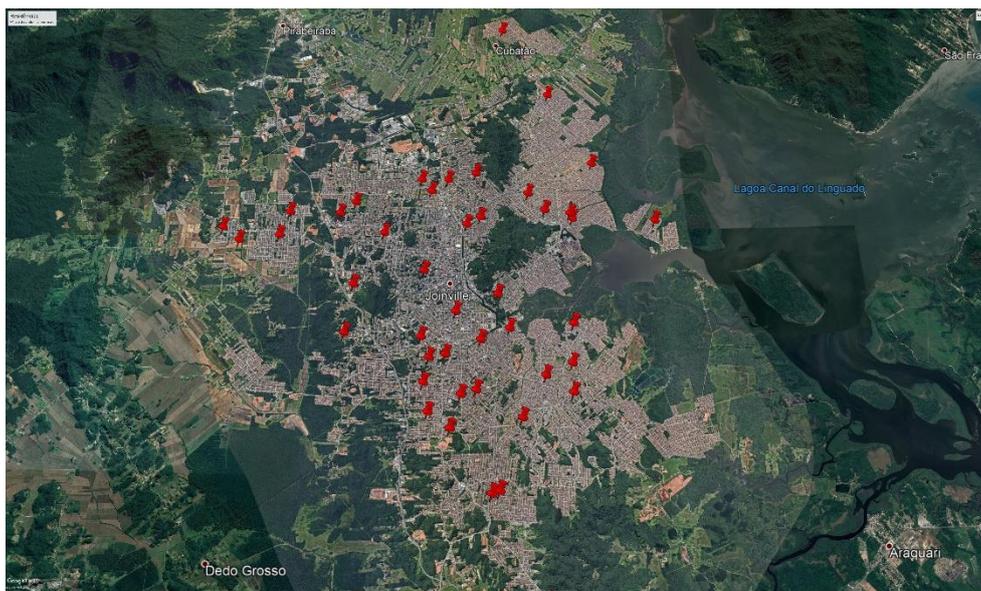


Figura 1: Mapa de Joinville com a localização das residências estudadas

O consumo de água de cada residência foi fornecido pela Companhia Águas de Joinville. A concessionária disponibilizou dados do período de outubro de 2014 até setembro de 2018, totalizando 48 meses de medição. Na análise estatística, em cada mês, foram consideradas a média e mediana do consumo de água das 43 casas, totalizando duas séries temporais. Isto porque pretende-se avaliar qual das duas medidas de tendência central melhor representará o conjunto de dados em análises futuras.

Inicialmente, verificou-se a hipótese de normalidade do consumo de água aplicando o teste Shapiro-Wilk. Foi escolhido esse teste em função do seu poder em avaliar a aderência dos dados à distribuição normal [15].

Os métodos de regressão são eficazes na análise de séries que podem apresentar sazonalidade determinística. O modelo de posto completo é dado por [13]:

$$Z_t = \sum_{i=0}^m \beta_i t^i + \sum_{j=1}^{n_j} \alpha_j D_{jt} + \alpha_t$$

Em que  $Z_t$  representa a variável resposta,  $D_{jt}$  são variáveis periódicas e  $\alpha_j$  e  $\beta_i$  indicam os coeficientes estimados da regressão. Dessa forma, não existe sazonalidade determinística na série temporal se a hipótese nula de que todos os  $\alpha_j$  são nulos não for rejeitada, ou seja, se  $H_0: \alpha_1 = \dots = \alpha_{n_j} = 0$  não for rejeitada [13].

Foram então criados seis modelos de regressão considerando como variável dependente o consumo médio de água ( $y$ ), sendo média ( $y_1$ ) e mediana ( $y_2$ ); e variáveis independentes qualitativas categóricas ( $x$ ): meses do ano, sendo janeiro, fevereiro, ..., dezembro ( $x_1$ ); estações do ano, sendo primavera, verão, outono e inverno ( $x_2$ ) e épocas semestrais do ano, sendo verão e inverno ( $x_3$ ).

Dado o modelo, a suposição de homocedasticidade da variância pode ser verificada pelos testes de Levene e Bartlett. O teste de Bartlett é o mais adequado nos casos em que a distribuição é normal. Já o teste de Levene é o melhor por ter baixa sensibilidade à não normalidade dos dados. Os testes avaliam a hipótese nula de que as variâncias são iguais  $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$  contra a hipótese alternativa  $H_1: \sigma_i^2 \neq \sigma_j^2$  para pelo menos um par  $(i, j)$  [19].

Ademais, dois testes formais para verificar a existência de sazonalidade na série são: teste ANOVA e o teste não-paramétrico de Kruskal-Wallis [13,[16]. No teste não paramétrico de Kruskal-Wallis, assume-se que os dados são provenientes de k amostras independentes  $A_1, A_2, \dots, A_k$  ( $k = 4$  se a série é trimestral;  $k = 12$  se é mensal) de tamanho  $n_1, n_2, \dots, n_k$  respectivamente [11]. O teste é baseado na estatística:

$$W = \frac{12}{n(n+1)} \sum_{j=1}^k \frac{S_j^2}{n_j} - 3(n+1)$$

Em que  $S_j$  representa a soma dos ranks das observações da amostra  $A_j$  dentre a amostra total de  $n = \sum_{j=1}^k n_j$  observações. Essa estatística segue a distribuição qui-quadrado com  $k - 1$  graus de liberdade. O software escolhido para tratamento e análise dos dados foi o R versão 3.5.1, com auxílio da interface *RStudio* [17]. O nível de significância adotado foi 5%.

### 3. Resultados e discussão

Nesta seção, são apresentados os resultados do estudo das séries temporais de consumo de água, além da verificação da presença de sazonalidade. Primeiramente, foi realizado um estudo preliminar observando o comportamento das séries no período de outubro de 2014 a setembro de 2018. Como ilustra a Figura 2, as séries de média e mediana possuem um comportamento estável ao longo dos anos, ou seja, não apresentam tendência reconhecível.

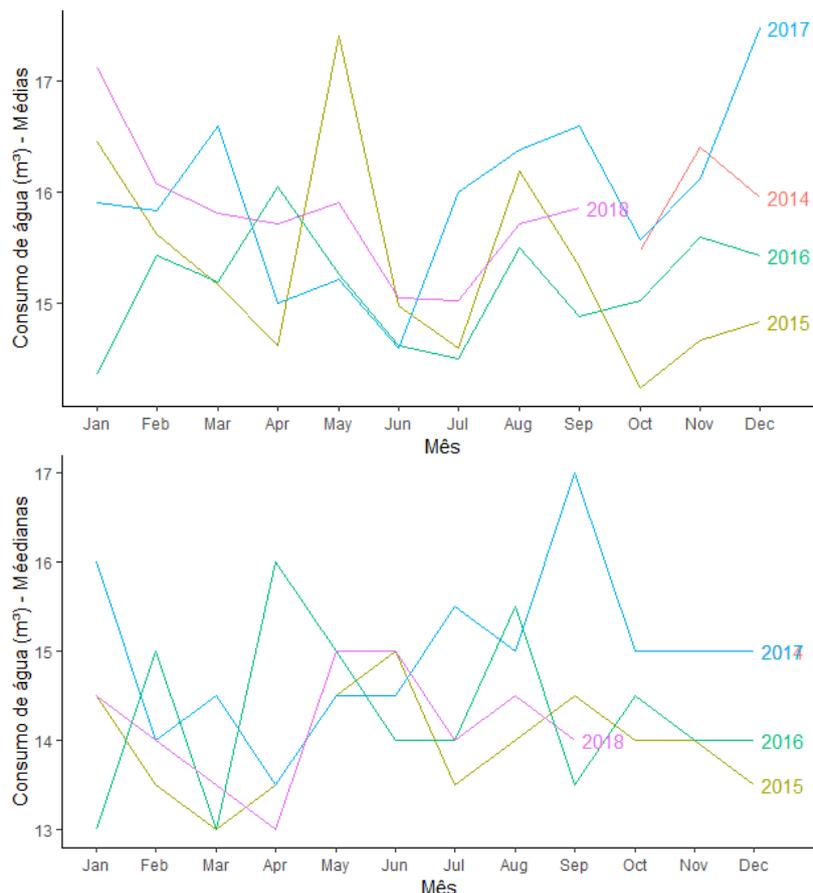


Figura 2: Gráfico de série temporal das médias e medianas do consumo de água (m³)

Em relação às características da amostra, a Tabela 1 apresenta a estatística descritiva das séries temporais de média e mediana do consumo bruto de água, em m<sup>3</sup> por mês.

Variável	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Desvio padrão
Consumo bruto (m <sup>3</sup> /mês) - Médias	14,24	15,02	15,54	15,57	16,01	17,48	0,768
Consumo bruto (m <sup>3</sup> /mês) - Medianas	13,00	14,00	14,50	14,40	15,00	17,00	0,850

Tabela 1: Estatística descritiva da variável dependente consumo bruto

A análise gráfica da Figura 3 revela uma alta variabilidade no consumo bruto de água considerando os diferentes meses do ano. Por exemplo, em janeiro (2015 a 2018), o consumo mediano de água estava próximo de 16,2m<sup>3</sup>. Já nos meses de junho e julho do período considerando, a mediana diminuiu para 14,8m<sup>3</sup> e novamente o consumo voltou a ficar próximo de 16m<sup>3</sup> em agosto, mantendo-se superior a 15m<sup>3</sup> até dezembro.

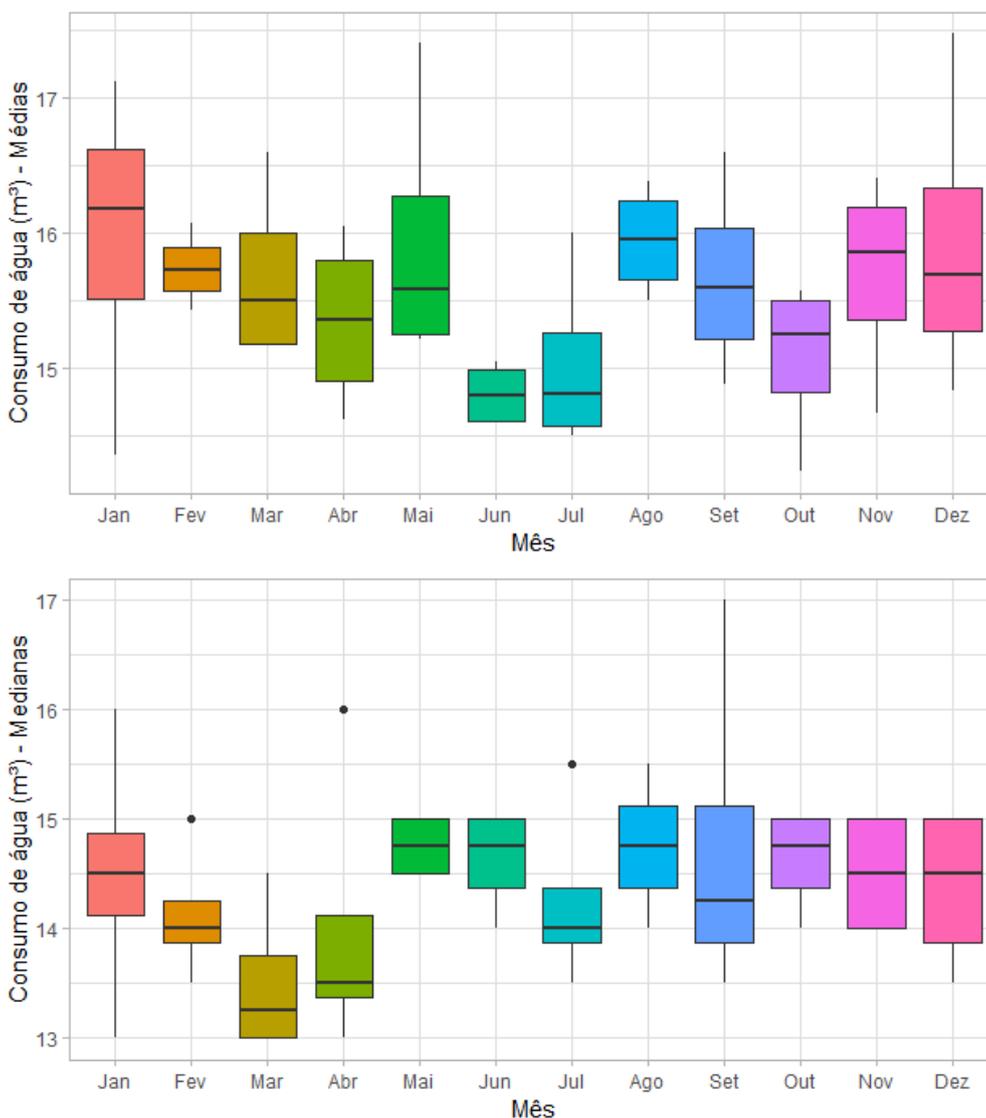


Figura 3: Boxplots do consumo médio e mediano de água (m<sup>3</sup>) em relação aos meses do ano

Com o intuito de averiguar as suposições de normalidade, a Figura 4 apresenta o histograma da variável dependente consumo de água, considerando a média e mediana das 43 residências. Ao se observar os histogramas (Figura 4), percebe-se que as médias do consumo de água apresentam indícios de serem superiores às medianas, possivelmente influenciados por valores altos de consumo. A aplicação do teste de normalidade Shapiro-Wilk relevou que os dados de consumo médio de água atenderam à suposição de normalidade ( $p$ -valor = 0,277).

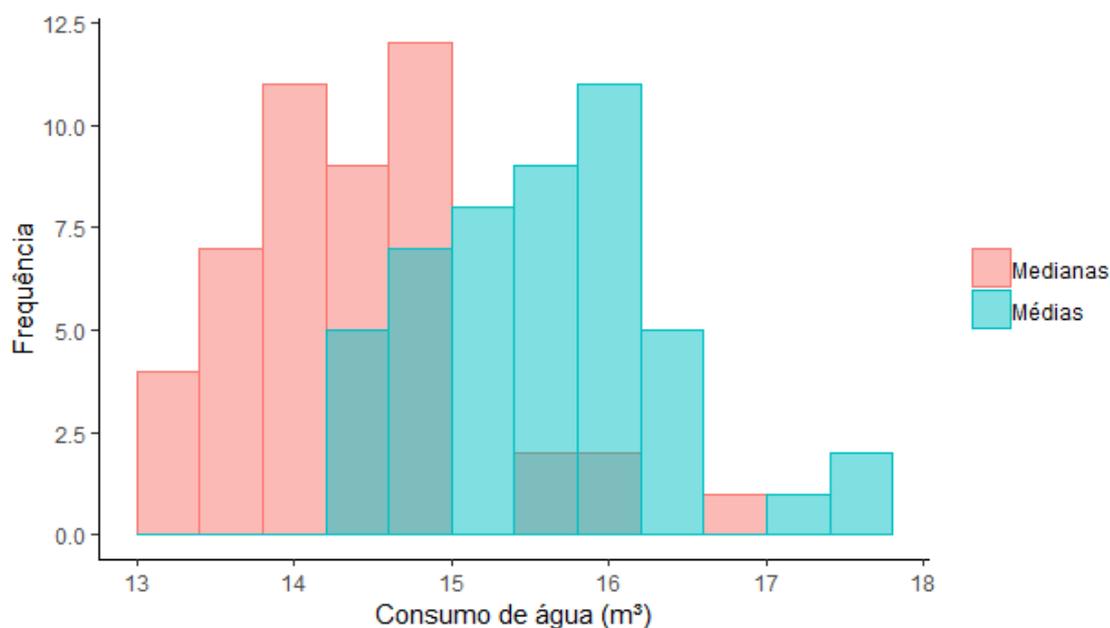


Figura 4: Histograma das médias e medianas do consumo de água (m<sup>3</sup>)

Todavia, ao considerar as medianas do consumo de água, o  $p$ -valor obtido foi de 0,020, demonstrando que a distribuição normal não era adequada para representar a amostra. Por esse motivo, a fim de avaliar a homocedasticidade da variância, foram aplicados o teste paramétrico de Bartlett e seu equivalente não paramétrico de Levene. Como resultado, as probabilidades de significância foram maiores que 0,05, sendo assim as variâncias das populações são iguais e a suposição de homocedasticidade das variâncias foi atendida (Tabela 2).

Utilizou-se também a análise de variância (ANOVA) e o teste não paramétrico de Kruskal-Wallis tendo em vista os seis modelos e o atendimento à suposição de normalidade. De acordo com a Tabela 2, em todas as situações, a probabilidade de significância ( $p$ -valor) resultante foi superior a 0,05, ou seja, não há evidências para rejeitar a hipótese nula de que tanto a série de consumo médio de água como a de medianas não possuem sazonalidade determinística, seja considerando os meses, estações ou semestres do ano.

Variável resposta	Fator	Variância Valor p	Sazonalidade Valor p
Consumo bruto em m <sup>3</sup> /mês Médias (y1)	Meses do ano (x1)	0,3685 <sup>a</sup>	0,3615 <sup>c</sup>
	Estações do ano (x2)	0,8408 <sup>a</sup>	0,2517 <sup>c</sup>
	Épocas semestrais do ano (x3)	0,6876 <sup>a</sup>	0,3370 <sup>c</sup>
Consumo bruto em m <sup>3</sup> /mês Medianas (y2)	Meses do ano (x1)	0,9308 <sup>b</sup>	0,5110 <sup>d</sup>
	Estações do ano (x2)	0,5194 <sup>b</sup>	0,4266 <sup>d</sup>
	Épocas semestrais do ano (x3)	0,7884 <sup>b</sup>	0,6363 <sup>d</sup>

<sup>a</sup> – Bartlett    <sup>b</sup> – Levene    <sup>c</sup> – ANOVA    <sup>d</sup> – Kruskal-Wallis

Tabela 2: Probabilidades de significância da ANOVA e dos testes Bartlett, Levene e Kruskal-Wallis

## 4. Conclusões

A análise de sazonalidade do consumo de água é fundamental uma vez que permite verificar a existência de padrões de consumo de acordo com as diferentes épocas do ano, auxiliando as empresas de abastecimento de água no planejamento adequado da demanda.

Pela análise estatística, o consumo de água nas residências unifamiliares não possui tendência de crescimento ou decréscimo. Além disso, concluiu-se que as séries temporais não apresentaram a componente sazonalidade, ou seja, o consumo de água nas residências unifamiliares analisadas não é influenciado pelos diferentes meses, estações e épocas do ano.

Por fim, essa pesquisa representa um passo inicial no estudo do consumo de água em residências unifamiliares da cidade de Joinville e pode ser útil para governos locais na criação de políticas sustentáveis.

## Referências

- [1] ANA – Agência Nacional de Águas, “Conjuntura dos recursos hídricos no Brasil 2018: relatório pleno”, Brasília: ANA, 2018. 72 p. Disponível em: <<http://arquivos.ana.gov.br/portal/publicacao/Conjuntura2018.pdf>>. Acesso em: 22 mar. 2019.
- [2] Box, George E.P.; Jenkins, Gwilym M.; Reinsel, Gregory C, “Time Series Analysis: forecasting and control”, 3. ed., New Jersey: Prentice-Hall, 1994.
- [3] Burek, Peter; Satoh, Yusuke; Fischer, Günther; Kahil, Taher; Jimenez, Luzma Nava; Scherzer, Angelika; Tramberend, Sylvia; Wada, Yoshihide; Eisner, Stefanie; Flörke, Martina; Hanasaki, Naota; Magnusziewski, Piotr; Cosgrove, William; Wiberg, David, “Water futures and solution: Fast track initiative”, Áustria: IIASA, 2016. Disponível em: <<http://pure.iiasa.ac.at/id/eprint/13008/1/WP-16-006.pdf>>. Acesso em: 18 abr. 2018.
- [4] Coelho, Marcelo; Fernandes, Cristovão Vicente Scapulatempo; Detzel, Daniel Henrique Marco; Mannich, Michael, Statistical validity of water quality time series in urban watersheds, Brazilian Journal of Water Resources, Porto Alegre, vol. 22, n. 51, (2017).
- [5] Comunello, Ana Carla; Daminelli, Lais Marques; Bortolotti, Silvana Ligia Vincenzi, Análise e previsão da demanda de água no município de Medianeira-PR usando o modelo de Holt-Wintersthematic, Revista Eletrônica Científica Inovação e Tecnologia, vol. 2, n. 8, pp. 10-19, (2013).
- [6] Devore, Jay L, “Probabilidade e Estatística para Engenharia e Ciências”, 8. ed., São Paulo: Cengage Learning, 2016.
- [7] Guimarães, Paulo Ricardo Bittencourt, “Métodos quantitativos estatísticos”, Curitiba: IESDE Brasil, 2012.
- [8] Hyndman, Rob J.; Athanasopoulos, George, “Forecasting: principles and practice”, Austrália: Springer, 2018. Disponível em: <<https://otexts.org/fpp2/>>. Acesso em: 04 nov. 2018.
- [9] Kelley, Kate; Clark, Belinda; Brown, Vivienne; Sitzia, John, Good practice in the conduct and reporting of survey research, International Journal for Quality in Health Care, vol. 15, n. 3, pp. 261-266, (2003).
- [10] Kothari, Chakravanti Rajagopalachari, “Research Methodology: Methods and Techniques”, Índia: New Age International, 2004.

- [11] Ladiray, Dominique; Quenneville, Benoit, “Seasonal Adjustment with the X-11 Method”, New York: Springer, 2001.
- [12] Makki, Anas A.; Stewart, Rodney A.; Beal, Cara D.; Panuwatwanich, Kriengsak, Novel bottom-up urban water demand forecasting model: Revealing the determinants, drivers and predictors of residential indoor end-use consumption, Resources, Conservation and Recycling, vol. 95, pp. 15-37, (2015).
- [13] Morettin, Pedro A.; Toloí, Clélia Maria de Castro, ”Análise de séries temporais” 2. ed., São Paulo: Blucher, 2006.
- [14] Nguyen, Linh; Novák, Vilém. Forecasting seasonal time series based on fuzzy techniques. Fuzzy Sets and Systems, vol. 361, pp. 114-129, (2019).
- [15] Öztuna, Derya; Elhan, Atilla Halil; Tüccar, Ersöz, Investigation of four different normality test in terms of type 1 error rate and power under different distributions, Turkish Journal of Medical Sciences, vol. 36, n. 3, pp. 171-176, (2006).
- [16] Portal Action, “Testes para sazonalidade determinística”, São Carlos: Estatcamp, 2019. Disponível em: <<http://www.portalaction.com.br/series-temporais/24-testes-para-sazonalidade-deterministica>>. Acesso em: 26 mar. 2019.
- [17] R Core Team, “R: A language and environment for statistical computing and graphics”, Version 3.5.1, Nova Zelândia: The R Foundation, 2018. Disponível em: <<https://www.r-project.org/>>. Acesso em: 18 abr. 2018.
- [18] Sabino, Claudia Vilhena Schayer; Lage, Ludmila Vieira; Noronha, César Vinicius, Variação sazonal e temporal da qualidade das águas em um ponto do Córrego Gameleiras usando técnicas quimiométricas robustas, Engenharia Sanitária e Ambiental, vol. 22, n. 5, pp. 969-983, (2017).
- [19] Tufféry, Stéphane, “Data mining and statistics for decision making”. United Kingdom: John Wiley & Sons, 2011.
- [20] Willis, Rachelle M.; Stewart, Rodney A.; Giurco, Damien P.; Talebpour, Mohammad R.; Mousavinejad, Alireza, End use water consumption in households: impacto of socio-demographic factors and efficient devices, Journal of Cleaner Production, vol. 60, pp. 107-115, (2013).
- [21] Wong, L. T.; Mui, K. W., Epistemic water consumption benchmarks for residential buildings, Building and Environment, vol. 43, n. 6, pp. 1031-1035, (2006).
- [22] WWAP - World Water Assessment Programme, “The United Nations World Water Development Report 2018: Nature-based solutions for water”, Paris: UNESCO, 2018. 139 p. Disponível em: <<http://unesdoc.unesco.org/images/0026/002614/261424e.pdf>>. Acesso em: 06 abr. 2018.
- [23] Xu, Guoce; Li, Peng; Lu, Kexin; Tantai, Zhan; Zhang, Jiabin; Ren, Zongping; Wang, Xiukang; Yu, Kunxia; Shi, Peng; Cheng, Yuting, Seasonal changes in water quality and its main influencing factors in the Dan River basin. Catena, vol. 173, pp. 131-140, (2019).



## Uma avaliação dos métodos semi-empíricos de cálculo de pressão de falha de dutos corroídos

**Thiago M. Fernandes, João E. Abdalla Filho**

Departamento Acadêmico de Construção Civil, UTFPR

81280-340, Curitiba, PR

E-mail: tfernandes@alunos.utfpr.edu.br, joaofilho@utfpr.edu.br

**Resumo:** A corrosão é um dos maiores problemas causadores de falhas e estruturas de oleodutos e gasodutos. Métodos analíticos e computacionais foram desenvolvidos para calcular a pressão de ruptura de dutos sujeitos à corrosão externa. Este trabalho investiga o conservadorismo dos métodos semi-empíricos de cálculo de pressão de falha de dutos com defeitos curtos de corrosão. Modelos numéricos com defeitos de formato retangular e arredondado, baseados no método dos elementos finitos, são desenvolvidos e comparados com testes experimentais. Uma série de defeitos são modelados por meio da variação do comprimento do defeito com relações fixas de profundidade. É considerada a não linearidade do material para representar o seu comportamento constitutivo.

### 1 Introdução

Os dutos são importantes meios de transporte de materiais fluidos e sólidos e encontram-se, muitas vezes em locais de difícil acesso como em montanhas ou em regiões de vegetação densa. Considerado um dos principais defeitos em estruturas de oleodutos e gasodutos, a corrosão pode levar ao colapso da peça, provocando enormes desastres ambientais e sociais, além de perdas financeiras. A corrosão é uma forma de degradação que reduz a resistência da estrutura. Pode ocorrer nas superfícies internas e externas do tubo ou na emenda da solda. Neste caso, a redução de material na seção transversal resulta na perda de eficiência operacional e de segurança da tubulação [3].

Alguns métodos semi-empíricos como o ASME B31G [2], 0,85dL [4], proposto por Kiefner e Vieth, DNV RP-F101 [6] e PCORRC [4], foram desenvolvidos para calcular a pressão de falha de dutos com defeitos de redução de espessura da parede. Tais procedimentos, são métodos de solução semi-empírica que derivam de princípios de mecânica dos sólidos, experimentos físicos até a ruptura, e também via simulações por elementos finitos. Procedimentos analíticos são fáceis de usar e são rotineiramente aplicados por engenheiros, a fim de decidir se o defeito de corrosão é crítico [1]. A análise por simulações computacionais via elementos finitos também tem sido usada como uma ferramenta para desenvolver soluções mais precisas da capacidade de carga para dutos com defeitos de corrosão. Em muitos casos, a solução numérica pode ser mais simples que através de métodos analíticos.

### 2 Métodos semi-empíricos

Os métodos foram desenvolvidos por meio da simplificação dos defeitos de corrosão. Ou seja, são considerados superfícies de geometrias retangulares ou parabólicas, desprezando a rugosidade do defeito. A expressão geral para cálculo da pressão de falha de defeitos com redução da espessura da parede do duto é definida em (1), (2), (3), (4):

$$P_f = P_i \left[ \frac{1 - \alpha \frac{A}{A_0}}{1 - \frac{A}{A_0} M^{-1}} \right] \quad (1)$$

$$A = L \cdot d \quad (2)$$

$$A_0 = L \cdot t \quad (3)$$

$$P_i = \frac{\sigma_y \cdot 2t}{D} \quad (4)$$

A visão geral esquemática do duto com defeito submetido à pressão interna é apresentada na Figura 1.

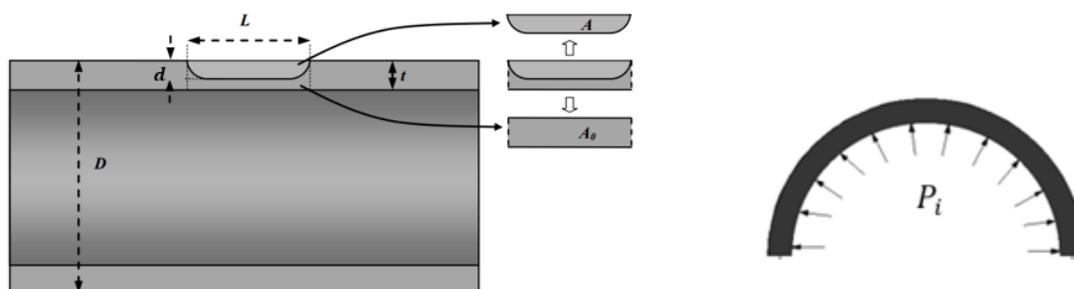


Figura 1: Visão esquemática geral do duto com defeito submetido à pressão interna.

Onde  $P_f$  é a pressão de falha do duto corroído;  $P_i$  é a pressão interna do duto;  $A$  é a área de projeção do defeito em um plano longitudinal através da espessura da parede;  $A_0$  é a área longitudinal do duto original na seção do defeito;  $A_0$  é o fator de abaulamento da folia, referente ao comprimento do defeito;  $L$  é comprimento do defeito;  $d$  é a profundidade máxima do defeito;  $t$  é a espessura da parede do duto;  $\alpha$  é um fator empírico referente ao formato da profundidade defeito ( $\alpha=1$  para formato retangular,  $\alpha=2/3$  para formato parabólico,  $\alpha=0,85$  para de formato intermediário entre retangular e parabólico;  $\sigma_y$  é a tensão de escoamento do material e  $D$  é o diâmetro externo do duto.

## 2.1 Método ASME B31G

Por meio de ensaios experimentais, estabeleceu uma área projetada da corrosão em formato parabólico escrito no defeito, assim como considera a tensão de escoamento,  $\sigma_y$ , em sua formulação.

O método estabelece o limite para defeitos curtos de acordo com (5) e o cálculo da pressão de falha em (6) e (7):

$$L \leq \sqrt{20Dt} \quad (5)$$

$$P_f = 1,1 \times \sigma_y \times \frac{2t}{D} \times \frac{1 - \frac{2}{3} \left( \frac{d}{t} \right)}{1 - \frac{2}{3} \left( \frac{d}{t} \right) M^{-1}} \quad (6)$$

$$M = \sqrt{1 + 0,8 \left( \frac{L^2}{Dt} \right)} \quad (7)$$

## 2.2 Método B31G Modificado ou 0,85 dL

Devido ao conservadorismo do B31G, algumas modificações foram introduzidas na equação original. O método 0,85 dL considera o formato intermediário entre retangular e parabólico para o defeito. Assim como o método ASME B31G, o 0,85dL também considera a tensão de escoamento,  $\sigma_y$ , em sua formulação.

Este método altera o limite para defeitos curtos em (8) e a pressão de falha é calculado por (9) e (10):

$$L \leq \sqrt{50Dt} \quad (8)$$

$$P_f = (\sigma_y + 68,95) \times \frac{2t}{D} \times \frac{1 - 0,85 \left(\frac{d}{t}\right)}{1 - 0,85 \left(\frac{d}{t}\right) M^{-1}} \quad (9)$$

$$M = \sqrt{1 + 0,6275 \left(\frac{L^2}{Dt}\right) - 0,003375 \left(\frac{L^2}{Dt}\right)^2} \quad (10)$$

## 2.3 Método DNV RP-F101

Este método considera o defeito de formato retangular e a tensão última,  $\sigma_u$ , como controladora da pressão de falha. Para qualquer tipo de defeito de corrosão, a pressão de falha é determinada de acordo com (11) e (12):

$$P_f = 1,05 \times \sigma_u \times \frac{2t}{D-t} \times \frac{1 - \left(\frac{d}{t}\right)}{1 - \left(\frac{d}{t}\right) M^{-1}} \quad (11)$$

$$M = \sqrt{1 + 0,31 \left(\frac{L^2}{Dt}\right)} \quad (12)$$

## 2.4 Método PCORRC

Formulação baseada em elementos finitos, o PCORRC é um método analítico que, assim como o método DNV RP-F101, considera a tensão última,  $\sigma_u$ , como determinante na pressão de falha em dutos com defeitos de corrosão.

Para qualquer tipo de defeito, a pressão de falha é determinada por (13):

$$P_f = \sigma_u \times \frac{2t}{D} \times \left\{ 1 - \frac{d}{t} \left[ 1 - \exp \left( -0,157 \frac{L}{\sqrt{\frac{D}{2}(t-d)}} \right) \right] \right\} \quad (13)$$

## 3 Metodologia

### 3.1 Material

Em 2003, Choi et. al. [5] publicaram um artigo baseado em uma série de testes de ruptura para dutos corroídos de diferentes geometrias feitos de aço API5L-X65 com as seguintes

características:  $D = 762$  mm,  $t = 17,5$  mm, módulo de elasticidade  $E = 203$  GPa; coeficiente de Poisson  $\nu = 0,3$ ,  $\sigma_y = 468$  MPa,  $\sigma_u = 673$  MPa. O mecanismo de falha foi controlado pelo colapso plástico em todos os testes. No presente trabalho, é utilizado o mesmo material, API5L-X65, utilizado por Choi et. al. É considerada a não linearidade do material para representar o seu comportamento constitutivo. A Figura 2 apresenta a curva Tensão x Deformação do aço X65.

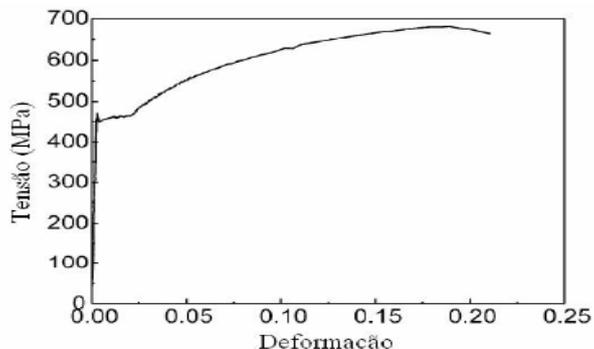


Figura 2: Curva Tensão x Deformação do aço X65 [9].

### 3.2 Geometria do defeito

A partir da tensão efetiva de Von Mises, Choi et. al. [5] observou que para defeitos de profundidades de formatos retangulares, a tensão de referência foi determinada como sendo 90% da tensão última do material.

Krujv [7] analisou numericamente defeitos de cantos retangulares e arredondados e observou uma melhor aproximação dos resultados com defeitos de canto arredondado. Neste trabalho, é modelado defeitos de profundidades retangulares. Em relação à superfície do defeito, é comparado os modelos de canto retangular e de canto arredondado, tendo como referência os ensaios realizados por Choi et. al. Os formatos superficiais do defeito são apresentados na Figura 3. As análises numéricas são realizadas por meio do software ANSYS, versão 19.1.

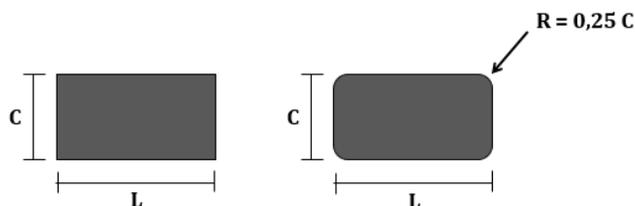


Figura 3: Formato superficial de canto retangular e canto arredondado.

A Tabela 1 apresenta a relação entre a pressão de falha dos modelos de canto retangular e de canto arredondado com os ensaios de Choi et.al. Os resultados mostram uma melhor aproximação entre as soluções numéricas com modelos de cantos arredondados.

Ensaio experimental		Geometria da corrosão			$P_{MEF}/P_{ENS}$	
Espécime	Pressão do ensaio (MPa)	l (mm)	c (mm)	d (mm)	Formato retangular	Formato arredondado
DA	24,11	200	50	4,4 (25%)	1,04	1,03
DB	21,76	200	50	8,8 (50%)	0,99	1,00
DC	17,15	200	50	13,1 (75%)	0,94	0,95
LA	24,3	100	50	8,8 (50%)	0,98	0,99
LC	19,8	300	50	8,8 (50%)	0,98	0,99
CB	23,42	200	100	8,8 (50%)	0,93	0,94
CC	22,64	200	200	8,8 (50%)	0,94	0,94

Tabela 1: Comparação da solução numérica com os ensaios experimentais.

### 3.3 Discretização do modelo

Há dois planos de simetria no modelo. Os planos de simetria representam uma simplificação da malha, que reduz o esforço computacional e diminui o tempo de solução do problema. Na discretização do modelo do duto, é utilizado o elemento de casca SHELL181, que tem como característica quatro nós com seis graus de liberdade em cada nó. Os elementos de casca são bidimensionais, porém aceitam a inclusão de uma espessura. Na região do defeito, a malha é mais refinada. Ao longo do duto íntegro, os elementos variam de tamanho e não necessitam ser muito refinados conforme vai se distanciando do defeito. A Figura 4 apresenta a discretização do modelo e suas condições de contorno.

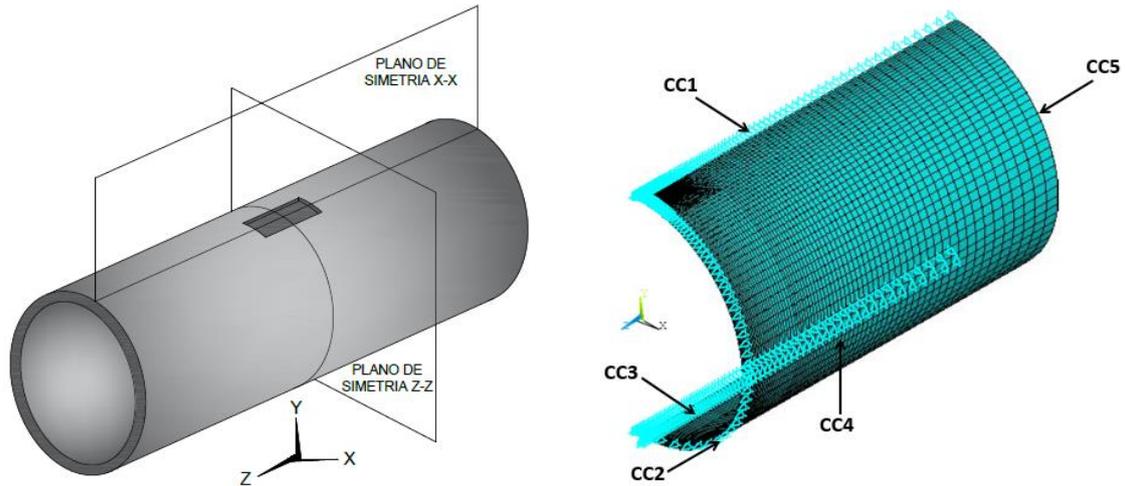
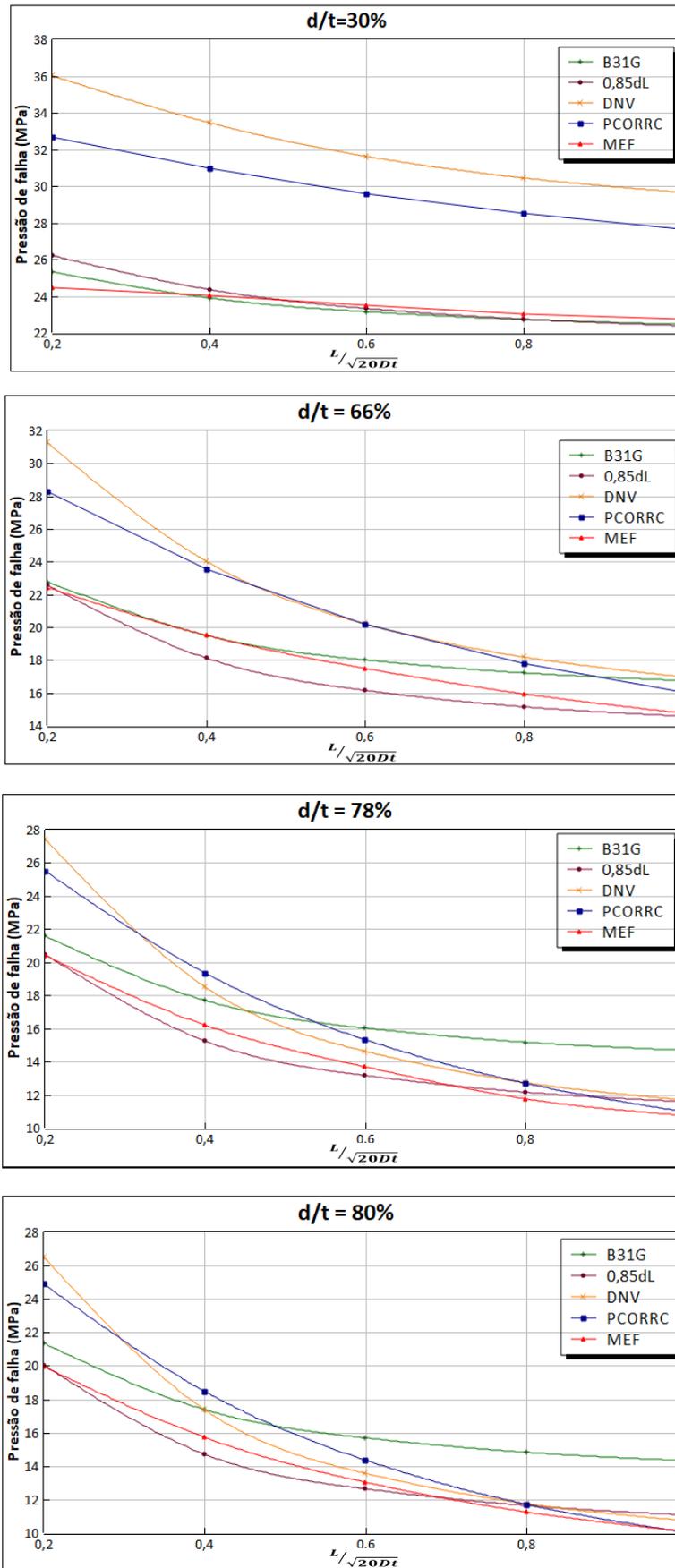


Figura 4: Discretização do modelo e condições de contorno.

- CC1 e CC3 – Pela condição do plano de simetria, estas duas linhas são restringidas na direção X.
- CC2 – Pela condição do plano de simetria, esta linha é restringida na direção Z.
- CC4 – Para não haver deslocamento de corpo rígido, esta linha é restringida na direção Y.
- CC5 – Uma pressão uniforme linear é aplicada para simular o efeito de um tampo físico. Esta pressão vale  $-177,47784 P_i$ , onde  $P_i$  é a pressão interna.

### 4 Resultados e Discussões

O material de estudo, duto da classe API-X65, cujas propriedades são módulo de Young  $E = 203$  GPa; coeficiente de Poisson  $\nu = 0,3$ ; tensão de escoamento  $\sigma_y = 468$  MPa e tensão última  $\sigma_u = 673$  MPa, baseadas no ensaio de Choi et. al. Dutos com diâmetro de 323,9 mm e parede com espessura de 8,4 mm são considerados. A pressão de falha é analisada a partir de uma relação  $d/t$  fixa, variando o comprimento em um intervalo de defeitos curtos de relação  $\frac{L}{\sqrt{20Dt}}$  variando de 0,2 a 1. Tal relação é considerada, pois abrange todo o limite para defeitos curtos. São estabelecidas as relações  $d/t$  iguais a 30%, 66%, 78% e 80%, apresentadas nas Figuras 4-7.



Figuras 5-8: Pressão de falha para as relações d/t de 30%, 66%, 78% e 80% variando o comprimento do defeito

De acordo com a Figura 5, para defeitos com profundidade de 30% em relação à espessura do duto, o método DNV RP-F101 e o método PCORRC são pouco conservadores, para qualquer comprimento de defeito. Já o método B31G e o método 0,85dL, apresentam valores próximos aos resultados numéricos para os diferentes comprimentos de defeito, sendo são mais seguros para defeitos de baixa profundidade.

Para defeitos com profundidade de 66%, a Figura 6 mostra que os métodos DNV RP-F101 e PCORRC continuam sendo pouco conservadores para defeitos de pequeno comprimento, no entanto, ao aumentar esse comprimento, se tornam mais conservadores que o método B31G. O método 0,85dL se apresenta com valores mais próximos aos resultados numéricos para qualquer comprimento do defeito.

Para as profundidades de 78% e 80%, as Figuras 7 e 8 mostram que o método DNV RP-F101 e o método PCORRC continuam sendo pouco conservadores para defeitos de pequeno comprimento, mas se tornam mais conservadores que os métodos B31G e 0,85dL com o aumento desse comprimento. O método B31G assume um patamar pouco conservador, com pressão de falha relativa muito maior que os outros métodos com o aumento do comprimento do defeito.

## 5 Conclusão

Este trabalho investiga por meio de análise computacional métodos semi-empíricos de cálculo de pressão de falha em dutos sujeitos à corrosão externa. Os métodos B31G e 0,85dL são conservadores para defeitos curtos e rasos, porém, para defeitos profundos e compridos, são contra a segurança. Já os métodos DNV RP-F101 e o PCORRC, são pouco conservadores para defeitos curtos e rasos, porém, se apresentam mais conservadores com o aumento da profundidade e comprimento do defeito. A análise numérica de defeitos curtos pelos métodos B31G; 0,85dL; DNV RP-F101 e PCORRC permite concluir que cada método apresenta um conservadorismo em função do comprimento e profundidade do defeito, ou seja, não é possível afirmar qual é o método semi-empírico mais conservador sem analisar a geometria da corrosão para defeitos curtos.

## Referências

- [1] Abdalla Filho, J. E.; Machado, R. D.; Bertin, R. J.; Valentini, M. D. On the failure pressure of pipelines containing wall reduction wall reduction and isolated pit corrosion defects. *Computers & Structures*, vol. 132, p.22-33, 2014.
- [2] ASME - American Society of Mechanical Engineers. Manual for Determining the Remaining Strength of Corroded Pipelines. ASME B31G, New York, 1991.
- [3] Basso, A. V.; Abdalla Filho, J. E.; Shang, H. Y. Assessment of DNV-RP-F101 method in estimating the failure pressure in corroded pipelines. *In: 23rd ABCM International Congress of Mechanical Engineering*. Rio de Janeiro, 2015.
- [4] Bertin, R. J. A numerical investigation of internal failure pressure of pipelines containing a single and double pit corrosion defect, considering plasticity. Tese (Doutorado em Engenharia Mecânica) - Pontifícia Universidade Católica do Paraná, Curitiba, 2015.
- [5] Choi, J. B.; Goo, B. K.; Kim, J. C. et. al. Development of limit load solution for corroded gas pipelines. *International Journal of Pressure Vessels and Piping*, Kyonggi-do, v.80, p.121-128, 2003.
- [6] DET NORSKE VERITAS. Recommended Practice DNVGL-RP-F101 - corroded pipelines. DNV-GL, 2017.
- [7] KUJV, J. F. R. Determinação da pressão de falha de dutos corroídos através do método dos elementos finitos. Trabalho de Conclusão de Curso (Graduação em Engenharia Civil) - Pontifícia Universidade Católica do Paraná, Curitiba, 2013.



## Controle do caos em sistemas dinâmicos. Aplicação à inicialização e previsão de nuvens em um modelo climático.

**Paul Krause,**

UFSC - Departamento de Matemática  
Campus Trindade, Florianópolis, SC, 88040-900  
E-mail: p.krause@ufsc.br

**Joseph Tribbia**

NCAR - Division of Climate and Global Dynamics  
1850 Table Mesa Drive, Boulder, CO, 80305, USA  
E-mail: tribbia@ucar.edu

**Resumo:** *O Influence Sampling (IS) é um método de controle do caos para variáveis não-observáveis de sistemas não-autônomos, como no caso da taxa de precipitação em modelos climáticos. Mostramos que o IS provê inicialização robusta e estabilização dinâmica da umidade e da taxa de precipitação em ciclos de assimilação de dados de um modelo climático aqui apresentado nos quais as componentes horizontais da velocidade de grade e a temperatura potencial são escolhidos como variáveis controladoras em cada ciclo.*

**Palavras-chave:** *caos, controle, inicialização, previsão*

### 1 Introdução

Nos modelos climáticos, as equações de subgrade geram *stiffness* no método numérico se não forem adequadas à escala das equações de grade [4]. Na seção 2, apresentamos um sistema de equações mesoescalares para a velocidade de subgrade em um modelo atmosférico climático. A dedução destas equações segue a linha de pensamento da Termodinâmica Irreversível Estendida (TIE) para processos de não-equilíbrio, na qual fluxos dissipativos são tratados como variáveis substituindo as relações constitutivas por equações de evolução. O fechamento de nossas equações, no entanto, pressupõe que a lei de equilíbrio para a produção de energia vale na meso-escala atmosférica. Neste artigo apenas apresentamos o modelo. Os detalhes de sua dedução serão apresentados em outro artigo.

Também nos modelos climáticos, a taxa de precipitação é uma variável não-observável e dinamicamente expansiva no sentido de [1]. Toda informação a respeito de sua distribuição na atmosfera em um dado instante se perde rapidamente. Em [6] foi apresentado um método de controle do caos para variáveis deste tipo: o Influence Sampling (IS). O método admite análise esparsa, amostra de tamanho 1 e paralelização. Na seção 3, o IS é aplicado à inicialização e previsão da taxa de precipitação com o modelo climático da seção 2. Neste modelo, a taxa de precipitação é controlável pela velocidade horizontal de grade e a temperatura potencial de grade em um sentido mais amplo que aquele sugerido em [7] e [2]. Por isso a noção de variável controlável é re-definida no início da seção.

## 2 Modelo atmosférico climático

Nesta seção apresentamos um modelo atmosférico climático composto pelas equações anelásticas de grade e um novo sistema de equações mesoescalares para a velocidade de subgrade. A dedução das equações da velocidade de subgrade será apresentada em outro artigo.

Equações anelásticas de grade:

$$\begin{aligned} \nabla^2 p_1 = & \partial_x[-\rho_0(\bar{u} + \check{u})\partial_x\bar{u} - \rho_0(\bar{v} + \check{v})\partial_y\bar{u} - \rho_0(\bar{w} + \check{w})\partial_z\bar{u} + 2\rho_0(\bar{v}\Omega_3 - \bar{w}\Omega_2) + \mu_0\partial_{zz}\bar{u}] \\ & + \partial_y[-\rho_0(\bar{u} + \check{u})\partial_x\bar{v} - \rho_0(\bar{v} + \check{v})\partial_y\bar{v} - \rho_0(\bar{w} + \check{w})\partial_z\bar{v} - 2\rho_0\bar{u}\Omega_3 + \mu_0\partial_{zz}\bar{v}] \\ & + \partial_z[-\rho_0(\bar{u} + \check{u})\partial_x\bar{w} - \rho_0(\bar{v} + \check{v})\partial_y\bar{w} - \rho_0(\bar{w} + \check{w})\partial_z\bar{w} + 2\rho_0\bar{u}\Omega_2 + \rho_0\frac{\theta_1}{\theta_0}g + \mu_0\partial_{zz}\bar{w}] \end{aligned} \quad (1)$$

$$\partial_t\bar{\mathbf{u}} + (\bar{\mathbf{u}} + \check{\mathbf{u}}) \cdot \nabla\bar{\mathbf{u}} = \bar{\mathbf{u}} \times 2\Omega + \frac{\theta_1}{\theta_0}g\mathbf{k} - \frac{1}{\rho_0}\nabla(p_1) + \nu\nabla^2\bar{\mathbf{u}} \quad (2)$$

$$\partial_t\theta_1 + (\bar{\mathbf{u}} + \check{\mathbf{u}}) \cdot \nabla\theta_1 + (\bar{w} + \check{w})\theta'_0 = s_c - s_r - k(\theta_1 - \theta_1^0) + \nu\nabla^2\theta_1 \quad (3)$$

$$\partial_t\bar{q} + (\bar{\mathbf{u}} + \check{\mathbf{u}}) \cdot \nabla\bar{q} = -s_p + \vartheta\nabla^2\bar{q} \quad (4)$$

Equações da velocidade de subgrade:

$$\begin{aligned} \nabla^2\phi = & \partial_x[2\bar{\rho}(\check{v}\Omega_3 - \check{w}\Omega_2) + \xi\partial_{zz}(\bar{u} + \check{u})] + \partial_y[-2\bar{\rho}\check{u}\Omega_3 + \xi\partial_{zz}(\bar{v} + \check{v})] \\ & + \partial_z[2\bar{\rho}\check{u}\Omega_2 + \frac{\rho_1^2}{\rho_0}g + \xi\partial_{zz}(\bar{w} + \check{w})] \end{aligned} \quad (5)$$

$$\partial_t(\bar{\rho}\check{\mathbf{u}}) + \partial_j(\bar{u}_j\bar{\rho}\check{\mathbf{u}} - \check{u}_j\bar{\rho}\bar{\mathbf{u}}) = \bar{\rho}\check{\mathbf{u}} \times 2\Omega + \frac{\rho_1^2}{\rho_0}g\mathbf{k} + \xi\nabla^2\bar{\mathbf{u}} - \nabla\phi + \xi\nabla^2\check{\mathbf{u}} \quad (6)$$

$$\rho_1 = (p_0 + p_1)/(\mathbf{R}\bar{\mathbf{T}}_v) - \rho_0 = \rho_0(\mathbf{T}_0/\bar{\mathbf{T}}_v - 1) + p_1/(\mathbf{R}\bar{\mathbf{T}}_v) \quad (7)$$

com

Domínio:

$$(x, y, z) \in [-\mathbf{L}_x, \mathbf{L}_x] \times [-\mathbf{L}_y, \mathbf{L}_y] \times [0, \mathbf{L}_z] \quad (8)$$

Condições de contorno:

- laterais:

$$\text{periodicidade em } x = \pm\mathbf{L}_x \text{ e } y = \pm\mathbf{L}_y \quad (9)$$

- superfície:

$$\partial_z p_1 = 2\rho_0\bar{u}\Omega_2 + \rho_0\frac{\theta_1}{\theta_0}g + \mu_0\partial_{zz}\bar{w} \text{ em } z = 0 \quad (10)$$

$$\partial_z\phi = 2\bar{\rho}\check{u}\Omega_2 + \frac{\rho_1^2}{\rho_0}g + \xi\partial_{zz}(\bar{w} + \check{w}) \text{ em } z = 0 \quad (11)$$

$$\partial_z a = c_a \sqrt{\bar{u}^2 + \bar{v}^2} (a - a_g) \text{ para } a = \bar{u}, \bar{v}, \check{u}, \check{v} \text{ em } z = 0. \text{ Aqui } a_g \text{ recebe o valor 0 para todo } a. \quad (12)$$

$$\partial_z b = c_b \sqrt{\bar{u}^2 + \bar{v}^2} (b - b_g) \text{ para } b = \theta_1, \bar{q} \text{ em } z = 0, \text{ onde } b_g = b_g(x, y, t) \text{ é o ciclo sazonal de } b \text{ no chão (especificado).} \quad (13)$$

$$\bar{w} = 0 \text{ em } z = 0 \quad (14)$$

$$\check{w} = 0 \text{ em } z = 0 \quad (15)$$

- topo:

$$\partial_z p_1 = 2\rho_0\bar{u}\Omega_2 + \mu_0\partial_{zz}\bar{w} \text{ em } z = \mathbf{L}_z \quad (16)$$

$$\partial_z\phi = 2\bar{\rho}\check{u}\Omega_2 + \frac{\rho_1^2}{\rho_0}g + \xi\partial_{zz}(\bar{w} + \check{w}) \text{ em } z = \mathbf{L}_z \quad (17)$$

$$\theta_1 = 0 \text{ em } z = \mathbf{L}_z \quad (18)$$

$$\bar{w} = 0 \text{ em } z = \mathbf{L}_z \quad (19)$$

$$\check{w} = 0 \text{ em } z = \mathbf{L}_z \quad (20)$$

onde

$$\bullet s_c = (c/(p_0 + p_1))^\kappa s_p \mathbf{L}_v/c_p \text{ é o calor de condensação.} \quad (21)$$

- $s_r$  é o calor de radiação de equilíbrio térmico (corpo negro) superposto com  $(c/(p_0 + p_1))^\kappa(az + b)$  onde houver precipitação. (22)

- $s_p = (\bar{w} + \check{w})(q_s g)((\rho_0 + \rho_1)\bar{T}/(p_0 + p_1))(L_v R - c_p R_v \bar{T})/(c_p R_v \bar{T}^2 + q_s L_v^2)$  é a taxa de precipitação. (23)

- $-k(\theta_1 - \theta_1^0)$  é um termo desestabilizador a ser usado quando as equações estiverem sendo resolvidas em baixa resolução. (24)

- $\bar{T}_v = ((p_0 + p_1)/c)^\kappa(\theta_0 + \theta_1)(1 + 0.6\bar{q})$  if  $\bar{w} + \check{w} > 0$  e  $\bar{q} \geq q_s$ , caso contrário  $s_p = 0$ . (25)

- $\bar{T} = ((p_0 + p_1)/c)^\kappa(\theta_0 + \theta_1)$  (26)

- $T_0 = p_0/(R\rho_0) = p_{00}/(R\rho_{00})$  é uma constante. (27)

- $p_0 = p_{00} \exp(-z/H)$  com  $p_{00} = 10^5$  Pa (28)

- $\rho_0 = \rho_{00} \exp(-z/H)$  com  $\rho_{00} = 1.2$  kg/m<sup>3</sup> (29)

- $\theta_0 = T_0(c/p_0)^\kappa$  (30)

- $\theta'_0 = (\kappa/H)\theta_0$  é a derivada de  $\theta_0(z)$ . (31)

- $H = RT_0/g$  (32)

- $\phi$  é a pressão de subgrade obtida por decomposição de Helmholtz. (33)

- $\Omega = (0, \Omega|\cos(2\pi y/L_y)|, \Omega\sin(2\pi y/L_y))$  (34)

- $\nu = \mu_0/\rho_0(z)$  onde  $\mu_0$  é uma constante. (35)

### 3 Inicialização e previsão de nuvens

O Influence Sampling (IS) é um método de controle do caos para variáveis não-observáveis de sistemas não-autônomos. Sua dedução, apresentada em [6], usa o método de características e a fórmula de Dyson para a equação do transporte linear. A fórmula de Dyson divide a ação do operador de evolução da equação. Em suma, o IS usa o lado esquerdo da aproximação de curto prazo

$$\tilde{\Phi}_t(\tilde{x} + d\tilde{x}, \hat{x}) = \tilde{\Phi}_t(\tilde{x}, \hat{x}_0) + \int_{t_0}^t [\partial_{\tilde{x}} \tilde{\Phi}_s(\tilde{x}, \hat{x}) - \tilde{I}] \tilde{f}'(\Phi_s(\tilde{x}, \hat{x}_0)) ds \quad (36)$$

com

$$d\tilde{x}(t_0, t) = \int_{t_0}^t \tilde{f}'(\Phi_s(\tilde{x}, \hat{x}_0)) ds, \quad (37)$$

onde

$$\tilde{f}'(\Phi_s(\tilde{x}, \hat{x}_0)) = \tilde{f}(\Phi_s(\tilde{x}, \hat{x}_0)) - \tilde{f}(\tilde{\Phi}_s(\tilde{x}, \hat{x}_0), \hat{x}), \quad (38)$$

para perturbar as variáveis controláveis a cada passo de uma integração que explora as direções instáveis destas variáveis em instantes futuros. Nestas expressões,  $f = (\tilde{f}, \hat{f})$  é o campo de um sistema de equações diferenciais ordinárias,  $\Phi = (\tilde{\Phi}, \hat{\Phi})$  é a dinâmica desse sistema,  $x = (\tilde{x}, \hat{x})$  é um valor amostral do estado do sistema no instante de análise  $t_0$ ,  $\hat{x}_0$  é o valor da assimilação de dados fornecido no instante  $t_0$  e  $\tilde{I}$  é a matriz identidade. O campo  $f$  pode depender do tempo, o que foi omitido na notação.

Nesta seção, o método IS é aplicado à inicialização e previsão da taxa de precipitação com o modelo climático apresentado na seção 2 e sua versão imperfeita na qual a flutuação de subgrade é omitida. Neste modelo, a taxa de precipitação é controlável pela velocidade horizontal de grade e a temperatura potencial de grade em um sentido mais amplo que aquele sugerido em [7] e [2]. Por isso a noção de variável controlável é aqui re-definida da seguinte forma: um grupo de variáveis é controlável no instante  $t_0$  por outro grupo de variáveis, dito controlador, se

seu fluxo de  $t_0$  a  $t$  se contrai para todo  $t > t_0$  sempre que um valor de assimilação de dados no atrator é fornecido ao grupo controlador. Em geral esta propriedade é local no espaço e no tempo.

Doravante nos referimos ao modelo climático e sua versão imperfeita como o modelo perfeito e o modelo imperfeito respectivamente. Também, um estado de referência é um estado da atmosfera tomado de uma trajetória de referência do modelo perfeito. E o método de inserção é qualquer método de assimilação de dados que não perturbe as variáveis não-observáveis. Para clareza dos resultados, o grupo de variáveis escolhido como sendo controlador receberá valores de assimilação de dados exatos, ou seja iguais à referência, em toda análise de inserção ou do IS.

Nas figuras 1 e 2 os perfis de erro obtidos com o IS em ciclos de assimilação de dados com os modelos perfeito e imperfeito são comparados aos perfis de erro obtidos com o método de inserção nestes ciclos. Em todos os casos os ciclos de assimilação de dados têm comprimento de 5 ciclos-solares (sc). Em ambos os casos o erro da taxa de precipitação cai e se estabiliza nos ciclos de assimilação de dados do IS, enquanto que ele se superpõe ao perfil obtido com integração livre nos ciclos de assimilação de dados do método de inserção. As figuras 3 e 4 mostram as inicializações da taxa de precipitação obtidas na latitude 1250km-Sul, instante 5sc destes ciclos. Em ambos os casos a inicialização do IS se superpõe ao perfil de referência, enquanto que a inicialização do método de inserção está longe disso.

Todas as integrações do modelo climático foram feitas usando o método de linhas com Runge-Kutta de 4a ordem.

## 4 Conclusão

Apresentamos um sistema de equações mesoescalares para a velocidade de subgrade de um modelo atmosférico climático. Definimos um grupo de variáveis como sendo controlável no instante  $t_0$  por outro grupo de variáveis, dito controlador, se seu fluxo de  $t_0$  a  $t$  se contrai para todo  $t > t_0$  sempre que um valor de assimilação de dados no atrator é fornecido ao grupo controlador. No modelo climático, a velocidade vertical de grade é controlável pela velocidade horizontal de grade e a temperatura potencial de grade. Mostramos, então, que o método Influence Sampling (IS) de controle do caos, apresentado em [6], provê inicialização robusta e estabilização dinâmica da umidade e da taxa de precipitação em ciclos de assimilação de dados do modelo climático nos quais a velocidade horizontal de grade e a temperatura potencial de grade são escolhidas como variáveis controladoras em cada ciclo. O trabalho subsequente consiste em estudar o impacto dos processos mesoscópicos de subgrade na previsibilidade climática com a versão esférica do modelo climático aqui estudado e um código IS paralelizado.

## Referências

- [1] N. C. Bernardes, P. R. Cirilo, U. B. Darji, A. Messaoudi, E. R. Pujals, Expansivity and shadowing in linear dynamics, *J. Math. An. Applic.*, 461:1 (2018) 796–816.
- [2] C. Foias, G. Prodi, Sur le comportement global des solutions non-stationnaires des équations de Navier-Stokes en dimension 2, *Rend. Sem. Mat. Univ. Padova*, 39 (1967) 1-34.
- [3] C. Foias, R. Témam, Determination of the solutions of the Navier-Stokes equations by a set of nodal values, *Math. Comp.*, 43:167 (1984) 117–133.
- [4] G. H. Golub, J. M. Ortega, “Scientific Computing and Differential Equations: an Introduction to Numerical Methods”, Academic Press, 1992.

- [5] D. Jou, J. Casas-Vazquez, G. Lebon, Extended irreversible thermodynamics, *Rep. Prog. Phys.*, 51 (1988) 1105–1179.
- [6] P. Krause, Influence sampling of trailing variables of dynamical systems, *Mathematics of Climate and Weather Forecasting*, 3 (2017) 51–63.
- [7] O. A. Ladyzhenskaya, “The Mathematical Theory of Viscous Incompressible Flow”, Gordon and Breach Sci. Pub., 1963.

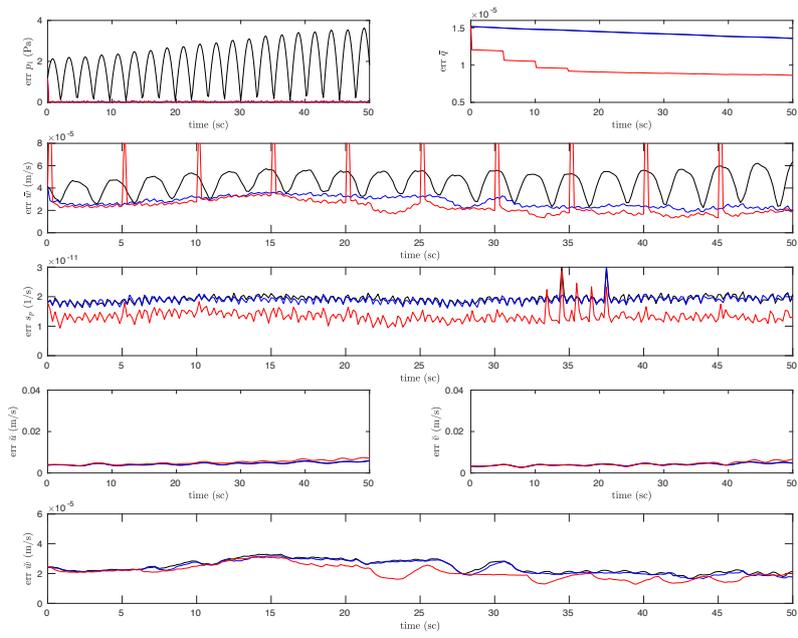


Figura 1: Perfis de erro obtidos com o modelo perfeito em ciclos de assimilação de dados de comprimento 5sc: integração livre (preto), método de inserção (azul) e método IS (vermelho).

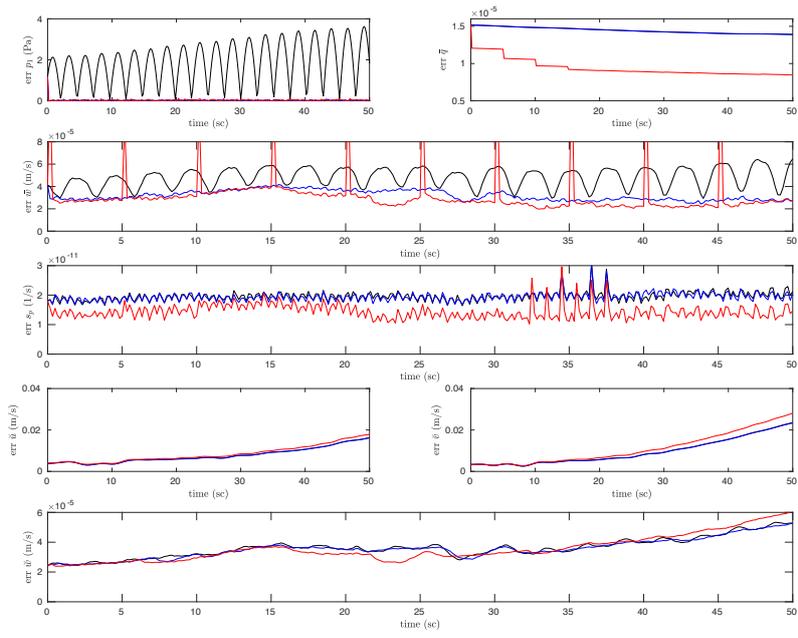


Figura 2: Perfis de erro obtidos com o modelo imperfeito em ciclos de assimilação de dados de comprimento 5sc: integração livre (preto), método de inserção (azul) e método IS (vermelho).

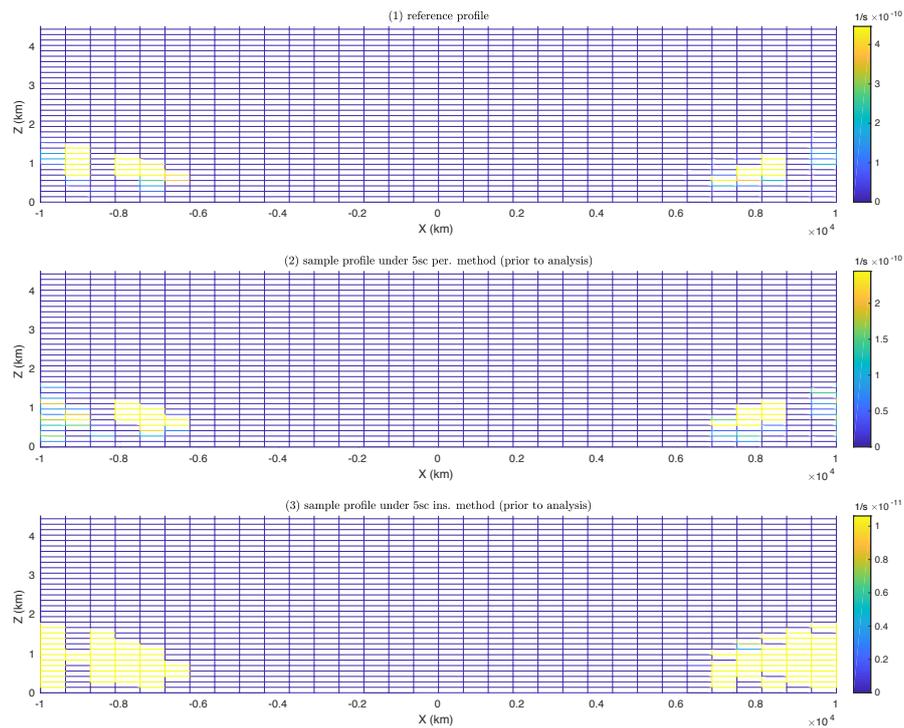


Figura 3: Inicializações de nuvens obtidas na latitude 1250km-Sul, instante 5sc dos ciclos da figura 1: referência (no alto), método IS (no meio), método de inserção (embaixo).

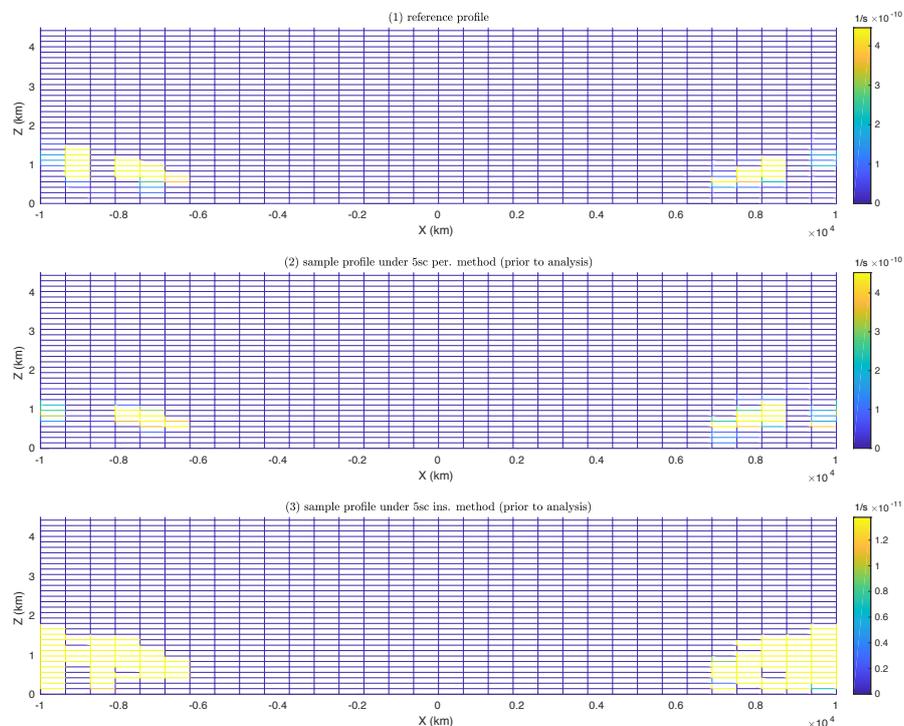


Figura 4: Inicializações de nuvens obtidas na latitude 1250km-Sul, instante 5sc dos ciclos da figura 2: referência (no alto), método IS (no meio), método de inserção (embaixo).

# Uma Proposta Matemática e Computacional para o Desenvolvimento da Metodologia da Matriz Densidade Tight-Binding Ordem-N

Julius Monteiro de Barros Filho ,      **Fernanda Lúcia Sá Ferreira**,

Coordenação de Disciplinas Básicas do Ensino Superior, CEFET-RJ,

26041-271, Nova Iguaçu, RJ

E-mail: julius.filho@cefet-rj.br,    fernanda.ferreira@cefet-rj.br,

**Moisés Augusto da Silva Monteiro de Araújo**

Departamento de Física, UFRRJ

23890-000, Seropédica, RJ

E-mail: mmaraujo@ufrj.br.

**Resumo:** *O objetivo deste trabalho é apresentar, em um contexto abrangente, uma proposta de fundamentação matemática que sirva de framework para o desenvolvimento da metodologia de cálculo de estrutura eletrônica chamada Método da Matriz Densidade Tight Binding (DMTB - Densit Matrix Tight Binding method), que foi apresentada originalmente na literatura pelo grupo liderado pelo físico David Vanderbilt em 1993 ([5]), bem como a sua relação com as modelagens computacionais que podem ser escolhidas para a sua implementação. A abordagem adotada deixa claro que a formulação matemática final desta metodologia é completamente dependente das estratégias computacionais que se escolhem para a sua efetiva implementação. Desta forma, colocamos o DMTB como um modelo matemático-computacional de formulação final variável. Por fim, propomos uma implementação baseada no método dos gradientes conjugados não linear. O modelo final obtido é ligeiramente diferente do DMTB que foi apresentado originalmente em 1993, estando em acordo com a versão apresentada por Millam e Scuseria, em 1997 ([7]). A abordagem utilizada desenvolve os aspectos matemáticos, visando a efetiva implementação computacional da metodologia.*

**Palavras-chave:** *matriz densidade Tight-Binding, modelagem matemática e computacional, mecânica quântica*

## 1 Introdução

Em geral, sistemas polieletrônicos possuem uma estrutura matemática bastante intrincada e as equações quantomecânicas só podem ser resolvidas com o uso de hipóteses simplificadoras, tais como a aproximação de Born-Oppenheimer e as aproximações de elétron independente. Atualmente as investigações teóricas de propriedades eletrônicas são feitas através de simulações computacionais de modelos matemáticos baseados nas metodologias básicas da teoria quântica da matéria condensada: as metodologias *semi-empíricas* e *de primeiros princípios*. Muitas propostas de resolução de modelos matemáticos que possuem descrição quântica em suas formulações, esbarram na alta complexidade computacional dos algoritmos envolvidos. Isto limita a aplicação destes modelos a sistemas com muitos átomos, impedindo que se façam simulações mais realistas. Várias estratégias matemático-computacionais com complexidade de ordem  $O(N)$  têm sido propostas, onde  $N$  é o número de átomos envolvidos. Uma delas é a metodologia semi-empírica chamada de Método da Matriz Densidade Tight-Binding (DMTB - *Densit Matrix Tight Binding method*). Este método foi desenvolvido por Li, Nunes e Vanderbilt e apresentado em [5]. Esta metodologia faz uso de hamiltonianos Tight-Binding parametrizados e, por apresentar possibilidade de implementação computacional de complexidade linear em relação ao número de

elétrons do sistema modelado, o DMTB tem sido largamente usado com sucesso para o estudo de estrutura eletrônica nas últimas duas décadas com formulações metodológicas diferentes das apresentadas do artigo original. Um exemplo pode ser visto em [7], por exemplo.

Neste trabalho é apresentado, em um contexto abrangente, uma proposta de fundamentação matemática para o DMTB. A abordagem utilizada desenvolve os aspectos matemáticos visando a efetiva implementação computacional da metodologia. Nesta construção, a formulação matemática final deste método é completamente dependente das estratégias computacionais que se escolhem para a sua implementação. Com isso, aqui, o DMTB fica estabelecido como um modelo matemático-computacional de formulação final variável.

## 2 Metodologia

Vamos considerar a representação de estruturas cristalinas por super-células. Podemos então formular o problema que estamos interessados da seguinte forma: dada uma estrutura gerada pela repetição de uma super-célula com  $N$  átomos e  $M$  orbitais eletrônicos por átomo, obter a energia quantomecânica do estado fundamental desta estrutura. Além disso, usaremos o formalismo do Operador Densidade Eletrônica,

$$\rho = \sum_n f_n |\psi_n\rangle \langle \psi_n|$$

onde  $n$  é o índice do espectro discreto do hamiltoniano,  $f_n$  é o número de ocupação e é usado o simbolismo introduzido por Dirac, no qual o símbolo  $|\rangle$ , chamado de *ket*, representa um vetor do espaço vetorial dos estados quantomecânicos, e o símbolo  $\langle |$ , chamado de *bra*, representa um vetor do dual topológico desse espaço. Para esse simbolismo, recomendamos [1], na página 18, e [2], na página 32.

Escrevendo os estados quantomecânicos como combinação linear da base finita do método Tight-Binding, obtem-se a matriz do operador  $\rho$ , chamada de matriz densidade eletrônica  $\rho_e$ . Admitindo-se que a base de orbitais forma um conjunto ortonormal, podemos escrever o número de elétrons -  $N_e$  - e a energia eletrônica total do sistema  $E$  em função da matriz densidade  $\rho_e$ , usando o operador “traço matricial”

$$N_e = \text{tr}(\rho_e) = \sum_i \rho_{eii}; \quad E = \text{tr}(\rho_e H) = \sum_{i,j} \rho_{eij} H_{ji}.$$

Além disso, como o operador densidade é um projetor, a matriz densidade é idempotente

$$\rho_e^2 = \rho_e.$$

Lembrando que nesse tipo de metodologia (Tigh-Binding + Matriz Densidade) a energia total de um sistema com  $N_e$  elétrons é usualmente dada por:

$$E_{tot} = E + E_{rep} + E_0 N, \quad (2.1)$$

onde  $E_{rep}$  representa o potencial repulsivo,  $N$  é número de átomos do sistema e  $E_0$  é uma constante de energia por átomo.

Para uma apresentação do método Tight-Binding usando o formalismo do Operador Densidade Eletrônica, bem como as demonstrações das identidades acima, recomendamos [8]. Na abordagem desta referência, assim como na nossa, os resultados foram obtidos admitindo-se que a base de orbitais forma um conjunto ortonormal.

### 2.1 Formulação do Problema

Como o estado quantomecânico fundamental é o estado de energia mínima e, uma vez que a energia do estado fundamental é função da densidade eletrônica ( $E = \text{tr}(\rho_e H)$ ), podemos agora

reformular o problema que estamos atacando: dada uma estrutura gerada pela repetição de uma super-célula com  $N$  átomos e  $M$  orbitais eletrônicos por átomo, com uma matriz hamiltoniana  $H$  (quadrada de ordem  $NM$ ), obter a matriz densidade eletrônica  $\rho_e$  que minimiza a função  $E(\rho) = \text{tr}(\rho H)$ . Note que estamos considerando a energia como uma função real de variável matricial. Daqui em diante, estaremos identificando o espaço das matrizes quadradas  $M_n(\mathbb{R})$  ( $n = NM^2$ ) com o espaço  $\mathbb{R}^{n^2}$  e, portanto, a energia  $E(\rho)$  será considerada um campo escalar em  $\mathbb{R}^{n^2}$ .

## 2.2 O Modelo de Minimização Irrestrita

A minimização indicada na formulação acima não pode ser feita de forma irrestrita. Como dissemos, a mecânica quântica usada na construção do problema impõe que o número de elétrons presentes na nossa estrutura precisa ser reproduzido pela matriz densidade do estado fundamental. Além disso, a matriz densidade do estado fundamental deve ser idempotente. Por fim, a matriz densidade eletrônica deve comutar com a matriz do hamiltoniano tight-binding. Tendo em vista as estratégias computacionais de interesse, nosso objetivo é obter um problema de minimização irrestrita. O vínculo do número de elétrons foi introduzido no modelo através do Teorema dos Multiplicadores de Lagrange:

**Teorema 2.1 (Teorema do Multiplicador de Lagrange)** *Seja  $f : U \rightarrow \mathbb{R}$  uma função de classe  $C^k$  ( $k \geq 1$ ) no conjunto aberto  $U \subset \mathbb{R}^{n+1}$ , e  $M = g^{-1}(0)$  uma hiperfície contida em  $U$ , imagem inversa do valor regular 0 por uma função  $g : U \rightarrow \mathbb{R}$ , também de classe  $C^k$ . Um ponto  $p \in M$  é um ponto crítico de  $f|_M$  se, e somente se, existe um número real  $\mu$  (chamado de multiplicador de Lagrange) tal que*

$$\nabla f(p) + \mu \nabla g(p) = 0.$$

Considerando a identificação de  $M_n(\mathbb{R})$  ( $n = NM^2$ ) com o espaço  $\mathbb{R}^{n^2}$ , o teorema anterior nos garante o seguinte:  $\rho_0$  é um ponto crítico da função  $E(\rho) = \text{tr}(\rho H)$ , restrita a hiperfície  $g^{-1}(0) = \{\rho \in M_n(\mathbb{R}); \text{tr}(\rho) - N_e = 0\}$ , se, e somente se, existe um número real  $\mu$  que torna nulo, em  $\rho_0$ , o gradiente da função

$$\Omega(\rho) = \text{tr}(\rho H) + \mu(\text{tr}(\rho) - N_e). \quad (2.2)$$

desde que o gradiente do vínculo  $g$  não seja nulo em  $\rho$ . Além disso, o gradiente de  $g$  pode ser obtido facilmente:

$$\nabla(g(\rho)) = \nabla((\text{tr}(\rho) - N_e)) = I^t = I,$$

mostrando que o uso dos multiplicadores de Lagrange é permitido.

Aqui vale observar que o grupo de Vanderbilt originalmente concebe o escalar  $\mu$  como um potencial químico em [5]. Com isso, é necessário algum procedimento para atualizar o valor do potencial químico durante a busca pela matriz densidade. Diferentemente disso, na nossa abordagem, a interpretação de  $\mu$  como um multiplicador de Lagrange, que tem o papel de selecionar matrizes que reproduzem o número correto de elétrons, permite usar a combinação da implementação computacional baseada no métodos dos gradientes conjugados não linear, que foi reportada em [3], com exigência da consistência física para o número de elétrons, para determinar explicitamente o multiplicador  $\mu$  como função de  $\rho$  em cada passo de iteração. Portanto, como veremos, no modelo acima,  $\mu$  não será uma incógnita independente. Esta mesma interpretação para  $\mu$  também é reportada em [7].

O vínculo da idempotência foi imposto, de forma aproximada e implícita, através da chamadas “transformações de purificação de McWeeny” ([6]), definida no espaço das matrizes quadradas, como se segue. Considere a aplicação

$$F(X) = 3X^2 - 2X^3,$$

definida no espaço das matrizes quadradas. É evidente que toda matriz idempotente é um ponto fixo desta aplicação. Além disso, é fácil ver que, se  $\lambda$  é um autovalor de uma matriz  $X$ , então  $f(\lambda)$  é autovalor de  $F(X)$ , onde  $f$  é o polinômio  $f(x) = 3x^2 - 2x^3$ .

Como sabemos, os autovalores de uma matriz idempotente pertencem ao conjunto  $\{0, 1\}$ . Seja  $B$  uma matriz “aproximadamente idempotente” com erro máximo absoluto da ordem de  $O(\epsilon)$ , no sentido de que, para um certo  $\epsilon > 0$ , seus autovalores são dados por  $h$  ou  $1 + h$ , com  $h \in \mathbb{R}$  e  $0 < |h| < \epsilon$ . Com isso, os autovalores de  $F(B)$  são  $f(h)$  ou  $f(1+h)$ . Pelo desenvolvimento de Taylor de ordem 1 para  $f$ , vemos que, a matriz  $F(B)$  será aproximadamente idempotente com erro máximo absoluto da ordem de  $O(\epsilon^2)$ . Portanto, usar  $F(\rho)$  no lugar de  $\rho$  na equação (2.2) se apresenta como uma proposta de introdução implícita e aproximada do vínculo da idempotência no nosso modelo.

A aplicação  $F$  é conhecida como Transformação de Purificação de MacWneey ([6]) e foi incorporada no modelo do grupo de Vanderbilt. Analisando o polinômio de purificação  $f$ , vemos que, estando os autovalores de uma matriz no intervalo  $[-0.5, 1.5]$ , os autovalores da matriz purificada permanecerão no intervalo  $[0, 1]$ . Com isso, a função que será minimizada, de forma irrestrita, é dada por

$$\Omega(\rho) = tr([3\rho^2 - 2\rho^3]H) + \mu(tr(\rho) - N_e), \quad (2.3)$$

onde o multiplicador  $\mu$  será definido a seguir.

Ainda levando em conta a identificação do espaço das matrizes e considerando-se  $\Omega$  também como um campo escalar, obtem-se que

$$\nabla\Omega = 3(H\rho + \rho H)^t - 2(H\rho^2 + \rho H\rho + \rho^2 H)^t + \mu I, \quad (2.4)$$

que é fundamental para os métodos computacionais de otimização irrestrita.

### 2.3 Estratégias Computacionais

Nos algoritmos básicos de otimização irrestrita de campos escalares  $\Omega(\rho)$  (máxima descida, Newton, Quase Newton, etc) as rotinas de busca linear (em linha) produzem uma sequência iterativa que converge para o ponto estacionário procurado

$$\rho_{k+1} = \rho_k + \alpha D_k,$$

onde  $\alpha$  é o passo dado na direção de busca  $D_k$ . No nosso contexto, devemos exigir que, em cada iteração, o número de elétrons do sistema seja mantido constante, ou seja, para todo  $k$ , devemos ter que  $tr(\rho_{k+1}) = tr(\rho_k)$ . Portanto, vamos impor, para todo  $k$ , que

$$tr(D_k) = 0.$$

Sem entrar em muitos detalhes, um método dos gradientes conjugados não linear (NLCG) resolve o problema de minimização irrestrita de uma função diferenciável  $\Omega : \mathbb{R}^d \rightarrow \mathbb{R}$  construindo uma sequência de vetores usando a busca em linha

$$\rho_{k+1} = \rho_k + \alpha_k D_k, \quad (2.5)$$

com direção de busca definida por

$$D_k = \begin{cases} -\nabla\Omega(\vec{x}_k), & k = 1 \\ -\nabla\Omega(\vec{x}_k) + \beta_k \vec{d}_{k-1}, & k \geq 2, \end{cases} \quad (2.6)$$

onde  $\beta_k$  são escalares reais. Existem várias propostas (não equivalentes) para o escalar  $\beta_k$  reportadas na literatura especializada em Otimização Numérica. Cada proposta gera um método NLCG diferente. Fixando um algoritmo de NLCG, usando a definição de direção de busca (de

descida) dada em (2.6) e a nossa exigência física de que  $tr(D_k) = 0$ , juntamente com o gradiente dado em 2.4 constatamos que devemos definir

$$\mu_k = \frac{6}{N_e} [tr(\rho_k^2 H) - tr(\rho_k H)]. \quad (2.7)$$

Métodos baseados no NLCG têm sido aplicados em Modelagem Matemática e Computacional de fenômenos da Física da Matéria Condensada desde, pelo menos, a década de 80 do século passado (ver [9], por exemplo). O método dos gradientes conjugados (não linear) tem algumas possibilidades de implementação com complexidade linear. O isomorfismo canônico entre o espaço das matrizes com o espaço euclidiano  $\mathbb{R}^d$  tem complexidade computacional da ordem  $d^2$ . Esta complexidade pode ser reduzida a complexidade linear fazendo-se uso de um raio de corte para a matriz densidade, como sugerido em [5], como se segue. Admitindo uma implementação para o NLCG em  $\mathbb{R}^d$  contendo  $d$  passos (algoritmo ótimo), levando em consideração a nossa identificação entre matrizes e vetores, a versão adaptada para o DMTB faria uso de  $d^2$  passos, uma vez que, para o cálculo da energia do sistema, em princípio seria necessário computar todos os elementos da matriz densidade. No entanto, sabe-se que

$$\lim_{R_{ij} \rightarrow \infty} \rho_{ij} = 0,$$

onde  $R_{ij}$  a distância entre os orbitais  $i$  e  $j$  ([5]). Esta convergência para zero tem ordem de convergência polinomial em metais e exponencial em isolantes ([5]). Diante deste fato, o grupo de Vanderbilt propôs a introdução de um raio de corte  $R_c > 0$  de modo que se  $R_{ij} > R_c$  então  $\rho_{ij} = 0$ . Sendo assim, fixado um átomo na rede, seja  $L$  o número de átomos da rede contidos na esfera de raio  $R_c$ , centrada no átomo que foi previamente fixado (devido à simetria da rede,  $L$  não depende do átomo fixado, sendo então uma constante para cada escolha do raio de corte). Com isso, para formarmos um elemento  $\rho_{ij}$  eventualmente não nulo,  $i$  pode ocupar qualquer uma das  $NM$  posições disponíveis, porém  $j$  deve variar dentre  $LM$  posições. Daí, se consideramos apenas estes elementos eventualmente não nulos da matriz densidade, o nosso processo de vetorização de matrizes produz um vetor com  $NLM^2$  coordenadas. Tendo este vetor como entrada, o algoritmo CG faria uso de  $NLM^2$  passos, apresentando, portanto, complexidade  $O(N)$ . É claro que  $R_c$  precisa ser escolhido de forma a obter matrizes de densidade próximas das densidades verdadeiras. Na implementação feita em [3] usou-se  $M = 4$  (um orbital  $s$  e três orbitais  $p$ ).

### 3 Considerações Finais

Para que esta construção de  $\mu$  (equação (2.7)) tenha o efeito que queremos, é fundamental que a matriz inicial  $\rho_0$  já possua o número correto de elétrons da super célula. Desta forma, a expressão

$$\mu(tr(\rho) - N_e),$$

que já era nula em  $\rho_0$ , se mantém nula durante todo o processo. Portanto a expressão do funcional  $\Omega$  se torna apenas

$$\Omega(\rho) = tr((3\rho^2 - 2\rho^3)H).$$

Com esta abordagem, o multiplicador de Lagrange não surte efeito algum sobre o funcional objetivo  $\Omega$ . Seu efeito, no entanto, se manifesta no gradiente  $\nabla\Omega$  em cada nova iteração.

Portanto, o problema final é a minimização irrestrita (via algoritmos dos gradientes conjugados não linear) de

$$\Omega(\rho) = tr((3\rho^2 - 2\rho^3)H),$$

usando-se as equações (2.4) e (2.7).

Por fim, a Mecânica Quântica ainda exige que tenhamos a comutatividade do produto  $\rho H$ , quando  $\rho = \rho_e$  for a densidade eletrônica. Esta propriedade afeta diretamente a escolha do

critério de parada para o nosso algoritmo baseado em NLCG. De fato, supondo que as matrizes construídas vão ficando aproximadamente idempotentes o suficiente para que possamos considerar  $\mu = 0$ , vemos que o gradiente se torna

$$\nabla\Omega = (H\rho + \rho H - 2\rho H\rho)^t.$$

É claro que, se  $\rho$  e  $H$  comutam, então  $\nabla\Omega = 0$ . Por outro lado, admitamos que  $\nabla\Omega = 0$ . Daí,

$$H\rho + \rho H = 2\rho H\rho.$$

Multiplicando-se esta equação por  $\rho$  à esquerda, vemos que

$$\rho H\rho + \rho^2 H = 2\rho^2 H\rho.$$

Considerando  $\rho^2 = \rho$ , obtemos que  $\rho H = \rho H\rho$ . De forma análoga, fazendo agora a multiplicação à direita, obtemos que  $H\rho = \rho H\rho$ . E, portanto  $\rho H = H\rho$ . Com isso, vemos que  $\rho$  será um ponto crítico de  $\Omega$  se, e somente se comutar com a matriz do hamiltoniano Tight-Binding. Isto nos leva a concluir que não devemos usar um número máximo de iterações no nosso algoritmo. O critério de parada deve estar ligado apenas à tolerância do valor  $\|\nabla\Omega\|$ .

Abaixo segue o pseudocódigo, baseado na metodologia do NLCG, para a minimização do funcional  $\Omega(\rho)$ , usado em [3].

---

**Algoritmo 1:** NLCG PARA O FUNCIONAL DE ENERGIA (POLAK-RIBIÈRE-TYPE)

---

**Entrada:**  $\Omega, \nabla\Omega, \rho_0, tol, it_{max}$  {funcional de energia, gradiente do funcional, estimativa inicial para a matriz densidade e número máximo de iterações }

**Saída:** matriz densidade eletrônica  $\rho$  para o estado fundamental

```

1 início
2    $it \leftarrow 0$ ;
3    $\vec{d}_0 \leftarrow -\nabla\Omega(x_0)$ ;
4    $\rho \leftarrow \rho_0$ ;
5    $\vec{d} \leftarrow \vec{d}_0$ ;
6    $G \leftarrow \vec{d}_0$ ;
7   enquanto  $\|G\| > tol$  e  $it < it_{max}$  faça
8      $G_0 \leftarrow \nabla\Omega(\rho_0)$ ;
9     Determine  $\alpha$  usando uma busca unidirecional
10     $\rho \leftarrow \rho + \alpha\vec{d}$ ;
11    Determine o Multiplicador de Lagrange  $\mu$ ;
12     $G \leftarrow \nabla\Omega(\rho)$ ;
13     $\beta \leftarrow \frac{tr(G(G-G_0)^t)}{tr(G_0G_0^t)}$ ;
14     $d \leftarrow -G + \beta d$ ;
15     $\rho_0 \leftarrow \rho$ ;
16     $it \leftarrow it + 1$ ;
17  fim
18 fim
19 retorna  $\rho$ 

```

---

Trata-se portanto de um processo iterativo que constrói a matriz densidade eletrônica do estado fundamental  $\rho_e$ .

Estes cálculos são realizados com hamiltonianos tight-binding parametrizados para espécies atômicas específicas. Deve ficar claro que, embora baseada em NLCG, esta abordagem computacional não se enquadra na definição geral de métodos dos gradientes conjugados, uma vez que o gradiente do campo em questão é atualizado a cada iteração do processo. Por tanto, os teoremas de convergência global conhecidos para esta categoria de algoritmos não podem ser aplicados.

A convergência e validação dos resultados de um caso particular pode ser apreciada em [3], onde a implementação com cristais de silício foi feita rigorosamente nos moldes do presente trabalho, com o uso do hamiltoniano de Kwon ([4]).

Para o silício, muitas medidas, sejam obtidas de forma experimental, sejam obtidas por métodos *ab initio*, estão em acordo com as previsões obtidas pelo DMTB. Sem entrar em muitos detalhes, o tipo validação proposta na referência [3] foi a seguinte: podemos tomar o parâmetro de rede experimental dos cristais de Si. Sabe-se que este parâmetro de rede vale aproximadamente 5.43 *angstroms*. Sendo assim, o algoritmo do DMTB aplicado ao silício (ou seja, com um hamiltoniano Tight-Binding parametrizado para o Si), quando alimentado com este valor de parâmetro de rede, deve apresentar energia total ( $E_{tot}$ ) com valor mais baixo do que outros valores para este parâmetro. Mais precisamente, a energia total dada pela equação (2.1), vista agora como função do parâmetro de rede, deve apresentar um mínimo em 5.43 Å. Como, tanto a matriz do hamiltoniano cristalino, quanto  $E_{rep}$  são dados de forma parametrizada, este tipo de validação não está ligada somente ao DMTB, mas também ao hamiltoniano escolhido. Abaixo vemos o gráfico obtido, utilizando o hamiltoniano de Kwon ([4]).

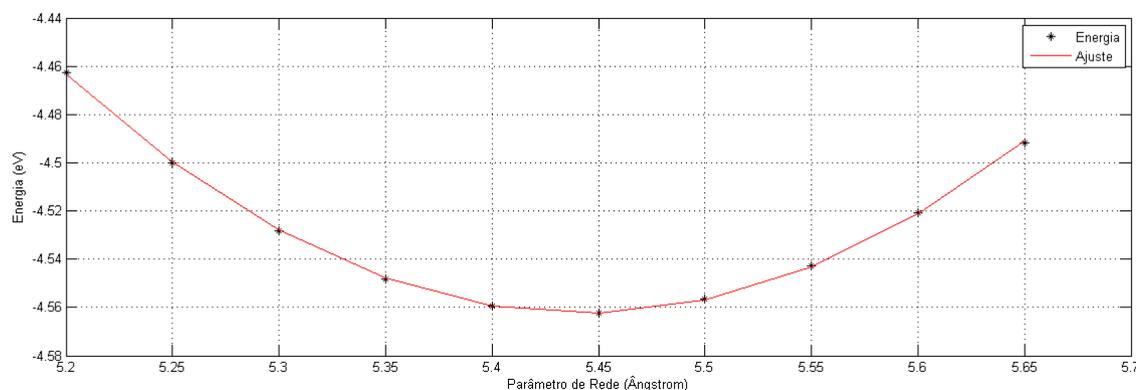


Figura 1: Cálculo do Parâmetro de Rede Experimental.

Para este gráfico, foram utilizados 10 valores de energia obtidos com o DMTB alimentado com valores de parâmetros de rede variando de 5.20 até 5.65, de 0.5Å em 0.5Å. Em seguida, ajustou-se, por mínimos quadrados, os pontos em uma parábola e obteve-se o seu mínimo. o valor mínimo ocorre em 5.4433Å. Com isso, vemos que o erro relativo é de aproximadamente 0.2%. Para maiores detalhes sobre este resultado, ver [3].

Ainda sobre a questão da complexidade computacional, vemos que, mesmo com o uso do raio de corte, o algoritmo acima não possui complexidade linear pois, na linha 13, temos o traço do produto de duas matrizes. No entanto, existem propostas de programação paralela bem estabelecidas para esta operação. Além disso, é necessário o uso de métodos de manipulação para matrizes esparsas. Portanto, a análise final de complexidade algorítmica deve ser feita sob todas estas considerações.

## Referências

- [1] P. Dirac, “The Principles of Quantum Mechanics”, Oxford University Press, Oxford, 1947.
- [2] J. M. Jauch, “Foundation of Quantum Mechanics”, ADDISON-WESLEY PUBLISHING COMPANY, Oxford, 1968.
- [3] F. L. S. Ferreira, J. M. B. Filho e M. A. S. M. Araújo, Gradiente Conjugado não Linear sobre um Ambiente “Matriz densidade Tight Binding ordem N” para o Estudo de Semi Condutores, em “Anais do ERMAC - Encontro Regional de Matemática Aplicada e Com-

- putacional / II Simpósio de Matemática Aplicada e Computacional da UFRRJ” (Sociedade Brasileira de Matemática Aplicada e Computacional) pp. 136–139, 2017.
- [4] I. Kwon, R. Biswas, C. Z. Wang, K. M. Ho and C. M. Soukoulis, Transferable tight-binding models for silicon, *Physical Reviews B*, 49 (1994) 7242–7250.
  - [5] X.-P. Li, R. W. Nunes and R. W. D. Vanderbilt, Density-matrix electronic-structure method with linear system-size scaling, *Phys. Rev. B*, 47 (1993) 10891–10894.
  - [6] R. McWeeny, Some Recent Advances in Density Matrix Theory, *Reviews of Modern Physics*, 32 (1960) 335–369.
  - [7] J. M. Millam and G. E. Scuseria, Linear scaling density matrix, *J. Chem. Phys.*, 106 (1997) 5569–5577.
  - [8] A. T. Paxton, An Introduction to the Tight Binding approximation-implementation by diagonalisation, *Multiscale Simulation Methods in Molecular Sciences*, 42 (2009) 145–176.
  - [9] M. P. Terter, Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients, *Physical Reviews B*, 40 (1989) 1225–1263.

## Sobre Funções Reais de Variável Matricial: Tratamento no Cálculo de Estrutura Eletrônica e Algoritmos de Busca

Fernanda Lúcia Sá Ferreira , Julius Monteiro de Barros Filho ,

Coordenação de Disciplinas Básicas do Ensino Superior, CEFET-RJ,

26041-271, Nova Iguaçu, RJ

E-mail: julius.filho@cefet-rj.br, fernanda.ferreira@cefet-rj.br,

Moisés Augusto da Silva Monteiro de Araújo

Departamento de Física, UFRRJ

23890-000, Seropédica, RJ

E-mail: mmaraujo@ufrj.br.

**Resumo:** *O objetivo deste artigo é usar o isomorfismo canônico existente entre o espaço vetorial das matizes quadradas de uma dada ordem  $n \in \mathbb{N}$  e o espaço vetorial  $\mathbb{R}^{n^2}$  para transportar o conceito de derivada, já bem estabelecido e com propriedades bem exploradas em funções reais de várias variáveis reais em  $\mathbb{R}^{n^2}$  para funções reais de variável matricial. Com a abordagem desenvolvida, tornamos disponíveis para o uso em funções reais de variável matricial, uma série de técnicas baseadas no Cálculo Diferencial, tais como os métodos de otimização nos espaços euclidianos  $\mathbb{R}^{n^2}$ . Devido a construção que fizemos, este uso se dá de forma nativa no espaço matricial, sem uso explícito do referido isomorfismo. Para exemplificar e motivar o desenvolvimento apresentado, mostramos a adaptação de um algoritmo de Gradiente Conjugado Não Linear para a minimização de uma função real de variável matricial. Esta função é oriunda de um método semi empírico da Mecânica Quântica para o cálculo de estrutura eletrônica conhecido por DMTB (Densit Matrix Tight Binding method).*

**Palavras-chave:** *Cálculo Diferencial Matricial, DMTB (Densit Matrix Tight Binding method), Gradiente Conjugado Não Linear*

### 1 Introdução

Este trabalho trata de um desenvolvimento de ferramentas básicas de cálculo diferencial para funções reais de uma variável matricial. A motivação deste trabalho repousa em uma metodologia semi empírica da Mecânica Quântica onde funções de variável matricial surgem de forma bastante natural. Esta metodologia é conhecida por “Método da Matriz Densidade Eletrônica Tight-Binding” (Density Matrix Tight Binding method - DMTB), desenvolvida nos anos 90 do século passado por um grupo liderado pelo físico David Vanderbilt [3], que é um método para a obtenção da energia do estado fundamental (energia mínima) de uma estrutura cristalina através da construção de uma matriz, chamada de *matriz densidade eletrônica*, por uma técnica de minimização irrestrita. Com isso, a função a ser minimizada é então uma função real de uma variável matricial. Sendo  $\rho_e$  a matriz densidade eletrônica, sabemos da Mecânica Quântica que a energia eletrônica é dada por

$$E = \text{tr}[\rho_e H], \quad (1.1)$$

onde “*tr*” representa o operador traço de uma matriz quadrada e  $H$  representa o hamiltoniano cristalino Tight Binding. A matriz densidade eletrônica deve respeitar algumas propriedades exigidas pela Mecânica Quântica, como ser idempotente e o seu traço corresponder ao número de elétrons da estrutura, o que nos leva a um problema de minimização restrita. No DMTB estes vínculos quânticos são acomodados de forma implícita e aproximada em uma função que

possa ser minimizada de forma irrestrita. Para maiores detalhes recomendamos [3]. Em [3] a função de energia a ser minimizada de forma irrestrita é dada por

$$\Omega(\rho) = \text{tr}[(3\rho^2 - 2\rho^3)(H - \mu I)], \quad (1.2)$$

onde  $I$  representa a matriz identidade de ordem  $n$ .

Nos últimos anos surgiram outras propostas não equivalentes para o DMTB na literatura especializada em Física da Matéria Condensada. Tanto em [4] como em [2], por exemplo, a função de energia é dada por

$$\Omega(\rho) = \text{tr}((3\rho^2 - 2\rho^3)H) + \mu(\text{tr}(\rho) - N_e). \quad (1.3)$$

Em ambas as formulações, a única variável é a matriz quadrada  $\rho$  de ordem  $n$  ( $n$  é igual ao produto do número de átomos pelo número de orbitais por átomo na estrutura cristalina). Na equação (1.2),  $H$  representa um hamiltoniano cristalino Tight Binding parametrizado para a espécie atômica de interesse,  $\mu$  é um potencial químico <sup>1</sup> (nível de Fermi) e  $N_e$  é o número de elétrons da estrutura cristalina.

De qualquer forma, a técnica de otimização utilizada é uma técnica construída nos espaços euclidianos  $\mathbb{R}^d$ , conhecida como Gradiente Conjugado Não Linear que, dentre outras questões, depende do conhecimento do gradiente da função a ser minimizada. Portanto, neste contexto, a necessidade de desenvolvimento de técnicas de Cálculo Diferencial para funções reais de variável matricial surge de forma bastante natural. Com isso, somos levados a construir identificações de conceitos e técnicas do espaço  $\mathbb{R}^{n^2}$  para o espaço das matrizes quadradas de ordem  $n$ .

## 2 Metodologia

Um fato básico e fundamental é que todos os espaços vetoriais de uma mesma dimensão finita são linearmente isomorfos. Portanto, dado  $n \in \mathbb{N}$ , podemos identificar, enquanto estrutura algébrica, o espaço vetorial das matrizes quadradas de ordem  $n$  com entradas reais,  $M_n(\mathbb{R})$ , com o espaço dos vetores de  $\mathbb{R}^{n^2}$ . Além disso, como as normas em  $\mathbb{R}^{n^2}$  induzem normas em  $M_n$ , as questões topológicas também estão bem identificadas.

Chamaremos qualquer isomorfismo linear entre  $M_n(\mathbb{R})$  e  $\mathbb{R}^{n^2}$  de *processo de vetorização de matrizes*. Vamos fixar um destes processos de vetorização. Escolhemos o isomorfismo linear entre estes dois espaços, definido por  $\text{vec}(X) = (X_{11}, \dots, X_{1n}, X_{21}, \dots, X_{2n}, \dots, X_{n1}, \dots, X_{nn})$ , onde estamos subentendendo a notação usual de matrizes  $X = (X_{ij})$ . Tendo sempre este processo de vetorização em mente, vamos definir o gradiente de funções reais de uma variável matricial e obter algumas regras de derivação para as mesmas.

Antes de tratar a questão diferencial, convém observar que usando um processo de vetorização fixado, podemos definir operações e conceitos em  $M_n(\mathbb{R})$  de forma muito natural. Por exemplo, um produto interno em  $M_n(\mathbb{R})$  pode ser definido da seguinte maneira: dados as matrizes  $A$  e  $B$ , definimos  $\langle A, B \rangle = \text{vec}(A) \cdot \text{vec}(B)$ . O lema 2.1 relaciona, o traço do produto de duas matrizes, uma operação que não possui uma identificação óbvia pelo processo de vetorização, com o produto interno em  $M_n(\mathbb{R})$  que mencionamos acima.

**Lema 2.1.** *Seja  $n \in \mathbb{N}$ . Consideremos o espaço vetorial real das matrizes quadradas de ordem  $n$ ,  $M_n(\mathbb{R})$ , e o isomorfismo linear  $\text{vec}$  entre  $M_n(\mathbb{R})$  e  $\mathbb{R}^{n^2}$  dado por*

$$X \in M_n(\mathbb{R}) \mapsto \text{vec}(X) = (X_{11}, \dots, X_{1n}, X_{21}, \dots, X_{2n}, \dots, X_{n1}, \dots, X_{nn}) \in \mathbb{R}^{n^2}.$$

*Então,  $\text{tr}(AB) = \text{vec}(A) \cdot \text{vec}(B^T)$ , quaisquer que sejam  $A, B \in M_n(\mathbb{R})$ , onde “ $\cdot$ ” representa o produto escalar em  $\mathbb{R}^{n^2}$*

<sup>1</sup>Em [2], o parâmetro  $\mu$  na equação 1.3 representa um Multiplicador de Lagrange, tendo, portanto, um significado diferente do apresentado na equação 1.2. Para maiores detalhes, ver [2].

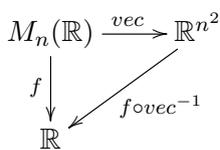
*Demonstração.* Usando a definição de produto matricial , temos que

$$\text{tr}[AB] = \sum_{i=1}^n (AB)_{ii} = \sum_{i=1}^n \sum_{k=1}^n A_{ik}B_{ki} = \text{vec}(A) \cdot \text{vec}(B^T). \square$$

Como definimos  $\langle A, B \rangle = \text{vec}(A) \cdot \text{vec}(B)$ , o lema 2.1 mostra que é possível encapsular o processo de vetorização para o cálculo do produto interno: dadas duas matrizes  $A$  e  $B$  pertencentes a  $M_n(\mathbb{R})$ , basta fazer, de forma nativa,  $\langle A, B \rangle = \text{tr}[AB^t]$ .

## 2.1 A Matriz Gradiente

A nossa vetorização de matrizes permite que consideremos uma função real de uma variável matricial como sendo um campo escalar. Mais precisamente, sendo  $f : M_n(\mathbb{R}) \rightarrow \mathbb{R}$ , vemos que  $f \circ \text{vec}^{-1} : \mathbb{R}^{n^2} \rightarrow \mathbb{R}$ , é um campo escalar.



Diremos então que  $f$  é diferenciável quando o campo escalar  $f \circ \text{vec}^{-1}$  for diferenciável. Sendo assim, a ideia é definirmos o gradiente de  $f$  como sendo uma matriz, que denotaremos por  $\frac{\partial f}{\partial X}$ , cuja vetorização seja o gradiente do campo escalar  $f \circ \text{vec}^{-1}$ .

Ou seja, estamos propondo que a definição de  $\frac{\partial f}{\partial X}$  satisfaça  $\text{vec}\left(\frac{\partial f}{\partial X}\right) = \nabla(f \circ \text{vec}^{-1})$ . De forma nativa, definimos então  $\frac{\partial f}{\partial X} = \text{vec}^{-1} \circ \nabla(f \circ \text{vec}^{-1})$ , que é uma função matricial de variável vetorial, cuja a expressão é bastante simples:

$$\frac{\partial f}{\partial X} = \begin{bmatrix} \frac{\partial f}{\partial X_{11}} & \frac{\partial f}{\partial X_{12}} & \cdots & \frac{\partial f}{\partial X_{1n}} \\ \frac{\partial f}{\partial X_{21}} & \frac{\partial f}{\partial X_{22}} & \cdots & \frac{\partial f}{\partial X_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial X_{n1}} & \frac{\partial f}{\partial X_{n2}} & \cdots & \frac{\partial f}{\partial X_{nn}} \end{bmatrix}$$

Mais uma vez, cabe a observação de que, o gradiente matricial definido acima tem como vetorização justamente o gradiente, digamos usual, do campo escalar  $f$ , quando  $X$  é posta na forma vetorizada que nos fixamos. Com isso, as propriedades básicas (lineares e com respeito ao produto interno) do operador gradiente são todas válidas para este novo operador.

No desenvolvimento do DMTB, é particularmente importante calcularmos o gradiente do traço da potência de matrizes (veja a equação (1.2)). A proposição abaixo é suficiente para esta necessidade.

**Proposição 2.2.** *Seja  $n \in \mathbb{N}$ . Consideremos o espaço vetorial real das matrizes quadradas de ordem  $n$ ,  $M_n(\mathbb{R})$ , e fixemos uma matriz  $A \in M_n(\mathbb{R})$ . Então valem as seguintes regras de derivação:*

1. Sendo a função real de variável matricial  $f$  dada por  $f(X) = \text{tr}(XA)$ , temos que

$$\frac{\partial f}{\partial X} = A^t;$$

2. Sendo a função real de variável matricial  $f$  dada por  $f(X) = \text{tr}(X^2A)$ , temos que

$$\frac{\partial f}{\partial X} = (AX + XA)^T;$$

3. Sendo a função real de variável matricial  $f$  dada por  $f(X) = \text{tr}(X^3A)$ , temos que

$$\frac{\partial f}{\partial X} = (AX^2 + XAX + X^2A)^T;$$

*Demonstração.* 1. Pela definição de traço e de produto matricial, temos que

$$f(X) = \text{tr}(AX) = \sum_{i=1}^n \sum_{k=1}^n A_{ik} X_{ki}.$$

Daí, fixados  $r$  e  $s$ ,  $\frac{\partial f}{\partial X_{rs}} = A_{sr}$ . Portanto,  $\frac{\partial f}{\partial X} = A^t$ .

2. Neste caso,

$$f(X) = \text{tr}(AX^2) = \sum_{i=1}^n \sum_{k=1}^n \sum_{j=1}^n A_{ik} X_{kj} X_{ji}.$$

Com isso, para o cálculo de  $\frac{\partial f}{\partial X_{rs}}$ , contribuem da expressão acima apenas as somas em que  $k = r$  e  $j = s$ , e  $j = r$  e  $i = s$ . Portanto, fixados  $r$  e  $s$ ,

$$\frac{\partial f}{\partial X_{rs}} = \sum_{i=1}^n A_{ir} X_{si} + \sum_{k=1}^n A_{sk} X_{kr} = (XA)_{sr} + (AX)_{sr},$$

mostrando que  $\frac{\partial f}{\partial X} = (AX)^t + (XA)^t$ .

3. Basta escrevermos,

$$\begin{aligned} f(X) &= \text{tr}(AX^3) = \text{tr}((AX)X^2) = \sum_{i=1}^n \sum_{k=1}^n (AX)_{ik} (X^2)_{ki} \\ &= \sum_{i=1}^n \sum_{k=1}^n \left( \sum_{l=1}^n A_{il} X_{lk} \right) \left( \sum_{j=1}^n X_{kj} X_{ji} \right) = \sum_{i=1}^n \sum_{k=1}^n \sum_{l=1}^n \sum_{j=1}^n A_{il} X_{lk} X_{kj} X_{ji}. \end{aligned}$$

Agora notemos que, fixados  $r$  e  $s$ , para o cálculo de  $\frac{\partial f}{\partial X_{rs}}$ , contribuem da expressão acima apenas as somas em que  $l = r$  e  $k = s$ ,  $k = r$  e  $j = s$ , e  $j = r$  e  $i = s$ . Portanto,

$$\begin{aligned} \frac{\partial f}{\partial X_{rs}} &= \sum_{i=1}^n \sum_{j=1}^n A_{ir} X_{sj} X_{ji} + \sum_{i=1}^n \sum_{l=1}^n A_{il} X_{lr} X_{si} + \sum_{k=1}^n \sum_{l=1}^n A_{sl} X_{lk} X_{kr} \\ &= (X^2 A)_{sr} + (XAX)_{sr} + (AX^2)_{rs}, \end{aligned}$$

revelando que  $\frac{\partial f}{\partial X} = (AX^2)^t + (XAX)^t + (X^2 A)^t$ . □

Sendo  $f(X)$  uma função de domínio matricial, a matriz gradiente  $\frac{\partial f}{\partial X}$  independe de processos de vetorização, no sentido que ela poderia ter sido definida, ainda que artificialmente, de forma direta. No entanto, a identificação  $\frac{\partial f}{\partial X} \longleftrightarrow \nabla f$ , onde  $f$  é encarada como um campo escalar definido em um subconjunto de  $\mathbb{R}^{n^2}$ , só é correta para a vetorização que fixamos aqui.

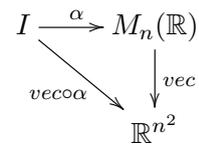
## 2.2 Uma Regra da Cadeia

No caso específico do DMTB, é necessária a derivada da composição de  $\Omega$  (dada pela equação (1.2)) com a linha  $\alpha(t) = \rho + tD$ , para matrizes  $\rho$  e  $D$  fixadas e  $t \in I$ , onde  $I$  é um intervalo aberto. Repare que  $\alpha(t)$  é um caminho matricial.

Definimos a derivada de caminhos matriciais, seguindo o espírito do nosso desenvolvimento, da seguinte forma:

Dado um caminho matricial  $\alpha : I \rightarrow M_n(\mathbb{R})$ , onde  $I$  é um intervalo aberto, dizemos que  $\alpha$  é diferenciável se o caminho vetorial  $vec \circ \alpha : I \rightarrow \mathbb{R}^{n^2}$  for diferenciável e, neste caso, definimos a derivada de  $\alpha$  como sendo matriz

$$\alpha'(t) = vec^{-1} \circ \frac{d}{dt}(vec \circ \alpha)(t).$$

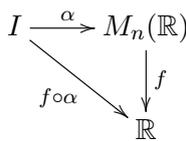


Podemos agora unificar o que desenvolvemos para estabelecer a seguinte *regra da cadeia*, de forma nativa:

**Teorema 2.3.** *Sejam  $f$  uma função real de variável matricial, diferenciável em um ponto  $P \in M_n(\mathbb{R})$ , e  $\alpha : I \rightarrow Dom f$  um caminho diferenciável em  $t$ , com  $\alpha(t) = P$ , onde  $I \subset \mathbb{R}$  é um intervalo aberto. Então a função real de variável real  $f \circ \alpha$  é diferenciável em  $t$  e*

$$\frac{d}{dt} f \circ \alpha(t) = tr[\alpha'(t)^T \frac{\partial f}{\partial X}(vec \circ \alpha(t))]. \tag{2.1}$$

*Demonstração.* Pela definição que demos,  $f$  ser diferenciável em  $P$  significa que  $f \circ vec^{-1}$  é diferenciável em  $vec(P)$ . Sendo assim, pelo teorema da regra da cadeia em  $\mathbb{R}^{n^2}$ , a função  $f \circ \alpha = f \circ vec^{-1} \circ vec \circ \alpha$  é diferenciável em  $t$  e  $\frac{d}{dt} f \circ \alpha(t) = \nabla f \circ vec^{-1}(vec(P)) \cdot vec(\frac{d}{dt} \alpha(t))$ , onde “ $\cdot$ ” denota o produto escalar de  $\mathbb{R}^{n^2}$ .



Agora, basta escrevermos esta expressão de forma nativa. Isto é feito facilmente ao considerarmos o lema 2.1 para o produto escalar e a definição de gradiente matricial, gerando  $\frac{d}{dt} f \circ \alpha(t) = tr[\frac{d}{dt} \alpha(t)^T \frac{\partial f}{\partial X}(vec \circ \alpha(t))]$ , como queríamos. □

### 3 Resultados e discussão

Para exemplificar o uso das ideias desenvolvidas aqui, voltemos ao DMTB na função de energia vista na equação (1.2).

O gradiente desta função real de variável matricial pode ser facilmente obtido com o uso da linearidade do operador traço e da proposição 2.2. Temos assim que

$$\nabla \Omega(\rho) = 3((H - \mu I)\rho + \rho(H - \mu I))^t \rho^2 - 2((H - \mu I) + \rho(H - \mu I)\rho + \rho^2(H - \mu I))^t. \tag{3.1}$$

#### 3.1 Adaptação ao Gradiente Conjugado

Métodos baseados no NLCG tem sido aplicados em Modelagem Matemática e Computacional de fenômenos da Física da Matéria Condensada desde, pelo menos, a década de 80 do século passado. Sem entrar em muitos detalhes, um método de NLCG resolve o problema de minimização irrestrita de uma função diferenciável  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  construindo uma sequência de vetores pela seguinte fórmula de recorrência

$$\vec{x}_{k+1} = \vec{x}_k + \alpha_k \vec{d}, \tag{3.2} \quad \vec{d}_k = \begin{cases} -\nabla f(\vec{x}_k), & k = 1 \\ -\nabla f(\vec{x}_k) + \beta_k \vec{d}_{k-1}, & k \geq 2, \end{cases} \tag{3.3}$$

onde o escalar  $\alpha_k$  é chamado de *comprimento do passo  $k$* , e o vetor  $\vec{d}_k$  é chamado de *direção de busca do passo  $k$* , e  $\beta_k$  são um escalares reais. É importante observar que existem várias propostas (não equivalentes) para o escalar  $\beta_k$  reportadas na literatura especializada em Otimização Numérica. Cada proposta gera um método NLCG diferente. Por exemplo, na proposta de Polak-Ribière [6] temos que  $\beta_k = \frac{d_k \cdot (d_k - d_{k-1})}{d_k \cdot d_k}$ .

A escolha do passo  $\alpha_k$  pode ser feita a partir de buscas unidirecionais exatas ou inexatas. Na proposta de busca exata, a estratégia é escolher o passo  $\alpha$  como sendo  $\min_t f(\vec{x} + t\vec{d})$ . Com isso, as buscas exatas são, elas mesmas, problemas de minimização que, em geral, podem se tornar bastante sofisticados.

A adaptação para o funcional de energia,  $\Omega$ , é feita, de forma não explícita, via a nossa vetorização de matrizes. Os produtos escalares são substituídos pelo traço do produto de matrizes, devido ao lema 2.1. Já o gradiente do campo  $f$  é substituído pela equação 3.1 e o vetor  $\vec{x}$  procurado é a vetorização da matriz densidade  $\rho$ .

### 3.2 A Busca Unidirecional Exata

Como dissemos, em processos de busca exata,  $\alpha$  deve ser escolhido de modo a produzir o ponto (matricial) da reta que passa por  $\rho$  com direção  $D$  onde  $\Omega$  assume o menor valor possível. Para isso, fixadas as matrizes  $\rho$  e  $D$ , deve-se encontrar os pontos críticos da função real, de uma variável real,  $\Omega(\rho + \alpha D)$ . A derivada desta função pode ser calculada de forma simples usando a regra da cadeia desenvolvida aqui (equação (2.1)) e o fato de que, além do operador traço matricial ser linear, o traço do produto de matrizes independe da ordem em que o produto é realizado. Assim, obtem-se que esta derivada é um polinômio de grau dois em  $\alpha$ , dado por  $\frac{d}{d\alpha}\Omega(\rho + \alpha D) = A\alpha^2 + B\alpha + C$ , onde  $A = -6tr(D^3(H - \mu I))$ ,  $B = 6tr(D^2(H - \mu I)) - 12tr(D^2\rho(H - \mu I))$  e  $C = tr(D\rho(H - \mu I)) - 6tr(D\rho^2(H - \mu I))$ . Os pontos críticos procurados são então os zeros deste polinômio. Casos em que este polinômio gere raízes complexas devem ser tratados com buscas inexatas, como pode ser visto em [1].

### 3.3 O Algoritmo

No algoritmo 19 temos um pseudocódigo, baseado na metodologia do NLCG, para a minimização do funcional  $\Omega(\rho)$ .

---

**Algoritmo 1:** NLCG PARA O FUNCIONAL DE ENERGIA (POLAK-RIBIÈRE-TYPE)

---

**Entrada:**  $\Omega, \nabla\Omega, \rho_0, tol, it_{max}$  {funcional de energia, gradiente do funcional, estimativa inicial para a matriz densidade e número máximo de iterações }

**Saída:** matriz densidade eletrônica  $\rho$  para o estado fundamental

```

1 início
2    $it \leftarrow 0$ ;
3    $D_0 \leftarrow -\nabla\Omega(\rho_0)$ ;
4    $\rho \leftarrow \rho_0$ ;
5    $D \leftarrow D_0$ ;
6    $G \leftarrow D_0$ ;
7   enquanto  $\|G\| > tol$  e  $it < it_{max}$  faça
8      $G_0 \leftarrow \nabla\Omega(\rho_0)$ ;
9     Determine  $\alpha$  usando uma busca unidirecional
10     $\rho \leftarrow \rho + \alpha D$ ;
11    Determine o Potencial Químico  $\mu$ ;
12     $G \leftarrow \nabla\Omega(\rho)$ ;
13     $\beta \leftarrow \frac{tr(G(G-G_0)^t)}{tr(G_0G_0^t)}$ ;
14     $D \leftarrow -G + \beta D$ ;
15     $\rho_0 \leftarrow \rho$ ;
16     $it \leftarrow it + 1$ ;
17 fim
18 fim
19 retorna  $\rho$ 

```

---

Note que, o que construímos foi um processo iterativo que constrói a matriz densidade eletrônica do estado fundamental. Como o nosso objetivo final é o valor da energia eletrônica, e não a matriz densidade, devido a fórmula (1.1), mais uma vez o processo de vetorização não se manifesta explicitamente.

## 4 Conclusões

Usando um isomorfismo linear existente entre os espaços vetoriais  $M_n$  e  $\mathbb{R}^{n^2}$  construímos, de forma bastante natural, o conceito de matriz gradiente de uma função real de variável matricial. Com as identificações obtidas, pudemos estabelecer alguns resultados fundamentais do Cálculo Diferencial, já consagrados em espaços  $\mathbb{R}^d$ , para as funções de uma variável matricial. Estes resultados ficaram apresentados de forma nativa em  $M_n$ , no sentido de não fazerem uso explícito do isomorfismo fixado. Por fim, aplicamos este formalismo a uma função real de variável matricial presente em uma metodologia para o cálculo da energia quanto mecânica de um cristal, conhecido por DMTB, derivando tal função e adaptando um método de Gradiente Conjugado Não Linear para a versão matricial. Devido a generalidade na abordagem apresentada, as ferramentas desenvolvidas aqui podem ser usadas para tratar outras versões do DMTB, tais como em [4], [2] e [5], bem como aplicações a quaisquer outras áreas que tratem de funções reais de uma variável matricial, sem maiores dificuldades. Especificamente para a versão apresentada em [2], recomendamos [1] como referência para a implementação computacional dos algoritmos construídos aqui, bem como os resultados da simulação para cristais de Silício. Finalmente, gostaríamos de mencionar que adotamos um enfoque que, em grande parte, não pode ser encontrado na literatura e que permite a obtenção de resultados profundos de maneira simples e com poucos pré-requisitos, tornando assim a exposição acessível a um grande número de leitores.

## Referências

- [1] F. L. S. Ferreira, J. M. B. Filho e M. A. S. M. Araújo, Gradiente Conjugado Não Linear Sobre um Ambiente “Matriz Densidade Tight -Binding Ordem-N” para o Estudo de Semicondutores, em “Anais do ERMAC - Encontro Regional de Matemática Aplicada e Computacional - UFRRJ (Seropédica) (Sociedade Brasileira de Matemática Aplicada e Computacional) pp. 136–139, 2017.
- [2] J. M. B. Filho, F. L. S. Ferreira e M. A. S. M. Araújo, Nova Proposta Matemática e Computacional para a Metodologia da Matriz Densidade Tight-Binding Ordem-N, em “Anais do ERMAC - Encontro Regional de Matemática Aplicada e Computacional - UFRRJ (Seropédica)” (Sociedade Brasileira de Matemática Aplicada e Computacional) pp. 41–44, 2017.
- [3] X.-P. Li, R. W. Nunes and R. W. D. Vanderbilt, Density-matrix electronic-structure method with linear system-size scaling, *Phys. Rev. B*, 47 (1993) 10891-10894.
- [4] J. M. Millam and G. E. Scuseria, Liner scaling density matrix, *J. Chem. Phys.*, 106 (1997) 5569–5577.
- [5] R. W. Nunes and D. Vanderbilt, Generalization of the density-matrix method to a nonorthogonal basis, *Phys. Rev. B*, 50 (1994) 17611–17614.
- [6] E. Polak and G. Ribiere, Note sur la convergence de methodes de directions conjuguées, *ESAIM: Mathematical Modelling and Numerical Analysis - Modelisation Mathematique et Analyse Numerique*, 3 (1969) 35–43.



# Números de Ramsey em Grafos Multipartidos para 4-Ciclos e Estrelas

Claudia J. F. Gonçalves, Emerson L. Monte Carmelo, Irene N. Nakaoka,

Departamento de Matemática, UEM,

87020-900, Maringá, PR

E-mail: claudiajulianafg@gmail.com, elmcarmelo@uem.br, innakaoka@uem.br,

**Resumo:** Os números de Ramsey multipartidos por classes foram introduzidos em 2004, generalizando os célebres números de Ramsey. Denote por  $C_4$  o 4-ciclo e seja  $K_{1,n}$  a estrela com  $n+1$  vértices. Neste trabalho, investigamos limitantes em  $C_4 - K_{1,n}$  números de Ramsey multipartidos por classes. São exploradas relações entre estes números e os  $C_4 - K_{1,n}$  números de Ramsey generalizados. Então, várias classes quase ótimas ou exatas são derivadas como aplicações.

**Palavras-chave:** Número de Ramsey, Grafo Multipartido, 4-Ciclo, Estrela

## 1 Introdução

Começamos introduzindo alguns conceitos. Um grafo (*simples*) é um par  $G = (V(G), E(G))$ , em que  $V(G)$  é um conjunto finito não-vazio e  $E(G)$  é uma família de subconjuntos de  $V(G)$  de cardinalidade 2. Os elementos de  $V(G)$  são denominados *vértices* e os elementos de  $E(G)$  de *arestas*. Quando não houver confusão, escrevemos simplesmente  $V$  e  $E$  no lugar de  $V(G)$  e  $E(G)$ , respectivamente. Dois vértices  $u$  e  $v$  são *adjacentes* (ou *vizinhos*) se  $\{u, v\} \in E(G)$ . Um grafo completo com  $n$  vértices, denotado por  $K_n$ , é um grafo em que dois vértices quaisquer são adjacentes (veja a Figura 1). Nas Figuras 2 e 3 temos mais dois exemplos de grafos sendo, respectivamente, um  $C_4$  (4-ciclo) e uma estrela  $K_{1,4}$ . Dados  $c$  e  $s$  inteiros positivos, denotamos por  $K_{c \times s}$  um grafo multipartido completo, cujo conjunto de vértices pode ser particionado em  $c$  classes, com  $s$  vértices em cada, de modo que vértices de uma mesma classe não são adjacentes, e de classes distintas são sempre adjacentes (por exemplo, a Figura 4 ilustra um  $K_{3 \times 2}$ ).



Figura 1:  $K_4$



Figura 2:  $C_4$



Figura 3:  $K_{1,4}$

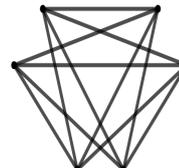


Figura 4:  $K_{3 \times 2}$

Um dos problemas mais estudado em Combinatória consiste em determinar o número de Ramsey para grafos, definido a seguir. Sejam  $G_1$  e  $G_2$  grafos simples. O número de Ramsey  $r(G_1; G_2)$  é o menor inteiro positivo  $c$  tal que, dado qualquer subgrafo  $H$  do grafo completo  $K_c$  com  $c$  vértices, tem-se que  $H$  contém uma cópia de  $G_1$  ou existe uma cópia de  $G_2$  no complementar de  $H$  (relativo a  $K_c$ ). Segue do teorema clássico de Ramsey [10] que  $r(G_1; G_2)$  é finito para cada par de grafos simples, mas seu valor exato é conhecido apenas em alguns casos especiais. O número  $r(G_1; G_2)$  também é chamado de número de Ramsey generalizado.

Os mais antigos e mais famosos exemplos de números de Ramsey são os que envolvem grafos completos, a saber,  $r(k, l) = r(K_k; K_l)$ . Esses números ficaram conhecidos como *números de Ramsey clássicos*. Desde seu surgimento em 1930, poucos valores exatos de  $r(k, l)$  foram determinados devido a dificuldade do assunto. De acordo com o “survey” [9] que é atualizado

periodicamente, as estimativas conhecidas sobre  $r(k, l)$ , para  $k, l \leq 10$ , podem ser vistas na Tabela 1.

$a \setminus b$	3	4	5	6	7	8	9	10
3	6	9	14	18	23	28	36	40-42
4		18	25	36-41	49-61	59-84	73-115	92-149
5			43-48	58-87	80-143	101-216	133-316	149-442
6				102-165	115-298	134-495	183-780	204-1171
7					205-540	217-1031	252-1713	292-2826
8						282-1870	329-3583	343-6090
9							565-6588	581-12677
10								798-23556

Tabela 1: Estimativas do número  $r(k, l)$

Devido às várias aplicações (geometria computacional, teoria da informação, algoritmos invariantes, análise harmônica, espaços métricos, Teoria dos Números, por exemplo), muitos pesquisadores têm estudado extensivamente a Teoria de Ramsey e suas conexões com muitas áreas da matemática e da ciência da computação (álgebra linear, Teoria dos Números, Teoria dos Grupos, algoritmos, estruturas discretas como geometrias finitas). Essas e outras aplicações da Teoria de Ramsey são dadas em [13]. Indicamos o livro [6] para uma visão geral sobre o tema.

Além do estudo dos números de Ramsey para  $G_1$  e  $G_2$  grafos completos, muitos outros tipos de grafos foram abordados, dentre os quais destacamos os ciclos, caminhos, árvores, grafos bipartidos, por exemplo. Veja o trabalho [9] para mais exemplos. Um desses números nos chama mais atenção. Falamos mais sobre tal número nos parágrafos seguintes.

Denote por  $C_4$  o 4-ciclo e por  $K_{1,n}$  a estrela com  $n + 1$  vértices. O estudo sistemático dos números  $r(n) := r(C_4; K_{1,n})$  foi iniciado por Parsons [11] em 1975, que calculou o limitante superior geral  $r(n) \leq n + \lceil \sqrt{n} \rceil + 1$  e computou as classes exatas  $r(n) = n + q + 1$  para  $n = q^2$  ou  $n = q^2 + 1$ , onde  $q$  denota uma potência de primo. Essencialmente, esses limitantes inferiores são derivados de grafos polaridades em planos projetivos.

Assim como para os números  $r(k, l)$ , a determinação de valores exatos para  $r(n)$  também é um problema difícil. Diversas ferramentas foram utilizadas a fim de contornar essa dificuldade e apresentar limitantes inferiores para  $r(n)$ . Com base em subgrafos de grafos polaridade, Parsons [12] obteve as seguintes classes exatas: para uma potência de primo ímpar  $q$  e qualquer  $k$  no intervalo  $1 \leq k \leq (q + 3)/4$ ,  $r(q^2 - 2k) = q^2 + q - 2k + 1$ ; se  $q$  é uma potência de 2, então  $r(q^2 - k) = q^2 + q - k + 1$  para qualquer  $k$  tal que  $1 \leq k \leq q - 1$  ou  $k = q + 1$ . Mais classes exatas foram apresentadas posteriormente por outros pesquisadores usando variadas ferramentas para a obtenção dos limitantes inferiores, tais como grafos polaridade a partir de uma família de quadrados latinos e grafos a partir de corpos finitos.

Em 2004, Burger et al. [2] introduziram uma generalização do número de Ramsey para grafos multipartidos. Mais especificamente, dado um inteiro  $s \geq 1$ , o número de Ramsey multipartido por classes  $M_s(G_1; G_2)$  é o menor inteiro positivo  $c$  com a propriedade: para qualquer subgrafo  $H$  do grafo multipartido completo  $K_{c \times s}$ ,  $H$  contém uma cópia de  $G_1$  ou existe uma cópia de  $G_2$  no complementar de  $H$  relativo a  $K_{c \times s}$ . Como  $K_{c \times 1}$  é isomorfo a  $K_c$ , esses números podem ser considerados como uma extensão dos números de Ramsey, mais formalmente,  $M_1(G_1; G_2) = r(G_1; G_2)$ .

Motivados pelo estudo de  $r(n)$  e pela possibilidade de uso das variadas ferramentas para tal estudo, neste trabalho investigamos os números  $M_s(n) := M_s(C_4; K_{1,n})$ . Assim,  $r(n)$  pode ser reformulado como  $M_1(n)$ .

Veremos que é possível estabelecer conexões entre  $r(n)$  e  $M_s(n)$ . Veremos, também, que o estudo dos números  $M_s(n)$  possibilita a utilização de várias ferramentas e técnicas da Teoria dos Números, Análise e Combinatória, por exemplo.

## 2 Preliminares

Enunciamos aqui as principais ferramentas utilizadas em nosso trabalho. Começamos definindo alguns conceitos da Teoria dos Números (para mais detalhes, indicamos o livro [8]).

Como pode ser visto na Introdução, o limitante superior do número  $r(n)$  é dado em termos da função  $\lceil x \rceil$ , chamada de *função teto*. Temos que  $\lceil x \rceil$  denota o menor inteiro  $j$  tal que  $j - 1 < x \leq j$ . De modo análogo, define-se a *função piso*  $\lfloor x \rfloor$  como sendo o maior inteiro  $i$  tal que  $i \leq x < i + 1$ . São funções muito comuns na Teoria de Ramsey.

Outro conceito é o de divisibilidade entre números inteiros. Dados dois inteiros  $a$  e  $b$ , dizemos que  $a$  divide  $b$ , e escrevemos  $a|b$ , quando existe um inteiro  $q$  tal que  $b = qa$ .

Dado um inteiro  $s \geq 1$ , lembremos que a função phi de Euler  $\phi(s)$  é definida como o número de inteiros  $a$  no intervalo  $1 \leq a \leq s$  para os quais  $\text{mdc}(a, s) = 1$ , onde  $\text{mdc}(a, s)$  denota o máximo divisor comum de  $a$  e  $s$ . Assim, se  $s$  é um número primo, é fácil ver que  $\phi(s) = s - 1$ . Na sequência, temos um resultado clássico da Teoria dos Números que envolve essa função.

**Teorema de Euler:** Dados inteiros  $a$  e  $s > 1$  tais que  $\text{mdc}(a, s) = 1$ , temos que  $s|(a^{\phi(s)} - 1)$ .

Um resultado simples mas que possui várias aplicações na resolução de problemas em Combinatória é o *princípio da casa dos pombos*, e recebe este nome pois uma maneira de enunciá-lo é a seguinte:

“Suponha que  $m + 1$  pombos vivem em  $m$  casas. Então, existe ao menos uma casa com pelo menos 2 pombos vivendo nela”.

Ferramentas da Análise também são utilizadas no estudo dos números de Ramsey. Em particular, destacamos as funções convexas. Nesse contexto, vale a seguinte desigualdade.

**Desigualdade de Jensen:** Sejam  $\lambda_1, \dots, \lambda_k \geq 0$  números reais tais que  $\lambda_1 + \dots + \lambda_k = 1$ , onde  $k \geq 2$ . Se  $f : \mathbb{R}^+ \rightarrow \mathbb{R}$  é uma função convexa, então para todos  $x_1, \dots, x_k \in \mathbb{R}^+$  vale

$$f(\lambda_1 x_1 + \dots + \lambda_k x_k) \leq \lambda_1 f(x_1) + \dots + \lambda_k f(x_k).$$

Sabendo-se que a função binomial  $f(x) = \binom{x}{2}$  é convexa, podemos aplicar a desigualdade de Jensen considerando  $\lambda_i = 1/k$ ,  $1 \leq i \leq k$ . Logo,

$$\sum_{i=1}^k \binom{x_i}{2} \geq k \binom{(\sum_{i=1}^k x_i)/k}{2}. \quad (1)$$

A desigualdade de Jensen e a função binomial são ótimas ferramentas que compõem argumentos de densidade. Esse tipo de argumento será retomado na próxima seção.

Por conveniência, dispomos aqui mais algumas notações da Teoria de Grafos que usamos ao longo do trabalho. Seja  $G = (V(G), E(G))$  um grafo. Dado  $W \subseteq V(G)$ ,  $G[W]$  representa o subgrafo de  $G$  induzido por  $W$ . Como usual, a vizinhança de um vértice  $v$  é representada por  $N_G(v) = \{u \in V(G) : \{u, v\} \in E(G)\}$  com grau  $d_G(v) = |N_G(v)|$ . A notação simplificada  $H \subseteq G$  indica que  $H$  é um subgrafo de  $G$ , enquanto que  $H \not\subseteq G$  significa que  $G$  não contém uma cópia de  $H$ . Referimos ao livro [1] para informações adicionais sobre Teoria de Grafos.

## 3 Um limitante superior

Argumentos de densidade têm sido um poderoso método na exploração de problemas da Teoria Extremal dos Grafos. Usando este método, Parsons [11] obteve o limitante superior geral  $r(n) \leq n + \lceil \sqrt{n} \rceil + 1$ . Uma adaptação para grafos multipartidos pode estender o limitante de Parsons, exceto para o caso em que  $n$  é um número quadrado, como é estabelecido na seguinte afirmação que provamos em [5, Proposição 3.1].

**Proposição 1:** Para inteiros  $s \geq 1$  e  $n \geq 2$ , seja  $t = n + s - 1$ . Valem os seguintes limitantes superiores:

$$M_s(n) \leq \begin{cases} \frac{n + \sqrt{t}}{s} + 2, & \text{se } \sqrt{t} \in \mathbb{Z} \text{ e } s \mid (n + \sqrt{t}); \\ \left\lceil \frac{n + \sqrt{t}}{s} \right\rceil + 1, & \text{caso contrário.} \end{cases}$$

Apresentamos a ideia da demonstração a partir de um exemplo. Vamos mostrar que  $c = 4$  é um limitante superior do número  $M_2(3)$ , isto é,  $M_2(3) \leq 4$ . Para isso, seja  $H = (V, E)$  um subgrafo qualquer de  $K_{4 \times 2}$  e seja  $\bar{H}$ , o complementar de  $H$  relativo a  $K_{4 \times 2}$ . Se  $K_{1,3} \subseteq \bar{H}$ , não há o que fazer. Assim, suponha que  $K_{1,3} \not\subseteq \bar{H}$ . Então,  $d_{\bar{H}}(v) \leq 2$  para todo  $v \in V(K_{4 \times 2})$ . Como  $d_H(v) + d_{\bar{H}}(v) = (c - 1)s = 6$ , então vale a desigualdade

$$d_H(v) \geq 4 \tag{2}$$

para todo  $v \in V$ . Para verificar que  $H$  contém uma cópia de  $C_4$ , o argumento de densidade é essencialmente baseado na afirmação a seguir. Seja  $\alpha$  o número de cópias de  $K_{1,2}$  em  $H$ . Observe que  $K_{4 \times 2}$  possui 8 vértices.

*Afirmção 1:* Se  $\alpha > \binom{8}{2} = 28$ , então existe uma cópia de  $C_4$  em  $H$ .

De fato, note que uma cópia de  $K_{1,2}$  em  $H$  pode ser representada pelo par ordenado  $(v, \{u, w\})$  tal que  $u \neq w$  e  $v$  é adjacente a ambos  $u$  e  $w$  em  $H$  (veja a Figura 5). Se  $\alpha > \binom{8}{2}$ , o princípio da casa dos pombos afirma que ao menos um par de vértices  $u \neq w$  tem mais de um vizinho em comum, formando uma cópia de  $C_4$  em  $H$  (veja a Figura 6).

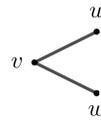


Figura 5:  $K_{1,2}$

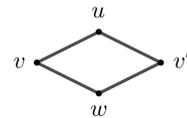


Figura 6:  $C_4$

Vamos determinar  $\alpha$ . Aplicando as desigualdades (1) e (2), obtemos

$$\alpha = \sum_{v \in V} \binom{d_H(v)}{2} \geq |V| \left( \frac{\sum_{v \in V} d_H(v)}{2} \right) \geq 8 \binom{4}{2} = 48.$$

As últimas desigualdades e a Afirmação 1 garantem que  $H$  contém uma cópia de  $C_4$  e, portanto,  $M_2(3) \leq 4$ .

As Figuras 7 e 8 exibem, respectivamente, subgrafos  $H$  e  $\bar{H}$  de  $K_{3 \times 2}$ , onde  $\bar{H}$  é o complementar de  $H$  relativo a  $K_{3 \times 2}$ . Observe que  $C_4 \not\subseteq H$  e  $K_{1,3} \not\subseteq \bar{H}$ , logo  $M_2(3) > 3$ . Portanto, temos  $M_2(3) = 4$ , o que mostra que o limitante superior dado nessa seção é atingido.

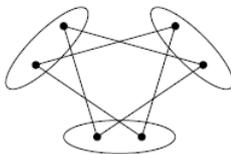


Figura 7:  $H \subseteq K_{3 \times 2}$

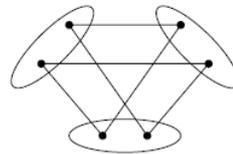


Figura 8:  $\bar{H} \subseteq K_{3 \times 2}$

## 4 Propriedade de crescimento

Consideremos, inicialmente, o número  $r(n)$ . Sabemos que a relação  $r(n) \leq r(n + 1)$  é verdadeira para todo inteiro  $n \geq 2$ , pois  $K_{1,n} \subseteq K_{1,n+1}$ . Em [3], Burr et al. questionaram se a desigualdade  $r(n + 1) \leq r(n) + 2$  é verdadeira para todo inteiro  $n \geq 2$ . Em 1997, G. Chen [4] mostrou que tal desigualdade é satisfeita para todo  $n \geq 2$ .

Pensando de modo análogo para o número  $M_s(n)$ , uma questão natural surge: qual é a relação entre  $M_s(n)$  e  $M_s(n+1)$ , além da trivial  $M_s(n) \leq M_s(n+1)$ ? Apresentamos uma resposta para tal questão na proposição abaixo, a qual generaliza o resultado de Chen.

**Proposição 2:** Para inteiros  $s \geq 2$  e  $n \geq 2$ ,  $M_s(n+1) \leq M_s(n) + 2$ .

Para a demonstração, denote por  $c = M_s(n)$  e suponha por contradição que  $M_s(n+1) > c+2$ . Logo, existe um subgrafo  $G$  de  $K_{(c+2) \times s}$  tal que  $G$  não contém uma cópia de  $C_4$ , nem  $\overline{G}$  (o complementar de  $G$  relativo a  $K_{(c+2) \times s}$ ) contém uma cópia de  $K_{1,n+1}$ . Escreva

$$V = V(K_{(c+2) \times s}) = V_1 \cup \dots \cup V_c \cup V_{c+1} \cup V_{c+2}.$$

Vamos analisar o subgrafo induzido  $H = G[V_1 \cup \dots \cup V_c]$  de  $G$ , obtido deletando-se os vértices de  $V_{c+1} \cup V_{c+2}$ . A escolha de  $c$  implica que  $C_4 \subseteq H$  ou  $K_{1,n} \subseteq \overline{H}$ . Como  $C_4 \subseteq H$ , o grafo  $\overline{G}$  contém uma estrela  $K_{1,n}$ . Denote por  $u_1$  um vértice em  $\overline{G}$  de grau  $n$ . Sem perda de generalidade, suponha  $u_1 \in V_1$ . Um argumento análogo mostra que existe outra estrela  $K_{1,n}$  em  $\overline{G}[V \setminus (V_1 \cup V_{c+2})]$ , cujo vértice de grau  $n$  é denotado por  $u_2$ . Tais estrelas estão contidas em  $\overline{G}[V \setminus (V_{c+1} \cup V_{c+2})]$  e  $\overline{G}[V \setminus (V_1 \cup V_{c+2})]$ , respectivamente, e  $K_{1,n+1} \not\subseteq \overline{G}$ , assim

$$V_{c+1} \cup V_{c+2} \subseteq N_G(u_1) \quad \text{e} \quad V_1 \cup V_{c+2} \subseteq N_G(u_2). \quad (3)$$

Sejam  $x, y \in V_{c+2}$ . Por (3), os vértices  $u_1, x, u_2$  e  $y$  induzem um  $C_4$  em  $G$ , uma contradição. Portanto,  $M_s(n+1) \leq c+2$ .

## 5 Uma classe exata

Uma abordagem computacional mostra que  $M_2(4) \leq 4$ , enquanto que o limitante superior da Seção 3 produz apenas  $M_2(4) \leq 5$ . Contudo, veremos que o limitante superior pode ser ótimo para várias classes aqui apresentadas. Em particular, se os parâmetros  $s$  e  $n$  satisfazem adequadas condições aritméticas, então o limitante superior é atingido. Mais especificamente, provamos em [5, Proposição 3.2] o seguinte resultado:

**Proposição 3:** Dados inteiros  $s, n \geq 2$ , seja  $t = n + s - 1$ . Suponha que  $(n + \sqrt{t})/s$  não é um inteiro e  $\lceil (n + \sqrt{t})/s \rceil \leq \lfloor (n+1)/s \rfloor + 1$ . Então,

$$M_s(n) = \left\lceil \frac{n + \sqrt{t}}{s} \right\rceil + 1.$$

Em quais situações as hipóteses da afirmação anterior se aplicam? Vejamos um exemplo.

**Exemplo:** Sejam  $s = 4$  e  $n = 4k - 1$ ,  $1 \leq k \leq 5$ , para os quais temos  $t = n + s - 1 = 4k + 2$ . Observe que, para cada  $k$  no intervalo citado,  $n, s$  e  $t$  satisfazem as condições da Proposição 3. Assim, obtemos os seguintes valores exatos:

$$M_4(3) = 3, M_4(7) = 4, M_4(11) = 5, M_4(15) = 6 \quad \text{e} \quad M_4(19) = 7.$$

O próximo resultado mostra que, se  $s$  e  $n$  satisfazem adequadas condições aritméticas, então valem as hipóteses da última proposição e uma classe exata do número  $M_s(n)$  é obtida.

**Teorema 4:** Dado  $s \geq 2$ , seja  $a$  um inteiro tal que  $-1 \leq a \leq \lfloor s/2 \rfloor - 1$ . Para todo inteiro  $k$  no intervalo  $1 \leq k \leq s - (2a + 1)$ ,

$$M_s(ks + a) = k + 2.$$

Apresentamos a prova completa em [5, Teorema A]. Aqui, damos uma ideia da demonstração. Fazendo  $n = ks + a$  e  $t = n + s - 1$ , com  $a$  e  $k$  nos respectivos intervalos indicados, vale a igualdade  $\lfloor (n+1)/s \rfloor + 1 = k + 1$  e, também, a desigualdade  $0 < a + \sqrt{(k+1)s + a - 1} < s$ . Assim, obtemos

$$\left\lceil \frac{n + \sqrt{t}}{s} \right\rceil = k + \left\lceil \frac{a + \sqrt{(k+1)s + a - 1}}{s} \right\rceil = k + 1 = \left\lfloor \frac{n+1}{s} \right\rfloor + 1,$$

e o resultado segue como uma aplicação da Proposição 3.

**Exemplo:** Usando a Proposição 4, vamos determinar as demais classes não contempladas no último exemplo no qual  $s = 4$ . De acordo com o resultado anterior, para cada  $a$  e  $k$  nos respectivos intervalos  $0 \leq a \leq 1$  e  $1 \leq k \leq -2a + 3$  (os valores calculados no último exemplo correspondem a  $a = -1$ ), obtemos os valores exatos

$$M_4(4) = 3, M_4(8) = 4, M_4(12) = 5 \text{ e } M_4(5) = 3.$$

## 6 Limitantes inferiores a partir de números de Ramsey

Para  $s$  fixo, o valor de  $n = ks + a$  no Teorema 4 pertence ao intervalo  $[s - 1; s^2 + s]$ . Para estender o intervalo do parâmetro  $n$ , aplicamos a seguinte conexão entre  $r(n)$  e  $M_s(n)$ , provada em [7] por Magnant e Yusko.

**Proposição 6:** Para quaisquer inteiros  $s \geq 1$  e  $n \geq 2$ ,  $\left\lfloor \frac{r(n) - 1}{s} \right\rfloor + 1 \leq M_s(n)$ .

Assim, a partir do conhecimento de classes exatas de  $r(n)$  juntamente com a proposição anterior e o limitante superior da Seção 3, é possível estimar valores para  $M_s(n)$ . Para esse propósito, listamos apenas algumas classes mais conhecidas sobre  $r(n)$ .

**Teorema 5:** Seja  $q$  uma potência de primo. Valem as seguintes classes exatas:

- (1) [11, Teoremas 1 e 2]  $r(q^2 + k) = q^2 + q + 1 + k$ , em que  $k \in \{0, 1\}$ ;
- (2) [12, Teorema 3]  $r(q^2 - q - 1) = q^2$ , se  $q$  é par.

Vamos discutir o impacto do resultado anterior. Para ilustrar a primeira consequência, certos números  $M_s(n)$  podem ser obtidos com um erro de, no máximo, uma unidade.

**Proposição 7:** Seja  $k \in \{0, 1\}$ . Dados um inteiro  $s \geq 2$  e uma potência de primo  $q \geq 4$  tais que  $s \leq 2(q + 1) - k$ , denote  $L_k = (q^2 + k + \sqrt{q^2 + k + s - 1})/s$ . Se  $L_k$  não é um inteiro, então  $\lceil L_k \rceil \leq M_s(q^2 + k) \leq \lfloor L_k \rfloor + 1$ .

O limitante inferior segue da Proposição 6 aplicada à classe  $r(q^2 + k) = q^2 + q + 1 + k$  (Item (1) do Teorema 5), enquanto que o limitante superior é obtido pela Proposição 1. Damos uma prova completa em [5, Proposição 4.3].

O resultado anterior reflete que ou a Proposição 6 ou a Proposição 1 é ótima para tais parâmetros. Isso nos leva a pensar se algo similar ocorre quando consideramos a classe exata  $r(q^2 - q - 1) = q^2$ , onde  $q$  é uma potência de 2 (Teorema 5 (2)). A Tabela 2 faz uma comparação entre os limitantes inferior  $l$  e superior  $L$  dados pelas Proposições 6 e 1, respectivamente.

$s \setminus n = q^2 - q - 1$	11	55	239	991	4031
2	8-9	32-33	128-129	512-513	2048-2049
3	6-6	22-22	86-86	342-342	1366-1366

Tabela 2:  $l \leq M_s(n) \leq L$  para  $s = 2, s = 3$  e  $q \in \{4, 8, 16, 32, 64\}$

Embora a diferença não exceda 1 quando  $s = 2$ , para  $s = 3$  temos  $l = L$ . Fazendo os cálculos para outros valores de  $s$  é possível perceber que, se  $s$  é ímpar, a igualdade  $l = L$  se verifica em vários casos, como é dado no resultado a seguir.

**Teorema 8:** Seja  $s$  um inteiro ímpar, com  $s \geq 3$ , e suponha que  $m$  é um inteiro positivo tal que  $s|(4^m - 1)$  e  $2^m \geq s$ . Para todo inteiro positivo  $k$ ,

$$M_s(2^{2mk} - 2^{mk} - 1) = \frac{2^{2mk} - 1}{s} + 1.$$

Damos uma prova para o Teorema 8 em [5, Teorema B].

Considere  $s = 3$ . Como  $\text{mdc}(4, 3) = 1$ , o Teorema de Euler garante que  $3|(4^{\phi(3)} - 1)$ . A desigualdade  $2^{\phi(3)} \geq 3$  também é satisfeita e o Teorema 8 pode ser aplicado fazendo-se  $m = \phi(3)$ . De um modo geral, demonstramos a seguinte afirmação em [5, Corolário 4.4].

**Corolário 9:** Seja  $s \geq 3$  um inteiro ímpar. Para qualquer inteiro positivo  $k$ ,

$$M_s(2^{2\phi(s)k} - 2^{\phi(s)k} - 1) = \frac{2^{2\phi(s)k} - 1}{s} + 1.$$

Nessa seção, apresentamos apenas algumas estimativas para  $M_s(n)$  obtidas via limitantes inferiores a partir dos números de Ramsey  $r(n)$ . Em [5, Teorema C], obtivemos mais estimativas através de outras classes exatas de  $r(n)$  não listadas no Teorema 5.

**Agradecimentos:** O primeiro autor foi parcialmente suportado pela Capes. O segundo autor é parcialmente suportado pelo CNPq/MCT: 311703/2016-0.

## Referências

- [1] B. Bollobás, “Modern Graph Theory”, Springer-Verlag, New York, 1998.
- [2] A.P. Burger, J.H. van Vuuren, Ramsey numbers in complete balanced multipartite graphs. Part I: Set numbers, *Discrete Math.*, 283 (2004) 37-43.
- [3] S. Burr, P. Erdős, R.J. Faudree, C.C. Rousseau, R.H. Schelp, Some complete bipartite graph-tree Ramsey numbers, *Ann. Discrete Math.*, 41 (1989) 79-90.
- [4] G. Chen, A result on  $C_4$ -star Ramsey numbers, *Discrete Math.*, 163 (1997) 243-246.
- [5] C.J.F Gonçalves, E.L. Monte Carmelo, I.N. Nakaoka, Projective planes and set multipartite Ramsey numbers for  $C_4$  versus star, arXiv:1710.06383 [math.CO].
- [6] R. L. Graham, B. L. Rothschild, J. H. Spencer, Ramsey Theory, John Wiley & Sons, New York, 1990.
- [7] C. Magnant, A. Yusko, Ramsey set numbers in balanced complete multipartite graphs, *J. Combin. Math. Combin. Comput.*, 77 (2011) 187-196.
- [8] F. C. P. Milies, S. P. Coelho, Números: Uma Introdução à Matemática, EdUSP, São Paulo, 2003.
- [9] S. P. Radziszowski, Small Ramsey numbers, *The Electronic Journal of Combinatorics*, 15 (2017).
- [10] F. P. Ramsey, On a problem of formal logic, *London Math. Soc.*, 30 (1930) 264-286.
- [11] T.D. Parsons, Ramsey graphs and block designs I, *Trans. Amer. Math. Soc.*, 209 (1975) 33-44.
- [12] T.D. Parsons, Graphs from projective planes, *Aequ. Math.*, 14 (1976) 167-189.
- [13] V. Rosta, Ramsey theory applications, *The electronic journal of combinatorics*, 1000 (2004) 13-7.



# Semântica relacional para a lógica proposicional do plausível

**Tiago Augusto dos Santos Boza**

UFSC - Colégio de Aplicação  
88040-900, Florianópolis, SC  
Campus Trindade  
E-mail: boza.tiago@ufsc.br

**Hércules de Araujo Feitosa**

UNESP - Departamento de Matemática  
17033-360, Bauru, SP  
Campus Bauru  
E-mail: haf@fc.unesp.br

15 de Abril de 2019

## Resumo

Esse trabalho tem como objetivo apresentar uma semântica de vizinhança para uma lógica modal, a saber, a lógica proposicional do plausível. A lógica do plausível foi proposta como uma forma de tratar um quantificador não clássico, no grupo das lógicas moduladas, que tem seus modelos conectados com uma variação do conceito de espaço topológico. Esta estrutura de espaço pseudo-topológico pode ser destacada e se transformar também num modelo para uma lógica proposicional, em que os seus operadores interpretam operadores lógicos de caráter modal. Por ser uma lógica modal não normal, a lógica proposicional do plausível não admite uma semântica de Kripke, porém têm semântica de vizinhança, uma semântica relacional.

**Palavras Chave:** lógica modal, lógica proposicional do plausível, semânticas de vizinhança

## Introdução

De um modo geral, as chamadas lógicas modais inserem um novo operador não definido à partir dos operadores clássicos e, assim, estendem as lógicas clássicas.

Há também extensões modais de lógicas distintas da lógica clássica, mas o sistema aqui considerado está de acordo com este modo geral.

As usuais lógicas modais normais têm como semânticas essenciais os modelos de Kripke, ou semântica dos mundos possíveis, que resgatam boas intuições das lógicas modais normais.

Também tratamos com modelos de mundos possíveis, mas com pequenas adequações. A lógica proposicional do plausível recebeu originalmente uma semântica algébrica, mas então buscamos modelos relacionais.

O elemento original deste trabalho é a arquitetura de uma semântica relacional para a lógica proposicional do plausível.

## 1 A motivação quantificacional

A formalização lógica quantificacional dos espaços pseudo-topológicos surgiu em [6], como uma extensão conservativa da lógica clássica de primeira ordem  $\mathcal{L}$ , ou seja, inserindo um novo quantificador não definido à partir dos quantificadores clássicos, existencial ( $\exists$ ) e universal ( $\forall$ ). Para mais detalhes sobre  $\mathcal{L}$  ver [3] ou [5].

Esta lógica de primeira ordem, apresentada por [6], foi denominada de *lógica do plausível* e denotada por  $\mathcal{L}(P)$ .

A lógica estendida  $\mathcal{L}(P)$  é determinada pelos seguintes acréscimos:

- a linguagem  $L$  de  $\mathcal{L}(P)$  conta com um novo símbolo de quantificador  $P$  e sentenças do tipo  $Px \varphi(x)$  são bem formadas em  $\mathcal{L}(P)$ .

- axiomas específicos do quantificador  $P$ :

$$(A_1) Px \varphi(x) \wedge Px \psi(x) \rightarrow Px(\varphi(x) \wedge \psi(x))$$

$$(A_2) Px \varphi(x) \wedge Px \psi(x) \rightarrow Px(\varphi(x) \vee \psi(x))$$

$$(A_3) \forall x \varphi(x) \rightarrow Px \varphi(x)$$

$$(A_4) Px \varphi(x) \rightarrow \exists x \varphi(x)$$

$$(A_5) \forall x(\varphi(x) \rightarrow \psi(x)) \rightarrow (Px \varphi(x) \rightarrow Px \psi(x))$$

$$(A_6) Px \varphi(x) \rightarrow Py \varphi(y), \text{ se } y \text{ é livre para } x \text{ em } \varphi(x).$$

- regras de dedução:

$$(MP) \text{ Modus Ponens: } \varphi, \varphi \rightarrow \psi \vdash \psi$$

$$(Gen) \text{ Generalização: } \varphi \vdash \forall x \varphi(x).$$

Os demais conceitos sintáticos são como os usuais, vide [3] ou [5].

**Definição 1.1** Um espaço pseudo-topológico é um par  $(E, \Omega)$ , em que  $E$  é um conjunto não-vazio e  $\Omega$  um subconjunto de  $\mathcal{P}(E)$ . Cada membro de  $\Omega$  é denominado um aberto de  $(E, \Omega)$ , de maneira que:

$$(i) \text{ se } A, B \in \Omega, \text{ então } A \cap B \in \Omega;$$

$$(ii) \text{ se } A, B \in \Omega, \text{ então } A \cup B \in \Omega;$$

$$(iii) E \in \Omega;$$

$$(iv) \emptyset \notin \Omega.$$

**Definição 1.2** Um subconjunto  $F$  de  $E$  é fechado em  $(E, \Omega)$  quando seu complementar  $F^C$  é um aberto em  $(E, \Omega)$ .

As estruturas adequadas para  $\mathcal{L}(P)$ , estruturas do plausível, são extensões das estruturas de primeira ordem  $\mathcal{A}$ .

Assim dada uma estrutura  $\mathcal{A}$ , consideremos que o seu domínio é denotado por  $A$ . Uma estrutura do plausível, denotada por  $\mathcal{A}^\Omega$ , é determinada a partir  $\mathcal{A}$  pelo acréscimo de um espaço pseudo-topológico  $\Omega$  sobre o universo  $A$ .

A interpretação dos símbolos de relação, função e constante é a mesma de  $\mathcal{L}$  com relação à  $\mathcal{A}$ .

**Definição 1.3** A satisfação de uma sentença do tipo  $Px \varphi(x)$  em  $\mathcal{A}^\Omega$  e definida indutivamente por:

- se  $\varphi$  é uma fórmula cujas variáveis livres estão em  $\{x\} \cup \{y_1, \dots, y_n\}$  e  $\bar{a} = (a_1, \dots, a_n)$  é uma sequência de elementos de  $A$ , então:

$$\mathcal{A}^\Omega \models Px \varphi[x, \bar{a}] \Leftrightarrow \{b \in A : \mathcal{A}^\Omega \models [b, \bar{a}]\} \in \Omega.$$

Para a sentença  $Px \varphi(x)$ , tem-se  $\mathcal{A}^\Omega \models Px \varphi(x) \Leftrightarrow \{a \in A : \mathcal{A}^\Omega \models \varphi(a)\} \in \Omega$ .

As outras noções semânticas são como as usuais, novamente vide [3] ou [5].

Grácio [6] demonstrou que as estruturas do plausível são modelos corretos e completos para  $\mathcal{L}(P)$ .

## 1.1 Uma formalização axiomática e proposicional

Com ênfase apenas nos espaços pseudo-topológicos apresentamos a versão proposicional da lógica do plausível, conforme [4].

Denotamos esta lógica por  $\mathbb{L}(\nabla)$ . Ela estende a lógica proposicional clássica (**LPC**) na linguagem  $L = \{\neg, \wedge, \vee, \rightarrow\}$  com o acréscimo do operador unário  $\nabla$ , donde obtemos a linguagem proposicional  $L = \{\neg, \wedge, \vee, \rightarrow, \nabla\}$ .

A lógica fica determinada pelo seguinte:

- Axiomas:

$$(Ax_0) \text{ tautologias da LPC}$$

$$(Ax_1) (\nabla \varphi \wedge \nabla \psi) \rightarrow \nabla(\varphi \wedge \psi)$$

$$(Ax_2) \nabla(\varphi \vee \neg \varphi)$$

$$(Ax_3) \neg \nabla \perp.$$

- Regras de dedução:

(MP) *Modus Ponens*  
 (RV)  $\vdash \varphi \rightarrow \psi / \vdash \nabla\varphi \rightarrow \nabla\psi$ .

Diante disso,  $\varphi$  e  $\psi$  são plausíveis se, e somente se,  $\varphi \wedge \psi$  é plausível. Toda tautologia é plausível, nenhuma contradição é plausível, e se  $\varphi$  é plausível, então é plausível  $\varphi \vee \psi$ . A regra (RV) diz que se  $\varphi \rightarrow \psi$  e  $\varphi$  são plausíveis, então também  $\psi$  é plausível.

As demonstrações dos resultados abaixo podem ser encontradas em [4].

**Proposição 1.4** (i)  $\vdash \nabla\perp \rightarrow \perp$   
 (ii)  $\vdash \nabla\varphi \rightarrow \nabla(\varphi \vee \psi)$   
 (iii)  $\vdash \varphi \Rightarrow \vdash \nabla\varphi$   
 (iv)  $\vdash \varphi \rightarrow \neg\nabla\neg\varphi$   
 (v)  $\vdash \nabla\varphi \rightarrow \neg\nabla\neg\varphi$   
 (vi)  $\vdash \nabla\neg\varphi \rightarrow \neg\nabla\varphi$ .

**Proposição 1.5**  $\nabla\varphi \rightarrow \nabla(\varphi \vee \psi) \Leftrightarrow (\nabla\varphi \vee \nabla\psi) \rightarrow \nabla(\varphi \vee \psi)$ .

## 1.2 Álgebras dos espaços pseudo-topológicos

Agora, apresentamos a álgebra do plausível, que corresponde à versão algébrica da lógica da seção anterior, conforme [4].

**Definição 1.6** *Álgebra do Plausível é uma estrutura  $\mathbb{P} = (P, 0, 1, \wedge, \vee, \sim, \#)$ , de maneira que  $(P, 0, 1, \wedge, \vee, \sim)$  é uma álgebra de Boole e  $\#$  é o operador do plausível, sujeito a:*

- (a<sub>1</sub>)  $\#a \wedge \#b \leq \#(a \wedge b)$
- (a<sub>2</sub>)  $\#a \leq \#(a \vee b)$
- (a<sub>3</sub>)  $\#0 = 0$
- (a<sub>4</sub>)  $\#1 = 1$ .

**Proposição 1.7** *Se  $\mathbb{P} = (P, 0, 1, \wedge, \vee, \sim, \#)$  é uma álgebra do plausível e  $a, b \in P$ , então:*

- (i)  $a \leq b \Rightarrow \#a \leq \#b$
- (ii)  $\#a \wedge \#b = \#(a \wedge b)$
- (iii)  $\#a \vee \#b \leq \#(a \vee b)$ .

**Proposição 1.8** *Para cada álgebra do plausível  $\mathbb{P} = (P, 0, 1, \wedge, \vee, \sim, \#)$  existe um monomorfismo  $h$  de  $P$  num espaço pseudo-topológico de conjuntos definidos em  $\mathcal{P}(\mathcal{P}(P))$ .*

**Definição 1.9** *A interpretação de  $\mathbb{L}(\nabla)$  numa pseudo-topologia  $(E, \Omega)$  é uma função  $v$  com as seguintes características:*

- $v(\perp) = \emptyset$
- $v(\top) = E$
- $v(\varphi \wedge \psi) = A \cap B$
- $v(\varphi \vee \psi) = A \cup B$
- $v(\neg\varphi) = A^C$
- $v(\nabla\varphi) = \overset{\circ}{A} \in \Omega$ .

$\overset{\circ}{A}$  é o interior de  $A$ , o maior aberto contido em  $A$ .

**Proposição 1.10** *Para a interpretação  $v$  acima, temos que:*

- $\vdash \varphi \rightarrow \psi \Leftrightarrow v(\varphi) \subseteq v(\psi)$ .

## 2 Semântica Relacional para $\mathbb{L}(\nabla)$

Agora, apresentamos uma semântica relacional para a lógica proposicional do plausível e, então, a adequação entre esta estrutura semântica relacional e a axiomática introduzida anteriormente.

**Definição 2.1** Um modelo relacional  $\mathcal{M}$  para a lógica proposicional do plausível é uma terna  $\mathcal{M} = \langle W, S, V \rangle$ , em que:

- $W$  é um conjunto não-vazio de mundos possíveis;
- $S$  uma função que associa cada  $x \in W$  um conjunto de subconjuntos de  $W$ , isto é,  $S(x) \subseteq \mathcal{P}(W)$ , de tal modo que satisfaz as seguintes condições:

(i) se  $X \in S(x)$  e  $Y \in S(x)$ , então  $X \cap Y \in S(x)$ ;

(ii) se  $X \in S(x)$ , então  $X \cup Y \in S(x)$ ;

(iii)  $\emptyset \notin S(x)$ ;

(iv)  $W \in S(x)$ ;

- $V$  é uma valoração em  $W$ , isto é, uma função do conjunto das fórmulas atômicas de  $\mathbb{L}(\nabla)$  em  $\mathcal{P}(W)$ .

Agora, temos de definir condições para que as fórmulas sejam válidas nesta nova estrutura.

**Definição 2.2** O conjunto verdade de  $\varphi$  em  $\mathcal{M}$  é denotado por  $\|\varphi\|^{\mathcal{M}}$  e definido por  $\|\varphi\|^{\mathcal{M}} = \{x \in W : \models_x^{\mathcal{M}} \varphi\}$ .

**Definição 2.3** Sejam  $\mathcal{M}$  um modelo e  $x \in W$ . Uma fórmula  $\varphi$  é verdadeira no mundo  $x$ , o que é denotado por  $\models_x^{\mathcal{M}} \varphi$ , se:

- as condições para os operadores Booleanos seguem como usualmente;

-  $\models_x^{\mathcal{M}} p_i \Leftrightarrow x \in V(p_i)$ , quando  $p_i$  é uma variável proposicional;

-  $\models_x^{\mathcal{M}} \nabla\varphi \Leftrightarrow \|\varphi\|^{\mathcal{M}} \in S(x)$ .

Em geral, denotamos o conjunto verdade  $\varphi$  apenas por  $\|\varphi\|$ .

**Definição 2.4** Uma fórmula  $\varphi$  é válida num modelo  $\mathcal{M}$ , quando é verdadeira para todo  $x \in W$ . A fórmula  $\varphi$  é válida se é verdadeira em todo modelo  $\mathcal{M}$ .

Denotamos que uma fórmula  $\varphi$  é válida num modelo  $\mathcal{M}$  por  $\models^{\mathcal{M}} \varphi$ , e que  $\varphi$  é válida apenas por  $\models \varphi$ .

**Definição 2.5** Se  $\Gamma$  é um conjunto de fórmulas e  $\mathcal{M}$  é um modelo, então dizemos que  $\models^{\mathcal{M}} \Gamma$  se  $\models^{\mathcal{M}} \varphi$ , para cada  $\varphi \in \Gamma$ .

**Definição 2.6** Seja  $\Gamma \cup \{\varphi\}$  um conjunto de fórmulas. Dizemos que  $\Gamma$  implica  $\varphi$ , ou que  $\varphi$  é uma consequência semântica de  $\Gamma$ , quando para todo modelo  $\mathcal{M}$ , temos que se  $\models^{\mathcal{M}} \Gamma$ , então  $\models^{\mathcal{M}} \varphi$ .

Denotamos a consequência semântica por  $\Gamma \models \varphi$ .

**Proposição 2.7** Se  $\mathcal{M}$  é um  $\mathbb{L}(\nabla)$ -modelo e  $\varphi, \psi$  são fórmulas quaisquer, então:

(i)  $\|\neg\varphi\| = -\|\varphi\|$ ;

(ii)  $\|\varphi \wedge \psi\| = \|\varphi\| \cap \|\psi\|$ ;

(iii)  $\|\varphi \vee \psi\| = \|\varphi\| \cup \|\psi\|$ ;

(iv)  $\|\varphi \rightarrow \psi\| = -\|\varphi\| \cup \|\psi\|$ ;

(v)  $\|\varphi \leftrightarrow \psi\| = (-\|\varphi\| \cup \|\psi\|) \cap (-\|\psi\| \cup \|\varphi\|)$ ;

(vi)  $\|\nabla\varphi\| = \{x \in W : \|\varphi\| \in S(x)\}$ .

**Demonstração:** O conjunto  $-\|\varphi\|$  coincide com  $W - \|\varphi\|$ , o complemento de  $\|\varphi\|$  relativo à  $W$ .

As condições de (i) a (v) seguem imediatamente do caráter Booleano da lógica clássica e da teoria dos conjuntos.

Resta mostrarmos a condição (vi). Por definição,  $x \models \nabla\varphi$  se  $x \in \|\nabla\varphi\|$ . E isto significa, portanto, que  $\{x \in W : \|\varphi\| \in S(x)\} = \|\nabla\varphi\|$ . ■

**Proposição 2.8 (Correção)**  $\Gamma \vdash \varphi \Rightarrow \Gamma \models \varphi$ .

**Demonstração:** Seja  $\mathcal{M}$  um modelo para  $\mathbb{L}(\nabla)$ . Fazemos a prova por indução sobre o comprimento da dedução  $\Gamma \vdash \varphi$ .

Quando  $n = 1$ , tem-se que  $\varphi$  pertence à  $\Gamma$  ou  $\varphi$  é um axioma. Desse modo, façamos em duas partes:

- se  $\varphi \in \Gamma$  ou  $\varphi$  é um axioma da **LPC**, nada temos a demonstrar;

- se  $\varphi$  é um dos axiomas modais, então:

(Ax<sub>1</sub>) Se  $x \in W$  e  $\models_x^{\mathcal{M}} \nabla\varphi \wedge \nabla\psi$ , por ser um modelo Booleano,  $\models_x^{\mathcal{M}} \nabla\varphi$  e  $\models_x^{\mathcal{M}} \nabla\psi$  e, assim,  $\|\varphi\| \in S(x)$  e  $\|\psi\| \in S(x)$ . Pela condição (i) da Definição 2.1, temos  $\|\varphi\| \cap \|\psi\| \in S(x)$  e, por definição,  $\|\varphi \wedge \psi\| \in S(x)$ . Desse modo,  $\models_x^{\mathcal{M}} \nabla(\varphi \wedge \psi)$ .

(Ax<sub>2</sub>) Para todo  $x \in W$ , temos que  $\models_x^{\mathcal{M}} \nabla\top$ , pois  $\|\top\| \in S(x)$ , para todo  $x \in W$ . Logo, para todo  $x \in W$ , temos que  $\models_x^{\mathcal{M}} \nabla\top$  ou  $\models_x^{\mathcal{M}} \nabla(\varphi \vee \neg\varphi)$ .

(Ax<sub>3</sub>) Nenhum  $x \in W$  é tal que  $\models_x^{\mathcal{M}} \nabla\perp$ , pois do contrário  $\|\perp\| \in S(x)$  para algum  $x \in W$ . Logo, por ser Booleana, para todo  $x \in W$ , temos que  $\models_x^{\mathcal{M}} \neg\nabla\perp$ .

Quando  $n \leq k$ , temos como hipótese de indução que:  $\Gamma \vdash \varphi_n \Rightarrow \Gamma \models \varphi_n$ .

Assim, há três possibilidades para o passo seguinte,  $k + 1$ , da indução: (i)  $\varphi_{k+1}$  é uma premissa; (ii)  $\varphi_{k+1}$  é um esquema de axiomas; (iii)  $\varphi_{k+1}$  é deduzida a partir das regras MP ou (R $\nabla$ ).

Para os itens (i) e (ii), a justificativa é como na base da indução.

Para a regra MP, sabemos da **LPC** que ela preserva a validade. Desse modo, falta-nos analisar a regra (R $\nabla$ ).

(R $\nabla$ ) Sejam  $x \in W$ , tal que  $\models_x^{\mathcal{M}} \varphi \rightarrow \psi$ . Por ser um modelo Booleano e  $\vdash \varphi \rightarrow \psi$ , temos que  $\|\varphi\| \subseteq \|\psi\|$ . Daí, para todo  $x \in W$ , se  $\|\varphi\| \in S(x)$ , então  $\|\psi\| \in S(x)$ . Assim, ou  $\not\models_x^{\mathcal{M}} \nabla\varphi$  ou  $\models_x^{\mathcal{M}} \nabla\psi$  e, por definição, temos  $\models_x^{\mathcal{M}} \nabla\varphi \rightarrow \nabla\psi$ . ■

Agora, daremos algumas definições e mostraremos alguns resultados para, por fim, demonstrarmos a completude de nosso sistema, relativo à semântica  $\mathcal{M}$ .

**Definição 2.9** Um conjunto de fórmulas  $\Delta$  é maximalmente consistente quando  $\Delta$  é consistente e nenhuma extensão conservativa de  $\Delta$  é consistente.

A demonstração do resultado abaixo pode ser encontrado em [2].

**Proposição 2.10 (Lindenbaum)** Todo conjunto maximalmente consistente  $\Gamma$  pode ser estendido à um conjunto maximalmente consistente  $\Delta$ .

A seguir, tratamos com modelos canônicos. Consideremos, o conjunto de todos os conjuntos de fórmulas de  $\mathbb{L}(\nabla)$  que são maximalmente consistentes. Denotamos este conjunto por  $\chi$ .

**Definição 2.11** O conjunto de demonstrações de  $\varphi$  é o conjunto  $|\varphi| = \{\Gamma \in \chi : \varphi \in \Gamma\}$ .

A demonstração da proposição seguinte pode ser encontrada em [2].

**Proposição 2.12** Se  $\varphi$  e  $\psi$  são fórmulas quaisquer  $\mathbb{L}(\nabla)$ , então:

- (i)  $|\neg\varphi| = \neg|\varphi|$ ;
- (ii)  $|\varphi \wedge \psi| = |\varphi| \cap |\psi|$ ;
- (iii)  $|\varphi \vee \psi| = |\varphi| \cup |\psi|$ ;
- (iv)  $|\varphi \rightarrow \psi| = \neg|\varphi| \cup |\psi|$ ;
- (v)  $|\varphi \leftrightarrow \psi| = (\neg|\varphi| \cup |\psi|) \cap (\neg|\psi| \cup |\varphi|)$ ;
- (vi)  $|\varphi| \subseteq |\psi| \Leftrightarrow \vdash \varphi \rightarrow \psi$ ;
- (vii)  $|\varphi| = |\psi| \Leftrightarrow \vdash \varphi \leftrightarrow \psi$ .

**Definição 2.13** A estrutura  $\mathcal{M} = \langle W, S, V \rangle$  é um modelo canônico para  $\mathbb{L}(\nabla)$ , quando satisfaz as seguintes condições:

- $W = \chi$ ;
- $|\varphi| \in S(\Gamma) \Leftrightarrow \nabla\varphi \in \Gamma$ , para todo  $\Gamma \in W$ ;
- $V(p_i) = |p_i|$ , para toda variável proposicional  $p_i$ .

**Proposição 2.14** Se  $\mathcal{M}$  é um modelo canônico, então, para toda fórmula  $\varphi$  e todo  $\Gamma \in W$ , tem-se  $\Gamma \models^{\mathcal{M}} \varphi \Leftrightarrow \varphi \in \Gamma$ .

**Demonstração:** Faremos por indução sobre a complexidade das fórmulas de  $\Gamma$ , para o conjunto de operadores  $\{\neg, \wedge, \nabla\}$ .

(i) Para  $\varphi \equiv p_i$ , com  $i \in \mathbb{N}$ : temos que  $\Gamma \models^{\mathcal{M}} p_i$  se  $\Gamma \in V(p_i)$  se  $\Gamma \in |p_i|$ . Por construção de  $|p_i|$ ,  $\Gamma$  é um conjunto em  $|p_i|$  se  $p_i \in \Gamma$ .

(ii) Para  $\varphi \equiv \neg\psi$ :  $\Gamma \models^{\mathcal{M}} \neg\psi$  se  $\Gamma \not\models^{\mathcal{M}} \psi$ . Pela hipótese de indução, temos que  $\psi \notin \Gamma$ , e como  $\Gamma$  é maximalmente consistente, segue que  $\neg\psi \in \Gamma$  e, portanto,  $\varphi \in \Gamma$ .

(iii) Para  $\varphi \equiv \psi \wedge \sigma$ :  $\Gamma \models^{\mathcal{M}} \psi \wedge \sigma$  se  $\Gamma \models^{\mathcal{M}} \psi$  e  $\Gamma \models^{\mathcal{M}} \sigma$ . Assim, por hipótese de indução,  $\psi, \sigma \in \Gamma$ . Pela **LPC**,  $\psi \wedge \sigma \in \Gamma$  e, portanto,  $\varphi \in \Gamma$ .

(iv) Para  $\varphi \equiv \nabla\psi$ :  $\Gamma \models^{\mathcal{M}} \nabla\psi$  se  $\|\psi\| \in S(\Gamma)$ . Desse modo, pela hipótese de indução, para todo  $\Phi \in W$ , temos que  $\Phi \models^{\mathcal{M}} \psi$  se  $\psi \in \Phi$ , isto é,  $\|\psi\| = |\psi|$ . Assim,  $\|\psi\| \in S(\Gamma)$  se  $|\psi| \in S(\Gamma)$ . Agora, por definição de  $S(\Gamma)$ ,  $|\psi| \in S(\Gamma)$  se  $\nabla\psi \in \Gamma$  e, assim,  $\varphi \in \Gamma$ . ■

Mais algumas definições.

**Definição 2.15** A estrutura  $\mathcal{M}]$  é o menor modelo canônico quando o conjunto  $S(\Gamma)$  contém apenas conjuntos de demonstrações.

**Definição 2.16** A suplementação de  $\mathcal{M}]$  é o modelo  $\mathcal{M}]^+ = \langle W, S^+, V \rangle$ , com a seguinte condição adicional para  $S^+$ . Para todo  $\Gamma \in W$  e todo  $X \subseteq W$ :

$X \in S^+(\Gamma) \Leftrightarrow$  existe  $Y \in S^+(\Gamma)$  tal que  $Y \subseteq X$ .

O conjunto  $S^+(\Gamma)$  descrito acima pode ser facilmente reescrito da seguinte maneira:

$S^+(\Gamma) = \{X \subseteq W : |\varphi| \subseteq X, \text{ para alguma } \nabla\varphi \in \Gamma\}$ .

Assim, claramente,  $S(\Gamma) \subseteq S^+(\Gamma)$ . Falta-nos mostrar que esta suplementação é um modelo canônico para  $\mathbb{L}(\nabla)$ .

**Proposição 2.17**  $\mathcal{M}]^+$  é um modelo canônico para  $\mathbb{L}(\nabla)$ .

**Demonstração:** Seja  $\mathcal{M}]$  o menor modelo canônico de  $\mathbb{L}(\nabla)$ .

Precisamos verificar a seguinte equivalência: para toda fórmula  $\varphi$  e todo  $\Gamma \in W$ :

$|\varphi| \in S^+(\Gamma) \Leftrightarrow \nabla\varphi \in \Gamma$ .

( $\Rightarrow$ ) Seja  $|\varphi| \in S^+(\Gamma)$ . Então, há  $Y \in S(\Gamma)$  tal que  $Y \subseteq |\varphi|$ . Como  $\mathcal{M}]$  é o menor modelo canônico, então  $Y = |\psi|$ , para alguma fórmula  $\psi$ . Desse modo,  $|\psi| \subseteq |\varphi|$  e  $\nabla\psi \in \Gamma$ . Pelos resultados anteriores, segue que  $\vdash \psi \rightarrow \varphi$  e, por (R $\nabla$ ) temos que,  $\vdash \nabla\psi \rightarrow \nabla\varphi$ . Portanto,  $\nabla\varphi \in \Gamma$ .

( $\Leftarrow$ ) Se  $\nabla\varphi \in \Gamma$ , então  $|\varphi| \in S(\Gamma)$  e como  $S(\Gamma) \subseteq S^+(\Gamma)$ , segue que  $|\varphi| \in S^+(\Gamma)$ . ■

**Proposição 2.18** Se  $\mathcal{M}]$  é o menor modelo canônico para  $\mathbb{L}(\nabla)$  e  $\mathcal{M}]^+$  é sua suplementação, então  $\mathcal{M}]^+ = \langle W, S^+, V \rangle$  é um modelo relacional para  $\mathbb{L}(\nabla)$ .

**Demonstração:** Pelo resultado anterior, segue que  $\mathcal{M}]^+$  é um modelo canônico para  $\mathbb{L}(\nabla)$ . Devemos verificar as condições (i), (ii), (iii) e (iv) da Definição 2.1.

(i) Sejam  $\Gamma \in W$  e  $X, Y \subseteq W$  tais que  $X \in S^+(\Gamma)$  e  $Y \in S^+(\Gamma)$ . Pela definição de  $S^+(\Gamma)$ , existe  $Z \subseteq X$  tal que  $Z \in S(\Gamma)$  e também que existe  $U \subseteq Y$  tal que  $U \in S(\Gamma)$ . Pela condição (i) para  $S$ , segue que  $Z \cap U \in S(\Gamma)$  e, portanto,  $X \cap Y \in S^+(\Gamma)$ .

(ii) Sejam  $\Gamma \in W$  e  $X \subseteq W$  tal que  $X \in S^+(\Gamma)$ . Pela definição de  $S^+(\Gamma)$ , existe  $Z \subseteq X$  tal que  $Z \in S(\Gamma)$ . Agora, para qualquer  $U \subseteq W$ , por (ii) para  $S$ , temos que  $Z \cup U \in S(\Gamma)$  e, assim,  $X \cup U \in S^+(\Gamma)$ .

(iii) Devemos mostrar que  $\emptyset \notin S^+(\Gamma)$ . Como  $\mathcal{M} = \langle W, S, V \rangle$  é um modelo canônico para  $\mathbb{L}(\nabla)$ , então  $\emptyset \notin S(\Gamma)$ . Logo, não existe  $Y \subseteq \emptyset$  tal que  $Y \in S(\Gamma)$  e, então,  $\emptyset \notin S^+(\Gamma)$ .

(iv) Devemos mostrar que  $W \in S^+(\Gamma)$ . Para  $\Gamma \in W$ , o conjunto  $S^+(\Gamma) \neq \emptyset$ . Se  $X \in S^+(\Gamma)$ , então  $X \subseteq W$  e, pela condição (ii),  $W \in S^+(\Gamma)$ . ■

**Definição 2.19** Seja  $\mathcal{M}]$  o menor modelo canônico e  $\mathcal{M}]^+$  sua suplementação. O modelo  $\mathcal{M}]_{\Gamma}^+ = \langle W_{\Gamma}, S_{\Gamma}, V_{\Gamma} \rangle$  é suplementado para  $\Gamma$  quando as seguintes condições são válidas:

-  $W_{\Gamma} = \{\Sigma \in \chi : \Gamma \subseteq \Sigma\}$ ;

- para todo  $\Sigma \in W_{\Gamma}$  e todo  $X \subseteq W_{\Gamma}$ :

$X \in S_{\Gamma} \Leftrightarrow$  para algum  $Y \in S^+(\Sigma)$ ,  $X = Y \cap W_{\Gamma}$ ;

-  $V_{\Gamma}(p_i) = \{\Sigma \in \chi : p_i \in \Sigma \text{ e } \Gamma \subseteq \Sigma\}$ , para toda variável proposicional  $p_i$ .

Intuitivamente, a estrutura acima é obtida excluindo de  $\mathcal{M}]^+$  todos os conjuntos maximalmente consistentes que não contém  $\Gamma$ .

Dessa forma,  $\mathcal{M}]_{\Gamma}^+$  é um modelo canônico e, portanto, as condições (i), (ii), (iii) e (iv) da definição de modelo para  $\mathbb{L}(\nabla)$  continuam válidas para esta estrutura.

Mais ainda, construído dessa forma, para todo  $\Sigma \in W_{\Gamma}$ ,  $\Gamma \subseteq \Sigma$ , segue-se que para  $\mathcal{K} = \mathcal{M}]_{\Gamma}^+$ , temos  $\Sigma \models^{\mathcal{K}} \Gamma$  e, portanto,  $\models^{\mathcal{K}} \Gamma$ .

Finalmente, mostraremos a completude.

**Proposição 2.20 (Completude)** Se  $\Gamma \models \varphi$ , então  $\Gamma \vdash \varphi$ .

**Demonstração:** Pela recíproca da contrária.

Se  $\Gamma \not\vdash \varphi$ , então  $\Gamma \not\vdash \neg\neg\varphi$  e, daí, que  $\Gamma \cup \{\neg\varphi\}$  é consistente.

Pelo Teorema de Lindenbaum, existe um conjunto maximalmente consistente  $\Delta$ , que estende  $\Gamma \cup \{\neg\varphi\}$ , isto é,  $\Gamma \cup \{\neg\varphi\} \subseteq \Delta$ , com  $\neg\varphi \in \Delta$  e  $\varphi \notin \Delta$ .

Como  $\Delta$  é um conjunto maximalmente consistente em  $\mathbb{L}(\nabla)$  e  $\Gamma \subseteq \Delta$ , considerando a estrutura  $\mathcal{M}|_{\Gamma}^+$ , temos que para o modelo  $\mathcal{K} = \mathcal{M}|_{\Gamma}^+$  vale  $\Delta \models^{\mathcal{K}} \Gamma$  e  $\Delta \not\models^{\mathcal{K}} \varphi$ . Deste modo,  $\Gamma \not\models \varphi$ . ■

Com isto, temos uma semântica de vizinhança que é completamente adequada à versão axiomática da lógica proposicional do plausível  $\mathbb{L}(\nabla)$ .

## Considerações finais

O presente trabalho mostra uma semântica relacional para a lógica proposicional do plausível.

Iniciamos procurando alguma semântica de Kripke. No entanto, as semânticas de Kripke usuais são aplicadas em lógicas modais normais. Dado que a lógica proposicional do plausível não é normal, precisamos de um outro para uma semântica de caráter relacional.

Chegamos nas semânticas relacionais ou de vizinhança. Elas tratam dos mundos possíveis e, no nosso caso, de algumas relações especiais impostas nas definições dos modelos que investigamos.

Para as relações impostas, olhamos para os axiomas e obtivemos algumas relações que de algum modo os espelham. Ao apresentarmos tais relações, mostramos a correção do sistema lógico com esta nova semântica.

Para a completude tivemos que seguir um caminho encontrado na literatura, com os conjuntos maximalmente consistentes, apoiados no Teorema de Lindenbaum.

A nova semântica para a lógica proposicional do plausível foi alcançada.

## Agradecimentos

Agradecemos apoio da FAPESP, do CA da UFSC - Câmpus Trindade e da UNESP, Bauru.

## Referências

- [1] CARNIELLI, W.; GRÁCIO M. C. C. Modulated logics and flexible reasoning. *Logic and Logical Philosophy*, v. 17, n. 3, p. 211-249, 2008.
- [2] CHELLAS, B. F. *Modal Logic: an introduction*. New York: University Press, Cambridge, 1999.
- [3] EBBINGHAUS, H. D.; FLUM, J.; THOMAS, W. *Mathematical logic*. New York: Springer-Verlag, 1984.
- [4] FEITOSA, H. A., NASCIMENTO, M. C., GRÁCIO, M. C. C. A propositional version of the logic of the plausible. In: Cezar Mortari e Luiz Henrique Dutra. (Org.) *Rumos da Epistemologia*. Florianópolis: NEL/UFSC, v. 9, p. 185-196, 2009.
- [5] FEITOSA, H. A.; PAULOVICH, L. *Um prelúdio à lógica*. São Paulo: Editora, UNESP, 2005.
- [6] GRÁCIO, M. C. C. *Lógicas moduladas e raciocínio sob incerteza*. Tese de Doutorado em Filosofia. Campinas: UNICAMP/IFCH, 1999.



## Estatística não-paramétrica aplicada na análise dos Fitoterápicos Meratrim e Morosil

Solange Regina dos Santos, Felipe de Oliveira Teixeira.

Colegiado de Matemática, UNESPAR

Campo Mourão, PR

E-mail: solaregina@gmail.com, feholi99@gmail.com.

**Resumo:** *A obesidade no Brasil se mostra cada vez mais presente entre os seres humanos e as suas principais causas são fatores genéticos e ambientais. Hábitos como refeições desorganizadas, às pressas ou ricas em calorias podem conduzir ao aumento de peso quando são repetidas durante muito tempo. Nas mulheres, a maior tendência ao sobrepeso e obesidade ocorre durante a menopausa, pois é uma fase em que ocorrem muitas alterações fisiológicas em todo o organismo, inclusive no metabolismo. Em situações de deficiência estrogênica, como ooforectomia bilateral ou falência ovariana, diversos estudos com modelos animais e em humanos demonstram que há aumento da massa total de gordura corporal. Com o aumento da obesidade, muitos medicamentos e tratamentos são produzidos com o objetivo de reduzir o peso. Entre esses medicamentos, os fitoterápicos vêm sendo utilizados com maior frequência devido aos seus benefícios. Assim, o objetivo desse trabalho é analisar a eficácia desses medicamentos utilizando métodos estatísticos não-paramétricos com base em um banco de dados fornecido por pesquisadores do Centro Universitário Integrado de Campo Mourão, onde ratas wistar foram submetidas à cirurgia de ovariectomia bilateral. Utilizamos o teste de Wilcoxon para verificar se a média do peso das populações que foram tratadas com os fitoterápicos é significativamente menor do que as que não foram submetidas à quaisquer tratamentos.*

**Palavras-chave:** *fitoterápicos, teste de hipóteses, métodos não-paramétricos*

### 1 Introdução

A obesidade é uma mazela que afeta milhões de pessoas ao redor do mundo e tornou-se, nos últimos anos, um grave problema de saúde pública. De acordo com Domenico *et al.* (2012), estudos com modelos animais e em humanos demonstram que em situações de deficiência estrogênica há alterações no tecido adiposo e aumento da massa total de gordura corporal.

Na busca de novas perspectivas para o tratamento da obesidade, a fitoterapia apresenta-se como uma ótima alternativa devido ao seu baixo custo e aos poucos efeitos colaterais. Entre os fitoterápicos, destacam-se o Meratrim, uma mistura de dois extratos vegetais obtidos a partir de cabeças de flores de *Sphaeranthus indicus* e cascas de frutas *Garcinia mangostana* e, o Morosil, um extrato da laranja vermelha. Nesse sentido, com o principal objetivo de verificar se esses medicamentos são eficazes na perda de peso, utilizamos um banco de dados fornecido pelos alunos da disciplina de Nutrição Experimental do Centro Universitário Integrado de Campo Mourão, em que um estudo de caráter experimental foi realizado durante 16 semanas com 22 ratas Wistar pesando entre 150g e 160g. Das 22 ratas, 17 foram submetidas a cirurgia de ovariectomia bilateral - OVX no intuito de induzir a obesidade por deficiência estrogênica. Desse modo, as ratas foram distribuídas em quatro grupos distintos:

Grupo OVX ( $n = 5$ ): ratas submetidas à cirurgia de ovariectomia bilateral;

Grupo OVX+Meratrim ( $n = 6$ ): ratas submetidas à cirurgia e induzidas ao tratamento com Meratrim;

Grupo OVX+Morosil ( $n = 6$ ): ratas submetidas à cirurgia e induzidas ao tratamento com Morosil;

Grupo controle ( $n = 5$ ): ratas que não foram submetidas a tratamento ou cirurgia.

Desse modo, o objetivo dessa pesquisa é verificar se esses medicamentos são eficazes no que diz respeito à perda de peso, ou seja, se os pesos das ratas das populações que foram tratadas com esses fitoterápicos é significativamente menor do que as que não foram submetidas à nenhum tipo de tratamento. Para tanto, métodos estatísticos, descritos a seguir, foram empregados na análise.

## 2 Testes estatísticos

Os métodos estatísticos auxiliam na tomada de decisão, por meio de embasamento científico, permitindo assim, que seja encontrada a solução mais eficaz para o problema. Decisões tomadas por meio da utilização de métodos estatísticos são denominadas de suposições fundamentadas, já no caso da não utilização de embasamento científico, representam meras suposições e, por essa razão, podem se revelar não confiáveis (Mann, 2006).

A estatística é uma ciência que utiliza a análise dos dados para testar hipóteses, verificar evidências, a existência de associações entre grupos e, até mesmo, a veracidade de fenômenos de interesse, sendo possível dividi-la em duas grandes áreas, a estatística descritiva e a inferencial. Enquanto a primeira está ligada à coleta, organização e descrição dos dados, a estatística inferencial é responsável pela redução, modelagem dos dados, a partir do que, finalmente, faz-se a análise e tomada de decisão para a população da qual os dados foram obtidos. Um aspecto importante dessa área da estatística é a possibilidade de fazer previsões, a partir das quais se podem tomar decisões.

A inferência estatística divide-se em duas grandes vertentes, a estatística paramétrica e a não-paramétrica. No primeiro caso, os parâmetros mais comuns são a média e a variância. De acordo com Morettin e Bussab (2017), com base nesses parâmetros podemos descrever características de uma população qualquer e testar hipóteses, de modo que, podemos saber se os resultados experimentais, provenientes de uma amostra contrariam ou não uma afirmação sobre um determinado parâmetro. No segundo caso, estatística não-paramétrica, não há nenhum tipo de suposição sobre quaisquer parâmetros de uma determinada população.

Em alguns casos, os testes paramétricos apresentam suposições muito rigorosas e sua aplicação torna-se praticamente inviável. Nesse contexto, os métodos não-paramétricos surgem como uma alternativa, uma vez que, eles são mais flexíveis em relação aos paramétricos e sua aplicabilidade é, em geral, mais fácil e a perda de eficácia é mínima. Nessa pesquisa utilizamos o *Teste de Wilcoxon* com o intuito de analisar os efeitos dos medicamentos para a redução de peso das ratas que induzidas à obesidade por deficiência estrogênica.

O teste de Wilcoxon é empregado com a finalidade de verificar se existe diferença significativa entre duas populações, ou seja, identificar se uma das populações tende a apresentar média e/ou mediana maiores do que a outra. Supondo que ambas as populações sejam independentes, a base desse teste são os postos dos valores obtidos na combinação de duas amostras. Esse processo é feito colocando-se os valores observados em ordem crescente, independentemente de qual população tais valores são provenientes. A estatística a ser utilizada é a soma dos postos associados aos valores amostrais de uma população  $P_1$ . Caso essa soma seja grande, temos um indício de que os valores dessa população são maiores do que de  $P_2$ , e, sendo assim, rejeitamos a hipótese nula de que ambas as populações têm mesma média e/ou mediana.

Seja  $P_1$  a população das ratas que não foram submetidas a nenhum tipo de tratamento (apenas a cirurgia de ovariectomia),  $P_2$  a população das ratas que foram submetidas ao tratamento com o fitoterápico Meratrim e  $P_3$  a população das ratas que foram submetidas ao tratamento com Morosil. Realizamos dois testes: o primeiro com o intuito de verificar se há diferença significativa entre a média das populações  $P_1$  e  $P_2$  e, o segundo, entre a média das populações  $P_1$  e  $P_3$ .

Dessa forma, para o primeiro caso, as hipóteses a serem testadas são:

$H_0$ : não existe diferença significativa entre os pesos médios das ratas das populações  $P_1$  e  $P_2$ ;

$H_1$ : existe diferença significativa entre os pesos médios das ratas das populações  $P_1$  e  $P_2$ .

A tabela 1 resume os dados fornecidos pelos Pesquisadores do Centro Universitário Integrado de Campo Mourão.

Semana	OVX ( $P_1$ )	OVX+Meratrim ( $P_2$ )	OVX+Morosil ( $P_3$ )
1	149,2	149,5	141,2
2	178,4	166	162,7
3	203,8	190,5	198,7
4	231,8	219,7	231,8
5	249,8	236	248,2
6	259,4	251,3	258,3
7	271,4	257,7	271,3
8	277,2	260,7	270
9	292,2	281,7	295,5
10	298,2	293,5	306,3
11	305,8	304,5	331
12	307	304,5	322,5
13	316,2	303,7	318,3
14	316,2	311,5	333,2
15	321,6	306,7	316,2
16	337	310,8	325,8

Tabela 1: Média dos pesos das populações.

No primeiro teste, colocamos em ordem crescente os valores observados de  $P_1$  e  $P_2$  de modo que as observações ficaram dispostas da seguinte maneira

149,2 149,5 166,0 178,4 190,5 203,8 219,7 231,8  
 236,0 249,8 251,3 257,7 259,4 260,7 271,4 277,2  
 281,7 292,2 293,5 298,2 303,7 304,5 304,5 305,8  
 306,7 307,0 310,8 311,5 316,2 316,2 321,6 337,0

Em virtude da ocorrência de empates nos valores observados, empregamos o teste de Wilcoxon para o caso em que as observações não são todas distintas.

Seja  $d_i$  a quantidade de observações empatadas e  $N$  a soma do tamanho das amostras  $n$  e  $m$ , ou seja,  $N = m + n$ . Como os testes foram feitos em 16 semanas e ocorreram 4 observações empatadas (duas na frequência 304,5 e duas na 316,2), temos que  $d_i = 4$  e  $N = 16 + 16 = 32$ . Note que a proporção  $d_i/N = 4/32 = 1/8$  não é próxima de 1. Para esse caso, Bussab e Morettin (2017) sugerem a seguinte aproximação normal

$$Z = \frac{W_S - E(W_S)}{\sqrt{Var(W_S)}} \tag{1}$$

em que a estatística  $W_S$  é calculada da seguinte forma

$$W_S = S_1 + S_2 + \dots + S_{15} + S_{16} = 2 + 3 + \dots + 27 + 28 = 245. \tag{2}$$

Conforme sugerido por Bussab e Morettin (2017), a esperança e variância da estatística  $W_S$  para o caso em que as observações não são todas distintas são dadas por

$$E(W_S) = \frac{m(N+1)}{2}, \quad (3)$$

$$Var(W_S) = \frac{nm(N+1)}{12} - \frac{mn}{12N(N-1)} \sum_{i=1}^e (d_i^3 - d_i), \quad (4)$$

onde  $e$  é a quantidade de observações empatadas.

Assim, para as amostras consideradas, obtemos

$$E(W_S) = \frac{16 \times (32+1)}{2} = 264, \quad (5)$$

$$Var(W_S) = \frac{16 \times 16 \times (32+1)}{12} - \frac{16 \times 16}{12 \times 32 \times (32-1)} \times [(2^3 - 2) + (2^3 - 2)] = 703,742. \quad (6)$$

Dessa maneira, substituindo (2), (5) e (6) em (1), temos

$$Z = \frac{W_S - E(W_S)}{\sqrt{Var(W_S)}} = Z = \frac{245 - 264}{\sqrt{703,742}} \cong -0,71. \quad (7)$$

Dessa maneira, denotamos  $Z_{obs} = -0,71$ . Utilizando o nível de significância  $\alpha$ , fixado em 5%, obtemos  $Z_{crit} = 1,64$  e a seguinte região crítica:  $RC = (-\infty; -1,64] \cup [1,64; \infty)$ . Como  $Z_{obs} \notin RC$ , não rejeitamos  $H_0$ , isto é, com  $\alpha = 5\%$  podemos concluir que não existe diferença significativa entre os pesos médios das ratas dos grupos amostrais OVX e OVX+Meratrim.

De modo análogo, procedemos o teste de hipóteses para as populações  $P_1$  e  $P_3$ . Assim, temos

$$W_S = 270, E(W_S) = 264, Var(W_S) = 703,484. \quad (8)$$

Dessa maneira, para a estatística  $Z$ , obtemos

$$Z = \frac{W_S - E(W_S)}{\sqrt{Var(W_S)}} = \frac{270 - 264}{\sqrt{703,484}} \cong 0,22. \quad (9)$$

Fixando novamente o nível de significância em 5%, determinamos a região crítica  $RC = (-\infty; -1,64] \cup [1,64; \infty)$ . Como novamente  $Z_{obs} \notin RC$ , aceitamos a hipótese nula, isto é, não existe diferença significativa entre a média e/ou mediana de ambas as populações. Logo, o teste indica um resultado equivalente para o caso anterior, isto é, o fitoterápico Morosil também não apresentou resultados otimistas quanto à perda de peso para a amostra considerada.

### 3 Considerações Finais

Destacamos a relevância da abordagem de testes não-paramétricos no contexto apresentado, justamente por serem mais simples de utilizar e de compreender, quando comparados aos testes paramétricos. Outra vantagem desses métodos é que eles não exigem que a população a partir da qual estejam sendo extraídas as amostras sejam normalmente distribuídas. Diante disso, concluímos que os métodos não-paramétricos demonstram ser ferramentas promissoras e com vasto campo de aplicação como, por exemplo, na investigação da eficácia dos fitoterápicos na redução de peso.

Apesar dos resultados dos testes indicarem não haver diferença significativa entre os pesos médios e/ou medianos dos grupos considerados, não podemos classificar os fitoterápicos Meratrim e Morosil como ineficientes na redução de peso, o que deve ficar claro é que para a amostra considerada os resultados não foram significativos. Para conclusão definitiva sobre a performance dos medicamentos é preciso que amostras maiores sejam consideradas. Todavia, não devemos descartar o índice de confiabilidade dos testes realizados.

Nesse sentido, destacamos aqui que o uso de métodos não-paramétricos são alternativas eficientes no estudo de novos tratamentos e medicamentos e, sua aplicabilidade tem se tornado abrangente.

## Referências

- [1] W. Bussab, and P. Morettin. *Estatística Básica, 9a. edição*. Saraiva, São Paulo, 2017.
- [2] P. S. Mann. *Introdução à estatística*. LTC, Rio de Janeiro, 2006.
- [3] J. M. Marques, and M. A. M. Marques. *Estatística básica para cursos de engenharia*. Domínio do saber, Curitiba, 2005.



## Aplicação do Delineamento Experimental DSD (*Definitive Screening Design*) em Otimização de Produto Industrial e Propostas para a Seleção dos Termos do Modelo

**Henrique Favre**      **Dr. Anselmo Chaves Neto**  
Programa de Pós-Graduação em Métodos Numéricos Aplicados à  
Engenharia – Programação Matemática – UFPR, Curitiba – PR  
E-mails: [henriquefavre@gmail.com](mailto:henriquefavre@gmail.com); [anselmo@ufpr.br](mailto:anselmo@ufpr.br)

**Resumo:** Este trabalho descreve a aplicação do delineamento experimental *Definitive Screening Design* (DSD) para a compreensão dos efeitos de seis (6) componentes na satisfação geral de um determinado produto. Por tratar-se de delineamento supersaturado, um método existente e específico para a seleção dos termos componentes do modelo foi utilizado, porém apresentando melhoria. Esse artigo propõe uma modificação no método e avalia seu resultado através de trabalhos de simulação.

### EXPOSIÇÃO DO PROBLEMA

Em 2011, JONES & NACHTSHEIM [8] introduziram uma nova categoria de experimento de triagem, denominada *Definitive Screening Designs* (DSD). Este trabalho trata da aplicação de um DSD em um caso de otimização dos níveis de seis (6) fatores (componentes) de um determinado produto perante avaliação da satisfação por consumidores. Entretanto, devido ao insucesso ao aplicar-se o método padrão criado pelos autores citados [9], na seleção dos termos do modelo, propõe-se aqui duas adaptações ao método. Estas modificações introduzem a intervenção do pesquisador na seleção de termos.

Um trabalho de simulação computacional foi realizado em R para avaliar o impacto das adaptações propostas no poder do método em detectar os efeitos ativos.

### METODOLOGIA

As técnicas de Delineamento de Experimentos permitem o estabelecimento da relação de causa-e-efeito de variáveis independentes, denominadas variáveis de entrada, em variáveis dependentes, denominadas variáveis-resposta. Isto se dá pelo uso do resultado de experimentos na definição de modelos do tipo  $Y = X\beta + \varepsilon$ . Os experimentos podem ser categorizados como de triagem (*screening*) e de otimização [12]. E, em experimentos de triagem busca-se, dentre diversas potenciais variáveis de entrada, a triagem daquelas com os maiores efeitos sobre a variável-resposta de interesse, sem necessariamente o estabelecimento de um modelo matemático preditivo. Já em experimentos de Otimização [2] buscam-se a obtenção de modelos preditivos de superfície de resposta (RSM).

Para dois fatores, os modelos de superfície de triagem e de superfície de resposta apresentam, respectivamente, as equações:

$$\text{Triagem: } f(x, \beta) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

$$\text{Superfície de Resposta: } f(X, \beta) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon$$

Sendo:  $\beta_0$  o intercepto,  $\beta_i$  o efeito do fator  $X_i$ ,  $\beta_{ij}$  o efeito da interação entre os fatores  $i$  e  $j$ ,  $\beta_{ii}$  o efeito quadrático do fator  $X_i$ ,  $\varepsilon$  o erro com distribuição  $N(0, \sigma^2)$ . Graficamente, os modelos tomam as formas da Figura 1, assumindo todos os efeitos ativos, positivos e de mesma magnitude:

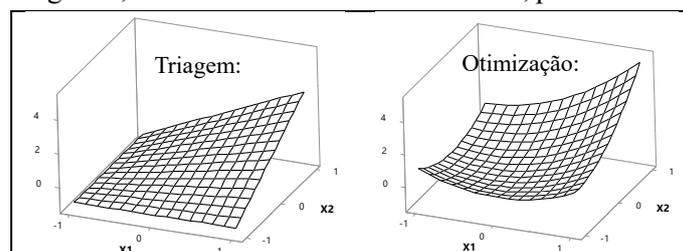


Figura 1: Gráficos de superfície de delineamento de triagem e de otimização com dois fatores

O delineamento experimental DSD foi criado em 2011 [8] por JONES & NACHTSHEIM como alternativa a delineamentos de triagem tais como os delineamentos fatoriais, permitindo a investigação do efeito de quatro (4) ou mais fatores.

O DSD é classificado como um delineamento de triagem que, contudo, possui a propriedade de estimar termos quadráticos.

A Tabela 1 mostra um esquema genérico das  $n$  corridas de um DSD. Além de um ponto central, cada fator  $X_i$  assume o valor médio zero (0) em duas ocasiões, com os demais fatores assumindo valores máximos ou mínimos.

Corridas	Fatores					
	Corrida (i)	$X_{i,1}$	$X_{i,2}$	$X_{i,3}$	...	$X_{i,m}$
1	1	0	$\pm 1$	$\pm 1$	...	$\pm 1$
	2	0	$\mp 1$	$\mp 1$	...	$\mp 1$
2	3	$\pm 1$	0	$\pm 1$	...	$\pm 1$
	4	$\mp 1$	0	$\mp 1$	...	$\mp 1$
3	5	$\pm 1$	$\pm 1$	0	...	$\pm 1$
	6	$\mp 1$	$\mp 1$	0	...	$\mp 1$
...	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
m	2m-1	$\pm 1$	$\pm 1$	$\pm 1$	...	0
	2m	$\mp 1$	$\mp 1$	$\mp 1$	...	0
Pt. Central	2m+1	0	0	0	...	0

Tabela 1: Estrutura do DSD para m fatores (JONES & NACHTSHEIM, 2011)

JONES & NACHTSHEIM [8] citam as vantagens do DSD sobre delineamentos fatoriais utilizados tradicionalmente e de tamanhos similares de corridas:

- Todos os fatores são investigados nos níveis mínimo, máximo e central, o que gera menos receio aos investigadores do que a condução de estudos somente em pontos mínimos e máximos, e permitindo a detecção da presença de efeitos quadráticos.
- Efeitos principais são ortogonais entre si e aos efeitos de segunda ordem (efeitos quadráticos e interações duplas).
- Efeitos de segunda ordem não são completamente ortogonais entre si, possuindo um nível de correlação umas com as outras, porém nunca em total confundimento.
- O número de corridas mínimo é de  $2m+1$  quando o número de fatores  $m$  é par
- Para estudos de 6 a 12 fatores, pode-se beneficiar da esparsidade (parcimônia), ou seja, a ideia de armazenar e operar somente os elementos de uma matriz que não sejam nulos. Outro benefício é a projetividade de efeitos, que é uma realidade em experimentos práticos [1][11]. O DSD projeta-se em um modelo RSM para quaisquer grupos de 2 ou 3 fatores quando os demais se mostram inativos.

Como limitação, destaca-se o fato dos DSDs serem supersaturados, ou seja, podem estimar mais termos que o permitido pelo número de corridas, sendo possível que estimem  $2m + \binom{m}{2}$  efeitos. Com  $m=6$ , torna-se possível estimar os efeitos de 27 fatores, o que dificulta a definição, durante a etapa de análise dos dados, de quais termos fazem parte do modelo.

Originalmente, JONES & NACHTSHEIM [8] sugerem o uso do método de seleção *Forward Stepwise* com uso do critério AICc (Critério de Informação de Akaike corrigido), porém estudos de simulação conduzidos por ERRETE ET AL. [4] demonstraram que os DSDs perdem o poder de detecção de efeitos de segunda ordem mesmo quando poucos destes estão ativos. JONES & NACHTSHEIM [9] sugerem então padronizar o uso de 4 corridas adicionais a cada experimento, com a adição, durante o delineamento, de 2 fatores *dummy*, também denominados *fakes*, ou seja, fatores artificialmente introduzidos e que são ignorados durante a etapa de análise, ao custo de quatro corridas adicionais.

No mesmo artigo [9], os autores propõem ainda um novo método para a seleção de termos de um modelo, implementado no software JMP 13 sob a função *Fit Definitive Screening Design* (aqui denominado "*Fit DSD*"). O método *Fit DSD* aproveita a estrutura ortogonal e espelhada

(*foldover*) do DSD particionando a resposta em dois espaços ortogonais entre si, dada a independência dos efeitos principais e dos efeitos de segunda ordem. Os passos do algoritmo são:

1 - Fatores *dummy* são utilizados para a estimativa do RMSE (Raiz do erro médio quadrático dos resíduos), visto que tais fatores não se relacionam com a variação de fatores reais do estudo;

2 – A variável resposta Y é dividida em duas,  $Y_{ME}$ : valores preditos após regressão de Y somente com um modelo de fatores principais e *dummies*, sem intercepto; e  $Y_{2nd}$ : resíduos da regressão anterior, ou seja,  $Y_{2nd} = Y - Y_{ME}$ ;

3 – Realiza-se o teste estatístico “t” de todos os efeitos principais em modelo ajustado, mantendo aqueles com significância a 5% e reutilizando os demais no refinamento da estimativa do RMSE;

4 – Ajusta-se todos os possíveis modelos de segunda ordem, desde modelos com um (1) termo de segunda ordem até  $c/2$  termos, onde c é a soma do número de fatores com o número de fatores dummy. O modelo de melhor ajuste é extraído. Esse trabalho cobre casos que respeitam o princípio da hereditariedade (*heredity*) descrito em [1], em que efeitos de ordem superior somente são considerados como potencialmente pertencentes ao modelo caso os efeitos principais que os compõem sejam ativos;

5 – Combinam-se os termos dos passos 2 e 3 em uma modelo final.

JONES & NACHTSHEIM [9] adaptam ainda o uso de tal algoritmo para diversos casos, por exemplo sem hereditariedade e sem a adição de fatores do tipo dummy. Nesse artigo, a aplicação é realizada em um DSD com  $m=6$  fatores, adição de 2 fatores dummy e um ponto central sem réplicas, em um total de 17 corridas.

## **RESULTADOS**

Uma aplicação do delineamento DSD e do método Fit DSD foi realizada em problema de otimização em um produto industrial, tendo o método Fit DSD apresentado ineficácia ao se ajustar modelo com sobreajuste (*overfit*), que HAWKINS [7] define quando um modelo é mais complexo que outro modelo cujo ajuste aos dados é igualmente bom sob o ponto de vista de prover uma explicação do fenômeno sob estudo, ou seja, quando o modelo explica o sinal apresentado pelos fatores, porém também parte do ruído.

O projeto em questão, denominado EGH102, busca a melhoria de Qualidade em uma das características de um produto industrial, incluindo otimização dos níveis de alguns dos componentes do produto, de modo a enriquecer a experiência sensorial dos consumidores. As etapas para definição do problema, condução e análise do experimento seguem as preconizadas por MONTGOMERY [12], sendo obtidos os resultados da Tabela 2:

<b>Código</b>	<b>X1</b>	<b>X2</b>	<b>X3</b>	<b>X4</b>	<b>X5</b>	<b>X6</b>	<b>F1</b>	<b>F2</b>	<b>Y1</b>
831	0	+1	+1	+1	+1	+1	+1	+1	6.360
724	0	-1	-1	-1	-1	-1	-1	-1	4.979
915	+1	0	+1	+1	-1	+1	-1	-1	5.395
948	-1	0	-1	-1	+1	-1	+1	+1	6.244
682	+1	-1	0	+1	+1	-1	+1	-1	4.163
185	-1	+1	0	-1	-1	+1	-1	+1	6.257
706	+1	-1	-1	0	+1	+1	-1	+1	3.799
250	-1	+1	+1	0	-1	-1	+1	-1	6.422
326	+1	+1	-1	-1	0	+1	+1	-1	6.122
806	-1	-1	+1	+1	0	-1	-1	+1	4.817
647	+1	-1	+1	-1	-1	0	+1	+1	3.972
514	-1	+1	-1	+1	+1	0	-1	-1	6.271
546	+1	+1	-1	+1	-1	-1	0	+1	6.138

Código	X1	X2	X3	X4	X5	X6	F1	F2	Y1
379	-1	-1	+1	-1	+1	+1	0	-1	4.238
151	+1	+1	+1	-1	+1	-1	-1	0	6.118
269	-1	-1	-1	+1	-1	+1	+1	0	4.415
498	0	0	0	0	0	0	0	0	5.998

Amostra de Validação (Padrão)									
Código	X1	X2	X3	X4	X5	X6	F1	F2	Y1
582	-0.303	1	-1	-0.333	0	-1	N/A	N/A	6.642

Tabela 2: Resultados do Experimento EGH102

As variáveis de X1 a X6 (renomeadas e recodificadas por razões de confidencialidade) representam os níveis dos componentes, F1 e F2 representam os fatores *dummy*, utilizados para estimativa do RMSE e Y1 representa a satisfação geral média do consumidor (medidas conforme DUTCOSKY [3] e na escala introduzida por JONES [10]). Em adição às amostras do delineamento o produto padrão, foi introduzido junto às demais, servindo como amostra para validação do modelo.

A utilização do método Fit DSD resulta no modelo ajustado da Figura 2, com todos os fatores principais, três interações de segunda ordem (X1\*X4, X2\*X5 e X4\*X6) e um efeito quadrático (X5\*X5).

Coefficientes:	Estimativa	Erro Padrão	Estatística t	P(> t )
(Intercepto)	6.01187	0.02606	230.666	4.48e-13 ***
X1	-0.21124	0.01022	-20.674	8.33e-07 ***
X2	0.95041	0.01022	93.020	1.04e-10 ***
X3	-0.04598	0.01022	-4.501	0.00410 **
X4	-0.02649	0.01022	-2.593	0.04105 *
X5	-0.02751	0.01022	-2.692	0.03594 *
X6	-0.16388	0.01022	-16.039	3.73e-06 ***
I(X5^2)	-0.74979	0.02958	-25.350	2.48e-07 ***
X1:X4	0.33593	0.01809	18.569	1.57e-06 ***
X2:X5	0.17312	0.02262	7.655	0.00026 ***
X4:X6	0.21352	0.01809	11.802	2.24e-05 ***

Residual standard error: 0.03823 on 6 degrees of freedom
Multiple R-squared: 0.9994, Adjusted R-squared: 0.9984
F-statistic: 1031 on 10 and 6 DF, p-value: 6.868e-09

Figura 2. Modelo selecionado para Exp. EGH102 via método Fit DSD

Foram identificados aqui os seguintes sintomas de sobreajuste e de ajuste de termos incorretos:

- O valor predito para Y de 7,97, quando as variáveis de entrada assumem os níveis: X1 = -1; X2 = 1; X3 = -1; X4 = -1; X5 = 0,097; X6 = -1, resultado muito alto em comparação a produtos já otimizados em estudos anteriores e mesmo com a amplitude de resultados do próprio experimento, entre 3,80 e 6,42.
- A seleção de fatores com efeitos normalmente tidos como negligenciáveis, como aqueles dos fatores X3, X4 e X5, respectivamente -0,046, -0,026 e -0,028.
- O acentuado efeito quadrático de X5, não antevisto pelos pesquisadores.
- O valor predito para a amostra 582 (7,35), distante de seu resultado real (6,64).
- Não inclusão no modelo de termos X1\*X2 e X2\*X2, significativos em estudos anteriores.

Tais resultados motivaram a proposta de alternativas à metodologia “Fit DSD” que possibilitem a intervenção pelo pesquisador que possuam conhecimento prévio sobre fatores ativos ou de tamanhos de efeito sabidamente negligenciáveis do ponto de vista prático.

Método “Fit DSD com intervenção interativa”: durante o passo de seleção dos efeitos principais, o pesquisador tem a possibilidade de remover fatores que, mesmo significativos, apresentem efeito prático pequeno e negligenciável. Mantendo-se a hereditariedade, removem-se da etapa seguinte, portanto, todos os efeitos de segunda ordem relacionados aos fatores em questão.

Método “Fit DSD com intervenção não interativa”: determina-se a priori fatores que deverão compor o modelo, independentemente de serem ou não significativos.

Feitas as aplicações das duas propostas, a Tabela 3 resume os modelos gerados, cujos passos detalhados do algoritmo estão descritos em [6]:

	Fit DSD	Fit DSD com Intervenção Interativa	Fit DSD com Intervenção Não-Interativa
Intervenções Realizadas	N/A	Efeitos de X3, X4 e X5 removidos após seleção dos efeitos principais	Efeitos de X1, X2, X2*X2 e X1*X2 forçados no modelo
Modelo Selecionado	$Y \sim X1 + X2 + X3 + X4 + X5 + X6 + X1*X4 + X2*X5 + X4*X6 + X5*X5$	$Y \sim X1 + X2 + X6 + X1*X1 + X2*X2 + X1*X2 + X2*X6$	$Y \sim X1 + X2 + X3 + X4 + X5 + X6 + X1*X1 + X2*X2 + X1*X2 + X3*X4$
R <sup>2</sup> Ajustado	99,8%	99,0%	99,6%
AICc	21,73	1,165	17,540
Imagem da função	[3,17 ; 7,97]	[3,73 ; 6,59]	[3,70 ; 6,89]
Predição da amostra 582 (y = 6.64)	7,350	6,558	6,776
Residual Std. Error	0,03823	0,09502	0,06294

Tabela 3. Comparativo de desempenho dos três métodos, no experimento EGH102

Todos os métodos de seleção forneceram modelos bons do ponto de vista estatístico, notadamente com R<sup>2</sup>-Ajustado acima de 99%. O conhecimento prévio dos pesquisadores de que os termos X1, X2, X2\*X2 e X1\*X2 são ativos foram confirmados, dado que seus efeitos são todos significativos nos modelos gerados por via interativa.

O ajuste Fit DSD foi descartado, principalmente pelo sobreajuste representado pela adição do termo X5\*X5, o que resulta no modelo que prediz a maior diferença em relação ao valor real da amostra de validação.

Dentre os modelos restantes, o escolheu-se utilizar o modelo gerado pelo método “Fit DSD com Intervenção interativa”, a partir do conceito de HAWKINS [7], que preconiza utilizar, dentre todos os modelos que possuam um bom ajuste aos dados, o mais simples dentre todos.

O modelo escolhido, por propriedade do delineamento DSD, é uma projeção em Superfície de Resposta para os fatores X1, X2 e X6. Portanto, os resultados podem ser utilizados para a obtenção de modelo preditivo de otimização mostrado na Equação:

$$\widehat{Y}_1 = 6,01 - 0,211 X_1 + 0,95 X_2 - 0,164 X_6 - 0,284 X_1^2 - 0,534 X_2^2 + 0,097 X_1 X_2 + 0,099 X_2 X_6$$

Trabalhos de simulação, seguindo as mesmas estratégias de simulação e apresentação de resultados realizados por JONES & NACHTSHEIM [9], aqui explicam os impactos das adaptações propostas ao método Fit DSD, tanto no poder de detecção dos efeitos ativos quanto no Erro do Tipo I (inclusão de termos não ativos), em uma gama de cenários. Os seguintes resultados abordam as simulações realizadas para o método Fit DSD com Intervenção Não Interativa (Fit DSD INI). Os resultados completos serão publicados em [5].

Foram variados, em delineamento fatorial completo, o número de fatores (6, 8 e 10), o número de efeitos principais ativos (3, 4, 5 e 6), a razão sinal/ruído (SN\_ratio) (1 ou 3), o número de efeitos de segunda ordem ativos (0, 2 e 3), o número de fatores principais forçados no modelo (0, 2 e 4), e o número de fatores principais forçados de forma incorreta (0 ou 2). Para avaliação do ajuste de efeitos de segunda ordem, foram variados ainda o número de fatores de segunda ordem forçados no modelo (0, 2 e 4, respeitando hereditariedade com termos principais também forçados), e o número de fatores principais forçados de forma incorreta (0 ou 2), sendo outras condições mantidas fixas (m = 8 e razão sinal/ruído =1).

Quanto ao poder de detecção para efeitos principais ativos, a Figura 3 exhibe, para o caso com m = 10 fatores (6 dos quais ativos), razão sinal/ruído=1 e 2 fatores de segunda ordem ativos, o

poder de detecção de fatores principais ativos aumenta de 64% (zero fatores forçados, equivalente ao Fit DSD) para 88,5% caso 4 fatores sejam corretamente forçados no modelo. A Figura 3 mostra ainda que forçar 2 termos de forma incorreta causa, como esperado, perda no referido poder, porém aos mesmos níveis encontrados quando não há termos forçados no modelo.

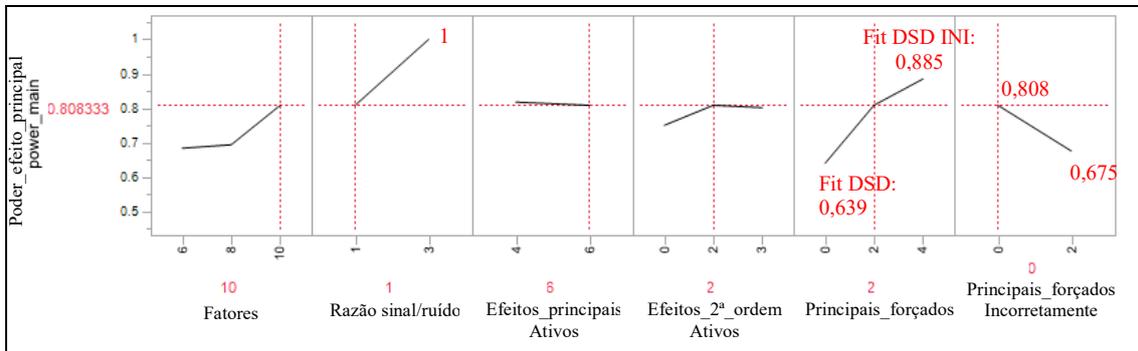


Figura 3. Poder de Detecção de Efeitos Principais para O Método Fit DSD Não Interativo (m=10; SN\_ratio=1 ou 3; mA=6; 2A=2; mF=0; mFE=0)

Quanto ao poder de detecção para efeitos de segunda ordem ativos, forçar efeitos principais de forma correta pode acarretar em aumento no poder de detecção de termos de segunda ordem. Por exemplo, a Figura 4 mostra que forçar todos os quatro (4) efeitos ativos aumenta o poder de detecção de efeitos de segunda ordem de 27% para 47%. Entretanto, quando dois dos quatro fatores principais são forçados de forma incorreta, o poder cai a 20%. Para 5 e 6 efeitos ativos, o efeito de forçar efeitos principais, mesmo que corretamente, não é observado.

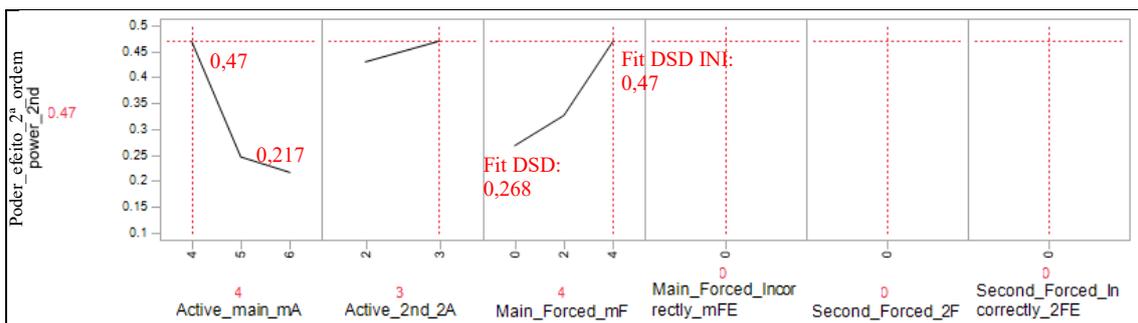


Figura 4. Poder de Detecção de Efeitos de Segunda Ordem para o Método Fit DSD Não Interativo (m=8; SN\_ratio=1; mA=4 ou 6; 2A=3; mF=0; mFE=0; 2F=0; 2FE=0)

Passando à simulação de casos em que há introdução forçada de dois (2) fatores segunda ordem, que dependem, pelo princípio da hereditariedade, de se forçar fatores principais, a Figura 5 mostra o incremento no poder para detecção de fatores de segunda ordem, chegando a 83% no caso com 4 efeitos principais ativos, todos forçados corretamente no modelo.

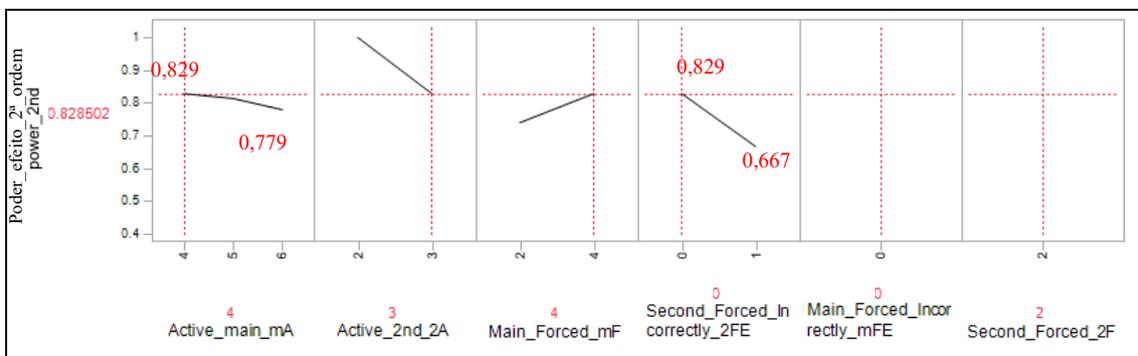


Figura 5. Poder de Detecção de Efeitos de Segunda Ordem para o Método Fit DSD Não Interativo (m=8; SN\_ratio=1; mA=4; 2A=3; mF=4; mFE=2; 2F=0; 2FE=0)

O caso da Figura 5 apresenta configurações que “facilitam” o algoritmo Fit DSD com Intervenção Não Interativa, e por esse motivo adicionalmente foram simulados casos com 6 fatores de segunda ordem ativos, com zero (0) ou dois (2) fatores de segunda ordem forçados no modelo e os demais mantidos fixos: Fatores principais ativos (mA) = 4; Fatores de segunda ordem ativos (2A) = 6; Fatores principais incorretamente forçados no modelo (mFE) = 0

Aqui, nota-se um bom incremento de poder para efeitos de segunda ordem, de 26% para 51%. Quando um dos fatores é incorretamente forçado, o poder cai para 41%. Os resultados são mostrados na Tabela 4.

<b>Fatores Variáveis</b>	<b>Níveis</b>		
Fatores principais forçados no modelo (mF)	0	2	2
Fatores de segunda-ordem forçados no modelo (2F)	0	2	2
Fatores de segunda-ordem incorretamente forçados no modelo (2FE)	0	0	1
<b>Poder de detecção de Fatores de Segunda Ordem</b>	26%	51%	41%

Tabela 4. Resultados de Simulações Adicionais do Método Fit DSD Não Interativo, na Avaliação do Poder para Detecção de Efeitos de Segunda Ordem

## CONCLUSÃO

Os métodos propostos apresentam resultados promissores e são, portanto, recomendados para pesquisadores que desejem utilizar DSDs quando de posse de conhecimento prévio sobre fatores ativos ou tamanhos de efeito sabidamente negligenciáveis do ponto de vista prático. Por tratar-se de tipo de delineamento relativamente novo e em constante evolução, recomenda-se o uso do delineamento DSD por pesquisadores em estudos de triagem que não possuam restrições ao espaço de estudo, pois além de possibilitar o estudo de curvatura em cada fator estudado, além das interações duplas, os benefícios do DSD parecem largamente superar suas limitações.

## **Referências**

- [1] Box, G.; Hunter, S.; Hunter, W. Statistics for Experimenters: design, discovery, and innovation. 2nd Ed. Wiley-Interscience, 2005.
- [2] Box, G.; Wilson, K. On the Experimental Attainment of Optimum Conditions. Journal of the Royal Statistical Society. Series B, 13, 1-45, 1951.
- [3] Dutcosky, S. D. Análise Sensorial de Alimentos, 4. ed – PucPress, Curitiba, 2013
- [4] Errore, A. et Al, “Using Definitive Screening Designs to Identify Active First-and Second-Order”, Journal of Quality Technology, Vol. 49, No.3, 244, 264, Jul 2017.
- [5] Favre, H., Procedimento de Seleção de Termos do Modelo para o Delineamento Definitive Screening Design, de Jones & Nachtsheim, com Intervenção do Experimentador, UFPR, Curitiba, Em processo de publicação.
- [6] Favre, H., Aplicação de métodos de seleção de termos de modelo em delineamento do tipo “Definitive Screening Design”, na avaliação da apreciação por consumidores de produto alimentício, Anais do III Simpósio de Métodos Numéricos em Engenharia, UFPR, Curitiba, 2018.
- [7] Hawkins, D., The Problem of Overfitting. J. Chem. Inf. Comput. Sci. 2004, 44, 1-12.
- [8] Jones, B.; Nachtsheim, C. A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects, Journal of Quality Technology, Vol. 43, No. 1, 2011.
- [9] Jones, B.; Nachtsheim, C., Effective Design-Based Model Selection for Definitive Screening Designs, Technometrics, 59:3, 319-329, 2017.
- [10] Jones, L. V.; Peryam, D. R.; Thurstone, L. L., Development of a Scale For Measuring Soldiers’ Food Preferences, Journal of Food Science, 20: 512–520, 1955.
- [11] Li, X et. Al, Regularities in data from factorial experiments. Complexity 11(5), p 32–45, 2006.
- [12] Montgomery, D., Design and Analysis of Experiments, 7<sup>a</sup> ed., Ed. John Wiley & Sons, Hoboken, NJ, 2008.



## Modelagem do IMCRB utilizando duas metodologias para extração de fatores

**Luana E. Felippi**

Bolsista de Desenvolvimento em Tecnologia e Inovação, CEASB  
85867-900, Foz do Iguaçu, PR  
E-mail: luana\_felippi@hotmail.com

**Sheila R. Oro**

Universidade Tecnológica Federal do Paraná - Departamento de Matemática e Estatística  
85601-970, Campus Francisco Beltrão, PR  
E-mail: sheiloro@utfpr.edu.br

*Resumo: O presente trabalho descreve os resultados obtidos na modelagem do Índice de Monitoramento Conjunto das Respostas dos Blocos da Barragem, utilizando dois métodos de extração de fatores na análise fatorial, a saber: método de componentes principais e máxima verossimilhança. Utilizando um teste de distribuição de probabilidade qui-quadrado, verificou-se que os dados estudados apresentam distribuição normal multivariada, condição essencial para a utilização da máxima verossimilhança. Após realizada a análise fatorial, utilizando as duas formas de extração de fatores, constatou-se que a utilização do método de componentes principais permitiu que as variáveis fossem explicadas de maneira mais satisfatória em comparação com a utilização da máxima verossimilhança. Os fatores obtidos pela análise fatorial apresentaram poucas mudanças em relação a mudanças de variáveis, sendo que 3 fatores não tiveram nenhuma variável alterada. As duas séries para o IMCRB obtidas utilizando os dois métodos de extração de fatores não apresentaram mudanças consideradas substanciais, apenas alguns amortecimentos localizados fazendo com que a série fique distribuída de maneira mais adequada dentro dos limites de confiança estipulados.*

### 1. Introdução

Toda barragem, independente de sua finalidade, exige cuidados especiais quanto à sua segurança estrutural, a fim de evitar acidentes e os eventuais prejuízos decorrentes, sejam eles de caráter estrutural, econômico, ambiental ou social [14].

A avaliação constante de sua segurança estrutural pode colaborar na determinação de anomalias, utilizando para este fim o auxílio da instrumentação de monitoramento e de inspeções visuais. Este conjunto de medidas permite que a equipe de engenheiros responsável pela segurança da barragem possa tomar decisões ou medidas protetivas acerca do problema identificado [16].

A instrumentação é uma importante ferramenta para colaborar no monitoramento das estruturas e fundação de uma barragem de concreto [5]. Os diferentes instrumentos captam as informações do comportamento da barragem [15] e alimentam uma grande base de dados [4], que são analisados graficamente e comparados com os valores estipulados em projeto.

Em diversas pesquisas [3], [6], [10], [12], são utilizadas técnicas estatísticas para analisar os dados de monitoramento de barragens. Por exemplo, em [13] a análise fatorial serviu de base para o desenvolvimento de uma metodologia de geração de um índice que represente o comportamento global de uma barragem, utilizando as séries de dados históricos da instrumentação.

A análise fatorial é uma técnica estatística multivariada que busca descrever muitas variáveis originais (observáveis) em termos de poucas variáveis latentes, denominadas fatores [9]. Por meio da aplicação desta técnica é possível identificar fontes subjacentes de variação comum entre as variáveis observáveis [1].

De acordo com [11], para aplicar a análise fatorial, primeiramente, organiza-se uma matriz de dados, e verifica-se a adequação destes para a aplicação desta técnica (Teste de Bartlett e

Coefficiente KMO). Em seguida, define-se o número de fatores (Critério de Kaiser). O processo de extração dos fatores pode ser realizado por diferentes formas, dentre elas, por componentes principais e por máxima verossimilhança. Por fim, os escores fatoriais são calculados, por meio da combinação linear das variáveis originais.

Com base nessas informações, o objetivo principal deste trabalho é analisar o comportamento global das estruturas do Bloco F19/20, do Trecho F, da Barragem de Itaipu, por meio do Índice de Monitoramento Conjunto das Respostas dos Blocos da Barragem (IMCRB).

Como objetivos específicos pretende-se utilizar a análise fatorial, empregando dois métodos de extração de fatores, para identificar os sensores da instrumentação com maior relevância em relação à variabilidade dos dados; gerar séries para o IMCRB com os fatores obtidos em cada um dos métodos de extração; utilizar o IMCRB para gerar séries que representem o comportamento conjunto da instrumentação do bloco estudado e fazer uso de cartas de controle para verificar a disposição dos dados em relação com os limites estipulados.

Nesta seção, é apresentada uma introdução sobre o assunto. A seção 2 apresenta os materiais e métodos, a qual descreve como foi realizada a pesquisa e quais os dados utilizados. Na sequência, são apresentados os resultados obtidos e, por fim, as considerações finais.

## 2. Materiais e Métodos

Do ponto de vista dos objetivos, este trabalho é caracterizado como descritivo, pois pretende descrever os resultados das análises realizadas na aplicação de métodos estatísticos em dados da instrumentação da Barragem de Itaipu. A pesquisa descritiva, de acordo com [7], "têm como objetivo primordial a descrição das características de determinada população ou fenômeno ou, então, o estabelecimento de relações entre variáveis".

A coleta de dados foi realizada utilizando as medições dos deslocamentos da estrutura e subpressões dos sensores da instrumentação instalada no bloco F19/20, durante o período de janeiro de 1990 a dezembro de 2017. Este bloco foi construído do tipo gravidade aliviada [8] e está localizado na barragem principal de Itaipu, como mostra a Figura 1.



Figura 1: Fotografia aérea da Barragem de Itaipu

Fonte: [8]

O conjunto de dados foi composto pelas médias mensais dos valores das medições manuais de deslocamentos e subpressões medidos por cinco tipos de instrumentos, sendo eles: pêndulos diretos, pêndulos invertidos, extensômetros múltiplos de hastes, piezômetros e bases de alongômetro, totalizando 36 sensores (variáveis observáveis).

A matriz de dados foi submetida a análise fatorial, com fatores extraídos por meio dos métodos de componentes principais e de máxima verossimilhança, com a finalidade de descrever a estrutura de covariância do relacionamento entre os sensores da instrumentação em função de um número reduzido de fatores.

Essencialmente, a diferença entre os dois métodos está no fato de que para o uso da máxima verossimilhança necessita-se a verificação da suposição de normalidade multivariada para as variáveis padronizadas, os fatores e os resíduos [11]. No entanto, os fatores e os resíduos do modelo fatorial são variáveis não observáveis. Desta forma, inicialmente realizou-se um teste de distribuição de probabilidade (Qui-quadrado), para verificar a normalidade multivariada dos dados dos 36 sensores.

Os escores fatoriais foram usados como dados de entrada para a modelagem do Índice de Monitoramento Conjunto das Respostas dos Blocos da Barragem (IMCRB), por meio de uma combinação linear dos fatores, com pesos determinados por seus respectivos autovalores da matriz de correlação dos dados, resultando numa série temporal que representa as respostas da barragem em relação à movimentação estrutural e às subpressões.

A partir da série obtida pelo IMCRB do Bloco F19/20, geraram-se cartas de controle para o índice. Uma carta de controle, oferece visualização das ocorrências de um processo ao longo do tempo, evidenciando, por meio de intervalos de confiança estipulados para a variável, quaisquer anormalidades que possam vir se manifestar durante o período analisado.

A carta de controle colabora na identificação de fatores particulares de variação, detectados por pontos fora dos limites de controle ou também pela presença de um comportamento sistemático não aleatório [2].

Por fim, realizou-se uma comparação dos resultados obtidos utilizando os dois métodos de extração de fatores: componentes principais e máxima verossimilhança.

### 3. Resultados

Inicialmente, aplicou-se um teste de adequabilidade, o qual resultou em um índice KMO = 0,93, superando o mínimo estipulado pela literatura de 0,6, indicando que os dados eram adequados à análise fatorial. Em seguida, usando o Critério de Kaiser, definiu-se a quantidade de 6 fatores, a serem extraídos.

#### 3.1. Método componentes principais para extração de fatores

Os fatores extraídos por componentes principais, permitiram a explicação de cerca de 87% da variabilidade nos dados, sendo esta considerada uma taxa satisfatória.

As comunalidades de cada variável, pelo método de componentes principais, foram superiores a 0,5 para todos os sensores abordados. Considera-se que esse valor representa uma boa proporção da variabilidade explicada pelo modelo fatorial.

Com os escores gerados na análise fatorial, gerou-se uma série para o Índice de Monitoramento Conjunto das Respostas dos Blocos da Barragem IMCRB, a qual apresenta o comportamento conjunto das respostas da instrumentação deste bloco.

A Equação 1 apresenta a combinação linear que gerou a série temporal do IMCRB, tal que  $F_i$  representa o vetor que compõe os escores do fator  $i$ , e os coeficientes são os autovalores determinados pela matriz de correlação dos dados dos sensores da instrumentação do Bloco F19/20.

$$\text{IMCRB} = 17,809.F_1 + 5,082.F_2 + 3,734.F_3 + 2,509.F_4 + 1,967.F_5 + 1,268.F_6 \quad (1)$$

A partir da série obtida foram calculados os limites, superior e inferior, à 95% e 99% de confiança, os quais foram utilizados na carta de controle para o índice, apresentada na Figura 3.

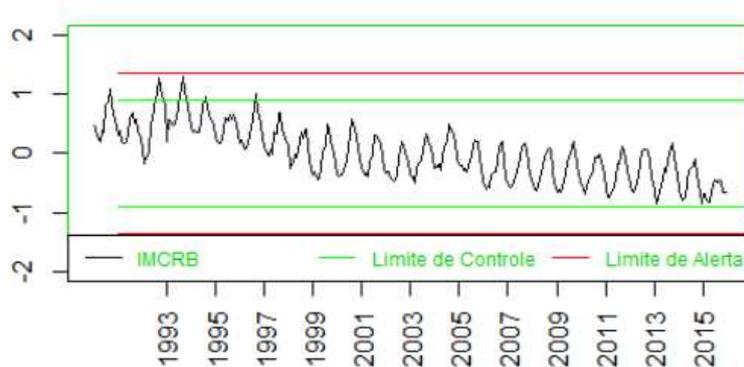


Figura 2: Carta de controle para o IMCRB utilizando componentes principais

A série apresentou sazonalidade, apresentadas pelas extremidades altas e baixas em períodos aproximadamente fixos de tempo. Essa condição é ocasionada devido à movimentação da estrutura e subpressões registradas, em condições da mudança de temperatura durante às estações do ano.

Esta série apresentou tendência decrescente na média. Porém, pode-se notar que nos últimos períodos de estudo a disposição dos dados fica em torno de uma mesma média, o que pode indicar deslocamentos mais estáveis.

### 3.2. Método da Máxima Verossimilhança para extração de fatores

O teste de normalidade dos dados, para a verificação da condição positiva de aplicação do método da máxima verossimilhança, foi realizado com os valores dos escores, gerados pelos 6 fatores na análise fatorial. O resultado do teste sinalizou como positivo para normalidade dos dados, como mostra o gráfico Q-Q Plot na Figura 2.

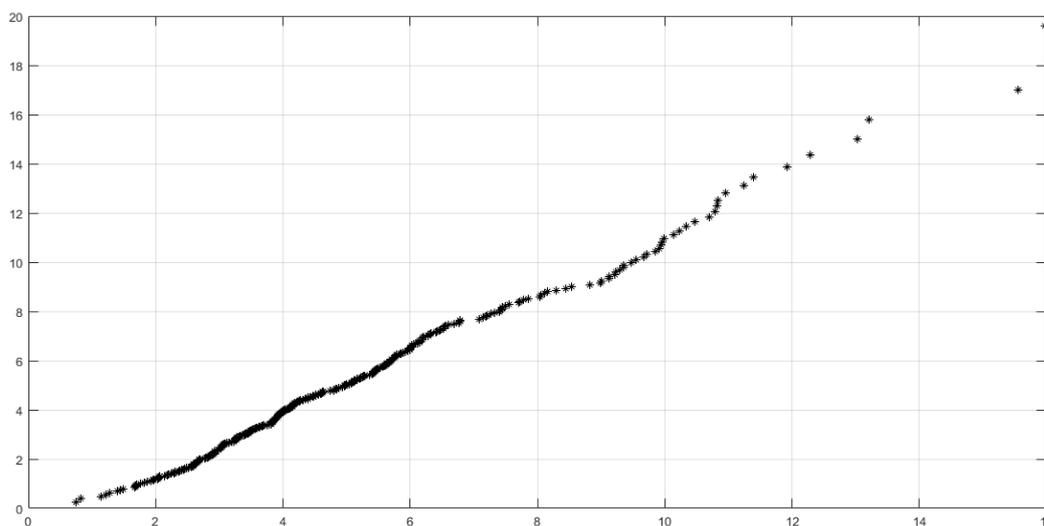


Figura 3: Q-Q Plot da distribuição normal multivariada dos dados

A porcentagem total de explicação dos dados pela análise fatorial utilizando o método da máxima verossimilhança foi de 84,85%. Utilizando o método de Máxima Verossimilhança, todos os sensores apresentaram comunalidades acima de 0,50, exceto dois sensores de piezômetros, os quais apresentaram comunalidades abaixo de 0,20. Desta forma, percebe-se que estes dois sensores não possuem uma boa proporção da variabilidade explicada pelo modelo fatorial.

Utilizando a Equação 1, apresentada anteriormente neste trabalho, gerou-se uma nova série para o IMCRB. Pode-se notar que os autovalores são os mesmos para a geração das duas séries,

logo esta é uma equação genérica, na qual serão alterados apenas os escores dos vetores  $F_i$ , de acordo com o método de extração de fatores utilizado.

A partir da série obtida gerou-se então novos limites de confiança, superior e inferior, à 95% e 99% de confiança, e posteriormente gerou-se uma nova carta de controle para o índice, a qual é apresentada na Figura 4, a seguir.

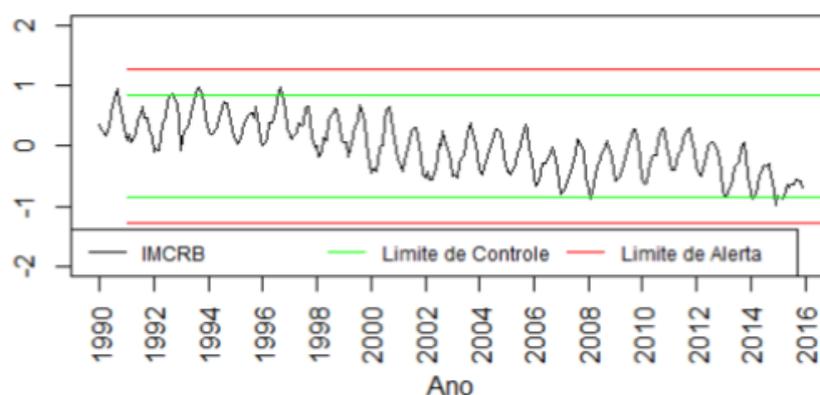


Figura 4: Carta de controle para o IMCRB utilizando máxima verossimilhança

Pode-se notar que o comportamento da série do IMCRB utilizando o método da máxima verossimilhança é semelhante ao comportamento da série anterior, apresentando sazonalidade e tendência decrescente na média.

### 3.3. Comparações entre os dois métodos

Com relação à extração dos fatores, percebeu-se que utilizando componentes principais existe uma pequena vantagem no percentual de explicação (2,53%). Contudo, é importante ressaltar que o método da máxima verossimilhança retorna um resultado com estimativas mais precisas para os escores fatoriais.

Em geral, observou-se que a maioria das comunalidades dos sensores aumentou quando se utilizou a máxima verossimilhança, isto é, esses sensores foram melhor representados pelo modelo fatorial obtido utilizando esse método para a extração de fatores.

A principal mudança observada foi na correlação dos sensores com os fatores (cargas fatoriais). Por componentes principais, o 3º fator teve maior relação com dois sensores de piezômetros e três de base de alongâmetro. Utilizando a máxima verossimilhança, estes sensores foram separados em dois fatores 6º e 4º, respectivamente. O 3º fator, utilizando a máxima verossimilhança, ficou composto então de uma junção dos sensores que antes faziam parte do 4º e 6º fatores pelo método de componentes principais.

A maior alteração nas cargas fatoriais, quando mudou-se o método de extração de fatores, foi observada em dois piezômetros (X1 e X2), que apresentaram redução significativa nas comunalidades com a utilização da máxima verossimilhança. A Tabela 1 apresenta os valores estimados para as comunalidades (Com) e as cargas fatoriais destes dois sensores de piezômetros pelos dois métodos de extração de fatores: componentes principais (CP) e máxima verossimilhança (MV).

Sensores	Com	Método	Cargas fatoriais					
			Fator 1	Fator 2	Fator 3	Fator 4	Fator 5	Fator 6
X1	0,7982	CP	-	-	-	-0,178	-	0,863
X2	0,9375	CP	-	-	-	-0,186	-	0,943
X1	0,2038	MV	-0,197	-0,372	0,120	-	-	-
X2	0,2098	MV	-0,188	-0,396	-	-	-	-

Tabela 1: Alterações nas comunalidades e cargas fatoriais de X1 e X2

Pode-se notar que, pelo método de componentes principais, estes dois sensores de piezômetros possuem uma boa explicação pelo 6º fator, apresentando valores de 0,863 e 0,943, de X1 e X2, respectivamente. Em contrapartida, utilizando o método da máxima verossimilhança as cargas fatoriais destes dois sensores ficaram distribuídas em três fatores: fator 2, fator 3 e fator 4, porém sem valores significativos, com módulo permanecendo abaixo de 0,4.

Em relação às maiores comunalidades geradas, utilizando o método da máxima verossimilhança verificou-se que três sensores apresentaram valores acima de 0,99, sendo estes: um sensor de pêndulo direto (Y1) e dois sensores de extensômetros (Y2 e Y3), os quais apresentaram valores iguais ou menores quando utilizado componentes principais. A Tabela 2 apresenta as comunalidades (Com) e as cargas fatoriais dessas três variáveis que pelos dois métodos de extração de fatores (CP e MV).

Sensores	Com	Método	Cargas Fatoriais					
			Fator 1	Fator 2	Fator 3	Fator 4	Fator 5	Fator 6
Y1	0,9752	CP	-0,612	0,720	0,178	-	0,185	-
Y2	0,9910	CP	-0,563	-	0,199	-0,394	0,445	0,203
Y3	0,9838	CP	0,981	-0,133	-	-	-	-
Y1	0,9935	MV	-0,574	0,770	-	0,166	0,169	0,122
Y2	0,9910	MV	0,964	-0,219	-	-	-	-
Y3	0,9923	MV	0,972	-0,195	-	-	-	-

Tabela 2: Alterações nas comunalidades e cargas fatoriais de Y1, Y2 e Y3.

A possível explicação para o aumento das comunalidades destes sensores quando utilizou-se a máxima verossimilhança é a concentração das cargas fatoriais em menor número de fatores.

Em relação às cartas de controle, pode-se observar que o comportamento do índice não sofreu grandes alterações. Nota-se, porém que por volta do ano de 1992 a 1994 houve um amortecimento na amplitude dos dados, quando utilizou-se a máxima verossimilhança, fazendo com que valores que antes ficavam fora dos limites de confiança a 95%, agora fiquem dentro desses limites. Em contrapartida, percebe-se uma pequena mudança aproximadamente no ano de 2014 a 2016, na qual os dados tendem a se aproximar mais do limite inferior de confiança a 95%.

#### 4. Considerações Finais

Este trabalho apresentou a comparação dos resultados obtidos a partir de dois métodos diferentes de extração de fatores, na análise fatorial, e modelagem do IMCRB, referentes às medições dos sensores da instrumentação instalados nas estruturas e fundação do Bloco F19/20, localizado no Trecho F da Barragem de Itaipu.

Constatou-se que utilizando o método de componentes principais foi ligeiramente melhor do que o da máxima verossimilhança para explicar a estrutura de covariância entre os sensores.

A estrutura do modelo fatorial pelos dois métodos apresentou diferenças, com a alternância de algumas variáveis entre os fatores e alterações nas comunalidades de alguns sensores.

O comportamento do índice IMCRB não apresentou grandes mudanças na comparação dos dois métodos de extração de fatores utilizados. Percebeu-se que utilizando o método da máxima verossimilhança, os valores permaneceram em maior quantidade dentro do limite de confiança a 95%, apresentando um aspecto de amortecimento na série em determinados períodos.

Desta forma, por meio deste estudo foi possível perceber que ambos os métodos são adequados para aplicação aos dados de monitoramento de barragens.

#### Referências

[1] CHAVES NETO, A. Análise Multivariada Aplicada à Pesquisa: Notas de aula. Foz do Iguaçu. 2014.

- [2] CHAVES NETO, A.; CHAVES, I. A. Técnicas Estatísticas Aplicadas à Engenharia da Qualidade: Notas de aula. Foz do Iguaçu. 2014.
- [3] CHOUINARD, L., LARIVIÈRE, R., CÔTE, P. e ZHAO, W. Analysis of irreversible displacements in multiple arch concrete dam using principal component analysis. Proceedings in: International Conference on Computing and Decision Making in Civil and Building Engineering, Montreal, 2006.
- [4] COELHO, D. P., PATIAS, J., GARAY, V. R. Sistema de otimização e análise de auscultação da Barragem de Itaipu. Proceedings in: XXX Seminário Nacional de Grandes Barragens. CBDB, Foz do Iguaçu, 2015.
- [5] CRUZ, P. T. 100 barragens brasileiras: casos históricos, materiais de construção, projeto. São Paulo: Oficina dos Textos, 2006.
- [6] DE SORTIS A, A. PAOLIANI, P. Statistical Analysis and Structural Identification in Concrete Dam Monitoring. V. 29, pp. 110-120. Engineering Structures, 2007.
- [7] GIL, A. C. Métodos e técnicas de pesquisa social. 6 ed. São Paulo: Atlas, 2008.
- [8] ITAIPU BINACIONAL. A Maior Geradora de Energia Limpa e Renovável do Planeta. Disponível em: <<https://www.turismoitaipu.com.br/pt/dados-tecnicos>>. Acesso em: 11 Abr. 2019.
- [9] LATTIN, J.; CARROLL, J. D.; GREEN, P. E. Análise de Dados Multivariados. São Paulo: Cengage Learning, 2011.
- [10] MATA, J. TAVARES, A. SÁ DA COSTA, J. (2013) - "Time–frequency analysis for concrete dam safety control: Correlation between the daily variation of structural response and air temperature", 7-f (658 - 665). Engineering Structures. V. 48.
- [11] MINGOTI, Sueli Aparecida. Análise de Dados Através de Métodos de Estatística Multivariada: Uma abordagem Aplicada. Belo Horizonte: Editora UFMG, 2005.
- [12] NEDUSHAN, B. A. (2002) – “Multivariate Statistical Analysis of Monitoring Data for Concrete Dams” (Doctoral dissertation, McGill University), Montreal.
- [13] ORO, S. R. Índice de Monitoramento do Comportamento Estrutural dos Blocos de Concreto de Barragens – Uma Abordagem Multivariada / Sheila Regina Oro. – Curitiba, 2016. 139 f.
- [14] SILVA, E. T. G. Barragens hidrelétricas e desastres: uma avaliação de metodologias de gestão de risco ambiental. Dissertação (Mestrado em Planejamento e Gestão Ambiental). Universidade Católica de Brasília, Brasília, 2012.
- [15] SILVEIRA, J. F. A. Instrumentação e comportamento de fundações de barragens de concreto. São Paulo: Oficina de Textos, 2003.
- [16] ZUCULIN, S. Curso Segurança de Barragens: Módulo II: Inspeção e Auscultação de Barragens. Unidade 2: Instrumentação de Barragens. 2013. 71 p.



## Otimização de parâmetros de processo de extração de antocianinas utilizando metodologia de superfície de resposta

**Cintia K. Imme; Cintia Marangoni; José Alexandre B. Valle; Rita de Cassia S. C. Valle**

Pós graduação em Engenharia Têxtil, Depto. De Engenharia Têxtil, Universidade Federal de Santa Catarina - Campus Blumenau, Blumenau-SC  
e-mail: [rita.valle@ufsc.br](mailto:rita.valle@ufsc.br)

**Luis A. Lourenço**

Pós graduação em Engenharia Química, Dpto. de Engenharia Química, Universidade Regional de Blumenau, Blumenau-SC

**Lucile C. Peruzzo**

Pós graduação em Engenharia Química, Dpto. Engenharia Química e Alimentos, Universidade Federal de Santa Catarina – Campus Trindade, Florianópolis-SC

### Resumo

*Antocianinas (ACYs) são pigmentos naturais derivados de frutas, flores e frutos que está atraindo o interesse e o interesse de indústrias têxteis, para uso como corante natural, e indústrias de alimentos, cosméticos e suplementos, para uso como antioxidante. Apesar das metodologias de extração serem bastante conhecidas, há necessidade de ajustar os parâmetros de forma a otimizar o processo para obter rendimentos e economia de materiais e energia. Métodos estatísticos surgem como uma ferramenta útil, visto que o uso de desenho de experimentos permite obter do sistema em análise toda a informação com um mínimo de ensaios. Neste trabalho é analisada a influência dos fatores de processo em duas etapas: seleção de variáveis por meio de um planejamento fatorial fracionado  $2^{4-1}$ , seguido de um método de superfície de resposta, considerando as variáveis temperatura, tipo de solvente, concentração do solvente e concentração de ácido, para a obtenção de ACYs de resíduo de vinículas. O desenho e a análise dos dados foram feitos no software Statistica. O uso de planejamento fatorial fracionado aliado ao conhecimento prévio do sistema permitiu uma tomada de decisão das variáveis que deveriam continuar a serem ensaiadas para atingir as condições ótimas do processo. O método de superfície de resposta (planejamento composto central) permite tirar conclusões sobre efeitos de interação e a visualização de gráficos com regiões ótimas de trabalho que oportuniza flexibilização e adequação dos processos em escala industrial, além de prever, por meio do modelo estatístico, as consequências de variações sobre as respostas, que nesse caso é rendimento de antocianinas no processo de extração.*

### Introdução

O interesse por produtos que contenham compostos bioativos, como as antocianinas, tem aumentado muito nos últimos anos. As antocianinas são, provavelmente, a classe de pigmentos naturais mais conhecidas, estando presentes na maior parte dos vegetais, conferindo coloração que variam entre o vermelho, violeta e púrpura. No entanto, a utilização industrial de antocianinas apresenta restrições em função de algumas limitações, como a disponibilidade de fontes comerciais destes pigmentos e dificuldades nos processos de obtenção.

A metodologia adotada para a extração destes pigmentos é dependente de alguns fatores, principalmente para qual finalidade será empregado o corante, a natureza dos compostos e da fonte ou matriz utilizada. Na maioria das frutas e demais vegetais, os pigmentos se encontram localizados em células próximas a superfície. Uma vez que o pigmento está dissolvido na seiva das células, a sua extração geralmente envolve o uso de solventes ácidos, os

quais desnaturam a membrana das células do tecido e dissolvem os pigmentos [2]. O caráter polar da molécula de antocianina, devido à presença de grupos substituintes (hidroxilas, carboxilas e metoxilas), permite a sua solubilidade em muitos solventes tais como alcoóis, acetona e água [4, 7].

Embora as metodologias encontradas na literatura sejam bastante consolidadas, as variáveis envolvidas nos processos de extração para obtenção de ACYS estão muito relacionadas com a matriz utilizada. Além disso, fazer um ajuste dos parâmetros, para um caso em particular, nos permite otimizar o processo extrativo, aliando aumento nos rendimentos e economia de insumos e energia. Neste âmbito, os métodos estatísticos surgem como uma ferramenta útil, onde segundo Barros Neto, Scarmino e Bruns [1], a utilização de planejamentos experimentais baseados em fundamentos de estatística, permite ao investigador obter do sistema em análise o máximo de informações úteis, fazendo um número mínimo de ensaios. Quando o objetivo principal é otimizar um sistema, isto é, maximizar uma resposta de interesse, que seja influenciada por diversas variáveis, ou fatores, uma técnica conveniente de análise é a metodologia de superfícies de resposta (MSR).

O software Statistica é um programa de métodos estatísticos que permite realizar análise e interpretação de dados experimentais, além de possuir um conjunto de ferramentas destinadas para o planejamento, análise e otimização de processos. A utilização de um planejamento experimental se mostra uma ferramenta muito interessante para avaliar os impactos dos fatores sobre as respostas e definir estratégias futuras. Além disso, permite avaliar se as faixas dos níveis escolhidos estão adequadas de modo a obter a máxima eficiência [6].

Neste sentido, este trabalho tem por objetivo utilizar o software Statistica para otimizar o processo de extração de antocianinas de um resíduo de uvas através de aplicação da MSR. Os extratos foram obtidos por extração sólido-líquido [3] e o teor de antocianinas extraídas foi determinado pelo método espectrofotométrico do pH único [8].

### Planejamento experimental

Para avaliar a influência dos fatores temperatura, tipo de solvente, concentração de solvente e concentração de HCl no processo de extração de antocianinas, um delineamento fatorial fracionado  $2^{4-1}$  foi utilizado (Tabela 1). Com o resultado desta análise preliminar foram selecionados os fatores a serem ajustados de modo a maximizar o teor de antocianinas extraídas.

Para esta etapa, um delineamento composto central foi realizado com 3 repetições no ponto central para verificar a linearidade da resposta e para o cálculo do erro experimental.

Tabela 1: Delineamento fatorial fracionado  $2^{4-1}$  para seleção das variáveis.

Ensaio	Temperatura (°C)	Solvente extrator	Concentração de solvente (%)	Concentração de HCl (N)
1	4	Etanol	30	0,5
2	4	Etanol	70	1,5
3	50	Etanol	70	0,5
4	50	Etanol	30	1,5
5	4	Acetona	70	0,5
6	4	Acetona	30	1,5
7	50	Acetona	30	0,5
8	50	Acetona	70	1,5

Através dos resultados do planejamento foram determinados os coeficientes de regressão para a resposta de interesse do processo e realizada análise de variância (ANOVA) a um nível de significância de 5%. Os dados experimentais foram ajustados num modelo polinomial, segundo a equação (1), considerando os componentes lineares, e de interação que apresentam efeito significativo sobre a resposta [5].

$$Y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i < j} \beta_{ij} x_i x_j \quad (1)$$

onde  $\beta_0$  é definido como uma constante,  $\beta_i$  é o coeficiente linear e  $\beta_{ij}$  é o coeficiente do produto de interação;  $x_i$  e  $x_j$  são os níveis das variáveis independentes (fatores) enquanto  $k$  é o número de fatores estudados.

Foram utilizados os critérios de coeficiente de determinação ajustado, teste de *Lack of Fit*, teste de verificação de curvatura e gráficos de distribuição dos resíduos para verificar a qualidade do ajuste dos dados experimentais ao modelo proposto. O planejamento experimental e a análise dos dados foram realizados no *software* Statistica 7.0 (StatSoft®, Inc. Tulsa, OK, EUA (1.02.01.032/6718)) e os efeitos de variabilidade nas respostas observadas devido a fatores não controláveis foram minimizados por randomização na ordem de execução dos ensaios.

## Resultados e Discussão

Diante da análise estatística dos dados, as variáveis elencadas neste estudo não foram significativas para a resposta a um nível de significância de 5%. Porém, avaliando os valores dos efeitos de cada variável pode-se constatar que a temperatura de extração exerce grande influência no teor de antocianinas, quando comparada aos demais fatores, sendo este um efeito negativo, ou seja, quanto maior a temperatura, menor o teor de pigmento extraído. Por outro lado, o tipo de solvente utilizado nas extrações tem pouco ou nenhum efeito sobre a resposta (Figura 1).

Uma vez identificados os efeitos que tem maior significância, se exclui a variável tipo de solvente, tornando o planejamento fatorial fracionado  $2^{4-1}$  num planejamento fatorial completo  $2^3$ . Além disso, foi adicionado ao planejamento experimental 3 repetições no ponto central para verificação da linearidade da superfície de resposta e cálculo do erro experimental.

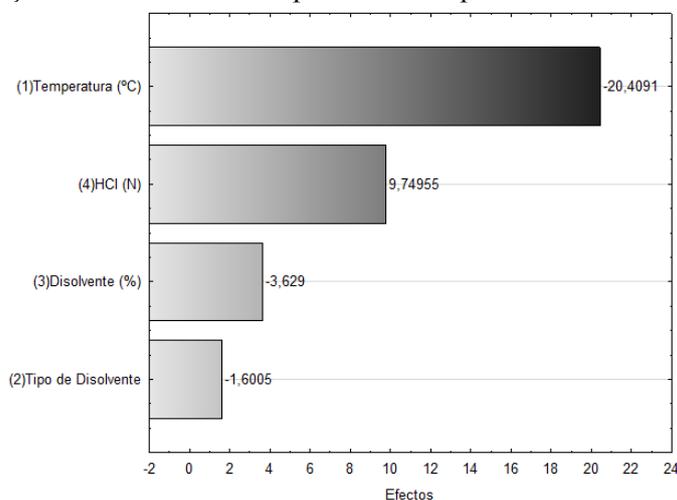


Fig. 1: Gráfico de Pareto para antocianinas totais (mg cianidina-3-glucosídeo/100 g)

A análise de variância do planejamento fatorial completo  $2^3$  confirma que o fator temperatura tem efeito significativo para a resposta, a um nível de significância de 5%. Além disso, o efeito de interação entre as variáveis temperatura e concentração de solvente também exerce influência significativa no teor de antocianinas. Ainda, pode-se concluir que o modelo proposto é adequado ao processo, visto que o teste *Lack of Fit* não foi significativo ( $p > 0,05$ ) e que o ajuste apresenta alto coeficiente de determinação ajustado ( $R^2 = 0,955$ ). A análise de variância indica que a curvatura não é significativa ( $p > 0,05$ ), de modo que a ordem do modelo proposto é adequada, não havendo a necessidade de incluir termos quadráticos ao modelo estatístico.

Tabla 2: ANOVA do planejamento fatorial completo 2<sup>3</sup>.

Fator	<sup>1</sup> SC	<sup>2</sup> GL	<sup>3</sup> CM	F-Ratio	p-Valor
Curvatura	106,552	1	106,552	6,68697	0,122635
(1) Temperatura (°C)	833,067	1	833,067	52,28130	0,018595
(2) Solvente (%)	26,339	1	26,339	1,65299	0,327316
(3) HCl (N)	190,107	1	190,107	11,93069	0,074564
(1) com (2)	1260,633	1	1260,633	79,11431	0,012405
(1) com (3)	283,541	1	283,541	17,79439	0,051864
(2) com (3)	2,819	1	2,819	0,17694	0,714907
Lack of Fit	5,123	2	5,123	0,32152	0,627850
Erro Puro	31,869	2	15,934	-	-
Total	2740,051	10	-	-	-
R <sup>2</sup> : 0,9865		R <sup>2</sup> adj: 0,955			

<sup>1</sup>SC: Soma dos quadrados<sup>2</sup>GL: Graus de liberdade<sup>3</sup>CM: Quadrado médio

O modelo de regressão obtido é apresentado na equação 2, onde  $x_1$  é a variável temperatura,  $x_2$  a variável concentração de solvente e  $x_3$  a variável concentração de HCl e  $Y$  é a concentração de antocianinas totais.

$$Y = -4.038 + 1,43847 x_1 + 0,58672 x_2 + 20,75879 x_3 - 0,02728 x_1 x_2 - 0,51768 x_1 x_3 + 0,05937 x_2 x_3 \quad (2)$$

A partir do modelo de regressão pode-se prever a resposta para o processo para qualquer condição experimental, dentro do intervalo das variáveis estipulado. A Tabela 3 apresenta os valores experimentais e os valores preditos pelo modelo.

Tabela 3: Valores observados e valores preditos pelo modelo.

Ensaio	Valor Experimental	Valor predito
1	27,08	26,28
2	70,21	69,41
3	13,76	12,96
4	40,34	39,54
5	45,77	46,56
6	45,95	46,75
7	42,08	42,88
8	11,19	11,99

A análise dos resíduos é necessária para confirmar que os pressupostos para a análise de variância sejam atendidos. Desta forma, pode-se verificar na Figura 2 que o gráfico de probabilidade normal dos resíduos indicam uma distribuição normal. Da plotagem dos resíduos conclui-se ainda que não há uma tendência positiva ou negativa dos dados, indicando natureza aleatória dos resíduos e variância constante (Figura 3).

Diante desta análise e sendo possível construir as superfícies de respostas (Figuras 4 a 6) pode-se definir a região de interesse para otimização do processo. As superfícies de respostas representam o efeito de duas variáveis independentes sobre a resposta, onde o valor da terceira variável é fixa em seu nível zero (valor central).

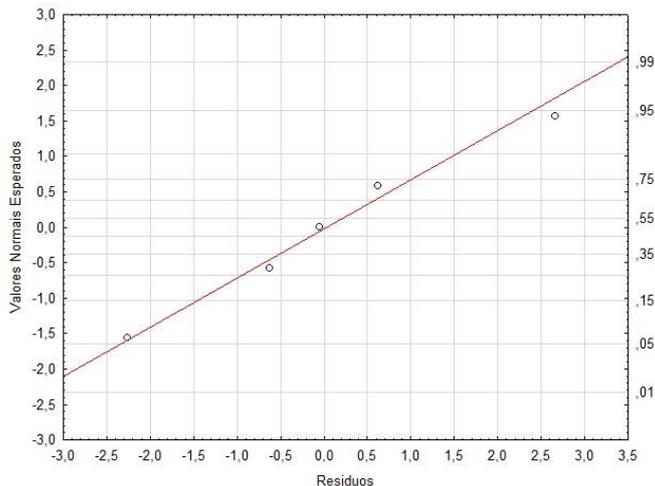


Fig. 2: Distribuição dos resíduos em torno da reta normal.

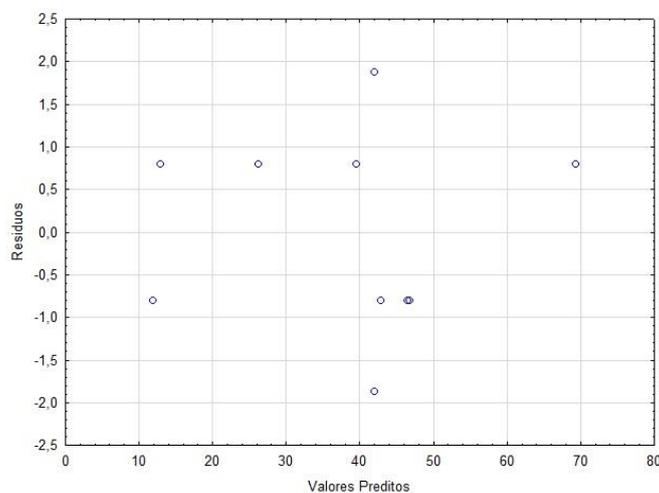


Fig. 3: Valores preditos *versus* resíduos.

Através da análise destas superfícies, fica evidente que o efeito das variáveis concentração de solvente e concentração de HCl não se mostraram significativos para a resposta (Figura 6), não havendo, portanto, relação entre estes dois fatores. Além disso, pode-se constatar uma tendência para uma região ótima do teor de antocianinas, sendo esta região mais evidente Figura 7.

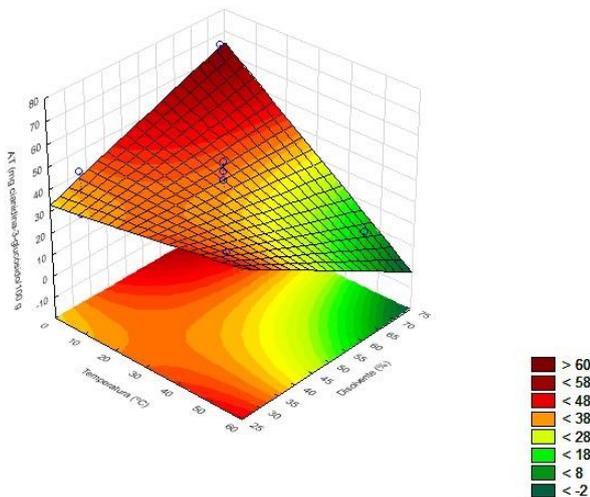


Fig. 4: Superfície de resposta para as variáveis temperatura e concentração de solvente.

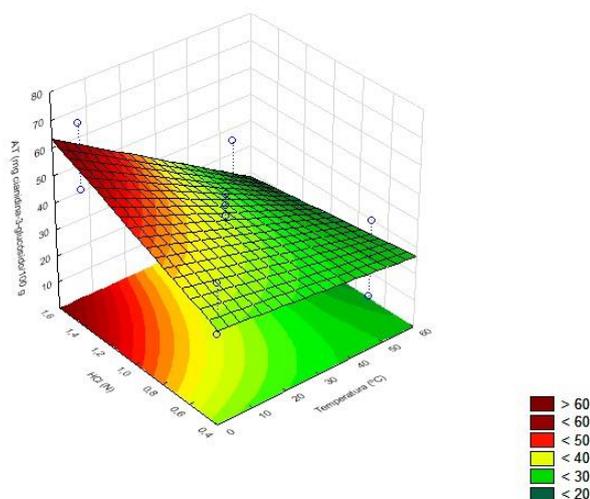


Fig. 5: Superfície de resposta para as variáveis temperatura e concentração de HCl.

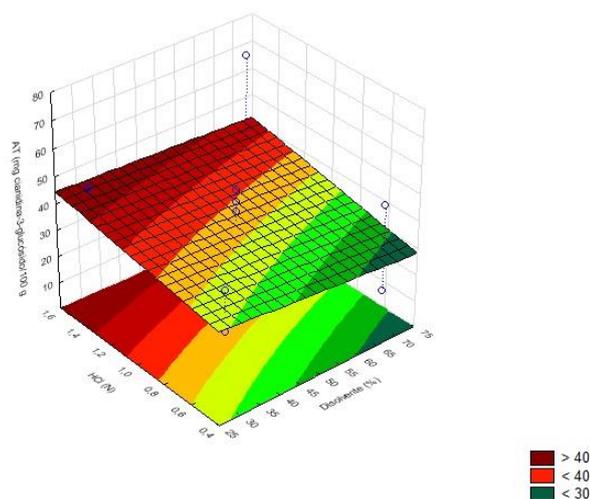


Fig. 6: Superfície de resposta para as variáveis concentração de HCl e concentração de solvente.

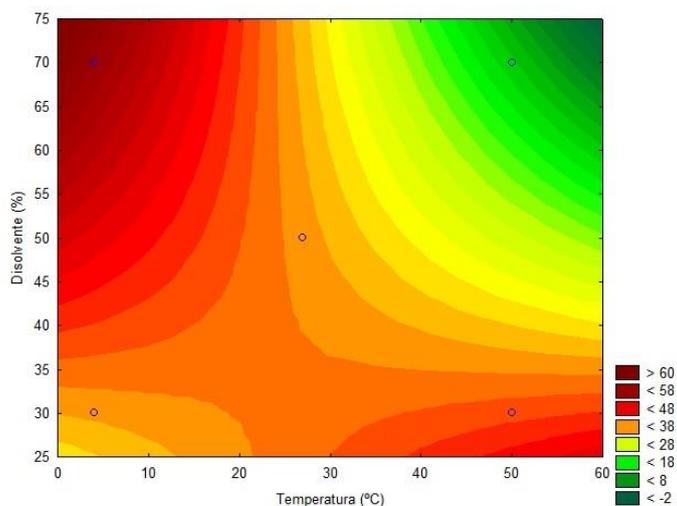


Fig. 7: Curva de contorno para as variáveis temperatura e concentração de solvente.

Ficam nítidas duas regiões para onde um futuro planejamento experimental pode ser deslocado, a fim de se otimizar o processo de extração. Porém, um deslocamento para a região de temperatura elevada (40 – 60°C) e de baixa concentração de solvente (25 – 30%) nos levaria

para condições mais restritas de operação, visto que os valores de temperatura não devem ser estendidos acima de 60 – 70°C devido a problemas de degradação térmica.

Diante destas constatações, fica claro que deve-se trabalhar na região de baixa temperatura e elevada concentração de solvente, indo ao encontro com a metodologia proposta por Fuleki e Francis [3]. Nestas condições, o teor de antocianinas extraídas foi de  $75,14 \pm 4,27$  mg cyn-3-glu/100 g de resíduo, sendo que o valor predito pelo modelo foi de 70,67 mg cyn-3-glu/100 g.

### Conclusões

A partir dos dados apresentados pode-se concluir que o programa Statistica é uma boa ferramenta para a realização de planejamento de experimentos, análise estatística dos dados e predição de condições operacionais de processos. A metodologia de superfície de resposta possibilitou a otimização do processo de extração, se mostrando uma ferramenta bastante eficaz no planejamento e na análise do processo. Os efeitos das variáveis selecionadas se mostraram significativas para a resposta, podendo-se ajustar um modelo de regressão polinomial para predição dos resultados, com bom coeficiente de correlação.

### Referências

- [1] BARROS NETO, Benício de; SCARMINIO, Ieda Spacino; BRUNS, Roy Edward. Como fazer experimentos: pesquisa e desenvolvimento na ciência e na indústria.4. ed. Porto Alegre: Bookman, 2010. 413 p.
- [2] CONSTANT, Patrícia B. L. Extração, caracterização e aplicação de antocianinas de açai (*Euterpe oleraea*, M.). 2003. 199 p. Tese (Doutorado em Ciência e Tecnologia de Alimentos), Universidade Federal de Viçosa, Viçosa, 2003.
- [3] FULEKI, T. & FRANCIS, F. J. Quantitative Methods for anthocyanins. Extraction and determination of total anthocyanin in cranberries. *Journal of Food Science*. v. 33, p. 72-77, 1968.
- [4] LOPES, Toni J. et al. Antocianinas: uma breve revisão das características estruturais e da estabilidade. *Bras. Agrociência*, v. 13, n.3, p.291-297, 2007.
- [5] MONTGOMERY, Douglas C; RUNGER, George C. Estatística aplicada e probabilidade para engenheiros.2. ed. Rio de Janeiro: LTC, 2003. xii, 463 p.
- [6] RODRIGUES, Maria Isabel; IEMMA, Antonio Francisco. Planejamento de experimentos e otimização de processos. São Paulo: Casa do Pão, 2005. 326 p.
- [7] RODRIGUEZ-SAONA, L. E.; WROLSTAD, R. E. Extraction, Isolation, and Purification of Anthocyanins. *Current Protocols in Food Analytical Chemistry*, 2001.
- [8] TEIXEIRA, L. N.; STRINGHETA, P. C.; OLIVEIRA, F. A. de. Comparação de métodos de extração para quantificação de antocianinas. *Revista Ceres*, v. 55, n. 4, p. 297-304, jul-ago 2008.



# SOFiA: Um Algoritmo Baseado na Dinâmica do Processo de Formação de Opinião

Luiz S. Ochi, Willian S. Girão,

Instituto de Computação, IC, UFF

24210-310, Niteroi, RJ

E-mail: williansg@id.uff.br, satoru@ic.uff.br

**Resumo:** *Apresentamos neste trabalho um novo método para otimização global intitulado Algoritmo de Formação Social de Opinião (Social Opinion Formation Algorithm - SOFiA), inspirado pela dinâmica social do processo de formação de opinião. O algoritmo é avaliado através de um estudo comparativo envolvendo dois outros algoritmos bem estabelecidos na literatura, Lobo Cinza e Enxame de Partículas, assim como um conjunto de funções teste utilizadas para avaliar o comportamento dos algoritmos em diferentes configurações de espaços de busca. Os resultados experimentais obtidos mostram que o algoritmo proposto não só é capaz de encontrar os ótimos globais para muitas das funções teste, entregando resultados altamente competitivos, mas também o faz com tempo de processamento consideravelmente inferior ao dos demais.*

**Palavras-chave:** *Otimização Global, Métodos Estocásticos, Meta-heurística*

## 1 Introdução

Nos últimos anos, meta-heurísticas<sup>1</sup> se tornaram populares e têm obtido sucesso em resolver vários problemas de otimização em diversas áreas, dada a presença comum de funções multimodais complexas que em muitos casos não podem ser otimizadas através de métodos analíticos ou numéricos. Fundamentalmente, há três razões para sua utilização: simplicidade, flexibilidade e capacidade favorável de lidar com mínimos locais. Métodos dessa classe são modelados em sua maioria a partir de mecanismos conceitualmente simples. Os aspectos de interesse associados à esses mecanismos observados no mundo natural são geralmente a nível biológico, como os conceitos de genética nos quais Algoritmos Genéticos (*Genetic Algorithms - GA*) são baseados, ou a nível de interações entre indivíduos, como a movimentação de passáros em um bando, modelada no algoritmo Enxame de Partículas (Particle Swarm Optimization - PSO) ou a matilha de lobos modelada no algoritmo Lobo Cinza (*Grey Wolf Optimizer - GWO*) [2].

Como mostrado no trabalho de Wolpert e Macready [5], "para qualquer algoritmo, qualquer performance elevada em uma classe de problemas é compensada negativamente em outra classe", ou seja, não existe um único algoritmo adequado para todos os problemas de otimização. Assim sendo, há necessidade de contínua pesquisa e desenvolvimento de algoritmos capazes de operar cada vez melhor na maior gama possível de classes de problemas. Guiado por essa necessidade, esse trabalho apresenta um novo algoritmo de Inteligência de Enxames (*Swarm Intelligence - SI*) [1] baseado nos aspectos sociais do processo de formação de opinião de indivíduos.

O trabalho é organizado como segue. Na seção 2, apresentamos a classe de métodos SI, assim como a descrição dos algoritmos utilizados para efeitos de comparação. Na seção 3, inicialmente introduzimos os conceitos associados ao processo de formação de opinião, prosseguindo com a descrição detalhada de nosso algoritmo. Na seção 4, descrevemos as configurações dos

---

<sup>1</sup>Meta-heurísticas são procedimentos que tentam, de forma genérica, encontrar soluções ótimas ou subótimas para problemas de otimização.

## 2 Inteligência de Enxames

Sendo uma das classes de algoritmos de otimização mais populares, SI se refere à técnicas que procuram reproduzir a emergência de comportamento inteligente através de interações entre agentes descentralizados e auto-organizáveis, assim como os vistos em colônias de formigas, ao em vez de depender de puro esforço cognitivo de um único indivíduo. Os algoritmos PSO e GWO são exemplos bem estabelecidos de técnicas pertencentes à essa classe de métodos.

Desenvolvido em 1995, PSO é uma técnica de otimização estocástica inicialmente desenvolvida para simular comportamento social, inspirada por simulações de várias interpretações dos movimentos de organismos em um cadurme ou bando. O algoritmo funciona gerando inicialmente uma população de  $N$  partículas (possíveis soluções), cada uma com configurações de posição e velocidade aleatórias. A cada iteração, cada partícula ajusta sua posição e velocidade ao longo das dimensões do espaço de busca do problema baseada na melhor posição encontrada até o momento pela própria partícula e na melhor posição encontrada por qualquer outra partícula em sua vizinhança.

Inspirado pela hierarquia social dos lobos Cinza (que consiste de lobos Alphas, Betas, Deltas e Ômegas), assim como seus mecanismos de caça, em 2014 o GWO foi proposto para problemas de otimização contínua. Nesse algoritmo, a melhor solução é considerada um Alpha, enquanto a segunda e terceira melhores são Beta e Delta, respectivamente, e essas soluções são as que guiam o processo de caça (navegação pelo espaço de busca), enquanto que as outras soluções são Ômegas. O processo de otimização é guiado por uma sequência de ações específicas dos membros da matilha, definido em três fases: rastreamento, caça e ataque.

## 3 Influencia Social e Formação de Opinião

Como definido por Mehdi Moussad [4], "*influência social é o processo pelo qual indivíduos adaptam suas opiniões, revisam suas crenças ou mudam seu comportamento como resultado de interações sociais*". Ou seja, indivíduos dependem da observação de outros para ajustar suas opiniões e comportamento. Quando expostos à opiniões de pessoas com certo nível de conhecimento em um assunto específico, pessoas tendem a analisar, filtrar e incorporar as novas informações às quais foram expostas para assim adaptar suas próprias opiniões. Esse tipo de processo pode resultar em padrões complexos de formação e polarização de opiniões [3].

As heurísticas subjacentes ao processo de adaptação de opinião foram estudadas e modeladas no trabalho "*Social Influence and the Collective Dynamics of Opinion Formation*" [4], onde métodos experimentais inspirados por psicologia social e conceitos teóricos de sistemas complexos, típicos de física e estatística, foram utilizados para conduzir experimentos controlados, que procuraram descrever os mecanismos da influência social, e embasar a elaboração de um modelo.

### 3.1 SOFiA

Inspirado pelo processo de adaptação de opinião mencionado anteriormente, desenvolvemos um algoritmo simples que requer o ajuste de apenas um parâmetro<sup>2</sup>. Nosso algoritmo é basicamente dirigido por duas variáveis específicas: a *distância normalizada* ( $\Delta O_{ij}$ ) entre opiniões e a *diferença em confiança* ( $\Delta C_{ij}$ ), que são definidas nas equações (1) e (2), respectivamente. No que segue,  $i$  e  $j$  representam duas soluções candidatas diferentes, codificadas como vetores *d-dimensionais*, sendo cada dimensão limitada pelo espaço de busca do problema. Sempre iremos

---

<sup>2</sup>Consideramos parâmetros que dependem do algoritmo (e não parâmetros comuns aos algoritmos da classe SI como, por exemplo, *tamanho da população*).

Anais do I Encontro Regional de Matemática Aplicada e Computacional do Vale do Itajaí(ERMAC-2019), Blumenau-SC nos referir à  $i$  como sendo o indivíduo sendo influenciado, enquanto que  $j$  será o influenciador. Usaremos os termos *solução candidata*, *indivíduo*, e *ponto* de forma intercambiável.

A *distância normalizada* mede o grau de similaridade entre duas respostas. Em nossa abstração, essa diferença entre opiniões é tratada como a distância euclidiana entre dois pontos no espaço de busca.

$$\Delta O_{ij} = \frac{1}{\sum_{x=1}^D \sqrt{(j_x - i_x)^2}} \times 0.2 \quad (1)$$

A capacidade de um indivíduo influenciar outro é computada como uma função da *distância normalizada* e a *diferença em confiança*. A *confiança* de uma solução candidata  $y$  corresponde ao valor da função  $f_\alpha$  avaliada nesse ponto, ou seja,  $C_i = f_\alpha(i)$ .

$$\Delta C_{ij} = \frac{|f_\alpha(i) - f_\alpha(j)|}{f_\alpha(i)} \quad (2)$$

O algoritmo é iniciado com uma população randômica de  $n$  pontos uniformemente distribuídos no espaço de busca de  $f_\alpha$ . Esses pontos são então ordenados com base em seus respectivos valores de confiança. A fase de exploração<sup>3</sup> consiste em selecionar os melhores indivíduos. Esses correspondem às regiões de interesse no espaço de busca. Na fase de exploração<sup>4</sup>, as áreas de interesse são vistas cada uma como um indivíduo  $j$  capaz de ter um impacto considerável na mudança de opinião de um outro indivíduo  $i$ . A exploração da área de interesse considerada é feita através da polarização da opinião de um indivíduo (que não seja um influenciador) baseada na opinião de  $j$ , o que se traduz em aproximar  $i$  de  $j$ .

A quantidade de movimento para uma dada componente do vetor solução é calculada como  $U_{i_x j_x} = (j_x - i_x) \times r$ , onde  $i_x$  e  $j_x$  são os  $x$ -ésimos valores da componente dos indivíduos  $i$  sendo influenciados pelo influenciador  $j$ , respectivamente, e  $r$  é um número randômico distribuído uniformemente no intervalo  $[0.5, 1]$ . Assim,  $i$  é aproximado de  $j$  por um valor em um intervalo proporcional à distância entre os dois pontos, o que torna possível não somente explorar o espaço imediatamente próximo de  $j$  (quando  $r \approx 1.0$ ) mas também explorar novas regiões de interesse que possam existir entre os dois pontos (quando  $r \approx 0.5$ ). Ao introduzir um procedimento que muda o sinal da coordenada de um ponto, nós também definimos um intervalo adicional na vizinhança de  $j$  que é simetricamente oposto ao definido na equação.

Após definir os influenciadores, o algoritmo procede a determinar quem da população será influenciado por qual dos influenciadores. Para um dado influenciador  $I_j$  e um indivíduo  $x_i$  ( $i, j = 1, 2, \dots, n$ ), a quantidade de influência que  $I_j$  têm sobre  $x_i$  é computada como  $I'_{ij} = \Delta C_{x_i I_j} + \Delta O_{x_i I_j}$ . A decisão de qual dos influenciadores  $I_j$  irá influenciar a opinião de  $x_i$  é probabilística: quanto maior  $I'_{ij}$ , maior a probabilidade dele ser escolhido como influenciador de  $x_i$ .

Para prevenir convergência prematura para soluções subótimas, a diversidade é mantida ao aplicar dois métodos. O primeiro deles, aplicado durante a etapa de atualização dos componentes dos vetores solução, muda o sinal do valor da componente atualizada com 30% de probabilidade. Testes experimentais durante a modelagem do algoritmo mostraram que esse procedimento ajuda a escapar de locais ótimos. Como acontece no fenômeno social, um indivíduo pode adaptar sua opinião ao assumir que a opinião de um perito está parcialmente correta, ou seja, o indivíduo pode manter aspectos de sua opinião intactos, enquanto que outros são modificados. O segundo método simula esse comportamento: o procedimento *Compromise* combina aleatoriamente pares de soluções para produzir novas. O Algoritmo 1 resume todo o processo algorítmico de SOFiA.

<sup>3</sup>A fase de *exploração* é referente ao processo de cobrimento do espaço de busca à procura de bons ótimos locais, idealmente próximos ao ótimo global.

<sup>4</sup>*Exploração* é o processo de refinamento de uma solução com o intuito de alcançar valores cada vez mais próximos do ótimo global.

**Algoritmo 1** SOFiA

---

```

Gera população inicial randômica  $X = \{x_1, x_2, \dots, x_n\}$ 
while iteração  $\leq iter_{max}$  do
  Ordena população X baseando-se nas confianças dos indivíduos
  for o  $k$  melhor valor de confiança do
     $I_j \leftarrow j$  melhor solução ( $j = 1, 2, \dots, k$ )
  end for
  for all  $x_i \in X$  que não sejam influenciadores do
    for all influenciadores do
       $I'_{ij} \leftarrow I_j$  influência em  $x_i$ 
      Calcule a probabilidade de  $I_j$  influenciar  $x_i$  proporcional a  $I'_{ij}$ 
    end for
     $\alpha \leftarrow$  selecione um dentre os influenciadores  $I_j$ 
     $\beta \leftarrow 1$ 
    Mude o sinal de  $\beta$  com 30% de probabilidade
    for cada componente  $y$  de  $x_i$  do
       $x_{i_y} \leftarrow (x_{i_y} + U_{x_{i_y}} \alpha_y) \times \beta$ 
    end for
  end for
  Compromise( $m, X$ )
  iteração  $\leftarrow$  iteração +1
end while
return Melhor solução

```

---

## 4 Experimentos e Resultados

Dada a natureza estocástica das meta-heurísticas apresentadas neste trabalho, se faz necessário um extensivo estudo experimental para avaliação de performance. Para tal avaliação, utilizamos um conjunto de vinte funções clássicas. Essas funções possuem diferentes níveis de complexidade, o que as tornam adequadas para examinar o comportamento dos algoritmos, mais especificamente como eles se comportam em diferentes espaços de busca. Esse grupo de funções é apresentado na Tabela 1.

Cada algoritmo utilizado nos experimentos foi executado cinquenta vezes para cada uma das funções. Os parâmetros *tamanho da população* e *número máximo de iterações* foram ajustados para vinte e quinhentos, respectivamente. Para o algoritmo GWO esses são os únicos parâmetros a serem ajustados. Experimentos durante o projeto do algoritmo SOFiA mostraram que o número estimado de influenciadores deve ser proporcional à 15% do tamanho da população (o que resulta em três influenciadores para uma população de vinte indivíduos).

O algoritmo PSO necessita do ajuste de cinco outros parâmetros: velocidade máxima permitida das partículas, coeficiente de aceleração cognitiva, coeficiente de aceleração social, e pesos de inércia máximo e mínimo. Os valores para esses parâmetros foram selecionados de acordo com os valores mais frequentes na literatura: 6, 2, 2, 0.9, 0.2, respectivamente.

### 4.1 Funções Unimodais

O primeiro grupo de funções é comumente utilizado [2] para avaliar a capacidade de exploração de um algoritmo. Dado seu caráter unimodal, as funções  $f_1$  até  $f_7$  servem para avaliar o quão perto do verdadeiro valor global ótimo os algoritmos conseguem chegar.

Com os resultados presentes na Tabela 2 mostra-se que SOFiA possui capacidade de exploração superior, dado que nosso algoritmo supera ambos os outros em seis de sete funções. É interessante notar que para as funções  $f_1$  até  $f_4$  SOFiA foi capaz de encontrar o ótimo global em todas as cinquenta execuções. Além disso, para a função  $f_5$  o algoritmo proposto encontra

Função	D	Intervalo	$f_{min}$
$f_1(x) = \sum_{i=1}^d x_i^2$	30	[-100,100]	0
$f_2(x) = \sum_{i=1}^d  x_i  + \prod_{i=1}^d  x_i $	30	[-10,10]	0
$f_3(x) = \sum_{i=1}^d (\sum_{j=i}^d x_j)^2$	30	[-100,100]	0
$f_4(x) = \max_i \{  x_i , 1 \leq i \leq d \}$	30	[-100,100]	0
$f_5(x) = \sum_{i=1}^{d-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$	30	[-30,30]	0
$f_6(x) = \sum_{i=1}^d [(x_i + 0.5)]^2$	30	[-100,100]	0
$f_7(x) = \sum_{i=1}^d i x_i^4 + random[0, 1]$	30	[-1.28, 1.28]	0
$f_8(x) = \sum_{i=1}^d -x_i \sin \sqrt{ x_i }$	30	[-500,500]	-12569.5
$f_9(x) = \sum_{i=1}^d [x_i^2 - 10 \cos(2\pi x_i) + 10]$	30	[-5.12, 5.12]	0
$f_{10}(x) = -20 \exp(-0.2 \sqrt{\frac{1}{d} \sum_{i=1}^d x_i^2}) - \exp(\frac{1}{d} \sum_{i=1}^d \cos(2\pi x_i)) + 20 + e$	30	[-32,32]	0
$f_{11}(x) = \frac{1}{4000} \sum_{i=1}^d x_i^2 - \prod_{i=1}^d \cos(\frac{x_i}{\sqrt{i}}) + 1$	30	[-600,600]	0
$f_{12}(x) = \frac{\pi}{d} \{ 10 \sin^2(\pi \mathbf{y}_i) + \sum_{i=1}^{d-1} (\mathbf{y}_i - 1)^2 [1 + 10 \sin^2(\pi \mathbf{y}_{i+1}) + (\mathbf{y}_n - 1)^2] \} + \sum_{i=1}^d \mathbf{u}(x_i, 10, 100, 4)$	30	[-50,50]	0
$f_{13}(x) = 0.1 \{ \sin^2(3\pi x_i) + \sum_{i=1}^d (x_i - 1)^2 [1 + \sin^2(3\pi x_i + 1)] + (\mathbf{x}_n - 1)^2 [1 + \sin^2(2\pi \mathbf{x}_n)] \} + \sum_{i=1}^d \mathbf{u}(x_i, 5, 100, 4)$	30	[-50,50]	0
$f_{14}(x) = (\frac{1}{500} + \sum_{j=1}^{25} \frac{1}{j + \sum_{i=1}^2 (x_i - a_{ij})^6})^{-1}$	2	[-65.536, 65.536]	$\approx 1$
$f_{15}(x) = \sum_{i=1}^{11} [a_i - \frac{x_1(b_i^2 + b_i x_2)}{b_i^2 + b_i x_3 + x_4}]^2$	4	[-5,5]	$\approx 0.0003075$
$f_{16}(x) = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_i^6 + x_1x_2 - 4x_2^2 + 4x_4^4$	2	[-5,5]	-1.0316285
$f_{17}(x) = (x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6)^2 + 10(1 - \frac{1}{8\pi}) \cos(x_1) + 10$	2	[-5,5]	$\approx 0.398$
$f_{18}(x) = [1 + (x_1 + x_2 + 1)^2 (19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)] \times [30 + (2x_1 - 3x_2)^2 \times (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)]$	2	[-2,2]	3
$f_{19}(x) = -\sum_{i=1}^4 c_i \exp[-\sum_{j=1}^3 a_{ij}(x_j - p_{ij})^2]$	3	[0,1]	-3.86278
$f_{20}(x) = -\sum_{i=1}^4 c_i \exp[-\sum_{j=1}^6 a_{ij}(x_j - p_{ij})^2]$	6	[0,1]	-3.32237

Tabela 1: Formas analíticas, número de dimensões (D), intervalos dos espaços de busca e seus respectivos valores ótimos globais ( $f_{min}$ ) do conjunto de funções teste [2].

valor médio bastante superior aos demais, assim como o melhor resultado geral dentre todas as execuções quando comparado aos outros, cujas soluções médias e melhores soluções não superam o pior valor encontrado por SOFiA.

## 4.2 Funções Multimodais

### 4.2.1 Locais Mínimos Massivos

De modo a melhor investigar o comportamento do algoritmo, a segunda série de experimentos foi executada com um grupo de funções multimodais. Esse grupo de funções é tido como o mais problemático para vários algoritmos de otimização [5]. As funções  $f_8$  até  $f_{13}$  possuem um número massivo de locais ótimos, o que as tornam adequadas para testar a capacidade de exploração dos algoritmos, assim como sua habilidade de lidar com os vales dos mínimos locais.

Novamente, os valores na Tabela 3 mostram que SOFiA continua performando de maneira competitiva, tendo melhores resultados para duas das seis funções e empatando em outras três.

Nós notamos que, dados os resultados mostrados até então, a capacidade de exploração de nosso método se mostra capaz de lidar com espaços de busca contendo grande número de locais ótimos sem debilitar consideravelmente sua capacidade de explorar as áreas de interesse encontradas. Em outras palavras, nosso algoritmo mostra um equilíbrio interessante entre exploração e exploração.

### 4.2.2 Locais Mínimos Reduzidos

O terceiro grupo de experimentos foi feito utilizando um conjunto de funções cujos espaços de busca possuem baixa dimensionalidade e poucos locais mínimos.

Os resultados observados na Tabela 4 mostram que todos os algoritmos obtiveram performance similar quando avaliadas suas acurácias médias, sem diferenças consideráveis, com exceção

F	SOFiA			GWO			PSO		
	Média	Melhor	Pior	Média	Melhor	Pior	Média	Melhor	Pior
$f_1$	0	0	0	1.30E-21	1.35E-30	5.66E-20	0.02	0.002	0.1
$f_2$	0	0	0	6.75E-16	3.23E-19	2.55E-14	7.3	0.1	40.1
$f_3$	0	0	0	4.92E-17	7.11E-25	1.09E-15	189.3	60.8	444.9
$f_4$	0	0	0	3.93E-10	4.92E-14	1.41E-08	1.64	0.88	2.39
$f_5$	16.8	0.01	28.9	28.3	28.1	28.9	1964.2	37.52	90067.6
$f_6$	0.009	0.000005	0.2	5.53	4.22	6.88	0.05	0.003	1.15
$f_7$	0.002	0.00009	0.01	0.0001	0.000004	0.0007	7.23	0.08	48.4

Tabela 2: Resultados para funções unimodais. Valores empatados são mostrados sublinhados.

F	SOFiA			GWO			PSO		
	Média	Melhor	Pior	Média	Melhor	Pior	Média	Melhor	Pior
$f_8$	-4069.2	-5355.8	-2957.9	-2430	-4050	-1740	-4914.6	-7789.8	-2387.9
$f_9$	<u>0</u>	0	0	<u>0</u>	0	0	125.02	61.77	208
$f_{10}$	4.00E-16	4.00E-16	4.00E-16	5.59E-09	4.00E-15	2.79E-07	1.01	0.05	2.04
$f_{11}$	<u>0</u>	0	0	<u>0</u>	0	0	0.01	0.0001	0.04
$f_{12}$	0.19	0.06	0.41	1.09	0.61	1.49	<u>0.26</u>	0.04	0.9
$f_{13}$	<u>0.02</u>	0.005	0.2	2.6	2.9	2.1	<u>0.02</u>	0.001	0.1

Tabela 3: Resultados para funções multimodais. Valores empatados são mostrados sublinhados.

da função  $f_{18}$ , onde SOFiA obteve solução média aproximadamente duas vezes inferior à melhor solução encontrada, mas compensa com os resultados para  $f_{15}$ , onde teve solução média significativamente melhor. Dentre as sete funções, SOFiA obteve soluções superiores para duas e empatou em três.

### 4.2.3 Tempo de Execução

Tendo acesso direto aos códigos desenvolvidos por [2], nosso último grupo de experimentos consistiu em calcular os tempos médios de execução de cada algoritmo. O tempo calculado (em segundos) consiste da média de cinquenta execuções independentes para cada uma das vinte funções.

As médias de tempo na Tabela 5 mostram que SOFiA não somente é capaz de entregar resultados altamente competitivos mas também o faz com tempo de execução consideravelmente inferior, sendo, em média, doze e dez vezes mais rápido que os algoritmos GWO e PSO, respectivamente.

## 5 Conclusão

A simplicidade da técnica proposta neste trabalho torna seu processo de otimização não somente rápido mas também bastante eficaz, dado que nosso algoritmo obteve melhor solução média em 50% das funções avaliadas, empatando em 30% delas e não obtendo o melhor resultado em somente quatro casos: enquanto as soluções encontradas por SOFiA para  $f_8$  e  $f_{18}$  são relativamente inferiores, os resultados para  $f_7$  e  $f_{19}$  não possuem diferenças estatísticas significantes quando comparados com os melhores resultados encontrados. Ainda para essas duas últimas funções mencionadas, nosso método foi de cinco a vinte e uma vezes mais rápido.

F	SOFiA			GWO			PSO		
	Média	Melhor	Pior	Média	Melhor	Pior	Média	Melhor	Pior
$f_{14}$	<u>4.34</u>	0.99	12.6	6.82	1	12.7	<u>4.34</u>	0.99	12.5
$f_{15}$	0.0008	0.0003	0.008	0.02	0.004	0.06	0.32	0.004	0.12
$f_{16}$	<u>-1.0316</u>	-1.0316	-1.03	-1.02	-1.0315	-0.98	<u>-1.0316</u>	-1.0316	-1.0316
$f_{17}$	<u>0.39</u>	0.39	0.39	0.56	0.4	3.51	<u>0.39</u>	0.39	0.39
$f_{18}$	9.04	2.9	33.1	3.7	3	28	4.78	2.99	91.81
$f_{19}$	-3.78	-3.86	-3.08	-3.73	-3.86	-3.36	-3.86	-3.86	-3.85
$f_{20}$	-3.26	-3.42	-1.86	-2.42	-3.08	-1.15	-3.09	-3.32	-1.71

Tabela 4: Resultados para funções de dimensões fixas. Valores empatados são mostrados sublinhados.

F	SOFiA	GWO	PSO	F	SOFiA	GWO	PSO
$f_1$	0.097	2.27	1.71	$f_{11}$	0.18	2.6	2.02
$f_2$	0.077	2.32	1.78	$f_{12}$	0.44	3.17	2.71
$f_3$	0.39	4.7	4.18	$f_{13}$	0.41	3.05	2.5
$f_4$	0.058	2.14	1.62	$f_{14}$	0.74	9.64	9.52
$f_5$	0.17	2.3	1.82	$f_{15}$	0.078	0.83	0.81
$f_6$	0.087	2.28	1.71	$f_{16}$	0.053	0.27	0.25
$f_7$	0.45	2.52	2.01	$f_{17}$	0.021	0.3	0.27
$f_8$	0.31	2.33	1.83	$f_{18}$	0.021	0.34	0.29
$f_9$	0.14	2.38	1.86	$f_{19}$	0.056	1.26	1.18
$f_{10}$	0.16	2.52	2.03	$f_{20}$	0.12	1.41	1.33

Tabela 5: Tempos de execução médios para cinquenta execuções independentes.

Dadas as capacidades demonstradas em nossos testes, SOFiA foi capaz de superar dois algoritmos de otimização bem estabelecidos, mostrando que nosso método possui potencial para ser empregado em outros problemas de otimização global, especialmente aqueles onde tempo de execução seja um fator limitante.

## Referências

- [1] B. Gerardo, “Robots and Biological Systems: Towards a New Bionics?”, Springer, Berlin, 1993.
- [2] M. Seyedali, Grey Wolf Optimizer, *Advances in Engineering Software*, 69 (2014) 46-61.
- [3] M. Michael, Individualization as Driving Force of Clustering Phenomena in Humans, *PLOS Computational Biology*, 6 (2010) 1-8.
- [4] M. Moussaïd, Social Influence and the Collective Dynamics of Opinion Formation, *PLOS ONE*, 8 (2013) 1-8.
- [5] R. Poli, Particle Swarm Optimization, *Swarm Intelligence*, 1 (1995) 33-57.



# Método Implícito de Diferenças Finitas Aplicado ao Processo de Conformação Eletromagnética

**Arthur S. Fagundes,**      **Eduardo V. da Silva**

Departamento de Engenharia Elétrica, UFRGS

90220-011, Porto Alegre, RS

E-mails: arthur.fagundes@ufrgs.br,    eduardo.varriale@ufrgs.br

**Rodrigo Rossi**

Departamento de Engenharia Mecânica, UFRGS

90050-170, Porto Alegre, RS

E-mail: rrossi@ufrgs.br

**Resumo:** Neste trabalho é desenvolvido um algoritmo que usa o método implícito das diferenças finitas para simular o funcionamento do processo de conformação eletromagnética. São realizados testes do algoritmo com comparações analíticas e experimentais.

**Palavras-chave:** Diferenças Finitas; Método Implícito; Conformação Eletromagnética;

**Notação:** Neste artigo,  $\vec{V}$  é um vetor no espaço,  $\tilde{V}$  é uma matriz,  $V_s$  é um escalar com subscrito  $s$ ,  $V_s^k$  é o valor do escalar no instante de tempo  $k$ ,  $V_s(i, j)$  é o valor do escalar nas coordenadas  $(i, j)$ .  $V_s^k(i, j)$  significa implicitamente  $\forall i, \forall j, \forall k$ .

## 1 Introdução

Processos de conformação por explosão e conformação eletromagnética são exemplos de “conformação em alta velocidade”. Estes processos aplicam em um curto intervalo de tempo elevadas quantidades de energia cinética em uma peça de trabalho, causando a sua deformação principalmente por inércia [6]. O processo de conformação eletromagnética, conhecido pela sigla EMF (do inglês *electromagnetic forming*), é realizado em peças de metal condutor. Nele, uma corrente de descarga de um banco de capacitores percorre uma bobina, que gera um campo magnético variável e conseqüentemente forças repulsivas pela interação deste com a peça de trabalho, que propõem a mesma contra um molde. Vários estudos sobre a simulação do processo já foram conduzidos [1], [2], [7]. A contribuição deste trabalho é o desenvolvimento de uma simulação numérica baseada em [1], utilizando o método implícito das diferenças finitas, devido à sua vantagem de ser estável independentemente do período de discretização do tempo [4].

## 2 Modelo matemático do sistema EMF

A máquina de conformação eletromagnética opera com peças de metal condutor, mais comumente cobre ou alumínio. Ela consiste em um banco de capacitores ligado a uma bobina posicionada no mesmo eixo da peça de trabalho. O conjunto pode ser visto na Fig. 1a, na configuração planar, e na figura 1b na configuração tubular.

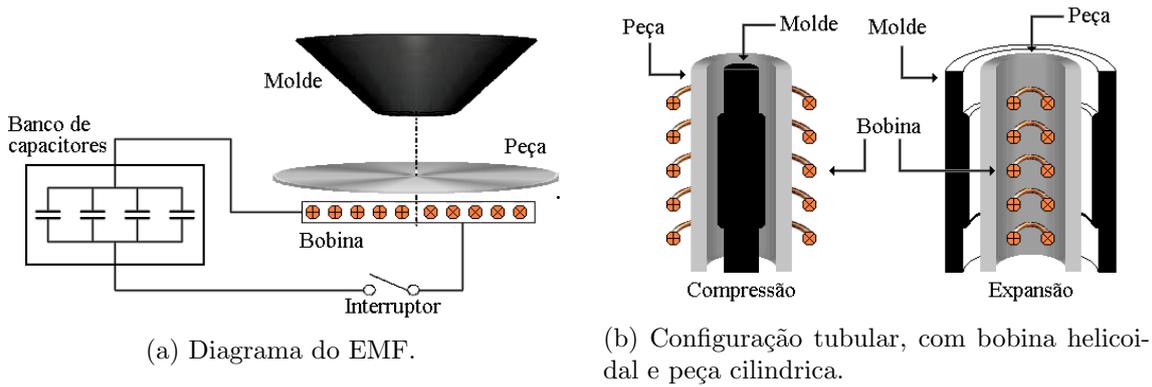


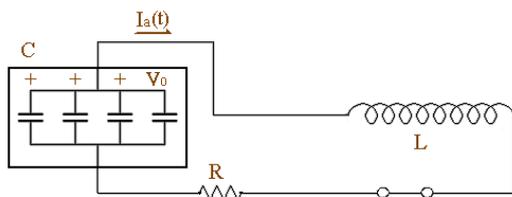
Figura 1

### Hipóteses

O modelo matemático do sistema EMF é baseado nas seguintes suposições:

1. Durante o processo, a peça de trabalho é deslocada. Esse deslocamento causa uma variação na indutância mútua entre bobina e peça. Todavia, como a maior parte da descarga de corrente no circuito ocorre antes de um movimento significativo da peça, o deslocamento da peça é considerado *quasi*-estacionário, e assim a indutância mútua entre bobina e peça é considerada constante. Manea et al. demonstraram que no processo EMF a velocidade só se torna significativa se está na ordem de  $10^7 m/s$  [3].
2. O enrolamento da bobina é discretizado em um número de círculos concêntricos.
3. A geometria de todos os componentes é simétrica em torno de um eixo (axissimétrica). Devido à axissimetria se manter durante a deformação, o campo magnético não varia na direção da circunferência, tornando apropriado um sistema de coordenadas cilíndrico para se descrever as coordenadas de qualquer ponto da bobina ou peça de trabalho.
4. A simulação é separada para cada componente do campo magnético. Os componentes não possuem relações de acoplamento, e podem ser resolvidos separadamente.
5. A variação de temperatura é desconsiderada por não ser suficiente para alterar as propriedades do material.

A modelagem do circuito elétrico é feita partindo das mesmas equações usadas em [1]. Baseando-se na hipótese #1 e que a indutância mútua permanece constante durante a descarga elétrica, o circuito elétrico da EMF é aproximado por um circuito com resistência, indutância e capacitância (RLC) em série equivalente (mostrado na figura abaixo, ao lado das equações), cuja resposta ao impulso (gerado pela comutação do interruptor da Fig. 1a) é dada pela Eq. (1) com as variáveis calculadas em (2), onde  $I_a(t)$  é a corrente de descarga ao longo do tempo,  $V_0$  é a tensão inicial (em  $t = 0$ ) do banco de capacitores,  $C$  a sua capacitância,  $L$  e  $R$  a indutância e resistência equivalentes do circuito.



$$I_a(t) = e^{-\frac{t}{\tau}} I_{a0} \text{sen}(\omega t) \quad (1)$$

$$I_{a0} = V_0 \sqrt{\frac{C}{L}} \quad \tau = 2 \frac{L}{R} \quad \omega = \frac{1}{\sqrt{LC}} \quad (2)$$

O equacionamento do campo eletromagnético, como em [1], parte das equações de Maxwell, mostradas no conjunto de equações (3). A modelagem do campo eletromagnético, ao contrário

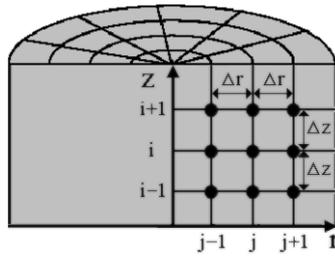


Figura 2: Discretização da peça.

da do circuito elétrico, é de parâmetros distribuídos. Em seu desenvolvimento, serão usadas as componentes de coordenadas cilíndricas das equações de Maxwell para o caso de axissimetria considerado pela hipótese #3, desenvolvidas na seção a seguir.

### Equações do campo eletromagnético em coordenadas cilíndricas

$$\begin{cases} \nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} & (3).1 \\ \nabla \cdot \vec{B} = 0 & (3).2 \\ \vec{B} = \mu \vec{H} & (3).3 \\ \nabla \times \vec{H} = \vec{J} & (3).4 \\ \vec{J} = \sigma(\vec{E} + \vec{v} \times \vec{B}) & (3).5 \end{cases} \quad (3)$$

Considerando-se novamente a suposição #1, a Eq. (3).5 tem o termo  $\vec{v}$  anulado. Substituindo  $\vec{J}$  e  $\vec{H}$  na Eq. (3).4 pelas suas igualdades das eqs. (3).3 e (3).5, aplicando o operador rotacional em ambos os lados da equação resultante, e eliminando  $\nabla \cdot \vec{B} = 0$ , o resultado é a Eq. (4), que é separada em componentes de um sistema de coordenadas cilíndrico dados pelas equações (5) e (6). Para resolver estas equações numericamente, o método implícito de diferenças finitas é utilizado em conjunto com condições de contorno e condições iniciais.

$$-\frac{1}{\mu\sigma} \nabla^2 \vec{B} + \frac{\partial \vec{B}}{\partial t} = 0 \quad (4)$$

$$-\frac{1}{\mu\sigma} \left( \frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{\partial^2}{\partial z^2} \right) B_z + \frac{\partial B_z}{\partial t} = 0 \quad (5)$$

$$-\frac{1}{\mu\sigma} \left( \frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{\partial^2}{\partial z^2} - \frac{1}{r^2} \right) B_r + \frac{\partial B_r}{\partial t} = 0 \quad (6)$$

### Discretização por diferenças finitas

A peça e o tempo devem ser discretizados para se usar o esquema de diferenças finitas nas equações (5) e (6) e resolvê-las numericamente. Para tanto, é criado o sistema de coordenadas  $r \times z$  da Fig. 2, cujas coordenadas se alinham respectivamente ao raio e à altura da peça de trabalho. As equações resultantes são mostradas na Eq. (7). Substituindo-as nas equações (5) e (6), obtém-se (8) e (9). Como  $B^k$  e  $B^{k+1}$  são vetores que carregam os valores de campo magnético instantâneos e sucessivos de todos nós da malha, e se deseja computar os valores sucessivos a partir de valores instantâneos, a solução para obtenção destes valores é isolar  $B^{k+1}$ . Mas alguns pontos contidos em  $B^{k+1}$  são predeterminados, por serem condições de contorno; isto significa que apenas a porção desvinculada a valores de contorno deve ser isolada e resolvida. Antes de prosseguir e criar um vetor com os nós desvinculados, as condições de contorno devem ser definidas.

$$\left\{ \begin{array}{l} \frac{\partial B_z}{\partial z} \approx \frac{\Delta B_z(i, j)}{\Delta z} = \frac{B_z(i+1, j) - B_z(i-1, j)}{2\Delta z} \end{array} \right. \quad (7).1$$

$$\frac{\partial B_z}{\partial r} \approx \frac{\Delta B_z(i, j)}{\Delta r} = \frac{B_z(i, j+1) - B_z(i, j-1)}{2\Delta r} \quad (7).2$$

$$\frac{\partial B_r}{\partial z} \approx \frac{\Delta B_r(i, j)}{\Delta z} = \frac{B_r(i+1, j) - B_r(i-1, j)}{2\Delta z} \quad (7).3$$

$$\frac{\partial B_r}{\partial r} \approx \frac{\Delta B_r(i, j)}{\Delta r} = \frac{B_z(i, j+1) - B_z(i, j-1)}{2\Delta r} \quad (7).4$$

$$\frac{\partial^2 B_z}{\partial z^2} \approx \frac{\Delta \frac{\Delta B_z(i, j)}{\Delta z}}{\Delta z} = \frac{B_z(i+1, j) - 2B_z(i, j) + B_z(i-1, j)}{\Delta z^2} \quad (7).5$$

$$\frac{\partial^2 B_z}{\partial r^2} \approx \frac{\Delta \frac{\Delta B_z(i, j)}{\Delta r}}{\Delta r} = \frac{B_z(i, j+1) - 2B_z(i, j) + B_z(i, j-1)}{\Delta r^2} \quad (7).6 \quad (7)$$

$$\frac{\partial^2 B_r}{\partial z^2} \approx \frac{\Delta \frac{\Delta B_r(i, j)}{\Delta z}}{\Delta z} = \frac{B_r(i+1, j) - 2B_r(i, j) + B_r(i-1, j)}{\Delta z^2} \quad (7).7$$

$$\frac{\partial^2 B_r}{\partial r^2} \approx \frac{\Delta \frac{\Delta B_r(i, j)}{\Delta r}}{\Delta r} = \frac{B_z(i, j+1) - 2B_z(i, j) + B_z(i, j-1)}{\Delta r^2} \quad (7).8$$

$$\frac{\partial B_z}{\partial t} \approx \frac{\Delta B_z^k}{\Delta t} = \frac{B_z^{k+1} - B_z^k}{\Delta t} \quad (7).9$$

$$\frac{\partial B_r}{\partial t} \approx \frac{\Delta B_r^k}{\Delta t} = \frac{B_r^{k+1} - B_r^k}{\Delta t} \quad (7).10$$

$$\tilde{M}_z B_z^{k+1} = \tilde{N}_z B_z^k \quad (8)$$

$$\tilde{M}_r B_r^{k+1} = \tilde{N}_r B_r^k \quad (9)$$

### Condições de contorno

As equações diferenciais devem ter limites espaciais definidos por condições de contorno e também condições iniciais para serem resolvidas. O conjunto de equações (10) mostra as condições de contorno e iniciais do sistema EMF de geometria planar para as equações dos componentes axial e radial do campo magnético. Estas condições são baseadas em aproximações físicas: as condições (10).3 a (10).6 se baseiam no efeito pelicular - *skin effect*. As equações (10).7 e (10).8 tratam da condição de axissimetria. Por último, as equações (10).9 e (10) mostram que o campo magnético na peça é nulo no início da simulação. As equações (10).1 e (10).2 impõem um “campo de contorno” na superfície inferior da peça. O cálculo do campo de contorno é feito com a corrente de descarga, obtida pela Eq. (1), e pelo uso do vetor potencial, conforme detalhado em [1]. As condições de contorno do sistema EMF em geometria tubular são similares, mudando a superfície do campo de contorno.

$$\left\{ \begin{array}{l} B_z^k(0, j) = B_{sz}^k \quad (10).1 \\ B_r^k(0, j) = B_{sr}^k \quad (10).2 \\ B_z^k(n_i, j) = 0 \quad (10).3 \\ B_r^k(n_i, j) = 0 \quad (10).4 \\ B_z^k(i, n_j) = 0 \quad (10).5 \\ B_r^k(i, n_j) = 0 \quad (10).6 \\ \frac{\partial B_z^k(i, j)}{\partial r} \Big|_{j=0} = 0 \quad (10).7 \\ B_r^k(i, 0) = 0 \quad (10).8 \\ B_z^0(i, j) = 0 \quad (10).9 \\ B_r^0(i, j) = 0 \quad (10).10 \end{array} \right. \quad i = 0, \dots, n_i \quad j = 0, \dots, n_j \quad (10)$$

## 2.1 Sistema de equações

As equações (8) e (9), quando reescritas levando em conta a diferença entre os nós desvinculados e os com valores de contorno, resultam nas equações (11) e (12). Ao se realizar a separação do vetor de campo magnético  $\vec{B}$  em vetor de nós desvinculados  $\vec{B}_{ub}$  e vetor de nós da superfície mais próxima da bobina  $\vec{B}_s$ , as matrizes  $\tilde{M}$  e  $\tilde{N}$  também são particionadas em  $\tilde{A}$ ,  $\tilde{C}$  e  $\tilde{D}$  para multiplicar os nós de  $B_{ub}^{k+1}$  e  $B_s^{k+1}$ . Assim, o vetor  $B_{ub}^{k+1}$  pode ser calculado se isolando os demais termos em um lado da equação, e então multiplicando ambos os lados pela inversa de  $\tilde{A}$ , como nas equações (13) e (14). Esta inversão de matriz é a razão deste método ser computacionalmente exigente; contudo, como  $\tilde{A}$  é invariante no tempo, a inversão é realizada apenas uma vez por simulação, e não a cada iteração, o que reduz significativamente o tempo de processamento.

$$\tilde{A}_z B_{ubz}^{k+1} + \tilde{C}_z B_{sz}^{k+1} = \tilde{D}_z B_{ubz}^k \quad (11)$$

$$\tilde{A}_r B_{ubr}^{k+1} + \tilde{C}_r B_{sr}^{k+1} = \tilde{D}_r B_{ubr}^k \quad (12)$$

$$B_{ubr}^{k+1} = \tilde{A}_z^{-1} \left( \tilde{D}_z B_{ubz}^k - \tilde{C}_z B_{sz}^{k+1} \right) \quad (13)$$

$$B_{ubr}^{k+1} = \tilde{A}_r^{-1} \left( \tilde{D}_r B_{ubr}^k - \tilde{C}_r B_{sr}^{k+1} \right) \quad (14)$$

Após o cálculo dos campos magnéticos, deve-se obter a distribuição de corrente induzida na peça. A Eq. (15) é o resultado da expansão do operador rotacional da Eq. (3).4 em coordenadas cilíndricas, da mesma forma que foi feito na Seção 2, e é usada para se calcular o campo magnético  $B_1$  no próximo instante da simulação. A Eq. (16) é a sua implementação no algoritmo usando-se os esquemas de diferenças finitas das equações (7).2 e (7).3.

$$J_\phi = \frac{1}{\mu} \left( \frac{\partial B_r}{\partial z} - \frac{\partial B_z}{\partial r} \right) \quad (15)$$

$$J_b^k(i, j) = \frac{1}{\mu} \left( \frac{B_r^k(i+1, j) - B_r^k(i-1, j)}{2\Delta z} - \frac{B_z^k(i+1, j) - B_z^k(i-1, j)}{2\Delta r} \right) \quad (16)$$

## 3 Resultados

### Comparação com solução analítica

Em [5], a propagação do campo magnético em um cilindro sólido de comprimento infinito, submetido a um campo magnético em sua superfície cuja magnitude segue uma função degrau de amplitude  $B_0$  e com direção em  $z$  é descrita pela Eq. (17), onde  $S_0()$  e  $S_1()$  são funções de Bessel de ordens zero e um,  $q$  é o raio do cilindro e os valores de  $\alpha_n$  podem ser encontrados por meio da Eq. (18).1. A amplitude do campo de contorno é dada pelas equações (18).2 e (18).3.

$$B_z(r, t) = B_0 \left( 1 - 2 \sum_{n=1}^{\infty} \frac{S_0(r\alpha_n)}{q\alpha_n S_1(r\alpha_n)} e^{-\frac{\alpha_n^2 t}{\mu\sigma}} \right) \quad (17)$$

$$S_0(q\alpha_n) = 0 \quad (18).1 \quad \begin{cases} B_{sr} = 0 & (18).2 \\ B_{sz} = step(k) & (18).3 \end{cases} \quad (18)$$

Em uma simulação com os parâmetros listados na Tabela 1, os resultados são gerados e comparados com a solução analítica. A comparação pode ser vista na Fig. 3a, com os resultados numérico e analítico obtidos aos pares em diferentes instantes do tempo.

Tabela 1: Parâmetros da comparação entre o algoritmo e a solução analítica.

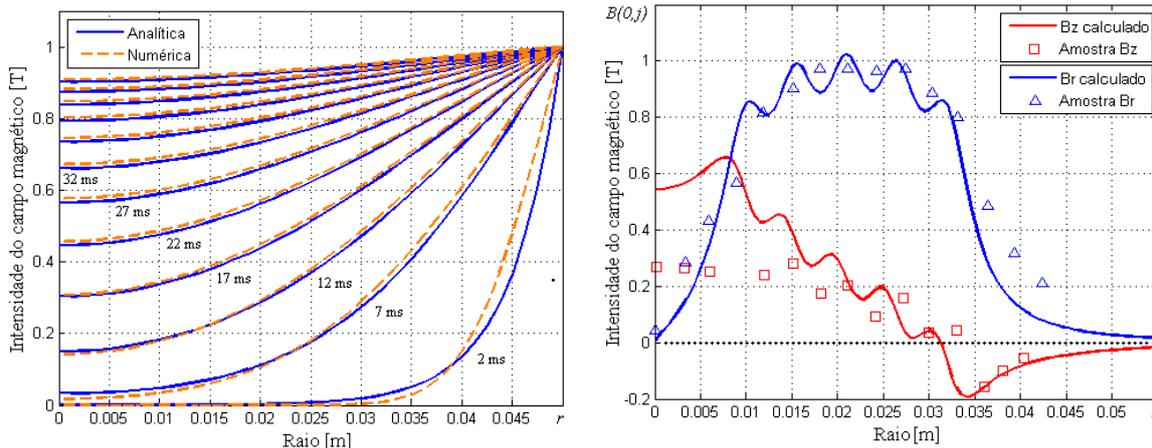
Parâmetros	
Raio da peça	55 mm
Altura (numérico)	$\infty$
Condutividade elétrica	36 MS/m
Permeabilidade magnética	$4\pi \times 10^{-7}$ H/m
Incrementos de malha ( $n_r$ )	20

### Comparação com experimento prático e método explícito

O algoritmo de diferenças finitas é comparado com o experimento conduzido por [2]. Os parâmetros do ensaio são expostos na Tabela 2. Em seu trabalho, sensores magnéticos foram usados para captar o campo magnético gerado por uma EMF no instante do pico de corrente. Os resultados são expostos na Fig. 3b. Também é feita uma comparação com o método explícito de [1]. Usando os mesmos parâmetros, é analisado o campo magnético em dois instantes após o pico. As figuras 4a e 4b mostram a vantagem da estabilidade do método implícito para períodos relativamente elevados de discretização do tempo.

Tabela 2: Tabela de parâmetros do ensaio.

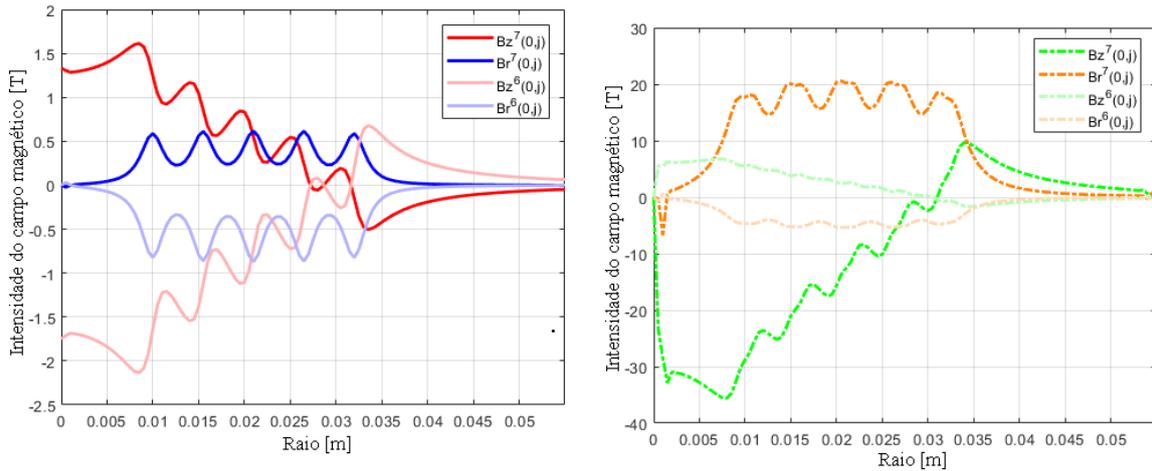
Circuito elétrico		Peça de Al 1050	
Numero de enrolamentos	5	Espessura	3 mm
Menor raio da bobina	10 mm	Raio	55 mm
Maior raio da bobina	32 mm	Condutividade elétrica	36 MS/m
Passo do enrolamento	5.5 mm	Permeabilidade magnética	$4\pi \times 10^{-7}$ H/m <sup>2</sup>
Espaço entre bobina e peça	1,6 mm	Incrementos de malha ( $n_r \times n_z$ )	110x6
Indutância equivalente	0.86 $\mu$ H		
Resistência equivalente	28.5 m $\Omega$		
Capacitância	40 $\mu$ F		
Tensão inicial	6000 V		



(a) Gráfico do campo magnético resultante das soluções analítica e numérica entre os instantes de 2 e 57 ms, em intervalos de 5 ms.

(b) Comparação do método implícito (linha) com resultados experimentais (quadrados e triângulos). [2]

Figura 3



(a) Método implícito em dois instantes de tempo após o pico. Pode-se notar um decaimento do campo da iteração 6 à 7. ( $100 \mu s$  e  $120 \mu s$ ), pois todos os valores do campo ao longo do raio, em ambas as componentes, são menores em módulo no segundo instante.

(b) Método explícito nos mesmos instantes de tempo. Pode-se notar a instabilidade da solução pelo campo estar aumentando após o pico de corrente (todos os valores estão maiores em módulo no segundo instante).

Figura 4

## 4 Conclusões

O método desenvolvido pôde ser validado por comparação com uma solução analítica, teve semelhança com um resultado particular obtido em experimento prático, e ao final foi exposta a conhecida vantagem do uso do método implícito, que é a sua estabilidade numérica para qualquer período de discretização. Esta vantagem permite a discretização da peça em mais pontos com o aumento arbitrário do período de discretização do tempo quando a capacidade computacional for limitada e exigir esta troca.

## Referências

- [1] Siddiqui, Muhammad Ali, "Numerical Modelling and Simulation of Electromagnetic Forming Process, Tese de Doutorado, Universidade de Strasbourg, 2009.
- [2] Nobuo Takatsu and M. Kato and Keijin Sato and Toshimi Tobe, "High-speed forming of metal sheets by electromagnetic force", The Japan Society of Mechanical Engineers, Japan, 1988.
- [3] T.E. Manea and M.D. Verweij and H. Blok, "The importance of the velocity term in the electromagnetic forming process", Proceedings of 27th General Assembly of the International Union of Radio Science, URSI, 2002.
- [4] Richard L. Burden and J. Douglas Faires, "Numerical Analysis", Brooks/Cole, 2010.
- [5] Heinz E. Knoepfel, "Magnetic Fields: A Comprehensive Theoretical Treatise for Practical Use", John Wiley & Sons, 2000.
- [6] Venkata S. Balanethiram and B.Tech. and M.S., "Hyperplasticity: enhanced formability of sheet metals at high workpiece velocity", Universidade de Ohio, 1996.
- [7] E. Paese and M. Geier and R. P. Homrich and J. L. Pacheco, "Simplified Mathematical Modeling for an Electromagnetic Forming System with Flat Spiral Coil as Actuator", Journal of the Brazilian Society of Mechanical Sciences and Engineering, 2011.



## **Modelagem Hidrodinâmica Bidimensional Aplicada a Inundações: Mapeamento e Estudo de Padrões no escoamento**

**Aldair Forster, Juliana Costi, William Correa Marques,**

Instituto de Matemática, Estatística e Física, IMEF. FURG

96203-000, Campus Carreiros, Rio Grande, RS

E-mail: aldairforster@gmail.com, juliana.costi@furg.br, wilianmarques@furg.br

**Antonio Raylton Rodrigues Bendô**

Escola de Engenharia, FURG

96203-000, Campus Carreiros, Rio Grande, RS

E-mail: rayltonbendo@gmail.com

***Resumo:** Dentre os desastres naturais, as inundações são as mais frequentes, e muitas vezes as que causam os maiores danos. Com o avanço das tecnologias, esses eventos podem ser estudados com detalhes. Uma das ferramentas de estudo é a modelagem numérica de inundações que vem sendo amplamente utilizada para antecipar riscos e auxiliar na tomada de decisão. O TELEMAT-2D, é um exemplar que é capaz de simular a hidrodinâmica de canais abertos pela resolução das equações das águas rasas num domínio discretizado por uma malha não estruturada de elementos finitos. Esse modelo foi aplicado em um estudo de caso nos rios que cruzam a cidade de Rio do Sul – SC, para simular a hidrodinâmica do local procurando mapear inundações que aconteceram na cidade nos anos de 2011 e 2013. Após a calibração e validação do modelo, 5 eventos com diferentes períodos de retorno puderam ser mapeados. Além do mapeamento, a metodologia permitiu encontrar outros padrões para os eventos, como duração de alagamento e identificar como esses padrões se repetem ou não em inundações com diferentes magnitudes. Os resultados podem ajudar a melhorar os sistemas locais de monitoramento e alerta e o planejamento de uso do solo.*

### **Introdução**

Com o passar dos anos as inundações apresentam-se mais frequentes embora as perdas materiais e mortes venham diminuindo [6]. Na microrregião do Alto Vale do Itajaí, localizado no estado de Santa Catarina, os problemas com inundações são recorrentes. A paisagem da região é formada por um relevo acidentado e encostas muito inclinadas, que formam pequenas várzeas em torno de ribeirões e rios, estando assim sujeita a frequentes inundações [2].

Em Rio do Sul, maior cidade da região, a população convive com as inundações desde os períodos iniciais de colonização, e nos últimos anos, principalmente após 2011, as inundações mostram-se mais frequentes. Em todo o período 1910-2018 a média de inundações por ano foi de 0.53, de 2011 até 2017 essa média é de 2 inundações por ano. Nesse último período duas das maiores inundações registradas aconteceram em setembro dos anos de 2011 e 2013.

Diante desse cenário de crescente número de inundações somado a não efetividade das medidas estruturais para contenção de cheias, medidas não estruturais podem ser adotadas para que os danos possam ser minimizados. Essas medidas, não são projetadas para dar uma proteção completa, mas para que a população possa conviver da melhor forma possível com a presença dos fenômenos. Elas compreendem a previsão e alerta de inundação, zoneamento das áreas de risco de inundação, seguro e proteção individual contra inundação [10].

Procurando atender à necessidade de mapear as áreas inundáveis na cidade, nesse trabalho o modelo hidrodinâmico bidimensional TELEMAT-2D é utilizado para simular a hidrodinâmica dos rios em uma parte da cidade de Rio do Sul, reproduzindo os eventos que aconteceram na cidade e mapeando assim as áreas de inundação nesses eventos. Essas informações podem ser utilizadas para atender algumas das exigências do Plano Diretor do Município como planejamento de uso de solo, prever rotas alternativas para o escoamento do trânsito [4]. Além disso, é previsto nessa lei que a ocupação dos solos deve ser feita de forma a "facilitar a convivência com as cheias dos rios através da expansão da cidade para regiões livres de enchentes ou pela verticalização controlada de algumas áreas já consolidadas através de

limitações à ocupação nas áreas mais baixas;". Portanto, essa pode ser uma ferramenta a ser utilizada como base para o desenvolvimento dessas atividades, tanto no local de estudo, como em qualquer região que disponha de dados suficientes para modelagem.

Nas simulações realizadas nesse trabalho, o modelo hidrodinâmico foi calibrado num período que compreendeu todo ano de 2013 e seus resultados foram validados com um mapa de extensão de inundação criado após essa inundação. Com os parâmetros calibrados, a hidrodinâmica da região no ano de 2011 foi simulada e os mapas de extensão de inundação para as inundações desse ano foram criados. Esses resultados, permitiram mapear outras quatro inundações ocorridas nesse ano com período de retorno de 1,24, 3,8, 4,2 e 21 anos.

Além do mapeamento das áreas inundáveis a modelagem hidrodinâmica bidimensional permite que outras informações a respeito do escoamento da área de estudo sejam obtidas. É apresentado nesse trabalho o número de dias que cada região permanece alagada, e, pela visualização dos padrões encontrados é possível apontar a suscetibilidade a inundação por região.

Com a metodologia proposta, as informações obtidas pelo modelo, associadas com os períodos de retorno de cada um dos cinco eventos permitiu a observação de algumas características do fluxo que são ou não recorrentes na área de estudo. Foi possível indicar e fazer algumas inferências que podem ajudar as autoridades na tomada de decisão, no gerenciamento de uso de solo indicando prováveis magnitudes dos eventos associadas com seu tempo de recorrência.

## Metodologia

A definição dos períodos de retorno (PR) mencionados nesse trabalho é dada por [9] que o estabelece como o intervalo médio de anos com que um evento hidrológico pode ser igualado ou superado pelo menos uma vez. É obtido através do inverso da probabilidade de ocorrência.

Para a criação das áreas de alagamento foi utilizado um modelo hidrodinâmico TELEMAC-2D [8] que necessita de dados fluviométricos e dados topográficos. O modelo foi calibrado através de simulações realizadas para todos o período do ano de 2013 e validado utilizando um mapa de extensão de inundação criado pela defesa civil da cidade após a inundação ocorrida em setembro. Depois de calibrados os parâmetros do modelo, o período de 2011 foi simulado e as extensões de inundação para as quatro inundações que aconteceram nos meses de julho (1), agosto (2) e setembro (1).

Os dados fluviométricos foram adquiridos na página web da ANA<sup>1</sup>. Eles são formados por dados de níveis do rio Itajaí-Açú, na estação localizada na cidade de Rio do Sul (estação número 83300200), e vazões nos rios Itajaí do Sul, Itajaí do Oeste e Trombudo, respectivamente nas cidades de Ituporanga (estação número 83250000), Taió (estação número 83050000) e Trombudo Central (estação número 83069900).

Os dados topográficos são formados pelo Modelo Digital do Terreno (MDT) além da batimetria dos canais, uma vez que o MDT não dá uma boa representação topográfica do leito dos rios. O MDT utilizado está disponível no Sistema de Informação Geográficas de Santa Catarina<sup>1</sup>. É um modelo de alta resolução espacial (1 m), gerado no levantamento aerofotogramétrico do estado de Santa Catarina, em 2010. Os dados batimétricos foram cedidos pela Defesa Civil do Estado de Santa Catarina onde foram levantadas 116 seções transversais de batimetria ao longo do Rio Itajaí do Sul, Rio Itajaí do Oeste e Rio Itajaí-açu, principalmente nas áreas próximas ao centro da cidade de Rio do Sul [1]. Os dados de topografia foram interpolados em malha não estruturada de elementos finitos através do *software Blue Kenue*<sup>2</sup>.

## Validação do modelo e avaliação dos resultados

A validação das simulações foi realizada de duas formas. Através da capacidade do modelo de simular as vazões na cidade de Rio do Sul e pela eficiência na simulação da extensão da inundação no evento ocorrido em 2013.

<sup>1</sup> Informação disponível em: <http://sigsc.sds.sc.gov.br/>. Acesso em 18 de maio de 2018.

<sup>2</sup> Informação disponível em: <https://nrc.canada.ca/en/research-development/products-services/software-applications/blue-kenuetm-software-tool-hydraulic-modellers> Acesso em 18 de maio de 2018.

Os dados para validação das vazões foram obtidos na estação da ANA na cidade de Rio do Sul (número 83300200). O polígono de extensão de inundação foi cedido pela defesa civil do município.

Para avaliar a extensão da inundação no tempo em que o mapa de inundação foi produzido, serão utilizadas as medidas de ajuste pontual  $F_1$  (eq. 1) e  $F_2$  (eq. 2), baseadas na Tabela 1, onde  $F_1$  varia de 0 a 1 enquanto que  $F_2$  varia de -1 a 1 e o termo  $-B$  no numerador desta última é usado para penalizar superestimações [3]. Nessas medidas, quanto mais próximo da unidade estiver o valor resultante, melhor foi o resultado encontrado pelo modelo.

Para realizar uma validação temporal das simulações, foram avaliadas as vazões calculadas pelo modelo com as vazões reais pelas medidas de eficiência: RMSE (Root Mean Square Error), e (eficiência de Nash e Sutcliffe (e) [7]), frequentemente utilizada em análises de vazões, além das medidas MAPE (Mean Absolute Percentage Error) e correlação de Pearson (r), que mostra se as vazões apresentam tendências de crescimento e decréscimo linearmente relacionados [3].

$$F_1 = \frac{A}{A+B+C} \quad (1)$$

$$F_2 = \frac{A-B}{A+B+C} \quad (2)$$

Tabela 1: Tabela de Contingência

	Observado Inundado	Observado Seco
Simulado Inundado	A	B
Simulado Seco	C	D

## Resultados

### Períodos de Retorno

Os períodos de retorno calculados nesse trabalho levaram em consideração os eventos que estão registrados na página da defesa civil do município de Rio do Sul<sup>2</sup>. Uma vez que a construção das barragens nos canais pode afetar o fluxo e a frequência de ocorrência de inundações, apenas os eventos pós operacionalização da barragem foram considerados nesses cálculos, ou seja, a partir do ano de 1975 até 2017.

Tabela 2: Período de retorno e magnitude de cada um dos eventos estudados.

	Cota (m)	PR (anos)
20/11/2013	10,39	7
03/07/2011	6,5	1,24
10/08/2011	8,83	4,2
31/08/2011	8,76	3,82
09/09/2011	12,96	21

Nesse período de 42 anos aconteceram 34 inundações. As inundações aqui estudadas, estão indicadas na Tabela 2, indicando a data da ocorrência do pico da cheia, a cota atingida e o período de retorno.

### Mapeamento das áreas de inundação em 2013

Após realizada a calibração a inundação ocorrida em setembro de 2013 foi mapeada. A Figura 1 mostra a área de inundação em contraste com o mapa criado pela defesa civil. Essa inundação teve cota máxima de 10.39 m relativo a uma vazão de 1364 m<sup>3</sup>/s no momento de pico. O modelo simulou como vazão máxima, no momento de pico 1068,5 m<sup>3</sup>/s. As medidas de eficiência para as vazões se encontram na Tabela 3, e apresentam bons resultados levando em consideração os resultados apresentados em [3] e [7].

Tabela 3: Medidas de eficiência para as vazões simuladas em 2011.

RMSE	MAPE	e	r
111,0095	0,8707	0,6309	0,9263

<sup>2</sup> Informação disponível em: <https://defesacivil.riodosul.sc.gov.br>. Acesso em 29 de junho de 2018.

Na Figura 1 nota-se que na maior parte da extensão da área, a inundação foi muito bem simulada ( $F_1 = 0,8043$  e  $F_2 = 0,6319$ ), se comparada a outros trabalhos como [5]. Os piores resultados se encontram na grande área de superestimação próxima ao ponto de coordenadas  $-49,64^\circ$ ,  $-27,21^\circ$ . Essa é uma área completamente urbanizada, portanto os sistemas de escoamento artificial, que não foram impostos no modelo, podem ter auxiliado no escoamento. Já as maiores subestimações aconteceram nas regiões a nordeste e sudoeste, são regiões próximas aos locais de imposição de condições de contorno, que podem ter "forçado" esses resultados.

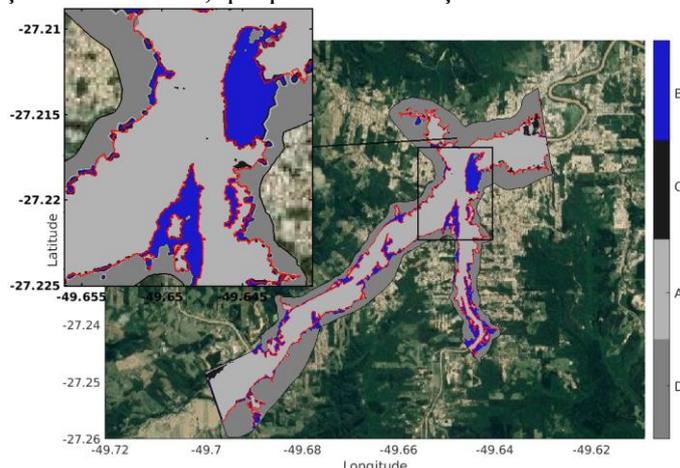


Figura 1: Resultados da simulação versus polígono de verificação. O polígono com contorno preto delimita a área de simulação. Os polígonos vermelhos representam a extensão da inundação real, com dados recolhidos em solo. Legenda: A: representa a área simulada inundada e observada inundada. B: representa subestimações. C: reporta as superestimações. D: representa a área simulada e observada seca.

De acordo com as informações expressas no site da defesa civil, as inundações registradas são aquelas que têm cotas maiores do que 6,5 m. Com essa informação, foram selecionados no período pré e pós pico de inundação os dias em que as cotas permaneceram maiores do que esse limiar. Com isso foi encontrada uma duração do alagamento pontualmente. No evento de setembro de 2013 a primeira cota de 6,5 m foi atingida no dia 20 e o rio voltou ao nível menor que esse limiar no dia 29. A Figura 2 (a) apresenta o número de dias de alagamento por região nesse período de tempo, e Figura 2 (b) apresenta o número de dias que cada região ficou alagada até 10 dias após a inundação, ou seja, até o dia 9 de outubro.

Ao analisar essas figuras, como esperado, as áreas adjacentes ao rio apresentam o maior número de dias de alagamento. Contudo o alagamento persiste em outras áreas. Percebe-se, por exemplo, que a inundação é mais persistente no rio Itajaí do Oeste (longitudes menores que  $49,65^\circ$ ). Os outros trechos com menor número de dias podem ser relacionados ao melhor escoamento da água.

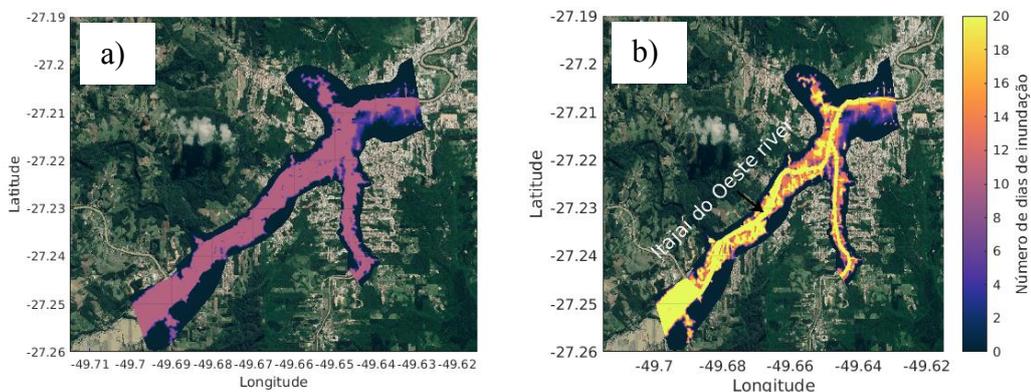


Figura 2: Dias de alagamento por região em no período de inundação em 2013. (a) Número de dias de alagamento entre 20 e 29 de setembro. (b) Número de dias que cada região ficou alagada entre 20 de setembro e 9 de outubro.

### Mapeamento das áreas de inundação em 2011

Uma vez que o modelo foi calibrado a hidrodinâmica da região foi simulada para todo o ano de 2011. Nesse ano aconteceram quatro episódios de inundação. O primeiro em 3 de julho, quando o rio atingiu a cota de 6,5 metros, o segundo e terceiro eventos aconteceram no mês de agosto, quando o rio atingiu as cotas de 8,83 m e 8,76 m, respectivamente nos dias 10 e 31. E por último a maior inundação que ocorreu no mês de setembro quando foi atingida a cota de 12,96 m, no dia 9, que é a segunda maior cota em todo o período de registros na estação.

A Figura 3 traz as vazões simuladas e observadas na estação fluviométrica da cidade de Rio do Sul. Os círculos coloridos apresentam o momento da vazão máxima em cada um dos eventos estudados nesse trabalho, listados na Tabela 2, em 2011.

A vazão máxima observada em todo o período foi de 1903.5 m<sup>3</sup>/s e a vazão máxima dada pelo modelo foi de 1902,4 m<sup>3</sup>/s, sendo que o modelo simulou esse máximo com uma hora de antecedência. As medidas de eficiência, encontram-se na Tabela 4.

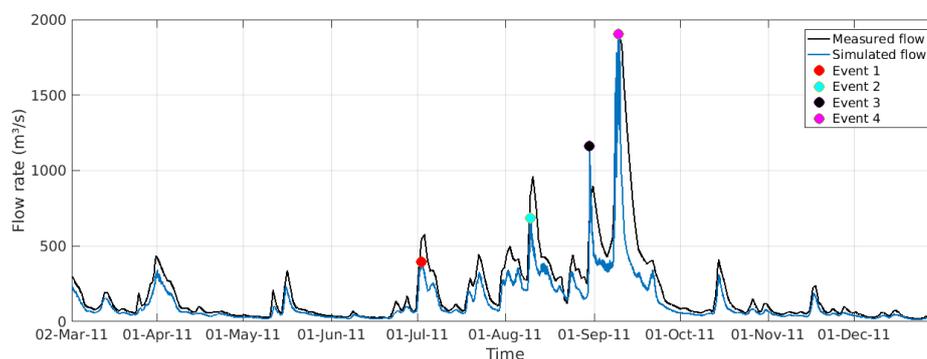


Figura 3: Vazões calculadas e observadas para o ano de 2011. As marcas coloridas representam os eventos de inundação registrados, e indicam os períodos de análise.

Ao comparar esses resultados com os obtidos para o período de 2013, percebe-se que nesses o RMSE apresentou resultados piores, enquanto que o MAPE, mostrou-se significativamente menor, reduzido em 49%. A eficiência de Nash e Sutcliffe (e) também mostrou melhores resultados, assim como o coeficiente r. Portanto, espera-se que as medidas de extensão de inundação sejam tão bem simuladas, ou melhor do que no evento de 2013.

Tabela 4: Medidas de eficiência para as vazões simuladas em 2011.

RMSE	MAPE	e	r
127,364	0,4453	0,7362	0,9282

As áreas de inundação para cada um dos eventos, marcados na Figura 3 foram demarcadas seguindo os mesmos procedimentos da seção anterior. A Figura 4 apresenta as extensões encontradas pelo modelo em cada período de pico nos quatro eventos. Nota-se que a esquerda da longitude -49.65°, no rio Itajaí do Oeste, a área está inundada de uma forma muito parecida em todos os eventos, independente da cota atingida, indicando propensão a inundações, mesmo quando as cotas atingidas não são muito altas.

No ano de 2011, dado que as inundações aconteceram de forma sequencial, com 68 dias entre os picos da primeira e da última, muitas áreas ficaram inundadas por muitos dias. Para se ter uma noção disso, a Figura 5(a) mostra o número de dias que cada área ficou alagada nesse período (de 03 de julho a 9 de setembro) e nesses acrescidos dos 10 dias seguintes (Figura 5 (b)).

### Contraste entre os eventos de 2011 e 2013 e análise padrões observados

Ao comparar a área total de inundação calculada pelo modelo no desastre de 2013 (6.66 km<sup>2</sup>) e no quarto evento de 2011 (7.42 km<sup>2</sup>), pode-se ter a impressão de que essas não diferem muito (0.76 km<sup>2</sup>) e, portanto, podem ter impactos parecidos. Contudo, as vazões máximas encontradas nos dois diferem em 539.5 m<sup>3</sup>/s, ou seja, as vazões do evento de 2011 apresentam velocidade consideravelmente maiores que podem causar mais danos.

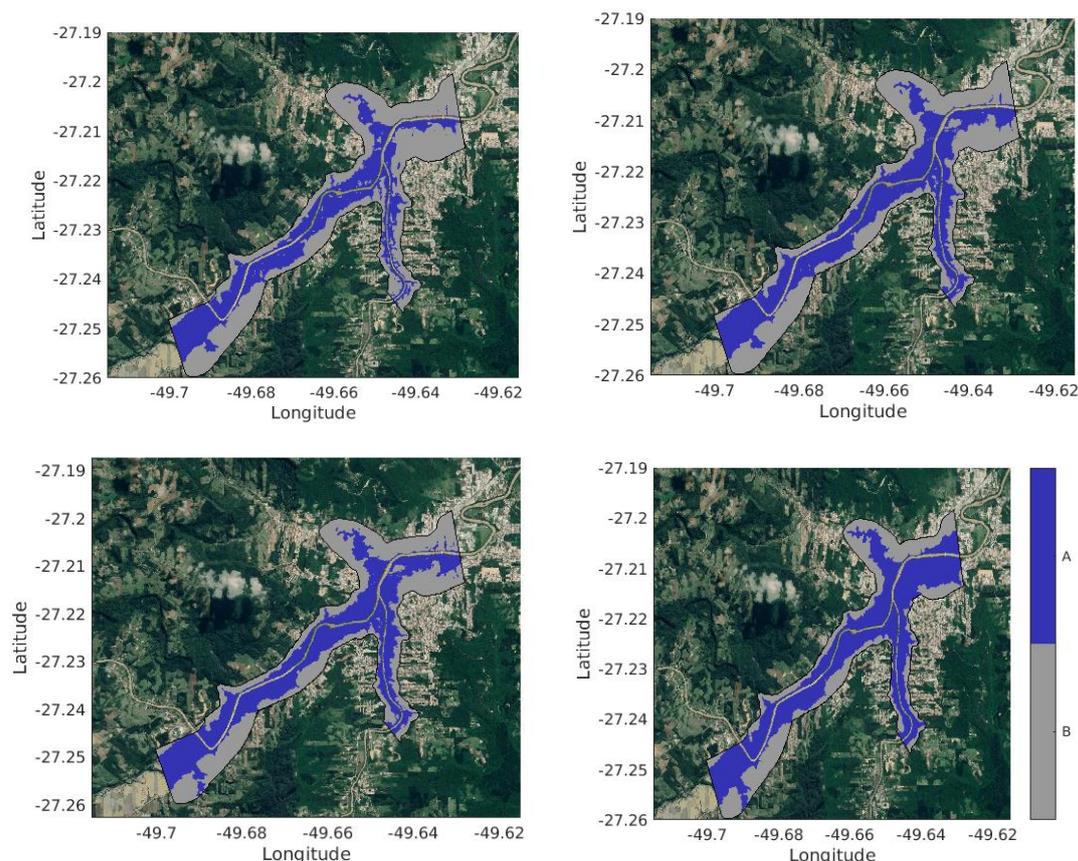


Figura 4: Resultados das extensões de inundações para os eventos ocorridos no ano de 2011. Legenda: A (azul) área simulada inundada; B (cinza) área simulada seca.

Os padrões de dias de alagamento encontrados no ano de 2013 (Fig. 2), foram similares às durações de inundação encontradas para os eventos de 2011 (Fig. 5). As áreas da região mais a nordeste apresentam durações menores mesmo que no evento de 2011 essa região tenha apresentado uma larga extensão de alagamento como observado na Figura 4.

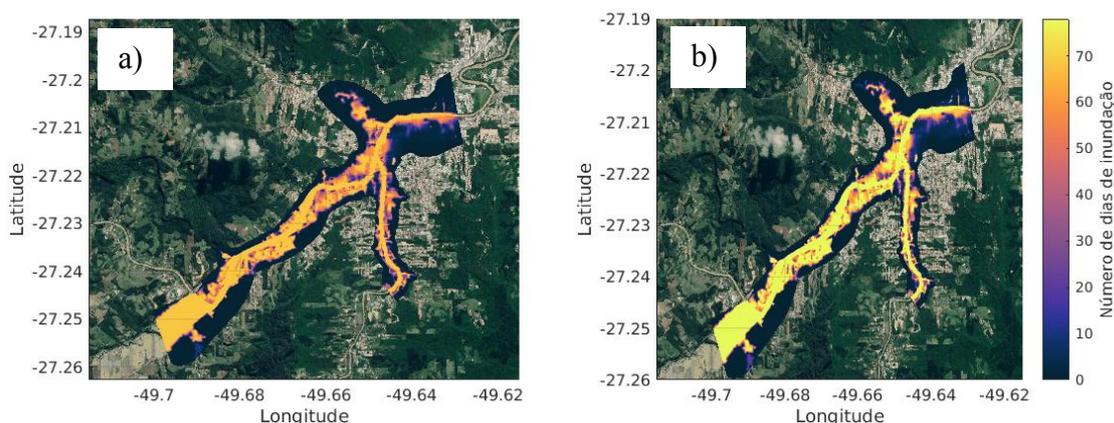


Figura 5: Dias de alagamento por região no período de inundações em 2011. (a) de 03 de julho a 9 de setembro. (b) de 3 junho a 19 de setembro.

As áreas a sudeste em torno do rio Itajaí do Sul, apresentam em todos os eventos locais com áreas de alagamento duradouras, entretanto essas áreas são menores e menos persistentes do que as encontradas nas áreas a sudoeste no rio Itajaí do Oeste, onde grandes regiões permaneceram alagadas durante todo o período em que o rio atingiu a cota de 6,5 m. E mesmo

com passados 10 dias do pico de inundação, no rio Itajaí do Oeste muitas áreas permaneceram alagadas (Figura 2(b) e Figura 5 (b)).

### Considerações finais

Pela calibração do modelo hidrodinâmico TELEMAC-2D foi possível reproduzir com acurácia a extensão de inundação e as vazões observadas na cidade de Rio do Sul no ano de 2013.

Através dessa calibragem, com parâmetros ótimos de fricção definidos e metodologia de validação dos resultados, os mesmos procedimentos foram utilizados para simular a hidrodinâmica da região no ano de 2011, quando aconteceram 4 inundações. Com isso esses eventos, com diferentes magnitudes e períodos de retorno distintos, puderam ser mapeados produzindo informações que podem ser utilizadas pelos sistemas de monitoramento e alerta, e pelas autoridades no planejamento de uso e ocupação do solo. Pela visualização dos padrões de duração das inundações, as seguintes considerações puderam ser feitas no caso de estudo apresentado.

A porção da área de estudo que margeia o rio Itajaí do Oeste apresenta grandes áreas de inundações mesmo com baixas cotas observadas na estação fluviométrica e; grandes aumentos nessas cotas não geram aumentos expressivos nas áreas de inundação. Na porção do rio Itajaí do Oeste existem várias regiões com grande persistência do alagamento (Fig. 2 e Fig. 5); Portanto espera-se que ao passo que as cotas aumentem essas áreas apresentem maior crescimento em extensão do que as áreas adjacentes ao rio Itajaí do Oeste.

Com a metodologia apresentada, além da informação de mapeamento de inundações com diferentes períodos de retorno, foi possível analisar a duração da inundação na área de estudo. Esses resultados dão um largo panorama das inundações no local de estudo. Além da duração das inundações, outros padrões podem ser encontrados com esse método, como profundidade da lâmina d'água e velocidade de fluxo em cada um dos pontos da malha, que não foram citados nesse trabalho devido ao espaço disponível. Portanto, esses procedimentos podem ser utilizados como auxiliares no gerenciamento de riscos, mapeamento de inundações e no planejamento e uso de solos.

### Referências

- [1] Ambiental, C.A.. “Relatório Técnico Municípios de Rio do Sul e Lontras / SC.” Consórcio Bacia do Rio Itajaí. 2015
- [2] Aumond, J.J.; Sevegnani, L.; Tachini, M.; Bacca, L.E. “Condições naturais que tornam o vale do Itajaí sujeito aos desastres.” In: Frank, b.; Sevegnani, l. (Org.). Desastre de 2008 no Vale do Itajaí: água, gente e política. Blumenau: Agência de Água do Vale do Itajaí, p. 22-37, 2009.
- [3] Di Baldassarre, G. “Floods in a changing climate: inundation modelling.” Vol. 3, Cambridge University Press. 2012.
- [4] Hobus, M. “Lei Complementar nº 163, de 12 de dezembro de 2006”. 2006
- [5] Horritt, M.; Bates, P. “Evaluation of 1D and 2D numerical models for predicting river flood inundation”. *Journal of hydrology*, 268, 87–99. 2002.
- [6] Jha, A.K.; Bloch, R.; Lamond, . “Cidades e Inundações: um guia para a gestão integrada do risco de inundação urbana para o século XXI.” Washington: Banco Internacional para Reconstrução e Desenvolvimento. 2002.
- [7] Nash, J.E.; Sutcliffe, J.V. “River flow forecasting through conceptual models part I—A discussion of principles.” *Journal of hydrology*, 10, 282–290. 2012.
- [8] Riadh, A.; Cedric, G.; Jean, M. “Telemac modeling system: 2d hydrodynamics telemac-2d software release 7.0 user manual.” Paris: R&D, Electricite de France, p. 134. 2014.
- [9] Santos, R.D. “Vulnerabilidade ambiental. Brasília:” Ministério do Meio Ambiente. 2007.
- [10] Tucci, C.E. “Gestão de águas pluviais urbanas; Programa de Modernização do Setor Saneamento.” Secretaria Nacional de Saneamento Ambiental, Ministério das Cidades. 2005



## Método de Galerkin descontínuo aplicado à problemas de escoamento em meios porosos com fraturas

Igor Mozolevski, Luciane I. A. Schuh,  
UFSC - Departamento de Matemática  
Campus Trindade  
88040-900, Florianópolis, SC  
E-mail: igor.mozolevski@ufsc.br. luciane.schuh@ufsc.br,

Marcio A. Murad  
LNCC - Laboratório Nacional de Computação Científica  
25651-075, Petrópolis, RJ  
E-mail: murad@lncc.br.

**Resumo:** *A simulação numérica de problemas de escoamento em meios porosos com fraturas é um problema relevante, uma vez que as heterogeneidades do meio poroso e a presença de fraturas tem forte impacto no processo de escoamento e tornam o problema multiescala. A modelagem de um escoamento em um meio poroso com fraturas está baseada na Lei de Darcy com a presença de fortes discontinuidades na permeabilidade, em regiões com espessura muito pequena em relação à escala do problema. Neste artigo estudamos a aplicação do método de Galerkin descontínuo para resolver problemas de escoamento monofásico em meios porosos com fraturas. Apresentamos resultados numéricos para problemas com fraturas que representam caminhos preferenciais e fraturas que representam uma barreira, e estudamos a influência da variação da espessura da fratura no processo de escoamento. Também apresentamos resultados numéricos para problemas com fraturas que se interceptam. Em todos os problemas abordados o método de Galerkin descontínuo apresentou resultados estáveis e precisos, capturando com eficiência os diferentes fenômenos presentes no processo de escoamento.*

**Palavras-chave:** *escoamento em meios porosos, fraturas, Galerkin descontínuo*

### 1 Introdução

A presença de fraturas em um meio poroso complica significativamente a modelagem numérica de problemas de escoamento. As fraturas ocorrem em diferentes escalas, com diversos tipos de geometria, e podem representar caminhos preferenciais para o escoamento do fluido ou podem representar barreiras. Desta forma as fraturas tem uma grande influência no processo de escoamento, tornando o fluxo muito mais rápido em certas direções ou, possivelmente, bloqueando o fluxo em outras direções. Tais tipos de problemas surgem nas mais diversas áreas, como engenharia de reservatório de petróleo, na gestão de reservatórios hídricos subterrâneos, na captura e armazenamento de carbono e resíduos radioativos, entre outras.

Na literatura podemos encontrar duas formas de modelar problemas de escoamento em meios porosos com fraturas [5, 3, 4, 1]. Uma das formas é considerar o modelo completo na modelagem, ou seja, as fraturas são consideradas como um subdomínio de espessura fina, em que a permeabilidade é muito alta (ou muito baixa) em relação ao restante do domínio. Outra forma, conhecida como formulação reduzida, consiste em considerar a fratura como uma interface entre subdomínios. Neste caso torna-se necessária a imposição de condições de transmissão na interface, que dependem de um parâmetro  $1/2 \leq \xi \leq 1$ , e permitem acoplar o problema ( $d - 1$ )-

dimensional com o problema  $d$ -dimensional. Porém na literatura não há um consenso de qual é o valor adequado a ser usado para o parâmetro  $\xi$ . Neste trabalho empregamos o método de Galerkin descontínuo (DG) proposto em [2] para resolver o modelo completo envolvendo diferentes problemas de escoamento em meios porosos com fraturas. Os resultados numéricos obtidos serão empregados em trabalhos futuros para validar resultados numéricos obtidos com a formulação reduzida, e assim permitir um estudo sobre qual o melhor valor a ser considerado para o parâmetro  $\xi$ . Nas seções a seguir apresentamos a descrição do problema, o método numérico empregado e resultados numéricos.

## 2 Formulação do Problema

Considere um domínio  $\Omega \subset \mathbb{R}^2$  com fronteira  $\Gamma = \partial\Omega$ . Dividimos o domínio  $\Omega$  em subdomínios  $\Omega_1, \Omega_2, \Omega_f$  (ver Figura 1), tal que  $\Omega \setminus \bar{\Omega}_f = \Omega_1 \cup \Omega_2$ ,  $\Omega_1 \cap \Omega_2 = \emptyset$ , onde  $\Omega_f$  representa uma fratura do meio poroso  $\Omega$ , com diâmetro  $d$ . Usaremos que  $\Gamma_i = \partial\Omega_i \cap \Gamma$ ,  $i = 1, 2, f$  e  $\gamma_i = \partial\Omega_i \cap \partial\Omega_f \cap \Omega$ ,  $i = 1, 2$ .

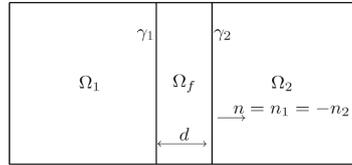


Figura 1: Domínio  $\Omega$  dividido em subdomínios  $\Omega_1, \Omega_2, \Omega_f$ .

Considere o escoamento monofásico e incompressível regido pela lei de Darcy em  $\Omega$

$$\begin{cases} \mathbf{div} u_i = q_i, & \text{em } \Omega_i, i = 1, 2, f \\ u_i = -K_i \nabla P_i, & \text{em } \Omega_i, i = 1, 2, f \\ P_i = P_{D_i}, & \text{sobre } \Gamma_i, i = 1, 2, f \\ P_i = P_f, & \text{sobre } \gamma_i, i = 1, 2 \\ u_i \cdot n = u_f \cdot n, & \text{sobre } \gamma_i, i = 1, 2 \end{cases} \quad (1)$$

onde  $K$  é a permeabilidade,  $q$  é o termo de fonte,  $u$  a velocidade de Darcy,  $p$  a pressão,  $n$  vetor normal unitário. Denotamos por  $p_i, u_i, K_i$  e  $q_i$  as restrições de  $p, u, K$  e  $q$ , respectivamente sobre  $\Omega_i, i = 1, 2, f$  e  $P_{D_i}$  a restrição de  $P_D$  sobre  $\Gamma_i, i = 1, 2, f$ .

## 3 Método de Galerkin Descontínuo

Empregamos o método de Galerkin descontínuo para modelar numericamente o problema de escoamento monofásico (1). Para tal fim consideramos uma partição  $\mathcal{T}_h$  do domínio  $\Omega$  composta por  $M$  elementos abertos, disjuntos,  $\mathcal{T}_h = \{T_i\}_{1 \leq i \leq M}$ , tal que  $\bar{\Omega} = \bigcup_{T \in \mathcal{T}_h} \bar{T}$ . Para cada elemento  $T$  da malha  $\mathcal{T}_h$ , associamos o diâmetro  $h_T$  e definimos o diâmetro da malha como sendo  $h := \max_{T \in \mathcal{T}_h} h_T$ . Denotamos por  $\mathcal{F}_h^i$  como sendo o conjunto formado pelas arestas interiores e  $\mathcal{F}_h^\partial$  o conjunto das arestas de fronteira,  $\mathcal{F}_h := \mathcal{F}_h^i \cup \mathcal{F}_h^\partial$  o conjunto de todas as arestas e para qualquer  $F \in \mathcal{F}_h$ ,  $h_F$  representa o diâmetro da aresta  $F$ .

Supomos que a malha  $\mathcal{T}_h$  é consistente com a divisão do domínio  $\Omega$  em subdomínios e com as condições de fronteira. Portanto, assumimos que as interfaces  $\gamma_i, i = 1, 2$  são exatamente cobertas por um conjunto de arestas armazenadas em  $\mathcal{F}_h^\gamma$ . Definimos  $\mathcal{F}_h^{i*} := \mathcal{F}_h^i \setminus \mathcal{F}_h^\gamma$ . Também consideramos os conjuntos  $\mathcal{F}_h^D$  e  $\mathcal{F}_h^N$  que correspondem respectivamente ao conjunto das arestas onde a condição de Dirichlet é satisfeita e onde a condição de Neumann deve ser imposta.

Para uma função  $v$  suficientemente suave, definimos o salto e a média em  $F \in \mathcal{F}_h^i$ , em que  $F = \partial T^- \cap \partial T^+$ , respectivamente como:

$$[v] := v|_{T^-} - v|_{T^+}, \quad \{v\} := \frac{1}{2}(v|_{T^-} + v|_{T^+}). \quad (2)$$

Extendemos essas definições para as faces da fronteira  $F \in \mathcal{F}_h^\partial$  como  $[v] = \{v\} := v|_T$ . Para a face  $F \in \mathcal{F}_h^i, F = \partial T^- \cap \partial T^+$  também definimos a normal  $\mathbf{n}_F$  como sendo exterior ao elemento  $T^-$  e para uma face de fronteira  $F \in \mathcal{F}_h^\partial$ , a normal  $\mathbf{n}_F$  coincide com a normal exterior  $\mathbf{n}_{\partial\Omega}$ .

Para lidar com as heterogeneidades do meio poroso empregamos as técnicas de média ponderada para uma função  $v$  e a média harmônica da permeabilidade  $K$ , respectivamente definidas por, para  $F \in \mathcal{F}_h^i$ ,

$$\{v\}_\omega := \omega_{T^-,F} v^- + \omega_{T^+,F} v^+, \text{ sendo } \omega_{T^-,F} := \frac{K_{T^+,F}}{K_{T^-,F} + K_{T^+,F}}, \omega_{T^+,F} := \frac{K_{T^-,F}}{K_{T^-,F} + K_{T^+,F}} \quad (3)$$

$$\langle K \rangle_F := \frac{2K_{T^-,F}K_{T^+,F}}{K_{T^-,F} + K_{T^+,F}}, \quad (4)$$

em que os pesos  $\omega_{T^-,F}$  e  $\omega_{T^+,F}$ , são definidos tais que  $\omega_{T^-,F} + \omega_{T^+,F} = 1$ . Para faces na fronteira,  $F \in \mathcal{F}_h^\partial$  extendemos as definições acima por  $\{v\}_\omega := v$  e  $\langle a \rangle_F := a$ . Definimos também o espaço de elementos finitos descontínuo para  $k \geq 1$ :

$$V_h^k := \{v_h \in L^2(\Omega); \forall T \in \mathcal{T}_h, v_h|_T \in \mathbb{P}_k(T)\}.$$

Resolvemos o problema (1) usando o método DG com penalização interior proposto em [2], ou seja, procuramos  $P_h \in V_h^k$  tal que  $\forall z_h \in V_h^k$ ,

$$\begin{aligned} & \sum_{T \in \mathcal{T}_h} \int_T K \nabla P_h^{m+1} \cdot \nabla z_h - \sum_{F \in \mathcal{F}_h^i \cup \mathcal{F}_h^D} \int_F (\mathbf{n}_F \cdot \{K \nabla P_h\}_\omega [z_h] + \vartheta \mathbf{n}_F \cdot \{K \nabla z_h\}_\omega [P_h]) \\ & + \sum_{F \in \mathcal{F}_h^i \cup \mathcal{F}_h^D} \int_F \gamma_F \frac{\sigma_F}{h_F} [P_h] [z_h] = \sum_{T \in \mathcal{T}_h} \int_T q z_h - \sum_{F \in \mathcal{F}_h^N} \int_F P_N z_h \\ & + \sum_{F \in \mathcal{F}_h^D} \int_F \left( -\vartheta \mathbf{n}_F \cdot K \nabla z_h + \gamma_F \frac{\sigma_F}{h_F} z_h \right) P_D \end{aligned} \quad (5)$$

em que  $P_D$  e  $P_N$  representam a condição de Dirichlet e de Neumann, respectivamente, e  $\vartheta \in \{-1, 0, 1\}$  permite migrar entre as formulações não-simétrica, incompleta ou simétrica do método DG. O parâmetro  $\sigma_F = \tilde{\sigma} k^2$ , sendo  $\tilde{\sigma} > 0$  suficientemente grande para garantir a estabilidade do método se  $\vartheta = 1$  e  $\gamma_F$  é dado pela média harmônica de acordo com a definição (4),  $\gamma_F = \langle K \rangle_F$ .

## 4 Resultados Numéricos

### 4.1 Exemplo 1

Neste exemplo, seguindo o artigo [3], testamos o método numérico para aproximar um problema com fratura permeável. Consideramos o domínio  $\Omega = (0, 2) \times (0, 1)$ , com fratura  $\Omega_f = (1-d, 1+d) \times (0, 1)$ . Os demais dados do problema foram considerados como apresentado na Figura 2.

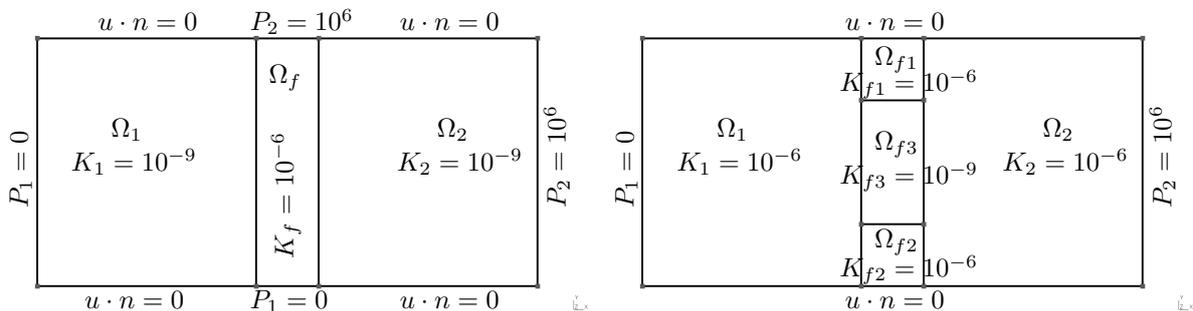


Figura 2: Domínio  $\Omega$  dividido em subdomínios  $\Omega_1, \Omega_2, \Omega_f$ , para o **Exemplo 1** à esquerda e para o **Exemplo 2** à direita sendo que neste  $\Omega_f$  é subdividido em  $\Omega_{f1}, \Omega_{f2}, \Omega_{f3}$ .

Na Figura 3 apresentamos aproximações para Pressão e Campo de velocidades, obtidas para diferentes diâmetros  $d$  da fratura  $\Omega_f$ . Na Figura 4 apresentamos o salto da velocidade na fratura  $([u \cdot n]_\gamma = (-K \nabla P)|_{\gamma_1} \cdot n - (-K \nabla P)|_{\gamma_2} \cdot n)$  e a velocidade tangencial na fratura,  $u_{f,T}$ . Podemos observar que no caso da fratura permeável o salto da pressão na interface é zero. Notamos também que no caso em que  $d = 0.01$  o perfil obtido com o método de Galerkin Descontínuo é semelhante ao apresentado no artigo [3].

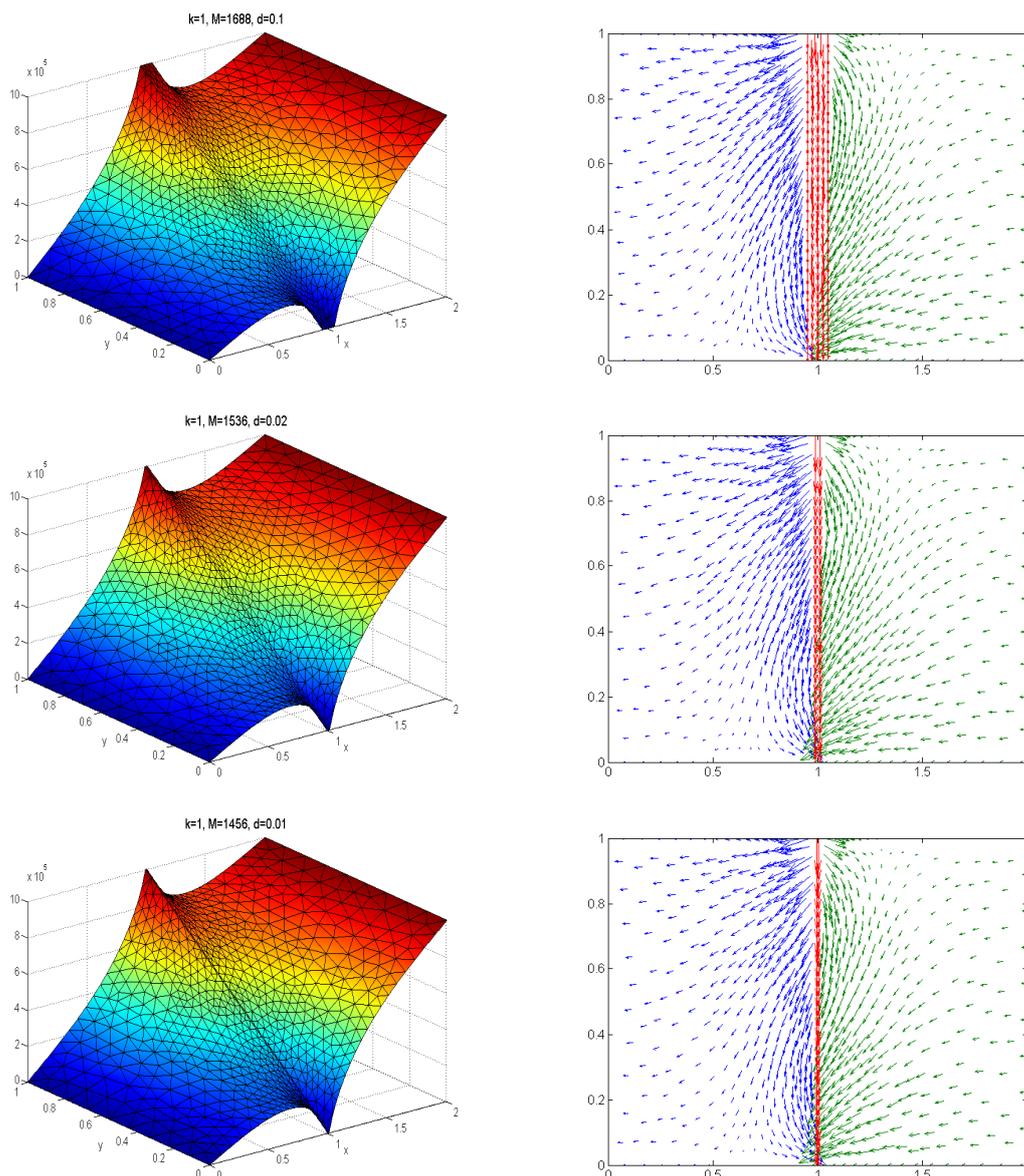


Figura 3: **Exemplo 1:** Aproximações numéricas para Pressão e velocidade em um domínio com fratura  $\Omega_f$  de diâmetro  $d = 0.1, 0.02, 0.01$ .

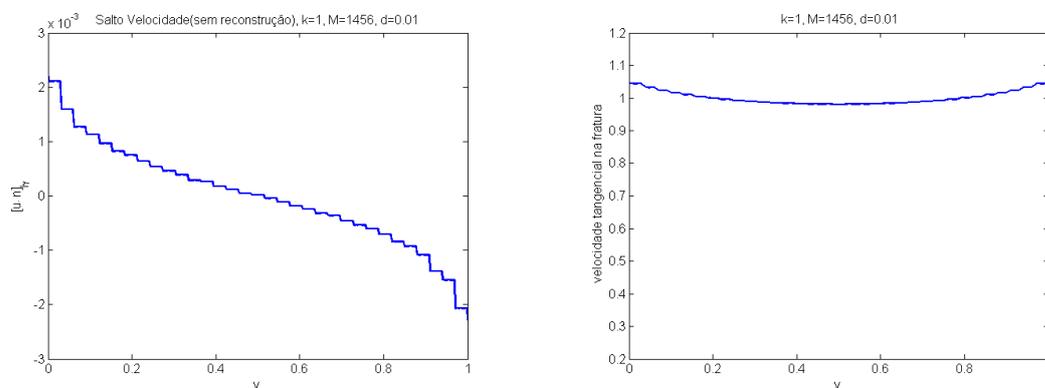


Figura 4: **Exemplo 1:** À esquerda Salto da velocidade normal na fratura. À direita: velocidade tangencial na fratura, ambos resultados para fratura  $\Omega_f$  com diâmetro  $d = 0.01$ .

## 4.2 Exemplo 2

Consideramos o domínio  $\Omega = (0, 2) \times (0, 1)$ , em que  $\Omega_f = (1 - d, 1 + d) \times (0, 1)$ . Os demais dados do problema foram considerados como apresentado na Figura 2. Neste exemplo (baseado

no artigo [4]) ilustramos a aplicação do método para problemas que possuem uma barreira geológica, ou seja, uma região com baixa permeabilidade na fratura. Na simulação podemos observar que o fluido contorna esta região de baixa permeabilidade, caracterizando assim uma barreira geológica. Os resultados numéricos são apresentados nas Figuras 5 e 6. Da Figura 6 observamos também, que no caso de uma fratura impermeável, a pressão apresenta um salto na fratura quando a espessura  $d$  tende a zero.

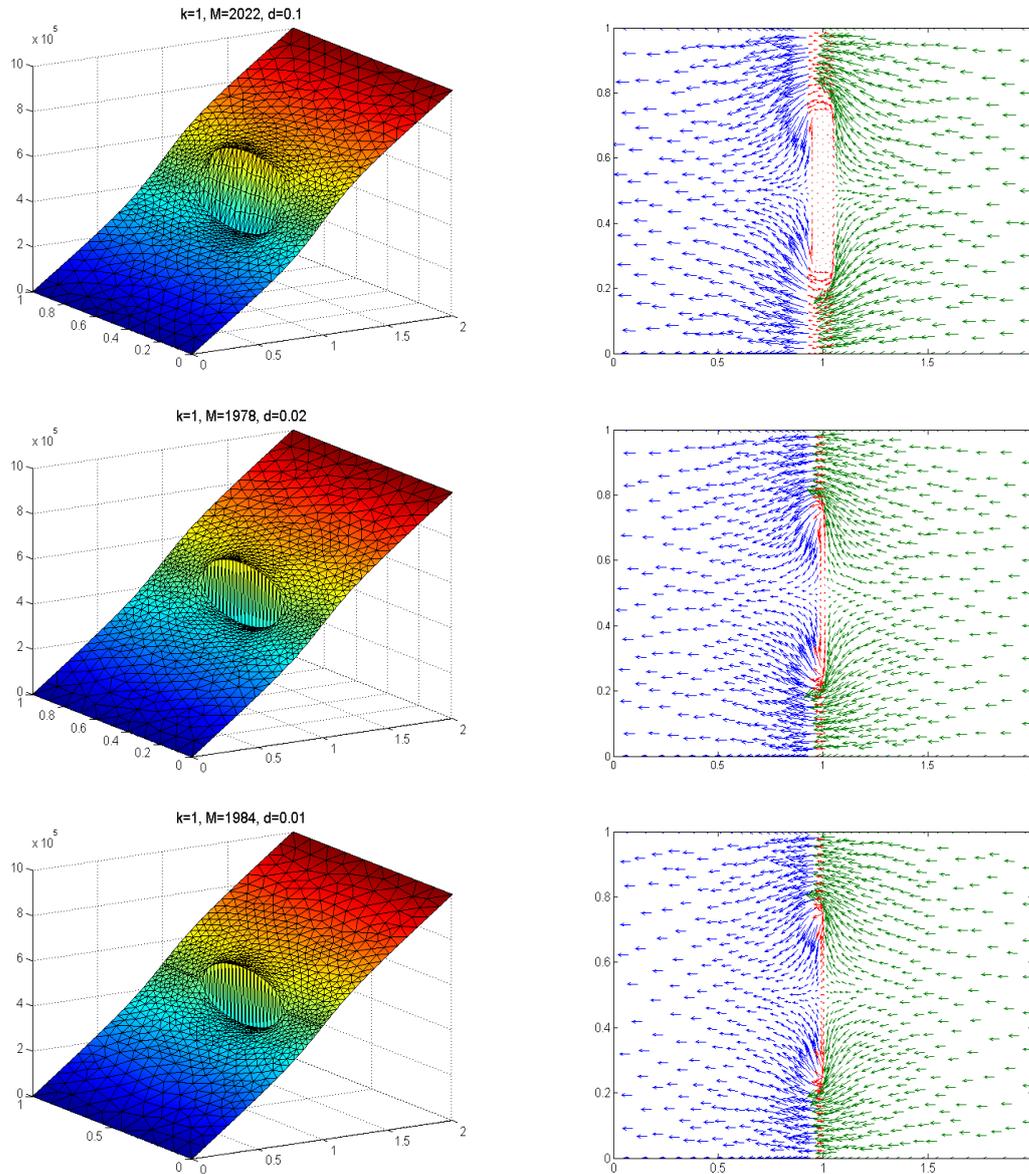


Figura 5: **Exemplo 2:** Aproximações numéricas para Pressão e velocidade em um domínio com fratura  $\Omega_f$  de diâmetro  $d = 0.1, 0.02, 0.01$ .

### 4.3 Exemplo 3

Consideramos o domínio  $\Omega = (0, 2) \times (0, 1)$ , com fraturas que se interceptam. Os demais dados do problema (baseado em [3]) foram considerados como apresentado na Figura 7 e os resultados numéricos são apresentados nas Figura 8 e 9. Como podemos notar, os resultados da simulação com método de Galerkin descontínuo são perfeitamente compatíveis com os resultados do artigo [3].

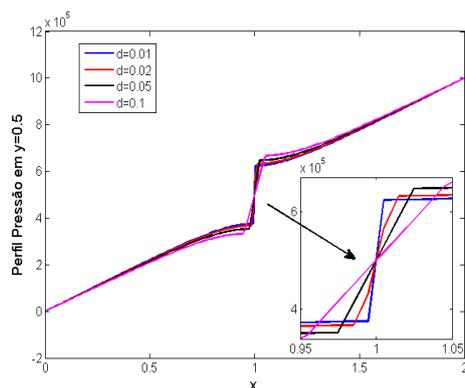


Figura 6: **Exemplo 2:** Perfil para pressão obtido ao longo de  $y = 0.5$  para fraturas com diferentes diâmetros  $d$ .

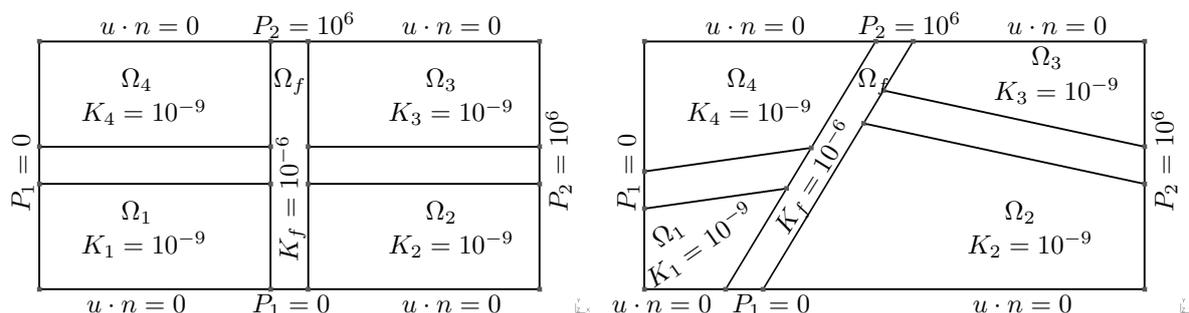


Figura 7: Domínio  $\Omega$  dividido em subdomínios  $\Omega_1, \Omega_2, \Omega_3, \Omega_4, \Omega_f$ , para o **Exemplo 3A** à esquerda e para o **Exemplo 3B** à direita.

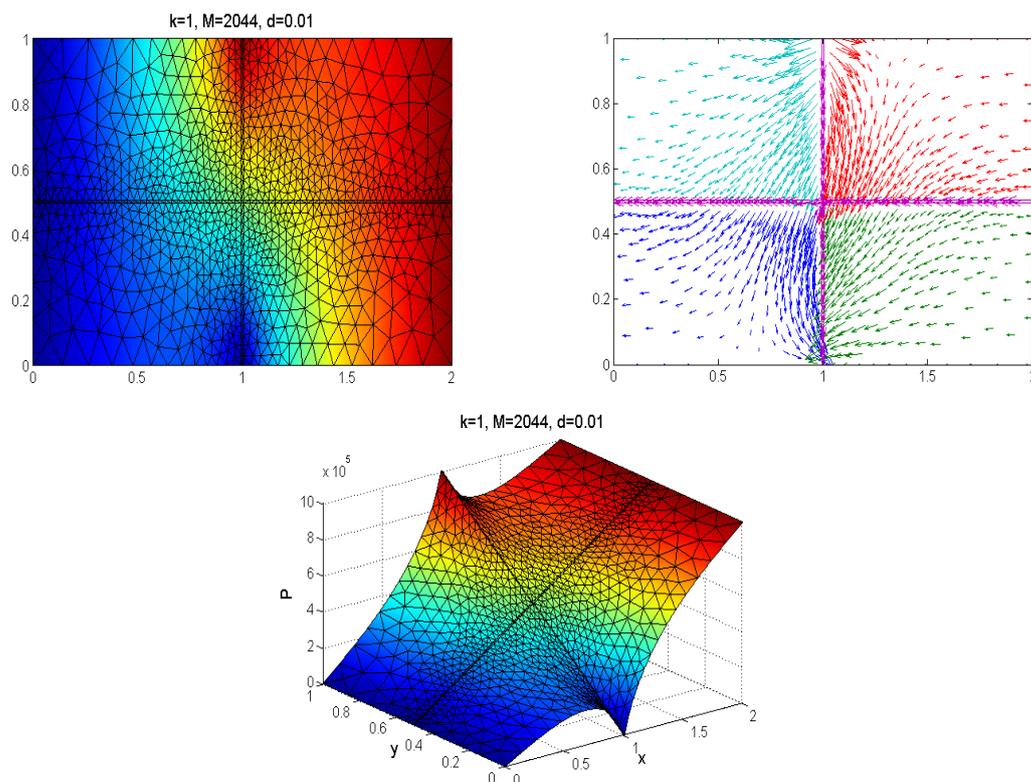


Figura 8: **Exemplo 3A:** Aproximações numéricas para Pressão e velocidade.

## 5 Conclusões

Os resultados numéricos apresentados comprovam que o método de Galerkin descontínuo é apropriado para resolver problemas de escoamento em meios porosos com fraturas. O método

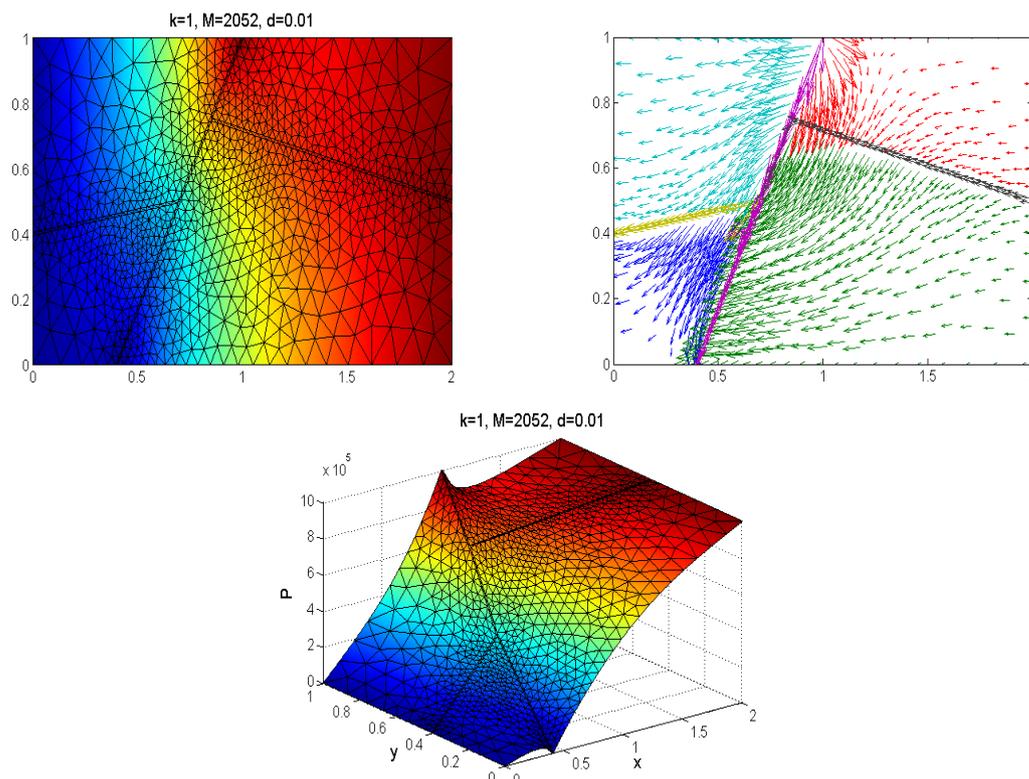


Figura 9: **Exemplo 3B:** Aproximações numéricas para Pressão e velocidade.

captou com precisão os mais diversos efeitos físicos envolvidos no processo de escoamento, como caminhos preferenciais para escoamento, descontinuidades na solução e bruscas variações no campo de velocidades. Porém a simulação em meios porosos com fraturas para problemas práticos em grande escala requer uma malha mais refinada na fratura, o que pode se tornar muito custoso. Para contornar esta dificuldade, torna-se necessário a utilização de modelos em que as fraturas são consideradas como interface. Esse modelo é conhecido como modelo reduzido. No modelo reduzido o fluxo dentro de fratura é modelado por uma equação de Darcy na variedade de codimensão um comparando com a dimensão do meio poroso. Simulações empregando o método de Galerkin descontínuo para esta formulação serão apresentadas em trabalhos futuros. E os resultados numéricos apresentados neste trabalho podem ser utilizados para validar tais simulações.

## Referências

- [1] P. Antonietti, C. Facciola, A. Russo, and M. Verani. Discontinuous Galerkin approximation of flows in fractured porous media on polytopic grids. *SIAM J. Sci. Comput.*, 41(1):A109–A138, 2019.
- [2] A. Ern, I. Mozolevski, and L. Schuh. Discontinuous Galerkin approximation of two-phase flows in heterogeneous porous media with discontinuous capillary pressures. *Computer Methods in Applied Mechanics and Engineering*, 199:1491–1501, 2010.
- [3] N. Frih, J. E. Roberts, and A. Saada. Modeling fractures as interfaces: a model for Forchheimer fractures. *Computational Geosciences*, 12:91–104, 2008.
- [4] V. Martin, J. Jaffre, and J. E. Roberts. Modeling fractures and barriers as interfaces for flow in porous media. *SIAM Journal on Scientific Computing*, 26(5):1667–1691, 2005.
- [5] M. Sahimi. *Flow and Transport in Porous Media and Fractured Rock: from Classical Methods to Modern Approaches*. John Wiley and Sons, 2011.



## Método iterativo baseado na iteração de Schultz para problemas discretos mal postos

Fermín S. V. Bazán, Everton Boos

UFSC - Departamento de Matemática

Campus Trindade

88040-900, Florianópolis, SC

E-mail: fermin@mtm.ufsc.br, everton.boos@gmail.com

**Resumo:** Propomos um método iterativo para a resolução de problemas discretos mal postos baseado em iterações matriciais geradas através do método de Schultz, visando também utilizar técnicas de projeção para contornar o custo operacional do método quando aplicado a problemas de grande porte. Essencialmente, tomando os iterados gerados pelo método de Schultz matricial por  $X_k$ , construímos os vetores  $x^{(k)} = X_k b$ , em que  $b$  é o vetor com dados do problema. Pela construção, como  $X_k$  converge à pseudo-inversa de  $A$ , estes iterados convergem para a solução de norma mínima do problema de mínimos quadrados  $\min_{x \in \mathbb{R}^n} \|Ax - b\|$ . Mostramos que a sequência  $\{x^{(k)}\}$  é quadraticamente convergente e ilustramos o seu comportamento numérico. No caso de dados de entrada com ruído, analisamos o comportamento de semi-convergência nos iterados e concluímos que as iterações devem ser truncadas para controlar a propagação do ruído. Como resultado, encontramos uma estimativa de erro para o caso em que o princípio da discrepância é utilizado para determinar o parâmetro de truncamento. Na tentativa de contornar o alto esforço computacional, estudamos estratégias de projeção em subespaços de Krylov e na base DCT (transformada discreta de cosseno), resultando em alternativas atrativas para aplicação em problemas de grande porte. Diversos resultados numéricos são apresentados para ilustrar a efetividade do método em problemas-teste bem conhecidos da literatura.

**Palavras-chave:** Problemas discretos mal postos, métodos iterativos, semi-convergência, critérios de parada.

### 1 Introdução

Diversas aplicações na ciência e na engenharia conduzem a sistemas de equações lineares de grande porte da forma  $Ax = \tilde{b}$  ou, mais geralmente, ao problema de mínimos quadrados associado

$$\min_{x \in \mathbb{R}^n} \|Ax - \tilde{b}\|, \quad (1)$$

para  $A \in \mathbb{R}^{m \times n}$  severamente mal condicionada e  $\tilde{b} = b + e$ , em que  $b \in \mathbb{R}^m$  denota os dados exatos e o vetor  $e$  corresponde ao ruído proveniente de imprecisões ou erros de medição. Problemas dessa forma são chamados de problemas discretos mal postos [11] e surgem naturalmente, por exemplo, na discretização de equações integrais de primeira espécie em processamento de sinais e restauração de imagens ou quando estamos interessados em conhecer a estrutura interna de um sistema através de medições externas, como acontece na tomografia computadorizada [9, 13, 21].

É fato conhecido que a solução para (1) gerada através da pseudo-inversa de Moore-Penrose  $A^\dagger$  [19], denotada por  $\tilde{x}_{LS} = A^\dagger \tilde{b}$ , é dominada pelo ruído nos dados de entrada. Neste cenário, métodos de regularização são essenciais para computar aproximações estáveis da solução livre de ruídos  $x_{LS} = A^\dagger b$  procurada. A literatura possui diversas abordagens com este intuito, como a regularização de Tikhonov [23], que consiste basicamente em trocar (1) por um problema relacionado bem condicionado cuja solução é aceita como aproximação para  $x_{LS}$ .

Outra maneira de lidar com o problema é através de métodos iterativos, tais como CGLS/LS-QR [4, 18], MINRES/MR-II [7, 17] e GMRES/RRGMRES [6, 20], entre muitos outros, para os quais o índice  $k$  da iteração atua como parâmetro de regularização. Isto ocorre porque, para muitos métodos iterativos, nos passos iniciais, os iterados tendem a convergir para  $x_{LS}$ , para então se deteriorarem com o crescimento de  $k$ , até convergirem finalmente para a solução de (1), dominada pelo ruído. Este fenômeno, conhecido como semi-convergência [16], sugere que a interrupção das iterações após um certo número de passos tem efeito regularizante. Para tanto, critérios de parada específicos são essenciais para determinar boas soluções.

Neste trabalho, propomos um método iterativo para computar aproximações estáveis para a solução livre de ruídos do problema (1),  $x_{LS} = A^\dagger b$ , baseado no fato de que se  $\{X_k\}$  é uma sequência matricial que converge para  $A^\dagger$ , então a sequência vetorial  $\{x^{(k)}\}$ , definida por

$$x^{(k)} = X_k b, \quad (2)$$

converge para a solução de mínimos quadrados de norma mínima  $x_{LS} = A^\dagger b$ . Entre uma vasta gama de possibilidades presentes na literatura, a geração da sequência matricial  $\{X_k\}$  se dá através da recorrência

$$X_{k+1} = X_k(2I - AX_k), \quad (3)$$

que consiste em um caso particular de um método de alta ordem [3] e é referenciado na literatura como método de Schultz [22]. Tais iterações possuem também um paralelo com as iterações geradas pelo método de Newton quando aplicado a um certo problema não linear, de modo que a recorrência matricial (3) é aqui denominada por iteração de Schultz ou Newton, de forma equivalente. A convergência quadrática característica do método de Newton é preservada nas iterações matriciais (3), desde que o iterado inicial seja escolhido apropriadamente.

Nas seções seguintes, abordaremos aspectos teóricos e práticos da sequência gerada por (2) através do método de Newton (3), tais como convergência e propriedades de monotonia. A seguir, apresentamos uma estimativa teórica para o caso em que o princípio da discrepância de Morozov [12, 14] (DP) é utilizado para determinar o parâmetro de truncamento  $k$ . Visando uma aplicação prática do método desenvolvido, surgem dificuldades com relação ao custo operacional e alguma estratégia de aceleração se faz necessária. Para isto, propomos uma técnica de projeção em subespaços de dimensão finita, essencialmente buscando um espaço de dimensão baixa que possa conter boas aproximações à solução procurada. Bases geradas para espaços de Krylov e a base DCT são alguns exemplos [9, 11]. De forma grosseira, uma vez determinado o espaço de projeção, aplicamos o método (2) ao problema projetado e calculamos uma solução aproximada para (1). Esta ideia resultou em métodos atrativos numericamente para aplicação em problemas de grande porte que, juntamente com os resultados teóricos, consistem das contribuições mais significativas do presente trabalho (veja [1]).

## 2 O método e resultados teóricos

Como esperado do método de Newton, a recorrência (3) possui convergência quadrática para  $A^\dagger$ , desde que o iterado inicial seja escolhido apropriadamente. Uma escolha usual na literatura (também considerada neste trabalho) costuma ser  $X_0 = \beta A^T$ , com  $0 < \beta < \frac{2}{\rho(AA^T)}$ , em que  $\rho(\cdot)$  denota o raio espectral de  $(\cdot)$ . O cálculo explícito de  $\beta$  pode ser feito tomando, por exemplo,  $\beta = 1/\|A\|_F^2$  ou  $\beta = \frac{1}{\|A\|_1\|A\|_\infty}$ , que satisfazem a condição necessária.

A primeira contribuição deste trabalho consiste de que, se  $0 < \beta < \frac{1}{\rho(AA^T)}$ , então podemos garantir que os iterados  $X_k$  capturam as informações da matriz  $A$  associadas aos maiores valores singulares preservando a ordem dos mesmos, isto é, adquirindo dados fornecidos pelos elementos de baixa frequência antes dos demais [1, Theorem 2.1]. Tal propriedade é de importância na construção de aproximações estáveis para  $x_{LS}$  a partir de dados com ruído, pelo fato de que a solução de problemas discretos mal postos é dominada pelos maiores valores singulares de  $A$  [9, 11].

Agora, com relação à sequência vetorial (2), provamos que a mesma satisfaz propriedades de monotonia na norma dos iterados  $x^{(k)}$  e dos resíduos  $r^{(k)} := b - Ax^{(k)}$ , a saber [1, Lemma 3.2]:

- (i)  $\|x^{(k+1)}\|_2 \geq \|x^{(k)}\|_2, k \geq 1,$
- (ii)  $\|r^{(k+1)}\|_2 \leq \|r^{(k)}\|_2, k \geq 0.$

Estas informações são essenciais para a aplicação de alguns critérios de parada heurísticos, tais como L-curve [8,12] e a regra do produto mínimo [2,5] (MPR), que são alternativas, por exemplo, ao princípio da discrepância, na detecção da propriedade de semi-convergência.

Neste âmbito, tomando o problema com perturbações (1) e os iterados gerados por estes dados,  $\tilde{x}^{(k)} = X_k b$ , e considerando que  $b$  pertence à imagem de  $A$ , podemos provar uma estimativa teórica para quando DP é utilizado como critério de parada. De fato, se  $\tilde{x}^{(k(\delta))}$  é o iterado gerado por DP, então [1, Theorem 4.1]

$$\frac{\|\tilde{x}^{(k(\delta))} - x_{LS}\|_2}{\|x_{LS}\|_2} = \mathcal{O}\left(\frac{\delta^{1/2}}{\|b\|_2^{1/2}}\right).$$

Em palavras, a estimativa infere que, quanto menor a perturbação nos dados de entrada, melhor será a aproximação de  $x_{LS}$  gerada pela sequência  $\{\tilde{x}^{(k)}\}$  quando DP determina o índice  $k$  de parada.

### 3 Aceleração via projeção

Agora, pensando num problema geral da forma  $Ax = b$ , nos aspectos computacionais buscamos computar  $x^{(k)} = X_k b$  sem calcular as matrizes  $X_k$  explicitamente em cada iteração. Para tanto, de (3), é fácil ver que podemos escrever

$$X_{k+1} = X_k + (I - X_k A)X_k = X_k + U_k X_k = X_k + U_0^{2^k} X_k,$$

em que  $U_k := I - X_k A$ . Desta forma, temos que

$$x^{(k+1)} = x^{(k)} + U_0^{2^k} x^{(k)}, \tag{4}$$

em que  $x^{(0)} = \beta A^T b$  e  $U_0 = (I - \beta A^T A) \in \mathbb{R}^{n \times n}$  e  $0 < \beta < \frac{2}{\rho(AA^T)}$ . Ainda assim, lidar com a formulação (4) pode ser uma tarefa de alto custo computacional, já que necessitamos operar com as matrizes  $U_0 \in \mathbb{R}^{n \times n}$ . Este custo é, para problemas de grande porte, uma limitação, e pode ser enormemente reduzido se alguma técnica de projeção for utilizada, ou seja, se buscarmos resolver o problema

$$\min_x \|Ax - b\|_2 \text{ sujeito a } x \in \mathcal{V}_\ell := \text{span}\{v_1, \dots, v_\ell\}, \tag{5}$$

para  $v_i$ 's apropriados e  $\ell \ll n$ . Tomando  $V_\ell := [v_1, \dots, v_\ell] \in \mathbb{R}^{n \times \ell}$ , resolver (5) é equivalente a solucionar o problema irrestrito

$$\min_{y \in \mathbb{C}^\ell} \|(AV_\ell)y - b\|_2, \tag{6}$$

para o qual podemos enfim aplicar o método de Newton. Neste cenário, note que a matriz  $U_0$  a ser gerada de (4) toma a dimensão  $\ell \times \ell$ , cujo custo operacional para efetuar operações matriz-matriz e matriz-vetor são reduzidos, se comparados com operar com uma matriz  $n \times n$ .

A grande questão é a escolha do subespaço  $\mathcal{V}_\ell$ , de modo que informações importantes do problema sejam representadas nesse espaço e, ao mesmo tempo, com poucos vetores na base (i.e., com  $\ell$  pequeno), assegurando a aplicação eficiente da sequência vetorial. Neste trabalho, consideramos duas possibilidades:

(i) Tome os seguintes subespaços de Krylov, para  $j \geq 1$ :

$$\mathcal{K}_j(A^T A, A^T r) = \text{span}\{A^T r, (A^T A)A^T r, \dots, (A^T A)^{j-1}A^T r\}, \quad (7)$$

em que  $r$  é um vetor randômico, e

$$\mathcal{K}_j(A, \tilde{b}) = \text{span}\{A\tilde{b}, A^2\tilde{b}, \dots, A^j\tilde{b}\}, \quad (8)$$

ambos com boas propriedades de aproximação tanto dos vetores singulares associados a elementos de baixa frequência como também da imagem de  $AV_j$  [9]. Especificamente, vetores de base para (7) e (8) podem ser gerados utilizando a bidiagonalização de Lanczos e o processo de Arnoldi/Lanczos, respectivamente, a partir dos quais se determina um tamanho  $j$  de subespaço adequado para a realização da projeção (veja [1] para maiores detalhes).

(ii) Subespaço dado a priori: aqui, tomamos os vetores da base DCT [11] (transformada discreta de cosseno), cujas oscilações dos vetores costumam gerar uma boa aproximação para o espaço gerado pelos vetores singulares de  $A$  associados aos elementos de baixa frequência.

Como mencionado, ao determinar o subespaço, projetamos o problema e então o resolvemos a partir do método proposto em (4). Note que para cada subespaço temos uma variante do método, aqui denominadas por SM-Bid, SM-Trid e SM-DCT, representando, respectivamente, o método de Schultz com os subespaços (7), (8) e DCT.

## 4 Exemplos numéricos

Nesta seção, buscamos ilustrar a aplicação das técnicas desenvolvidas aqui em alguns problemas-teste. Daqui em diante,  $NL$  representa o nível de ruído (relativo) presente nos dados, isto é,  $NL = \|\tilde{b} - b\|/\|b\|$ . As implementações foram todas efetuadas em MATLAB.

### 4.1 Problemas de porte médio

Ilustramos as técnicas desenvolvidas em problemas provenientes do pacote *Regularization Tools* [10] e da Galeria do MATLAB, a saber:

- 1) foxgood, 2) phillips, 3) heat, 4) shaw, 5) gravity,  
 6) baart, 7) deriv2, 8) moler, 9) lotkin, 10) prolate,  
 11) lehmer, 12) cauchy, 13) fiedler, 14) frank, 15) hilb.

	Problema	CGLS	SM-DCT	SM-Bid	SM-Trid	LW	CIM	SART
1	foxgood	0.0070	0.0077	<b>0.0037</b>	0.0070	0.0368	0.0273	0.1086
2	phillips	<b>0.0033</b>	0.0050	0.0049	0.0055	0.0232	0.0233	0.0272
3	heat	<b>0.0145</b>	-	0.0163	-	0.2917	-	0.2052
4	shaw	0.0474	0.0472	<b>0.0398</b>	0.0475	0.1649	0.1032	0.3999
5	gravity	0.0077	0.0104	<b>0.0073</b>	0.0099	0.0421	0.0373	0.0289
6	baart	0.1145	0.1149	0.0638	<b>0.0358</b>	0.3217	0.3363	0.4027
7	deriv2	0.1071	0.0261	0.1192	0.1281	0.3531	0.1721	<b>0.0003</b>
8	moler	0.0139	<b>0.0119</b>	0.0147	0.0151	0.4197	0.4570	-
9	lotkin	0.4496	0.4497	<b>0.4487</b>	-	-	0.4699	0.4531
10	prolate	0.0003	0.0018	-	0.0003	<b>0.0001</b>	0.0002	0.0002
11	lehmer	0.0054	<b>0.0040</b>	0.0071	0.0076	0.3591	0.3648	0.3766
12	cauchy	0.4393	0.4393	<b>0.4392</b>	0.4393	-	0.4700	0.4513
13	fiedler	0.0050	<b>0.0033</b>	0.0060	0.0047	0.3514	0.3673	0.3441
14	frank	0.0062	<b>0.0051</b>	0.0112	0.0289	0.1013	0.3801	0.2659
15	hilb	0.4394	<b>0.4393</b>	0.4394	0.4394	-	0.4700	0.4513

Tabela 1: Média de erros para  $NL = 0.1\%$ .

Os exemplos aqui são considerados com matrizes de ordem  $10000 \times 10000$ , com  $NL = 0.1\%$ . Por simplicidade, o único critério de parada apresentado aqui é DP, mesmo que testes com MPR e L-curve tenham sido efetuados [1]. Por vias de notação nas tabelas e gráficos, consideramos resultados comparativos com o método CGLS, método de Landweber (LW), método de Cimmino (CIM) e *Simultaneous Algebraic Reconstruction Technique* (SART). Veja a Tabela 1 para as médias de erros obtidas e a Figura 1 para conferir os tempos médios de operação, após 30 resoluções de cada problema. Ressaltamos que a base DCT foi tomada fixa com 30 vetores, ao passo que SM-Bid e SM-Trid determinam internamente um tamanho de subespaço apropriado [1].

Note que os resultados obtidos são competitivos com os apresentados pelas estratégias comparadas e, mais ainda, que os tempos de execução se mostram atrativos para aplicação em problemas de ordem maior. Especialmente para SM-Trid e SM-DCT, veja que os tempos de execução foram ainda inferiores aos obtidos por CGLS.

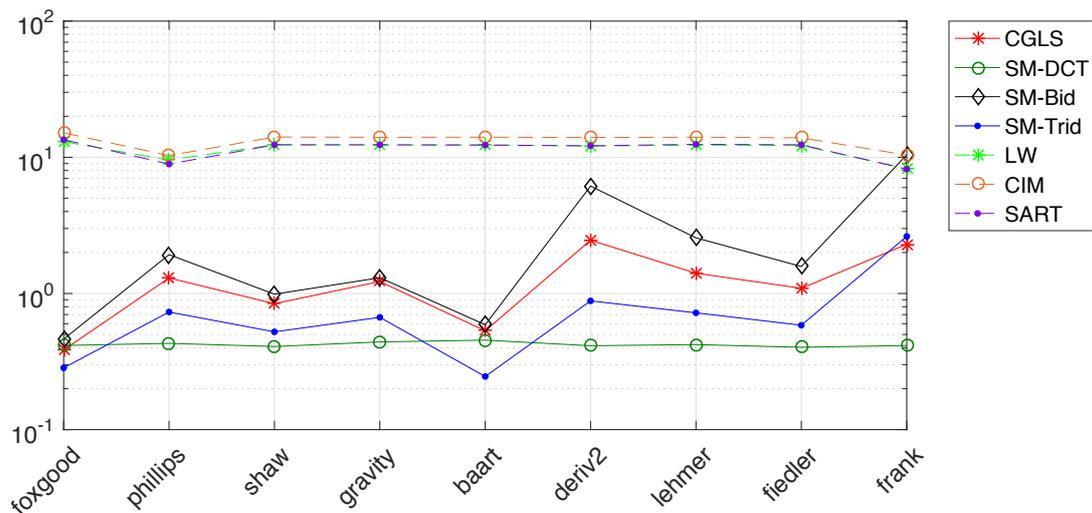


Figura 1: Tempo operacional, em segundos, para nível de ruído  $NL = 0.1\%$ .

## 4.2 Problema de restauração de imagem

Considere a imagem *satellite*, que contém  $256 \times 256$  pixels, do pacote *Restore Tools* [15], à qual adicionamos *blur* e ruído. O problema resultante possui matriz de ordem  $256^2 \times 256^2$ , de modo que são 65536 variáveis para serem determinadas.

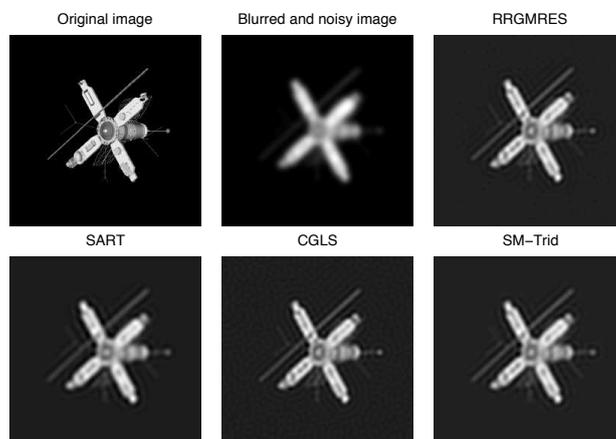


Figura 2: Melhores erros relativos para imagem *satellite*,  $NL = 0.1\%$ .

Na Figura 2 apresentamos tanto a imagem original (solução exata) e a imagem com *blur* e

ruído  $\tilde{b}$ , bem como também os melhores resultados para os métodos utilizados (note a inclusão de RRGMRES nestes testes). Perceba que, visualmente, as imagens recuperadas são virtualmente as mesmas, a despeito do erro relativo inferior apresentado por CGLS (veja Tabela 2). Em questão de tempo operacional, no entanto, SM-Trid apresenta vantagem mais clara (neste quesito), ao menos para este problema, se tornando uma alternativa atrativa para problemas de maior porte.

	RRGMRES	CGLS	SM-Trid	SART
Erro (relativo)	0.2636	<b>0.2618</b>	0.2727	0.2899
Tempo (s)	0.5969	0.9682	<b>0.0571</b>	0.1317

Tabela 2: Resultados para o problema de restauração de imagem.

## 5 Conclusão

Propomos um novo método iterativo para resolver problemas discretos mal-postos baseado em iterações matriciais geradas pelo método de Schultz. Como resultado, provamos que as iterações são quadraticamente convergentes e, com o comportamento de semi-convergência surgindo quando os dados de entrada são corrompidos por ruído, aplicamos regularização através de truncamento das iterações. Disto, exibimos uma análise de erro para quantificar a qualidade das soluções computadas e derivamos uma estimativa de erro para o caso em que o parâmetro de truncamento é gerado pelo princípio da discrepância.

Em adição, ao aliarmos o método proposto com estratégias de projeção em subespaços, conseguimos alternativas aplicáveis em problemas de grande porte, equiparando (e eventualmente melhorando) os resultados gerados por algoritmos presentes na literatura. Neste âmbito, novos focos de estudo podem envolver formas de escolher subespaços de interesse à projeção e também a aplicação das mesmas ideias com outros métodos já existentes na literatura. Os exemplos numéricos exibidos mostram que o método de Schultz funciona bem quando acoplado com critérios de parada apropriados, indicando que o método é atrativo para problemas discretos mal postos envolvendo dados com ruído em níveis realísticos.

Propostas de trabalhos futuros envolvem: estender os resultados teóricos obtidos para outras versões de métodos iterativos de ordem mais alta; melhorar a estimativa de erro para DP considerando condições de fonte do tipo Hölder e logarítmicas; e analisar técnicas para inclusão de informações a priori no processo do método proposto, que tendem a introduzir dados acerca da suavidade e/ou comportamento da solução exata, podendo gerar melhorias significativas nas soluções aproximadas.

## Referências

- [1] F. S. V. Bazán, E. Boos, Schultz matrix iteration based method for stable solution of discrete ill-posed problems, *Linear Algebra and its Applications*, vol. 554, pp. 120–145, DOI: 10.1016/j.laa.2018.05.022, (2018).
- [2] F. S. V. Bazán, L. S. Borges, GKB-FP: an algorithm for large-scale discrete ill-posed problems, *BIT*, DOI 10.1007/s10543-010-0275-3, (2010).
- [3] A. Ben-Israel, T. N. E. Greville, “Generalized Inverses: Theory and Applications”, 2nd. ed., New York: Springer, 420 p., 2003.
- [4] A. Björck, “Numerical Methods for Least Squares Problems”, Philadelphia: SIAM, 1996.
- [5] L. S. Borges, F. S. V. Bazán, M. C. Cunha, Automatic stopping rule for iterative methods in discrete ill-posed problems, *Comp. Appl. Math.*, DOI 10.1007/s40314-014-0174-3, (2014).

- [6] D. Calvetti, B. Lewis, L. Reichel, GMRES-type methods for inconsistent systems, *Linear Algebra and its Applications*, vol. 316, pp. 157–169, (2000).
- [7] M. Hanke, “Conjugate gradient type methods for ill-posed problems”, Harlow, UK: Longman, 1995.
- [8] P. C. Hansen, Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Rev.*, vol. 34, pp. 561–580, (1992).
- [9] P. C. Hansen, “Rank-Deficient and Discrete Ill-Posed Problems”, Philadelphia: SIAM, 247 p., 1998.
- [10] P. C. Hansen, Regularization Tools: A MATLAB package for analysis and solution of discrete ill-posed problems. *Numerical Algorithms*, vol. 6, pp. 1–35, (1994).
- [11] P. C. Hansen, “Discrete Inverse Problems: Insight and Algorithms, 2nd. ed. Philadelphia: SIAM, 213 p., 2010.
- [12] M. E. Kilmer, D. P. O’Leary, Choosing regularization parameters in iterative methods for ill-posed problems, *SIAM J. Matrix Anal. Appl.*, vol. 22, pp. 1204–1221, (2001).
- [13] A. Kirsch “An Introduction to the Mathematical Theory of Inverse Problems”, Applied Mathematical Sciences, vol. 120, New York: Springer, 1996.
- [14] V. A. Morozov, On the solution of functional equations by the method of regularization, *Soviet Math. Dokl.*, vol. 7, pp. 414–417, (1966).
- [15] J. G. Nagy, K. Palmer, L. Perrone, Iterative methods for image deblurring: a Matlab object-oriented approach, *Numerical Algorithms*, vol. 36, pp. 73–93, (2004).
- [16] F. Natterer, “The Mathematics of Computerized Tomography”, New York: Wiley, 226 p., 1986.
- [17] C. C. Paige, M. A. Saunders, Solution of sparse indefinite systems of linear equations, *SIAM Journal on Numerical Analysis*, vol. 12, pp. 617–629, (1975).
- [18] C. C. Paige, M. A. Saunders, LSQR: An algorithm for sparse linear equations and sparse least squares, *ACM Trans. Math. Softw.*, vol. 8, pp. 43–71, (1982).
- [19] R. Penrose, A generalized inverse for matrices. *Proc. Cambridge Phil. Soc.*, vol. 51, pp. 406–413, (1955).
- [20] Y. Saad, M. Schultz, GMRES: A Generalized Minimum Residual Algorithm for Solving Nonsymmetric Linear Systems., *SIAM J. Scientific and Stat. Comp.*, vol. 7, pp. 856–869, (1986).
- [21] A. J. Silva Neto, F. D. Moura Neto, “Problemas Inversos: Conceitos fundamentais e aplicações”, Rio de Janeiro: UERJ, 2005.
- [22] G. Schultz, Iterative Berechnung der Reciproken Matrix, *Z. Angew. Meth. Mech.*, vol. 13, pp. 57–59, (1933).
- [23] A. N. Tikhonov, Solution of incorrectly formulated problems and the regularization method, *Soviet Math. Dokl.*, vol. 4, pp. 1035–1038, (1963).



# Previsibilidade pluviométricas: Um estudo matemático analítico e estatístico sobre o volume das precipitações de São Paulo

Giselle Couto Falcão\*

Programa de Pós Graduação em Modelagem Matemática Computacional, CEFET-MG, Belo Horizonte, MG

Edgar Lacerda Aguiar†

Programa de Pós Graduação em Modelagem Matemática Computacional, CEFET-MG, Belo Horizonte, MG

**Resumo:** *A previsibilidade das inundações tem sido um desafio recorrente nas últimas décadas nas grandes cidades e principalmente em São Paulo, e diversas outras cidades do Brasil. Nas estações de primavera e verão algumas determinadas regiões do Brasil sofrem com constantes alagamentos, enchentes e inundações. Tais fenômenos ocorrem principalmente devido ao aumento das precipitações intensas e falta de investimento dos órgãos responsáveis em prevenção. Este trabalho buscou analisar as previsões do volume máximo de precipitações do bairro Jardim de São Paulo, as análises dos dados foram realizados utilizando os métodos estatísticos não paramétricos de séries temporais não estacionárias reescaladas para a observação de sazonalidade. Com objetivo de localizar uma possível linearidade das médias móveis auto-regressivas no período sazonal. Tendo como modelo de apresentação estrutural a simulação dos dados em sistemas complexos de previsibilidade e instabilidade, o qual terá como base, dados históricos mensais dos volumes das precipitações dos meses de novembro, dezembro e janeiro coletados pela estação 83781 no Mirante de Santana. Por meio de métodos estatísticos, bem como análises residuais, que garantiram uma defasagem adequada para conferir. Este trabalho buscou testar e validar as hipóteses sugeridas sobre o volume máximo das precipitações, considerando as séries históricas dos últimos 58 anos de São Paulo. Quanto as reincidências periódicas não anuais dos altos volumes de precipitações, e avaliar se tal volume esta em crescimento.*

**Palavras-chave:** *Autocorrelação, Clima, Estudo tendência, Previsibilidade pluviométricas, Precipitações.*

## 1 Introdução

Por meio da análise de sazonalidade relacionada ao volume das precipitações vem sendo amplamente discutido, entre estes assuntos existe corrente discussão quanto ao comportamento cíclico de períodos de grande volume de precipitação e além disso apresenta um crescimento linear, que pode ser evidenciadas de forma eficaz, por meio testes não paramétricos como Dickey-Fuller e Phillips-Perron [4],[17], [5].

O dimensionamento e comportamento fractal de campos geofísicos, são usados para modelar diversos fenômenos naturais, tais como o volume da precipitação, temperatura, níveis do mar, ventos, entre outros. Este dimensionamento nos permite o uso de técnicas matemáticas que descrevem o modelo em ajustes de transformadas Fourier (TF), que permite o ajuste da série, em busca de períodos com maior incidência de precipitações no domínio do tempo. Esta ferramenta permite uma verificação mais eficiente quanto a sua previsibilidade e ciclicidade. [11].

As séries temporais com comportamento não estacionário, como é o caso da série objeto desta pesquisa, após submetidas a ajustes adequados, tais como TF. Que nos permite a elaboração e verificação de modelo matemático capaz de ajustar de forma concisa os dados da série temporal,

---

\* giselle\_falcao@hotmail.com

† edgarlaguiar@gmail.com

observando a possibilidade pontual de uma análise de tendência da série. dentre estes, o modelo Holt-Winter, representa um algoritmo eficaz que garante as características da series quanto a sazonalidade, estacionariedade e principalmente a normalidade. [10]

Outra análise que garante a observação de resultados é quanto aos resíduos, estes podem ser observados assim como a serie principal por meio do uso de modelos de médias móveis (SMA), autocorrelação, autocovariância entre outros, que permite o estudo do ruído da série, que garante diversas informações sobre a série, ajuste de modelos entre outros. [2] [3].

Deste modo o tratamento analítico da série temporal, apos ajustes, reescalonamento e transformações adequadas, podem ser submetidos a uma análise de tendência cuidadosa e seletiva que por ser realizada com o auxilio do modelo auto-regressivo integrado de médias móveis (ARIMA). Este modelo comumente usado na econometria, permite a análise de tendência em séries temporais, baseando-se em alguns parâmetros, como o nível de defasagem da autocorrelação da série, as medias anteriores e posteriores e a diferença. Com estas características o modelo torna-se uma ferramenta muito interessante para observação a longo prazo da série além de auxiliar no calculo das previsões do modelo ajustado [18].

Utilizando a interação entre estes métodos, observando os ajustes, os resíduos e as transformações, este estudo busca evidencias do crescimento do volume de precipitação na cidade de São Paulo, focado na média, desvio padrão e variância das flutuações em séries históricas.

Este trabalho foi realizado utilizando o refinamento dos dados de volume das precipitações, de forma estatísticas obtidos por meio das análises de longas séries de dados coletados mensalmente, dos registros históricos da precipitação do Instituto Nacional de Meteorologia (INMET), colhidos durante 58 anos na estação 83781 [9].

## 2 Metodologia

Foi utilizado longas séries históricas dos volumes de precipitação mensais em São Paulo, coletados na estação 83781 no Mirante de Santana no Bairro Jardim, a estação é a principal estação do estado. Ela composta de vários sensores isolados que registram continuamente os parâmetros meteorológicos como a pressão atmosférica, temperatura, umidade relativa do ar, precipitação, radiação solar, direção e velocidade do vento, etc. Tais dados são anotados por um observador a determinado intervalo e enviado ao um centro coletor por um meio de comunicação.

Os dados foram submetidos inicialmente ao teste de normalidade e estacionariedade, a sazonalidade dos dados foi observada pela autocorrelação. Apos esta análise a série foi submetida a TF, para uma análise espectral dos dados, a partir deste resultados foram extraídos os períodos de maior incidência de chuva que corresponde ao meses de novembro, dezembro e janeiro dos últimos 58 anos.

A utilização do método aditivo de HoltWinters garante uma suavização exponencial a série. com a equação de previsão e três equações de suavização - uma para o nível , uma para a tendência e uma para o componente sazonal, com o correspondente parâmetros de suavização. Dado pelo modelo apresentado na equação 1.

$$s_t = \gamma(1 - \alpha)(y_t - \beta_{(t-1)} - b_{(t-1)}) + [1 - \gamma(1 - \alpha)]s_{(t-m)}. \quad (1)$$

Se  $0 \leq \gamma \leq 1$ ,

Já para a utilização do ARIMA temos t, que representa um índice pertencentes aos números reais, o qual o modelo pode ser descrito pela equação 2, a seguir. é importante salientar que o processo ARIMA dado por  $(p, d, q)$ .

$$\left(1 - \sum_{i=1}^{p'} \alpha_i L^i\right) = \left(1 - \sum_{i=1}^{p'-d} \phi_i L^i\right) (1 - L)^d \cdot \left(1 - \sum_{i=1}^{p'} \alpha_i L^i\right) = \left(1 - \sum_{i=1}^{p'-d} \phi_i L^i\right) (1 - L)^d. \quad (2)$$

Neste trabalho utilizamos os dados mensais para todo o ano, quanto os dados do meses de novembro, dezembro e Janeiro. em busca de comportamento crescente linear ao não para o volume máximo de precipitação, para isto utilizamos os resultados do ARIMA para analisar a tendencia de forma associada aos dados de HoltWinters, por meio de uma analise logarítmica.

### 3 Resultados Obtidos

A partir dos métodos realizados e gráficos desenvolvidos, pode-se afirmar a não estacionariedade da série, bem como sua não normalização. A decomposição da série e a flutuação da SMA pode indicar a incidência crescente no volume das precipitações dos meses novembro, dezembro e janeiro. Em conjunto com autocorrelação foi possível observar a persistência dos altos volumes das precipitações após períodos extensos de baixo volume de precipitação em São Paulo. Em sequência o ajuste exponencial aditivo demonstrado pela equação (1) foi possível modelar a série que foi submetida a um melhoramentos considerando a defasagem de 5 anos.

A necessidade e possibilidade de melhoramento do modelo foi possível por meio da autocorrelação residual, que apontou persistência em 5 anos. Uma vez ajustado o modelo a série foi submetida a uma transformada de Fouries para garantir sua estacionariedade e normalidade, deste modo foi possível utilizamos para calculo ARIMA as informações obtidas pelo modelo HoltWintrs modificado, Deste modo o modelo ARIMA utilizado (9,0,0) (1,0,1) pode nos garantir por meio dos parâmetros tanto a defasagem da série, quanto a defasagem residual. Um crescimento dentro das médias esperadas de previsão considerando o estimativa calculado pelo modelo ARIMA, em que foi detectado um coeficiente  $\rho = 0.687564$ . Isto garante O que indica que essa persistência é sazonal, ou seja, obedecem uma periodicidade de aproximadamente 6-8 anos entre períodos de alto volume e baixo volume de precipitações.

Este aumento pode ser identificado de modo temporal para as precipitações de alto volume ocorridas em São Paulo por meio de análise de tendências que o considerando os dados históricos coletados.

Na Figura 1, observa-se a decomposição de série temporal, é possível verificar que os volumes mínimos e máximos nos meses de novembro, dezembro e janeiro de 1961 a 2018.

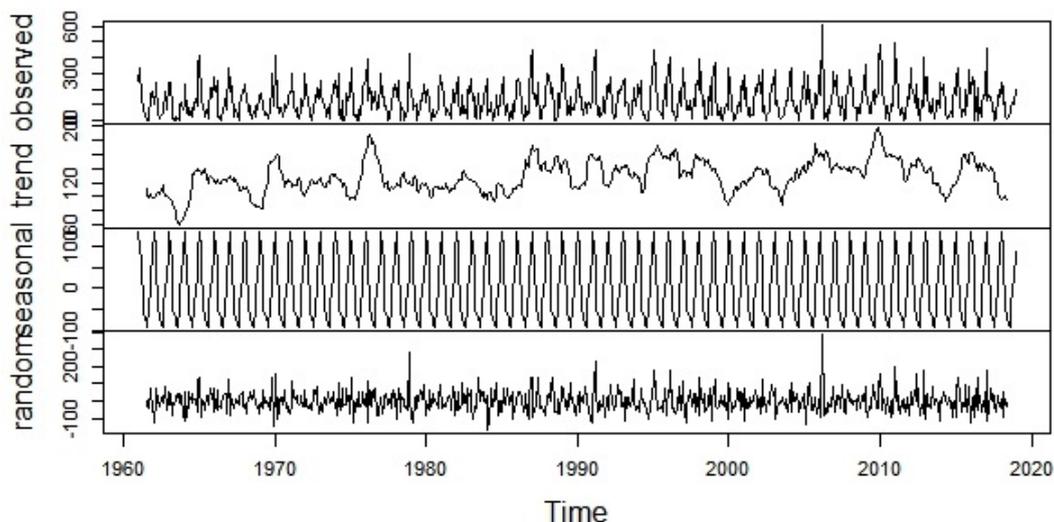


Figura 1: Decomposição da série

Após as análises dos dados e os ajustes podemos descrever um modelo que se aproxima muito bem a série original, como é possível verificar na Figura 2. No gráfico em questão a linha preta indica a série original e em vermelho o modelo proposto para ajuste da série utilizando o método Holt Winter observado a equação (1).

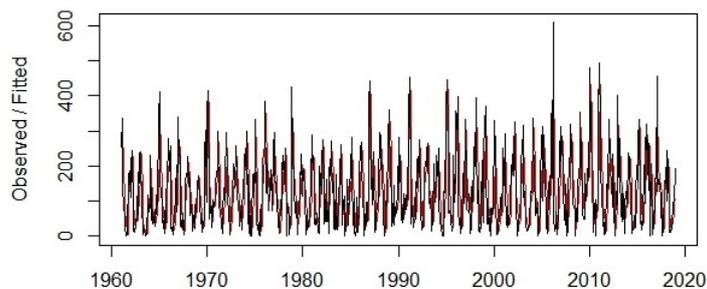


Figura 2: Ajuste da série do volume de precipitação

Este ajuste contou com o coeficiente de autorregressivo, para inferir ao modelo resultado melhor. Para o desenvolvimento do coeficiente em questão, utilizamos a autocorrelação dos dados originais com 20 defasagens distintas, onde cada resíduo de uma autocorrelação foi armazenado e posteriormente correlacionados, em seguida foi realizado o cálculo do atraso da série, como é possível observar na figura 3 abaixo.

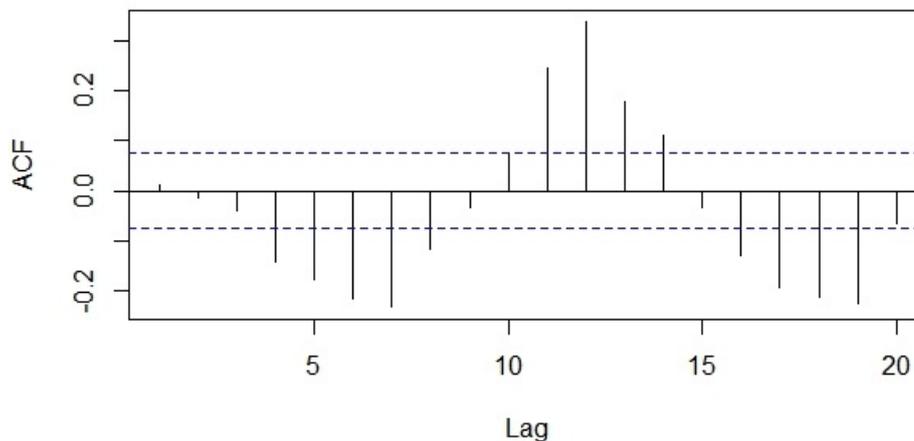


Figura 3: Correlação residual

Após esta verificação foi aplicado a análise de tendências utilizando os dados obtidos pelo modelo ARIMA, que se encontra abaixo na Figura 4. Na mesma é possível observar uma forte tendência ascendente (trecho em azul) para altos volumes de precipitações da Cidade de São Paulo no período de 2020 a 2030.

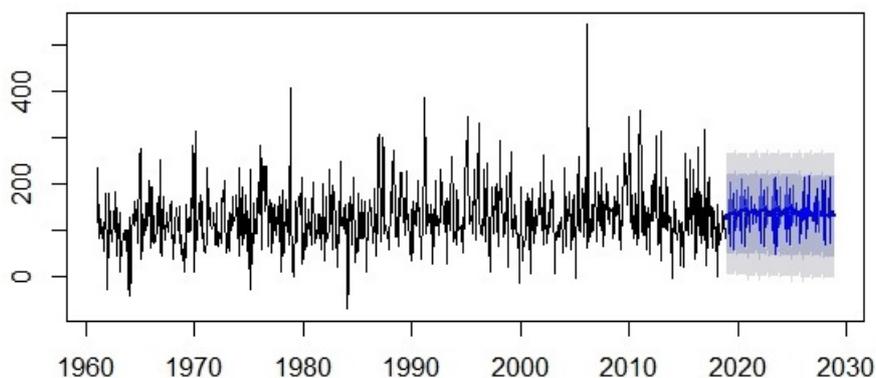


Figura 4: Análise de Tendência

Tal informação auxilia na validação do modelo, que se encontra bem ajustado para os dados utilizados e o mesmo teve uma validação excelente com uma alta margem superior a 95%, para o período de 11 anos contando entre 2019 a 2030.

## 4 Conclusão

A partir dos resultados obtidos e das discussões realizadas, foi possível identificar uma ocorrência de sazonalidade no volume das precipitações ocorridas nos meses de novembro, dezembro e janeiro dos últimos 57 anos (1961-2018). Esta sazonalidade obteve uma avaliação de 95% no grau de confiança, no qual foi observado um período médio de 7,2 anos, entre os volumes máximos de precipitações em São Paulo de acordo com as medições da Estação 83587 Mirante de Santana. Através das análises de tendências obtidas, foi possível realizar previsões dos volumes máximos das precipitações.

Graças as elevadas confiabilidade obtidas por meio das análises de tendências das futuras precipitações que podem vir a auxiliar em diversas áreas engenharias e recursos hídricos e outras. Tais métricas são de elevada importância para os órgãos e governos para auxiliar e prever os futuros investimentos em infraestrutura e saneamento objetivando uma minimização dos danos causados pelas enchentes e nas melhorias sociais ambientais que ocorrem na cidade de São Paulo.

## Referências

- [1] BACK, Álvaro José. Aplicação de análise estatística para identificação de tendências climáticas. Pesquisa agropecuária brasileira, v. 36, n. 5, p. 717-726, 2001.
- [2] BOLLERSLEV, Tim et al. A conditionally heteroskedastic time series model for speculative prices and rates of return. Review of economics and statistics, v. 69, n. 3, p. 542-547, 1987.
- [3] BRAGA, A. C. et al. Sobre a Detecção de Autocorrelações em Séries Temporais: Uma Comparação Objetiva entre Análise de Flutuações, Transformações Wavelet e Análise Entrópica. v. 3, n. Iii, p. 1-7, 2015.
- [4] CONOVER, W.J. Practical Nonparametric Statistics. New York, Wiley, 1980
- [5] DICKEY, D.A. and W.A. FULLER. "Distribution of the Estimators for Autogressive Time Series with A Unit Rot", Journal of the American Statistical Association, 74, p. 427-431, 1979.
- [6] FARIAS, A. et al. ESTIMATIVA DO EXPOENTE DE HURST EM DADOS SÍSMICOS UTILIZANDO ANÁLISE FRACTAL. I congresso Nacional de Engenharia de Petroleo, Gas Natural e Biocombustíveis. Anais 2015.
- [7] FAVARETTO, A. B. Estimativa do Expoente de Hurst de Série Temporais de Chuvas do Estado de São Paulo Usando as transformadas de Fourier, Wavelets e Análise R/S. [s.l.] Universidade Estadual Paulista, 2004.
- [8] GNEDENKO, B.V, AND A.N. KOLMOGOROV Limit Distributions for Sums of Independent Random Variables, Addison-Wesley: Reading, 1968.
- [9] INMET, MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO, Disponível em: [www.inmet.gov.br/](http://www.inmet.gov.br/). Acesso em: 20 de Março de 2019.
- [10] KALEKAR, Prajakta S. Time series forecasting using holt-winters exponential smoothing. Kanwal Rekhi School of Information Technology, v. 4329008, n. 13, 2004.

- [11] OLIVEIRA, H. M. DE. Análise de Fourier e Wavelets: sinais estacionário e não estacionários. Recife-PE: 2007, 2007.
- [12] SAYAMA, Hiroki. Introduction to the modeling and analysis of complex systems. Open SUNY Textbooks, 2015.
- [13] SPROTT, J.C. Chaos and Time-Series Analysis, Oxford University Press: Oxford, 2003.
- [14] TAN, X.; GAN, T. Y. Multifractality of Canadian precipitation and streamflow. International Journal of Climatology, v. 37, p. 1221–1236, 2017.
- [15] THOM, H. C. S. Some methods of climatological analysis. Genève : World Meteorological Organization, 1966. 54 p. (WMO Technical Note, 81).
- [16] VITOI, D. A. DE O. Coeficiente de Hurst na Comparação de séries de nível do mar. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, 2016.
- [17] YU, PAO-SHAN, TAO-CHANG YANG, and CHUN-CHAO KUO. "Evaluating long-term trends in annual and seasonal precipitation in Taiwan." Water Resources Management 20.6 (2006): 1007-1023.
- [18] ZHANG, G. Peter. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, v. 50, p. 159-175, 2003.

# Um Estudo sobre a Sincronização do Ritmo Circadiano: Modelo PIM

**Stefânia da S. Glaeser,**

Programa de Pós Graduação em Modelagem Computacional, FURG,

Rio Grande, RS

E-mail: stefaniaglaeser@gmail.com,

**Adriano De Cezaro**

Instituto de Matemática, Estatística e Física, FURG,

Rio Grande, RS

E-mail: decezaromtm@gmail.com.

**Resumo:** *O ritmo circadiano regula diversas funções importantes do corpo humano. Já é bem conhecido da literatura que alterações na sua ritmicidade pode desencadear transtornos indesejados. Portanto, compreender como se comporta o ritmo circadiano e como fatores externos afetam na sua sincronização ou dessincronização podem gerar procedimentos médicos futuros mais eficientes. Neste trabalho, estudaremos condições suficientes para a sincronização e dessincronização do ritmo circadiano, sendo modelado com os ciclos da temperatura corporal, atividade-reposo e da dor, como um fator externo. A principal novidade consiste na solução analítica dos casos de sincronizado total e parcial dos ciclos, cujas forças de acoplamento do modelo são distintas. Por fim, apresentaremos alguns resultados numéricos que descrevem características de sincronização e dessincronização para o modelo proposto.*

**Palavras-chave:** *Ritmo Circadiano, Osciladores não-lineares, Sincronização, Modelagem*

## 1 Introdução

É fascinante como um mecanismo simples pode controlar muitas funções vitais de um organismo vivo - o relógio biológico interno. De fato, nosso relógio biológico interno é o responsável por adaptar precisamente nossa fisiologia, independente das fases de movimento drástico do dia, regulando funções importantes como o comportamento, o metabolismo, os níveis hormonais, o sono e a temperatura corporal [2].

A funcionalidade do relógio biológico interno e como este pode acarretar na sincronização ou dessincronização do ritmo circadiano, gerando, eventualmente, certos distúrbios, vem gradativamente sendo estudado por pesquisadores de diversas áreas, as quais englobam desde bio-medicina - na qual a importância do estudo associado ao ritmo circadiano pode ser comprovada, haja vista que os agraciados com o Prêmio Nobel em Fisiologia ou Medicina 2017 estudaram e descobriram mecanismos moleculares que controlam o ritmo circadiano - até a modelagem matemática, para o qual busca-se por modelos que descrevem o comportamento do ritmo circadiano ou que tentem descrever características de interesse deste, e.g., [1, 2, 6]. Dentre estes, temos a modelagem do ritmo circadiano por simulação multiagente, que é também uma de nossas linhas de pesquisas [3].

De vital importância nos estudos associados ao ritmo circadiano é o entendimento de como fatores externos influenciam na sincronização ou dessincronização do mesmo. Uma vez que, como comentado anteriormente, a dessincronização do ritmo circadiano pode desencadear alterações indesejadas de diversas funções vitais do corpo humano. Um apanhado mais geral sobre ritmo circadiano pode ser visto em [1, 2] e referências.

Sendo assim, nossa hipótese científica se divide em dois pontos: o ritmo circadiano pode ser modelado utilizando sistemas acoplados de três osciladores não lineares, em que a dor é considerado como um fator externo que está relacionada com a dessincronização deste ritmo; e pacientes com ritmo circadiano sincronizado se recuperam mais rápido em relação a pacientes com o ritmo circadiano dessincronizado.

Portanto o objetivo deste trabalho consiste em modelar aspectos de sincronização e dessincronização do ritmo circadiano, tendo a dor como um agente externo. Para tal, assumiremos que ritmo circadiano é definido pelos ciclos da temperatura corporal, atividade-reposo e dor, acoplados por osciladores não lineares de fase, cujas frequências intrínsecas, forças de acoplamento e condições iniciais são consideradas na modelagem, resultando em um modelo similar ao proposto por Kuramoto [1, 4] e Strogatz [6].

As principais novidades deste trabalho são:

- Propusemos um modelo para o ritmo circadiano definido pelos ciclos da temperatura corporal, atividade-reposo sob a influência externa da dor, o qual pode ser visto como uma generalização do modelo de Strogatz [6]. Veja a Seção 2.
- Provamos que o modelo é bem posto no sentido de Hadamard. Veja o Teorema 2.1.
- Apresentamos propriedades de sincronização e dessincronização, parcial e total, para o modelo proposto. Em particular, nos caso de sincronização total e parcial, apresentamos a solução analítica do problema estudado, generalizando assim, os resultados de [6]. Veja as Subseções 2.1 e 2.2.
- Os resultados obtidos permitem entender como a dor pode influenciar na sincronização do ritmo circadiano.

## 2 Modelo PIM

Os modelos que utilizamos como embasamento na modelagem do ritmo circadiano são os modelos de Kuramoto [4] e Strogatz [6]. Strogatz considera dois osciladores não lineares fracamente acoplados, sendo eles os ciclos da temperatura corporal e sono-vigília, para modelar o ritmo circadiano. Kuramoto propõe um modelo geral, cujo número de osciladores acoplados é  $N$ . Os casos estudados do modelo de Kuramoto geralmente consideram alto o número de osciladores fracamente acoplados, ou seja,  $N$  tende ao infinito. Além disso, no modelo de Kuramoto as forças de acoplamento entre os osciladores são consideradas iguais. Uma revisão completa sobre o assunto pode ser encontrada em [1, 4, 6] e referências.

Nossa proposta de modelagem, intitulada modelo PIM, consiste no acoplamento de três osciladores não lineares fracamente acoplados, cujas forças de acoplamento são consideradas distintas. O primeiro oscilador conduz o ciclo da temperatura corporal, o segundo o ciclo da atividade-reposo e o terceiro, que será considerado como o oscilador que influência de forma externa, será definido como oscilador da dor. A figura 1 traz a topologia de acoplamento.

Podemos observar na figura 1 os osciladores  $\theta_1(t)$  e  $\theta_2(t)$  acoplados - estes equivalem ao modelo de Strogatz. O terceiro oscilador,  $\theta_3(t)$ , acoplado aos outros dois, conduz o ciclos da dor. Na figura 1, podemos ver ainda, que as forças de acoplamento das ligações entre os osciladores não são necessariamente iguais.

Optamos por utilizar tal proposta de modelagem, visto que a dinâmica de fase de osciladores acoplados é um caso limitativo de esquemas dinâmicos gerais e plausíveis [6], muito semelhante ao relógio biológico interno que regula o ritmo circadiano.

As equações dinâmica do modelo PIM são

$$\dot{\theta}_1(t) = \omega_1 + B_1 \sin(2\pi(\theta_2(t) - \theta_1(t))) + C_1 \sin(2\pi(\theta_3(t) - \theta_1(t))) \quad (1)$$

$$\dot{\theta}_2(t) = \omega_2 + A_1 \sin(2\pi(\theta_1(t) - \theta_2(t))) + C_2 \sin(2\pi(\theta_3(t) - \theta_2(t))) \quad (2)$$

$$\dot{\theta}_3(t) = \omega_3 + A_2 \sin(2\pi(\theta_1(t) - \theta_3(t))) + B_2 \sin(2\pi(\theta_2(t) - \theta_3(t))), \quad (3)$$

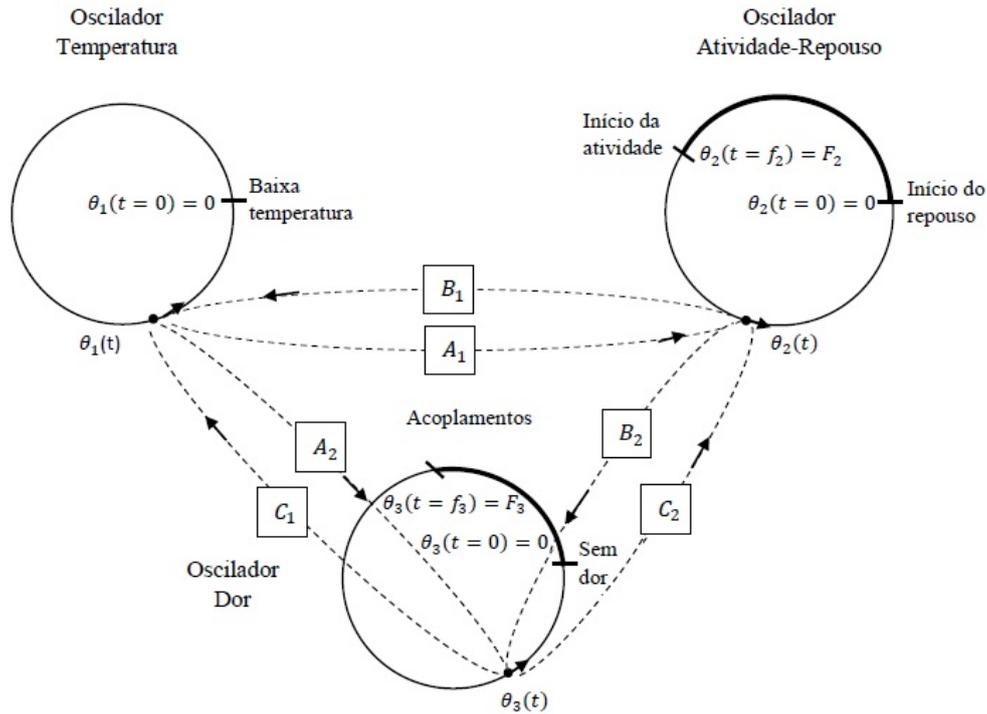


Figura 1: Modelo PIM

onde  $\omega_i = \frac{1}{\tau_i}$  são as frequências intrínsecas, com  $\tau_i$  representando o período em horas do ciclo  $i$ ,  $i = 1, 2, 3$ ;  $A_1, A_2, B_1, B_2, C_1$  e  $C_2$  são as forças de acoplamento constantes definidas no intervalo  $[-1, 1]$ ;  $\theta_1(t = 0) = 0, \theta_2(t = 0) = 0, \theta_3(t = 0) = 0$  são as condições iniciais.

Para o problema formulado acima, podemos provar o seguinte resultado:

**Teorema 2.1.** *Para o modelo PIM, com condições iniciais homogêneas dadas acima, podemos garantir que existe uma única solução  $(\theta_1, \theta_2, \theta_3)^T \in C^1([0, \infty[, \mathbb{R}^3)$ , que depende continuamente dos parâmetros e dos dados iniciais.*

**Demonstração:** Como  $|\sin X| \leq |X|$ , segue que o lado direito de (1)-(3) é contínuo com relação a  $t \in [0, T]$  e Lipschitz contínuo com relação a  $(\theta_1, \theta_2, \theta_3)^T$ . Isso implica ainda que a solução (se existir), cresce no máximo linearmente. Assim, segue da teoria clássica de EDO's (e.g., [5]) que a afirmativa do teorema é verdadeira.  $\square$

Nas próximas subseções apresentaremos os resultados mais importantes desta contribuição.

## 2.1 Sincronização Total

Dizer que o ritmo circadiano está sincronizado significa dizer que os ciclos biológicos possuem uma relação entre si e com o dia. Considerando os osciladores dois a dois, suas forças de acoplamento e frequências intrínsecas, segundo Winfree [7], temos que se o módulo da soma das forças de acoplamento for maior que o módulo da diferença das frequências intrínsecas, então a relação entre tais osciladores é sincronizada, caso contrário, é dessincronizada. Ou ainda, olhando para fase dos osciladores, temos que se dois osciladores estão sincronizados, então a diferença de fase entre estes é nula ou constante.

Desta forma, para que ocorra a sincronização total é necessário que os osciladores dois a dois satisfaçam a definição de sincronização (e assim  $\theta_1(t) = \theta_2(t) = \theta_3(t)$ ), logo as equações dinâmicas do modelo PIM podem ser escritas como

$$\dot{\theta}_1(t) = \dot{\theta}_2(t) = \dot{\theta}_3(t) = \omega,$$

onde  $\omega$  é a frequência intrínseca dos osciladores - igual para todos.

Assim, segue diretamente do Teorema Fundamental do Cálculo que, para o caso totalmente sincronizado, a solução analítica do sistema (1)-(3) é dada por

$$\theta_1(t) = \omega t \quad (4)$$

$$\theta_2(t) = \omega t - \omega_2 f_2 + F_2 \quad (5)$$

$$\theta_3(t) = \omega t - \omega_3 f_3 + F_3. \quad (6)$$

## 2.2 Sincronização Parcial

Nesta subseção, analisaremos a condição que é mais interessante do ponto de vista da aplicação que temos em mente, ou seja, o caso em que  $\theta_1(t)$  e  $\theta_2(t)$  estão sincronizados (e assim  $\theta_1(t) = \theta_2(t)$  pela definição de sincronização), e estes são influenciados de forma externa pelo oscilador correspondente a dor,  $\theta_3(t)$ , o qual está dessincronizado com os demais.

**Proposição 2.2.** *Assuma que os osciladores  $\theta_1(t)$  e  $\theta_2(t)$  estejam sincronizados. Então, existe uma solução analítica para o modelo PIM (1)-(3), dada por*

$$\theta_1(t) = \left( \frac{\omega_1 C_2 - \omega_2 C_1}{C_2 - C_1} \right) t \quad (7)$$

$$\theta_2(t) = \theta_1(t) - \left( \frac{\omega_1 C_2 - \omega_2 C_1}{C_2 - C_1} \right) f_2 + F_2 \quad (8)$$

$$\theta_3(t) = \left( \frac{(\omega_1 - \omega_2)(-A_2 - B_2)}{C_2 - C_1} + \omega_3 \right) t - \left( \frac{(\omega_1 - \omega_2)(-A_2 - B_2)}{C_2 - C_1} + \omega_3 \right) f_3 + F_3, \quad (9)$$

**Demonstração:** Da hipótese temos que  $\theta_1(t) - \theta_2(t) = 0$ . Assim, as equações dinâmicas do modelo PIM (1)-(3) podem ser reescritas como

$$\dot{\theta}_1(t) = \omega_1 + C_1 \sin(2\pi(\theta_3(t) - \theta_1(t))) \quad (10)$$

$$\dot{\theta}_2(t) = \omega_2 + C_2 \sin(2\pi(\theta_3(t) - \theta_2(t))) \quad (11)$$

$$\dot{\theta}_3(t) = \omega_3 + A_2 \sin(2\pi(\theta_1(t) - \theta_3(t))) + B_2 \sin(2\pi(\theta_2(t) - \theta_3(t))). \quad (12)$$

Defina a diferença de fase

$$\psi_1(t) = \theta_1(t) - \theta_3(t) = \theta_2(t) - \theta_3(t), \quad (13)$$

obtemos

$$\dot{\psi}_1(t) = \dot{\theta}_1(t) - \dot{\theta}_3(t) = \dot{\theta}_2(t) - \dot{\theta}_3(t). \quad (14)$$

Da substituição de (13) e (14) em (10)-(12), obtemos que  $\dot{\psi}_1(t) = 0$  satisfaz

$$\begin{aligned} \omega_1 + C_1 \sin(2\pi\psi_1(t)) &= \omega_2 + C_2 \sin(2\pi\psi_1(t)) \\ \implies \psi_1(t) &= -\frac{1}{2\pi} \arcsin \frac{\omega_1 - \omega_2}{C_2 - C_1}. \end{aligned} \quad (15)$$

Substituindo a equação (15) nas equações (10)-(12) obtemos

$$\dot{\theta}_1(t) = \dot{\theta}_2(t) = \frac{\omega_1 C_2 - \omega_2 C_1}{C_2 - C_1} \quad (16)$$

$$\dot{\theta}_3(t) = \frac{(\omega_1 - \omega_2)(-A_2 - B_2)}{C_2 - C_1} + \omega_3. \quad (17)$$

Agora o resultado segue do Teorema Fundamental do Cálculo e a utilização das condições de iniciais.  $\square$

Seque de (15) que os valores definidos aos parâmetros para o caso parcialmente sincronizado deve satisfazer as restrições

$$\begin{aligned} |\omega_1 - \omega_2| &< |A_1 + B_1| \\ |\omega_1 - \omega_3| &> |A_2 + C_1| \\ |\omega_2 - \omega_3| &> |B_2 + C_2|, \end{aligned} \tag{18}$$

caso contrário, se as relações ente as magnitudes do acoplamento e das frequências intrínsecas do oscilador da dor ( $\theta_3$ ) com relação aos dos osciladores da atividade-repouso e temperatura, farão com que o sistema todo passe a ser dessincronizado, como pode ser visto na Figura 3.

### 2.3 Dessincronização Total

Dada a não-linearidade do modelo PIM (1)-(3), a única maneira de obtermos uma solução analítica seria em circunstâncias especiais, nos quais consideramos mais de uma força de acoplamento nula, de maneira similar ao feito por Strogatz em [6]. Como isso é um caso muito particular da dessincronização total, optamos por apresentar os resultados de forma numérica somente, cuja existência, unicidade e dependência contínua dos dados iniciais é garantida pelo Teorema 2.1.

## 3 Resultados Numéricos

Gostaríamos de enfatizar mais uma vez que as simulações numéricas que apresentaremos abaixo, para o modelo PIM, descrevem o comportamento de sincronização e dessincronização do ritmo circadiano e não o ritmo circadiano em si.

Além disso, as simulações para as soluções problemas totalmente dessincronizado (o qual não conhecemos uma solução analítica) foram feitas utilizando a rotina ODE45 do Matlab.

A tabela 1 trás os valores estimados aos parâmetros que foram usados nas simulação numéricas das situações de sincronização e dessincronização abordados na Seção 2. Tais valores são escolhidos de forma a satisfazerem as restrições de sincronização e dessincronização de Winfree [7] e não são fruto de um processo de calibração a partir de dados reais. A calibragem dos dados será fruto de trabalhos futuros. Os valores das condições iniciais foram fixados como  $f_2 = \frac{1}{3}$ ,  $F_2 = 6$ ,  $f_3 = \frac{1}{8}$  e  $F_3 = 2$  para todos os casos.

Parâmetro	Sinc. Total	Sinc. Parcial	Dessinc. Total
$\tau_1$	23,5	25,5	22
$\tau_2$	23,5	29	31,4
$\tau_3$	23,5	15	25,8
$A_1$	-	0,54	0,0911
$B_1$	-	-0,071	-0,084
$A_2$	-	-0,0131	-0,0072
$C_1$	-	0,0005	0,0034
$B_2$	-	-0,0098	-0,0026
$C_2$	-	0,0015	0,00019

Tabela 1: Parâmetros utilizados nas simulações, para a sincronização total, parcial e dessincronização.

Note que, de (4)-(6), as solução  $(\theta_1(t), \theta_2(t), \theta_3(t))^T$  independe das forças de acoplamento. Por isso, tais valores das constantes de acoplamento não aparecem na primeira coluna da tabela 1. Fica evidente ainda por (4)-(6) que a solução são retas paralelas. Assim, preferimos não apresentar os resultados numéricos para tal caso.

Para a simulação do caso sincronizado parcial utilizamos os valores dos parâmetros da segunda coluna da tabela 1. É importante notar que tais parâmetros satisfazem a restrição de sincronização parcial (18), e a solução analítica (7)-(9), disponível na seção 2.2. A figura 2 mostra como se comportam as soluções  $(\theta_1(t), \theta_2(t), \theta_3(t))^T$ .

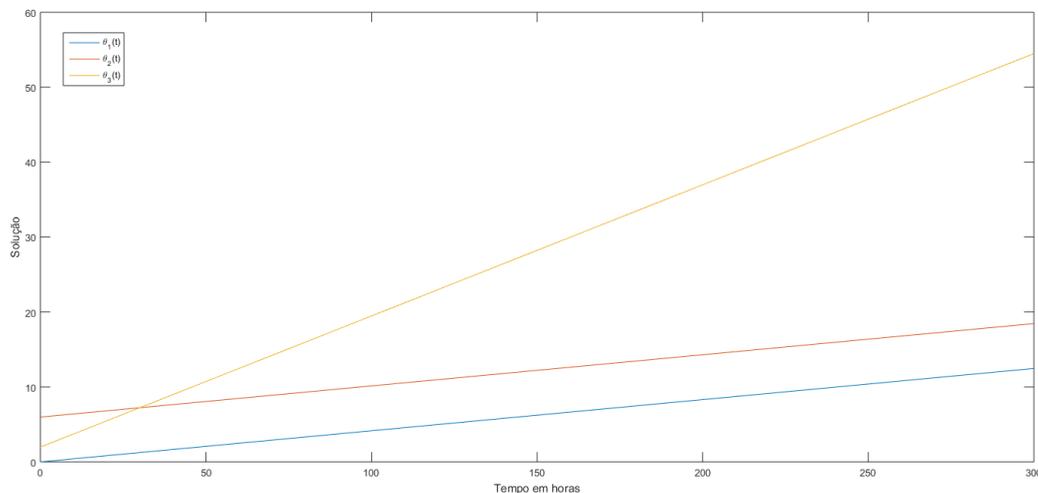


Figura 2: Solução do Modelo PIM com Sincronização Parcial

Podemos observar pela figura 2 que de fato os osciladores  $\theta_1(t)$  e  $\theta_2(t)$  estão sincronizados, uma vez que a diferença de fase entre tais osciladores é constante para qualquer  $t \in [0, 300]$ . Entretanto, a diferença de fase entre o oscilador  $\theta_3(t)$  e o oscilador  $\theta_1(t)$  (ou  $\theta_2(t)$ ) não é nula nem mesmo constante para  $t \in [0, 300]$ , descrevendo, portanto, a dessincronização entre o oscilador  $\theta_3(t)$  e os demais. Nesta situação os ciclos da temperatura corporal e da atividade-reposo estão sincronizados entre si, mas estão dessincronizados com o ciclo da dor, gerando assim um ritmo circadiano parcialmente sincronizado. Uma interpretação que segue deste caso é a seguinte: se os osciladores  $\theta_1(t)$  e  $\theta_2(t)$  estão fortemente acoplados em relação a  $\theta_3(t)$ , então embora a influência externa  $\theta_3(t)$  aumente significativamente, esta não afetará a sincronização entre  $\theta_1(t)$  e  $\theta_2(t)$ .

Por outro lado, para os valores dos parâmetros que estão dispostos na terceira coluna da tabela 1 temos a dessincronização total. A solução  $(\theta_1(t), \theta_2(t), \theta_3(t))^T$  para este caso é apresentada na figura 3.

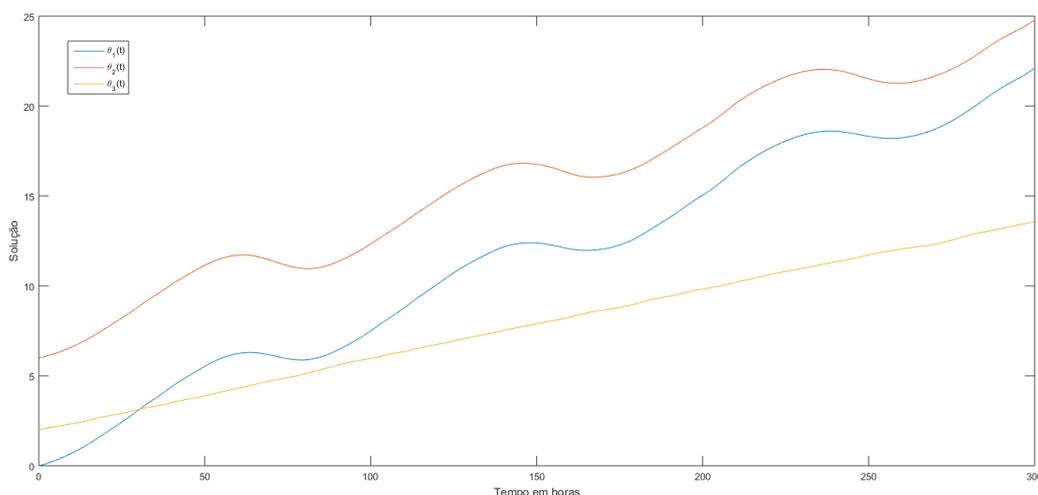


Figura 3: Solução do Modelo PIM com Dessincronização Total

Podemos observar no gráfico da figura 3 que a diferença de fase entre os osciladores (para

$t \in [0, 300]$ ), considerando-os dois a dois, não é constante nem nula, ou seja, o ritmo circadiano sob tais condições está totalmente dessincronizado. Em outras palavras, podemos interpretar tal resultado como: se os os ciclos da temperatura corporal e atividade-reposo não estão fortemente acoplados, a influência externa da dor age no sistema de forma a dessincronizar o ritmo.

## 4 Conclusões

Nesta contribuição propomos uma modelagem simplificada para descrever efeitos de sincronização para o ritmo circadiano. Utilizando como base os ciclos da temperatura corporal, atividade-reposo e a dor como uma influência externa, sob a hipótese de serem osciladores cíclicos não-linearmente acoplados. Denominamos tal modelo de Pain Influenced Model (PIM).

Provamos a boa colocação do modelo PIM. Além disso, dada a escolha da topologia de acoplamento, obtivemos resultados de sincronização e dessincronização (total e parcial) para o modelo PIM. Utilizamos para tal a relação entre as forças de acoplamento e as frequências intrínsecas, as quais podem ser vistos como uma extensão dos resultados de Strogatz em [6]. Para os casos de sincronizado total e sincronizado parcial, obtivemos soluções analíticas para o modelo PIM. Para o caso totalmente dessincronizado, resolvemos o PVI correspondente ao modelo PIM utilizando uma estratégia numérica, a qual mostrou satisfazer os resultados previstos pelas restrições entre as forças de acoplamento e as frequências intrínsecas.

Dos resultados acima podemos deduzir que, se  $\theta_1(t)$  e  $\theta_2(t)$  estão fortemente (fracamente) sincronizados, uma influência externa  $\theta_3(t)$  que está fracamente (fortemente) sincronizados não influência (influência) na sincronização dos osciladores  $\theta_1(t)$  e  $\theta_2(t)$ .

## Agradecimentos

Agradecemos a Universidade Federal do Rio Grande (FURG) e ao Programa de Pós Graduação em Modelagem Computacional (PPGMC) pelo espaço, e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo incentivo financeiro.

## Referências

- [1] S. Boccaletti, A. N. Pisarchik, C. I. Del Genio, A. Amann, “Synchronization: from coupled systems to complex networks”, Cambridge University Press, Cambridge, 2018.
- [2] J. J. Chiesa, “Patrones del ritmo de actividade motora como expresión de la plasticidade del Sistema Circadiano”, Tese de Doutorado, Universitat de Barcelona, 2005.
- [3] S. S. Glaeser, A. T. Santos, A. De Cezaro, C. M. S. Machado, D. F. Adamatti, Modeling of Circadian Rhythm under influence of Pain: an approach based on Multi-agent Simulation, *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, 7/2 (2018) 13 -25.
- [4] Y. Kuramoto, “Chemical Oscillations, Waves, and Turbulence”, Springer-Verlag, New York, 1984.
- [5] J. M. Sotomayor, “Lições de Equações Diferenciais Ordinárias”, Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1979.
- [6] S. H. Strogatz, “Lecture Notes in Biomathematics: The mathematical structure of the human sleep-wake cycle”, Springer-Verlag, 69 (1986), DOI: 10.1007/978-3-642-46589-5.
- [7] A. T. Winfree, Biological Rhythms and the Behavior of Populations of Coupled Oscillators, *Journal of Theoretical Biology*, 16 (1967) 15-42.



## Solução das Equações da Cinética Pontual de nêutrons pelo método de Fator de Integração Implícito

**Jaime C. Zandoná, Claudio Z. Pertesen, Marcelo Schramm**

Universidade Federal de Pelotas – Instituto de Física e Matemática

Programa de Pós-Graduação em Modelagem Matemática

96001-970, Capão do Leão, RS

E-mail:jaime.zandona@ufpel.edu.br, claudio.petersen@ufpel.edu.br, schrammmarcelo@gmail.com

**Resumo:** *O presente trabalho tem por objetivo resolver as equações da cinética pontual de nêutrons considerando 6 grupos de precursores de nêutrons atrasados e um termo fonte. Considerar termos fontes nas equações da cinética é algo tratado nos reatores do tipo Accelerator Driven System (ADS), na qual são sistemas acionados por fontes em estado subcrítico mantidas em estado estacionário por uma fonte externa de nêutrons. Para resolver esse problema utiliza-se o método de Fator de Integração Implícito (FII), que consiste em multiplicar a equação por um fator de integração, tratando a parte não homogênea de forma implícita, construindo um esquema que se baseia em aproximar a exponencial da matriz e o termo fonte através de uma interpolação polinomial por Lagrange. Verifica-se a metodologia realizando dois casos testes, um com fonte externa igual a zero, na qual os resultados são comparados com outros dois métodos presentes na literatura, e outro com fonte externa constante, na qual o modelo é testado para verificar sua aplicabilidade física.*

**Palavras-chave:** *Reator ADS, modelo da cinética pontual de nêutrons, método de Fator de Integração Implícito.*

### Introdução

A energia nuclear tem um espaço importante na matriz energética mundial devido à crescente demanda de energia e as inúmeras vantagens proporcionadas. Dentre as mais relevantes, vale destacar a sua aplicação na área da medicina e indústria com a produção de radioisótopos por reatores de pesquisa. Além disso, os reatores nucleares de potência, que geram energia elétrica, utilizam pouco material para produzir grandes quantidades de energia e atendem aos objetivos do Protocolo de Kyoto e, mais recentemente, o Acordo de Paris [8], por não emitir dióxido de carbono (CO<sub>2</sub>), provocando pouco ou quase nenhum impacto sobre a biosfera. Entretanto, a sustentabilidade ainda é o maior desafio no desenvolvimento de tecnologias nucleares inovadoras como opção para longo prazo.

Além disso, para suprir o aumento do consumo de energia elétrica, evitando impactos ambientais causados pelo emprego de combustíveis fósseis, é de suma importância promover estudos no campo da física de reatores nucleares. Uma das questões imprescindíveis é dispor de segurança e controle dos reatores. Para tanto, necessita-se compreender todos os parâmetros que envolvem as reações no núcleo do reator para então constituir um balanço entre as perdas e produções de nêutrons.

Dentro da teoria de difusão de nêutrons, as equações que governam a dinâmica temporal e espaço-temporal da população de nêutrons são as chamadas equações da cinética. As equações da cinética da teoria da difusão dividem-se em: equações da cinética pontual e equações da cinética espacial.

Neste trabalho dá-se ênfase ao modelo da cinética pontual, mais especificamente as variações do fluxo de nêutrons devido a escalas pequenas de tempo, ou seja, o afastamento da criticalidade devido a mudanças nos parâmetros nucleares para tempos pequenos.

As equações da cinética pontual estão interessadas exclusivamente na variação da amplitude do fluxo com o tempo, ou seja, assumem total separabilidade no tempo e no espaço, na qual a forma espacial do fluxo é conhecida o que torna essas equações exclusivamente dependentes do tempo. Esse modelo é válido para núcleos pequenos, onde a distribuição

espacial não é sensível às mudanças locais nas propriedades do reator. O modelo da cinética pontual ainda tem um papel relevante em física de reatores na medida em que pode ser utilizado, quando devidamente resolvido, para uma previsão de tempo quase real da potência do reator, o que permite um controle em tempo útil e de intervenção na planta, a fim de evitar a ocorrência de acidentes graves [6].

A rigidez é uma grande dificuldade em resolver numericamente as equações da cinética devido à grande diferença em magnitude entre os tempos de vida média dos nêutrons prontos e atrasados, o que resulta na exigência de incrementos muito pequenos no intervalo de tempo na resolução por métodos numéricos. Tem havido um grande esforço na investigação sobre a eliminação do problema da rigidez e formas alternativas para resolver as equações da cinética pontual.

Com o passar do tempo, novos modelos, métodos e parâmetros nucleares tem sido propostos para estimar a população de nêutrons dentro do núcleo de um reator nuclear de potência, por isso deve-se sempre realizar testes para verificar, tanto a capacidade do modelo, quanto ao do método a ser empregado de prever situações específicas, definindo limites de validade (qualificação do modelo e método). Esses testes são realizados utilizando-se problemas de referência (padrões) em física de reatores chamados de problemas *benchmark* e mais recentemente uma atenção especial tem sido dado ao *Accelerator Driven System* (ADS) [1], [3], [2].

Neste sentido, neste trabalho resolvem-se as equações da cinética pontual de nêutrons considerando 6 grupos de precursores de nêutrons atrasados considerando um termo fonte. Utiliza-se o método de Fator de Integração Implícito (**FI**), recentemente utilizado com sucesso no modelo geral de reação difusão [5]. A ideia é multiplicar a equação da cinética pontual de nêutrons por um fator de integração tratando a parte não homogênea de forma implícita, construindo um esquema que se baseia em aproximar a exponencial da matriz e o termo fonte através de uma interpolação polinomial por Lagrange.

Cabe ressaltar que, considerar termos fontes nas equações da cinética é algo relativamente novo, tratado nos reatores chamados ADS, na qual são sistemas acionados por fonte em estado subcrítico mantida em estado estacionário por uma fonte externa de nêutrons. Reatores do tipo ADS com núcleos subcríticos ainda é um assunto pouco explorado e precisa ser melhor entendido e investigado. Por isso, neste trabalho apresenta-se a metodologia completa para tratar as equações da cinética pontual de nêutrons admitindo uma fonte externa de nêutrons que possibilitem prever o comportamento da população de nêutrons com satisfatória precisão e confiabilidade.

## Metodologia

Partindo das equações da cinética pontual de nêutrons considerando seis grupos de nêutrons atrasados com fonte externa tem-se:

$$\begin{aligned} \frac{dn(t)}{dt} &= \frac{\rho(t) - \beta}{\Lambda} n(t) + \sum_{i=1}^6 \lambda_i C_i(t) + S(n(t)) \\ \frac{dC_i(t)}{dt} &= \frac{\beta_i}{\Lambda} n(t) - \lambda_i C_i(t) \quad , \end{aligned} \quad (1)$$

com as seguintes condições iniciais

$$\begin{aligned} n(0) &= n_0 \\ C_i(0) &= \frac{\beta_i}{\lambda_i \Lambda} n_0 \quad , \end{aligned} \quad (2)$$

para  $i=1:6$ , na qual,  $n(t)$  dado em  $[cm^{-3}]$  é a densidade de nêutrons no tempo  $t$ ,  $C_i(t)$  dado em  $[cm^{-3}]$  é a concentração de precursores de nêutrons atrasados para o grupo  $i$  de precursores

no tempo  $t$ ,  $\rho(t)$  é a reatividade no instante  $t$ ,  $\beta$  é a fração de nêutrons atrasados,  $\beta_i$  são as frações de nêutrons atrasados para o grupo  $i$  de precursores,  $\lambda_i$ , dado em  $[s^{-1}]$  são as constantes de decaimento para o grupo  $i$  de precursores,  $\Lambda$  dado em  $[s]$  é o tempo médio de geração entre o nascimento de nêutron e posterior absorção, e  $S(n(t))$  dado em  $[cm^{-3}]$ , é o termo fonte podendo representar uma fonte externa de nêutrons.

Reescrevendo a equação (1) na forma matricial, tem-se:

$$\frac{d\bar{y}(t)}{dt} = A(t)\bar{y}(t) + \bar{S}(y(t)) \quad , \quad (3)$$

onde

$$\bar{y}(t) = \begin{pmatrix} n(t) \\ c_1(t) \\ c_2(t) \\ c_3(t) \\ c_4(t) \\ c_5(t) \\ c_6(t) \end{pmatrix}, \quad A(t) = \begin{pmatrix} \frac{\rho(t)-\beta}{\Lambda} & \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & \lambda_5 & \lambda_6 \\ \frac{\beta_1}{\Lambda} & -\lambda_1 & 0 & 0 & 0 & 0 & 0 \\ \frac{\beta_2}{\Lambda} & 0 & -\lambda_2 & 0 & 0 & 0 & 0 \\ \frac{\beta_3}{\Lambda} & 0 & 0 & -\lambda_3 & 0 & 0 & 0 \\ \frac{\beta_4}{\Lambda} & 0 & 0 & 0 & -\lambda_4 & 0 & 0 \\ \frac{\beta_5}{\Lambda} & 0 & 0 & 0 & 0 & -\lambda_5 & 0 \\ \frac{\beta_6}{\Lambda} & 0 & 0 & 0 & 0 & 0 & -\lambda_6 \end{pmatrix} e \quad \bar{S}(y(t)) = \begin{pmatrix} S_0(n(t)) \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad (4)$$

Para resolver essa equação utiliza-se o Método de Fator de Integração Implícito (FII) [5]. Esse método é obtido a partir da multiplicação pelo fator de integração  $e^{-At}$ , na equação (3). Sem perda de generalidade, vamos considerar  $\rho(t) = \rho$ , logo a matriz  $A(t)=A$ . Assim, tem-se a seguinte equação:

$$e^{-At} \frac{d\bar{y}(t)}{dt} = e^{-At} A \bar{y}(t) + e^{-At} \bar{S}(y(t)) \quad . \quad (5)$$

Reescrevendo a equação (5) e utilizando a regra do produto, temos:

$$\frac{d(e^{-At} \bar{y}(t))}{dt} = e^{-At} \bar{S}(y(t)) \quad . \quad (6)$$

Integrando entre  $t_n$  e  $t_{n+1}$ , sendo  $t_{n+1} = t_n + \Delta t$ , em ambos os lados desta equação, vamos obter:

$$\int_{t_n}^{t_{n+1}} \frac{d(e^{-At} \bar{y}(t))}{dt} dt = \int_{t_n}^{t_{n+1}} e^{-At} \bar{S}(y(t)) dt \quad , \quad (7)$$

Fazendo uma mudança de variável,  $\tau = t - t_n$ , em (7), temos:

$$e^{-At} \bar{y}(t) \Big|_{t_n}^{t_{n+1}} = e^{-At_n} \int_0^{\Delta t} e^{-A\tau} \bar{S}(y(t_n + \tau)) d\tau \quad , \quad (8)$$

Aplicando os limites de integração em (8), ficamos com:

$$e^{-A(t_n+\Delta t)}\bar{y}(t_n+\Delta t)-e^{-At_n}\bar{y}(t_n)=e^{A\Delta t}\int_0^{\Delta t}e^{-A\tau}\bar{S}(y(t_n+\tau))d\tau \quad (9)$$

Multiplicando a equação (9) por  $e^{A(t_n+\Delta t)}$ , temos que:

$$\bar{y}(t_n+\Delta t)-e^{A\Delta t}\bar{y}(t_n)=e^{A\Delta t}\int_0^{\Delta t}e^{-A\tau}\bar{S}(y(t_n+\tau))d\tau \quad (10)$$

Assim:

$$\bar{y}_{n+1}=e^{A\Delta t}\bar{y}_n+e^{A\Delta t}\int_0^{\Delta t}e^{-A\tau}\bar{S}(y(t_n+\tau))d\tau \quad (11)$$

onde  $\bar{y}(t_n)=\bar{y}_n$  é utilizado, usualmente, para aproximação de  $\bar{y}(t_n)$ . Queremos tratar a parte não homogênea de forma implícita construindo um esquema que se baseia em aproximar  $e^{-A\tau}\bar{S}(y(t_n+\tau))$  através de uma interpolação polinomial que envolve  $t_{n+1}$ , definindo

$$G(\tau)=e^{-A\tau}\bar{S}(y(t_n+\tau)) \quad (12)$$

Para construir um esquema de integração de ordem  $r$  de erro de truncamento, aproximamos  $G(\tau)$  por um polinômio de Lagrange de ordem  $(r-1)$ , com pontos de interpolação  $t_{n+1}, t_n, \dots, t_{n+2-r}$ .

Na mudança de variáveis trabalhamos com os pontos de interpolação  $\Delta t, 0, -\Delta t, \dots, (2-r)\Delta t$ : então:

$$P(\tau)=\sum_{i=1}^r L_{r,i}(\tau)G((2-i)\Delta t)=\sum_{i=1}^{r-2} L_{r,i}(\tau)G((-i)\Delta t) \quad (13)$$

onde  $L_{r,i}$  são os polinômios de Lagrange, dados por:

$$L_{r,i}(\tau)=\prod_{\substack{j=-1 \\ j \neq i}}^{r-2} \frac{\tau-(-j\Delta t)}{(-i\Delta t-(-j\Delta t))}=\prod_{\substack{j=-1 \\ j \neq i}}^{r-2} \frac{\tau+j\Delta t}{(j-i)\Delta t} \quad (14)$$

Considerando em particular  $r=3$  temos que:

$$P(\tau)=\sum_{i=-1}^1 L_{3,i}(\tau)G((-i)\Delta t)=L_{3,-1}(\tau)G(\Delta t)+L_{3,0}(\tau)G(0)+L_{3,1}(\tau)G(-\Delta t) \quad (15)$$

onde

$$\left\{ \begin{array}{l} G(\Delta t)=e^{-A\Delta t}\bar{S}(y(\Delta t+t_n))=e^{-A\Delta t}\bar{S}(y_{n+1}) \\ G(0)=\bar{S}(y(t_n))=\bar{S}(y_n) \\ G(-\Delta t)=e^{A\Delta t}\bar{S}(y(t_n-\Delta t))=e^{A\Delta t}\bar{S}(y(t_{n-1}))=e^{A\Delta t}\bar{S}(y_{n-1}) \end{array} \right\} \quad (16)$$

Assim, para o esquema de ordem  $r$ , temos:

$$P(\tau)=\sum_{i=-1}^{r-2} e^{iA\Delta t}\bar{S}(y_{n-i})L_{r,i}(\tau) \quad (17)$$

ou

$$P(\tau)=\sum_{i=-1}^{r-2} e^{iA\Delta t}\bar{S}(y_{n-i})\prod_{\substack{j=-1 \\ j \neq i}}^{r-2} \frac{\tau+j\Delta t}{(j-i)\Delta t} \quad ; 0 \leq \tau \leq \Delta t \quad (18)$$

Portanto, da equação (11), tem-se que:

$$\bar{y}_{n+1}=e^{A\Delta t}\bar{y}_n+e^{A\Delta t}\int_0^{\Delta t}\sum_{i=-1}^{r-2} e^{iA\Delta t}\bar{S}(y_{n-i})\prod_{\substack{j=-1 \\ j \neq i}}^{r-2} \frac{\tau+j\Delta t}{(j-i)\Delta t}d\tau \quad ; 0 \leq \tau \leq \Delta t \quad (19)$$

Desta forma, obtemos o esquema do método **FII** geral como:

$$\vec{y}_{n+1} = e^{A\Delta t} \vec{y}_n + \Delta t \left( \alpha_{n+1} \vec{S}(y_{n+1}) + \sum_{i=0}^{r-2} \alpha_{n-i} \vec{S}(y_{n-i}) \right), \quad (20)$$

onde:

$$\alpha_{n-i} = \frac{e^{(i+1)A\Delta t}}{\Delta t} \int_0^{\Delta t} \prod_{\substack{k=1 \\ k \neq i}}^{r-2} \frac{\tau + k\Delta t}{(k-i)\Delta t} d\tau \quad ; \quad -1 \leq i \leq r-2 \quad . \quad (21)$$

Assim, utilizando o método de Lagrange de segunda ordem, ou seja,  $r=3$ , o que resulta em

$\alpha_{n+1} = \frac{1}{2}$  e  $\alpha_n = \frac{e^{A\Delta t}}{2}$ , ficamos com:

$$\vec{y}_{n+1} = e^{A\Delta t} \underbrace{\left( \vec{y}_n + \frac{\Delta t}{2} \vec{S}(y_n) \right)}_{\text{Termo Explícito}} + \underbrace{\frac{\Delta t}{2} \vec{S}(y_{n+1})}_{\text{Termo Implícito}} \quad . \quad (22)$$

Note que, para resolver a exponencial da matriz vamos utilizar neste trabalho a aproximação pela decomposição espectral da matriz  $A$ , uma vez que a matriz  $A$  é diagonalizável, então existe uma matriz  $P$  dos autovetores inversível, tal que:

$$A = P \Lambda P^{-1} \quad , \quad (23)$$

onde  $\Lambda$  é uma matriz diagonal dos autovalores da matriz  $A$ ,  $P$  é a matriz associada dos autovetores e  $P^{-1}$  é a matriz inversa dos autovetores da matriz  $A$ , respectivamente. Logo, em vez de determinar  $e^{A\Delta t}$ , avaliamos  $P e^{A\Delta t} P^{-1}$ .

## Resultados

Neste trabalho para fins de verificação da metodologia foram realizados dois casos testes. No primeiro foram considerados valores constantes da reatividade para quatro tempos distintos. Dois deles subcríticos ( $\rho = -1\beta$  e  $\rho = -0,5\beta$ ), ou seja, neste caso, o número de fissões e a população de nêutrons de uma geração é menor que a geração anterior e o número de nêutrons diminuem com o passar do tempo; e outros dois casos supercríticos (subpronto-crítico:  $\rho = 0,5\beta$  e pronto-crítico:  $\rho = 1\beta$ ), ou seja, neste caso, o número de fissões e a população de nêutrons de uma geração é maior que a geração anterior e o número de nêutrons aumenta com o passar do tempo. Já para o segundo caso teste foi considerado uma reatividade constante ( $\rho = -0,5\beta$ ) e duas fontes externas constantes: uma considerando  $S = 20\text{cm}^{-3}$  ligada em 20 s e outra  $S = 30\text{cm}^{-3}$  ligada em 40 s. Os resultados foram gerados até 140 segundos e a densidade de nêutrons é apresentada em forma gráfica.

Para ambos casos testes foram utilizados os parâmetros cinéticos estudados em [7]. Considera-se a densidade de nêutrons inicial como  $n_0 = 1\text{cm}^{-3}$ . O software utilizado para implementação do algoritmo foi o Scilab 6.0.1 juntamente com um computador que possui essas configurações: Intel(R) Core(TM) i5-7200U 2.5GHz, 8GB de RAM, Sistema operacional de 64 bits e processador com base em x64.

Para o primeiro caso teste os resultados obtidos pelo método de Fator de Integração Implícito (**FII**), proposto neste trabalho, são comparados com dois métodos: o Método de Aproximação Polinomial (PAM) presente em [7] e o Método Backward Euler Finite Difference (BEFD) presente em [4]. Os resultados para esse caso teste são apresentados na Tabela 1.

Reatividade	t (s)	FII	PAM $\Delta t = 0,0001s$	BEFD
$\rho = -1 \beta$	0,1	<b>0,5205642866</b>	0,520564066	0,5205642866
	1	<b>0,4333334453</b>	0,433333445	0,4333334453
	10	<b>0,2361106508</b>	0,236110651	0,2361106508
	100	<b>0,0286676425</b>	0,028667642	0,0286676424
$\rho = -0,5 \beta$	0,1	<b>0,6989252256</b>	0,698925094	0,6989252256
	1	<b>0,6070535656</b>	0,607053566	0,6070535656
	10	<b>0,3960776907</b>	0,396077691	0,3960776907
	100	<b>0,0715828544</b>	0,396077691	0,07158285444
$\rho = 0,5 \beta$	0,1	<b>1,5331126460</b>	1,533112714	1,533112646
	1	<b>1,5331126460</b>	2,511494293	2,511494291
	10	<b>14,215025242</b>	14,21502524	14,21502524
	100	<b>80061435,62</b>	80061435,76	80061435,62
$\rho = 1 \beta$	0,1	<b>2,5157661414</b>	2,515766171	2,515766141
	0,5	<b>10,362533810</b>	10,362534	10,36253381
	1	<b>32,183540945</b>	32,1835419	32,18354095
	10	<b>3246978898</b>	3246979823	3246978898
	100	<b>2,59648464E+89</b>	2,59649204E+89	2,596484646E+89

Tabela 1: Densidade de nêutrons em  $cm^{-3}$  com inserção de reatividade constante.

Podemos observar que, para todos os tempos e reatividades consideradas o método **FII** concorda em todas as casas decimais com BEFD, que é considerado um método analítico de solução exata. Já quando comparado com o PAM, temos uma precisão de 6 a 8 casas para todos os tempos e reatividades consideradas.

Para o segundo caso teste os resultados são apresentados na figura 2.

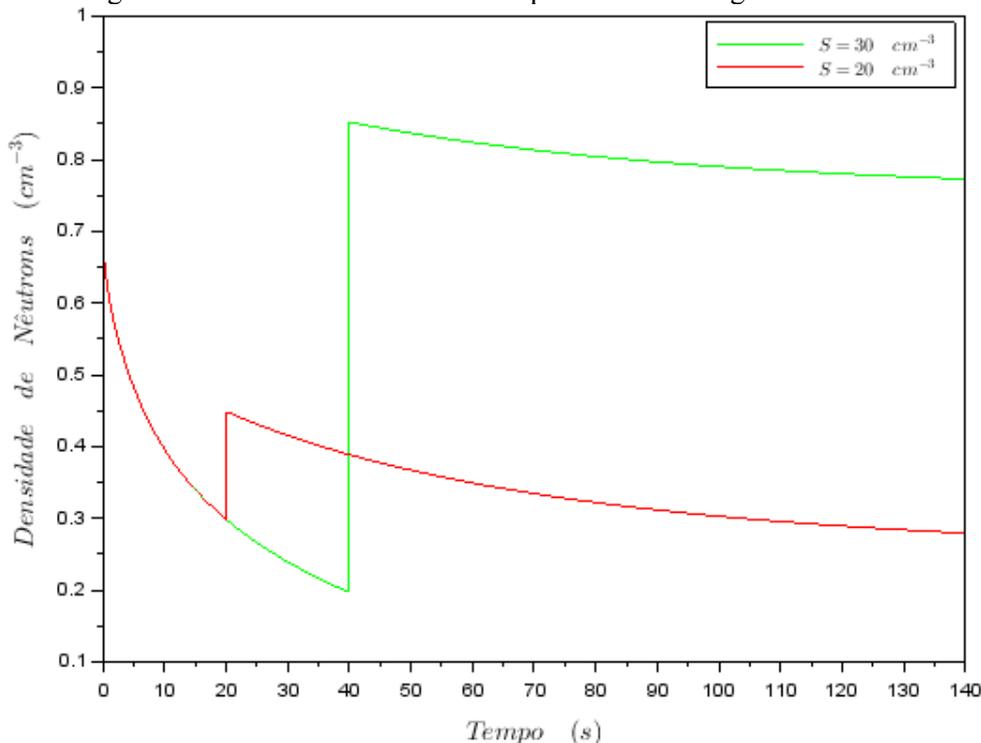


Figura 1: Densidade de nêutrons para  $(\rho = -0,5\beta)$  e duas fontes externas constantes ligadas  $S = 20cm^{-3}$  em 20 segundos e  $S = 30cm^{-3}$  em 40 segundos.

Pode-se observar que o ligamento de fontes externas em diferentes tempos faz a densidade de nêutrons aumentar dando um salto de densidade, estabilizando para um valor constante acima de zero. O aumento da fonte faz aumentar a amplitude desse aumento, fazendo a densidade estabilizar num valor maior.

## Conclusão

Neste trabalho foi apresentada uma solução das equações da cinética pontual de nêutrons, considerando reatividade constante, 6 grupos de energia e fonte externa constante. Os resultados apresentados mostraram que o método de Fator de Integração Implícito (FII) associado com a decomposição espectral da matriz através dos autovalores e autovetores da matriz associada apresentou excelentes resultados quando comparados com a literatura, principalmente com o método BEFD considerado analítico. Os testes realizados com fontes externas foram plausíveis com a realidade física, mostrando o papel das fontes externas para manter a densidade de nêutrons estabilizada com o passar do tempo. Cabe salientar que esses reatores ADS são considerados reatores mais seguros, pois trabalham em estado subcrítico, na qual a população de nêutrons diminui com o passar do tempo, necessitando de uma fonte externa para manter a população de nêutrons constante. Isso significa que na ausência de fontes externas o reator se desliga, uma vez que a população de nêutrons naturalmente reduz a zero com o passar do tempo. Neste sentido, estudos com fontes externas são importantes para simular quais tipos de fontes manteriam a densidade de nêutrons constante num patamar seguro de operação. Cabe ressaltar que esse assunto é relativamente novo devendo ser melhor investigado. Como perspectivas futuras pretendemos estudar melhor quais tipos de fontes externas podem ser aplicadas mais realisticamente no modelo da cinética pontual de nêutrons para prever a população de nêutrons com certa previsibilidade e eficácia. Além disso, pretendemos estender essa metodologia para modelos mais realísticos como difusão e cinética espacial de nêutrons que vão ampliar o estudo e análise em física de reatores.

## Referências

- [1] B. M. Santos, R. G. Cabral, M. Gonçalves, and S. B. Duarte. "Neutron Generation and Diffusion Process in Reactors Stimulated by High Energy Proton Beam — The Accelerator Driven System", American Institute of Physics, 2011, doi: 10.1063/1.3608996.
- [2] B. Merk and V. Glivici-Cotruta, "Solutions without Space-Time Separation for ADS Experiments: Overview on Developments and Applications", Science and Technology of Nuclear Installations, Article ID 140946, 11 pages, 2012.
- [3] B. Merk, F.P. Weiß, A two group analytical approximation solution for an external source problem without separation of space and time. *Annals of Nuclear Energy*, 37 (2010) 942–952.
- [4] B. D. Ganapol, A highly accurate algorithm for the solution of the point kinetics equations. *Annals of Nuclear Energy*, 62 (2013) 564-571.
- [5] E. F. Y. Medina. "Método do Fator de Integração Implícito para Problemas de Reação-Difusão", Dissertação de Mestrado, PPGMC-UFJF, 2016.
- [6] C. Z. Petersen, "Solução Analítica das Equações da Cinética Pontual e Espacial da Teoria de Difusão de Nêutrons pelas Técnicas da GITT e Decomposição", Tese de Doutorado, PROMEC-UFRGS, 2011.
- [7] F. Tumelero, Polynomial approach method to solve the neutron point kinetics equations with use of the analytic continuation. *Kerntechnik*, 81 (2016) 1-29.
- [8] UNITED NATIONS. Paris Agreement. 2015. Disponível em: <[http://unfccc.int/files/essential\\_background/convention/application/pdf/english\\_paris\\_agreement.pdf](http://unfccc.int/files/essential_background/convention/application/pdf/english_paris_agreement.pdf)>, Acesso em: 23 jul. 2016.



## ERMAC 2019

### Modelagem das Concentrações das Espécies Químicas Existentes na Combustão do Metanol Via Mecanismo Reduzido

**Guilherme Janecke Weymar, Letícia Machado Kaufmann,**

Depto de Matemática e Estatística, PPGMMat, UFPel  
96160-000, Capão do Leão, RS

E-mail: guilhermejanecke@gmail.com, leticiamachadokaufmann@hotmail.com,

**Régis Sperotto de Quadros**

Depto de Matemática e Estatística, PPGMMat, UFPel  
96160-000, Capão do Leão, RS

E-mail: quadros99@gmail.com

**Resumo:** *As emissões veiculares tem uma contribuição significativa na poluição atmosférica. Assim, muitos estudos estão sendo desenvolvidos com a finalidade de reduzir a emissão dos gases de efeito estufa. Uma alternativa é a utilização do biocombustível metanol. Este possui uma cadeia bem detalhada, com grande número de reações químicas. Para modelar as concentrações das espécies químicas existentes na combustão do metanol, este trabalho, apresenta uma estratégia para se obter mecanismos cinéticos reduzidos aplicando as hipóteses de equilíbrio parcial e de estado estacionário. O mecanismo reduzido obtido nesse processo, será verificado utilizando uma análise assintótica.*

## 1 Introdução

Os biocombustíveis são originários de biomassa renovável que podem ou não substituir os combustíveis oriundos de petróleo e gás natural. A biomassa, tida como qualquer material orgânico que pode ser utilizado para algum tipo de produção de energia, é uma alternativa econômica e ambiental para reduzir a queima dos combustíveis fósseis. O emprego de metanol, etanol e outros combustíveis derivados de biomassa, vem encontrando um mercado crescente como consequência de políticas econômicas ou de esforços em prol da redução da poluição atmosférica, causada por emissões veiculares. Uma ação importante e conhecida para essa redução de poluição é o Protocolo de Kyoto que tem como objetivo firmar acordos e discussões internacionais para conjuntamente estabelecer metas de redução na emissão de gases do efeito estufa na atmosfera, principalmente por parte dos países industrializados, além de criar formas de desenvolvimento de maneira menos impactante àqueles países em pleno desenvolvimento (FREITAS, 2018).

São muitas as vantagens de se usar biocombustíveis, tais como: redução da poluição com a sua queima e processamento; podem ser cultivados e, assim sendo renováveis; geram empregos em sua cadeia produtiva; minimizam a dependência em relação aos combustíveis fósseis; além de elevar os índices de exportações do país, favorecendo a balança comercial. No entanto, existem algumas desvantagens, como a necessidade de amplas áreas agricultáveis, podendo intensificar o desmatamento pela expansão da fronteira agrícola; pressão sobre o preço dos alimentos, que podem ter sua produção diminuída para dar lugar à produção de biomassa; entre outros fatores.

O metanol ou álcool metílico é um dos principais compostos do grupo orgânico dos alcoóis. Sua fórmula molecular é  $CH_3OH$ , ele possui ponto de fusão igual a  $-97\text{ }^\circ\text{C}$  e ponto de ebulição de  $64,7\text{ }^\circ\text{C}$ . Durante muitos anos, essa substância foi obtida unicamente pela destilação da madeira a seco e na ausência de ar, o que tornou o metanol conhecido como álcool de madeira. Atualmente a substância é obtida sinteticamente a partir do monóxido de carbono (processo carbocímico) ou por oxidação controlada do metano (processo petroquímico). Sua solubilidade em água é infinita, pois, por possuir moléculas pequenas, como o grupo  $OH$ , grupo característico

dos alcoóis, suas moléculas formam ligações de hidrogênio com a água (JOVELINO, 2015).

Espécies como o metanol possuem cadeias bem detalhadas, com grande número de reações e de variáveis. Esses modelos produzidos por cinéticas químicas das reações, quando feitas com mecanismos muito detalhados têm difícil resolução, por conta da existência de radicais altamente reativos, o que leva a rigidez do sistema de equações. Logo, se faz necessário desenvolver mecanismos reduzidos com menor número de variáveis e com rigidez moderada, mantendo boa precisão e o comportamento do mecanismo (BJUREL,1970).

## 2 Metodologia

A estratégia usada para se obter a redução do mecanismo é dividida em quatro etapas, que são apresentadas a seguir:

### 2.1 Taxa de reação

Para determinar a velocidade (taxa de reação)  $k$  de cada reação que acontece no mecanismo do metanol, utiliza-se a seguinte relação:

$$k_k = AT^\beta \exp\left(\frac{-E_A}{RT}\right) \quad (1)$$

onde  $A$  é o fator de frequência,  $T$  a temperatura,  $\beta$  o expoente de temperatura,  $E_A$  a energia de ativação e  $R$  a constante universal dos gases (PETERS, 1983).

### 2.2 Cadeia principal

Define-se a cadeia principal através do cálculos das velocidades realizado no item anterior gerado pelo mecanismo completo do metanol, com 129 reações e 31 espécies (MARINOV, 1999), na qual obtem-se a Tabela 1.

### 2.3 Hipóteses de estado estacionário e de equilíbrio parcial

Em um sistema homogêneo, a hipótese de estado estacionário é válida para as espécies intermediárias que são produzidas por reações lentas e consumidas por reações rápidas, o que faz com que suas concentrações permaneçam pequenas (TURNS, 2000). A hipótese do equilíbrio parcial é justificada quando as velocidades das reações de ida e de retorno são muito maiores do que as outras velocidades específicas do mecanismo (PETERS, 1988).

### 2.4 Análise assintótica

A análise assintótica consiste em assumir o estado estacionário para determinadas espécies, obtendo equações algébricas entre as taxas de reações. Desse modo, o mecanismo reduzido será determinado através da estequiometria dessas reações (ANDREIS, 2011).

Consideram-se as reações descritas na Tabela 1, apresentada por PETERS (1993), o sub-mecanismo para a combustão do metanol, que é também composta pelos submecanismos do hidrogênio, do monóxido de carbono e do metano.

Tabela 1: Mecanismo para chamas do metanol (unidades são  $mol$ ,  $cm^3$ ,  $s$ ,  $K$ , e  $cal/mol$ ), (PETERS, 1993)

Reação	A	$\beta$	$E_A$
1f. $O_2 + H \rightarrow OH + O$	2.000E+14	0.00	7.03E+01
1b. $OH + O \rightarrow O_2 + H$	1.568E+13	0.00	3.52E+00
2f. $H_2 + O \rightarrow OH + H$	5.060E+04	2.67	2.63E+01
2b. $OH + H \rightarrow H_2 + O$	2.222E+04	2.67	1.83E+01
3f. $H_2 + OH \rightarrow H_2O + H$	1.000E+08	1.60	1.38E+01
3b. $H_2O + H \rightarrow H_2 + OH$	4.312E+08	1.60	7.65E+01
4f. $OH + OH \rightarrow H_2O + O$	1.500E+09	1.14	4.20E-01
4b. $H_2O + O \rightarrow OH + OH$	1.473E+10	1.14	7.10E+01
5f. $O_2 + H + M \rightarrow HO_2 + M$	2.300E+18	-0.80	0.00E+00
5b. $HO_2 + M \rightarrow O_2 + H + M$	3.190E+18	-0.80	1.95E+02
6. $HO_2 + H \rightarrow OH + OH$	1.500E+14	0.00	4.20E+00
7. $HO_2 + H \rightarrow H_2 + O_2$	2.500E+13	0.00	2.90E+00
8. $HO_2 + OH \rightarrow H_2O + O_2$	6.000E+13	0.00	0.00E+00
9. $HO_2 + H \rightarrow H_2O + O$	3.000E+13	0.00	7.20E+00
10. $HO_2 + O \rightarrow OH + O_2$	1.800E+13	0.00	-1.70E+00
11. $HO_2 + HO_2 \rightarrow H_2O_2 + O_2$	2.500E+11	0.00	-5.20E+00
12f. $OH + OH + M \rightarrow H_2O_2 + M$	3.250E+22	-2.00	0.00E+00
12b. $H_2O_2 = M \rightarrow OH + OH + M$	1.692E+24	-2.00	2.02E+02
13. $H_2O_2 + H \rightarrow H_2O + OH$	1.000E+13	0.00	1.50E+01
14f. $H_2O_2 + OH \rightarrow H_2O + HO_2$	5.400E+12	0.00	4.20E+00
14b. $H_2O + HO_2 \rightarrow H_2O_2 + OH$	1.802E+13	0.00	1.35E+02
15. $H + H + M \rightarrow H_2 + M$	1.800E+18	-1.00	0.00E+00
16. $OH + H + M \rightarrow H_2O + M$	2.200E+22	-2.00	0.00E+00
17. $O + O + M \rightarrow O_2 + M$	2.900E+17	-1.00	0.00E+00
18f. $CO + OH \rightarrow CO_2 + H$	4.400E+06	1.50	-3.10E+00
18b. $CO_2 + H \rightarrow CO + OH$	4.956E+08	1.50	9.00E+01
19. $CH + O_2 \rightarrow CHO + O$	3.000E+13	0.00	0.00E+00
20. $CO_2 + CH \rightarrow CHO + CO$	3.400E+12	0.00	2.90E+00
21. $CHO + H \rightarrow CO + H_2$	2.000E+14	0.00	0.00E+00
22. $CHO + OH \rightarrow CO + H_2O$	1.000E+14	0.00	0.00E+00
23. $CHO + O_2 \rightarrow CO + HO_2$	3.000E+12	0.00	0.00E+00
24f. $CHO + M \rightarrow CO + H + M$	7.100E+14	0.00	7.03E+01
24b. $CO + H + M \rightarrow CHO + M$	1.136E+15	0.00	9.97E+00
25f. $CH_2 + H \rightarrow CH + H_2$	8.400E+09	1.50	1.40E+00
25b. $CH + H_2 \rightarrow CH_2 + H$	5.830E+09	1.50	1.30E+01
26. $CH_2 + O \rightarrow CO + H + H$	8.000E+13	0.00	0.00E+00
27. $CH_2 + O_2 \rightarrow CO + OH + H$	6.500E+12	0.00	6.30E+00
28. $CH_2 + O_2 \rightarrow CO_2 + H + H$	6.500E+12	0.00	6.30E+00
29. $CH_2O + H \rightarrow CHO + H_2$	2.500E+13	0.00	1.67E+01
30. $CH_2O + O \rightarrow CHO + OH$	3.500E+13	0.00	1.46E+01
31. $CH_2O + OH \rightarrow CHO + H_2O$	3.000E+13	0.00	5.00E+00
32. $CH_2 + M \rightarrow CHO + H + M$	1.400E+17	0.00	3.20E+02
33f. $CH_3 + H \rightarrow CH_2 + H_2$	1.800E+14	0.00	6.30E+01
33b. $CH_2 + H_2 \rightarrow CH_3 + H$	3.680E+13	0.00	4.43E+01
35. $CH_3 + O \rightarrow CH_2O + H$	7.000E+13	0.00	0.00E+00

Reação	A	$\beta$	$E_A$
37. $\text{CH}_3 + \text{O}_2 \rightarrow \text{CH}_2\text{O} + \text{OH}$	3.400E+11	0.00	3.74E+01
83. $\text{CH}_2\text{OH} + \text{H} = \text{CH}_2\text{O} + \text{H}_2$	3.000E+13	0.00	0.00E+00
84. $\text{CH}_2\text{OH} + \text{O}_2 = \text{CH}_2\text{O} + \text{HO}_2$	1.000E+13	0.00	3.00E+01
85. $\text{CH}_2\text{OH} + \text{M} = \text{CH}_2\text{O} + \text{H}$	5.000E+13	0.00	1.05E+02
86. $\text{CH}_3\text{OH} + \text{M} = \text{CH}_3 + \text{OH} + \text{M}$	3.160E+18	0.00	3.36E+02
87. $\text{CH}_3\text{OH} + \text{H} = \text{CH}_2\text{OH} + \text{H}_2$	4.000E+13	0.00	2.55E+01
88. $\text{CH}_3\text{OH} + \text{O} = \text{CH}_2\text{OH} + \text{OH}$	1.000E+13	0.00	1.96E+01
89. $\text{CH}_3\text{OH} + \text{OH} = \text{CH}_2\text{OH} + \text{H}_2\text{O}$	1.000E+13	0.00	7.10E+00
90f. $\text{CH}_3\text{OH} + \text{HO}_2 \rightarrow \text{CH}_2\text{OH} + \text{H}_2\text{O}_2$	0.620E+13	0.00	8.11E+01
90b. $\text{CH}_2\text{OH} + \text{H}_2\text{O}_2 \rightarrow \text{CH}_3\text{OH} + \text{HO}_2$	0.100E+08	1.70	4.79E+01

As espécies envolvidas neste mecanismo, são:  $\text{OH}, \text{O}, \text{HO}_2, \text{H}_2\text{O}_2, \text{CH}, \text{HCO}, \text{CH}_2, \text{CH}_3, \text{CH}_2\text{OH}, \text{O}_2, \text{H}, \text{H}_2, \text{H}_2\text{O}, \text{CO}, \text{CO}_2, \text{CH}_2\text{O}, \text{CH}_3\text{OH}$ . Dessa maneira, o sistema de equações diferenciais ordinárias (EDO'S) será dado por:

$$L(\text{OH}) = \omega_1 + \omega_2 - \omega_3 - \omega_4 + 2\omega_6 - \omega_8 + \omega_{10} - 2\omega_{12} + \omega_{13} - \omega_{14} - \omega_{16} - \omega_{18} - \omega_{22} + \omega_{27} + \omega_{30} - \omega_{31} + \omega_{37} + \omega_{86} + \omega_{88} - \omega_{89}. \quad (2)$$

$$L(\text{O}) = \omega_1 - \omega_2 + \omega_4 + \omega_9 - \omega_{10} - 2\omega_{17} + \omega_{19} - \omega_{26} - \omega_{30} - \omega_{35} - \omega_{88}. \quad (3)$$

$$L(\text{HO}_2) = \omega_5 - \omega_6 - \omega_7 - \omega_8 - \omega_9 - \omega_{10} - 2\omega_{11} + \omega_{14} + \omega_{84} - \omega_{90} \quad (4)$$

$$L(\text{H}_2\text{O}_2) = \omega_{11} + \omega_{12} - \omega_{13} - \omega_{14} + \omega_{90} \quad (5)$$

$$L(\text{CH}) = -\omega_{19} - \omega_{20} + \omega_{25} \quad (6)$$

$$L(\text{CHO}) = -\omega_{21} - \omega_{22} - \omega_{23} + \omega_{29} + \omega_{30} + \omega_{31} + \omega_{32} \quad (7)$$

$$L(\text{CH}_2) = -\omega_{26} - \omega_{27} - \omega_{28} \quad (8)$$

$$L(\text{CH}_3) = \omega_{33} - \omega_{37} + \omega_{86} \quad (9)$$

$$L(\text{CH}_2\text{OH}) = -\omega_{83} - \omega_{84} - \omega_{85} + \omega_{87} + \omega_{88} + \omega_{89} \quad (10)$$

$$L(\text{O}_2) = -\omega_1 - \omega_5 + \omega_8 + \omega_{10} + \omega_{11} + \omega_{17} - \omega_{23} - \omega_{27} - \omega_{28} - \omega_{37} - \omega_{84} \quad (11)$$

$$L(\text{H}) = -\omega_1 + \omega_2 + \omega_3 - \omega_5 - \omega_6 - \omega_7 - \omega_9 - \omega_{13} - 2\omega_{15} - \omega_{16} + \omega_{18} - \omega_{21} + \omega_{24} - \omega_{25} + 2\omega_{26} + \omega_{27} + 2\omega_{28} - \omega_{29} - \omega_{33} + \omega_{35} - \omega_{83} + \omega_{85} - \omega_{87} \quad (12)$$

$$L(\text{H}_2) = -\omega_2 - \omega_3 + \omega_7 + \omega_{15} + \omega_{21} + \omega_{25} + \omega_{29} + \omega_{33} + \omega_{83} + \omega_{87} \quad (13)$$

$$L(\text{H}_2\text{O}) = \omega_3 + \omega_4 + \omega_8 + \omega_9 + \omega_{10} + \omega_{14} + \omega_{16} + \omega_{22} + \omega_{31} + \omega_{89} \quad (14)$$

$$L(\text{CO}) = -\omega_{18} + \omega_{20} + \omega_{21} + \omega_{22} + \omega_{23} + \omega_{24} + \omega_{26} + \omega_{27} \quad (15)$$

$$L(\text{CO}_2) = -\omega_{20} + \omega_{28} \quad (16)$$

$$L(\text{CH}_2\text{O}) = -\omega_{29} - \omega_{30} - \omega_{31} - \omega_{32} + \omega_{35} + \omega_{37} + \omega_{83} + \omega_{84} + \omega_{85} \quad (17)$$

$$L(\text{CH}_3\text{OH}) = -\omega_{86} - \omega_{87} - \omega_{88} - \omega_{89} - \omega_{90} \quad (18)$$

$L(i)$  é operador diferencial aplicado a concentração da espécie  $i$  e  $\omega_k$  representa a taxa da reação  $k$ , sendo  $\omega_i = ki[A][B]$ . Considera-se o sinal positivo para as espécies do lado direito da

reação (produto) e o sinal negativo para as espécies do lado esquerdo da reação (reagente).

Sendo  $\omega_{if}$  e  $\omega_{ib}$ , taxas das reações de ida e de volta, respectivamente. Usa-se:

$$\begin{aligned}
 \omega_1 &= \omega_{1f} - \omega_{1b} \\
 \omega_2 &= \omega_{2f} - \omega_{2b} \\
 \omega_3 &= \omega_{3f} - \omega_{3b} \\
 \omega_4 &= \omega_{4f} - \omega_{4b} \\
 \omega_5 &= \omega_{5f} - \omega_{5b} \\
 \omega_{12} &= \omega_{12f} - \omega_{12b} \\
 \omega_{14} &= \omega_{14f} - \omega_{14b} \\
 \omega_{18} &= \omega_{18f} - \omega_{18b} \\
 \omega_{24} &= \omega_{24f} - \omega_{24b} \\
 \omega_{25} &= \omega_{25f} - \omega_{25b} \\
 \omega_{33} &= \omega_{33f} - \omega_{33b} \\
 \omega_{89} &= \omega_{89f} - \omega_{89b}
 \end{aligned} \tag{19}$$

### 3 Resultados

Utiliza-se a hipótese de estado estacionário para as espécies:  $OH, O, HO_2, H_2O_2, CH, CHO, CH_2, CH_3, CH_2OH$ , desta forma, os seus operadores diferenciais ( $L_{(i)}$ ) para essas espécies serão iguais a zero, obtendo um conjunto de nove equações algébricas, representadas a seguir:

$$\begin{aligned}
 0 = L(OH) &= \omega_1 + \omega_2 - \omega_3 - \omega_4 + 2\omega_6 - \omega_8 + \omega_{10} - 2\omega_{12} + \omega_{13} - \omega_{14} - \omega_{16} - \omega_{18} \\
 &\quad - \omega_{22} + \omega_{27} + \omega_{30} - \omega_{31} + \omega_{37} + \omega_{86} + \omega_{88} - \omega_{89}
 \end{aligned} \tag{20}$$

$$0 = L(O) = \omega_1 - \omega_2 + \omega_4 + \omega_9 - \omega_{10} - 2\omega_{17} + \omega_{19} - \omega_{26} - \omega_{30} - \omega_{35} - \omega_{88} \tag{21}$$

$$0 = L(HO_2) = \omega_5 - \omega_6 - \omega_7 - \omega_8 - \omega_9 - \omega_{10} - 2\omega_{11} + \omega_{14} + \omega_{84} - \omega_{90} \tag{22}$$

$$0 = L(H_2O_2) = \omega_{11} + \omega_{12} - \omega_{13} - \omega_{14} + \omega_{90} \tag{23}$$

$$0 = L(CH) = -\omega_{19} - \omega_{20} + \omega_{25} \tag{24}$$

$$0 = L(CHO) = -\omega_{21} - \omega_{22} - \omega_{23} + \omega_{29} + \omega_{30} + \omega_{31} + \omega_{32} \tag{25}$$

$$0 = L(CH_2) = -\omega_{26} - \omega_{27} - \omega_{28} \tag{26}$$

$$0 = L(CH_3) = \omega_{33} - \omega_{37} + \omega_{86} \tag{27}$$

$$0 = L(CH_2OH) = -\omega_{83} - \omega_{84} - \omega_{85} + \omega_{87} + \omega_{88} + \omega_{89} \tag{28}$$

Isolando  $\omega_{10}, \omega_{27}, \omega_{25}, \omega_{21}, \omega_{22}, \omega_{30}$  e  $\omega_{12}$  nas equações (22), (26), (24), (25), (20), (21) e (26), respectivamente, tem-se o seguinte conjunto de equações algébricas:

$$\omega_{10} = \omega_5 - \omega_6 - \omega_7 - \omega_8 - \omega_9 - 2\omega_{11} + \omega_{14} + \omega_{23} + \omega_{84} - \omega_{90} \tag{29}$$

$$\omega_{27} = -\omega_{25} - \omega_{26} - \omega_{28} + \omega_{33} \tag{30}$$

$$\omega_{25} = \omega_{19} + \omega_{20} \tag{31}$$

$$\omega_{21} = \omega_{19} + \omega_{20} - \omega_{22} - \omega_{23} - \omega_{24} + \omega_{29} + \omega_{30} + \omega_{31} + \omega_{32} \tag{32}$$

$$\begin{aligned}
 \omega_{22} &= \omega_1 + \omega_2 - \omega_3 - 2\omega_4 + 2\omega_6 - \omega_8 + \omega_{10} - 2\omega_{12} + \omega_{13} - \omega_{14} - \omega_{16} - \omega_{18} + \omega_{27} \\
 &\quad + \omega_{30} - \omega_{31} + \omega_{37} + \omega_{86} + \omega_{88} - \omega_{89}
 \end{aligned} \tag{33}$$

$$\omega_{30} = \omega_1 - \omega_2 + \omega_4 + \omega_9 - \omega_{10} - 2\omega_{17} + \omega_{19} - \omega_{26} - \omega_{35} - \omega_{88} \quad (34)$$

$$\omega_{12} = -\omega_{11} + \omega_{13} + \omega_{14} - \omega_{90} \quad (35)$$

Faz-se as taxas:

$$W_I = \omega_{86} + \omega_{87} + \omega_{88} + \omega_{89} + \omega_{90} \quad (36)$$

$$W_{II} = \omega_{29} + \omega_{30} + \omega_{31} + \omega_{32} + \omega_{33} \quad (37)$$

$$W_{III} = \omega_{18} - \omega_{20} + \omega_{28} \quad (38)$$

$$W_{IV} = \omega_5 - \omega_{11} + \omega_{13} + \omega_{14} + \omega_{15} + \omega_{16} + \omega_{17} + \omega_{21} + \omega_{22} + \omega_{23} - \omega_{32} - \omega_{85} - \omega_{86} - \omega_{90} \quad (39)$$

$$W_V = \omega_1 + \omega_6 + \omega_9 + \omega_{11} - \omega_{14} - \omega_{17} - \omega_{20} - \omega_{26} + \omega_{33} + \omega_{37} + \omega_{90} \quad (40)$$

Tem-se as seguintes combinações lineares para as equações (18), (11), (15), (16), (17), (14), (12) e (13):

$$L(CH_3OH) = -W_I \quad (41)$$

$$L(O_2) = -W_V \quad (42)$$

$$L(CO) = W_{II} - W_{III} \quad (43)$$

$$L(CO_2) = W_{III} \quad (44)$$

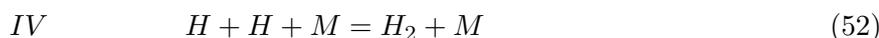
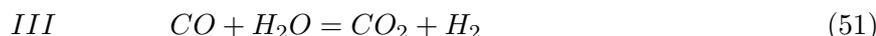
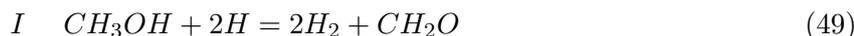
$$L(CH_2O) = W_I - W_{II} \quad (45)$$

$$L(H_2O) = 2W_V - W_{III} \quad (46)$$

$$L(H) = -2W_I - W_{IV} + 2W_V \quad (47)$$

$$L(H_2) = 2W_I + W_{II} + W_{III} + W_{IV} - 3W_V \quad (48)$$

A estequiometria dessas equações corresponde às reações do mecanismo reduzido do metanol:



## 4 Conclusões

Neste trabalho foi realizado a redução do mecanismo de combustão do metanol de 55 reações e 17 espécies para apenas 5 reações e 8 espécies, obtendo uma redução de aproximadamente uma ordem de magnitude. Como perspectiva futura pretende-se simular numericamente o mecanismo reduzido a fim de verificar se o conjunto de 5 reações e 8 espécies, consegue representar o mecanismo do metanol com uma boa precisão e assim obter maiores informações sobre o processo de combustão.

## Referências

- [1] E. Freitas, Protocolo de Kyoto, Brasil Escola, (2018).
- [2] G. Bjurel; G. Dahlquist; L. Odén; B. Lindberg; S. Linde, "Survey of stiff ordinary differential equations", Stockholm Univ. Royal Inst. Technol. Inst. Inform., 1970.
- [3] G. PACZKO; P. LEFDAL; N. PETERS, Reduced reaction schemes for methane, methanol and propane flames, Symposium (International) on Combustion, vol 21, pp. 739-748, (1988).
- [4] G. S. L. Andreis, "Solução via LES de chamas difusivas de metano, metanol e etanol", Tese de Doutorado, Universidade Federal do Rio Grande do Sul, 2011.
- [5] J.R. Jovelino, Análise do processo de produção do Metanol a partir da gaseificação da biomassa para aplicações industriais, III Workshop de Engenharia de Petróleo, 2015.
- [6] M. N. Marinov, A detailed chemical kinetic model for high temperature ethanol oxidation, International Journal of Chemical Kinetics, vol 31, pp. 183-220, (1999).
- [7] N.Peters, B. Rogg, Reduced Kinetic Mechanisms for Applications in Combustion Systems, Springer-Verlag, Germany, 1993.
- [8] S. R. TURNS, An introduction to combustion. Boston: McGraw Hill, Techbooks, New York, 2000.



## Aplicação do Método de Diferenças Finitas na Equação da Difusão de Nêutrons Unidimensional

**Bruno C. Bartz, Marcelo Schramm, Cláudio Z. Petersen**

Centro de Engenharias / Programa de Pós-graduação em Modelagem Matemática, UFPel  
96010-450, Pelotas, RS

E-mail: brunocunhabartz@gmail.com, schrammmarcelo@gmail.com, claudiopetersen@yahoo.com.br

**Resumo:** Neste trabalho os autores aplicaram o método de diferenças finitas na equação unidimensional da difusão de nêutrons. Foi construído um código capaz de simular o comportamento do fluxo de nêutrons e da criticalidade em meios homogêneos e heterogêneos, para um e para dois grupos de energia de nêutrons, no interior de um reator nuclear. Aqui é apresentada a metodologia empregada, além dos resultados obtidos e a comparação destes com trabalhos semelhantes encontrados na literatura.

**Palavras-chave:** Difusão de nêutrons, Fluxo de nêutrons, Diferenças finitas, Método da potência

### 1. INTRODUÇÃO

Foi em 1942, sob um estádio abandonado de futebol americano, que o primeiro reator nuclear, o Chicago Pile 1, conseguiu realizar uma série de reações em cadeia e entrar em funcionamento. Desde o ocorrido, o interesse de pesquisadores ao redor do mundo voltou-se à energia nuclear, maravilhados com a promessa da enorme quantidade de energia contida no núcleo de um átomo [4]. A prova disso é que, atualmente, existem 436 usinas nucleares distribuídas entre 34 países, 16 destes dependendo da energia nuclear para suprir um quarto de suas necessidades elétricas, além de encontrarem-se em construção 53 outras usinas, dentre as quais uma é localizada no Brasil, a chamada Angra 3 [7].

Mesmo com grande potencial hidrelétrico, o interesse do Brasil em energia nuclear remonta à década de 1950, quando em 1952 foi criado o Instituto de Pesquisas Radioativas, em Minas Gerais. Hoje temos a Central Nuclear Almirante Álvaro Alberto, que fica no município Angra dos Reis [3]. Sob a posse da empresa Eletronuclear, estão em funcionamento as unidades Angra 1 e Angra 2, além do andamento das obras da supracitada Angra 3. Angra 1 é atualmente capaz de gerar energia suficiente para suprir as necessidades de um milhão de habitantes, enquanto Angra 2 possui o dobro desta capacidade [6].

O princípio de funcionamento de uma usina nuclear é muito semelhante a usinas termelétricas: uma fonte de calor aquece um líquido ou gás a altas temperaturas. O gás então se expande, girando as pás de uma turbina que converte a energia mecânica gerada em energia elétrica [9]. O diferencial de uma usina nuclear é que o calor utilizado para aquecer o fluido que irá mover a turbina é obtido por meio da energia liberada ao dividir-se o núcleo de um átomo que possui muitos prótons e nêutrons em dois núcleos menores através do impacto de um nêutron, um fenômeno chamado de fissão nuclear. Logo, as energias que estavam envolvidas na união deste núcleo são então liberadas na forma de calor [1].

Cada fissão nuclear, além dos dois núcleos menores, gera também de dois a três outros nêutrons, que irão partir e fissionar outros núcleos, dando início a uma reação em cadeia [1]. O modo como uma reação em cadeia irá proceder pode ser descrito conforme um fator  $k$ , onde  $k$  é o número médio de nêutrons que são diretamente oriundos de um nêutron localizado previamente no sistema [9]. Para que um reator funcione a uma taxa constante, sem que as fissões nucleares cresçam a uma taxa desordenada, é necessário manter  $k$  o mais próximo possível de 1.

É por este motivo que a pesquisa na área nuclear tem se concentrado no controle da

evolução populacional de nêutrons em sistemas nucleares, com destaque aos métodos analíticos, pois permitem verificar se as equações cinéticas dos nêutrons são corretamente resolvidas pelos algoritmos numéricos, permitindo ainda que se adicione os limites da equação [5].

O presente trabalho teve como objetivo o desenvolvimento de um algoritmo com maior precisão para calcular a população de nêutrons em reatores nucleares com um controle legítimo de erro.

## 2. METODOLOGIA

O estudo foi realizado no município de Pelotas, localizado no Rio Grande do Sul, no período de agosto de 2017 a agosto de 2018.

Inicialmente, o código foi construído para controlar um grupo de nêutrons em um meio completamente homogêneo, através da resolução da equação abaixo (1) contendo Problemas de Valor de Contorno (PVC) em  $\phi(0)=0$  e  $\phi(L)=0$ , onde  $L$  representa o  $x$  máximo da malha utilizada:

$$-\phi''(x) + \phi(x) = \frac{1}{k} \cdot 0,5 \cdot \phi(x) \quad (1)$$

Em um segundo caso, desta vez mais completo, a equação cinética dos nêutrons utilizada para construir o código foi a equação (2), com PVC idênticos ao caso anterior:

$$-D(x)\phi''(x) + \Sigma_A(x)\phi(x) = \frac{1}{k} \cdot \nu\Sigma_F \cdot \phi(x) \quad (2)$$

onde:

- $\phi$ : fluxo escalar de nêutrons;
- $D$ : coeficiente de difusão de nêutrons;
- $\Sigma_A$ : seção de choque macroscópica de absorção;
- $k$ : criticalidade;
- $\nu$ : número médio de nêutrons emitido por fissão;
- $\Sigma_F$ : seção de choque macroscópica de fissão.

Para este caso, o código foi desenvolvido de modo a simular o ambiente heterogêneo no interior de um sistema nuclear para um único grupo de nêutrons, com um total de três regiões distintas, nas quais  $D$ ,  $\Sigma_A$  e  $\nu\Sigma_F$  possuem diferentes valores. Para isto foi empregado o princípio analítico de *benchmarking* que, de acordo com [5], pode ser definido um método de solução para problemas que são significantes para a situação física de interesse, mantendo um controle integral de erros em todos os passos.

Para a resolução do PVC foi utilizado o Método das Diferenças Finitas (MDF), que consiste no rearranjo de um problema contínuo em um problema discreto ao utilizar fórmulas de diferenças finitas na malha desejada, seguindo os seguintes passos: 1. Construção da malha; 2. Construção do problema discreto; 3. Resolução do problema discreto; e 4. Visualização e interpretação dos resultados [8].

Dentro do passo 3, a resolução do problema discreto é feita escrevendo-a na forma matricial  $A\phi = b$ , onde  $A$  é uma matriz tridiagonal de coeficientes. Sua forma matricial pode ser vista abaixo:



pelo usuário na preferência 2.

$$erro = (k_i - k_{i-1})/k_i \tag{4}$$

### 3. RESULTADOS E DISCUSSÃO

Para mostrar os resultados em parâmetros semelhantes, os códigos dos meios homogêneo e heterogêneo receberam valores idênticos para o comprimento da malha e o critério de parada, sendo eles, respectivamente,  $L = 10$  e  $erro < 10^{-4}$ . O código foi implementado e executado em um computador com sistema operacional Windows 10 (64-Bit), com processador Intel Core i3-6100U de 2,30 GHz de processamento, 4Gb de memória RAM e utilizando-se o programa SciLab versão 5.5.2. Abaixo, podemos ver lado a lado os gráficos gerados, mostrando-nos o comportamento do fluxo de nêutrons conforme a posição em um meio homogêneo (Figura 1) e heterogêneo (Figura 2).

Figura 1 Comportamento do fluxo de nêutrons em um meio homogêneo

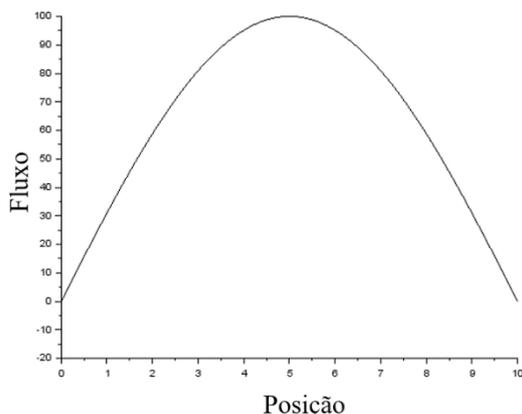
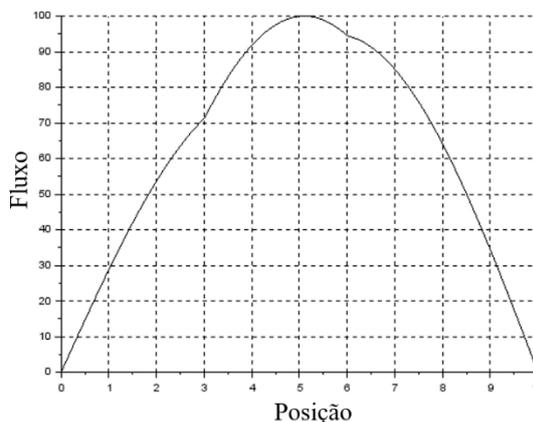


Figura 2 Comportamento do fluxo de nêutrons em um meio heterogêneo



Podemos notar que o comportamento do fluxo no meio homogêneo apresenta-se sem descontinuidades aparentes, uma vez que os seus parâmetros permanecem inalterados ao longo de toda malha. Em contrapartida, o comportamento do fluxo em um meio heterogêneo manifesta descontinuidades acentuadas em  $L = 3$  e  $L = 6$ , que são justamente os pontos de interface onde há mudança dos parâmetros  $D$ ,  $\Sigma_A$  e  $\nu\Sigma_F$ . A localização das interfaces é determinada pelo método de *benchmarking*, onde o código é capaz de abrigar vários casos com diferentes parâmetros e pontos de interface. Assim, o usuário pode adotar o caso que lhe seja mais conveniente. Executando o código para um meio homogêneo, em 19 iterações tivemos que  $k$  convergiu para 0,455086 em 0,01 segundo.

Ao final da execução do código para o meio heterogêneo utilizando os valores escolhidos de  $L$  e  $erro$ , foi impresso na tela que, após 10 iterações,  $k$  convergiu para 0,135478 e o processo durou 0,041 segundos.

Para fins de comparação, posteriormente foram adotados os parâmetros utilizados por [2] que, por sua vez, aplicou e resolveu um problema de *benchmark* disponibilizado pelo Argonne National Laboratory, um centro de pesquisa em engenharia e ciência multidisciplinar localizado nos Estados Unidos da América. Neste caso temos um  $L = 240\text{ cm}$ , com fronteiras em  $L = 40\text{ cm}$  e em  $L = 200\text{ cm}$ . Os parâmetros se encontram abaixo, na Tabela 1.

Região	$D_g^{[i]}$ (cm <sup>-1</sup> )		$\Sigma_{rg}^{[i]}$ (cm <sup>-1</sup> )		$\nu\Sigma_{fg}^{[i]}$ (cm <sup>-1</sup> )		$\Sigma_{s12}^{[i]}$ (cm <sup>-1</sup> )
	$g=1$	$g=2$	$g=1$	$g=2$	$g=1$	$g=2$	-
1 e 3	1,5	0,5	0,026	0,18	0,01	0,2	0,015
2	1,0	0,5	0,02	0,08	0,005	0,099	0,01

Tabela 1. Parâmetros utilizados pelo Argonne National Laboratory

Com estes parâmetros inseridos no código, este gerou o gráfico mostrado na Figura 3. Ao lado deste, pode-se ver o gráfico gerado pelo código criado por [2] (Figura 4). O critério de parada adotado foi o mesmo de [2], *erro* < 10<sup>-15</sup>.

Figura 3 Comportamento do fluxo de nêutrons em um meio heterogêneo

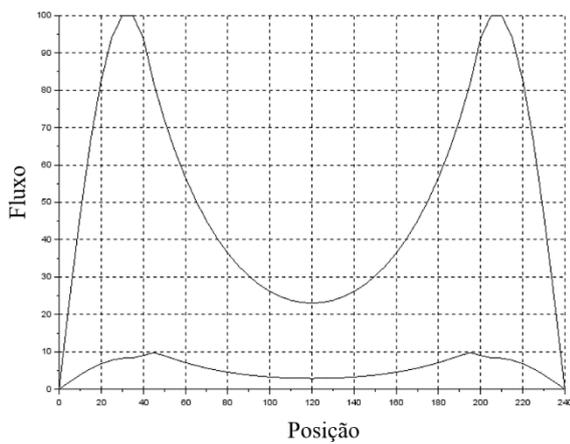
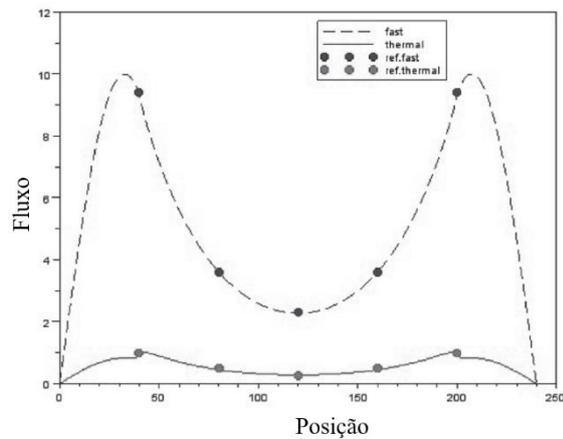


Figura 4 Comportamento do fluxo de nêutrons em um meio heterogêneo [2]



Segundo o código de [2], com  $dx = 10$  cm, ao final  $k$  convergiu para 0,90155. Na Tabela 2 abaixo mostramos os resultados que obtivemos com o código construído aplicando três diferentes refinamentos de malha.

Refinamento	$dx$ (cm)	Tempo de processo (s)	Número de iterações	$k$
1	10	86,08	14662	0,901119
5	2	1300,89	128008	0,901551
10	1	8583,80	406187	0,901584

Tabela 2. Resultados obtidos utilizando três diferentes refinamentos de malha

#### 4. CONCLUSÕES

Averiguando os resultados obtidos e apresentados nas Figuras 1 e 2, podemos afirmar que os mesmos foram satisfatórios, comportando-se conforme o esperado e suportando malhas com um bom grau de refinamento, além de respeitar critérios de parada que se enquadram com o objetivo.

Com relação à comparação realizada com [2], o resultado foi satisfatório, uma vez que o valor para o qual  $k$  convergiu em ambos os códigos mostrou-se semelhante. Todavia, é importante destacar a sensibilidade do código ao aumento do refinamento da malha, gerando  $dx$  cada vez menores. Ao aumentarmos o refinamento de 1 para 5, o tempo de processamento para respeitar o critério de parada aumenta cerca de 15 vezes, e leva 113346 iterações a mais. A

discrepância é ainda maior ao observarmos um *refinamento* = 10 em detrimento de um *refinamento* = 5, quando o processamento levou 6 vezes mais que para o *refinamento* = 5 e o número de iterações praticamente quadruplicou.

Estes fatos observados ao variarmos  $dx$  constituem um ponto fraco do código, que acaba tornando o processo mais demorado a medida que exigimos um refinamento maior dele.

Porém, apesar de alguns resultados satisfatórios, ainda há progressões a serem realizadas na construção dos códigos. Para os deixar mais próximos de casos reais e passíveis de uso em outras condições diversas, ainda é um objetivo a ser alcançado a implementação de um código capaz de executar casos com multigrupos de nêutrons e em regimes que sejam transientes, não apenas uni, mas bidimensionais e tridimensionais.

## REFERÊNCIAS

- [1] E. Cardoso, A Energia Nuclear, 3.ed., Rio de Janeiro: CNEN, 2012. (Apostila Educativa).
- [2] C. Ceolin; M. Schramm; B. Bodmann; M. Vilhena; S. Leite, On an analytical evaluation of the flux and dominant eigenvalue problem for the steady state multi-group multi-layer neutron diffusion equation, *Kerntechnik*, 79 (2014) 430-435.
- [3] CNEN. Biblioteca Digital Memória da CNEN. . Acesso em: 23 de agosto de 2018. Online. Disponível em: <http://memoria.cnen.gov.br/memoria/Cronologia.asp?Unidade=Brasil>.
- [4] J. Duderstadt; L. Hamilton, “Nuclear Reactor Analysis”, John Wiley & Sons, New York, 1976.
- [5] S. Dulla.; P. Ravetto; D. Picca; D. Tomatis, Analytical Benchmarks for the Kinetics of Accelerator-driven Systems, em “Joint International Topical Meeting on Mathematics & Computation and Supercomputing in Nuclear Applications”, Monterey, 2007.
- [6] Eletronuclear. Angra 3. Acesso em: 23 de agosto de 2018. Online. Disponível em: <http://www.eletronuclear.gov.br/Nossas-Atividades/Paginas/Angra-3.aspx>
- [7] L. Guimarães, “A indústria nuclear hoje”, Valor Econômico, São Paulo, 2010.
- [8] C. R. Maliska, Transferência de Calor e Mecânica dos Fluidos Computacional, 2.ed., LTC, Rio de Janeiro, 2017.
- [9] P. Reuss, “Neutron Physics”, EDP Sciences, France, 2008.

# SOLUÇÃO DA CINÉTICA PONTUAL DE NÊUTRONS VIA EQUAÇÃO IN-HOUR PARTICULARIZADA

**Marcelo S.T.A Lourenço, Lucas S. Rocha,**

**Marcelo Schramm, Claudio Z. Petersen**

Centro de Engenharias / Programa de Pós-graduação em Modelagem Matemática, UFPel 96010-450, Pelotas, RS

E-mail: marcelotortolero16@gmail.com, lsrochad@gmail.com

schrammmarcelo@gmail.com, claudiopetersen@yahoo.com.br

**Resumo:** As equações clássicas da cinética pontual de nêutrons são escritas em forma matricial, deixando implícito o operador derivativo dentro da matriz de coeficientes. É utilizada a regra de Laplace para obter a expansão por cofatores. Uma forma geral para a equação in-hour é deduzida para qualquer reatividade transiente linear

**Palavras-chave:** Cinética pontual de nêutrons, equação In-Hour, reatividade constante

## Introdução

A cinética de nêutrons é a área do conhecimento que engloba o comportamento temporal dos nêutrons em reatores nucleares. Classicamente, é categorizada considerando o comportamento espaço-temporal (cinética espacial) ou apenas temporal (cinética pontual) dos nêutrons. Esta última é abordada neste trabalho.

O modelo da cinética pontual é formado a partir de um balanço de nêutrons, onde são consideradas as perdas e produções dentro de um volume de controle. Através da fissão nuclear de um nuclídeo pesado ocorre a produção dos nêutrons. A reação nuclear gera produtos de fissão, energia (que é o objetivo dos reatores nucleares de potência) e nêutrons livres – chamados nêutrons prontos. Esses produtos de fissão, por sua vez, também podem emitir nêutrons via decaimento radioativo – chamados nêutrons atrasados. O modelo abordado (cinética pontual), então, mostra não apenas o balanço entre ganho e perda dos nêutrons prontos e atrasados, mas também dos precursores de nêutrons atrasados. A diferença entre os tempos de geração dos nêutrons prontos e atrasados tornam o problema rígido do ponto de vista matemático, ou seja, exige tratamento algébrico mais adequado que os não rígidos. Para mais detalhes, ver [2] e [4].

No decorrer dos anos, diversos métodos foram utilizados para resolver as equações da cinética pontual de nêutrons. Em [1] é utilizado o modelo de Cattaneo e Vernotte na relação constitutiva, resultando em uma derivada segunda na equação diferencial, resolvido analiticamente com auxílio de software de computação simbólica. Já em [3] utilizando um método de diferenças finitas modificado, são averiguados os efeitos do modelo da cinética pontual fracionária de nêutrons no start-up de um reator. Através de uma abordagem semianalítica, utilizando uma definição apropriada de derivada de ordem fracionária, [5], foi mostrado que os efeitos deste modelo alterado de derivada fracionária para casos não lineares, mais realísticos, são irrelevantes. Um importante resultado obtido em [7] é a dedução da equação in-hour para a apresentada equação da cinética pontual de nêutrons para nêutrons em diferentes faixas de energia.

Vale ressaltar que um dos objetivos do estudo de modelos e metodologias para a cinética pontual de nêutrons é a geração de nova tecnologia a ser implementada na cinética espacial, modelo mais utilizado em reatores nucleares térmicos. Assim, o objetivo deste trabalho é obter uma metodologia eficiente para a obtenção da densidade de nêutrons via equação in-hour, de maneira que seja possível facilmente ser modificada para casos de reatividade não linear como feedback de temperatura. Neste sentido, é abordado o modelo clássico da cinética pontual de nêutrons, juntamente com uma forma generalizada da equação in-hour, que representa os autovalores da matriz de coeficientes, na qual suas raízes são obtidas via método numérico, a partir das quais é construída a densidade de nêutrons via aplicação das condições iniciais

## Metodologia

Para a realização do trabalho, sem perda de generalidade, foram utilizadas as equações da cinética pontual de nêutrons (NPK), levando-se em consideração seis grupos de precursores de nêutrons atrasados, escritas como:

$$\frac{dn(t)}{dt} = \frac{\rho(t) - \beta}{\Lambda} n + \sum_{i=1}^6 \lambda_i C_i(t) \quad , \quad (1)$$

$$\frac{dC_i(t)}{dt} + \lambda_i C_i(t) = \frac{\beta_i}{\Lambda} n(t) \quad . \quad (2)$$

Em que  $n$  é a densidade de nêutrons,  $\rho$  corresponde à reatividade,  $\Lambda$  é o tempo médio de geração instantânea de nêutrons,  $\lambda$  é a constante de decaimento do grupo  $i$  de precursores de nêutrons atrasados,  $\beta_i$  é a fração de nêutrons atrasados emitidos por precursores do grupo  $i$ ,  $\beta$  é o somatório dos  $\beta_i$ ,  $C_i$  é a concentração de precursores de nêutrons atrasados e  $t$  é a variável temporal.

Classicamente são levados em consideração seis grupos de precursores de nêutrons atrasados, porque esta quantidade de grupos já permite que se obtenha a generalização para quaisquer quantidades de grupos.

Estas equações são passíveis de solução através de um sistema de equações diferenciais ordinárias, de modo a utilizar o operador  $D(\bullet) = \frac{d(\bullet)}{dt}$ . Este sistema de equações diferenciais ordinárias foi escrito de forma matricial, omitindo as dependências temporais como:

$$\begin{bmatrix} \frac{\rho - \beta}{\Lambda} - D & \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & \lambda_5 & \lambda_6 \\ \frac{\beta_1}{\Lambda} & -\lambda_1 - D & 0 & 0 & 0 & 0 & 0 \\ \frac{\beta_2}{\Lambda} & 0 & -\lambda_2 - D & 0 & 0 & 0 & 0 \\ \frac{\beta_3}{\Lambda} & 0 & 0 & -\lambda_3 - D & 0 & 0 & 0 \\ \frac{\beta_4}{\Lambda} & 0 & 0 & 0 & -\lambda_4 - D & 0 & 0 \\ \frac{\beta_5}{\Lambda} & 0 & 0 & 0 & 0 & -\lambda_5 - D & 0 \\ \frac{\beta_6}{\Lambda} & 0 & 0 & 0 & 0 & 0 & -\lambda_6 - D \end{bmatrix} \begin{pmatrix} n \\ C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad . \quad (3)$$

Claramente os valores de  $D$  que satisfazem (3) são os autovalores da matriz  $A$ , e a densidade de nêutrons pode ser escrita em função destes autovalores da forma:

$$n(t) = c_1 e^{(D_1 t)} + c_2 e^{(D_2 t)} + c_3 e^{(D_3 t)} + c_4 e^{(D_4 t)} + c_5 e^{(D_5 t)} + c_6 e^{(D_6 t)} + c_7 e^{(D_7 t)} \quad , \quad (4)$$

Em que  $c_i$  para  $i=1, \dots, 7$  são constantes a serem determinadas pelas condições iniciais. O que é mostrado a partir deste ponto é uma forma para se obter os valores de  $D$  (ou os autovalores da matriz  $A$ ).

Chama-se de  $A$  a matriz  $7 \times 7$  apresentada. Para obter o vetor de soluções, é necessário obter os autovalores de  $A$ , ou seja, os valores de  $D$  na qual  $\det(A) = 0$ . A fim de obter uma expressão para o determinante da matriz  $A$  foi utilizada a Regra de Laplace escrita como

$$\det A = \left(\frac{\rho - \beta}{\Lambda} - D\right)\Delta_{11} - \left(\frac{\beta_1}{\Lambda}\right)\Delta_{12} + \left(\frac{\beta_2}{\Lambda}\right)\Delta_{13} - \dots + \left(\frac{\beta_6}{\Lambda}\right)\Delta_{17} \quad , \quad (5)$$

em que  $\Delta_{11}$  ,  $\Delta_{12}$  , (...),  $\Delta_{17}$  são

$$\Delta_{11} = \begin{vmatrix} -\lambda_1 - D & 0 & 0 & \dots & 0 \\ 0 & -\lambda_2 - D & 0 & \dots & 0 \\ 0 & 0 & -\lambda_3 - D & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -\lambda_6 - D \end{vmatrix} \quad , \quad (6a)$$

$$\Delta_{12} = \begin{vmatrix} \lambda_1 & \lambda_2 & \lambda_3 & \dots & \lambda_6 \\ 0 & -\lambda_2 - D & 0 & \dots & 0 \\ 0 & 0 & -\lambda_3 - D & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -\lambda_6 - D \end{vmatrix} \quad , \quad (6b)$$

$$\Delta_{13} = \begin{vmatrix} \lambda_1 & \lambda_2 & \lambda_3 & \dots & \lambda_6 \\ -\lambda_1 - D & 0 & 0 & \dots & 0 \\ 0 & 0 & -\lambda_3 - D & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -\lambda_6 - D \end{vmatrix} \quad , \quad (6c)$$

$$\Delta_{14} = \begin{vmatrix} \lambda_1 & \lambda_2 & \lambda_3 & \dots & \lambda_6 \\ -\lambda_1 - D & 0 & 0 & \dots & 0 \\ 0 & -\lambda_2 - D & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -\lambda_6 - D \end{vmatrix} \quad , \quad (5d)$$

$$\Delta_{15} = \begin{vmatrix} \lambda_1 & \lambda_2 & \lambda_3 & \dots & \lambda_6 \\ -\lambda_1 - D & 0 & 0 & \dots & 0 \\ 0 & -\lambda_2 - D & 0 & \dots & 0 \\ 0 & 0 & -\lambda_3 - D & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -\lambda_6 - D \end{vmatrix} \quad , \quad (6e)$$

$$\Delta_{16} = \begin{vmatrix} \lambda_1 & \lambda_2 & \lambda_3 & \dots & \lambda_6 \\ -\lambda_1 - D & 0 & 0 & \dots & 0 \\ 0 & -\lambda_2 - D & 0 & \dots & 0 \\ 0 & 0 & -\lambda_3 - D & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -\lambda_6 - D \end{vmatrix} \quad , \quad (6f)$$

$$\Delta_{17} = \begin{vmatrix} \lambda_1 & \lambda_2 & \lambda_3 & \dots & \lambda_6 \\ -\lambda_1 - D & 0 & 0 & \dots & 0 \\ 0 & -\lambda_2 - D & 0 & \dots & 0 \\ 0 & 0 & -\lambda_3 - D & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{vmatrix} . \quad (6g)$$

Estes determinantes das submatrizes,  $\Delta_{1j}$ , podem ser obtidos aplicando novamente a Regra de Laplace da forma:

$$\Delta_{11} = (-\lambda_1 - D)(-\lambda_2 - D)(-\lambda_3 - D) \dots (-\lambda_6 - D) , \quad (7a)$$

$$\Delta_{12} = \lambda_1(-\lambda_2 - D)(-\lambda_3 - D)(-\lambda_4 - D) \dots (-\lambda_6 - D) , \quad (7b)$$

$$\Delta_{13} = -\lambda_2(-\lambda_1 - D)(-\lambda_3 - D)(-\lambda_4 - D) \dots (-\lambda_6 - D) , \quad (7c)$$

$$\Delta_{14} = \lambda_3(-\lambda_1 - D)(-\lambda_2 - D)(-\lambda_4 - D) \dots (-\lambda_6 - D) , \quad (7d)$$

$$\Delta_{15} = -\lambda_4(-\lambda_1 - D) \dots (-\lambda_3 - D)(-\lambda_5 - D)(-\lambda_6 - D) , \quad (7e)$$

$$\Delta_{16} = \lambda_5(-\lambda_1 - D)(-\lambda_2 - D) \dots (-\lambda_4 - D)(-\lambda_6 - D) , \quad (7f)$$

$$\Delta_{17} = -\lambda_6(-\lambda_1 - D)(-\lambda_2 - D)(-\lambda_3 - D) \dots (-\lambda_5 - D) . \quad (7g)$$

Observa-se um padrão apresentado nas equações (7b), (7c), (7d), (7e), (7f) e (7g), o que permite a representação genérica de  $\Delta_{12}$  a  $\Delta_{17}$ . Ressalta-se também que devido ao padrão observado a partir  $\Delta_{12}$  é possível generalizar a forma para  $P$  grupos de precursores de nêutrons atrasados, que está representado a seguir, em (7b). Logo,  $\Delta_{11}$  pode ser escrito genericamente da forma apresentada em (8<sup>a</sup>) como:

$$\Delta_{11} = \prod_{i=1}^P (\lambda_i + D) , \quad (8a)$$

$$\Delta_{1j} = \lambda_{j-1} (-1)^{(P-1)} \prod_{\substack{i=1 \\ i \neq j-1}}^P (\lambda_i + D) , \quad (8b)$$

em que  $j=1, 2, \dots, P$ .

Substituindo (8a) e (8b) na equação (5), é encontrada uma equação geral (para  $P$  grupos de precursores de nêutrons) polinomial cujas raízes podem ser obtidas através da utilização de métodos numéricos, que possibilita a obtenção de uma solução para o determinante da matriz  $A$  da forma:

$$\det(A) = \left[ (-1)^P \left( \frac{\rho(t) - \beta}{\Lambda} - D \right) \prod_{i=1}^P (\lambda_i + D) \right] + \sum_{i=1}^P \left[ (-1)^{P+i-1} \frac{\beta_i \lambda_i}{\Lambda} \prod_{\substack{j=1 \\ j \neq i}}^P (\lambda_j + D) \right] = 0 . \quad (9)$$

## Conclusão

Tendo em vista a metodologia desenvolvida no presente trabalho, é necessário que sejam encontradas as raízes (valores de  $D_i$  ou autovalores da matriz  $A$ ) da equação (9).

Um código em Scilab já está estágio final de elaboração com o intuito de obter, numericamente, os valores de  $D_i$  que correspondem aos autovalores da matriz  $A$ . Consequentemente, a solução  $n(t)$  dada pela equação (4) será obtida.

O método numérico utilizado para a obtenção dos valores de  $D_i$  é o método da bissecção, porque este método é um método intervalar, procurando por raízes dentro do intervalo escolhido. Além disso, é um método robusto (eficiente) para o que é proposto porque, diferentemente do método de Newton-Raphson, não necessita que se encontre a derivada da equação (9), que não é uma tarefa simples de ser realizada.

A tabela a seguir apresenta alguns resultados parciais no que diz respeito aos valores de  $D_i$  (ou autovalores da matriz  $A$ ) obtidos utilizando o método de aproximação da bissecção na equação (9), com uma tolerância equivalente a  $10^{-8}$ , com uma varredura de 0,01 percorrendo de -100 até 1000, e conseguindo encontrar as 7 raízes em um tempo computacional equivalente a 7,138 segundos. Vale ressaltar que esta tabela com os autovalores da matriz  $A$  correspondem aos resultados piloto do código, em que foi considerado  $\rho = -0,5\beta$  :

---

Valores de  $D_i$  : -22,471064 -3,9488377 -1,2658064 -0,4026771 -0,1036417 -0,03592 -0,0124538

---

Esses valores encontrados correspondem aos valores encontrados utilizando o comando “spec”(um comando nativo do Scilab para encontrar raízes) após gerar uma matriz  $A$  no ambiente matemático do scilab.

No que diz respeito ao futuro do trabalho, os autores pretendem considerar a reatividades dependente do tempo, o que torna o modelo mais realístico para simular mudanças na potência do reator para pequenos intervalos de tempo.

## Referências

- [1] Altahhan, Muhammad Ramzy *et al.* Formulation of a point reactor kinetics model based on the neutron telegraph equation. *Annals of Nuclear Energy*, [S. l.], vol. 91, p. 176-188, mai. 2016.
- [2] Duderstadt, James J.; Hamilton, Louis J. *Nuclear Reactor Analysis*. New York: John Wiley & Sons, 1974.
- [3] Polo-Labarríos, M.-A; Espinosa-Paredes, G. Application of the fractional neutron point kinetic equation: Start-up of a nuclear reactor. *Annals of Nuclear Energy*, [S. l.], vol. 46, p. 37-42, ago. 2012.
- [4] Reuss, Paul. *Neutron Physics*. Les Ulis Cedex A, France: EDP Sciences, 2008.
- [5] Schramm, M. *et al.* The neutron point kinetics equation: Suppression of fractional derivative effects by temperature feedback. *Annals of Nuclear Energy*, [S. l.], vol. 87, p. 479-485, jan. 2016.
- [6] Tumelero, Fernanda. *Solução das Equações da Cinética Pontual de Nêutrons com e sem Retroalimentação de Temperatura pelo Método da Aproximação Polinomial*. 2015. Dissertação (Pós-Graduação em Modelagem Matemática) - Universidade Federal de Pelotas, Pelotas, RS, 2015.
- [7] Yamoah, S.; Akaho, E.H.K; Nyarko, B.J.B. An accurate solution of point reactor neutron kinetics equations of multi-group of delayed neutrons. *Annals of Nuclear Energy*, [S. l.], vol. 54, p. 104-108, abr. 2013.



# Avaliação do comportamento efetivo de compósitos microperiódicos não lineares, com contato imperfeito entre as fases, através do Método de Homogeneização Assintótica: uma introdução com o caso unidimensional

**Roberto M. S. Décio Júnior**

Programa de Pós Graduação em Modelagem Matemática, IFM, UFPel,  
96160-000 Pelotas, RS

E-mail: roberto.decio.jr@ufpel.edu.br

**Leslie D. Pérez-Fernández**

Instituto de Física e Matemática, UFPel,  
96160-000 Pelotas, RS

E-mail: leslie.fernandez@ufpel.edu.br

**Julian Bravo-Castillero**

Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas, UNAM  
Ciudad de México, Mexico

E-mail: julian@mym.iimas.unam.mx

**Resumo:** *Neste trabalho o Método de Homogeneização Assintótica é aplicado em problemas de valores de contorno estáticos e unidimensionais, com coeficientes rapidamente oscilantes e constantes por partes, com o objetivo de avaliar o comportamento efetivo de compósitos microperiódicos não lineares, levando-se em conta ainda as situações de contato perfeito e imperfeito entre as fases. Para esta avaliação, considerou-se um fluxo não linear do tipo potencial, que apresenta interessantes interpretações físicas. O método foi aplicado com sucesso, levando-se a resultados passíveis de relevantes interpretações: os diferentes expoentes de não linearidade considerados influenciaram o comportamento efetivo; a variação das concentrações das fases nos compósitos consideramos reproduziu uma variação significativa no comportamento global; e quanto ao contato imperfeito, a avaliação de casos limites forneceu resultados coerentes com os significados físicos considerados. Diante disso, conclui-se que o MHA é uma ferramenta eficaz na avaliação do comportamento efetivo de compósitos periódicos, mesmo que não lineares e sob condição de contato imperfeito, e dessa forma, se faz necessário expandir esta aplicação principalmente a outros tipos de não linearidades, e considerar outras metodologias a fim de validar os resultados aqui obtidos*

**Palavras-chave:** *Compósitos, Método de Homogeneização Assintótica, Materiais Não Lineares*

## Exposição do problema

Os materiais compósitos têm mostrado interessantes aplicações, em diversas áreas, estando hoje o seu mercado difundido na área dos transportes (31%), construção civil (19,7%), marinha (12,4%), equipamento eléctrico/electrónico (9,9%), produtos de consumo (5,8%), aeroespacial e de aeronaves (0,8%), entre outros, segundo [15]. De forma geral, compósitos são materiais heterogêneos formados pela união de dois ou mais materiais homogêneos, os quais serão suas fases. Aqui, serão considerados compósitos periódicos cuja escala de distribuição das fases

(microescala) é muito maior que a escala atômica e muito menor que a escala macroscópica, caracterizando assim a microperiodicidade.

A micro-heterogeneidade destes materiais faz com que os problemas matemáticos que modelam seu comportamento físico apresentem coeficientes rapidamente oscilantes, dificultando a aplicação direta de métodos numéricos usuais (como Diferenças Finitas e Elementos Finitos)[11]. Entretanto, simultaneamente, é satisfeita a Hipótese da Homogeneidade Equivalente, e o material pode ser considerado fisicamente equivalente a um material homogêneo ideal, cujos problemas apresentam coeficientes constantes e cuja solução é suficientemente próxima do problema original. O processo de obtenção deste material homogêneo é chamado de Homogeneização.

Dentre os métodos de homogeneização está o Método de Homogeneização Assintótica (MHA) [2], o qual considera uma aproximação assintótica da solução do problema original, em forma de série de potências de um parâmetro pequeno positivo  $\varepsilon$  e em dupla escala (macro e micro). Ao aplicar esta série assintótica no problema original, obtém-se uma sequência recorrente de problemas para os coeficientes das potências de  $\varepsilon$ , que darão origem à problemas locais cujas soluções serão os termos da série assintótica inicialmente proposta. Dentre as vantagens do MHA, destacam-se os fatos de apresentar baixo custo computacional no uso de métodos numéricos para resolver os problemas da sequência recorrente, e produzir boas aproximações da solução exata do problema original. O método apresenta duas aplicações principais: aproximar a solução do problema original através de soluções assintóticas e determinar o comportamento efetivo do material heterogêneo através do homogêneo equivalente. Neste trabalho, a abordagem será da segunda aplicação.

Tradicionalmente, encontram-se relações constitutivas lineares para modelar fenômenos físicos, como a Lei de Hooke da teoria de elasticidade, a Lei de Fourier no caso do fluxo do calor, a Lei de Nernst para o fluxo de partículas de um gás, entre outros. Porém, existem muitos fenômenos físicos de natureza não linear, os quais não podem ser estudados através de modelos lineares, por exemplo: plasticidade, viscoelasticidade, hiperelasticidade, eletrostrição, magnetostrição, termoplasticidade, entre outros [13]. Desta forma, se faz estudar tais modelos, esperando-se obter resultados tão proeminentes quanto os obtidos para os casos lineares. Alguns destes podem ser encontrados em [7], [5], [8] e [6].

Outra característica importante dos compósitos é a adesão das suas fases constituintes. Muitos trabalhos consideravam que os materiais constituintes eram perfeitamente ligados, implicando na continuidade da solução do problema nesta interface [3]. Entretanto, na prática, a adesão não é perfeita, ocorrendo imperfeições mecânicas, isolamentos térmicos, reações químicas ou então diferença de potencial elétrico. Devido a isso, se faz importante considerar este fato nos problemas que modelam o comportamento destes materiais. De fato, tem-se encontrado que é muito relevante a influência do contato imperfeito no comportamento efetivo de materiais compósitos ([14], [9], [10], [12], [1]).

Diante disto, o presente trabalho tem por objetivo aplicar o MHA em um problema de valores de contorno para a equação de difusão estática e unidimensional, a fim de avaliar em diferentes situações o comportamento efetivo de um compósito bifásico microperiódico não linear, nas situações de contato perfeito e imperfeito entre as fases.

## Metodologia

Considera-se um compósito bifásico  $\varepsilon$ -periódico, unidimensional de comprimento  $l$ , como um intervalo  $[0, l]$ , equivalente a uma união finita de  $N$  células periódicas  $\Omega_1 \cup \Omega_2$  e dos pontos de interface  $x_j$  ( $j = 1, 2, \dots, N$ ). O problema  $P_O$  de valores de contorno, que modela um fenômeno difusivo estacionário, com fluxo não linear, para um meio microperiodico, com propriedades constantes por partes, com condições de contato imperfeito para a solução e perfeito para o

fluxo, será da seguinte forma:

$$P_O : \begin{cases} \frac{d}{dx} \left[ \sigma \left( \frac{x}{\varepsilon}, \frac{du^\varepsilon}{dx} \right) \right] = f(x), \quad x \in (0, l) - \{x_j\}_{j=1}^N \\ \sigma \left( \frac{x}{\varepsilon}, \frac{du^\varepsilon}{dx} \right) = \begin{cases} \sigma_1 \left( \frac{du^\varepsilon}{dx} \right), \quad x \in \Omega_1 \\ \sigma_2 \left( \frac{du^\varepsilon}{dx} \right), \quad x \in \Omega_2 \end{cases} \\ \left[ \left[ \sigma \left( \frac{x}{\varepsilon}, \frac{du^\varepsilon}{dx} \right) \right] \right]_{x=x_j} = 0 \\ \left[ [u^\varepsilon(x)] \right]_{x=x_j} = \frac{\varepsilon}{\beta} \sigma_1 \left( \frac{du^\varepsilon}{dx} \right) \Big|_{x=x_j} \\ u^\varepsilon|_{x=0} = g_1, \quad u^\varepsilon|_{x=l} = g_2 \end{cases} \quad (1)$$

onde o fluxo  $\sigma^\varepsilon$  é  $\varepsilon$ -periódico, continuamente diferenciável por partes em  $[0, l]$ ;  $K_1$  e  $K_2$  são constantes que representam as propriedades das fases constituintes; e a fonte  $f(x)$  contínua por partes em  $[0, l]$ . A não linearidade do tipo potencial considerada modela importantes fenômenos, como deformação plástica e termoplasticidade, apresentando geralmente  $n < 1$  ([13], [4]).

Em um contexto mecânico,  $\sigma^\varepsilon$  é a tensão de deformação,  $u^\varepsilon$  o deslocamento e seu gradiente a deformação. Já considerando os casos condutivos térmico e elétrico,  $\sigma^\varepsilon$  é o fluxo de calor ou a corrente elétrica, e  $u^\varepsilon$  a temperatura ou o potencial elétrico, respectivamente. O parâmetro  $\beta > 0$  pode ser entendido como uma condutância na interface, de forma que  $1/\beta$  representa uma resistência. Chama-se atenção para o fato de que, quando  $\beta \rightarrow \infty$ , tem-se uma condição de contato para  $u^\varepsilon$  nula, ou seja, a condição de contato perfeito.

Iniciando a abordagem do MHA neste problema, propõe-se uma solução assintótica em dupla escala para a solução  $u^\varepsilon(x)$ , na forma:

$$u^{(2)}(x, y) = v_0(x) + \varepsilon u_1(x, y) + \varepsilon^2 u_2(x, y), \quad (2)$$

com  $y = x/\varepsilon$ ,  $v_0(x), u_k(x, y) \in C^2([0, l])$  para  $k = 1, 2$ , 1–periódicas em  $y$ . Substituindo  $u^{(2)}(x, y)$  na equação diferencial do problema  $P_O$ , considerando o fluxo linearizado por um polinômio de Taylor no ponto  $\zeta = \frac{dv_0}{dx} + \frac{\partial u_1}{\partial y}$ , obtém-se:

$$\begin{aligned} \frac{d}{dx} \left[ \sigma \left( \frac{x}{\varepsilon}, \frac{du^{(2)}}{dx} \right) \right] - f(x) &= \varepsilon^{-1} \left[ \frac{\partial \sigma}{\partial y}(y, \zeta) \right] + \varepsilon^0 \left[ \frac{\partial \sigma}{\partial x}(y, \zeta) + \right. \\ &\quad \left. + \frac{\partial}{\partial y} \left[ \left( \frac{\partial u_1}{\partial x} + \frac{\partial u_2}{\partial y} \right) \frac{\partial \sigma}{\partial \varepsilon}(y, \zeta) \right] - f(x) \right] + O(\varepsilon). \end{aligned} \quad (3)$$

E para que se cumpra a igualdade assintótica, ou seja, para que se tenha:

$$\left| \frac{d}{dx} \left[ \sigma \left( \frac{x}{\varepsilon}, \frac{du^{(2)}}{dx} \right) \right] - f(x) \right| = O(\varepsilon), \quad (4)$$

toma-se  $u_2 = 0$ , e obtém-se a seguinte equação para  $\varepsilon^{-1}$ :

$$\frac{\partial \sigma}{\partial y}(y, \zeta) = 0. \quad (5)$$

Aplicando  $u^{(2)}(x, y)$  também nas condições de contato e contorno, obtém-se a seguinte família uniparamétrica de problemas para  $u_1(x, y)$ , a saber:

$$P_L^{\frac{dv_0}{dx}} : \begin{cases} \frac{\partial \sigma}{\partial y} \left( y, \frac{dv_0}{dx} + \frac{\partial u_1}{\partial y} \right) = 0, \quad y \in (0, 1) - \{c_1\} \\ \left[ \left[ \sigma \left( y, \frac{dv_0}{dx} + \frac{\partial u_1}{\partial y} \right) \right] \right]_{y=c_1} = 0 \\ \left[ [u_1] \right]_{y=c_1} = \frac{1}{\beta} \sigma_1 \left( y, \frac{dv_0}{dx} + \frac{\partial u_1}{\partial y} \right) \Big|_{y=c_1} \\ u_1(0, 0) = u_1(l, 0) = 0 \end{cases} \quad (6)$$

A existência e unicidade da solução 1-periódica deste problema é dado pelo seguinte Lema, generalizado para a condição de contato imperfeito do lema análogo encontrado em [2]:

**Lema:** *Seja  $\bar{\epsilon}$  um parâmetro e  $\sigma(y, \epsilon)$  continuamente diferenciável por partes em  $[0, 1]$ . Então, para todo  $x$  fixo, existem funções  $\mathcal{N}_1(y, \bar{\epsilon})$  1-periódicas em  $y$  que são solução de uma família uniparamétrica de problemas  $P_L^{\bar{\epsilon}}$  com parâmetro  $\bar{\epsilon}$ , definida por:*

$$\begin{cases} \frac{\partial \sigma}{\partial y} \left( y, \bar{\epsilon} + \frac{\partial \mathcal{N}_1}{\partial y} \right) = 0, & y \in (0, 1) - \{c_1\} \\ \left[ \left[ \sigma \left( y, \bar{\epsilon} + \frac{\partial \mathcal{N}_1}{\partial y} \right) \right] \right]_{y=c_1} = 0; & \left[ \mathcal{N}_1 \right]_{y=c_1} = \frac{1}{\beta} \sigma_1 \left( y, \bar{\epsilon} + \frac{\partial \mathcal{N}_1}{\partial y} \right) \Big|_{y=c_1} \end{cases} \quad (7)$$

A condição  $\mathcal{N}_1(0, \bar{\epsilon}) = 0$  garante a unicidade da solução.

Garantindo a existência e unicidade da solução do problema na equação 6, se faz possível encontrar a lei efetiva (ou comportamento efetivo), que será a relação funcional entre a média de  $\sigma(\bar{\sigma})$  e a média do gradiente de  $u_1(\bar{\epsilon})$ .

Do problema local para  $u_1$  (equação 6), tem-se que o fluxo não depende de  $y$  no ponto  $\zeta = \frac{dv_0}{dx} + \frac{\partial u_1}{\partial y}$ , logo:

$$\sigma \left( y, \frac{dv_0}{dx} + \frac{\partial u_1}{\partial y} \right) = \sigma(x) \equiv \bar{\sigma}. \quad (8)$$

Agora, de isolar  $\partial u_1 / \partial y$  na equação 8, obtém-se:

$$\frac{\partial u_1}{\partial y} = \varphi(y, \bar{\sigma}) - \frac{dv_0}{dx}, \quad (9)$$

onde  $\varphi(y, \bar{\sigma})$  é a inversa do fluxo  $\sigma$ , garantida pelo Teorema da Função Implícita. Agora, aplicando a média sobre a célula periódica (representada por  $\langle \cdot \rangle$ ), tem-se que:

$$u_1|_{y=1} - [u_1]_{y=c_1} - u_1|_{y=0} = \langle \varphi(y, \bar{\sigma}) \rangle - \frac{dv_0}{dx}. \quad (10)$$

Da condição de contato para  $u_1$ , chega-se a:

$$\frac{\bar{\sigma}}{\beta} = \langle \varphi \rangle(\bar{\sigma}) - \frac{dv_0}{dx}. \quad (11)$$

O fluxo efetivo  $\hat{\sigma} \left( \frac{dv_0}{dx} \right)$  é determinado encontrando a relação, que está implícita em (11), entre  $\bar{\sigma}$  e  $dv_0/dx$  (que será denotado por  $\bar{\epsilon}$ ).

## Resultados

Será considerado um compósito bifásico, periódico, com uma fase (1) linear e a outra (2) não linear. A não linearidade considerada será a do tipo potencial, com expoente  $n$ , a qual modela importantes fenômenos, como deformação plástica e termoplasticidade, apresentando geralmente  $n < 1$  ([13], [4]):

$$\sigma \left( \frac{x}{\epsilon}, \frac{du^\epsilon}{dx} \right) = \begin{cases} K_1 \frac{du^\epsilon}{dx}, & x \in \Omega_1 \\ K_2 \left( \frac{du^\epsilon}{dx} \right)^n, & x \in \Omega_2 \end{cases} \quad (12)$$

onde  $K_1$  e  $K_2$  são as propriedades de cada fase (módulo de elasticidade, condutividade térmica ou elétrica, por exemplo).

Considerando o procedimento de aplicação do MHA no problema  $P_O$  considerando o fluxo como na equação 12, chega-se a seguinte equação para determinar a lei para o fluxo efetivo (análoga a 11):

$$\frac{c_1}{K_1} \bar{\sigma} + \frac{c_2}{K_2^{1/n}} \bar{\sigma}^{1/n} + \frac{\bar{\sigma}}{\beta} - \bar{\epsilon} = 0, \quad (13)$$

onde  $c_1$  é a concentração da fase 1 (linear), e  $c_2$  da fase 2 (não linear). Adimensionalizando-se os termos com  $\lambda = K_1/K_2$  e  $\bar{\sigma}/K_1$ , a equação 13 torna-se:

$$\lambda^{1/n} c_2 \left( \frac{\bar{\sigma}}{K_1} \right)^{1/n} + \left( c_1 + \frac{1}{\beta_{k_1}} \right) \frac{\bar{\sigma}}{K_1} - \bar{\epsilon} = 0, \tag{14}$$

com  $\beta_{k_1} = \beta/K_1$ . A fim de obter a lei efetiva, esta foi resolvida para diferentes valores dos parâmetros  $n$ ,  $c_1$ ,  $\lambda$  e  $\beta$  envolvidos através de iteração de Newton.

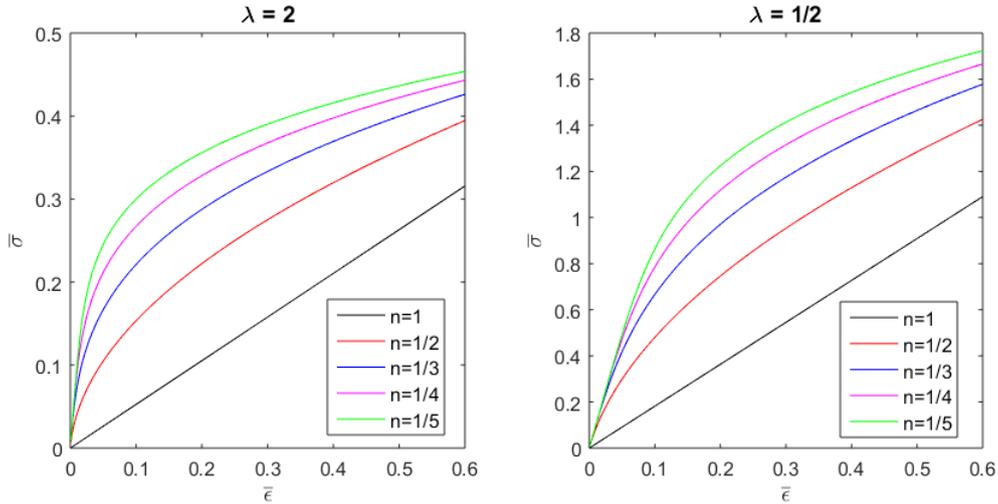


Figura 1: Lei Efetiva para a condição de contato perfeito, em relação à diferentes valores do expoente de não linearidade  $n$ .

Na Figura 1, é considerado o caso do contato perfeito ( $\beta \rightarrow \infty$ ), e determinou-se  $\hat{\sigma}$  para  $c_1 = 0.1$ , e para diferentes valores de  $n$  ( $n \leq 1$ ) e  $\lambda$ . Observando-a, é visível o comportamento puramente linear para  $n = 1$ . Este resultado confirma que o comportamento global do compósito será linear, desde que ambas as fases sejam lineares. Além disso, a medida que  $n$  decresce, o comportamento global se "afasta" do comportamento linear, em ambos casos de  $\lambda$  considerados. O caso de  $\lambda = 2$  representa um compósito com uma família de materiais bifásicos com propriedade linear duas vezes mais intensa, e  $\lambda = 1/2$  o mesmo, mas para a propriedade não linear.

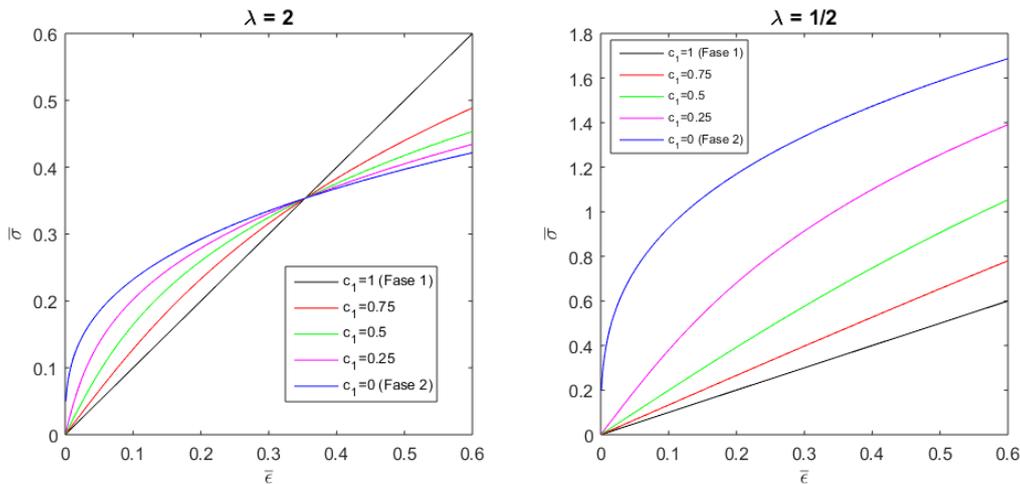


Figura 2: Lei Efetiva para a condição de contato perfeito, em relação à diferentes valores de concentração das fases.

Agora, na Figura 2, é fixado  $n = 1/3$ , e ainda considerado o contato perfeito, variando os valores de concentrações das fases através de  $c_1$ . Diante dos resultados obtidos, é visto que para os dois casos de  $\lambda$  avaliados, a medida que aumenta-se a concentração  $c_1$ , o comportamento global se aproxima do comportamento linear da fase 1, da mesma forma que a medida que diminui-se  $c_1$ , o comportamento global tende ao não linear da fase 2. Tendo em vista que os dois casos extremos,  $c_1 = 1$  e  $c_1 = 0$ , representam 100% de presença do material 1 e do material 2, respectivamente, estes resultados mostram a influência das concentrações das fases no comportamento efetivo.

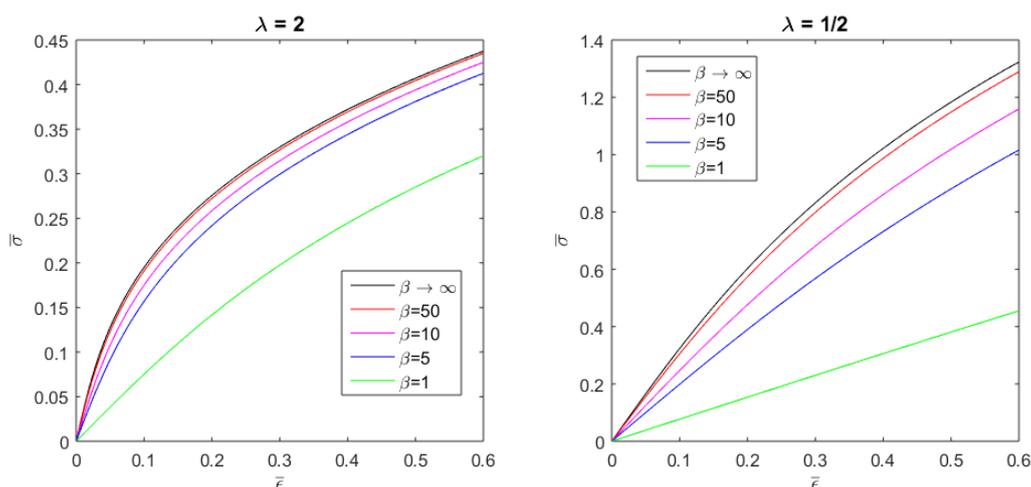


Figura 3: Lei Efetiva para a condição de contato imperfeito, em relação à diferentes valores do parâmetro  $\beta$ .

Finalmente, na Figura 3, é avaliado o comportamento efetivo para diferentes casos de contato imperfeito, fixando-se  $c_1 = 0.3$  e  $n = 1/3$ , e variando-se o parâmetro  $\beta$ . Fica claro o comportamento assintótico das leis efetivas obtidas para o caso do contato perfeito, quando  $\beta \rightarrow \infty$ , visto que a diferença entre as curvas na Figura 3 é praticamente imperceptível para  $\beta > 50$ . Além disso, percebe-se que no outro extremo, quando o valor de  $\beta$  é tomado menor ( $\beta \rightarrow 1$ ), a lei efetiva assume valores menores também.

Estes resultados para o contato imperfeito são coerentes quando pensa-se nas aplicações. No caso de condutividade, por exemplo, quanto menor o valor de  $\beta$ , maior é a resistência ao fluxo de calor na interface, ou seja, menor será o fluxo efetivo do material. E por maior que seja tomado este valor, o fluxo efetivo não poderá superar o fluxo da situação ideal, que é o caso de contato perfeito.

Em um contexto mecânico, o aumento da resistência da interface pode ser interpretado como a diminuição da rigidez nesta. Assim, quanto menor o valor de  $\beta$ , menor essa rigidez, o que resulta numa menor rigidez efetiva do compósito. Ou seja, o material se deforma mais, quando submetido às mesmas tensões. Isto resulta em curvas de tensão-deformação com menor inclinação, o que é visto também na Figura 3.

## Conclusões

De forma geral, conclui-se primeiramente que o MHA é uma ferramenta eficaz na obtenção da lei efetiva de compósitos microperiódicos, ainda que não lineares e apresentando contato imperfeito entre as fases. A sua aplicação nestes problemas fornece resultados muito relevantes, mostrando que a concentração de cada fase, o tipo de não linearidade e o contato imperfeito entre as fases influencia diretamente o comportamento global do compósito.

Em suma, é vista a necessidade de expandir os resultados aqui obtidos, primeiramente para diferentes tipos de não linearidade, a fim de aumentar a abrangência das aplicações do MHA.

Ademais, se faz necessário considerar outras metodologias de homogeneização a afim de comparar e validar os resultados aqui obtidos.

## Referências

- [1] F. E. Alvarez-Borges; J. Bravo-Castillero; M. E. Cruz; R. Guinovart-Diaz; L. D. Pérez-Fernández; R. Rodríguez-Ramos; F. J. Sabina, Reiterated homogenization of a laminate with imperfect contact: gain-enhancement of effective properties, *Appl. Math. Mech. -Engl. Ed.*, 39(8) (2018) 1119-1146.
- [2] N. S. Bakhvalov; G. P. Panasenko, “Homogenisation: Averaging Processes in Periodic Media”, Kluwer Academic Publishers, Dordrecht, 1989.
- [3] Y. Benveniste, The effective mechanical behaviour of composite materials with imperfect contact between the constituents, *Mechanics of Materials*, 4 (1985) 197-208.
- [4] G. Dai; F. Jarrar; F. Ozturk; J. Sheik-Ahmad, On the Effect of the Complexity of the Constitutive Model in Simulating Superplastic Forming, *Defect and Diffusion Forum*, 385 (2018) 379384.
- [5] S. Ghosh; J. Bai; D. Paquet, Homogenization-based continuum plasticity-damage model for ductile failure of materials containing heterogeneities, *J. of the Mech. and Phys. of Sol.*, 57 (2009) 1017-1044.
- [6] . A. L. Kalamkarov; F. Tornabene; P. M. C. L. Pacheco; M. A. Savi; G. C. Saha, Geometrically non-linear elastic model for a thin composite layer with wavy surfaces, *Z. Angew. Math. Mech.*, 97(11) (2017) 1381-1392.
- [7] Z. F. Khisaeva; M. Ostoja-Starzewski, Scale effects in infinitesimal and finite thermoelasticity of random composites, *Journal of Thermal Stresses*, 30 (2007) 587-603.
- [8] T. Lemaire; J. Kaiser; S. Naili; V. Sansalone, Textural versus electrostatic exclusion-enrichment effects in the effective chemical transport within the cortical bone: A numerical investigation, *Int. J. Numer. Meth. Biomed. Engng.*, 29 (2013) 1223-1242.
- [9] R. Lipton; B. Vernescu, Composites with imperfect interface, *Proc. R. Soc. Lond. A*, 452 (1996) 329-358.
- [10] J. C. López-Realpozo; R. R. Ramos; R. G. Díaz; J. B. Castillero; F. J. Sabina, Compuestos elásticos no lineales con condiciones de contacto imperfectas, *Revista Electrónica Nova Scientia*, 1(1) (2008) 53-64.
- [11] G. P. Panasenko, Homogenization for Periodic Media: from Microscale to Macroscale, *Physics of Atomic Nuclei*, 71(4) (2008) 681-694.
- [12] R. Penta; A. Gerisch, The asymptotic homogenization elasticity tensor properties for composites with material discontinuities, *Continuum Mech. Thermodyn*, 29 (2017) 187-206.
- [13] P. Ponte-Castañeda; P. Suquet, Nonlinear Composites, *Advances in Applied Mechanics*, 34 (1998) 171303.
- [14] S. Torquato; M. D. Rintoul, Effect of the Interface on the Properties of Composite Media, *Physical Review Letters*, 75(22) (1995) 4067-4070.
- [15] A. F. M. Ventura, Os Compósitos e a sua aplicação na Reabilitação de Estruturas metálicas, *Ciência & Tecnologia de Materiais*, 21(3-4) (2009) 10-19.



## **Modelagem Matemática de uma Bancada Robotizada com Acionamento Pneumático para Reabilitação Física**

### **Roberta Goergen,**

Instituto Federal Farroupilha, IFFAR, Campus Panambi,  
E departamento de Ciências Exatas e Engenharias, UNIJUI,  
98280-000, Panambi, RS  
E-mail: betinhamtm@gmail.com,

### **Marcia Regina Maboni Hoppen Porsch, Luiz Antonio Rasia**

Departamento de Ciências Exatas e Engenharias, UNIJUI,  
98280-000, Panambi, RS  
E-mail: marcia\_porsch@hotmail.com, rasia@unijui.edu.br

### **Antonio Carlos Valdiero**

Faculdade de Horizontina, FAHOR,  
98920-000, Horizontina, RS  
E-mail: antoniocvaldiero@gmail.com

### **Maurício Oberdörfer, Jocarly Patrocínio de Souza**

Universidade de Passo Fundo, UPF,  
99052-900, Passo Fundo, RS  
E-mail: mauriciooberdorfer@hotmail.com, jocarly@upf.br

**Resumo:** *Atualmente muitas pessoas tem sofrido acidentes ou doenças que comprometem a mobilidade das diferentes condições motoras. A função do membro superior ou inferior é um dos principais problemas que interrompem as atividades do cotidiano destas pessoas. Com a finalidade de auxiliar essas pessoas a recuperar o membro afetado, muitos dispositivos robóticos de apoio para movimentos dos membros superiores e inferiores estão sendo desenvolvidos. O presente trabalho tem como objetivo a pesquisa e o desenvolvimento da modelagem matemática de um mecanismo robótico, com atuadores pneumáticos, aplicado na recuperação física de pacientes. Utiliza-se um atuador pneumático para que seja possível obter segurança para a operação e o controle da força. A metodologia utilizada na modelagem consiste da definição dos sistemas de coordenadas de referência e das relações cinemáticas a partir da convenção de Denavit-Hartenberg, da dedução das equações da dinâmica pela formulação de Lagrange e da inclusão da dinâmica do atuador pneumático com a não linearidade de atrito. Esta pesquisa aborda o desenvolvimento do robô para reabilitação a partir de uma metodologia de integração de modelagem matemática com as fases do processo de projeto. Portanto os resultados obtidos são um desenho modular para reabilitação de membros inferiores e que pode ser estendido para membros superiores. E também apresentar a proposta metodológica para facilitar o entendimento das relações matemáticas não lineares em mecanismos robóticos que é importante para fins de simulação, análise de desempenho e estratégias de controle.*

**Palavras-chave:** Modelagem Matemática, Atuadores Pneumáticos, Reabilitação Física, Modelagem Dinâmica

## **1 – Introdução**

Diversos dispositivos robóticos estão sendo pesquisados e desenvolvidos para auxiliar o processo de reabilitação. A ciência da reabilitação mostrou que, na maioria dos casos, movimentos repetidos de membros humanos inferiores e superiores afetados, podem ajudar o paciente a recuperar as funções da lesão do membro. Existem três tipos de sistemas mecânicos usados para reabilitação de movimentos: robôs, manipuladores baseados em cabos e exoesqueletos. [2], [3], [6], [7].

A população de pacientes que necessitam de reabilitação física de membros inferiores e superiores está aumentando constantemente devido a capacidades perdidas em algum tipo de doença, lesão muscular ou acidente. Trabalhar em técnicas de reabilitação influencia a independência e qualidade de vida desses pacientes. Os dispositivos robóticos têm o potencial de ser uma ferramenta útil na reabilitação. No entanto, a disponibilidade desses dispositivos em ambientes clínicos é limitada, deixando muito espaço para melhorias e estimulando o desenvolvimento de novas estruturas.

Estudos recentes mostraram que os exoesqueletos robóticos são usados no treinamento de reabilitação, a fim de melhorar as limitações de mobilidade e, conseqüentemente, contribuir para mudanças significativas na qualidade de vida. [9]

Em [8], encontra-se um esboço otimizado de um manipulador a cabo, que se destina a reabilitação ou exercício de pacientes com problemas no ombro, como doenças, eventos traumáticos ou para idosos que precisam exercitar seus membros.

O projeto e controle de um robô de três graus de liberdade para reabilitação de membros inferiores de um paciente após uma lesão medular, acidente vascular cerebral, distúrbio muscular ou cirurgia foi abordado em [1]. Neste caso o robô manipulador pode executar todos os exercícios ativos e passivos, bem como aprender movimentos de exercícios específicos e executá-los sem o fisioterapeuta através da interface homem-máquina.

A formulação da modelagem matemática da dinâmica é importante e necessária, para que as simulações possam ser feitas e as leis de controle desenvolvidas para robôs manipuladores. Além disso, o modelo dinâmico também é útil no dimensionamento de peças e juntas de mecanismo, e para selecionar atuadores de robô.

Este trabalho visa contribuir no campo da robótica de reabilitação, apresentando um projeto conceitual inicial de um dispositivo robótico modular robótico para reabilitação de membros inferiores. Primeiro, a descrição da bancada robotizada para reabilitação física de membros inferiores é apresentada seguida da formulação matemática para o dispositivo proposto. Finalmente, as conclusões deste trabalho e sugestões futuras são apresentadas.

## 2 – Descrição da Bancada Robotizada para Reabilitação Física

Esta seção tem como objetivo mostrar o projeto conceitual de um dispositivo para reabilitação física de membros inferiores. O mecanismo robótico proposto apresenta um movimento independente (flexão / extensão do joelho). Observa-se pela Figura 1, o sistema robótico de reabilitação de membros inferiores que é dividido em três principais componentes: mecanismo, sistema pneumático e sistema de controle. [4]

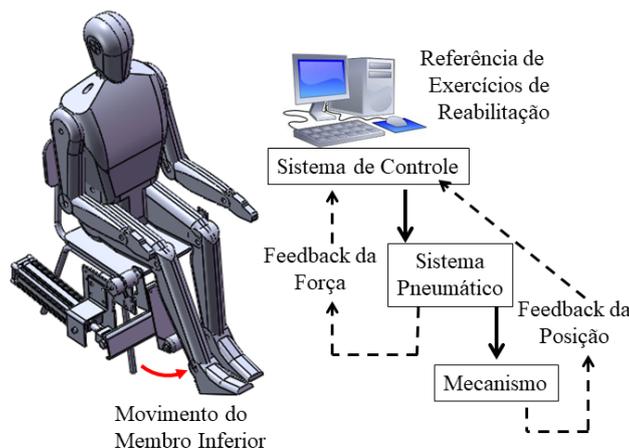


Figura 1: Os principais componentes do protótipo do robô reabilitação de membros inferiores com acionamento pneumático. Adaptado [5]

Na Figura 1 apresenta-se a descrição do protótipo em desenvolvimento, o robô com acionamento pneumático que será utilizado como aplicação na reabilitação física de membros inferiores, construído em uma estrutura fixa com um grau de liberdade, de junta rotativa.

Na Figura 2, mostra-se uma vista com os movimentos propostos para a reabilitação dos membros inferiores. O projeto prevê a construção de um protótipo, que será montado sobre uma cadeira convencional e a programação das trajetórias de movimento de referência por um fisioterapeuta. O mecanismo consiste de um conjunto de elos rígidos unidos por articulações e montados sobre uma cadeira, que tem a função de realizar os movimentos para os exercícios de reabilitação.

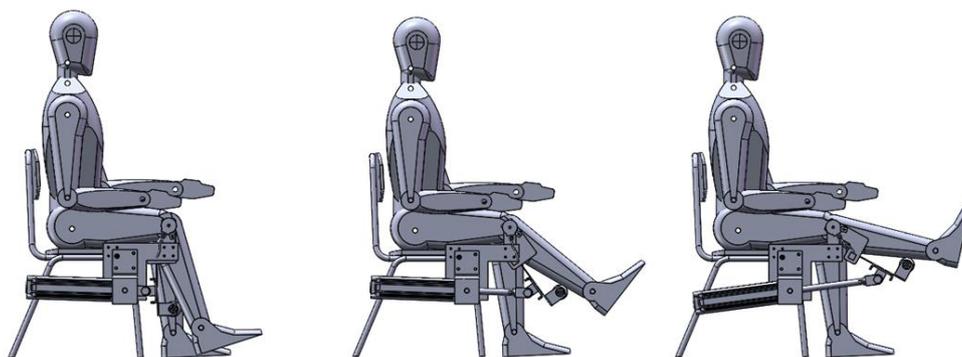


Figura 2: Os movimentos do protótipo do robô reabilitação de membros inferiores com acionamento pneumático.

O protótipo está sendo desenvolvido, projetado e testado no Núcleo de Inovação em Máquinas Automáticas e Servo Sistemas (NIMASS), que possui infraestrutura computacional e experimental adequada para a construção de uma bancada de testes para verificar/validar o desempenho da modelagem e do controle em atuadores pneumáticos.

### 3 –Modelagem Matemática

Nesta seção apresenta-se o modelo dinâmico de um robô para reabilitação física com um grau de liberdade (1 dof), mas podem ser estendidos para múltiplos graus de liberdade em função dos movimentos requeridos pela reabilitação. Este modelo matemático proporciona as equações que relacionam os torques gerados pelos atuadores nas juntas com o movimento da estrutura. Além disso, o conhecimento do modelo dinâmico de um manipulador robótico e de suas propriedades tem um papel fundamental nas estratégias de controle.

O modelo dinâmico do mecanismo robótico proposto pode ser derivado pela Formulação de Lagrange. O estudo do comportamento dinâmico do mecanismo é muito útil para o projeto mecânico da estrutura, escolha de atuadores, determinação de estratégias de controle e simulação computacional do movimento do manipulador. [10]

A Figura 3 mostra uma vista do protótipo com a indicação dos principais parâmetros conforme a convenção de Denavit-Hartenberg (D-H). [5], [10], [11]

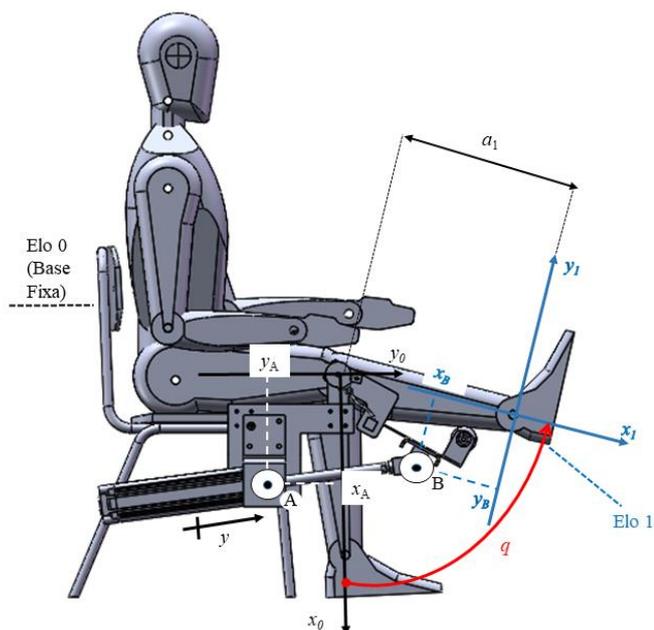


Figura 3: Parâmetros de D-H para o robô reabilitação de membros inferiores com acionamento pneumático

O atuador é localizado por meio dos pontos A e B, onde o ponto A é solidário e fixo ao elo 0 (Base Fixa) e pode ser descrito pelas coordenadas  $(x_A, y_A)$  no sistema de referência do elo 0. E analogamente o ponto B é fixo ao elo 1 e pode ser descrito pelas coordenadas  $(x_B, y_B)$  no sistema de referência do elo1. Onde  $q$  é a variável de junta e  $a_l$  é a distância normal comum entre o eixos das juntas.

As equações de Lagrange são expressas por

$$\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}} \right) - \frac{\partial L}{\partial q} = \xi \quad (1).$$

Onde  $\xi$  é a força generalizada associada à coordenada generalizada  $q$ ,  $L = K - U$  é o lagrangeano do sistema mecânico;  $K$  é a energia cinética e  $U$  é a energia potencial do sistema.

A Energia Cinética  $K_{l_1}$  para o elo 1 é dada pela equação (2).

$$K_{l_1} = \frac{1}{2} m_{l_1} \dot{p}_{l_1}^T \dot{p}_{l_1} + \frac{1}{2} w_1^T R_1 I_{l_1}^1 R_1^T w_1 \quad (2)$$

Onde  $m_{l_1}$  é a massa do elo 1;  $l_1$  a distância entre o centro de massa e eixo da junta 1;  $\dot{p}_{l_1}$  é a velocidade linear do centro de massa;  $w_1$  é a velocidade angular do elo 1;  $R_1$  é a matriz de rotação do elo 1 para a base fixa;  $I_{l_1}^1$  é uma matriz simétrica e representa o tensor de inércia relativo ao centro de massa do elo 1, em que o vetor de coordenadas generalizadas é  $q = \theta_1$ .

Então, observando-se a Figura 4, deduzimos a Cinemática através da Convenção Denavit-Hartenberg:

$$\dot{p}_{l_1} = \begin{bmatrix} -l_1 \sin \theta_1 \cdot \dot{\theta}_1 \\ l_1 \cos \theta_1 \cdot \dot{\theta}_1 \\ 0 \end{bmatrix}, w_1 = \begin{bmatrix} 0 \\ 0 \\ \dot{\theta}_1 \end{bmatrix}, R_1 = \begin{bmatrix} \cos \theta_1 & -\sin \theta_1 & 0 \\ \sin \theta_1 & \cos \theta_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ e } I_{l_1}^1 = \begin{bmatrix} I_{xx}^1 & -I_{xy}^1 & -I_{xz}^1 \\ -I_{yx}^1 & I_{yy}^1 & -I_{yz}^1 \\ -I_{zx}^1 & -I_{zy}^1 & I_{zz}^1 \end{bmatrix}.$$

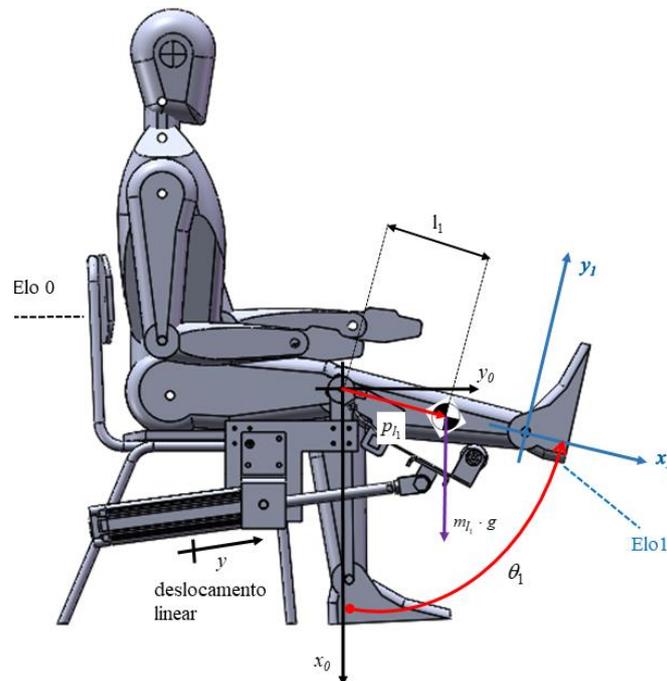


Figura 4: Descrição da Cinemática do elo 1 para a Formulação de Lagrange

Portanto a Energia Cinética  $K_l$  no elo 1 é dado pela equação 3.

$$K_l = \frac{1}{2} m_l l_1^2 \dot{\theta}_1^2 + \frac{1}{2} I_{zz}^1 \cdot \dot{\theta}_1^2 \quad (3)$$

Agora calcula-se a energia Potencial  $U_l$  para o caso do elo 1:

$$U_l = -m_l g_0^T p_l \quad (4)$$

Onde  $g_0$  é a aceleração da gravidade. E neste caso  $g_0 = \begin{bmatrix} g \\ 0 \\ 0 \end{bmatrix}$  e  $p_l = \begin{bmatrix} l_1 c \theta_1 \\ l_1 s \theta_1 \\ 0 \end{bmatrix}$ .

Logo a energia Potencial  $U_l$  para o elo 1 é dado pela equação 5

$$U_l = -m_l \cdot g \cdot l_1 \cdot c \theta_1 \quad (5)$$

Substituindo na Equação 1, a equação da energia Cinética obtida em (3) e a equação da energia Potencial obtida em (5) chega-se à formulação de Lagrange apresentada na equação (6).

$$(m_l l_1^2 + I_{zz}^1) \ddot{\theta}_1 + m_l \cdot g \cdot l_1 \cdot s \theta_1 = \tau_1 \quad (6)$$

Onde  $\tau_1$  é o vetor das forças do movimento da junta 1.

Combinando a modelagem dinâmica do elo 1 dado por (6), com a dinâmica do atuador pneumático dado por  $\tau = \bar{J}^T f_L$ , que expressa o vetor dos torques  $\tau$  de acionamento da articulação como uma função do vetor de força de carga  $f_L$  no atuadores lineares. Desta forma, podemos escrever a equação (7).

$$(m_l l_1^2 + I_{zz}^1) \ddot{\theta}_1 + m_l \cdot g \cdot l_1 \cdot s \theta_1 = \bar{J}^T f_L \quad (7)$$

Onde  $\bar{J}^T$  é definida como a matriz Jacobiana do atuador, a qual mapeia as velocidades dos atuadores com as velocidades no espaço das juntas.

Nesta seção, explicamos brevemente a modelagem matemática de um robô para reabilitação física com um grau de liberdade acionado por atuador pneumático, mas pode ser estendido para sistemas de múltiplos graus de liberdade em função dos movimentos requeridos pela reabilitação.

Considerando-se um robô de  $n$  graus de liberdade (veja Figura 5), pode-se obter o modelo dinâmico no espaço das juntas por meio da equação de Lagrange e escrever na forma de matriz compacta, conforme equação (8).

$$H(q)\ddot{q} + C(q, \dot{q})\dot{q} + G(q) = \tau. \quad (8)$$

Onde  $\dot{q}$  é o vetor de velocidades das juntas;  $\ddot{q}$  é o vetor de aceleração das juntas;  $H(q)$  é a matriz de inércia simétrica,  $C(q, \dot{q})$  é a matriz que representa os efeitos centrífugos e de Coriolis e  $G(q)$  é o vetor que representa o momento gerado em cada eixo de junta referente a gravidade.

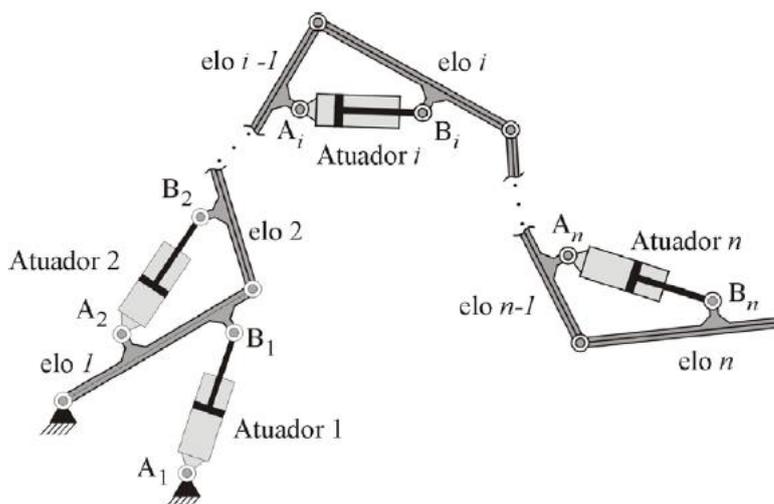


Figura 5: Desenho de um robô de juntas rotativas acionado por atuadores lineares. Adaptado [11]

Portanto, mostrou-se neste trabalho a dedução da modelagem matemática da relação dinâmica do acionamento pneumático de uma bancada robotizada como estratégia de pesquisa empregada em uma potencial aplicação na reabilitação de membros inferiores.

## 4 – Conclusões

Este trabalho mostrou o desenvolvimento e a modelagem matemática de um novo mecanismo robótico, acionado pneumaticamente, para reabilitação de membros inferiores. As relações matemáticas não-lineares deduzidas para o mecanismo robótico permitiram obter o movimento funcional desejado para a reabilitação do membro inferior. Estudos futuros devem investigar estratégias para o controle e assistência ao movimento articular humano.

## Referências

- [1] E. Akdoğan, M. A. Adli. The design and control of a therapeutic exercise robot for lower limb rehabilitation: Physiotherobot. Mechatronics, 21(3), 509-522, 2011

- [2] J. Baier, L. Kuchinke, M. Neumann and B. Bender, "Form and function — Exemplary analysis of the significance for the design of rehabilitation devices," 2017 International Conference on Rehabilitation Robotics (ICORR), London, pp. 740-745, 2017
- [3] I. Díaz, J. J. Gil, and E. Sánchez, "Lower-Limb Robotic Rehabilitation: Literature Review and Challenges", *Journal of Robotics*, vol. 2011, Article ID 759764, 11 pages, 2011.
- [4] R. Goergen, M. R. M. H. Porsch, A. C. Valdiero, L. A. Rasia, M. Oberdörfer, J. P. Souza de, R. S. Gonçalves. Design of a Pneumatically Driven Robotic Workbench for Rehabilitation. *International Journal of Development Research*, Vol. 08, pp. 22408-22413, Aug. 2018
- [5] R. Goergen, M. R. M. H. Porsch, L. A. Rasia, A. C. Valdiero, M. Oberdörfer and J. P. de Souza. Modelagem Matemática da Relação Cinemática do Acionamento Pneumático de uma Bancada Robotizada para Reabilitação Física. *Proceeding series of the Brazilian Society of Computational and Applied Mathematics*, 7(1), 2018
- [6] R. S. Gonçalves, J. C. M. Carvalho, *Robot Modeling for Physical Rehabilitation. Service Robots and Robotics Design and Application*. IGI Global, p. 154-175. DOI: 10.4018/978-1-4666-0291-5.ch009, 2012
- [7] R. S. Gonçalves, T. Hamilton and H. I. Krebs, MIT – Skywalker: On the use of a markerless system. *IEEE – RAS – EMBS International Conference on Rehabilitation Robotics*, 2017.
- [8] E. Hernandez, S. I. Valdez, G. Carbone, M. Ceccarelli, M. Design Optimization of a Cable-Driven Parallel Robot in Upper Arm Training-Rehabilitation Processes. In: *Multibody Mechatronic Systems: Proceedings of the MUSME Conference held in Florianópolis, Brazil, October 24-28, 2017*. Springer, 2017. p. 413.
- [9] H. Hsieh, D. Chen, L. Chien and C. Lan, "Design of a Parallel Actuated Exoskeleton for Adaptive and Safe Robotic Shoulder Rehabilitation," in *IEEE/ASME Transactions on Mechatronics*, vol. 22, no. 5, pp. 2034-2045, Oct. 2017.
- [10] B. Siciliano, L. Sciavicco, L. Villani, G. Oriolo, "Modelling, Planning and control". Springer: London, 2009.
- [11] A. C. Valdiero. Modelagem matemática de robôs hidráulicos, Ijuí: UNIJUI, 1:200, 2012.



## **Modelagem Estatística do Processo de Destilação de Membranas aplicado à Recuperação de Águas Residuais da Indústria Têxtil**

**Regilene de Sousa Silva<sup>1</sup>; Miriam Albara<sup>2</sup>, Rita de Cassia S. C. Valle<sup>2</sup>, Cintia Marangoni<sup>1,2</sup>**

<sup>1</sup>Programa de Pós-Graduação em Engenharia Química, Departamento de Engenharia Química e Engenharia de Alimentos, Campus Reitor João David Ferreira Lima, Universidade Federal de Santa Catarina, Florianópolis, SC

<sup>2</sup>Programa de Pós-Graduação em Engenharia Têxtil, Departamento de Engenharia Têxtil, Universidade Federal de Santa Catarina - Campus Blumenau, Blumenau-SC  
e-mail: [regilenelela@hotmail.com](mailto:regilenelela@hotmail.com)

### **Resumo**

*No presente trabalho buscou-se realizar a modelagem estatística de um processo de Destilação por Membrana por Contato Direto (DMCD) aplicado ao tratamento de águas residuais da indústria têxtil com soluções sintéticas dos corantes preto reativo e preto disperso e membrana plana comercial de poli(tetrafluoretileno) (PTFE). Foi empregado um delineamento composto central rotacional constituído de um planejamento fatorial 2<sup>3</sup>, com repetições no ponto central e ao nível de 95% de confiança. Vazão de alimentação, (Va), vazão de permeado (Vp) e temperatura de alimentação (Ta) foram as variáveis independentes, fluxo final (J) e a taxa de rejeição de cor (TR) as variáveis dependentes. Observou-se que a temperatura de alimentação é o efeito mais significativo para as duas classes de corantes estudadas. No caso do corante reativo, a vazão de permeado seguida da interação entre esta variável e a vazão de alimentação foram os efeitos de maior significância. Já com o corante disperso, a vazão de permeado também se apresentou significativa, mas a interação observada foi entre esta variável foi a temperatura de alimentação. Foi observado que a taxa de rejeição de cor não depende das condições operacionais e sim das características das membranas e dos corantes. Por fim, obteve-se o modelo de regressão para o fluxo do permeado, sendo este considerado como adequado para representação do fluxo de permeado do processo devido ao elevado ajuste (R<sup>2</sup>) obtido e a distribuição normal de resíduos.*

### **Introdução**

A indústria têxtil produz uma grande quantidade de águas residuais, onde o efluente gerado está contaminado com corantes e diversas outras substâncias nocivas. Mokhtar e colaboradores [1] indicam que o processo de tingimento é a maior fonte poluidora das águas residuais têxteis. O tecido retém apenas certa quantidade de corante durante o processo, resultando na emissão de efluentes altamente coloridos. Embora o consumo de água varie dependendo do processo e tipo de tecido, geralmente é em torno de 2,5-932 L por kg de produto [2]. Portanto, as indústrias têxteis precisam de alternativas para impedir que uma alta produção de águas residuais seja descartada sem tratamento adequado. Sob estas condições, projetos para diminuição da geração de efluentes aquosos e a possibilidade de reuso de água vem sendo estudados.

A Destilação por Membrana (DM) é uma tecnologia avançada de separação térmica, em que uma solução aquosa é destilada através de uma membrana hidrofóbica porosa, permitindo a completa rejeição dos componentes não voláteis na alimentação, como sais. O processo pode ser operado a temperaturas relativamente baixas e pode, portanto, utilizar fontes térmicas sustentáveis, como a energia solar, ou mesmo fontes potencialmente livres, como fluxos de calor residual de fábricas [3].

O processo vem conquistando espaço como uma técnica de separação proporcionando recuperação e tratamento de efluentes da indústria têxtil. Várias configurações são possíveis

para os processos de DM, porém a de contato direto (DCMD) é a forma mais simples, econômica e eficiente para o tratamento de águas residuais têxteis. Uma das vantagens é que as águas residuais têxteis são geralmente descarregadas em torno de 90°C e, portanto, não há necessidade de energia térmica para aquecê-las no processo [2]. E uma desvantagem é o de incrustações (*fouling*) que é também presente na DCMD, porém em intensidade significativamente menor.

O desempenho dos sistemas de destilação por membrana pode ser incrementado por meio da avaliação das condições de operação ou configuração do sistema. Tais esforços podem melhorar a permeabilidade da membrana ou reduzir a polarização da temperatura para alcançar maiores fluxos, taxas de rejeição de cor e eficiência energética. Dessa forma, a utilização de metodologias como planejamento experimental permite o conhecimento do processo de forma sistemática, levando a possibilidade de exercer melhorias sobre o mesmo. É possível obter informações específicas de todos os tratamentos e observar efeitos de interação, o que não tem sido realizado nos estudos envolvendo destilação por membranas. Apesar de já ser conhecida a influência de algumas variáveis no desempenho do processo, pouco se tem estudado sobre a interação entre efeitos. Pode-se destacar também que as condições operacionais avaliadas são testadas em uma faixa relativamente pequena, o que não se observa ao utilizar um planejamento fatorial.

Assim, o objetivo deste trabalho foi avaliar a influência das variáveis do processo e de suas possíveis interações e, obter desta forma a modelagem estatística do processo de destilação por membranas de contato direto aplicado a águas residuais da indústria têxtil. Duas classes de corante foram avaliadas: reativo e disperso, que correspondem a maioria dos corantes utilizados em tingimentos.

## Material e métodos

Foi aplicado um delineamento composto central rotacional  $2^3$  com três repetições no ponto central utilizando os seguintes fatores: vazão de alimentação ( $V_a$ ), vazão de permeado ( $V_p$ ) e temperatura de alimentação ( $T_a$ ). Foram empregadas soluções sintéticas de corante preto reativo e preto disperso e membranas planas comerciais de PTFE que apresentaram os melhores resultados permeabilidade com o processo de DMCD de acordo com a literatura [4]. Como variáveis dependentes foram definidas: fluxo final ( $J$ ) e taxa de rejeição ( $TR$ ). Na Tabela 1 são apresentados os níveis das variáveis independentes do delineamento experimental seguido neste trabalho. Por meio dos resultados do planejamento foram determinados os coeficientes de regressão para a resposta de interesse do processo e realizada análise de variância (ANOVA) a um nível de significância de 5%. Os dados obtidos por meio dos experimentos foram analisados utilizando o software Statistica versão 13 para Windows.

Tabela 1: Valores dos níveis das variáveis independentes do planejamento fatorial empregado no processo DMCD aplicado às águas residuais têxteis com corantes reativo e disperso.

Condições	Axiais (- $\alpha$ )	Inferior (-1)	Central (0)	Superior (+1)	Axiais (+ $\alpha$ )
$V_a$	1,3	2	3	4	4,7
$V_p$	0,2	0,4	0,7	1	1,2
$T_a$	60	66	75	84	90

Na Figura 1 é ilustrado um diagrama esquemático da unidade utilizada nos experimentos. A unidade apresenta um lado de alimentação (quente) e um lado oposto de permeado (frio), ambos mantidos em circulação durante operação. No lado da alimentação a corrente flui tangencialmente à camada ativa da membrana e o permeado, tem o fluxo tangencial do lado oposto da camada ativa da membrana. Amostras dos reservatórios de alimentação e de permeado foram retiradas para análise do fluxo de permeado e da rejeição de cor. Assim, foi avaliado o fluxo final de permeado,  $J$  ( $\text{kg}\cdot\text{m}^{-2}\cdot\text{h}^{-1}$ ) segundo a Equação 1, e a taxa rejeição de cor da membrana,  $TR$  (%) segundo a Equação 2.

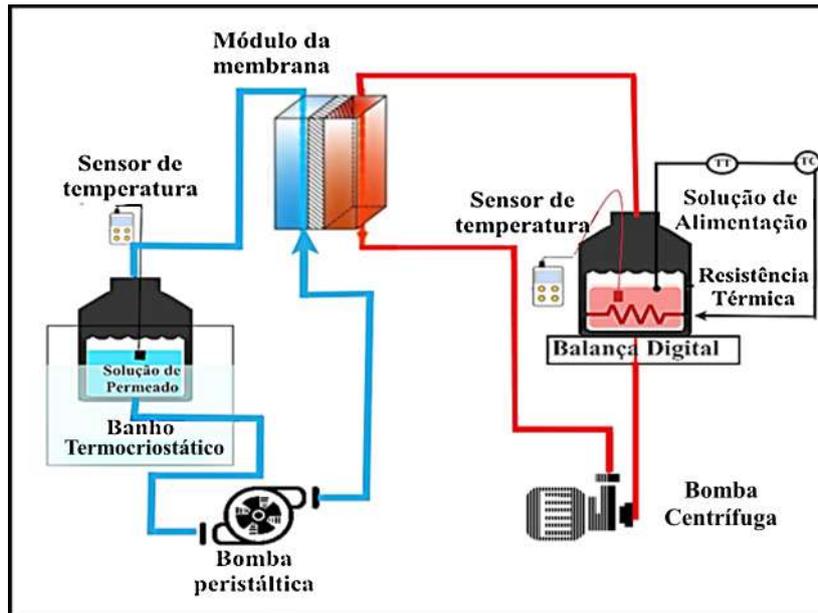


Figura 1: Diagrama esquemático da unidade experimental utilizada

$$J = \frac{\Delta M}{A \times \Delta t} \quad (1)$$

onde M é a massa de permeado coletado (kg), A a área efetiva da membrana (m<sup>2</sup>) e Δt o tempo total do processo (h).

$$TR = \frac{C_A - C_P}{C_A} \times 100 = 1 - \frac{C_P}{C_A} \times 100 \quad (2)$$

onde CA é a concentração do corante no lado de alimentação e CP a concentração do corante no lado do permeado.

### Resultados e Discussão

Na Figura 2 é mostrado os resultados do diagrama de Pareto para o corante preto reativo, onde a magnitude dos efeitos é representada pelas colunas, enquanto que a linha transversal às colunas representa a magnitude dos efeitos ao nível de significância de 5%.

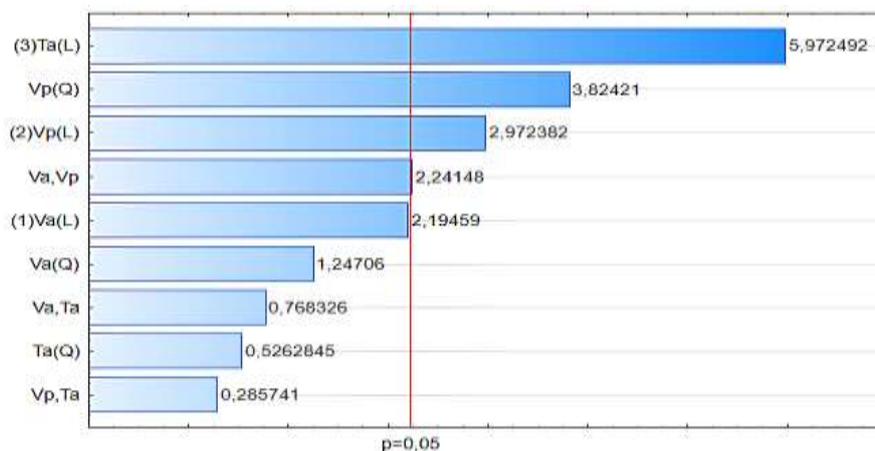


Figura 2: Diagrama de Pareto com a estimativa dos efeitos significativos (α=0,05) do efluente de corante PR das variáveis independentes estudadas relacionadas ao fluxo de permeado.

Analisando a figura observa-se os efeitos e a ordem de influência das variáveis independentes na resposta. As variáveis Ta linear e Vp quadrático e linear apresentam efeitos significativos ao nível de confiança de 95% em relação ao fluxo final, bem como interação Va e Vp. Observa-se também que os efeitos são positivos, ou seja, ao passar do nível mais baixo para o mais alto há um aumento no valor de fluxo de permeado.

A temperatura de alimentação é um dos parâmetros mais importantes. O incremento desta variável aumenta a pressão parcial interfacial e, subsequentemente o fluxo transmembrana [5]. Ali et al. [6] explicam que o fluxo de permeado aumenta exponencialmente com o acréscimo da temperatura da alimentação. Isso indica que a temperatura de alimentação contribuiu principalmente para maior gradiente de pressão de vapor, aumentando a transferência de massa que corresponde ao padrão de fluxo de alta permeabilidade. A principal razão para o aumento do fluxo de água é que o aumento da temperatura entre a alimentação e o permeado aumentou a força motriz da pressão de vapor através da membrana, resultando em uma condução significativa e perda de calor latente, promovendo uma grande diferença de temperatura entre a solução em massa e a superfície da membrana.

Para o efluente de corante preto disperso a análise estatística dos fatores estudados ao nível de significância de 95% também é apresentada por meio de Diagrama de Pareto, conforme a Figura 3.

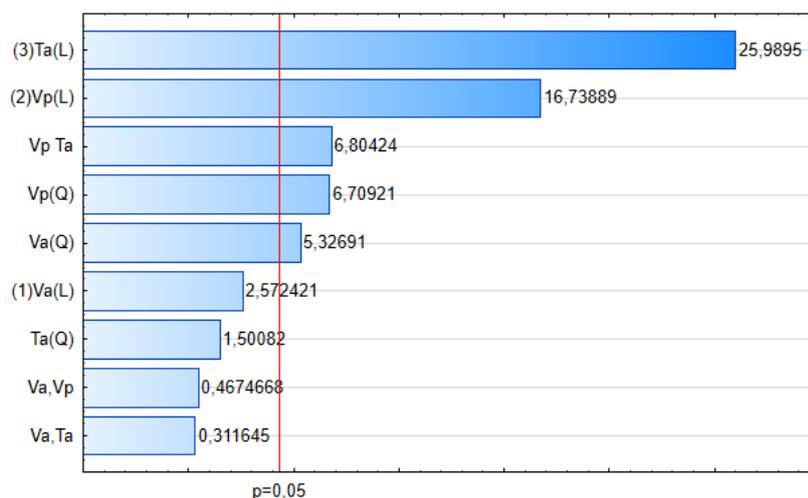


Figura 3: Diagrama de Pareto com a estimativa dos efeitos significativos ( $\alpha=0,05$ ) do efluente de corante PD das variáveis independentes estudadas relacionadas ao fluxo de permeado.

As variáveis Ta linear, Vp quadrático e linear, Va quadrático e a interação Vp e Ta apresentam efeitos significativos em relação ao fluxo final ao nível de 95% de confiança. Assim como observado para o estudo realizado com o corante reativo, a temperatura de alimentação se apresentou como a variável de maior influência sobre o processo. Porém, os ensaios com o corante disperso estudado apresentou diferente fluxo de permeado assim como magnitude dos efeitos dos fatores estudados. Este resultado é atribuído ao fato de que o fluxo de permeado está fortemente relacionado às propriedades químicas das classes dos corantes e à sua interação com a membrana.

Em relação a taxa de rejeição de cor, observou-se que os fatores não são estatisticamente significativos ao nível de 95% de confiança dos experimentos dos dois corantes estudados. Vale ressaltar que o efeito de não significância nesse caso é válido, já que a taxa de rejeição de cor apresentou seus valores máximos em todos os experimentos para os dois corantes. Obteve-se o valor máximo de 99,34% para o corante preto reativo e 99,35% para o corante preto disperso, e os valores mínimos de 98,76% e 98,73% respectivamente com os corantes preto reativo e disperso. Esse fato pode ser explicado devido à alta hidrofobicidade da membrana de PTFE, que provoca que o efeito de repulsão das gotas da solução de alimentação sobre as membranas o que minimizou o contato tanto do corante reativo, como disperso com a

membrana, dificultando a sua permeação através dos poros da membrana. Dessa forma, foi observado que a rejeição de cor não depende das condições operacionais e sim das características das membranas e dos corantes.

O ajuste do modelo foi realizado por análise de variância (ANOVA), apresentada nas Tabelas 2 e 3 para os dois corantes avaliados. O valor de F calculado foi superior ao tabelado, portanto é válido com o nível de confiança de 95 % estudado.

Tabela 2: Análise de Variância para o corante preto reativo

Fator	Soma dos quadrados	Graus de Liberdade	Quadrado médio	Fcal	p
(1)Va (L)	149,313	1	149,313	4,81623	0,052923
Va (Q)	48,213	1	48,213	1,55516	0,240792
(2)Vp (L)	273,904	1	273,904	8,83505	0,013989
Vp (Q)	453,391	1	453,391	14,62456	0,003350
(3)Ta (L)	1105,862	1	1105,862	35,67066	0,000137
Ta (Q)	8,587	1	8,587	0,27698	0,610164
Va, Vp	155,761	1	155,761	5,02423	0,048881
Va,Ta	18,301	1	18,301	0,59032	0,460053
Vp,Ta	2,531	1	2,531	0,08165	0,780908
Erro	310,020	10	31,002		
<b>Total</b>	<b>3037,770</b>	<b>19</b>			

Tabela 3: Análise de Variância para o corante preto disperso

Fator	Soma dos quadrados	Graus de Liberdade	Quadrado médio	Fcal	p
(1)Va (L)	12,264	1	12,264	1,4059	0,274418
Va (Q)	52,590	1	52,590	6,0288	0,043762
(2)Vp (L)	519,286	1	519,286	59,5293	0,000115
Vp (Q)	83,425	1	83,425	9,5636	0,017505
(3)Ta (L)	1251,841	1	1251,841	143,5070	0,000006
Ta (Q)	4,175	1	4,175	0,4786	0,511362
Va, Vp	0,405	1	0,405	0,0464	0,835544
Va,Ta	0,180	1	0,180	0,0206	0,889827
Vp,Ta	85,805	1	85,805	9,8364	0,016464
Erro	61,062	7	8,723		
<b>Total</b>	<b>2040,195</b>	<b>16</b>			

Nas Equações 3 e 4 são apresentados os modelos de regressão obtidos para o fluxo de permeado para os efluentes de corante preto reativo e preto disperso, respectivamente. Somente as variáveis que apresentaram significância estatística são consideradas.

$$J = 42,37 + 4,19Vp + 8,43Ta + 5,57Vp^2 + 4,41VaVp \quad (3)$$

$$J = 40,23 + 6,20Vp + 9,60Ta + 2,11Va^2 + 2,76Vp^2 + 3,26VpTa \quad (4)$$

Para verificar a qualidade do ajuste do modelo linear calculou-se o coeficiente de determinação ( $R^2$ ). Tem-se:  $R^2 = 0,89794$  para o modelo referente ao corante preto reativo e  $R^2 = 0,97007$  para o modelo que corresponde ao corante preto disperso. A partir do modelo matemático observa-se que as variáveis estudadas e que se apresentaram significativamente importante são positivas, indicando que quando maior forem Ta, Va e Vp, maior será o fluxo de permeado.

Nas Figuras 4 e 5 são apresentados os gráficos de resíduos em torno da reta normal, respectivamente para modelos obtidos a partir dos resultados com os corantes reativo e disperso.

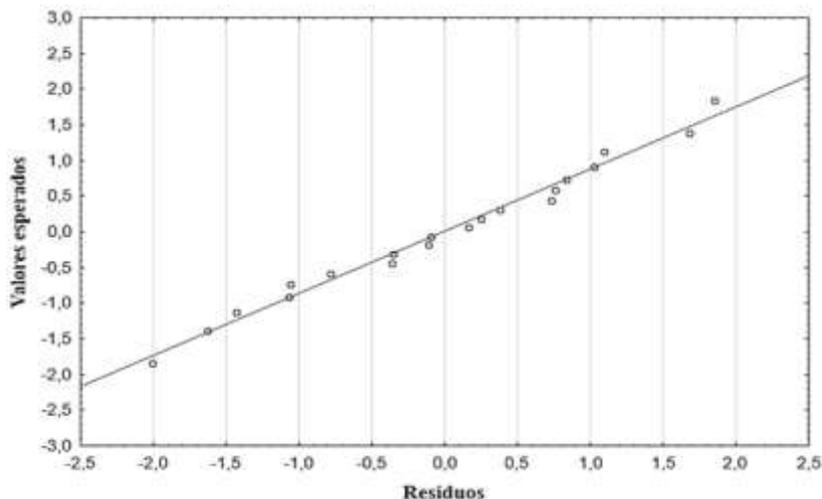


Figura 4: Distribuição dos resíduos em torno da reta normal para o modelo obtido com o corante reativo

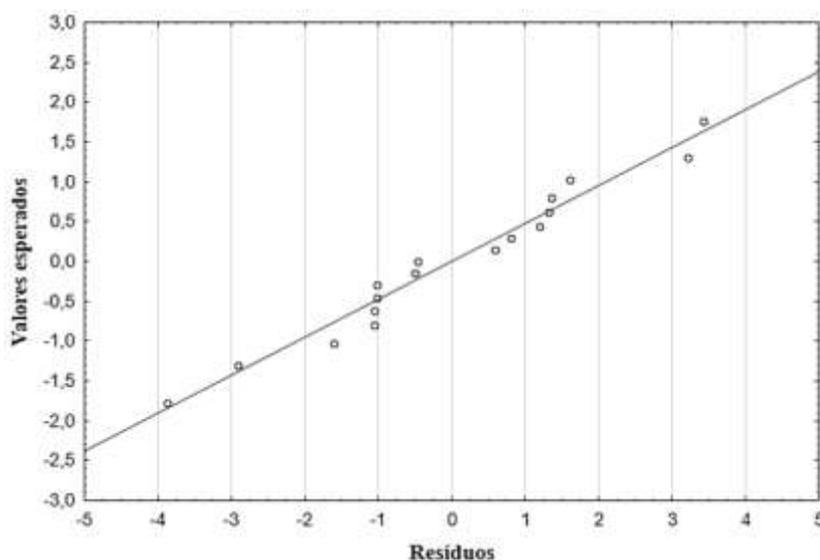


Figura 5: Distribuição dos resíduos em torno da reta normal para o modelo obtido com o corante disperso

Observa-se em ambos os casos que não há uma tendência positiva ou negativa dos dados, indicando natureza aleatória dos resíduos e variância constante.

### Conclusões

Os resultados obtidos permitiram identificar o comportamento do fluxo de permeado e da taxa de rejeição ao corante no processo de destilação por membranas aplicado a recuperação de águas residuais da indústria têxtil. Verificou-se que em relação a taxa de rejeição, nenhuma das variáveis estudadas apresentou significância ao nível de 95%, e que quando o fluxo de permeado a temperatura de alimentação é o efeito mais significativo para as duas classes de corantes preto estudadas: reativo e disperso.

O modelo obtido permite avaliar prever que quando maior forem temperatura e vazão de alimentação, bem como vazão de permeado, maior será o fluxo de permeado. Ainda, a identificação de que existe interação entre determinadas variáveis foi possível por meio da metodologia estatística empregada.

## Referências

- [1] Mokhtar, N. M.; Lau, W. J.; ismail, A. F.; Youravong, W.; Khongnakorn, W.; Lertwittayanon, K. Performance evaluation of novel PVDF–Cloisite 15A hollow fiber composite membranes for treatment of effluents containing dyes and salts using membrane distillation. *Rsc Advances*, vol. 5, n. 48, pp.38011-38020, (2015)
- [2] Dilaver, M.; Hocaoglu, S. M.; Soydemir, G.; Dursun, M.; Keskinler, B.; Koyuncu, İ.; Ağtaş, M. Hot wastewater recovery by using ceramic membrane ultrafiltration and its reusability in textile industry. *Journal of Cleaner Production*, vol. 171, pp. 220–233, (2018)
- [3] Hitsov, I.; Eykens, L.; Schepper, W. De; Sitter, K. De; Dotremont, C.; Nopens, I. Full-scale direct contact membrane distillation (DCMD) model including membrane compaction effects. *Journal of Membrane Science*, vol. 524, pp. 245-256, (2017)
- [4] Ramlow, H; Machado, R. A. F.; Bierhalz, A. C. K; Marangoni, C. Dye synthetic solution treatment by direct contact membrane distillation using commercial membranes. *Environmental Technology*, pp.1-13, (2018)
- [5] Hayer, H; Bakhtiari, O; Mohammadi, T. Simulation of momentum, heat and mass transfer in direct contact membrane distillation: A computational fluid dynamics approach. *Journal of Industrial and Engineering Chemistry*, vol. 21, pp. 1379-1382, (2015)
- [6] Ali, A.; Macedonio, F.; Drioli, E.; Aljlil, S.; Alharbi, O. A. Experimental and theoretical evaluation of temperature polarization phenomenon in direct contact membrane distillation. *Chemical Engineering Research and Design*, vol. 91, pp. 1966–1977, (2013)



# Modelagem Empírica de uma Unidade Piloto de Destilação de Película Descendente Assistida por Termossifão: Determinação de Funções de Transferência

**José Luiz Francisco Alves, Valdemar Francisco da Silva Filho,**  
**Ricardo Antonio Francisco Machado, Cintia Marangoni**

Programa de Pós-Graduação em Engenharia Química, Departamento de Engenharia Química e Engenharia de Alimentos, Campus Reitor João David Ferreira Lima, Universidade Federal de Santa Catarina, 88040-900, Florianópolis, SC

E-mail: zeluz\_alves@hotmail.com; valdemardasilvafilho@gmail.com; ricardo.machado@ufsc.br; cintia.marangoni@ufsc.br

**Resumo:** *Recentemente, colunas de destilação de película descendente com baixo consumo energético e dimensões reduzidas foram propostas, como uma nova configuração energeticamente intensificada de destilação. Porém o comportamento dinâmico destas novas colunas de destilação intensificadas energeticamente ainda não avaliado. Este trabalho propõe a identificação de um modelo empírico (funções de transferência) que represente de forma adequada a resposta transiente não-linear de uma unidade piloto multitubular de destilação de película descendente assistida por termossifão. Uma mistura modelo de etanol e água foi utilizada nos testes experimentais em condição atmosférica. Experimentos sob condições transientes foram realizados a partir de perturbações do tipo degrau na temperatura pré-aquecimento da corrente de alimentação, na potência fornecida para a câmara de vapor, e na composição mássica de etanol na corrente de alimentação. Para determinação dos parâmetros das funções de transferências foi utilizado o “Método de Otimização” que é baseado na técnica de regressão não-linear. Um modelo empírico simplificado de primeira ordem foi capaz de descrever adequadamente a resposta dinâmica não-linear das temperaturas da base e do topo da unidade de um sistema intensificado de destilação de película descendente. O maior ganho do processo em estado estacionário foi observado para a perturbação na composição mássica de etanol na corrente de alimentação, porém com um efeito negativo sobre a resposta dinâmica da unidade. Um novo estado estacionário é alcançado mais lentamente para o distúrbio na temperatura de pré-aquecimento. Como uma consequência da forma de operação e das características construtivas do sistema estudado, a resposta dinâmica deste sistema ocorre instantaneamente, e dessa forma o tempo morto pode ser negligenciado. Estes resultados são ferramentas úteis no entendimento das características essenciais do processo durante condições transientes, e também no projeto apropriado de estratégias de controle para esta nova configuração energeticamente intensificada de destilação.*

**Palavras-chave:** *Sistema energeticamente intensificado; Destilação de película descendente; Resposta transiente; Funções de transferência; Dinâmica de primeira ordem.*

## 1. Introdução

A destilação é uma operação unitária de transferência de calor e massa tradicionalmente com baixa eficiência termodinâmica. Essa limitação tem motivado uma preocupação em relação a melhoria na eficiência da operação de destilação, seja aumentando-se o rendimento de um determinado produto ou diminuindo-se a energia necessária para efetivar a separação. Ao longo dos anos, várias configurações de colunas intensificadas energeticamente de destilação foram propostas visando a redução do consumo energético do processo de destilação. Processos intensificados energeticamente possuem várias vantagens em relação à destilação convencional como redução do tamanho dos equipamentos e plantas industriais (baixo investimento de capital), melhorias na eficiência e segurança dos processos, e redução do consumo de energia (baixo custo operacional) [1].

Recentemente, foi proposta uma nova configuração intensificada energeticamente de destilação de película descendente que emprega um sistema de fornecimento de calor do tipo câmara de vapor (termossifão bifásico) para efetuar o processo de separação com melhor eficiência energética [2]. Na configuração proposta, a película de líquido formada pela corrente de alimentação escoar em fluxo descendente e em contracorrente o vapor é gerado, enquanto que a temperatura da parede do tubo de destilação é mantida aquecida e radialmente constante através do calor da câmara de vapor, termossifão bifásico [3]. Algumas vantagens dessa tecnologia é operar em pressão atmosférica, dimensões reduzidas e menor consumo energético [4,5].

Para a configuração intensificada energeticamente em questão, os poucos estudos relatados na literatura são baseados em análises do estado estacionário, a fim de avaliar o desempenho de separação deste novo sistema intensificado de destilação de película descendente frente a diferentes misturas e condições experimentais [3,4,6]. No entanto, em uma operação em escala comercial, o processo de destilação pode estar sujeito a flutuações ou perturbações nas variáveis de entrada. Portanto, é essencial avaliar o comportamento dinâmico das variáveis de saída deste processo com base nas mudanças nas variáveis de entrada em função do tempo. Com relação a colunas de destilação convencionais, várias pesquisas tratam do desempenho dinâmico sob condições operacionais transitórias [7–10], entretanto, para sistemas intensificados energeticamente de destilação de filme descendente o comportamento dinâmico ainda é desconhecido.

O principal objetivo deste trabalho foi determinar em malha aberta os parâmetros das funções de transferência a partir da resposta transiente das temperaturas do topo e da base de uma unidade intensificada de destilação de película descendente assistida por termossifão frente a distúrbios aplicados na temperatura pré-aquecimento da corrente de alimentação, na potência fornecida para a câmara de vapor, e na composição mássica de etanol na corrente de alimentação. Estes parâmetros serão ferramentas úteis no entendimento das características essenciais do processo durante condições transientes, e também no projeto adequado de estratégias de controle para este novo processo intensificado de destilação.

## 2. Material e Métodos

### 2.1 Descrição da unidade de destilação piloto

Os experimentos sob condições transientes foram realizados em uma unidade piloto multitubular (nove tubos verticais de destilação de película descendente) equipada com um termossifão bifásico, processando uma mistura de etanol e água. Esta unidade multitubular intensificada energeticamente de destilação de película descendente foi construída em aço 306 e opera sob regime contínuo, pressão atmosférica e uma condição de refluxo externo zero. Alguns periféricos foram incorporados a unidade piloto multitubular como tanque acumulador, tanque pulmão, tanque de mistura, condensador, termossifão bifásico, bomba helicoidal e banho termostático.

A mistura a ser separada é pré-aquecida por meio de um banho termostático modelo USH 400 (Lauda, Lauda-Königshofen, Alemanha), e depois alimentada no topo da coluna de destilação por meio de uma bomba de cavidade progressiva (helicoidal) de modelo NM021BY02S12B (Netzsch, Pomerode-SC, Brasil) e então forma-se um filme líquido descendente que escorre sobre a superfície interna de cada um dos nove tubos verticais de transferência de calor e massa, cada tubo possui 1 m de altura, 26,4 mm de diâmetro e com 3 mm de espessura. Esse filme escoar em contracorrente com o vapor gerado a uma determinada pressão e temperatura, enquanto que radialmente a temperatura da parede que envolve os tubos verticais de transferência de calor e massa é mantida constante por meio do evaporador. Uma fotografia da unidade piloto multitubular incluindo os principais periféricos associados a essa unidade e o arranjo dos nove tubos de destilação em relação à câmara de vapor são apresentados na Figura 1.

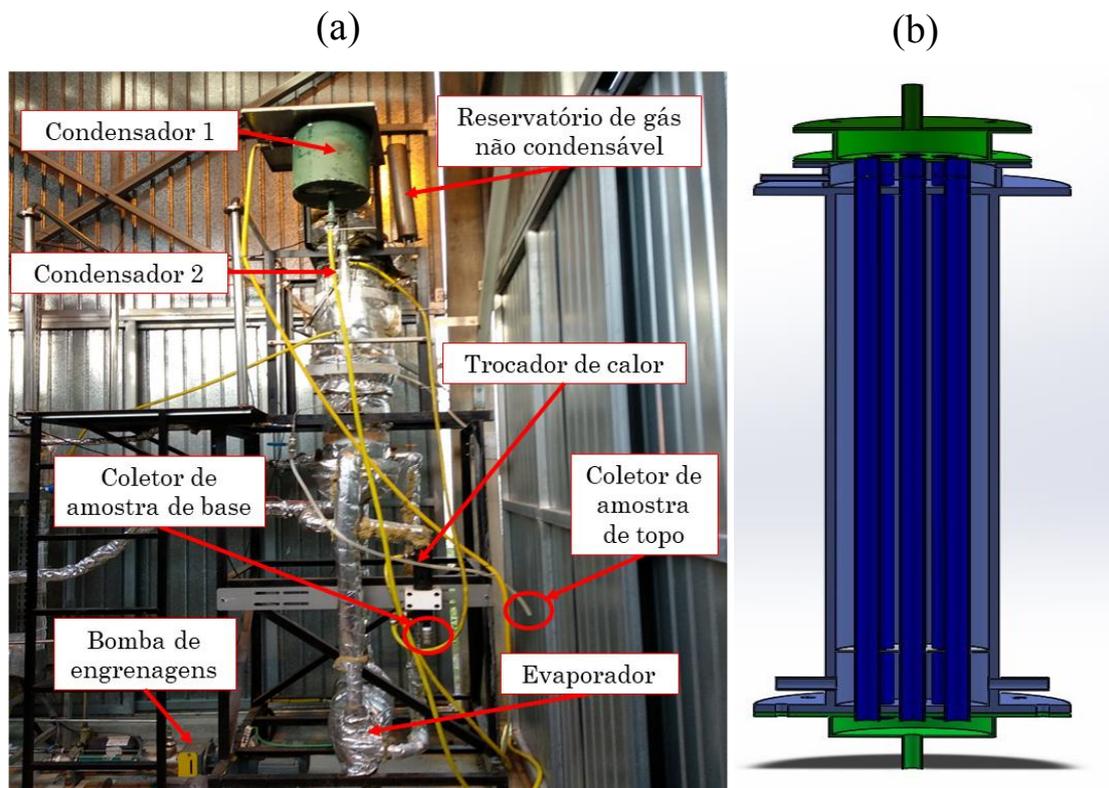


Figura 1: Fotografia da unidade piloto multitubular de filme descendente incluindo os principais periféricos (a) e o arranjo dos nove tubos de destilação na unidade piloto multitubular de filme descendente em relação à câmara de vapor (b).

O fornecimento de calor da unidade piloto multitubular de filme descendente é realizado por meio de uma câmara de vapor do tipo termossifão empregando-se um evaporador que utiliza vapor de água como fluido térmico. O evaporador é formado por resistências elétricas totalizando uma potência máxima de 36 kW. Devido às características do termossifão, o modo de fornecimento de energia tanto pode ser realizado isotermicamente quanto impondo um perfil de temperatura ao longo do tubo de destilação (ver mais detalhes em Da Silva et al. [3]). Neste trabalho apenas o modo isotérmico de fornecimento de energia foi avaliado. Com intuito de evitar a perda de calor para o ambiente a unidade e seus periféricos são isolados utilizando lã de rocha.

Pontos de aquisição de temperatura utilizando termopares do tipo K (modelo TT-K-36, Omega, Stamford, Estados Unidos) foram instalados no topo e na base da coluna de destilação. Estas temperaturas foram monitoradas em tempo real por meio de um sistema de aquisição de dados modelo CR1000, combinado com multiplexador modelo AM25T (ambos da Campbell Scientific, Utah, Estados Unidos), e utilizadas na identificação dos parâmetros empíricos das funções de transferência.

## 2.2 Procedimento experimental em malha aberta

Perturbações do tipo degrau aplicadas na temperatura pré-aquecimento da corrente de alimentação, na potência fornecida para a câmara de vapor, e na composição mássica de etanol na corrente de alimentação foram empregadas para estimar os parâmetros empíricos das funções de transferência. Uma “condição de referência” foi utilizada como “ponto inicial do processo”: 87,50 °C (temperatura de pré-aquecimento -  $T_{P-A}$ ); 245,00 kg·h<sup>-1</sup> (vazão mássica da corrente de alimentação -  $\dot{m}_A$ ); 33,70% (potência fornecida para a câmara de vapor -  $P_{EVAP}$ ); e 13,00% (composição mássica de etanol na corrente de alimentação -  $X_{EtOH,A}$ ).

Para identificação dos parâmetros das funções de transferência, foi considerado um período de 0,50 h antes do distúrbio (referido como “ponto inicial do processo”) e os ensaios experimentais foram conduzidos por aproximadamente 2,00 h após os distúrbios aplicados. Um período total de 2,50 h foi utilizado na identificação das funções de transferência a partir do comportamento transiente do processo. O comportamento dinâmico da unidade piloto de destilação em termos da temperatura da base ( $T_B$ ) e a da temperatura do topo ( $T_T$ ) foram estudadas a partir da aplicação de perturbações positivas na temperatura de pré-aquecimento da corrente de alimentação (+20,00 °C), na potência fornecida para a câmara de vapor (+25,60%) e na composição mássica de etanol na corrente de alimentação (+8,55%).

### 2.3 Método para obtenção dos parâmetros das funções de transferência de primeira ordem

Os modelos empíricos de primeira ordem foram obtidos a partir dos dados experimentais utilizando a técnica de regressão não-linear. O modelo de primeira ordem com tempo morto pode ser expresso seguindo a Equação 1.

$$\tau_p \cdot \frac{dy(t)}{dt} = -y(t) + K_p \cdot u(t - \theta) \quad (1)$$

Onde  $y(t)$  é a variável dependente de saída em função do tempo,  $u(t)$  é a variável independente de entrada em função do tempo,  $\tau_p$  é a constante de tempo do processo,  $K_p$  é o ganho do processo em estado estacionário e  $\theta$  é tempo morto do processo.

A Equação 2 pode ser utilizada para estimar o ganho do processo em estado estacionário e a constante de tempo do processo, ambos do modelo empírico de primeira ordem.

$$u_0 = e^{\frac{-\Delta t}{\tau_p}} \cdot (y_{j-1} - y_0) + \left(1 - e^{\frac{-\Delta t}{\tau_p}}\right) \cdot K_p \cdot (u_{j-\theta-1} - u_0) + y_0 \quad (2)$$

Onde  $\Delta t$  é o comprimento do passo de tempo,  $y_0$  é o valor de saída inicial,  $u_0$  é o valor de entrada inicial,  $y_{j-1}$  é o valor de entrada do passo anterior,  $u_{j-1}$  é o valor de saída do passo anterior,  $\theta$  é tempo morto do processo.

Os parâmetros desconhecidos do modelo de primeira ordem representado pela Equação 2 foram identificados a partir da minimização da soma dos quadrados dos resíduos (SQres). A ferramenta Solver do Microsoft Excel 2016 foi empregada para minimizar o SQres, que é representado pela Equação 3 [11,12].

$$SQres = \sum_{j=1}^N (y_j - \hat{y}_j)^2 \quad (3)$$

Onde o subscrito  $j$  na Equação 3 representa o ponto experimental,  $N$  é o número total de pontos experimentais utilizados,  $y_j$  é o dado experimental de saída e  $\hat{y}_j$  é o dado previsto pelo modelo correspondente ao dado experimental de saída.

A validação dos parâmetros do modelo empírico (funções de transferência de primeira ordem) foi realizada a partir da comparação entre a resposta transiente experimental e a resposta transiente prevista pelo modelo, esta comparação é mensurada por meio do ajuste ou “Fit” (Equação 4).

$$Fit = 1 - \frac{\sum_{j=1}^N (y_j - \hat{y}_j)^2}{\sum_{j=1}^N (y_j - \bar{y})^2} \quad (4)$$

Onde o subscrito  $j$  na Equação 4 representa o ponto experimental,  $N$  é o número total de pontos experimentais utilizados,  $y_j$  é o dado experimental de saída,  $\hat{y}_j$  é o dado previsto pelo modelo correspondente ao dado experimental de saída e  $\bar{y}$  é a média de todos os dados experimentais.

### 3. Resultados e Discussão

Na Figura 2 é apresentado os perfis da resposta transiente experimental e da prevista pelo modelo empírico. É possível notar um ajuste satisfatório entre o comportamento dinâmico experimental e a função de transferência de primeira ordem. O modelo empírico de primeira ordem foi adequado para descrever a resposta dinâmica para as temperaturas da base ( $T_B$ ) e do topo ( $T_T$ ) da coluna de destilação a partir do distúrbio na temperatura de pré-aquecimento apresentando um valor de ajuste maior que 0,9583 (ver Figura 2 (a)). Como pode ser observado na Figura 2 (b), para o distúrbio na potência fornecida para a câmara de vapor, o modelo dinâmico de primeira ordem também foi adequado para representar a resposta transiente das variáveis de saída. Para este caso, o coeficiente de ajuste foi maior que 0,9880. Os coeficientes de ajuste, na Figura 2 (c), foram maiores que 0,9391 para o distúrbio na composição mássica de etanol na corrente de alimentação em relação as temperaturas da base e do topo avaliadas.

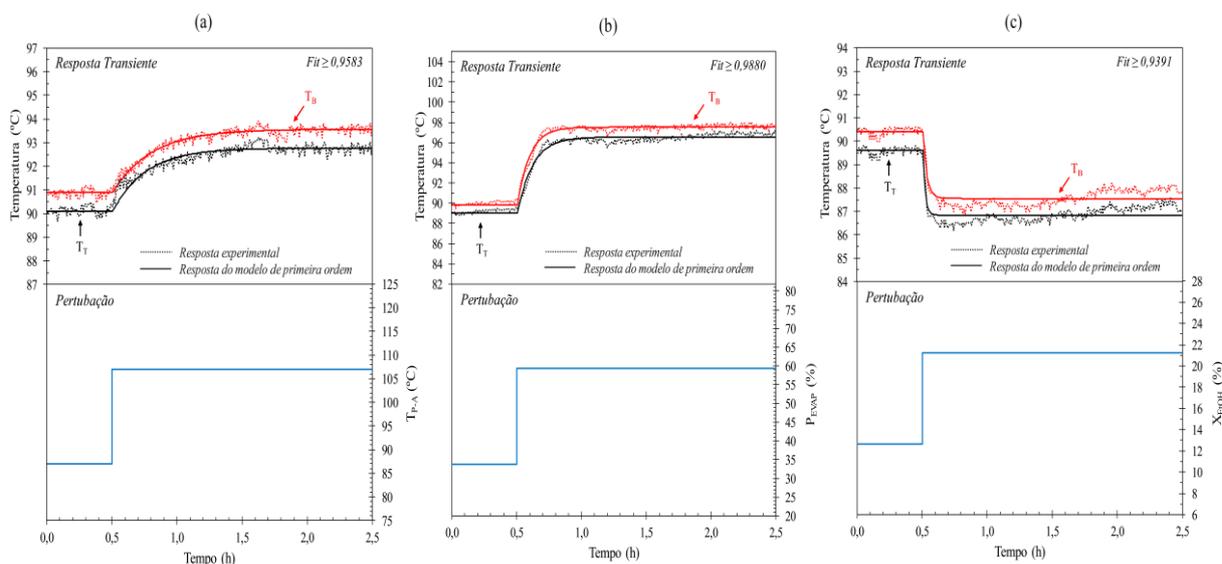


Figura 2: Comparação entre a resposta dinâmica experimental (linha tracejada,  $\cdots$ ) e a resposta prevista pelas funções de transferência de primeira ordem (linha sólida,  $—$ ) para os distúrbios na temperatura de pré-aquecimento (a), na potência fornecida à câmara de vapor (b), e na composição mássica de etanol na corrente de alimentação (c).

É importante notar também que após a aplicação das perturbações, a resposta dinâmica do sistema é instantânea (conforme pode ser visualizado na Figura 2). Pode-se então considerar que o valor do tempo morto é zero. Características deste novo sistema intensificado de destilação, como baixo tempo de residência (em torno de 7,88 s), e o fornecimento de calor da unidade piloto que é realizado isotermicamente ao longo da altura da coluna de destilação, são responsáveis por esse comportamento. O valor calculado para o tempo de residência aparece ser

consistente quando comparado com os resultados de Stefanov e Hoo [13], que reportaram o valor de tempo de residência em torno de 7,00 s para um evaporador de película descendente.

Na Tabela 1 são resumidos os parâmetros das funções de transferência de primeira ordem obtidos a partir da regressão não-linear utilizando os dados experimentais de resposta transiente para cada distúrbio aplicado. Todos os valores para a constante de tempo do processo são apresentados em minutos. A partir da análise dos valores das constantes de tempo do processo é possível concluir que o distúrbio na temperatura de pré-aquecimento apresenta um longo período transiente, ou seja, uma resposta dinâmica lenta, enquanto que a resposta dinâmica para o distúrbio na composição mássica de etanol na corrente de alimentação apresenta um curto período transiente, ou seja, uma resposta dinâmica mais rápida. Para o distúrbio na temperatura de pré-aquecimento o valor do ganho do processo em estado estacionário foi próximo a 0,13 °C por unidade de temperatura de pré-aquecimento. O sinal negativo observado no ganho do processo em estado estacionário para o distúrbio na composição mássica de etanol na corrente de alimentação indica uma interação negativa entre o distúrbio e a resposta transiente das temperaturas da base e do topo.

Variável	Distúrbio								
	$T_{PA}$ (°C)			$P_{EVAP}$ (%)			$X_{EtOH,A}$ (%)		
	$K_p$	$\tau_p$	Fit	$K_p$	$\tau_p$	Fit	$K_p$	$\tau_p$	Fit
$T_T$ (°C)	0,1333	15,3869	0,9583	0,2945	7,0288	0,9880	-0,3225	1,0660	0,9399
$T_B$ (°C)	0,1333	18,0466	0,9722	0,3018	5,6467	0,9926	-0,3333	1,8190	0,9391

Tabela 1: Parâmetros do modelo dinâmico de primeira ordem a partir da resposta experimental e ajuste para cada distúrbio aplicado.

Dessa forma, um modelo empírico simplificado pode ser uma ferramenta adequada para descrever a resposta dinâmica não-linear de um sistema intensificado de destilação de destilação de película descendente. Este modelo empírico de primeira ordem é baseado somente nas relações de causa e efeito entre as variáveis de entrada e de saída do sistema, e não necessita de informações fenomenológicas detalhadas sobre o processo.

## 4. Conclusões

Com base na avaliação dos ajustes do modelo empírico utilizado é possível concluir que a unidade intensificada de destilação com nove tubos de destilação operando com aquecimento isotérmico apresentou o comportamento dinâmico de um sistema não linear de primeira ordem. Além disso, os parâmetros das funções de transferência encontradas podem fornecer informações de entrada para o projeto adequado de um sistema de controle para esta nova configuração energeticamente intensificada de destilação.

Características dinâmicas essenciais do sistema estudado foram elucidadas a partir da interpretação dos valores calculados para os parâmetros das funções de transferência, tais quais um novo estado estacionário é alcançado mais lentamente para o distúrbio na temperatura de pré-aquecimento; o tempo morto pode ser negligenciado devido a resposta dinâmica instantânea do sistema estudado; existe uma interação negativa entre o distúrbio na composição mássica de etanol na corrente de alimentação e a resposta transiente das temperaturas da base e do topo; e uma resposta dinâmica rápida (curto período transiente) para o distúrbio na composição mássica de etanol na corrente de alimentação.

## Agradecimentos

Os autores agradecem o suporte financeiro disponibilizado a partir da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), e Centro de Pesquisas e Desenvolvimento Leopoldo Américo Miguez de Mello (CENPES/PETROBRAS).

## Referências

- [1] N. Sharma, K. Singh, Control of Reactive Distillation Column: A Review, *Int. J. Chem. React. Eng.*, vol. 8, pp. 1–55, (2010). doi:10.2202/1542-6580.2260.
- [2] K.W. Milanez, M.B.H. Mantelli, L.C.M. Oliveira, M.C. Santos, R.A.F. Machado, C. Marangoni, A.P. Meneguelo, Equipamento e Processo de Destilação com Controle de Energia pela Demanda. Brasil. Número do registro: BR10201402976850, Instituição de registro: INPI - Instituto Nacional da Propriedade Industrial, (2014).
- [3] V.F. da Silva Filho, J.L.F. Alves, G.F. Reus, R.A.F. Machado, C. Marangoni, A. Bolzan, Experimental evaluation of the separation of aromatic compounds using falling film distillation on a pilot scale, *Chem. Eng. Process. - Process Intensif.*, vol. 130, pp. 296–308, (2018). doi:10.1016/j.cep.2018.06.026.
- [4] M. V Querino, C. Marangoni, R.A.F. Machado, Indirect Series of Falling Film Distillation Column to Process Synthetic Naphtha, *Chem. Eng. Trans.*, vol. 69, pp. 679–684, (2018). doi:10.3303/CET1869114.
- [5] C. Marangoni, A.P. Meneguelo, J.G. Teleken, I.G.B. Parisotto, L.O. Werle, R.A.F. Machado, M.C. dos Santos, A.O. Gomes, L.C. Medina, New Configuration of a Distillation Process with Reduced Dimensions, *Chem. Eng. Trans.*, vol. 24, pp. 799–804, (2011). doi:10.3303/CET1124134.
- [6] C. Marangoni, A.P. Meneguelo, J.G. Teleken, L.O. Werle, K.W. Milanez, M.B.H. Mantelli, M.B. Quadri, A. Bolzan, M.C. dos Santos, L.C. Medina, R.A.F. Machado, Falling film distillation column with heat transfer by means of a vapor chamber – part I: isothermal operation, *Chem. Eng. Commun.*, vol. xx, pp. 1–12, (2019). doi:10.1080/00986445.2018.1542250.
- [7] C. Marangoni, R.A.F. Machado, A. Bolzan, Distributed heat supply for distillation control to reduce feed composition disturbance effects, *Chem. Eng. Technol.*, vol. 36, pp. 2071–2079, (2013). doi:10.1002/ceat.201300266.
- [8] C. Marangoni, G.N. Mello, A. Bolzan, R.A.F. Machado, Dynamic Study of Distillation Column Operated with Tray Heat Source Combined with Reboiler, *Chem. Eng. Commun.*, vol. 203, pp. 364–371, (2016). doi:10.1080/00986445.2014.1001888.
- [9] C. Marangoni, E.J. Zierhut, I.G.B. Parisotto, R.A.F. Machado, A. Bolzan, Influence of the Location of the Internal Temperature Control Loop on the Performance of the Dual Temperature Control for Feed Temperature Disturbance, *Chem. Biochem. Eng. Q. J.*, vol. 30, pp. 411–418, (2017). doi:10.15255/CABEQ.2016.871.
- [10] G.N. Mello, R.A.F. Machado, C. Marangoni, Use of stripping or rectifying trays for a distributed control strategy in distillation column, *Chem. Eng. Trans.*, vol. 43, pp. 1099–1104, (2015). doi:10.3303/CET1543184.
- [11] J.L.F. Alves, J.C.G. Da Silva, V.F. da Silva Filho, R.F. Alves, W.V. de Araujo Galdino, S.L.F. Andersen, R.F. De Sena, Determination of the Bioenergy Potential of Brazilian Pine-Fruit Shell via Pyrolysis Kinetics, Thermodynamic Study, and Evolved Gas Analysis, *BioEnergy Res.*, vol. 12, pp. 168–183, (2019). doi:10.1007/s12155-019-9964-1.
- [12] J.L.F. Alves, J.C.G. da Silva, V.F. da Silva Filho, R.F. Alves, W.V. de Araujo Galdino, R.F. De Sena, Kinetics and thermodynamics parameters evaluation of pyrolysis of invasive aquatic macrophytes to determine their bioenergy potentials, *Biomass and Bioenergy.*, vol. 121, pp. 28–40, (2019). doi:10.1016/j.biombioe.2018.12.015.
- [13] Z. Stefanov, K.A. Hoo, A distributed-parameter model of black liquor falling film evaporators. Part I. Modeling of a single plate, *Ind. Eng. Chem. Res.*, vol. 42, pp. 1925–1937, (2003). doi:10.1021/ie020483a.



# O Método de Levenberg-Marquardt para Problemas de Otimização de Menor Valor Ordenado

Anderson E. Schwertner, Emerson V. Castelani,

Departamento de Matemática, UEM  
87020-900, Maringá, PR

E-mail: andersonschwertner@hotmail.com, evcastelani@uem.br,

Francisco N. Calmon Sobral

Departamento de Matemática, UEM  
887020-900, Maringá, PR

E-mail: fncsobral@uem.br.

**Resumo:** *Proposto inicialmente por Andreani et al [2], o problema de Otimização de Menor Valor Ordenado (LOVO) consiste em, dadas  $r$  funções reais definidas em  $\Omega \subset \mathbb{R}^n$  e um inteiro  $p$ , com  $1 \leq p \leq r$ , minimizar a soma destas funções, tomados os  $p$  menores valores. Neste sentido, apresentamos uma nova estratégia para solução de problemas LOVO irrestritos com convergência para pontos fracamente críticos, via método de Levenberg-Marquardt com Busca Linear, assim como resultados de convergência do método. Também são sugeridos dois novos parâmetros de damping e uma estratégia híbrida. Por fim, são apresentados dados sobre o desempenho do novo método, em comparação com uma técnica já existente.*

**Palavras-chave:** *Método de Levenberg-Marquardt, Otimização de Menor Valor Ordenado, Quadrados Mínimos Não Linear*

## 1 O Método de Levenberg-Marquardt

Seja  $R : \mathbb{R}^n \rightarrow \mathbb{R}^m$  uma função residual e considere o problema de quadrados mínimos não linear

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} \|R(x)\|_2^2.$$

Dado um ponto inicial  $x_0$ , o método de Levenberg-Marquardt visa encontrar, a cada iteração  $k$ , uma direção  $d(\lambda_k) \in \mathbb{R}^n$  que seja solução do sistema linear

$$(J(x_k)^T J(x_k) + \lambda_k I) d(\lambda_k) = -J(x_k)^T R(x_k), \quad (1)$$

onde  $J(x) \in \mathbb{R}^{m \times n}$  é a matriz jacobiana de  $R(x)$  e  $\lambda \in \mathbb{R}_+$  é uma constante denominada parâmetro de Levenberg-Marquardt ou parâmetro de *damping*. A direção de Levenberg-Marquardt  $d(\lambda)$  também será denotada por  $d^{LM}$ .

Diferentemente do que ocorre no método de Gauss-Newton, o qual pode não estar bem definido caso a matriz  $J(x_k)^T J(x_k)$  não seja definida positiva, se  $\lambda_k > 0$ , a correção  $\lambda_k I$  aplicada a matriz  $J(x_k)^T J(x_k)$ , torna esta não singular. Sendo assim, em cada iteração  $k$  o sistema linear (1) tem solução e esta é única. Portanto, o método está bem definido.

Vale salientar ainda que se  $\lambda_k = 0$ , a direção  $d^{LM}$  obtida pelo método de Levenberg-Marquardt coincide com a direção de Gauss-Newton  $d^{GN}$ . Neste caso, se  $J(x_k)^T J(x_k)$  é uma boa aproximação para a matriz hessiana, então o método apresentará taxa de convergência quadrática, partindo de um ponto suficientemente próximo da solução. Entretanto, se  $\lambda_k$  tender ao infinito, teremos que a direção  $d^{LM}$  tenderá à direção de máxima descida  $-\nabla f(x_k)$ , a qual denotaremos por  $-g_k$  [7, Teoremas 2 e 3].

## 2 O Problema de Otimização de Menor Valor Ordenado

Sejam dadas  $r$  funções  $F_i : \Omega \rightarrow \mathbb{R}$ ,  $i = 1, \dots, r$ , onde  $\Omega \subseteq \mathbb{R}^n$  e  $p \in \{1, 2, \dots, r\}$  um inteiro. O conjunto  $\{1, 2, \dots, r\}$  contém exatamente  $m = \binom{r}{p} = \frac{r!}{p!(r-p)!}$  subconjuntos distintos  $C_1, C_2, \dots, C_m$  com cardinalidade  $p$ . Construímos então a família de funções

$$f_i(x) = \sum_{j \in C_i} F_j(x), \forall i = 1, \dots, m, \forall x \in \Omega$$

e a função

$$f_{min}(x) = \min\{f_1(x), f_2(x), \dots, f_m(x)\}.$$

O problema de Otimização de Menor Valor Ordenado (LOVO) consiste em

$$\min_{x \in \Omega} f_{min}(x).$$

Considere ainda as seguintes definições:

**Definição 2.1.** [2, p.4] Para cada  $x \in \Omega$ , o conjunto  $I_{min}(x)$  é o conjunto de índices das funções  $f_i$  que minimizam a função  $f_{min}$  no ponto  $x$ , isto é,

$$I_{min}(x) = \{i \in 1, \dots, m \mid f_i(x) = f_{min}(x)\}.$$

**Definição 2.2.** [2, p.6] Seja  $x_* \in \Omega$ . Dizemos que:

- $x_*$  é fortemente crítico se, para todo  $i \in I_{min}(x_*)$ ,  $x_*$  é ponto estacionário de  $f_i$ .
- $x_*$  é fracamente crítico se existe  $i \in I_{min}(x_*)$  tal que  $x_*$  é ponto estacionário de  $f_i$ .

Note que, por definição, o domínio da função  $f_{min}$  pode conter pontos nos quais a função não é diferenciável. Deste modo, o problema LOVO é um problema de otimização não-suave. Contudo, conforme afirmam Andreani et al [2, p.03], uma alternativa viável para sua solução é a adaptação de métodos essencialmente suaves. Assim, a cada vez que  $\nabla f_{min}(x)$  é requerida, podemos substituí-la por  $\nabla f_i(x)$ , com  $i \in I_{min}(x)$ . De modo análogo, podemos proceder desta maneira caso o método exija informações acerca da hessiana de  $f_{min}$ . Tal estratégia pode ser catastrófica se aplicada a problemas não-suaves, entretanto, neste caso específico, as consequências são menos graves e, necessariamente, temos convergência para pontos fracamente críticos [2, p.06].

## 3 Um algoritmo de Levenberg-Marquardt para problemas LOVO irrestritos

Dadas  $r$  funções  $F_1, F_2, \dots, F_r$ , continuamente diferenciáveis em  $\mathbb{R}^n$ , isto é,  $F_i \in C^1(\mathbb{R}^n)$ ,  $i = 1, 2, \dots, r$ , e um inteiro  $p \in \{1, 2, \dots, r\}$ , considere a função vetorial  $R : \mathbb{R}^n \rightarrow \mathbb{R}^r$ , cujas entradas correspondem a cada uma das funções  $F_i$ ,  $i = 1, \dots, r$ , e sua matriz jacobiana  $J(x)$ :

$$R(x) = \begin{bmatrix} F_1(x) \\ F_2(x) \\ \vdots \\ F_r(x) \end{bmatrix}, J(x) = \begin{bmatrix} \frac{\partial F_1(x)}{\partial x_1} & \dots & \frac{\partial F_1(x)}{\partial x_n} \\ \frac{\partial F_2(x)}{\partial x_1} & \dots & \frac{\partial F_2(x)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_r(x)}{\partial x_1} & \dots & \frac{\partial F_r(x)}{\partial x_n} \end{bmatrix}.$$

Assim, dado  $i \in I_{min}(x)$ , denotamos:

- O vetor coluna  $R^i(x) := R|_{C_i}(x)$ , isto é,  $[R^i(x)]_j = F_j(x)$  para todo  $j \in C_i$ .
- A função  $f^i(x) = \frac{1}{2} \|R^i(x)\|_2^2$ .

- A matriz  $J^i(x) := J|_{C_i}(x)$ , isto é,  $[J^i(x)]_{j,l} = \frac{\partial F_j(x)}{\partial x_l}$  para todo  $j \in C_i$  e  $l = 1, \dots, n$ .

Com as notações acima,  $\nabla f^i(x) = J^i(x)^T R^i(x)$ .

O algoritmo a seguir corresponde a uma adaptação do método de busca linear com Levenberg-Marquardt, no contexto do problema LOVO irrestrito.

---

**Algoritmo 1:** LOVO LEVENBERG-MARQUARDT IRRESTRITO COM CONVERGÊNCIA PARA PONTOS FRACAMENTE CRÍTICOS

---

**Entrada:**  $R(x)$ ,  $J(x)$ , uma aproximação inicial  $x_0 \in \mathbb{R}^n$ , uma constante  $\alpha \in (0, 1)$  e uma tolerância  $\varepsilon > 0$ .

**Saída:**  $x_* \in \mathbb{R}^n$ , ponto fracamente crítico de  $f_{min}(x)$ .

```

1 início
2    $k \leftarrow 0$  ;
3   repita
4     Escolha  $i = v(k) \in I_{min}(x_k)$  ;
5     Calcule  $\lambda_k$ ;
6     Obtenha  $d^i(\lambda_k)$  solução de:
           
$$(J^i(x_k)^T J^i(x_k) + \lambda_k I) d^i(\lambda_k) = -J^i(x_k)^T R^i(x_k); \quad (2)$$

           Faça  $t = 1$ ;
7     enquanto  $f^i(x_k + t d^i(\lambda_k)) > f^i(x_k) + \alpha t \nabla f^i(x_k)^T d^i(\lambda_k)$  faça
8       | Escolha novo  $t \in [0.1t, 0.9t]$ .
9     fim
10    Faça:
           
$$x_{k+1} \leftarrow x_k + t d^i(\lambda_k);$$

           
$$k \leftarrow k + 1;$$

11   até  $\|\nabla f^i(x_k)\|_2 \leq \varepsilon$ ;
12 fim
13 retorna  $x_k$ 

```

---

**Teorema 3.1.** *Seja  $x_*$  ponto limite de uma sequência gerada pelo Algoritmo 1. Suponha que  $\lim_{k \in K} x_k = x_*$  e um mesmo índice  $i = v(k) \in I_{min}(x_k)$  é escolhido para infinitos índices  $k \in K$ , então:*

1.  $i \in I_{min}(x_*)$ ;
2.  $\|d^i(\lambda)\|_2^2$  é uma função contínua decrescente na variável  $\lambda$ , tal que  $\lim_{\lambda \rightarrow \infty} \|d^i(\lambda)\|_2^2 = 0$ ;
3. A função angular  $\gamma^i(\lambda)$  é contínua monótona decrescente na variável  $\lambda$  e  $\lim_{\lambda \rightarrow \infty} \gamma^i(\lambda) = 0$ . Além disso,  $d^i(\lambda)$  rotaciona na direção de  $-g^i$ , quando  $\lambda \rightarrow \infty$ .

**Demonstração:** Ver [8, Teorema 6.1]. □

As hipóteses listadas a seguir são importantes para assegurar a convergência do Algoritmo 1 para pontos fracamente críticos. Mais especificamente, tais hipóteses nos permitem garantir a existência de constantes  $\beta > 0$  e  $\theta \in (0, 1)$  que satisfaçam as chamadas Condição  $\beta$  ( $\|d_k\|_2 \geq \beta \|\nabla f(x_k)\|_2$ ) e Condição do Ângulo ( $\nabla f(x_k)^T d_k \leq -\theta \|\nabla f(x_k)\|_2 \|d_k\|_2$ ), na determinação da direção  $d_k$ .

**Hipótese 3.2.** *A escolha de  $\lambda$  é limitada inferiormente e superiormente por constantes não-negativas, isto é, existem constantes  $L_{min}, L_{max} \in \mathbb{R}_+$ , com  $L_{min} \leq L_{max}$  tais que  $\lambda \in [L_{min}, L_{max}]$ .*

**Hipótese 3.3.** Suponha que  $x_*$  é ponto limite de uma sequência  $(x_k)_{k \in \mathbb{N}}$  gerada pelo Algoritmo 1 tal que um mesmo índice  $i = v(k) \in I_{min}(x_k)$  é escolhido para infinitos índices  $k \in K$  e  $\lim_{k \in K} x_k = x_*$ . Considere ainda que:

- i) Existe  $M \in \mathbb{R}$  tal que  $\sigma_1^i(x) < M$  para todo  $x \in \mathbb{R}^n$ ; onde  $\sigma_1^i(x)$  é o maior autovalor da matriz  $J^i(x)^T J^i(x)$ .
- ii) O ângulo  $\gamma_x^i(\lambda)$  entre a direção de Levenberg-Marquardt e a direção de máxima descida é limitado superiormente por um constante positiva  $\bar{\gamma}^i$  tal que  $0 \leq \gamma_x^i(\lambda) \leq \bar{\gamma}^i < \frac{\pi}{2}$  para todo  $\lambda \in [L_{min}, L_{max}]$  e para todo  $x \in \mathbb{R}^n$ .

**Lema 3.4.** Suponha que o Algoritmo (1) é implementado de forma a satisfazer as Hipóteses (3.2) e (3.3). Então existem constantes  $\beta > 0$  e  $\theta \in (0, 1)$  tais que:

$$\begin{aligned} \|d^i(\lambda_k)\|_2 &\geq \beta \|\nabla f^i(x_k)\|_2, \\ \nabla f^i(x_k)^T d^i(\lambda_k) &\leq -\theta \|\nabla f^i(x_k)\|_2 \|d^i(\lambda_k)\|_2. \end{aligned}$$

**Demonstração:** De fato, combinando  $\nabla f^i(x) = J^i(x)^T R^i(x)$  e (2),

$$\begin{aligned} (J^i(x)^T J^i(x) + \lambda I) d^i(\lambda) &= -\nabla f^i(x), \\ \|J^i(x)^T J^i(x) + \lambda I\|_2 \|d^i(\lambda)\|_2 &\geq \|\nabla f^i(x)\|_2, \\ \|d^i(\lambda)\|_2 &\geq \frac{1}{\|J^i(x)^T J^i(x) + \lambda I\|_2} \|\nabla f^i(x)\|_2, \\ &\geq \frac{1}{\|J^i(x)^T J^i(x)\|_2 + \lambda} \|\nabla f^i(x)\|_2. \end{aligned}$$

Pela definição da 2-Norma,

$$\|d^i(\lambda)\|_2 \geq \frac{1}{\sigma_1^i(x) + \lambda} \|\nabla f^i(x)\|_2.$$

Assim, aplicando as Hipóteses (3.2) e (3.3),

$$\|d^i(\lambda)\|_2 \geq \frac{1}{M + \lambda} \|\nabla f^i(x)\|_2 \geq \frac{1}{M + L_{max}} \|\nabla f^i(x)\|_2.$$

Portanto, basta tomar  $\beta = \frac{1}{M + L_{max}}$ .

Pela definição de  $\gamma_x^i(\lambda)$ , temos que

$$\cos \gamma_x^i(\lambda) = \frac{-\nabla f^i(x)^T d^i(\lambda)}{\|\nabla f^i(x)\|_2 \|d^i(\lambda)\|_2}.$$

O Teorema 3.1 e a Hipótese (3.2) implicam que  $\gamma_x^i(\lambda) \leq \gamma_x^i(L_{min})$ , para todo  $\lambda \in [L_{min}, L_{max}]$  e para todo  $x \in \mathbb{R}^n$ . Desse modo, como a função cosseno é contínua e decrescente no intervalo  $[0, \pi/2]$ , a Hipótese (3.3) garante que para todo  $\lambda \in [L_{min}, L_{max}]$  e para todo  $x \in \mathbb{R}^n$ ,

$$\cos \gamma_x^i(\lambda) \geq \cos \gamma_x^i(L_{min}) \geq \cos \bar{\gamma}^i.$$

Portanto,

$$\frac{-\nabla f^i(x)^T d^i(\lambda)}{\|\nabla f^i(x)\|_2 \|d^i(\lambda)\|_2} \geq \cos \bar{\gamma}^i \iff \nabla f^i(x)^T d^i(\lambda) \leq -\cos \bar{\gamma}^i \|\nabla f^i(x)\|_2 \|d^i(\lambda)\|_2.$$

Note que  $\cos(\bar{\gamma}^i) \neq 0$ , pela Hipótese (3.3). Assim, tomando  $\theta = \cos(\bar{\gamma}^i)$ , a condição do ângulo é satisfeita.  $\square$

**Teorema 3.5.** *Se  $x_*$  é ponto limite de uma sequência gerada pelo Algoritmo 1 sob as Hipóteses (3.2) e (3.3), então  $\nabla f^i(x_*) = 0$ .*

**Demonstração:** Pelo Lema 3.4, existem constantes  $\beta > 0$  e  $\theta \in (0, 1)$ , satisfazendo as chamadas condições  $\beta$  e do ângulo  $\theta$ , na escolha da direção  $d(\lambda_k)$ . Então, o Algoritmo 1 pode ser visto como um algoritmo de descida com busca linear, tal como apresentado em [6, Algoritmo 6.1.5], e converge globalmente [6, Teorema 6.1.6].  $\square$

## 4 Testes e resultados numéricos

Nesta seção apresentamos algumas conclusões obtidas contrapondo o Método de Gauss-Newton, devido a Andreani et al [1, Algoritmo 1], e a adaptação do Método de Levenberg-Marquardt no contexto dos problemas LOVO, proposto na Seção 3 por meio do Algoritmo 1. Para fins de comparação, os testes envolvendo este último método foram realizados com diversas escolhas para o parâmetro de *damping*.

O Método de Gauss-Newton para problemas LOVO, está indicado pela sigla GN, já o Método de Levenberg-Marquardt pela sigla LM. Os diferentes parâmetros de *damping* testados estão indicados pelas siglas PD1, . . . , PD8H3, conforme indicado na tabela abaixo.

Sigla	Autoria	Parâmetro de <i>damping</i>
PD1	Levenberg, ver [5]	$\lambda_k = \frac{\ J(x_k)^T R(x_k)\ _2^2}{f(x_k)}$
PD2	Levenberg modificado, ver [5], [3]	$\lambda_k = \ J(x_k)^T R(x_k)\ _2^2$
PD3	Levenberg modificado 2, ver [5], [3]	$\lambda_k = \ J(x_k)^T R(x_k)\ _2$
PD4	Yamashita & Fukushima, ver [9]	$\lambda_k = \ R(x_k)\ _2^2 = 2f(x)$
PD5	Fan & Yuan, ver [4]	$\lambda_k = \ R(x_k)\ _2 = \sqrt{2f(x)}$
PD6	Benatti, ver [3]	$\lambda_k = \frac{2\ J(x_k)^T R(x_k)\ _2}{3k}$
PD7	Schwertner 01	$\lambda_k = \frac{2\sqrt{f(x_k)}}{3k}$
PD7H1	Schwertner 01, híbrido 01	$\lambda_k = \max \left\{ \frac{2\sqrt{f(x_k)}}{3k}, 10^{-2} \right\}$
PD7H2	Schwertner 01, híbrido 02	$\lambda_k = \max \left\{ \frac{2\sqrt{f(x_k)}}{3k}, 1 \right\}$
PD7H3	Schwertner 01, híbrido 03	$\lambda_k = \max \left\{ \frac{2\sqrt{f(x_k)}}{3k}, 10 \right\}$
PD8	Schwertner 02	$\lambda_k = \frac{\ R(x_k)\ _2^2}{3k} = \frac{2f(x)}{3k}$
PD8H1	Schwertner 02, híbrido 01	$\lambda_k = \max \left\{ \frac{2f(x)}{3k}, 10^{-2} \right\}$
PD8H2	Schwertner 02, híbrido 02	$\lambda_k = \max \left\{ \frac{2f(x)}{3k}, 1 \right\}$
PD8H3	Schwertner 02, híbrido 03	$\lambda_k = \max \left\{ \frac{2f(x)}{3k}, 10 \right\}$

Tabela 1: Parâmetros de *damping* empregados nos testes numéricos.

Note que, assim como exposto na Seção 1, a cada iteração do método de Levenberg-Marquardt, o parâmetro de *damping*  $\lambda_k$  é um valor real positivo. Contudo, tal valor pode se aproximar arbitrariamente de zero, de modo que a matriz  $J(x_k)^T J(x_k) + \lambda_k I$ , apesar de definida positiva, pode apresentar mal condicionamento. Deste modo, propomos as estratégias híbridas PDH1, PD7H2,

PD7H3, PD8H1, PD8H2, PD8H1, baseadas em PD7 e PD8, nas quais limitamos inferiormente a escolha de  $\lambda_k$  pelas constantes  $10^{-2}$ , 1 e 10, assim como mostrado na Tabela 1.

Para a realização dos testes, consideramos o subproblema LOVO de quadrados mínimos generalizado [2] e uma família com 6 funções modelo (duas lineares e quatro não lineares). Os testes foram executados em conjuntos contendo 100, 1000 e 5000 pontos, com uma taxa de confiança de 90%, isto é, apenas 10% de todos os pontos considerados consistiam de avaliações erradas dos modelos. Os conjuntos com dados de 100 e 1000 pontos foram obtidos através da avaliação da função modelo em um mesmo intervalo. Deste modo, podemos ver o conjunto maior como um refinamento das informações presentes no menor. Já para o caso dos conjuntos com 5000 pontos, foram considerados intervalos cinco vezes maiores que os originais. Durante a execução dos algoritmos, foram tomados como critérios de parada a norma do gradiente da função objetivo menor que uma tolerância ( $\varepsilon = 10^{-4}$ ), o número máximo de iterações ( $\mathcal{K}_{max} = 500$ ) e o tempo máximo de execução (1 hora).

Por fim, ao analisarmos comparativamente os dados relativos a performance do método proposto e do Método de Gauss-Newton, evidencia-se que os problemas lineares foram solucionados mais rapidamente por este último. Já ao observarmos os dados relacionados ao conjunto de problemas não lineares, há uma tendência de que o Método de Levenberg-Marquardt com Busca Linear resolva mais rapidamente, em termos de tempo e número de iterações, os problemas de maior dimensão, principalmente ao escolher os parâmetros PD7 e PD7H1. Tal comportamento justifica-se pois o método de Levenberg-Marquardt associa o baixo custo computacional do método de Gauss-Newton (as segundas derivadas da função objetivo não são necessárias) à boa definição dos iterandos (as direções  $d(\lambda_k)$  sempre estão bem definidas, dada positividade da matriz  $J(x_k)^T J(x_k) + \lambda_k I$ ).

Sigla	Nº. de Sucessos (%)	Velocidade (Iterações)	Velocidade (Tempo)
GN	18 (94,74 %)	11	8
PD1	15 (78,95 %)	0	0
PD2	1 (5,26 %)	0	0
PD3	15 (78,95 %)	1	2
PD4	12 (63,16 %)	0	0
PD5	15 (78,95 %)	0	0
PD6	18 (94,74 %)	0	0
PD7	17 (89,47 %)	6	5
PD7H1	17 (89,47 %)	5	3
PD7H2	17 (89,47 %)	1	0
PD7H3	16 (84,21 %)	0	0
PD8	16 (84,21 %)	0	0
PD8H1	16 (84,21 %)	0	0
PD8H2	16 (84,21 %)	0	0
PD8H3	15 (78,95 %)	0	0

Tabela 2: Desempenho dos métodos empregados nos testes numéricos.

Na Tabela 2, estão reunidas informações acerca do desempenho dos métodos de Gauss-Newton e Levenberg-Marquardt com Busca Linear. Note que entre os algoritmos mais robustos (maior número de sucessos), encontram-se GN, LM PD6, LM PD7, LM PD7H1 e LM PD7H2. Dentre os algoritmos que apresentam maior velocidade de execução (em termos de menor número de iterações e menor tempo), destacam-se principalmente GN, LM PD7, LM PD7H1. Deste modo, podemos concluir que os parâmetros de *damping* propostos PD7 (Schwertner 01) e PD7H1 (Schwertner 01, híbrido 01) apresentaram os melhores desempenhos quando comparados a Gauss-Newton.

## 5 Conclusão

Este trabalho teve como foco o estudo do problema de Otimização de Menor Valor Ordenado (LOVO), onde dadas uma família  $\mathcal{F}$  composta por  $r \in \mathbb{N}$  funções reais definidas em  $\Omega \subset \mathbb{R}^n$ , e um inteiro  $p \in \{1, 2, \dots, r\}$ , desejamos minimizar a soma das funções, tomados os  $p$  menores valores, isto é, pretendemos resolver o problema de minimização

$$\min_{x \in \Omega} f_{min}(x) = \min \left\{ \sum_{j \in C_i} F_j(x), i = 1, \dots, m \right\},$$

onde  $m = \binom{r}{p}$  e cada  $C_i$  é um subconjunto distinto de cardinalidade  $p$  do conjunto  $\{1, \dots, r\}$ . De modo particular, nos debruçamos sobre uma generalização do problema de quadrados mínimos não linear, o qual pode ser visto como um subproblema LOVO.

Apresentamos brevemente o Método de Levenberg-Marquardt, assim como introduzimos o problema LOVO e os principais conceitos à ele relacionados. Em seguida, no âmbito do problema LOVO irrestrito, sugerimos a implementação de uma modificação do Método de Levenberg-Marquardt com Busca Linear, o qual converge globalmente para pontos fracamente críticos caso o problema satisfaça as hipóteses de convergência. Vale destacar que sistemas lineares satisfazem naturalmente tais hipóteses e portanto, há um conjunto não trivial de problemas cujas especificidades satisfazem as condições de convergência global. Também propomos dois novos parâmetros de *damping*, além de seis estratégias híbridas, dos quais dois deles, PD7 e PD7H1, mostram-se promissores quando comparados a Gauss-Newton.

## Referências

- [1] Andreani, R., Cesar, G., Cesar-Jr, R. M., Martínez, J. M., e Silva, P. J. S., Efficient curve detection using a Gauss-Newton method with applications in agriculture, em “Proc. 1st International Workshop on Computer Vision Applications for Developing Regions in Conjunction with ICCV”, CVDR, Rio de Janeiro, 2007.
- [2] Andreani, R., Martínez, J. M., Martínez, L., e Yano, F. S., Low order-value optimization and applications, em “Technical Report MCDO 051013”, IMECC-Unicamp, Campinas, 2005.
- [3] Benatti, K. A., “O Método de Levenberg-Marquardt para o Problema de Quadrados Mínimos não Linear”, Dissertação de Mestrado, UFPR, 2017.
- [4] Fan, J. Y., e Yuan, Y. X., On the convergence of a new Levenberg-Marquardt method, em “Technical Report, AMSS, Chinese Academy of Sciences”, pp. 1-11, AMMS, Beijing, 2001.
- [5] Levenberg, K., A method for the solution of certain non-linear problems in least squares, *Quarterly of Applied Mathematics*, 2(2)(1944), 164-168.
- [6] Martínez, J. M., e Santos, S. A., Métodos computacionais de otimização, “Colóquio Brasileiro de Matemática”, Apostilas, 20, Unicamp, Campinas, 1995.
- [7] Marquardt, D. W., An algorithm for least-squares estimation of nonlinear parameters, *Journal of the Society for Industrial and Applied Mathematics*, 11(2)(1963), 431-441.
- [8] Schwertner, A. E. “O Método de Levenberg-Marquardt para problemas de Otimização de Menor Valor Ordenado”, Dissertação de Mestrado, UEM, 2019.
- [9] Yamashita, N., e Fukushima, M., On the rate of convergence of the Levenberg-Marquardt method, “Topics in Numerical Analysis”(Alefled, G. e Chen, X., eds.) pp.239-249, Springer, Viena, 2001.



# Um algoritmo geométrico para o problema de inclusão no envoltório convexo\*

**Rafaela Filippozzi, Douglas Soares Gonçalves,**

UFSC - Departamento de Matemática

Campus Universitário Trindade

88040-900, Florianópolis, SC

E-mail: rafaela.filippozzi@gmail.com, douglas.goncalves@ufsc.br

**Luiz Rafael dos Santos**

UFSC - Departamento de Matemática

Campus Blumenau

89036-004, Blumenau, SC

E-mail: l.r.santos@ufsc.br.

**Resumo:** *O problema de inclusão no envoltório convexo consiste em determinar se um certo ponto pertence ao envoltório convexo de um conjunto de  $n$  pontos. Este problema encontra importantes aplicações em geometria computacional e programação linear. Apresentamos um estudo teórico e prático de um Algoritmo Geométrico, que tem como base um teorema de separação chamado de Dualidade de Distâncias. Utilizando conceitos de otimização contínua estabelecemos a relação do Algoritmo Geométrico com o clássico método de Frank-Wolfe e propomos duas variações do primeiro. Estudamos o comportamento prático dos algoritmos em instâncias artificiais do problema, e comparamos suas performances com algoritmos clássicos de otimização para reformulações lineares e quadráticas. Os resultados obtidos sugerem o Algoritmo Geométrico, e variantes, como uma alternativa promissora para o problema de inclusão no envoltório convexo.*

**Palavras-chave:** *Algoritmo Geométrico; Envoltório Convexo; Dualidade de Distâncias; Teoremas de Separação; Frank-Wolfe.*

## 1 Introdução

Dado  $S \subset \mathbb{R}^m$ , um conjunto de  $n$  pontos em  $\mathbb{R}^m$ , o problema de inclusão no envoltório convexo (PIEC) consiste em determinar se um dado  $p \in \mathbb{R}^m$  pertence ao  $\text{conv}(S)$  (o envoltório convexo de  $S$ ). Se as coordenadas dos  $n$  pontos de  $S$  são colocadas como colunas de uma matriz  $S \in \mathbb{R}^{m \times n}$ , e denotando por  $e \in \mathbb{R}^n$  o vetor cujas componentes são todas iguais a um, podemos formalmente descrever o PIEC como o seguinte problema de decisão: existe  $x \in \mathbb{R}^n$  tal que  $Sx = p$ ,  $e^T x = 1$ ,  $x \geq 0$ ?

Este problema está relacionado a conceitos fundamentais em programação linear [1, 7, 9] e encontra importantes aplicações em geometria computacional [3, 5, 7]. Por exemplo um caso particular do PIEC, é o problema de irredundância, que consiste em determinar todos os pontos extremos do  $\text{conv}(S)$  [5].

Embora o PIEC possa ser formulado como um problema de otimização (mais especificamente um problema de programação quadrática) e então resolvido por métodos clássicos como Gradiente Projetado ou Frank-Wolfe [2], em [7], Kalantari apresenta um algoritmo simples, inspirado em ideias geométricas, ao qual chamaremos de *Algoritmo Geométrico* (Algoritmo 1).

---

\*Este trabalho tem como base a dissertação de mestrado [4].

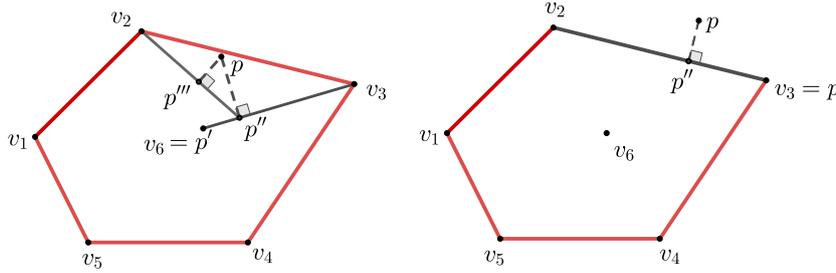


Figura 1: Ilustração das iterações do Algoritmo Geométrico.

## 2 Algoritmo Geométrico

---

### Algoritmo 1: ALGORITMO GEOMÉTRICO

---

- Dados:**  $S = \{v_1, \dots, v_n\}, p \in \mathbb{R}^m, \varepsilon \in (0, 1)$
- 1 Escolha  $p' \in \arg \min\{d(v_j, p) : v_j \in S\}$ .
  - 2 Se  $\forall v_j \neq p', d(v_j, p) > d(v_j, p')$ , pare.
  - 3 Escolha  $v_j \neq p'$ , tal que  $d(v_j, p) \leq d(v_j, p')$ .
  - 4 Calcule  $\bar{\alpha} = \arg \min\{d(p, (1 - \alpha)p' + \alpha v_j) : 0 \leq \alpha \leq 1\}$ , defina  $p'' = (1 - \bar{\alpha})p' + \bar{\alpha}v_j$  e atualize  $p' \leftarrow p''$ .
  - 5 Se  $d(p', p) < \varepsilon$ , pare.
  - 6 Retorne ao passo 2
- 

A cada iteração do Algoritmo 1, buscamos  $v_j \in S \setminus \{p'\}$  tal que  $d(v_j, p) \leq d(v_j, p')$  (veja Passo 3). Um elemento de  $S$  com essa propriedade será chamado de *p-pivô para p'*, *pivô simples*, ou simplesmente *pivô*. Se existe um pivô  $v_j$ , então  $p'$  é atualizado para  $p''$ : o ponto no segmento  $p'v$  mais próximo de  $p$ . Se em uma dada iteração não existe pivô, então o algoritmo para, e retorna  $p'$  como uma *testemunha* de que  $p \notin \text{conv}(S)$ : o hiperplano que bissecta ortogonalmente o segmento  $p'p$  separa  $p$  de  $\text{conv}(S)$ .

Na Figura 1 podemos ver uma ilustração das iterações do Algoritmo Geométrico em ambos os casos. Se  $\forall v_j \neq p', d(v_j, p) > d(v_j, p')$  a resposta para o PIEC é que  $p \notin \text{conv}(S)$  (veja Teorema 1). Caso contrário, se o Algoritmo Geométrico para no Passo 5, temos um ponto  $p' \in \text{conv}(S)$  que está a uma distância menor que  $\varepsilon$  de  $p$ . Temos então que  $p_\varepsilon$  é uma  $\varepsilon$ -solução pertencente ao  $\text{conv}(S)$ , e classificamos  $p$  como elemento de  $\text{conv}(S)$ .

A boa definição e corretude do Algoritmo 1 seguem de um teorema de separação denominado *dualidade de distâncias* que será apresentado a seguir, como consequência do seguinte lema:

**Lema 1.** Sejam  $S = \{v_1, v_2, \dots, v_n\} \subset \mathbb{R}^m$  e  $p \in \mathbb{R}^m$ . Temos que  $p \notin \text{conv}(S)$  se, e somente se, existe  $p' \in \text{conv}(S)$  tal que  $d(v_j, p) > d(v_j, p'), \forall v_j \in S$ .

*Demonstração.* Como  $\text{conv}(S)$  é um conjunto não-vazio, fechado, e convexo, pelo Teorema da projeção [2], se  $p \notin \text{conv}(S)$ , existe um único  $p^+ \in \text{conv}(S)$  tal que  $\|v - p\| > \|v - p^+\|, \forall v \in \text{conv}(S) \setminus \{p^+\}$ . Como todo  $v_j \in S$  pertence ao  $\text{conv}(S)$ , usando  $p' = p^+$  provamos que  $d(v_j, p) > d(v_j, p'), \forall v_j \in S$ .

Por outro lado, considere agora que existe  $p' \in \text{conv}(S) \setminus \{p\}$  tal que  $d(v_j, p) > d(v_j, p'), \forall v_j \in S$ , isto é,  $\|v_j - p\| > \|v_j - p'\|, \forall v_j \in S$ . Em particular,  $\|v_j - p\| > \|v_j - p'\|, \forall v_j \in E$ , em que  $E \subset S$  é o subconjunto de  $S$  dos pontos extremos de  $\text{conv}(S)$ . Daí, de  $\|v_j - p\|^2 > \|v_j - p'\|^2$ , segue que:

$$\frac{\|p\|^2 - \|p'\|^2}{2} > (p - p')^T v_j, \quad \forall v_j \in E. \tag{1}$$

Além disso,  $\forall v \in \text{conv}(S)$ ,  $v$  é combinação convexa dos  $v_j$ 's em  $E$ . Então, de (1), temos que  $\forall v \in \text{conv}(S)$

$$(p - p')^T v = (p - p')^T \left( \sum_{j=1}^{|E|} \alpha_j v_j \right) < \frac{\|p\|^2 - \|p'\|^2}{2}. \quad (2)$$

Em particular, para  $v = p'$  temos  $\frac{\|p\|^2 - \|p'\|^2}{2} > (p - p')^T p'$ , do que segue que

$$(p - p')^T p + \frac{\|p\|^2 - \|p'\|^2}{2} > \|p\|^2 - \|p'\|^2,$$

e portanto

$$(p - p')^T p > \frac{\|p\|^2 - \|p'\|^2}{2}. \quad (3)$$

De (2) e (3) concluímos que  $p \notin \text{conv}(S)$ . □

Assim, se em um dado  $p' \in \text{conv}(S)$  não existir pivô, temos que o hiperplano bissetor ortogonal ao segmento  $p'p$ , dado por  $H = \{x \in \mathbb{R}^m : 2(p - p')^T x = \|p\|^2 - \|p'\|^2\}$ , separa estritamente  $p$  de  $\text{conv}(S)$ .

Segue deste lema o seguinte teorema de alternativas (Dualidade de distâncias).

**Teorema 1.** Sejam  $S = \{v_1, v_2, \dots, v_n\} \subset \mathbb{R}^m$  e  $p \in \mathbb{R}^m$ . Exatamente uma das duas condições é satisfeita

1. Existe  $p' \in \text{conv}(S)$  tal que  $d(v_j, p) > d(v_j, p')$ ,  $\forall v_j \in S$ ;
2. Para todo  $p' \in \text{conv}(S)$ , existe  $v_j \in S$  tal que  $d(v_j, p) \leq d(v_j, p')$ .

Do Teorema 1 temos a justificativa para o critério de parada do Passo 2 e a boa definição do Passo 3 do Algoritmo 1.

### 3 Formulações Alternativas

A fim de verificar a performance prática do Algoritmo Geométrico, apresentamos formulações alternativas para o PIEC, modelando-o como problemas de programação linear e quadrática. Com isso, podemos comparar o desempenho de métodos clássicos para as respectivas formulações com o Algoritmo Geométrico. Além disso, também propomos variantes do Algoritmo Geométrico que podem apresentar um desempenho prático melhor que o original.

#### 3.1 Formulações Lineares

Note que o problema de inclusão no envoltório convexo é um caso especial de problema de factibilidade em Programação Linear (PL). Decidir se  $p \in \text{conv}(S)$  é equivalente a testar a factibilidade de  $Sx = p, x \geq 0, e^T x = 1$ , em que (abusando da notação)  $S$  é uma matriz na qual as colunas são os vetores  $\{v_1, \dots, v_n\}$ ,  $e^T = (1, 1, \dots, 1) \in \mathbb{R}^n$  e  $p \in \mathbb{R}^m$  é o ponto que queremos consultar.

Em [8] foi mostrado que o PIEC pode ser formulado como o seguinte PL:

$$\begin{aligned} \min_x \quad & \sum_{i=1}^{m+1} z_i \\ \text{s.a.} \quad & Sx + z = p \\ & e^T x + z_{m+1} = 1 \\ & x \geq 0, z \geq 0, z_{m+1} \geq 0, \end{aligned} \quad (4)$$

em que  $z \in \mathbb{R}^m$  e  $z_{m+1}$  são variáveis de folga para cada uma das  $m + 1$  restrições.

Com efeito, se assumirmos que  $p$  possui todas as componentes não-negativas<sup>1</sup> e definindo  $z_i = p_i, i = 1, \dots, m, z_{m+1} = 1$  e  $x = 0$ , temos uma solução básica factível para (4). Note que, se o valor ótimo de (4) for zero para uma solução ótima  $(x, z, z_{m+1})$ , temos que todas as variáveis de folga são nulas e então  $p = Sx, x \geq 0$  e  $e^T x = 1$ , ou seja,  $p$  é uma combinação convexa dos elementos de  $S$  e portanto  $p \in \text{conv}(S)$ . No entanto, se o valor ótimo da função objetivo for maior que zero, isso implica que algum  $z_i \neq 0, i = 1, \dots, m$ , e neste caso  $p \notin \text{conv}(S)$ .

Neste trabalho, propomos uma outra formulação linear para o PIEC, usando como base o Teorema do hiperplano separador [9]. Considere o seguinte problema de otimização:

$$\begin{aligned} \min \quad & -w_{m+1} \\ \text{s.a.} \quad & (x - p)^T w + w_{m+1} \leq 0, \quad \forall x \in \text{conv}(S), \\ & \|w\|_\infty \leq 1, w_{m+1} \geq 0, \end{aligned} \tag{5}$$

em que assumimos que  $p \neq 0$  é o ponto a ser consultado.

Observe que se existe um hiperplano separador  $w^T x = \beta$ , com  $w \neq 0$ , entre  $p$  e  $\text{conv}(S)$ , tal que  $w^T x \leq \beta, \forall x \in \text{conv}(S)$  e  $w^T p > \beta$ , então necessariamente  $(x - p)^T w \leq 0$ , o que leva a formulação (5). Se na solução ótima de (5) tivermos  $w_{m+1} = 0$ , então não existe tal hiperplano e concluímos que  $p \in \text{conv}(S)$ .

Note que a formulação (5) apresenta infinitas restrições. Para contornar esta dificuldade basta perceber que  $x \in \text{conv}(S)$  é combinação convexa de  $\{v_1, \dots, v_n\}$ . Logo, podemos escrever (5) de forma equivalente como

$$\begin{aligned} \min \quad & -w_{m+1} \\ \text{s.a.} \quad & (v_j - p)^T w + w_{m+1} \leq 0, \quad j = 1, 2, \dots, n, \\ & -1 \leq w_i \leq 1, \quad i = 1, 2, \dots, m, \\ & w_{m+1} \geq 0. \end{aligned} \tag{6}$$

Para resolver as formulações lineares (4) e (6) podemos usar o método Simplex [1, 9], por exemplo. Por outro lado, podemos verificar que o PIEC pode também ser tratado como um problema de programação quadrática:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \frac{1}{2} \|Sx - p\|^2 \\ \text{s.a.} \quad & x \in \Delta_n, \end{aligned} \tag{7}$$

em que  $\Delta_n = \{x \in \mathbb{R}^n : e^T x = 1, x \geq 0\}$  denota o simplex unitário, e  $x \in \Delta_n$  corresponde aos coeficientes de uma combinação convexa.

Seja  $x_*$  a solução de (7). Se o valor ótimo for estritamente positivo, i.e.,  $\|Sx_* - p\| > 0$ , então  $p \notin \text{conv}(S)$  e  $Sx_*$  corresponde a projeção de  $p$  em  $\text{conv}(S)$ . Se o valor ótimo for zero então temos que  $p \in \text{conv}(S)$ .

Para resolver a formulação quadrática (7), vamos considerar os clássicos métodos do Gradiente Projetado e Gradiente Condicional [9].

### 3.2 Variantes do Algoritmo Geométrico

Sabemos que os algoritmos de busca direcional utilizam a condição do ângulo para evitar que as direções de busca se tornem ortogonais ao gradiente negativo. Por outro lado, podemos mostrar que no Algoritmo Geométrico, se  $v_j$  é pivô para  $p'$ , então  $d = v_j - p'$  é uma direção de descida para  $f(y) = \frac{1}{2} \|y - p\|^2$ , e que  $\sin \theta = \sin \angle pp'v_j$  determina a redução na distância em relação a  $p$  [4].

<sup>1</sup>Se houvesse alguma componente negativa, bastaria multiplicar a equação correspondente por  $(-1)$ .

Então, ao invés de simplesmente escolher um pivô, propomos, a cada iteração do Algoritmo Geométrico, tomar  $v_j \in S$  tal que

$$(p' - p)^T(v_j - p') \leq -\kappa \|p' - p\| \|v_j - p'\|, \quad (8)$$

para  $\kappa \in (0, 1]$ .

---

**Algoritmo 2:** ALGORITMO GEOMÉTRICO COM CONDIÇÃO DO ÂNGULO

---

- Dados:**  $S = \{v_1, \dots, v_n\}, p \in \mathbb{R}^m, \varepsilon \in (0, 1), \kappa \in (0, 1]$
- 1 Escolha  $p' \in \arg \min\{d(v_j, p) : v_j \in S\}$ .
  - 2 Determine  $S_\kappa = \{v_j \in S \setminus \{p'\} : (p' - p)^T(v_j - p') \leq -\kappa \|p' - p\| \|v_j - p'\|\}$ .
  - 3 Se  $S_\kappa \neq \emptyset$ , escolha  $v_j \in S_\kappa$  e vá para o passo 6.
  - 4 Se  $\forall v_j \neq p', d(v_j, p) > d(v_j, p')$ , pare.
  - 5 Escolha  $v_j \neq p'$ , tal que  $d(v_j, p) \leq d(v_j, p')$ .
  - 6 Calcule  $\bar{\alpha} = \arg \min\{d(p, (1 - \alpha)p' + \alpha v_j) : 0 \leq \alpha \leq 1\}$ , defina  $p'' = (1 - \bar{\alpha})p' + \bar{\alpha}v_j$  e atualize  $p' \leftarrow p''$ .
  - 7 Se  $d(p', p) < \varepsilon$ , pare.
  - 8 Retorne ao passo 2
- 

No Algoritmo 2, buscamos na lista  $S$  algum  $v_j$  que satisfaça a condição (8). Dependendo de  $p'$  e do valor de  $\kappa$  pode ser que não exista  $v_j$  em  $S$  satisfazendo este critério, e então simplesmente tomamos um pivô simples.

É claro que para concluir que  $S_\kappa = \emptyset$  temos que percorrer toda a lista  $S$  e, neste processo, já podemos aproveitar para verificar quais  $v_j$  são pivôs.

É importante notar que a condição (8) só implica que  $v_j$  é pivô para valores suficientemente grandes de  $\kappa = \cos \theta$ . Por outro lado, dependendo de  $\kappa$ , nem todo pivô cumpre a condição do ângulo.

Além disso, notamos uma relação entre o Algoritmo Geométrico e o Gradiente Condicional (Frank-Wolfe): o método de Frank-Wolfe, aplicado ao problema (7), equivale ao Algoritmo Geométrico com a escolha de um pivô  $v_j \neq p'$  tal que

$$v_j^T(p' - p) = \min\{v_i^T(p' - p) : v_i \in S \setminus \{p'\}\}. \quad (9)$$

Assim, temos que o método de Frank-Wolfe (com as particularidades do PIEC) equivale a um Algoritmo Geométrico Ganancioso, em que ao invés de escolher  $v_j$  tal que

$$v_j^T(p' - p) \leq \frac{\|p'\|^2 - \|p\|^2}{2}, \quad (10)$$

escolhemos  $v_j$  tal que o lado esquerdo de (10) é mínimo.

---

**Algoritmo 3:** ALGORITMO GEOMÉTRICO GANANCIOSO

---

- Dados:**  $S = \{v_1, \dots, v_n\}, p \in \mathbb{R}^m, \varepsilon \in (0, 1)$
- 1 Escolha  $v_j \in S \setminus \{p'\}$  tal que  $v_j^T(p' - p) = \min\{v_i^T(p' - p) : v_i \in S \setminus \{p'\}\}$ .
  - 2 Se  $v_j^T(p' - p) > (\|p'\|^2 - \|p\|^2)/2$ , pare.
  - 3 Calcule  $\bar{\alpha} = \arg \min\{d(p, (1 - \alpha)p' + \alpha v_j) : 0 \leq \alpha \leq 1\}$ , defina  $p'' = (1 - \bar{\alpha})p' + \bar{\alpha}v_j$  e atualize  $p' \leftarrow p''$ .
  - 4 Se  $d(p', p) < \varepsilon$ , pare.
  - 5 Retorne ao passo 1.
- 

A desvantagem do Algoritmo 3 é que a cada iteração é necessário percorrer toda a lista  $S$  para assegurar o mínimo. Logo, o custo por iteração é sempre  $O(nm)$  enquanto no Algoritmo Geométrico original a busca na lista para tão logo um pivô simples seja encontrado.

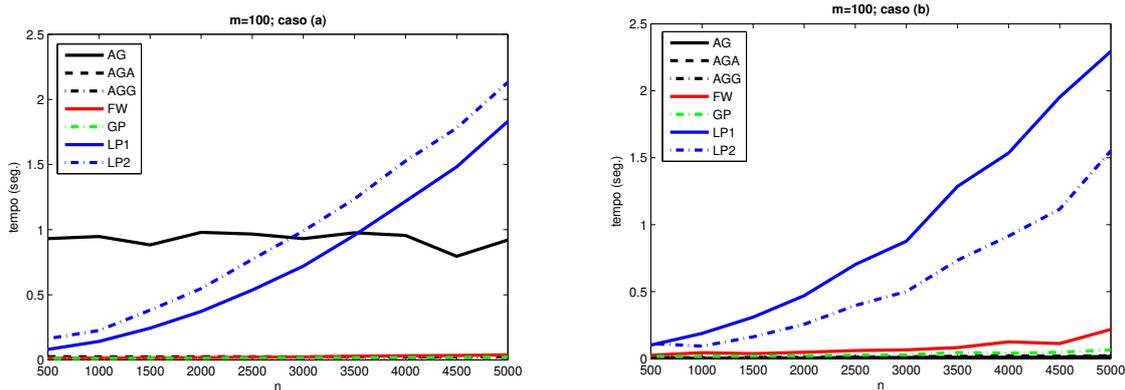


Figura 2: O gráfico a esquerda representa o tempo médio em 10 instâncias para o caso (a):  $p = 0 \in \text{conv}(S)$ , com tolerância  $\varepsilon = 10^{-2}$ . A direita temos o caso (b):  $\|p\| = 2 \notin \text{conv}(S)$ , com  $\varepsilon = 10^{-4}$ .

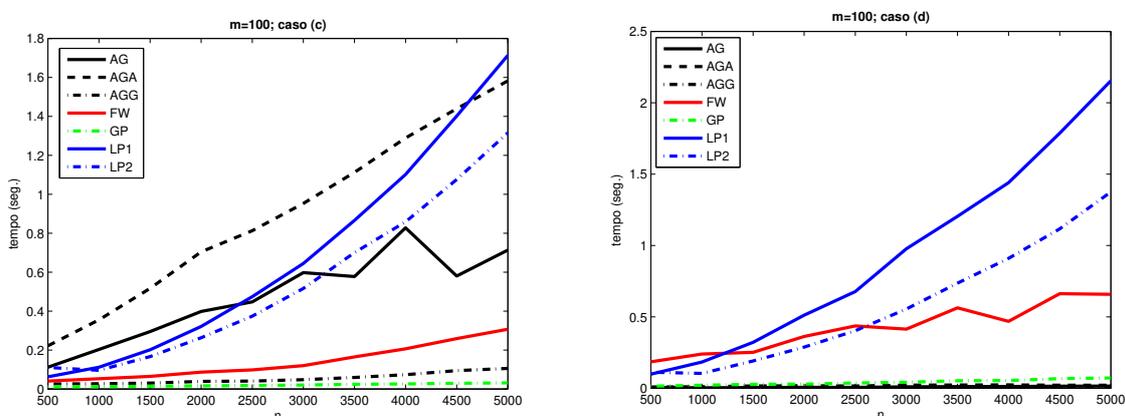


Figura 3: A esquerda temos o tempo médio no caso em que  $p \in \text{conv}(S)$  e está próximo da fronteira (caso (c)), tolerância:  $\varepsilon = 10^{-2}$ . O gráfico a direita representa o tempo médio para o caso (d):  $\|p\| = 1.01$ , com  $\varepsilon = 10^{-4}$ .

### 3.3 Experimentos Numéricos

Para avaliar o desempenho do Algoritmo Geométrico e suas variantes frente a algoritmos clássicos de otimização para as formulações lineares e quadrática, realizamos testes computacionais para instâncias do PIEC geradas de forma artificial. Os pontos do conjunto  $S = \{v_1, \dots, v_n\}$ , são gerados segundo uma distribuição uniforme na bola unitária em  $\mathbb{R}^m$  [6]. O ponto  $p$  foi gerado de quatro maneiras distintas para analisar diferentes cenários, como discutiremos a seguir.

Os algoritmos foram implementados e os experimentos realizados em ambiente Matlab R2018b em um computador com processador Intel Core i5 1.8Ghz, com 8GB de RAM. Esses algoritmos foram: Algoritmo Geométrico (AG), Algoritmo Geométrico com a condição do ângulo (AGA), Algoritmo Geométrico Ganancioso (AGG), `linprog`<sup>2</sup> para a formulação (4) (LP1), `linprog` para formulação (6) (LP2), Algoritmo de Frank Wolfe (FW) e Algoritmo Gradiente Projetado (GP) para a formulação (7). Para todos consideramos a dimensão  $m = 100$ . Para cada cenário e para cada valor de  $n > m$  foram geradas 10 instâncias.

Os experimentos foram organizados em quatro cenários: (a)  $p \in \text{conv}(S)$  longe da fronteira, (b)  $p \notin \text{conv}(S)$  longe da fronteira, (c)  $p \in \text{conv}(S)$  perto da fronteira e (d)  $p \notin \text{conv}(S)$  perto da fronteira.

<sup>2</sup>utilizamos o solver de PL `linprog` do Matlab com a opção 'Dual-simplex'.

Analisando os resultados obtidos nos quatro cenários propostos, representados nas Figuras 2 e 3, podemos notar que, em geral, o Algoritmo Geométrico e suas variantes obtiveram um desempenho melhor em relação às formulações lineares (resolvidas de forma exata pelo Dual-Simplex) e quadráticas (resolvidas com Frank-Wolfe e Gradiente Projetado), principalmente nos casos (b) e (d), quando  $p \notin \text{conv}(S)$ . Podemos mostrar que nesse caso a complexidade do Algoritmo Geométrico e suas variantes depende exclusivamente da geometria do problema, sendo pouco sensível à tolerância  $\varepsilon$  [7].

## 4 Considerações finais

Neste trabalho realizamos um estudo teórico e prático do Algoritmo Geométrico, proposto em [7], para o problema de inclusão no envoltório convexo (PIEC).

A corretude do algoritmo é baseada em um teorema de alternativas chamado Dualidade de Distâncias (Teorema 1) para o qual fornecemos uma nova demonstração inspirada no Teorema da Projeção e no Teorema do Hiperplano Separador.

Além disso, interpretando o PIEC como um problema de otimização, propomos variantes do algoritmo original: o Algoritmo Geométrico com condição do ângulo e o Algoritmo Geométrico Ganancioso. Este último foi motivado pela relação entre o algoritmo original e o método de Gradiente Condicional (Frank-Wolfe).

Para ter uma ideia do desempenho prático do Algoritmo Geométrico, considerados formulações de programação linear e quadrática para o PIEC, e comparamos o AG com métodos clássicos para estas formulações. Os resultados obtidos indicam que o Algoritmo Geométrico e variantes apresentam um desempenho bom frente aos métodos clássicos, pelo menos nos problemas teste artificiais considerados neste trabalho.

Assim, concluímos que o Algoritmo Geométrico figura como um alternativa promissora para o problema de inclusão no envoltório convexo, apresentando um bom desempenho prático.

## Referências

- [1] Bazaraa, M. S. e John J. J. “Linear programming and network flows”. WILEY, New York, 1977.
- [2] Bertsekas, D. P. “Nonlinear programming”, Athena Scientific Belmont, 1999.
- [3] Chen, L.L e Woo, T. C. “Computational geometry on the sphere with application to automated machining”. Journal of Mechanical Design, American Society of Mechanical Engineers, vol. 114, n.2, pp. 288-295,(1992).
- [4] Filippozzi, R. “Um algoritmo geométrico para o problema de inclusão no envoltório convexo.” Dissertação de Mestrado, UFSC, 2019.
- [5] Goodman, E. J. e O’Rourke,F. J. “Handbook of discrete and computational geometry”, Chapman and Hall/CRC, 2017.
- [6] Harman, R. e Lacko, V. “On decompositional algorithms for uniform sampling from n-spheres and n-balls”. Journal of Multivariate Analysis, vol. 101, n.10, pp. 2297-2304, (2010).
- [7] Kalantari,B. “A characterization theorem and an algorithm for a convex hull problem”, Annals of Operatons Research, Springer, v.226, n.1, p.301-349, (2015).
- [8] Li, M e Kalantari, B. “Experimental study of the convex hull decision problem via a new geometric algorithm”. 23rd Annual Fall Workshop on Computational Geometry, City College of New York, 2013.
- [9] Luenberger, D. G. e Ye, Y. “Linear and Nonlinear Programming”, Springer, 2008.

