Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Programa de Pós-Graduação em Matemática

Julio César Agustín Sangay

Aplicação do método de complementaridade mista para problemas parabólicos não lineares

Juiz de Fora
2015

Julio César Agustín Sangay

Aplicação do método de complementaridade mista para problemas parabólicos não lineares

Dissertação apresentada ao Programa de Pós-Graduação em Matemática da Universidade Federal de Juiz de Fora, na área de concentração em Matemática , como requisito parcial para obtenção do título de Mestre em Matemática.

Orientador: Sandro Rodrigues Mazorche.

Coorientador: Grigori Chapiro.

Juiz de Fora

2015

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Agustín S., Julio C. .
Aplicação do método de complementaridade mista para problemas parabólicos não lineares / Julio César Agustín Sangay. – 2015.
94 f. : il.

Orientador: Sandro Rodrigues Mazorche.
Coorientador: Grigori Chapiro.
Dissertação (Mestrado) – Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Programa de Pós-Graduação em Matemática, 2015.

1. Combustão 2. Algoritmo de Complementaridade Mista. 3. Diferenças Finitas 4. Leis de conservação. I. Mazorche, Sandro Rodrigues, orient. II. Chapiro, Grigori, coorient. III. Título.

JULIO CÉSAR AGUSTÍN SANGAY

APLICAÇÃO DO MÉTODO DE COMPLEMENTARIDADE MISTA PARA PROBLEMAS PARABÓLICOS NÃO LINEARES

Dissertação aprovada pela Comissão Examinadora abaixo elencada como requisito para a obtenção do título de Mestre em Matemática pelo Mestrado Acadêmico em Matemática do Instituto de Ciências Exatas da Universidade Federal de Juiz de Fora.

_____

Prof. Dr. Sandro Rodrigues Mazorche
(Orientador)
Mestrado Acadêmico em Matemática
UFJF

_____

Prof. Dr. Grigori Chapiro
(Coorientador)
Mestrado Acadêmico em Matemática
UFJF

_____

Prof. Dr. Wilhelm Passarella Freire
UFJF

_____

Prof. Dr. Eduardo Cardoso De Abreu
UNICAMP

Juiz de Fora, 29 de Maio de 2015.

AGRADECIMENTOS

Ao finalizar esta Dissertação, indicando o final do percurso do Mestrado, meta conseguida com muito esforço e trabalho, para o qual interviram pessoas que colaboraram para cumprir este objetivo, quero agradecer o meu orientador professor Dr. Sandro Rodrigues Mazorche pela sua paciência, tempo, apoio e dedicação para me orientar. Agradeço a ele por me aceitar para trabalhar na dissertação de mestrado sob sua orientação. Sua confiança, amizade e capacidade para guiar minhas idéias foram de muito valor, não só no desenvolvimento da dissertação, mas também na minha formação como pesquisador. Agradeço as ideias, rigorosidade e críticas com a finalidade de realizar um bom trabalho. Muito obrigado Professor.

Quero expressar também meu mais sincero agradecimento ao Dr. Grigori Chapiro pelo seu apoio e participação na conclusão da dissertação, sua disponibilidade e paciência.

Agradeço ao Departamento de Matemática da UFJF, aos professores, pessoal administrativo e funcionários que contribuíram na minha formação. Muito obrigado Professores e Funcionários.

A minha mãe Raquel, que sempre acreditou e me apoiou nos meus objetivos, minha família que apesar de estar longe de mim sempre me apoiaram para alcançar meus objetivos. especialmente meu pai José, avô Evaristo e Antenor que cuidam de mim lá do céu. A todos meus amigos, que compartilharam comigo tempo de estudos, conversas e brincadeiras. Aos amigos que fiz no mestrado: Marianna, Sandra, Eliza, Livia, Yulia, Marianna, Talita, Camila, Natália, Taís, Wesley, Carlos, Juan Torres, Yamashita, Eduardo, Miguel, Jesús, Angel, Daniel, Adalton, Erasmo, Pedretti, Pavel, Mariano, aos meus parceros de futebol e festas Vladimir, Walberto, Oscar. Obrigado a todos meus amigos que fiz nestes dois anos em Juiz de Fora.

Agradeço o apoio financeiro da Capes.

RESUMO

Neste trabalho realizamos um estudo do método de complementaridade mista para problemas parabólicos não lineares, devido ao fato de que alguns podem ser escritos como problema de complementaridade mista e aparecem em muitas aplicações como fluxo de líquidos em um meio poroso, difusão, fluxo de calor envolvendo mudança de fases. Estes tipos de problemas apresentam dificuldades para obter as soluções analíticas.

Estuda-se leis de conservação e os tipos de soluções associadas ao Problema de Riemann, essencialmente leis de balanço que expressam o fato de que alguma substância é conservada. O estudo desta teoria é importante pois frequentemente as leis de conservação aparecem quando nos problemas parabólicos são desprezados os termos difusivos de segunda ordem.

Estudaremos um método numérico que permita a busca de uma solução aproximada da solução exata, o qual é uma variação do método de Newton para resolver sistemas não lineares que estão baseados num esquema de diferenças finitas implícito e um algoritmo de complementaridade mista não linear, FDA-MNCP. O método tem a vantagem de fornecer uma convergência global em relação ao método de diferenças finitas como o método de Newton que só tem convergência local.

A teoria é aplicada ao modelo de combustão in-situ, que pode ser reescrito na forma de problema de complementaridade mista, além disso faremos uma comparação com o método FDA-NCP.

Palavras-chave: Combustão, Algoritmo de Complementaridade Mista. Diferenças Finitas. Leis de Conservação.

ABSTRACT

In this work, we study the mixed complementarity method for nonlinear parabolic problems, because some can be written as mixed complementarity problems and appear in many applications such as fluid flow in porous media, diffusion, heat flow wrapping phase change. These types of problems have difficulty obtaining the analytical solution.

We study the conservation laws and the types of solutions associated with the Riemann Problem, these types of laws are essentially balance laws that express the fact that some substance is balanced. The study of this theory is important because the conservation laws often appear when the parabolic problems are neglected the diffusive terms of second order.

We will study a numerical method that allows finding an approximate solution of the exact solution, which is a variation of the Newton's method for solving nonlinear systems based on an implicit finite difference scheme and a nonlinear algorithm mixed complementarity, FDA-MNCP. The method has the advantage of provide a global convergence in relation to the finite difference method and method of Newton that only has local convergence.

The theory is applied to model in-situ combustion, which can be rewritten in the form of mixed complementarity also we do a comparison with the FDA-NCP method.

Key-words: Combustion. Mixed Nonlinear Complementarity Algorithm. Finite Diference. Conservation Laws.

LISTA DE ILUSTRAÇÕES

Figura 1 – Superfície Ω. 14
Figura 2 – S_1 : Meio pelo qual a massa ingressa a Ω e S_2 : Meio pelo qual a massa sai de Ω. 15
Figura 3 – Elemento da Superfície S_2. 15
Figura 4 – Supporte compacto de f. 18
Figura 5 – Curva que divide o plano em duas regiões, Ω_l e Ω_r. 19
Figura 6 – Descomposição da região Ω. 21
Figura 7 – Orientação positiva de B_l e B_r. 21
Figura 8 – Ilustração dos comportamentos: genuinamente não linear e linearmente degenerado. 26
Figura 9 – Curvas características para uma onda de contato, onde $f'(u) > 0$ quando o salto ocorre em $x_0 = 0$. 31
Figura 10 – Representação da malha do esquema de diferenças finitas. 33
Figura 11 – Molécula computacional do método de Crank-Nicolson. 34
Figura 12 – Discretização para a análise de erro 39
Figura 13 – Comparação dos métodos de FDA-MNCP e FDA-NCP para $M = 50$ nos instantes de tempos $t = 0.000,\ 0.002,\ 0.004,\ 0.006,\ 0.008,\ 0.010$. Os valores de θ são representados por bolinhas verdes e linha contínua vermelha, os valores de η são representados por bolinhas rosas e linha contínua azul. 69
Figura 14 – Comparação dos métodos de FDA-MNCP e FDA-NCP para $M = 100$ nos instantes de tempos $t = 0.000,\ 0.002,\ 0.004,\ 0.006,\ 0.008,\ 0.010$. Os valores de θ são representados por bolinhas verdes e linha contínua vermelha, os valores de η são representados por bolinhas rosas e linha contínua azul. 70
Figura 15 – Comparação dos métodos de FDA-MNCP e FDA-NCP para $M = 200$ nos instantes de tempos $t = 0.000,\ 0.002,\ 0.004,\ 0.006,\ 0.008,\ 0.010$. Os valores de θ são representados por bolinhas verdes e linha contínua vermelha, os valores de η são representados por bolinhas rosas e linha contínua azul. 71
Figura 16 – Comparação dos métodos de FDA-MNCP e FDA-NCP para $M = 400$ nos instantes de tempos $t = 0.000,\ 0.002,\ 0.004,\ 0.006,\ 0.008,\ 0.010$. Os valores de θ são representados por bolinhas verdes e linha contínua vermelha, os valores de η são representados por bolinhas rosas e linha contínua azul. 72
Figura 17 – Diferença entre os métodos FDA-MNCP e FDA-NCP. 73
Figura 18 – Tempo (t)vs$E_{\Delta x}$. Erro Relativo do Método FDA-MNCP. Aqui $\Delta x = \frac{1}{50},\ \frac{1}{100},\ \frac{1}{200}$. 76

Figura 19 – Tempo (t)vs$E_{\Delta x}$. Erro Relativo do Método FDA-NCP. Aqui $\Delta x = \frac{1}{50}, \frac{1}{100}, \frac{1}{200}$. 76
Figura 20 – Modelo 80

LISTA DE TABELAS

Tabela 1 – Erro Relativo para o método de Crank-Nicolson 39
Tabela 2 – Parâmetros dimensionais para combustão in-situ e seus valores típicos [23]. 63
Tabela 3 – Comparação do tempo de processo computacional com $M = 50$ para os métodos FDA-MNCP e FDA-NCP. O $t(n)$ é o tempo medido em segundos que demorou o método em encontrar a solução no instante de tempo t. 73
Tabela 4 – Comparação do tempo de processo computacional com $M = 100$ para os métodos FDA-MNCP e FDA-NCP. O $t(n)$ é o tempo medido em segundos que demorou o método em encontrar a solução no instante de tempo t. 74
Tabela 5 – Comparação do tempo de processo computacional com $M = 200$ para os métodos FDA-MNCP e FDA-NCP. O $t(n)$ é o tempo medido em segundos que demorou o método em encontrar a solução no instante de tempo t. 74
Tabela 6 – Comparação do tempo de processo computacional com $M = 400$ para os métodos FDA-MNCP e FDA-NCP. O $t(n)$ é o tempo medido em segundos que demorou o método em encontrar a solução no instante de tempo t. 75
Tabela 7 – Erro Relativo para θ com FDA-MNCP e $h = \frac{1}{50}$ para os instantes de tempo "t" indicados na primeira coluna. 77
Tabela 8 – Erro Relativo para θ com FDA-NCP e $h = \frac{1}{50}$ para os instantes de tempo "t" indicados na primeira coluna. 77
Tabela 9 – Erro Relativo para η com FDA-MNCP e $h = \frac{1}{50}$ para os instantes de tempo "t" indicados na primeira coluna. 78
Tabela 10 – Erro Relativo para η com FDA-NCP e $h = \frac{1}{50}$ para os instantes de tempo "t" indicados na primeira coluna. 78
Tabela 11 – Parâmetros dimensionais para a injeção de vapor num meio poroso saturado de água e seus valores típicos 81

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
UFJF	Universidade Federal de Juiz de Fora
put	Por unidade de tempo
PCM	Problema de Complementariedade Mista
MLCP	Problema de Complementariedade Mista Linear
MNCP	Problema de Complementariedade Mista Não Linear
FDA-MNCP	Algoritmo de Direções Admissíveis para Problema de Complementariedade Mista Não Linear
FDA-NCP	Algoritmo de Direções Admissíveis para Problema de Complementariedade Não Linear
EDP	Equação Diferencial Parcial
supp f	Supporte de f que pode ser compacto o não.
iter	Numero de iterações
BL(t)	Busqueda linear

LISTA DE SÍMBOLOS

$\forall$	Para todo
$\in$	Pertence
$\mathbb{R}_0^+$	Números reais não negativos
Ω	Superficie em $\mathbb{R}^3$
$\partial\Omega$	Fronteira da superfície
S_1	Meio pelo qual a massa ingressa a Ω
S_2	Meio pelo qual a massa sai de Ω
ρ	Densidade do meio
v	Velocidade
$\eta_{1,\ 2}$	Vetor normal exterior e unitario a $S_{1,\ 2}$
V	Volume
Δt	Incremento no tempo
Δx	Incremento no espaço
λ_j	Autovalor de uma matriz
r_j	Autovetor associado ao autovalor λ_j
∇	Gradiente
$\bullet$	Producto de Hadamard

SUMÁRIO

1 INTRODUÇÃO 13

2 SISTEMAS HIPERBÓLICOS DE LEIS DE CONSERVAÇÃO 14
2.1 LEI DE CONSERVAÇÃO 14
2.2 SISTEMAS HIPERBÓLICOS 17
2.3 SOLUÇÃO FORTE E SOLUÇÃO FRACA 18
2.4 EXISTÊNCIA E UNICIDADE DE SOLUÇÕES FRACAS 23
2.4.1 **Condição de Entropia** 23
2.4.2 **Critério de Viscosidade** 24
2.5 SOLUÇÃO DO PROBLEMA DE RIEMANN 25
2.5.1 **Campos Característicos** 25
2.5.2 **Ondas de Choque** 27
2.5.3 **Ondas de Rarefação** 28
2.5.4 **Descontinuidade de Contato** 31

3 ESQUEMAS DE DIFERENÇAS FINITAS 33
3.1 CONSISTÊNCIA, ESTABILIDADE E CONVERGÊNCIA 35
3.2 ERRO DE TRUNCAMENTO 38
3.3 MÉTODO DE NEWTON 40

4 PROBLEMA DE COMPLEMENTARIDADE MISTA NÃO LINEAR 42
4.1 CONCEITOS BÁSICOS 42
4.2 ALGORITMO DE PONTO INTERIOR VIÁVEL PARA MNCP 43
4.3 DESCRIÇÃO DO ALGORITMO FDA-MNCP 47
4.4 CONVERGÊNCIA GLOBAL DO FDA-MNCP 48
4.4.1 **Análise de Convergência Assintótica** 59

5 COMBUSTÃO IN-SITU 62
5.1 MODELAGEM DO PROBLEMA FÍSICO 62
5.2 DESCRIÇÃO DO MÉTODO FDA-MNCP PARA O MODELO SIMPLES DE COMBUSTÃO IN-SITU 64
5.3 COMPARAÇÃO DOS MÉTODOS FDA-MNCP COM O MÉTODO FDA-NCP 68
5.4 ANÁLISE DE ERRO 75

6 CONCLUSÕES 79

7 TRABALHOS FUTUROS - MODELO SIMPLES DA INJEÇÃO DE VAPOR NUM MEIO POROSO SATURADO DE ÁGUA 80
7.1 MODELAGEM DO PROBLEMA FÍSICO 80
7.2 EQUAÇÕES ADIMENSIONAIS 81
7.3 DESCRIÇÃO DOS ESQUEMAS NUMÉRICOS 84
7.3.1 **Método FDA-MNCP** 85

REFERÊNCIAS 89

APÊNDICE A – Definições e resultados gerais 91
A.1 CONCEITOS BÁSICOS PARA PROBLEMAS DE OTIMIZAÇÃO . . . 91
A.2 BUSCA LINEARES INEXATAS 93
A.3 O MÉTODO DE NEWTON PARA EQUAÇÕES 93

1 INTRODUÇÃO

Vários modelos matemáticos em diferentes disciplinas, como engenharia, física, economia e outras ciências estudam equações diferenciais parciais do tipo parabólico. Estes modelos podem levar ao problema de complementaridade mista, isto é, o caso do modelo de combustão in-situ, o qual será nosso modelo neste trabalho. Outras aplicações de problemas de complementaridade são descritos em [12].

Como o objetivo é encontrar uma aproximação da solução analítica, desenvolveremos um método numérico que nos permita cumprir nosso objetivo. Esta técnica será aplicada ao modelo simples de combustão in-situ o qual será escrito como um problema de complementaridade mista.

O modelo de combustão in-situ é um caso particular do modelo tratado em [4]. Neste caso, o modelo considera a injeção de ar num meio poroso que contém combustível sólido e consiste de um sistema de duas equações diferenciais de tipo parabólico não lineares.

As contribuições do trabalho são: Estudo do modelo simples de combustão in-situ e as simulações para o modelo proposto aplicando método de Crank-Nicolson e o algoritmo FDA-MNCP [20]. Além disso, estudaremos o modelo de injeção de vapor num meio poroso via leis de conservação e no futuro fazer as simulações para este modelo. A implementação será feita em Matlab, o qual é um programa destinado para fazer cálculos numéricos e é amplamente utilizado no campo da engenharia.

Este trabalho está organizado como segue. No Capítulo 2 faremos uma introdução às leis de conservação. No Capítulo 3 descreveremos o método de diferenças finitas que será usado nas simulações do modelo simples de combustão in-situ. No Capítulo 4 apresentaremos a descrição e funcionamento do algoritmo FDA-MNCP que será usado nos capítulos seguintes. Além disso, apresentaremos os resultados que apontam que a sequência de pontos viáveis gerada está contida numa região viável e verificaremos que as direções obtidas são viáveis e de descida para uma função associada ao problema de complementaridade e também veremos a prova da convergência global para o FDA-MNCP seguindo o feito para FDA-NCP [14]. No Capítulo 5 aplicamos o método FDA-MNCP ao Problema de Combustão in-situ, descrevemos o procedimento de discretização usando a técnica de diferenças finitas estudadas no Capítulo 3, para o problema de complementaridade mista associado ao problema e também apresentamos os resultados numéricos e a análise de erro correspondente com a comparação com o método FDA-NCP. Finalmente apresentaremos algumas conclusões, trabalhos futuros, bibliografia e apêndice.

2 SISTEMAS HIPERBÓLICOS DE LEIS DE CONSERVAÇÃO

Os sistemas hiperbólicos de equações diferenciais parciais são empregados para modelar uma grande quantidade de fenômenos físicos que surgem nas mais diversas áreas como dinâmica dos gases, geofísicia, acústica, ótica, entre outras, que envolvem, por exemplo, movimento de onda ou o transporte advectivo de substância, onde a propagação de singularidade está quase sempre presente e são chamados *Leis de Conservação*, os quais tem a seguinte forma [24], [25]:

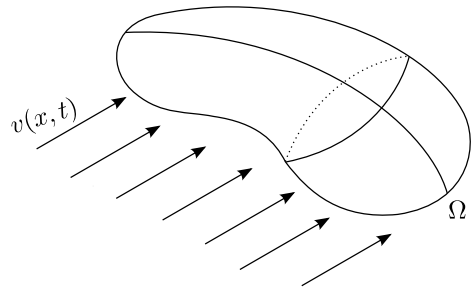
$$\frac{\partial}{\partial t}u(x,t) + \frac{\partial}{\partial x}f(u((x,t)) = 0, \tag{2.1}$$

onde, $u : \mathbb{R} \times \mathbb{R}_0^+ \to \mathbb{R}^m$ representa as variáveis de estado no ponto x, instante t e $f(u)$ é o fluxo dessas variáveis, [17]. Consideraremos aqui um princípio básico, chamado lei de conservação da massa.

2.1 LEI DE CONSERVAÇÃO

Suponhamos que tem-se um meio contínuo com um campo de velocidades $v(x,t)$ e sejam $\Omega \subset \mathbb{R}^3$, $x \in \Omega$, ver Figura 1. De acordo com a lei da conservação da massa [17],

Figura 1 – Superfície Ω.



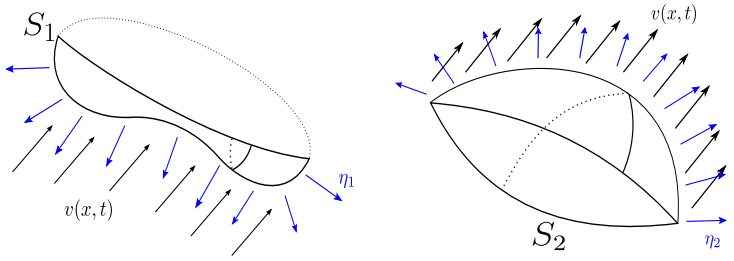
Fonte: Julio Agustín

sabemos o que ingressa menos o que sai de Ω por unidade de tempo (*put*), é o que se acumula em Ω, o que poderíamos escrever como:

$$\textbf{ENTRADA} - \textbf{SAÍDA} = \textbf{ACUMULAÇÃO}. \tag{2.2}$$

Seja $S = \partial\Omega$, uma superfície suave. Dividimos S em duas partes: S_1 que é o meio pelo qual a massa ingressa a Ω e S_2 é o meio pelo qual a massa sai de Ω, como mostra a Figura 2.

Figura 2 – S_1 : Meio pelo qual a massa ingressa a Ω e S_2 : Meio pelo qual a massa sai de Ω.

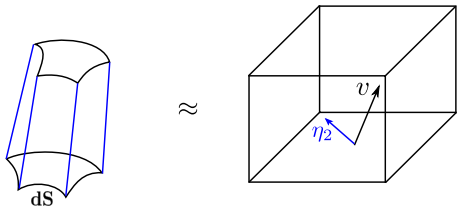


Fonte: Julio Agustín

onde η_1 é o vetor normal exterior à superficie S_1 e η_2 o vetor normal exterior à superficie S_2.

Calcule-se a quantidade de material que sai de Ω, (put). Seja dS um elemento da superfície em S_2. O volume do meio que passa por dS num intervalo de tempo é aproximadamente: $(v \cdot \eta_2)dS$.

Figura 3 – Elemento da Superfície S_2.



Fonte: Julio Agustín

Seja $\rho(x,t)$ a densidade do meio, logo a quantidade de massa que passa por dS, (put) é:

$$\rho(v \cdot \eta_2)dS.$$

Portanto, a quantidade de massa que **SAI** por toda a superfície S_2, (put) é:

$$\int_{S_2} \rho(v \cdot \eta_2)dS, \qquad \text{com} \quad v \cdot \eta_2 > 0.$$

Analogamente, a quantidade de massa que **INGRESSA** por S_1, (put) é:

$$-\int_{S_1} \rho(v \cdot \eta_1)dS, \qquad \text{com} \quad v \cdot \eta_1 < 0.$$

Portanto, o que passa menos o que sai por a superfície $S = \partial\Omega$, é:

$$-\int_{S_1} \rho(v \cdot \eta_1)dS - \int_{S_2} \rho(v \cdot \eta_2)dS = -\int_S \rho(v \cdot \eta)dS,$$

onde η é o vetor normal exterior e unitário a S.

Agora calculemos a quantidade de massa que se acumula em Ω (put). Seja dV um elemento do volume de massa dentro de Ω. No instante t, a massa de dV é dada por:

$$\rho(x,t)dV.$$

O que se acumula em dV, no intervalo de tempo $[t, t+\Delta t]$ é:

$$[\rho(x, t+\Delta t) - \rho(x,t)]dV.$$

Logo, o que se acumula em Ω no intervalo de tempo $[t, t+\Delta t]$ é:

$$\int_\Omega [\rho(x, t+\Delta t) - \rho(x,t)]dV.$$

A quantidade de massa que se acumula em todo Ω (put) é:

$$\frac{1}{\Delta t}\int_\Omega [\rho(x, t+\Delta t) - \rho(x,t)]dV,$$

tomando limite, quando $\Delta t \to 0$, obtemos:

$$\frac{d}{dt}\int_\Omega \rho(x,t)dV,$$

que é a quantidade de massa que se acumula em Ω (put).

Portanto, da expressão da lei de conservação da massa, temos:

$$\frac{d}{dt}\int_\Omega \rho(x,t)dV = -\int_S \rho(v\eta)dS,$$

onde a integral da direita é o fluxo que atravessa a superfície S.

Se denotamos a densidade desta substância como $u(x,t)$ e por $f(x,t)$ o fluxo, então a lei de conservação pode-se escrever como:

$$\frac{d}{dt}\int_\Omega u(x,t)dV = -\int_{\partial\Omega} f\eta dS, \tag{2.3}$$

onde S é a superfície regular e η é o vetor normal exterior unitário de S. Esta equação é a forma integral de uma lei de conservação. Supondo que u e f são diferenciáveis e aplicando o teorema da divergência [18], temos:

$$\int_\Omega \frac{\partial}{\partial t}u(x,t)dV + \int_\Omega div(f(u))dV = 0.$$
$$\int_\Omega \left[\frac{\partial}{\partial t}u(x,t) + div(f(u))\right]dV = 0.$$

Como Ω é uma região arbitrária, obtemos a equação de continuidade

$$\frac{\partial}{\partial t}\left(u(x,t)\right) + div(f(u)) = 0, \tag{2.4}$$

a qual é chamada de **Forma Diferencial da Lei de Conservação.**

Se repetimos todos os cálculos anteriores sobre a hipótese de existência de uma função **fonte** ou **sumidouro** $g(x,t,u)$, a forma diferencial da lei de conservação (2.4) ficará

$$\frac{\partial u(x,t)}{\partial t} + div(f(u)) = g(x,t,u). \tag{2.5}$$

A equação (2.5) relaciona a função u, a função fluxo f e o termo fonte g a qual é denominada **Lei do Balanço** [1]. Por outro lado, para a modelagem ter sentido precisamos de uma condição inicial $u(x,0) = u_0(x)$.

Nosso estudo estará restringido ao caso de $(x,t) \in \mathbb{R} \times \mathbb{R}_0^+$ no qual a equação (2.5) se escreve como:

$$u_t(x,t) + [f(u)]_x = g(x,t,u). \tag{2.6}$$

Na seção seguinte estudaremos com maior detalhe a equação com termo fonte $g = 0$.

2.2 SISTEMAS HIPERBÓLICOS

Nesta seção estudaremos a equação (2.6) que representa um sistema de leis de conservação, onde $u : \mathbb{R} \times \mathbb{R}_0^+ \longrightarrow \mathbb{R}^m$ e a função de fluxo $f : \mathbb{R}^m \longrightarrow \mathbb{R}^m$. No caso linear temos que $f(u) = Au$ então $f'(u) = A$ com $A \in \mathbb{R}^{m\times m}$ uma matriz constante. O sistema (2.6) é dito **Hiperbólico** se os autovalores $\lambda_1, \lambda_2, ..., \lambda_m$ da matriz jacobiana $f'(u)$ de f, são reais para todo $u \in \mathbb{R}^m$. Se além de reais, os autovalores são distintos em Ω, dizemos que (2.6) é **Estritamente Hiperbólico** e sendo $r_1(u), r_2(u), ..., r_m(u)$ os respectivos autovetores.

Consideraremos, como exemplo, o seguinte sistema, que será abordado com mais detalhes mais na frente:

$$\begin{aligned} \theta_t + (\mathrm{v}\theta)_x &= 0, \\ \omega_t \ + \ (\omega)_x &= 0, \\ \sigma_t \ + \ (\sigma)_x &= 0, \end{aligned} \tag{2.7}$$

onde v é uma constante real positiva.

Sendo $U = \begin{pmatrix} \theta \\ \omega \\ \sigma \end{pmatrix} \in \mathbb{R}^3$ temos que:

$$f(U) = \begin{pmatrix} \mathrm{v}\theta \\ \omega \\ \sigma \end{pmatrix} \quad \Rightarrow f'(U) = \begin{pmatrix} \mathrm{v} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

tem autovalores $\lambda_1 = \lambda_1(U) = \mathrm{v}$, $\lambda_{2,3} = \lambda_{2,3}(U) = 1$. Por tanto, o sistema (2.7) é um sistema **Hiperbólico**. Além disso, os autovetores associados são:

$$r_1 = r_1(U) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad r_2 = r_2(U) = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \ e \ \ r_3 = r_3(U) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

para λ_1, λ_2 e λ_3, respectivamente.

2.3 SOLUÇÃO FORTE E SOLUÇÃO FRACA

Estudaremos problemas de valor inicial formado por uma lei de conservação da forma:

$$u_t + [f(u)]_x = 0, \quad (x,t) \in \mathbb{R} \times \mathbb{R}_0^+$$

e com dado inicial:

$$u(x,0) = u_0(x) = \begin{cases} u_l \text{ se } x \leq 0, \\ u_r \text{ se } x > 0, \end{cases} \tag{2.8}$$

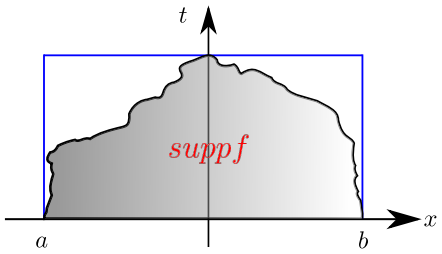
onde $u : \mathbb{R} \times \mathbb{R}_0^+ \longrightarrow \mathbb{R}^m$, $f : \mathbb{R}^m \longrightarrow \mathbb{R}^m$, u_r é o estado inicial à direita e u_l é o estado inicial à esquerda, com $u_r, u_l \in \mathbb{R}^m$. O problema (2.8) é chamado **Problema de Riemann.** Podemos ver que a condição inicial pode ter uma descontinuidade em $x = 0$. O caso trivial acontece quando $u_l = u_r$. Em [23], [25] temos que mesmo no caso escalar ($m = 1$) e a condição inicial $u_0(x)$ seja bastante regular, não poderemos supor que a solução $u(x,t)$ seja regular para todo $t > 0$.

A discussão precedente motiva a generalizar a definição de solução, pois a noção de uma solução da equação diferencial parcial (2.8), conhecida como **solução forte** ou também **solução clássica**, é uma função de classe C^1 que satisfaz (2.8). Logo, como uma função descontínua não pode ser solução no sentido clássico.

No que segue definimos um novo tipo de solução mas precisamos das seguintes definições:

Definição 2.1 *Uma função $\phi : \mathbb{R} \times \mathbb{R}_0^+ \longrightarrow \mathbb{R}$ tem supporte compacto se $\phi \equiv 0$ fora de um conjunto compacto de $\mathbb{R}$. Uma função teste é uma função $\phi : \mathbb{R} \times [0,\infty) \to \mathbb{R}$ de supporte compacto e $\phi \in C^\infty(\mathbb{R} \times [0,\infty))$. O conjunto das funções teste será denotado por $\mathcal{D}(\mathbb{R} \times [0,\infty))$.*

Figura 4 – Supporte compacto de f.



Fonte: Julio Agustín

Como [17], [25], multiplicando e integrando $u_t + [f(u)]_x = 0$ por uma função teste $\phi(x,t) \in \mathcal{D}(\mathbb{R} \times [0,\infty))$, temos

$$\int_{-\infty}^{\infty} \int_0^{\infty} [u\phi_t + f(u)\phi_x] dt dx + \int_{-\infty}^{\infty} \phi(x,0)u(x,0)dx = 0. \tag{2.9}$$

Definição 2.2 (Solução fraca) *[17]. Dizemos que $u \in L^\infty(\mathbb{R} \times (0, \infty))$ é uma solução fraca da lei de conservação definida para o problema (2.8), se ela satisfaz a equação (2.9) para todo $\phi \in \mathcal{D}(\mathbb{R} \times [0, \infty))$.*

Teorema 2.1 *[17]. Se u é uma solução de classe C^1 do problema (2.8), então satisfaz a equação (2.9), para todo $\phi \in C_0^\infty(\mathbb{R} \times [0, \infty))$.*

Demonstração:

Como u é solução clássica de (2.8), então u é continuamente diferenciável e

$$u_t + [f(u)]_x = 0, \quad x \in \mathbb{R},\ t \in \mathbb{R}_0^+$$

$$u(x, 0) = u_0(x), \quad x \in \mathbb{R},$$

Além disso, para qualquer função suave $\phi \in \mathcal{D}(\mathbb{R} \times [0, \infty))$ temos que:

$$\int_{-\infty}^{\infty} \int_0^{\infty} [u_t + f(u)_x]\phi \; dtdx = 0,$$

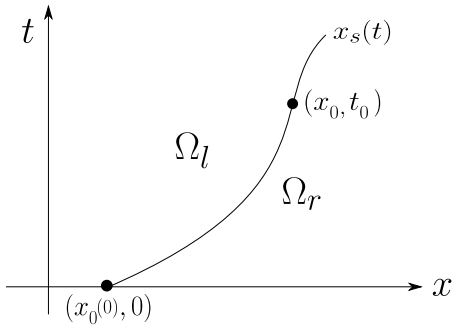
integrando por partes e usando o fato que ϕ é de suporte compacto, temos que:

$$\int_{-\infty}^{\infty} \int_0^{\infty} [u\phi_t + f(u)\phi_x] dtdx + \int_{-\infty}^{\infty} \phi(x, 0)u(x, 0)dx = 0.$$

Portanto, a solução clássica u é também uma solução fraca [17]. ■

Como as soluções fracas geralmente não são contínuas, vamos descrever essas soluções da seguinte forma. Seja $x_s(t)$ uma curva no plano xt que divide o semiplano $t > 0$ em duas partes: Ω_r à direita da curva e Ω_l à esquerda da curva, como mostra a Figura 5 e definimos uma solução contínua por partes [25].

Figura 5 – Curva que divide o plano em duas regiões, Ω_l e Ω_r.



Fonte: [15]

Definição 2.3 *Uma função $u(x,t)$ é denominada solução contínua por partes do problema (2.8) com descontinuidade do tipo salto ao longo da curva $x_s(t)$, se $u(x,t)$ tem as seguintes propriedades:*

1) $u(x,t)$ tem derivadas primeiras contínuas em Ω_r e Ω_l e é solução do problema

$$u_t + [f(u)]_x = 0, \quad x \in \Omega_r, \ t > 0,$$

$$u(x,0) = u_0(x), \quad x < x_s(0),$$

em Ω_r e do problema

$$u_t + [f(u)]_x = 0, \quad x \in \Omega_l, \ t > 0,$$

$$u(x,0) = u_0(x), \quad x > x_s(0),$$

em Ω_l.

2) Em cada ponto (x_0,t_0) de $x_s(t)$, os limites de $u(x,t)$ quando $(x,t) \longrightarrow (x_0,t_0)$, $(x,t) \in \Omega_l$ e $(x,t) \longrightarrow (x_0,t_0)$, $(x,t) \in \Omega_r$ existem mas não são necessariamente iguais.

O seguinte resultado é de grande importância quando consideramos uma solução fraca do problema de Riemann a qual pode ser visto em [15].

Teorema 2.2 *Seja $\mathcal{C}$ uma curva no plano xt ($\mathbb{R} \times \mathbb{R}_0^+$) diferenciável, $x = x(t)$ a qual divide $\Omega \in \mathbb{R} \times \mathbb{R}_0^+$ em Ω_l e Ω_r, onde u é uma solução diferenciável em Ω_l, Ω_r e u tem uma descontinuidade* ***(u é solução fraca)*** *do problema (2.8) a qual tem uma descontinidade de salto em $\mathcal{C}$, então:*

$$x'(t) = \frac{[f(u)]}{[u]}, \tag{2.10}$$

onde: $[f(u)] = f(u_r) - f(u_l)$ e $[u] = u_r - u_l, \ (u_l \neq u_r)$.

Demonstração:

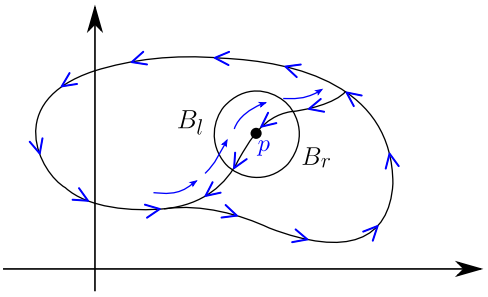
Seja p um ponto, tal que $p \in \mathcal{C}$ e B uma bola com centro p tal que $B \in \Omega$ tal que $(x(0),0) \notin B$.
Seja u uma solução fraca de (2.8) e tomando uma $\phi \in C_0^\infty$ com suporte compacto em B, temos

$$\int_0^\infty \int_{-\infty}^\infty [u\phi_t + f(u)\phi_x] dxdt = 0. \tag{2.11}$$

então

$$0 = \int_0^\infty \int_{-\infty}^\infty [u\phi_t + f(u)\phi_x] dxdt = \iint_{B_l} [u\phi_t + f(u)\phi_x] dxdt + \iint_{B_r} [u\phi_t + f(u)\phi_x] dxdt.$$

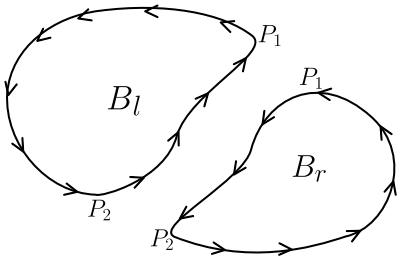
Figura 6 – Descomposição da região Ω.



Fonte: Julio Agustín

Pelo teorema de Green em B_l e B_r cujas fronteiras ∂B_l e ∂B_r estão orientadas positivamente.

Figura 7 – Orientação positiva de B_l e B_r.



Fonte: Julio Agustín

Então, temos:

$$\begin{aligned}
0 &= \int_{\partial B_l}[-u\phi dx + f(u)\phi dt] + \int_{\partial B_r}[-u\phi dx + f(u)\phi dt].\\
&= \int_{\partial B_l}\phi[-udx + f(u)dt] + \int_{\partial B_r}\phi[-udx + f(u)dt].\\
&= \int_{P_1}^{P_2}\phi[-u_l dx + f(u_l)dt] + \int_{P_1}^{P_2}\phi[-u_r dx + f(u_r)dt].\\
&= \int_{P_1}^{P_2}\phi[-(u_l - u_r)dx + (f(u_l) - f(u_r))dt].
\end{aligned}$$

Portanto como $\phi \neq 0$, então $-(u_l - u_r)dx + (f(u_l) - f(u_r))dt = 0$, logo

$$\begin{aligned}
(f(u_l) - f(u_r))dt &= (u_l - u_r)dx.\\
\frac{f(u_l) - f(u_r)}{u_l - u_r} &= \frac{dx}{dt}.\\
\frac{[f(u)]}{[u]} &= x'(t).
\end{aligned}$$

■

O Teorema (2.2) sugere que para estender a solução obtida pelo método das características numa região Ω, a curva $\mathcal{C}$ que separa as descontinuidades com velocidade $s = x'(t)$, chamada choque ou descontinuidade, na solução do problema de Riemann (2.8), a curva $\mathcal{C}$ é caracterizada pelos valores u_l e u_r como sendo os estados à direita e à esquerda, respectivamente, e com velocidade de propagação **s** [21], [24], [25].

Denotamos esta caracterização do choque por (u_l, u_r, s). Dentre as várias curvas com esta propriedade, escolhemos a que satisfaz a condição (2.10), a qual motiva a seguinte definição [21], [24], [25].

Definição 2.4 *Um choque (u_l, u_r, s) cumpre a condição de Rankine-Hugoniot se e somente se:*

$$f(u_r) - f(u_l) = s(u_r - u_l). \tag{2.12}$$

Para um estado fixo $u^\star$ temos a seguinte definição, [25].

Definição 2.5 *O lugar geométrico de Hugoniot para um estado fixo $u^\star$, ou simplesmente Hugoniot Locus, é o seguinte conjunto:*

$$H(u^\star) = \{u \in \mathbb{R}^m / \exists s \in \mathbb{R} : f(u) - f(u^\star) = s(u - u^\star)\}. \tag{2.13}$$

Voltando para o exemplo (2.7), o Locus de Hugoniot para um estado fixo $U^\star = \begin{pmatrix} \theta^\star \\ \omega^\star \\ \sigma^\star \end{pmatrix}$ é:

$$\begin{pmatrix} \mathrm{v}\theta \\ \omega \\ \sigma \end{pmatrix} - \begin{pmatrix} \mathrm{v}\theta^\star \\ \omega^\star \\ \sigma^\star \end{pmatrix} = s \left[\begin{pmatrix} \theta \\ \omega \\ \sigma \end{pmatrix} - \begin{pmatrix} \theta^\star \\ \omega^\star \\ \sigma^\star \end{pmatrix} \right]$$

então:

$$\mathrm{v}\theta - \mathrm{v}\theta^\star = s[\theta - \theta^\star] \tag{2.14}$$

$$\omega - \omega^\star = s[\omega - \omega^\star] \tag{2.15}$$

$$\sigma - \sigma^\star = s[\sigma - \sigma^\star] \tag{2.16}$$

Da primeira equação obtemos que $s = \mathrm{v}$ e das outras duas $s = 1$.

2.4 EXISTÊNCIA E UNICIDADE DE SOLUÇÕES FRACAS

Soluções fracas não são únicas e, para um mesmo dado inicial, pode ser possível obter-se mais de uma solução fraca para o problema (2.8). A condição de Rankine-Hugoniot não é suficiente para garantir a unicidade de tais soluções. É preciso, então, que outras condições sejam impostas, permitindo escolher entre todas as soluções fracas possíveis, aquela que seja fisicamente correta. Duas das mais importantes condições são as condições de entropia e do critério de viscosidade evanescente, discutidos logo a seguir.

2.4.1 **Condição de Entropia**

Em algumas situações a solução fraca não é única, sendo requerida uma condição adicional que permita a escolha da solução fisicamente admissível, ou seja, aquela na qual a solução de viscosidade tende a desaparecer. A condição que define esta solução é que deve ser a solução limite da condição de viscosidade quando $\epsilon \longrightarrow 0$, em uma norma convenientemente escolhida. No caso de equações escalares, as características do choque devem-se interceptar à medida que o tempo avança [22].

Definição 2.6 *Uma descontinuidade propagando-se com velocidade s dada pela equação (2.12) satisfaz a condição de entropia de Lax se:*

$$f'(u_l) > s > f'(u_r), \tag{2.17}$$

sendo $f'(u)$ a velocidade característica.

Se a função f for estritamente convexa, a velocidade s da condição de Rankine-Hugoniot deve estar situada entre $f'(u_l)$ e $f'(u_r)$, então, a última inequação reduz-se à exigência $f'(u_l) > f'(u_r)$ a qual devido à convexidade, exige que $u_l > u_r$ [24], [25].

Uma outra condição de entropia para o caso escalar é devido a Oleinik [21], onde a função de fluxo f deve satisfazer a seguinte definição.

Definição 2.7 (Condição de Entropia de Oleinik) *Se $u : \mathbb{R}\times\mathbb{R}_0^+ \longrightarrow \mathbb{R}^m$ é a solução entrópica, então todas as descontinuidades têm a seguinte propiedade:*

$$\frac{f(u) - f(u_l)}{u - u_l} \geq s \geq \frac{f(u) - f(u_r)}{u - u_r}, \quad \forall\ u_l < u < u_r. \tag{2.18}$$

No caso da função de fluxo f ser convexa, a equação (2.18) reduz-se à equação (2.17).

Segundo [24], a condição de entropia que relaciona a velocidade do choque com as velocidades características à direita e à esquerda do própio choque mostra-se na seguinte definição.

Definição 2.8 (Condição de Entropia de Lax) . *Diz-se que um choque (u_l, u_r, s) do problema de Riemann satisfaz o critério de entropía de Lax se:*

$$\begin{array}{ccccc} \lambda_i(u_r) & < & s & < & \lambda_{i+1}(u_r) \\ \lambda_{i-1}(u_l) & < & s & < & \lambda_i(u_l) \end{array} \tag{2.19}$$

para algum $i = 1, 2, ..., m$. Neste caso, este choque é denominado de i-choque de Lax.

Uma outra forma de condição de entropia é aquela baseada no espalhamento das características num leque de rarefação, [24], [25]. Se $u(x,t)$ for uma função crescente de x em alguma região, então as características se espalham se $f'' > 0$. A taxa de espalhamento pode ser quantificada, segundo Oleinik [21].

Definição 2.9 *$u(x,t)$ é uma solução que satisfaz a condição de entropia, se existir uma constante $E > 0$, tal que, para todo $a > 0$, $t > 0$, e $x \in \mathbb{R}$:*

$$\frac{u(x+a,t) - u(x,t)}{a} < \frac{E}{t}. \tag{2.20}$$

Em [24], [25], para uma descontinuidade propagando-se entre estados à esquerda e à direita u_l e u_r, esta última inequação será satisfeita se e somente se $u_r - u_l \leq 0$, o que leva à equação (2.17).

A inequação (2.20) é mais facilmente aplicável ao estudo de métodos numéricos que devem convergir para uma solução correta. Fazendo $a = \Delta x$, deve-se assegurar a existência de uma constante $E > 0$ tal que:

$$u_{j+1} - u_j < \left(\frac{E}{t}\right)\Delta x, \quad \forall t > 0 \text{ e } \Delta x \longrightarrow 0. \tag{2.21}$$

2.4.2 Critério de Viscosidade

Uma forma de escolher a solução correta fisicamente e matematicamente para (2.8) surge em estudar a solução viscosa. Esta solução é definida como o limite quando $\epsilon \longrightarrow 0$ das funções $u^\epsilon(x,t)$ onde $u^\epsilon(x,t)$ é solução da equação parabólica,

$$u_t^\epsilon + f(u^\epsilon)_x = \epsilon D u_{xx}^\epsilon, \tag{2.22}$$

com condição inicial $u^\epsilon(x,0) = u_0^\epsilon$, onde $u = u(x,t) \in \mathbb{R}^m, \varepsilon > 0$, e D é uma matriz constante definida positiva, denominada matriz de viscosidade. Em geral, as soluções de (2.22) são suaves devido ao termo de viscosidade $\varepsilon D u_{xx}$, [25].

Existem várias razões para escolher a solução viscosa como a correta, [21] [24], [25] . Uma delas é que as equações que estamos resolvendo modelam situações físicas que incluem algum tipo de dissipação ou difusão (natural). Uma das características mais importante da solução viscosa é o seguinte resultado.

Definição 2.10 *Um choque (u_l, u_r, s) para o sistema de leis de conservação (2.8) é chamado admissível, segundo o critério de viscosidade evanescente em relação à matriz D, se existir uma onda viajante da equação parabólica (2.22) com velocidade s e extremos u_l em $x = -\infty$ e u_r em $x = +\infty$*

No caso particular de uma lei de conservação escalar:

$$\begin{aligned} u_t + [f(u)]_x &= 0, \\ u(x,0) &= u_0(x), \end{aligned} \tag{2.23}$$

onde $u : \mathbb{R} \times \mathbb{R}_0^+ \longrightarrow \mathbb{R}^m$, $f \in C^2(\mathbb{R})$, deve-se considerar a equação parabólica (ou lei de conservação viscosa) associada:

$$\begin{aligned} u_t + [f(u)]_x &= \varepsilon u_{xx}, \quad x \in \mathbb{R},\ t > 0 \\ u(x,0) &= u_0(x), \quad x \in \mathbb{R}. \end{aligned} \tag{2.24}$$

Proposição 2.1 *Se uma solução viscosa existe, ela é uma solução fraca de (2.23).*

A demonstração da proposição anterior pode ser encontrada em [21],[23], [25].

2.5 SOLUÇÃO DO PROBLEMA DE RIEMANN

Como [23], dado o Problema de Riemann (2.8) definimos as curvas características como:

$$\begin{aligned} x'(t) &= \lambda_i(u(x(t),t)), \\ x(0) &= x_0. \end{aligned} \tag{2.25}$$

Os autovalores $\lambda_i = \lambda_i(u), i = 1, 2, ..., m$ da matriz jacobiana $f'(u)$ são denominadas **velocidades características**, mas (2.25) mostra que λ_i depende da solução do problema u, que é desconhecida a priori. Portanto, não é possível resolver o sistema (2.25) para determinar as curvas características e encontrar a solução de (2.8).

2.5.1 Campos Característicos

Considere o sistema (2.6) estritamente hiperbólico, com autovalores $\lambda_i(u)$ e autovetores $r_i(u)$ para todo $i = 1, 2, ..., m$, tal que a velocidade característica $\lambda_i(u)$ defina um λ_i-campo, chamado $i-$Campo Característico [24].

Definição 2.11 *Um $i-$campo característico é chamado* ***linearmente degenerado*** *em $u^\star \in \mathbb{R}^m$ se:*

$$\nabla\lambda_i(u^\star)r_i(u^\star) = 0. \tag{2.26}$$

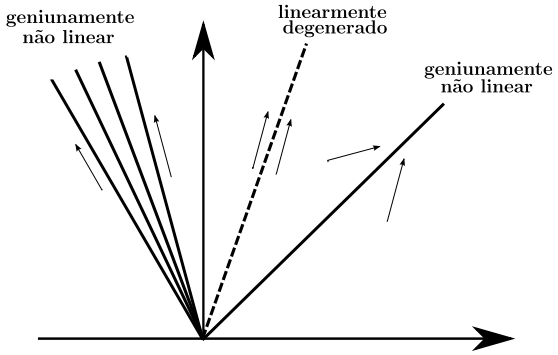
O $i-$campo carcaterístico é dito linearmente degenerado se é para todo $u \in \mathbb{R}^m$.

Definição 2.12 *Um $i-$campo característico é chamado* ***genuinamente não linear*** *em $u^\star \in \mathbb{R}^m$ se:*

$$\nabla\lambda_i(u^\star)r_i(u^\star) \neq 0. \tag{2.27}$$

O $i-$campo carcaterístico é dito genuinamente não linear se é para todo $u \in \mathbb{R}^m$.

Figura 8 – Ilustração dos comportamentos: genuinamente não linear e linearmente degenerado.



Fonte: [22]

O **plano de fase** é o espaço de vetores $u = (u_1, u_2, ..., u_m) \in \mathbb{R}^m$. Vejamos o que acontece com os autovalores de nosso sistema (2.7), onde podemos classificar os autovalores:

$$\nabla\lambda_i(u)r_i(u) = 0,\ i = 1, 2, 3 \ \text{ para todo } u \in \mathbb{R}^3, \tag{2.28}$$

então, os 1,2,3-campos característico são linearmente degenerados.

Teorema 2.3 *[24]. Se todos os autovalores de f' do sistema (2.6) são genuinamente não lineares ou linearmente degenerados, então, existe $\eta > 0$ tal que o problema de Riemann (2.8) tem solução para todo dado inicial (u_l, u_r) com $\mid u_l - u_r \mid \leq \eta$.*

Pode-se encontrar a demonstração do Teorema 2.3 em [24].

Para resolver o problema de Riemann (2.8) no caso que a função de fluxo f é não linear, primeiro estudamos a parte da solução associada a um autovalor. Existem diferentes possibilidades. Se o autovalor λ_k é linearmente degenerado, então a condição de

Rankine-Hugoniot é satisfeita. Este tipo de descontinuidade é chamada descontinuidade de contato. Se λ_k é um autovalor genuinamente não linear, têm-se duas possibilidades:

$$\lambda_k(u_l) < \lambda_k(u_r) \quad ou \qquad \lambda_k(u_l) > \lambda_k(u_r). \tag{2.29}$$

Onde as soluções recebem os nomes de **onda de rarefação** e **onda de choque**, respectivamente.

2.5.2 Ondas de Choque

No caso de sistemas lineares, os autovalores são constantes e $\nabla\lambda_i(u) = 0$, ou seja, todos os campos característicos de um sistema hiperbólico linear com coeficientes constantes são linearmente degenerados.

Seguindo [24], [25], suponha que u seja uma solução suave de um sistema hiperbólico cujos autovalores de $f'(u)$ são tais que $\lambda_1(u) \leq \lambda_2(u) \leq \cdots \leq \lambda_m(u)$. Se $x(t) \in C^1([0,\tau])$ for uma solução de $x'(t) = \lambda_i(u)$, então a curva $\{(x(t),t) : 0 \leq t \leq \tau\}$ é chamada $i-$característica.

Suponhamos que os estados à esquerda e à direita estejam próximos, no sentido que a diferença $u_l - u_r$ seja pequena. Desta forma, os estados constantes $u_0, u_1, u_2, \cdots u_n$ estão separados por pequenos saltos. A $i-$ésima descontinuidade deve satisfazer a condição de salto de Rankine-Hugoniot, isto é,

$$f(u_{i-1}) - f(u_i) = s(u_{i-1} - u_i), \tag{2.30}$$

onde s é a velocidade de propagação. Deseja-se encontrar $u_{i-1} - u_i$ para cada i.

Inicialmente, construimos uma solução fraca do problema de Riemann que consiste em i descontinuidades que se propagam com velocidades $s_1 < s_2 < s_3 < \cdots < s_m$, lembrando que uma descontinuidade que se propaga com velocidade $\mathbf{s}$ e possua valores constantes u_l e u_r em cada lado da descontinuidade deve satisfazer a condição de Rankine-Hugoniot. Supondo que o ponto $u_l \in \mathbb{R}^m$ esteja fixo, determinamos o conjunto de pontos u_r que podem ser conectados a u_l por uma descontinuidade que satisfaça tal condição.

A condição de Rankine-Hugoniot estabelece que:

$$f((u_r)_i(\xi)) - f(u_l) = s_i(\xi)((u_r)_i(\xi) - u_l), \quad i = 1, 2, \cdots, m \tag{2.31}$$

onde $(u_r)_i(\xi)$ é uma parametrização da $i-$ésima curva que passa por u_l e que satisfaz $(u_r)_i(0) = u_l$, sendo ξ o parâmetro.

Diferenciando a igualdade (2.31) com respeito a ξ e substituindo por $\xi = 0$, temos:

$$f'(u_l)(u_r)_i'(0) = s_i(0)(u_r)_i'(0). \tag{2.32}$$

Isto significa que $(u_r)_i'(0)$ deve ser um múltiplo escalar do autovetor $r_i(u_l)$ de $f'(u_l)$, enquanto $s_i(0)$ deve ser o autovalor $\lambda_i(u_l)$. Portanto, concluímos que a curva $(u_r)_i(\xi)$ deve ser tangente a $r_i(u_l)$ no ponto u_l.

Uma solução do problema de Riemann pode não ser única. Em geral são requeridas condições iniciais para o problema. Para uma única lei de conservação, no caso em que f é estritamente convexa, esta condição se dá pela seguinte desigualdade,[17]:

$$f'(u_l) > s > f'(u_r). \tag{2.33}$$

Para o caso de sistemas de leis de conservação, esta condição aplicada a um campo genuinamente não linear permite afirmar que o salto no $k-$ésimo campo (de u_r a u_l, por exemplo) é admissível se:

$$\lambda_i(u_l) > s > \lambda_i(u_r). \tag{2.34}$$

Cada uma das leis de conservação deve satisfazer a equação

$$s\,[u_i] = [f_i], \quad i = 1, 2, \cdots, m \tag{2.35}$$

Assim, dado um $i-$campo característico genuinamente não linear, definimos a solução do problema de Riemann (2.8) como:

$$u(x,t) = \begin{cases} u_r, & x < st, \\ u_l, & x > st. \end{cases} \tag{2.36}$$

A solução (2.36) é chamada $i-$**onda de choque**.

2.5.3 Ondas de Rarefação

O Problema de Riemann (2.8) além de possuir soluções de tipo ondas de choque, também pode possuir soluções conhecidas como ondas de rarefação que estudaremos a seguir. Segundo [17], estas soluções tem a seguinte propriedade: a solução é constante ao longo de todos os raios da forma $x = \xi t$. Logo, a solução é uma função apenas de $\frac{x}{t}$, e diz-se uma solução da EDP. Uma **onda de rarefação** é solução de um sistema de equações que também tem essa propriedade e tem a forma

$$u(x,t) = \begin{cases} u_l & \text{se } x \leq \xi_1 t, \\ w(x,t) & \text{se } \xi_1 t < x < \xi_2 t, \\ u_r & \text{se } x \geq \xi_2 t, \end{cases} \tag{2.37}$$

onde w é uma função suave com $w(\xi_1) = u_l$ e $w(\xi_2) = u_r$.

Agora podemos fazer a seguinte pergunta: Quando é que um sistema de equações tem uma solução deste tipo? Como no caso de choques, para estados arbitrários u_l e u_r o

sistema (2.25) pode não ter solução desta forma. Mas, em geral, a partir de cada ponto u_l existem $m-$curvas que consistem em pontos de u_r que podem ser ligados a u_l por uma onda de rarefação, os quais são chamadas de *curvas integrais* do campos de vetores $r_i(u)$.

Definição 2.13 (Curva Integral) *Uma curva integral para $r_i(u)$ é uma curva que tem a propriedade de que a tangente à curva no ponto de qualquer u encontra-se na direção $r_i(u)$. A existência de curvas suaves desta forma segue a partir da suavidade de f e estritamente hiperbólica, desde que $r_i(u)$ seja uma função suave de u. Se $u_p(\xi)$ é uma parametrização de uma curva integral na $i-th$ família, então, o vetor tangente é proporcional à $r_i(u_p(\xi))$, em cada ponto, isto é,*

$$u_i'(\xi) = \alpha(\xi) r_i(u_i(\xi)), \tag{2.38}$$

onde $\alpha(\xi)$ é um fator escalar.

Para ver que as curvas de rarefação se encontram ao longo das curvas integrais e determinar explicitamente a função $w(x/t)$ em (2.37), diferenciamos $u(x,t) = w(x/t)$, obtendo:

$$\begin{array}{rcl} u_t(x,t) & = & -\frac{x}{t^2} w'(x/t), \\ u_x(x,t) & = & \frac{1}{t} w'(x/t), \end{array} \tag{2.39}$$

de modo que $u_t + f'(u)u_x = 0$, resulta

$$-\frac{x}{t^2} w'(x/t) + \frac{1}{t} f'(w(x/t)) w'(x/t) = 0, \tag{2.40}$$

multiplicando por t, obtemos

$$f'(w(\xi)) w'(\xi) = \xi w'(\xi), \tag{2.41}$$

onde $\xi = x/t$. Uma possível solução da última igualdade é $w'(\xi) = 0$, isto e, w constante. Qualquer função constante é uma solução semelhante da lei de conservação e na verdade a onda de rarefação (2.37) toma esta forma para $\xi < \xi_1$ e $\xi > \xi_2$. No meio, w é presumivelmente de variação suave e $w' \neq 0$. Então a última igualdade nos diz que $w'(\xi)$ deve ser proporcional a algum autovetor $r_i(w(\xi))$ de $f'(w(\xi))$,

$$w'(\xi) = \alpha(\xi) r_i(w(\xi)), \tag{2.42}$$

consequentemente o valor de $w(\xi)$ se encontra na curva integral r_i. Em particular, ambos estados $u_l = w(\xi_1)$ e $u_r = w(\xi_2)$, encontram-se na mesma curva integral. Esta condição é necessária para a existência de uma onda de rarefação conectada a u_l e u_r, mas note que isso não é suficiente. Precisamos que $\xi = (x/t)$ seja monótona crescente enquanto $w(\xi)$ se move de u_l para u_r ao longo da curva integral; caso contrário, a função (2.37) não tem um

valor único. Observe que a parametrização da curva integral não é arbitrária desde que ξ é um valor próprio de $f'(w(\xi))$,

$$\xi = \lambda_i(w(\xi)). \tag{2.43}$$

Esta parametrização é forçada pela definição $\xi = x/t$. Note ainda que a igualdade acima implica que w é constante ao longo do raio $x = \lambda_i(w)t$ e, portanto, cada valor constante de w se propaga com a velocidade $\lambda_i(w)$. Pela última igualdade, acima a monoticidade de ξ é equivalente à monoticidade de $\lambda_i(w)$ quando w se move de u_l para u_r. A partir de um determinado estado u_l pode-se passar ao longo da curva integral no sentido que aumenta λ_i. Se λ_i tem um máximo local em u_l na direção r_i então não há onda de rarefação com estado u_l à esquerda. No caso não-linear, há uma família de parâmetros que podem ser conectados a u_l por uma $i-$rarefação.

Se o $i\,th-$campo é genuinamente não-linear, então, λ_i varia monotonamente ao longo de toda a curva integral. E não precisamos nos preocupar com máximos locais, pois u_l e u_r sempre podem ser conectados por uma onda de rarefação desde que se encontrem na mesma curva integral e

$$\lambda_i(u_l) < \lambda_i(u_r). \tag{2.44}$$

Se o $i\,th-$campo é linearmente degenerada, então, λ_i é constante em cada curva integral e não existem possíveis ondas rarefação nesta família. A fim de determinar a função $w(\xi)$, determinamos primeiro o fator de escala $\alpha(\xi)$ em (2.42) pela diferenciação (2.43) com respeito a ξ, temos

$$\begin{aligned} 1 &= \nabla\lambda_p(w(\xi)) \cdot w'(\xi) \\ &= \alpha(\xi)\nabla\lambda_p(w(\xi)) \cdot r_p(w(\xi)) \end{aligned}$$

e usando (2.42), obtemos

$$\alpha(\xi) = \frac{1}{\nabla\lambda_p(w(\xi)) \cdot r_p(w(\xi))}. \tag{2.45}$$

Aplicando em (2.42), obtemos um sistema de equações diferenciais ordinárias para $w(\xi)$:

$$w'(\xi) = \frac{r_p(w(\xi))}{\nabla\lambda_p(w(\xi)) \cdot r_p(w(\xi))}, \quad \xi_1 \leq \xi \leq \xi_2 \tag{2.46}$$

com dado inicial:

$$w(\xi_1) = u_l$$

onde $\xi_1 = \lambda_p(u_l)$ e $\xi_2 = \lambda_p(u_r)$. Note que o denominador em (2.46) é finito para $\xi_1 \leq \xi \leq \xi_2$ só se λ_p é monótona entre ξ_1 e ξ_2.

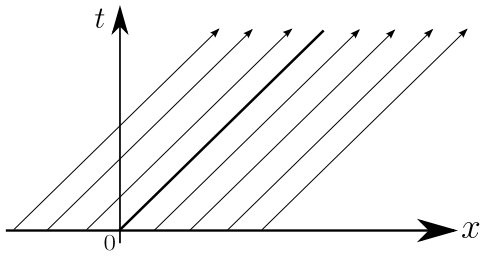
2.5.4 **Descontinuidade de Contato**

Como outro tipo de descontinuidade para o problema de Riemann (2.8), temos a Descontinuidade de Contato.

Definição 2.14 *Diz-se ser uma Descontinuidade de Contato quando a velocidade $f'(u)$ é constante para todo $u \in \Omega$ no problema de valor inicial (2.8).*

As curvas características são paralelas, pois $f'(u_l) = f'(u_r)$ [15]. Podemos ver isso na seguinte Figura 9.

Figura 9 – Curvas características para uma onda de contato, onde $f'(u) > 0$ quando o salto ocorre em $x_0 = 0$.



Fonte: [15]

Se consideramos $f'(u) = c$, com $c > 0$, temos pelo método das características

$$\begin{cases} \frac{dx}{dt} = f'(u) = c \\ x(0) = x_0. \end{cases} \tag{2.47}$$

Desta forma as curvas características são

$$x_s(t) = \begin{cases} f'(u_l)t + x_0 = ct + x_0, & \text{se } x_0 < 0 \\ f'(u_r)t + x_0 = ct + x_0, & \text{se } x_0 > 0, \end{cases} \tag{2.48}$$

isto é, $x_s = ct + x_0$, se $x_0 \neq 0$.

Pela condição Rankine-Hugoniot, temos que

$$s = \frac{f(u_l) - f(u_r)}{u_l - u_r} = c. \tag{2.49}$$

Além disso, cumpre a condição de Entropia (2.18), pois

$$
\begin{array}{rcccccl}
 & & c & \geq & s & \geq & c \\
c\left(\frac{u-u_l}{u-u_l}\right) & = & c & \geq & s & \geq & c \quad = \quad c\left(\frac{u-u_r}{u-u_l}\right) \\
 & & \frac{cu-cu_l}{u-u_l} & \geq & s & \geq & \frac{cu-cu_r}{u-u_r} \\
 & & \frac{f(u)-f(u_l)}{u-u_l} & \geq & s & \geq & \frac{f(u)-f(u_r)}{u-u_r}
\end{array}
$$

para todo u entre u_l e u_r.

Logo a solução será em forma de choque. Portanto, a solução é

$$
u(x,t) = u_0(x-ct) = \begin{cases} u_l, & \text{se} \quad x < ct, \\ u_r, & \text{se} \quad x > ct. \end{cases} \tag{2.50}
$$

Voltando ao exemplo (2.7), já vimos em (2.28) que todos os λ_i são linearmente degenerados, então temos a descontinuidade de contato dado por:

$$
U(x,t) = \begin{cases} U_l, & x < st, \\ U_r, & x > st. \end{cases} \tag{2.51}
$$

onde $U_l = \begin{pmatrix} \theta_l \\ \omega_l \\ \sigma_l \end{pmatrix}$, $U_r = \begin{pmatrix} \theta_r \\ \omega_r \\ \sigma_r \end{pmatrix}$ e $s = \begin{pmatrix} \text{v} \\ 1 \\ 1 \end{pmatrix}$. ■

3 ESQUEMAS DE DIFERENÇAS FINITAS

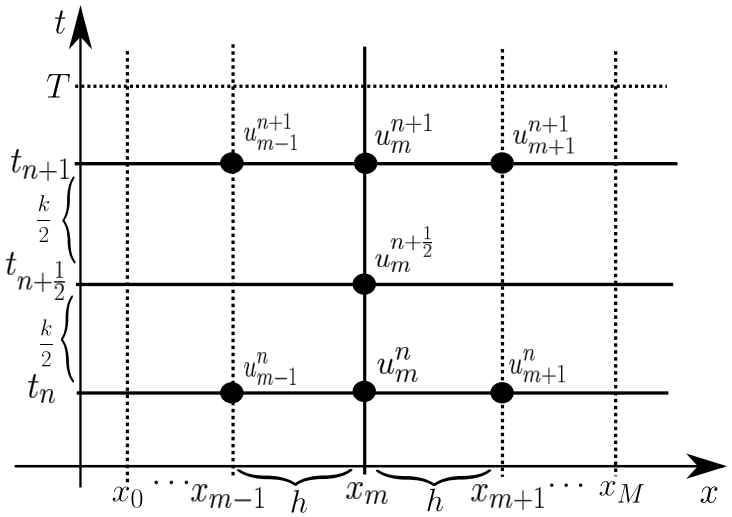
Apresentamos seguidamente a ideia geral do método de diferenças finitas que é a discretização do domínio e a substituição das derivadas presentes na equação diferencial por aproximações envolvendo somente valores numéricos da função.

Devido ao fato que nem sempre é possível obter uma solução analítica de uma EDP do tipo parabólica ou hiperbólica não linear, mas existem métodos numéricos que nos permitem encontrar uma aproximação da solução exata, em busca de resolver numericamente uma EDP, é necessário, primeiramente, expressá-la na forma de operações aritméticas para que o computador possa resolvê-la. Para isso, basta representá-las por expressões algébricas, ou seja, discretizar a região onde se procura a solução (chamado de passo de discretização do domínio) em função dos pontos de malha.

A equação de diferenças finitas deve representar a solução exata da EDP em cada ponto da região discretizada em que se deseja obter a solução do problema.

Do Capítulo 2, sabemos que nossa região de estudo é um retângulo $R = \Omega \times [0, T]$ contida no semiplano superior $\mathbb{R} \times \mathbb{R}_0^+$. Seja $\Omega = [x_0, x_M] \subset \mathbb{R}$ e seja $(x, t) \in R$ onde x denota a variável espacial e t a variável temporal. Seguindo [6], subdividimos os intervalos $[x_0, x_M]$ e $[0, T]$ em M e N subintervalos de comprimentos $h = \Delta x = \frac{(x_M - x_0)}{M}$ e $k = \Delta t = \frac{T}{N}$ respectivamente, formando-se assim uma malha na região R como mostra a Figura 10. Seguindo [6], [17], [19], [26] denotaremos os pontos da malha como $(x_m, t_n) = (x_0 + hm, nk)$.

Figura 10 – Representação da malha do esquema de diferenças finitas.



Fonte: [17]

Então, o valor de uma função u num ponto da malha (x_m, t_n) será denotado por $u_m^n = u(x_m, t_n)$.

Agora, descreveremos brevemente o esquemas de diferenças finitas para a equação

seguinte:

$$\frac{\partial u}{\partial t}(x,t) = \frac{\partial^2 u}{\partial x^2}(x,t), \quad x \in [x_0, x_M], \ t \in [0,T] \tag{3.1}$$

O esquema para a equação (3.1) será descrito com maior detalhe no Capítulo 5, mas por agora, vejamos como funciona o método de diferenças finitas na equação (3.1).

A ideia é substituir as derivadas presentes na equação diferencial pelas suas aproximações em diferenças finitas, a qual podem ser feitas de muitas maneiras. Neste trabalho usamos o método de Crank-Nicolson, o qual estabelece que as derivadas espaciais no ponto $\left(x_m, t_{n+\frac{1}{2}}\right)$ é a média das derivadas espaciais nos pontos (x_m, t_n) e (x_m, t_{n+1}), isto é:

$$\frac{\partial^2 u}{\partial x^2}\left(x_m, t_{n+\frac{1}{2}}\right) \approx \frac{u_{m+1}^{n+1} - 2u_m^{n+1} + u_{m-1}^{n+1}}{2h^2} + \frac{u_{m+1}^{n} - 2u_m^{n} + u_{m-1}^{n}}{2h^2}. \tag{3.2}$$

A derivada temporal será aproximada mediante diferença central com passo de comprimento $k/2$, isto é:

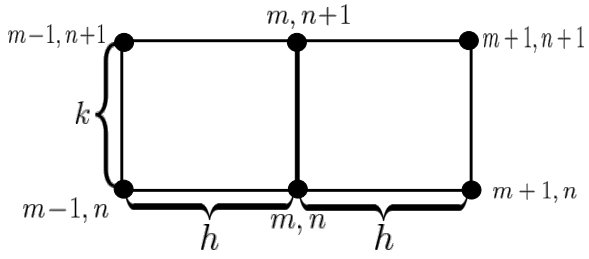
$$\frac{\partial u}{\partial t}\left(x_m, t_{n+\frac{1}{2}}\right) \approx \frac{u_m^{n+1} - u_m^{n}}{k}. \tag{3.3}$$

Substituindo as equações (3.2), (3.3) em (3.1) temos:

$$\frac{u_m^{n+1} - u_m^{n}}{k} = \frac{1}{2h^2}(u_{m+1}^{n+1} - 2u_m^{n+1} + u_{m-1}^{n+1} + u_{m+1}^{n} - 2u_m^{n} + u_{m-1}^{n}),$$

cuja molécula computacional [6], é dada pela figura (11):

Figura 11 – Molécula computacional do método de Crank-Nicolson.



Fonte: [6]

e considerando a condição de fronteira, obtemos o seguinte esquema de diferenças finitas, que é diagonalmente dominante:

$$AU^{n+1} = BU^n + c^n, \tag{3.4}$$

onde as matrizes A, B e c serão dadas com mais detalhe no Capítulo 5. Então para se obter a solução em cada estágio, é preciso resolver um sistema tridiagonal. Note que, sendo A tridiagonalmente dominante, o problema discreto tem solução única, [3].

Seguindo [26], definamos os vetores:

$$U^n = \{U_m^n / m \in J_\Omega\} \quad \text{e} \quad u^n = \{u_m^n / m \in J_\Omega\}, \tag{3.5}$$

onde U^n é a solução do esquema (3.4) em cada passo de tempo "n" e J_Ω representa o conjunto de índices dos pontos $u(x_m; t_n)$ em que são desconhecidos. Isto é, se as condições de fronteira em x_0 e x_M são tipo Dirichlet então $J_\Omega = \{1, 2, ..., M-1\}$, se as condições de fronteira são tipo Neuman em x_0 e x_M então $J_\Omega = \{0, 1, 2, ..., M\}$.

Esperando que em cada passo de tempo "n" o vetor U^n seja uma boa aproximação do vetor u^n, onde u representa solução exata da equação (3.1), isto é:

$$U^n \approx u^n \Longleftrightarrow U_m^n \approx u_m^n \quad \text{para todo} \quad m \in J_\Omega \tag{3.6}$$

Assim, para comparar os valores U_m^n com os valores reais da solução $u_m^n = u(x_m, t_n)$ do problema (3.1) em cada instante de tempo "n", seguindo [17] e [19], faremos uso da **norma em** l^2 ($\| \cdot \|_2$), isto é:

$$\| U^n - u^n \|_2 = \left\{ h \sum_{m \in J_\Omega} | U^n - u^n |^2 \right\}^{\frac{1}{2}} \tag{3.7}$$

A seguir estudaremos os conceitos de: *Consistência, Estabilidade e Convergência*, para um esquema de diferenças finitas.

3.1 CONSISTÊNCIA, ESTABILIDADE E CONVERGÊNCIA

Consideremos o seguinte esquema de equações diferenciais parciais, como em [19], [26]:

$$\begin{aligned} \frac{\partial u}{\partial t} &= L(u) \quad \text{em} \quad \Omega \times (0, T), \\ g(u) &= g_0 \quad \text{em} \quad \partial\Omega_1 \subset \Omega, \\ u &= u^0 \quad \text{em} \quad \Omega \ \text{ quando } \ t = 0, \end{aligned} \tag{3.8}$$

onde:

- O operador L depende das derivadas parciais espaciais de u mas não depende da derivada temporal.
- $g(u)$ descreve o comportamento de u na fronteira $\partial\Omega$ ou numa parte dela, $\partial\Omega_1 \subset \partial\Omega$.
- u^0 é o comportamento de u no instante inicial na região Ω, isto é, em $t = 0$.

O esquema de diferenças finitas associado a (3.8) é dado por (3.4) onde as matrizes A e B não dependem de n, pois por hipótese, o operador L não depende do tempo. As matrizes A e B podem depender dos pontos da malha onde são aplicadas. O vetor c^n é chamado *termo livre* e pode depender do espaço e tempo.

Para o esquema (3.4) suponhamos que a matriz A é invertível, então obtemos:

$$U^{n+1} \;=\; A^{-1}\left[BU^n + c^n\right]. \tag{3.9}$$

Suponhamos que a matriz A é bem condicionada, isto é, existe uma constante $K > 0$ tal que:

$$\| A^{-1} \| \;\leq\; K\Delta t, \tag{3.10}$$

para qualquer norma matricial.

Definição 3.1 *O Erro de Truncamento é definido como:*

$$\tau^n = Au^{n+1} - (Bu^n + c^n), \tag{3.11}$$

onde u é uma solução exata do problema (3.8)

O erro de truncamento é uma medida de como a equação em diferenças finitas se aproxima localmente da EDP [19]. Além disso, um esquema numérico é dito ser consistente com a EDP se no limite quando h e k tendem a zero, a equação discretizada aproxima-se da equação original. Em termos matemáticos, a consistência fica definida como em [19].

Definição 3.2 *O esquema (3.4) é dito* ***consistente*** *com o problema (3.8) se e somente se:*

$$\tau_j^n \longrightarrow 0 \quad \textit{quando} \quad \Delta t \longrightarrow 0 \quad \textit{para todo} \quad j \in J_\Omega, \tag{3.12}$$

para toda solução u de (3.8).

Definição 3.3 *Sejam p e q os maiores inteiros tais que existe uma constante positiva C que satisfaz:*

$$| \tau_j^n | \;\leq\; C[(\Delta x)^q + (\Delta t)^p] \quad \textit{quando} \quad \Delta t \longrightarrow 0, \quad \forall j \in J_\Omega, \tag{3.13}$$

para toda solução u do sistema (3.8). Então, o esquema (3.4) é dito de ordem de aproximação (p,q).

Agora, interpretaremos a relação (3.6).

Definição 3.4 *Dizemos que a solução U^n do esquema de diferenças finitas (3.4) é* ***convergente*** *à solução u^n de (3.8) se e somente se:*

$$\| U^n - u^n \| \longrightarrow 0 \quad quando \quad \Delta t \longrightarrow 0,\ n\Delta t \longrightarrow t \in (0, T], \tag{3.14}$$

para todo u^0 para o qual (3.8) esteja bem posto.

Observamos que a consistência é uma condição necessária para convergência, pois ela nos diz quão bom o esquema de diferenças finitas representa a EDP original [17], [19], [26].

Outro conceito importante é a estabilidade, que descreve como uma perturbação na solução é amplificada ao longo da simulação.

Definição 3.5 *Sejam V^n e W^n soluções de (3.9) com diferentes pontos iniciais V^0 e W^0 e mesmo termo livre c^n. Dizemos que o esquema de diferenças finitas é* ***estável*** *se exite uma constante $K > 0$ tal que:*

$$\| V^n - W^n \| \quad \leq \quad K \| V^0 - W^0 \|, \quad n\Delta t \leq T, \tag{3.15}$$

onde a constante K não depende de V^0 e W^0 nem de Δt da malha dada.

Definição 3.6 *Dizemos que o problema (3.8) está bem posto numa norma dada se para todo h (discretização espacial) suficientemente pequeno cumpre-se as seguintes condições:*

i) Existe uma solução para todo u^0 para o qual $\| u^0 \|$ é limitado independente de h.

ii) Existe uma constante $K > 0$ tal que para qualquer par de soluções v e w temos:

$$\| v^n - w^n \| \leq \ K \| v^0 - w^0 \|, \quad t_n \leq T. \tag{3.16}$$

As definições de consistência, convergência e estabilidade vistas foram relacionadas no seguinte teorema devido a Lax em 1953, [19].

Teorema 3.1 (Teorema de Equivalência de Lax-Richtmayer) *Seja um problema evolutivo linear bem posto e uniformemente bem condicionado e um esquema de diferenças finitas consistente segundo (3.10), então a estabilidade do esquema é necessária e suficiente para a convergência.*

A demonstração do Teorema 3.1 pode ser encontrada em [19]. Mencionamos que o Teorema 3.1 é válido só para problemas lineares como é o caso da EDP (3.1). Para problemas não lineares não temos um teorema que garanta a convergência, [27].

3.2 ERRO DE TRUNCAMENTO

Agora, vamos estudar a análise de erro para o método de Crank-Nicolson aplicado à equação (3.1). Definimos τ_m^n por:

$$\frac{u_m^{n+1}-u_m^n}{k} = \frac{1}{2h^2}(u_{m+1}^{n+1}-2u_m^{n+1}+u_{m-1}^{n+1}+u_{m+1}^n-2u_m^n+u_{m-1}^n)+\tau_m^n.$$

Expandindo em série de Taylor em torno do ponto $(x_m, t_{n+\frac{1}{2}})$, obtemos:

$$\frac{u_m^{n+1}-u_m^n}{k} = u_t + \frac{1}{6}\left(\frac{k}{2}\right)^2 u_{ttt} + O(k^4),$$

$$u_{m+1}^n - 2u_m^n + u_{m-1}^n = h^2 u_{xx} + \frac{h^4}{12}u_{xxxx} + O(h^6),$$

$$u_{m+1}^{n+1} - 2u_m^{n+1} + u_{m-1}^{n+1} = h^2 u_{xx} + \frac{h^4}{12}u_{xxxx}\frac{h^2k^2}{8}u_{xxtt} + \frac{h^4k^2}{96}u_{xxxxtt} + O(h^6).$$

Logo,

$$\begin{aligned}\tau_m^n &= \frac{k^2}{24}u_{ttt} - \frac{h^{12}}{12}u_{xxxx} - \frac{k^2}{8}u_{xxtt} - \frac{h^2k^2}{96}u_{xxxxtt} + O(h^4) + O(k^4).\\ &= O(h^2+k^2).\end{aligned} \tag{3.17}$$

A expressão acima nos diz que o método de Crank- Nicolson é consistente de ordem 2 em $h = \Delta x$ e $k = \Delta t$. Além disso, observe que quando h e k tendem para zero temos que o erro de truncamento tende para a EDP original, assim temos que o método de Crank-Nicolson é consistente com a equação (3.1). Em [19], [26], podemos encontrar que o método de Crank-Nicolson é *incondicionalmente estável*. Então pelo Teorema 3.1, o esquema (3.4) é convergente a solução exata de (3.1).

Deste modo, como em [19], [26], com Δt constante, podemos analisar o erro em relação à variação de Δx. A partir de (3.17) temos para cada Δx que a solução exata u no tempo "t_n" denotada por u^n pode ser escrita como segue:

$$u^n = U_{\Delta x}^n + O((\Delta x)^2), \tag{3.18}$$

onde $U_{\Delta x}^n$ representa a solução aproximada U^n, obtida pelo esquema de diferenças finitas de Crank-Nicolson com comprimento de subintervalo Δx.

Considerando $\Delta x = h, \frac{h}{2}, \frac{h}{4}$ obtemos:

$$u^n = U_h^n + O(h^2). \tag{3.19}$$

$$u^n = U^n_{2\frac{h}{2}} + O\left(\frac{h}{2}\right)^2 = U^n_{2\frac{h}{2}} + \frac{1}{4}O(h^2). \tag{3.20}$$

$$u^n = U^n_{4\frac{h}{4}} + O\left(\frac{h}{4}\right)^2 = U^n_{4\frac{h}{4}} + \frac{1}{16}O(h^2). \tag{3.21}$$

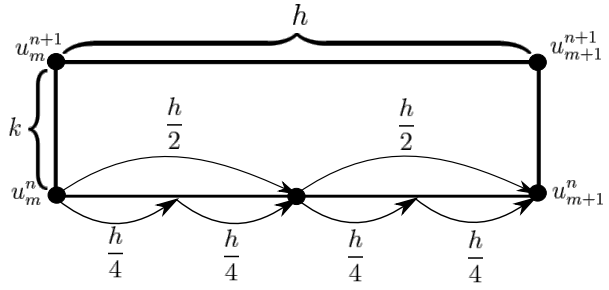
Não conhecemos a solução exata mas sabemos que sempre é a mesma para qualquer Δx, então subtraindo (3.20) de (3.19) e (3.21) de (3.20) obtemos:

$$\| U^n_{2\frac{h}{2}} - U^n_h \| = \frac{3}{4}O(h^2), \tag{3.22}$$

$$\| U^n_{4\frac{h}{4}} - U^n_{2\frac{h}{2}} \| = \frac{3}{16}O(h^2). \tag{3.23}$$

Observe que na solução aproximada $U^n_{2\frac{h}{2}}$ temos o dobro de pontos que em U^n_h, segundo pode-se deduzir da Figura 12, que para (3.22) tenha sentido devemos considerar os mesmos pontos em $U^n_{2\frac{h}{2}}$ e U^n_h. O mesmo acontece para $U^n_{4\frac{h}{4}}$ e $U^n_{2\frac{h}{2}}$ e assim sucessivamente.

Figura 12 – Discretização para a análise de erro



Fonte: [6]

As equações (3.22) e (3.23) são chamadas de *Erro Relativo* que são denotadas por $E_{\Delta x}$. Em relação ao erro relativo faremos a análise de erro para o método de Crank-Nicolson, o qual é aplicado no Capítulo 5. A Tabela 1, mostra os valores de erro relativo $E_{\Delta x}$ para os Δx indicados na primeira coluna.

Tabela 1 – Erro Relativo para o método de Crank-Nicolson

Δx	$E_{\Delta x}$	
h	$\| U^n_{2\frac{h}{2}} - U^n_h \|$	$\frac{3}{4}O(h^2)$
$\frac{h}{2}$	$\| U^n_{4\frac{h}{4}} - U^n_{2\frac{h}{2}} \|$	$\frac{3}{16}O(h^2)$
$\frac{h}{4}$	$\| U^n_{8\frac{h}{8}} - U^n_{4\frac{h}{4}} \|$	$\frac{3}{64}O(h^2)$
$\frac{h}{8}$	$\| U^n_{16\frac{h}{16}} - U^n_{8\frac{h}{8}} \|$	$\frac{3}{256}O(h^2)$

Na tabela 1 vemos que $\frac{E_{\Delta x}}{E_{\frac{\Delta x}{2}}} = 4$, isto é, observamos que $E_{\frac{\Delta x}{2}}$ seja a quarta parte do erro $E_{\Delta x}$.

Vemos que para resolver uma EDP por um esquema de diferenças finitas, obtemos de maneira natural um sistema de equações a ser resolvido. Como em [23], para resolver o sistema de equações resultante usaremos o método de Newton que descrevemos a seguir.

3.3 MÉTODO DE NEWTON

Consideremos o seguinte problema:

$$\text{Encontrar} \quad x^\star \in \mathbb{R}^n \quad \text{tal que} \quad F(x^\star) = 0, \tag{3.24}$$

onde $F : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ é uma função não linear.

Suponhamos que $F \in C^1(D)$, isto é, F é uma função continuamente diferenciável em D. Denotamos por $J_F(x)$ a matriz jacobiana da função F aplicada no ponto $x = (x_1, x_2, ...x_n) \in \mathbb{R}^n$, definida como segue:

$$(J_F(x))_{ij} = \left(\frac{\partial F_i}{\partial x_j}(x)\right), \quad \text{para todo} \quad i, j = 1, 2, ..., n;$$

O método de Newton pode ser descrito como segue:

Algoritmo 3.1 (Algoritmo Método de Newton.)

Passo 1. *Dar um ponto inicial $x_0 \in \mathbb{R}^n$, uma tolerância tol e fazer $k = 0$.*

Passo 2. *Resolver o sistema linear: $J(x_k)d_k = -F(x_k)$.*

Passo 3. *Fazer $x_{k+1} = x_k + d_k$.*

Passo 4. *Se $\| d_k \| < tol$ então a solução é x_k. Caso contrário, fazer $k = k + 1$ e voltar ao Passo (2).*

Observe que em cada iteração precisamos da matriz jacobiana $J(x_k)$.

Teorema 3.2 *Seja $F : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ uma função de clase C^2 num conjunto aberto e convexo $D \subset \mathbb{R}^n$ que contém $x^\star$. Suponhamos que $J_F(x^\star)$ exista e que além disso existam constantes positivas R, C e L tais que:*

$$\| J_F(x^\star) \| \leq C, \tag{3.25}$$

$$\| J_F(x) - J_F(y) \| \leq L \| x - y \| \quad \textit{para todo} \quad x, y \in B(x^\star, R). \tag{3.26}$$

Então existe $r > 0$ tal que para qualquer ponto inicial $x_0 \in B(x^\star, r)$ a sequência gerada pelo algoritmo (3.1) converge para a solução $x^\star$ com:

$$\| x_{k+1} - x^\star \| \leq \| x_k - x^\star \|^2 . \tag{3.27}$$

A demostração do teorema 3.2, pode ser encontrado em [23].

Do Teorema (3.2) podemos deduzir que uma característica pela qual o método de Newton não é melhor, é porque há o problema de que a vizinhança onde o método converge é pequena.

No seguinte capítulo apresentaremos uma variação do método de Newton para resolver um sistema de equações não lineares.

4 PROBLEMA DE COMPLEMENTARIDADE MISTA NÃO LINEAR

Uma grande variedade de problemas que aparecem em muitos campos da ciência, como por exemplo, Engenharia, Física, Economia, Programação Matemática, Mecânica, teoria da Elasticidade, podem ser formulados como Problemas de Complementaridade Mista(PCM), [12], [13], [16], [20]. O estudo de resolver numericamente os PCM tem muitos enfoques e entre os mais comuns temos o uso de funções de mérito ou ver o PCM como um sistema de equações não lineares. Neste último sentido, é desenvolvido em [20] o Algoritmo de Direções Admissíveis para Problemas de Complementaridade Mista Não Linear (FDA-MNCP), o qual é extensão do FDA-NCP [14] para problemas que envolvam mais variáveis e uma condição de igualdade, além da condição de complementaridade.

O objetivo deste capítulo é descrever o algoritmo FDA-MNCP, o qual gera uma sequência de pontos que cumprem duas condições:

1) Uma condição de viabilidade.

2) Uma condição de decrescimento para uma função associada ao Problema de Complementaridade Mista.

Além disso, se apresentam os teoremas que garantem a convergência do algoritmo. No Capítulo 5 será estudado um modelo de equações em derivadas parciais de tipo parabólico, os que pode ser tratado como problema de complementaridade mista não linear respectivamente, portanto, será possível aplicar o algoritmo FDA-MNCP apresentado neste capítulo. Desta maneira, o Capítulo 5 fará uso dos aspectos teórico e computacional dos Capítulos 2, 3 e 4, previamente apresentados.

4.1 CONCEITOS BÁSICOS

A seguir apresentaremos os conceitos básicos para compreensão do algoritmo FDA-MNCP. Tal como em [20], consideremos $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ e $Q : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$, aplicações de classe C^1 em $\mathbb{R}^n \times \mathbb{R}^m$.

Definição 4.1 *O problema de complementaridade mista é: encontrar* $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ *tal que:* $x \geq 0$, $F(x, y) \geq 0$ *e*

$$\begin{cases} x \bullet F(x, y) = 0, \\ Q(x, y) \quad = 0, \end{cases} \tag{4.1}$$

onde $x \bullet F(x, y) = \begin{pmatrix} x_1 F_1(x, y) \\ \vdots \\ x_n F_n(x, y) \end{pmatrix}$ *representa o produto de Hadamard.*

Se $F(x,y)$ e $Q(x,y)$ são lineares temos o Problema de Complementaridade Mista Linear (MLCP), caso contrário temos um Problema de Complementaridade Mista Não Linear (MNCP).

Definição 4.2 (Ponto Viável). *Seja $\Omega \subset \mathbb{R}^n \times \mathbb{R}^m$. Chamaremos de conjunto de pontos viáveis do problema de complementaridade mista dado por F o seguinte conjunto:*

$$\Omega := \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m / x \geq 0, F(x,y) \geq 0\}. \tag{4.2}$$

Definição 4.3 (Ponto Estritamente Viável). *Se $(x,y) \in \Omega$ verifica as seguintes condições $x > 0$ e $F(x,y) > 0$, então diremos que este ponto é estritamente viável para o problema de complementaridade mista. E denotaremos o conjunto dos pontos estritamente viáveis por Ω^0.*

Definição 4.4 (Direção Viável). *Dado $(x,y) \in \Omega$, se diz que "d" $\in \mathbb{R}^n \times \mathbb{R}^m$ é uma direção viável em (x,y), com respeito a Ω, quando:*

$$\exists\ \theta > 0 \quad tal \quad que \quad ((x,y) + td) \in \Omega, \quad \forall t \in [0,\theta].$$

Definição 4.5 (Campo de direções). *Um campo vetorial $d(\cdot)$ é um campo de direções uniformemente viáveis se existe $\theta > 0$ tal que:*

$$((x,y) + td(x,y)) \in \Omega, \quad \forall t \in [0,\theta].$$

Definição 4.6 (Direção de Descida). *Uma direção "d" $\in \mathbb{R}^n \times \mathbb{R}^m$ é uma direção de descida da função f em (x,y) se existe $\theta > 0$ tal que:*

$$f((x,y) + td) < f(x,y), \quad \forall t \in [0,\theta].$$

Definição 4.7 (Solução Degenerada). *Uma solução de (4.1) é dita solução degenerada se para algum índice i, $(x_i, y_i) = 0 \quad e \quad F_i(x,y) = 0$.*

Definição 4.8 (Solução não Degenerada). *Uma solução de (4.1) é dita solução não degenerada se para todo índice i, $(x_i, y_i) + F_i(x,y) \neq 0$.*

4.2 ALGORITMO DE PONTO INTERIOR VIÁVEL PARA MNCP

Nesta seção apresentaremos um algoritmo de ponto interior para o problema de complementaridade mista (4.1). Este algoritmo gera uma sequência de pontos que satisfaz as condições do problema de complementaridade (4.1), além da condição de igualdade.

A idéia do algoritmo é resolver, dentro da região $\Omega \subset \mathbb{R}^n \times \mathbb{R}^m$, o seguinte sistema de equações:

$$S(x,y) = \begin{cases} x \bullet F(x,y) = 0, \\ Q(x,y) \quad = 0. \end{cases} \tag{4.3}$$

Para esta finalidade, utilizaremos o método de Newton para construir uma sequência de **"pontos viáveis"**, a qual convergirá para a solução do problema de complementaridade mista (4.1). Tomando

$$\Omega = \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m / x \geq 0 \ \text{ e } \ F(x,y) \geq 0\}$$

como o conjunto de pontos viáveis e uma função

$$f(x,y) = \phi(x,y) + \parallel Q(x,y) \parallel^2,$$

onde $\phi(x,y) = x^T F(x,y)$.

Vamos procurar uma solução para o sistema (4.3) na região

$$\Omega_c = \{(x,y) \in \Omega / f(x,y) \leq c\},$$

e construir uma sequência de pontos convergente para uma solução do sistema (4.3).

Então, (x,y) é solução do problema de complementaridade mista não linear se, e só se, é solução do sistema (4.3) em Ω.

Assim, temos que desenvolver um algoritmo que gera uma sequência de pontos em Ω e convirja para a solução do sistema (4.3) e, portanto, tende a convergir para uma solução do problema de complementaridade mista. Para gerar esta sequência utilizaremos uma iteração de Newton no sistema (4.3), mas veremos que isso não será suficiente para garantir que a sequência gerada está contida em Ω, [20].

Antes de começarmos a construção do algoritmo veremos algumas notações que ajudarão no decorrer do texto, como em [20].

O gradiente de $S(x,y) = \begin{cases} x \bullet F(x,y) = 0, \\ Q(x,y) \quad = 0, \end{cases}$ é dado por:

$$\nabla S(x,y) = \begin{pmatrix} \nabla_x (x \bullet F(x,y)) & \nabla_y (x \bullet F(x,y)) \\ \nabla_x Q(x,y) & \nabla_y Q(x,y) \end{pmatrix}, \tag{4.4}$$

onde

$$\nabla_x (x \bullet F(x,y)) = Diag(F(x,y)) + Diag(x) \nabla_x F(x,y)$$

e

$$\nabla_y (x \bullet F(x,y)) = Diag(x).\nabla_y F(x,y)$$

com $Diag(F(x,y))$, $Diag(x)$ matrizes diagonais e a matriz jacobiana de $F(x,y)$ é:

$$\nabla F(x,y) = \begin{pmatrix} \nabla_x F \\ \nabla_y F \end{pmatrix}^T,$$

onde:

$$\nabla_x F(x,y) = \begin{pmatrix} \frac{\partial F_1(x,y)}{\partial x_1} & \frac{\partial F_1(x,y)}{\partial x_2} & \cdots & \frac{\partial F_1(x,y)}{\partial x_n} \\ \frac{\partial F_2(x,y)}{\partial x_1} & \frac{\partial F_2(x,y)}{\partial x_2} & \cdots & \frac{\partial F_2(x,y)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_n(x,y)}{\partial x_1} & \frac{\partial F_n(x,y)}{\partial x_2} & \cdots & \frac{\partial F_n(x,y)}{\partial x_n} \end{pmatrix}$$

e

$$\nabla_y F(x,y) = \begin{pmatrix} \frac{\partial F_1(x,y)}{\partial y_1} & \frac{\partial F_1(x,y)}{\partial y_2} & \cdots & \frac{\partial F_1(x,y)}{\partial y_m} \\ \frac{\partial F_2(x,y)}{\partial y_1} & \frac{\partial F_2(x,y)}{\partial y_2} & \cdots & \frac{\partial F_2(x,y)}{\partial y_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_n(x,y)}{\partial y_1} & \frac{\partial F_n(x,y)}{\partial y_2} & \cdots & \frac{\partial F_n(x,y)}{\partial y_m} \end{pmatrix}.$$

Agora, dado um ponto $(x^k, y^k) \in \Omega$ que não seja solução de (4.1), pela iteração de Newton em $S(x,y) = 0$, temos:

$$\nabla S(x^k, y^k) d_1^k = -S(x^k, y^k). \tag{4.5}$$

Para que o sistema fique bem definido, supondo $\nabla S(x^k, y^k)$ não singular em (x^k, y^k). Se ocorre $S_i(x^k, y^k) = 0$ sem ser solução do problema de complementaridade mista, para algum índice $i \in \{1, 2, ..., (n+m)\}$, segue a seguinte situação.

Seja a i-ésima linha do sistema (4.5), onde ocorre $S_i(x^k, y^k) = 0$, então:

$$\left[F_i(x^k, y^k) e_i + x_i^k \nabla F_i(x^k, y^k)\right] d_1^k = 0,$$

$$\nabla Q_i(x^k, y^k) d_1^k = 0,$$

onde $e_i \in \mathbb{R}^n \times \mathbb{R}^m$.

Pela não singularidade da matriz $\nabla S(x^k, y^k)$, temos as seguintes possibilidades:

a) $x_i^k = 0$ e $F_i(x^k, y^k) > 0$, isso implica que d_1^k é tangente a restrição $x_i \geq 0$.

b) $x_i^k > 0$ e $F_i(x^k, y^k) = 0$, isso implica que d_1^k é tangente a restrição $F_i(x^k, y^k) \geq 0$.

Nestes dois casos não pode-se garantir a viabilidade da direção d_1^k em Ω.

A seguinte proposição nos fornece as condições que uma direção $d \in \mathbb{R}^n \times \mathbb{R}^m$ deve satisfazer para ser uma direção viável em Ω, [20].

Proposição 4.1 *Seja $d \in \mathbb{R}^n \times \mathbb{R}^m$ e $(x,y) \in \Omega$. Se a direção d satisfaz as condições:*

i) $d_i > 0$ *para todo índice* i *tal que* $(x_i, y_i) = 0$,

ii) $\nabla F_i(x^k, y^k)d > 0$ *para todo índice* i *tal que* $F_i(x, y) = 0$.

Então d *é uma* ***direção viável*** *no ponto* (x, y).

Deste modo, usaremos uma direção de restauração para que, em combinação com a direção d_1 de Newton gere uma nova direção d que seja viável em Ω, [20]. Para isso tomemos o seguinte sistema:

$$\nabla S(x^k, y^k)d_2^k = \rho^k \overline{E}, \tag{4.6}$$

onde o vetor $\overline{E} \in \mathbb{R}^n \times \mathbb{R}^m$ tal que $\overline{E} = \begin{bmatrix} E_1 \\ E_0 \end{bmatrix}$ com:

$$E_1 = [1, 1, ..., 1] \in \mathbb{R}^n \quad \text{e} \quad E_0 = [0, 0, ..., 0] \in \mathbb{R}^m \quad \text{e} \quad \rho^k > 0.$$

Então para todo $(x^k, y^k) \in \Omega$ que não seja solução do problema de complementaridade mista mas que para algum índice i temos $S_i(x^k, y^k) = 0$ segue que:

- $x_i^k = 0$ e $F_i(x^k, y^k) > 0$, então $d_{2_i}^k > 0$.
- $x_i^k > 0$ e $F_i(x^k, y^k) = 0$, então $\nabla F_i(x^k, y^k)d_2^k > 0$.

Assim, a direção d_2^k é uma direção viável em Ω pela proposição 4.1.

Portanto, a combinação linear, $d = d_1^k + \rho^k d_2^k$ com $\rho^k > 0$, faz com que a direção d^k seja uma direção viável em Ω.

Vamos definir:

$$\rho^k = \frac{\rho_0 \phi^\beta(x^k, y^k)}{n} \in (0, 1),$$

onde $\phi(x^k, y^k) = (x^k)^T F(x^k, y^k)$, $\rho_0 = \alpha \min\{1; \frac{1}{c^{\beta-1}}\}$ com $\alpha \in (0, 1)$, $\beta \in [1, 2]$ para todo $(x^k, y^k) \in \Omega_c$ que não seja solução do problema de complementariedade mista [20].

Desta forma, calculamos a direção de busca d^k resolvendo:

$$\nabla S(x^k, y^k)d^k = -S(x^k, y^k) + \rho^k \overline{E}. \tag{4.7}$$

Em cada iteração, veremos que a direção d^k tem a propriedade de ser viável em Ω e de ser uma direção de descida para a função:

$$f(x, y) = \phi(x, y) + \| Q(x, y) \|^2 . \tag{4.8}$$

Proposição 4.2 (Viabilidade da direção). *Sejam* $(x^k, y^k) \in \Omega$ *tal que* $f(x^k, y^k) > 0$ *e* $\nabla S(x^k, y^k)$ *não singular, então a direção* d^k *obtida pela resolução do sistema (4.7) é viável em* Ω.

Proposição 4.3 (Direção de descida). *Em todo ponto* $(x^k, y^k) \in \Omega$ *tal que* $f(x^k, y^k) > 0$, *a direção* d^k *obtida pela resolução do sistema (4.7) é uma direção de descida para* $f(x^k, y^k)$ *se* $\rho_0 \phi^{\beta-1}(x^k, y^k) < 1$.

Da Proposição 4.2 e da Proposição 4.3 temos que a direção d^k solução do sistema (4.7) é viável e de descida para o problema de complementaridade (4.1), desde que tenhamos $\nabla S(x^k, y^k)$ não singular e $\rho_0 \phi^{\beta-1}(x^k, y^k) < 1$ para todo $(x^k, y^k) \in \Omega$.

4.3 DESCRIÇÃO DO ALGORITMO FDA-MNCP

Neste algoritmo vamos obter uma sequência de pontos interiores na região Ω que converge para a solução do problema de complementaridade mista, mediante a resolução de um sistema de equações e uma busca linear nas restrições $x \geq 0, F(x, y) \geq 0$ e na função $f(x, y)$.

Algoritmo 4.1 (Algoritmo FDA-MNCP)

Passo 1. *Dados de entrada:* $c, \epsilon > 0$, $\alpha, \eta, \nu \in (0, 1)$, $\beta \in [1, 2]$.

Dados inciais: $(x^0, y^0) \in \Omega^0$ *tal que* $f(x^0, y^0) \leq c$ *e* $k = 0$.

Passo 2. *Com* $\rho^k = \rho_0 \, \phi^{(\beta)}(x^k, y^k)/n$, *resolver o sistema:*

$$\nabla S(x^k, y^k) d^k = -S(x^k, y^k) + \rho^k \overline{E}.$$

Passo 3. *Busca linear. Encontrar* t^k *como o primeiro valor da sequência* $\{1, \nu, \nu^2, \nu^3, \cdots\}$, *tal que:*

$$\begin{aligned} x^k + t^k d_x^k &\geq 0, \\ F((x^k, y^k) + t^k d^k) &\geq 0, \\ f(x^k, y^k) + t^k \eta \nabla f(x^k, y^k)^t d^k &\geq f((x^k, y^k) + t^k d^k). \end{aligned}$$

Passo 4. *Atualização:*

$$(x^{k+1}, y^{k+1}) := (x^k, y^k) + t^k d^k \quad e \quad k := k + 1.$$

Passo 5. *Se* $f(x^{k+1}, y^{k+1}) < \epsilon$ *então* ***PARAR.***

Se não voltar ao ***Passo 2.***

Podemos destacar que no Passo 2 determinamos a direção de busca d^k, por meio da resolução de um sistema linear. Este sistema fornece uma direção d^k que é uma combinação da direção de Newton d_1^k e da direção de restauração d_2^k. Logo d^k tem a propriedade de ser uma direção de descida da função $f(x^k, y^k)$ e viável na região Ω, pela Proposição 4.2 e Proposição 4.3, que foram mostrados em [20].

4.4 CONVERGÊNCIA GLOBAL DO FDA-MNCP

Agora apresentamos resultados teóricos sobre a convergência global do algoritmo FDA-MNCP e uma solução do problema de complementaridade mista (4.1). Mostraremos que o algoritmo FDA-MNCP gera uma sequência de pontos $(x^k, y^k) \in \Omega$ a partir de um ponto estritamente viável, e a cada iteração o valor da função $f(x^k, y^k)$ vai diminuindo. Para tal, o algoritmo produzirá uma direção de busca d^k, que tem as propriedades de ser uma direção viável em Ω e de descida para a função $f(x, y)$, e esta direção d^k é um campo uniforme de direções viáveis e limitado. O passo da busca para a função $f(x, y)$ é limitado inferiormente por um valor positivo.

Vamos supor condições para a função $F(x, y)$, $Q(x, y)$, a matriz $\nabla S(x, y)$ e o conjunto de pontos viáveis em Ω, [20].

Suposição 4.1 *O conjunto* $\Omega_c = \{(x, y) \in \Omega / f(x, y) \leq c\}$ *é compacto e possui interior não vazio, tal que para cada* $(x, y) \in \Omega_c^0$ *satisfaz* $x > 0$ *e* $F(x, y) > 0$.

Isso significa dizer que é garantida a existência de pontos estritamente viáveis em Ω_c.

Suposição 4.2 *As funções* $F(x, y)$ *e* $Q(x, y)$ *são de classe* $C^1(\mathbb{R}^n \times \mathbb{R}^m)$ *e ainda os* $\nabla F(x, y)$ *e* $\nabla Q(x, y)$ *satisfazem a condição de Lipschitz,*

$$\| \nabla F(x_2, y_2) - \nabla F(x_1, y_1) \| \leq l_0 \| (x_2, y_2) - (x_1, y_1) \|$$

e

$$\| \nabla Q(x_2, y_2) - \nabla Q(x_1, y_1) \| \leq L \| (x_2, y_2) - (x_1, y_1) \|,$$

para qualquer (x_1, y_1) *e* $(x_2, y_2) \in \Omega_c$, *onde* l_0 *e* L *são constantes positivas.*

Em outras palavras, $\nabla S(x, y)$ satisfaz a condição de Lipschitz e ainda $Q_i^2(x, y)$ também é Lipschitz, então sem perda de generalidade podemos supor que existe uma constante $\gamma > 0$ que verifica a condição de Lipschitz para $\nabla S(x, y)$ e $Q_i^2(x, y)$, para $i = 1, 2, ..., m$.

Suposição 4.3 *A matriz* $\nabla S(x, y)$ *é não singular para todo* $(x, y) \in \Omega_c$, *isto é, existe* $(\nabla S(x, y))^{-1}$ *para todo* $(x, y) \in \Omega_c$.

Ou seja, o sistema linear do algoritmo (4.7) possui solução única, logo a direção d^k está bem definida.

Alem disso, como $\nabla F(x,y)$ e $\nabla Q(x,y)$ são contínuas, então $\nabla S(x,y)$ e $(\nabla S(x,y))^{-1}$ são contínuas e limitadas em Ω_c, ou seja, existem constantes positivas k_0 e k tal que $\parallel \nabla S(x,y) \parallel \leq k_0$ e $\parallel (\nabla S(x,y))^{-1} \parallel \leq k$, para todo $(x,y) \in \Omega_c$.

Suposição 4.4 *Existe constante real $\sigma > 0$ tal que o seguinte subconjunto $\Omega^\star$ é não vazio:*

$$\Omega^\star = \{(x,y) \in \Omega_c \quad tal\ que \quad \sigma \parallel Q(x,y) \parallel \leq \phi(x,y)\}.$$

A qual nos garanta que quando estamos num ponto que satisfaz a complementaridade $x \bullet F(x,y) = 0$, implica que também irá satisfazer $Q(x,y) = 0$. Assim podemos controlar a convergência do algoritmo só com a condição de complementaridade.

Agora vejamos uma sequência de resultados que provará a convergência global do algoritmo FDA-MNCP, [20].

Proposição 4.4 *Dado um ponto $(x^k, y^k) \in \Omega^\star$ a direção d^k dada pelo sistema (4.7) é viável em Ω desde que $f(x^k, y^k) \neq 0$.*

Prova:

Como [20], dado $(x^k, y^k) \in \Omega^\star$, a direção d^k cumpre:

$$\nabla S(x^k, y^k) d^k = -S(x^k, y^k) + \rho^k \overline{E},$$

então as $i-$ésimas linhas com $i = 1, 2, ..., n$ do sistema anterior, ficam:

$$\left[e_i F_i(x^k, y^k) + x_i^k \nabla F_i(x^k, y^k)\right] d^k \ = \ -(x^k, y^k) F_i(x^k, y^k) + \rho^k, \tag{4.9}$$

onde: $e_i \in \mathbb{R}^n \times \mathbb{R}^m$ e $\nabla F_i = [\nabla_x F_i \quad \nabla_y F_i]_{1\times(n+m)}$ com $(x^k, y^k) \in \Omega^\star \subset \Omega_c \subset \Omega$.

Sejam os conjuntos [20]:

$$K = \{i / x_i = 0 \text{ e } F_i(x^k, y^k) > 0\},$$

$$J = \{i / x_i > 0 \text{ e } F_i(x^k, y^k) = 0\},$$

então, para $i \in K$, temos: $x_i^k = 0$ e $F_i(x^k, y^k) > 0$. Logo, em (4.9), obtemos:

$$[e_i F_i(x^k, y^k)] d^k = \rho^k$$

e como $\rho^k > 0$, implica que $d_i^k > 0, \quad \forall \rho^k > 0$.

Por outro lado, se $i \in J$, temos $x_i^k > 0$ e $F_i(x^k, y^k) = 0$. Logo, em (4.9), obtemos:

$$x_i^k \nabla F_i(x^k, y^k) d^k = \rho^k,$$

então $\nabla F_i(x^k, y^k) d^k > 0, \quad \forall \rho^k > 0$.

Portanto, pela proposição (4.1) temos que d^k é uma DIREÇÃO VIÁVEL no ponto (x^k, y^k). ■

Proposição 4.5 *Dado* $(x^k, y^k) \in \Omega^\star \subset \Omega_c$, $f(x^k, y^k) \neq 0$, *a direção* d^k *dada pelo sistema (4.7) é uma direção de descida para a função:*

$$f(x,y) \;=\; \phi(x^k,y^k) \,+\, \parallel Q(x^k,y^k) \parallel^2 \,=\, \sum_{i=1}^{n} x_i^k.F(x^k,y^k) \,+\, \sum_{j=1}^{m} Q_i^2(x^k,y^k),$$

com $\rho_0\phi^{\beta-1}(x^k,y^k) < 1$, *onde* $\beta \in [1,2]$, $\rho_0 = \alpha\, min\{1, \frac{1}{c^{\beta-1}}\}$ *e* $\alpha \in (0,1)$.

Prova:

Todo ponto de Ω que cumpre $f(x^k, y^k) = 0$ é uma solução de 4.1, isto é, $x_k \geq 0$, $F(x^k, y^k) \geq 0$, $x^k \bullet F(x^k, y^k)$ e $Q(x^k, y^k) = 0$, e seguindo como [20].

O gradiente de $f(x^k, y^k)$ é:

$$\begin{aligned}
\nabla f(x^k,y^k) \;&= \nabla\left[\phi(x^k,y^k) + \parallel Q(x^k,y^k) \parallel^2\right].\\
&= \nabla\left[x^k \bullet F(x^k,y^k) + \parallel Q(x^k,y^k) \parallel^2\right].\\
&= \left[\nabla_x\left(x^k \bullet F(x^k,y^k) + \parallel Q(x^k,y^k) \parallel^2\right)\, \nabla_y\left(x^k \bullet F(x^k,y^k) + \parallel Q(x^k,y^k) \parallel^2\right)\right].\\
&= \left[E_1^T \;\; 2Q^T(x^k,y^k)\right] \begin{bmatrix} \nabla_x(x \bullet F(x,y)) & \nabla_y(x \bullet F(x,y)) \\ \nabla_x Q(x,y) & \nabla_y Q(x,y) \end{bmatrix}.
\end{aligned}$$

então:

$$\nabla f(x^k,y^k) = \left[E_1^T \;\; 2Q^T(x^k,y^k)\right] \nabla S(x^k,y^k).$$

Dada a direção d^k, temos:

$$\begin{aligned}
\nabla f(x^k,y^k)d^k \;&=\; \left[E_1^T \;\; 2Q^T(x^k,y^k)\right] \nabla S(x^k,y^k)d^k.\\
&=\; \left[E_1^T \;\; 2Q^T(x^k,y^k)\right] \left(-S(x^k,y^k) + \rho^k \overline{E}\right).\\
&=\; \left[E_1^T \;\; 2Q^T(x^k,y^k)\right] \begin{bmatrix} -x^k \bullet F(x^k,y^k) \\ -Q(x^k,y^k) \end{bmatrix} + \rho^k \left[E_1^T \;\; 2Q^T(x^k,y^k)\right] \begin{bmatrix} E_1 \\ E_0 \end{bmatrix}.\\
&=\; -x^k F(x^k,y^k) - 2 \parallel Q(x^k,y^k) \parallel^2 + n\rho^k.\\
&\leq\; -\phi(x^k,y^k) - \parallel Q(x^k,y^k) \parallel^2 + \rho_0\phi^\beta(x^k,y^k).\\
&<\; -f(x^k,y^k) + \rho_0\phi^\beta(x^k,y^k) + \rho_0\phi^{\beta-1}(x^k,y^k) \parallel Q(x^k,y^k) \parallel^2.\\
&=\; -f(x^k,y^k) + \rho_0\phi^{\beta-1}(x^k,y^k)\left(\phi(x^k,y^k) + \parallel Q(x^k,y^k) \parallel^2\right).\\
&=\; -f(x^k,y^k)\left(1 - \rho_0\phi^{\beta-1}(x^k,y^k)\right) < 0.
\end{aligned}$$

de onde podemos obter que:

$$1 - \rho_0\phi^{\beta-1}(x^k,y^k) \geq 0,$$

então

$$\rho_0\phi^{\beta-1}(x^k,y^k) < 1 \implies \nabla f(x^k,y^k)d^k < 0.$$

Portanto, a direção d^k é uma DIREÇÃO DE DESCIDA para $f(x^k, y^k)$. ■

Destes dois últimos resultados obtemos que a direção d^k é de DESCIDA E VIÁVEL para o problema de complementariedade mista desde que $\nabla S(x^k, y^k)$ seja não singular e $\rho_0 \phi^{\beta-1}(x^k, y^k) < 1$ para todo $(x^k, y^k) \in \Omega$.

Proposição 4.6 *Dado um ponto $(x^k, y^k) \in \Omega^\star$, a direção d^k calculada pelo FDA-MNCP satisfaz a seguinte desigualdade:*

$$\| d^k \| \leq \overline{\kappa} \phi(x^k, y^k), \tag{4.10}$$

consequentemente $\| d^k \| \leq \overline{\kappa} c$ para todo $(x^k, y^k) \in \Omega^*$.

Prova:

Como [20], considerando a equação (4.7) e sabendo que a matriz $(\nabla S(x^k, y^k))^{-1}$ é não singular, então podemos calcular $\| d^k \|$ através de

$$\begin{aligned} \nabla S(x^k, y^k) d^k &= -S(x^k, y^k) + \rho^k \overline{E}. \\ (\nabla S(x^k, y^k))^{-1} (\nabla S(x^k, y^k))^{-1} d^k &= (\nabla S(x^k, y^k))^{-1} (-S(x^k, y^k)) + (\nabla S(x^k, y^k))^{-1} \rho^k \overline{E}. \\ d^k &= \left(\nabla S(x^k, y^k)\right)^{-1} \left[-S(x^k, y^k) + \rho^k \overline{E}\right]. \end{aligned}$$

Calculando sua norma, obtemos a seguinte desigualdade:

$$\begin{aligned} \| d^k \| &= \| \left(\nabla S(x^k, y^k)\right)^{-1} \left[-S(x^k, y^k) + \rho^k \overline{E}\right] \|. \\ &\leq \| \left(\nabla S(x^k, y^k)\right)^{-1} \| \, \| -S(x^k, y^k) + \rho^k \overline{E} \|. \end{aligned} \tag{4.11}$$

Da segunda norma da desigualdade (4.11),

$$\begin{aligned} \| -S(x^k, y^k) + \rho^k \overline{E} \| &= \left\langle -S(x^k, y^k) + \rho^k \overline{E}, \; -S(x^k, y^k) + \rho^k \overline{E} \right\rangle. \\ &= \| S(x^k, y^k) \|^2 - 2\rho^k \left\langle S(x^k, y^k), \overline{E} \right\rangle + \left(\rho^k\right)^2 \| \overline{E} \|. \\ &= \| S(x^k, y^k) \|^2 - 2\rho^k \phi(x^k, y^k) + n \left(\rho^k\right)^2. \end{aligned} \tag{4.12}$$

onde:

$$\| S(x^k, y^k) \|^2 = \left\| \begin{pmatrix} x^k \bullet F(x^k, y^k) \\ Q(x^k, y^k) \end{pmatrix} \right\|^2 = \| x^k F(x^k, y^k) \|^2 + \| Q(x^k, y^k) \|^2. \tag{4.13}$$

Substituindo (4.13) em (4.12), obtemos:

$$\| -S(x^k, y^k) + \rho^k \overline{E} \| = \| x^k F(x^k, y^k) \|^2 - 2\rho^k \phi(x^k, y^k) + n \left(\rho^k\right)^2 + \| Q(x^k, y^k) \|^2, \tag{4.14}$$

vamos ver as três primeiras parcelas da parte direita da igualdade (4.14), somando e subtraindo $\left(\rho^k\right)^2$:

$$\begin{aligned}
\| x^k F(x^k,y^k) \|^2 - 2\rho^k\phi(x^k,y^k) + n\left(\rho^k\right)^2 &= \| x^k F(x^k,y^k) \|^2 - 2\rho^k\phi(x^k,y^k) + \\
&\quad + n\left(\rho^k\right)^2 - \left(\rho^k\right)^2 + \left(\rho^k\right)^2. \\
&= \left[\left(x^k \bullet F(x^k,y^k)\right) - \rho^k\right]^2 + \left(\rho^k\right)^2 (n-1). \\
&= \left[\phi(x^k,y^k) - \frac{\rho_0\phi^\beta(x^k,y^k)}{n}\right]^2 + \\
&\quad + \left(\rho^k\right)^2 (n-1). \\
&= \phi^2(x^k,y^k)\left(\frac{n-\rho_0\phi^{\beta-1}(x^k,y^k)}{n}\right)^2 + \\
&\quad + \left(\frac{\rho_0\phi^\beta(x^k,y^k)}{n}\right)^2 (n-1). \\
&= \phi^2\left[\frac{\left(n-\rho_0\phi^{\beta-1}\right)^2 + (n-1)\left(\rho^0\phi^{\beta-1}\right)}{n^2}\right].
\end{aligned} \tag{4.15}$$

Da proposição anterior, obtemos $\rho_0\phi^{\beta-1}(x^k,y^k) < 1$, multiplicando por -1 e adicionando n obtemos:

- $n-1 < n-\rho_0\phi^{\beta-1}(x^k,y^k) < n.$
- $\left[\rho_0\phi^{\beta-1}(x^k,y^k)\right]^2 < \rho_0\phi^{\beta-1}(x^k,y^k) < 1.$

então, substituindo estas duas últimas desigualdades em (4.15) obtemos:

$$\begin{aligned}
\| x^k \bullet F(x^k,y^k) \|^2 - 2\rho^k\phi(x^k,y^k) + n\left(\rho^k\right)^2 &< \phi^2(x^k,y^k)\left(\frac{n^2-n}{n^2}\right). \\
&< \phi^2(x^k,y^k)\left(\frac{n-1}{n}\right). \\
&< \phi^2(x^k,y^k).
\end{aligned}$$

Logo, substituindo (4.16) na igualdade (4.14), e pela suposição 4.4 obtemos:

$$\begin{aligned}
\| -S(x^k,y^k) + \rho^k\overline{E} \|^2 &< \phi^2(x^k,y^k) + \frac{1}{\sigma^2}\phi^2(x^k,y^k). \\
&< \left(\frac{\sigma^2+1}{\sigma^2}\right)\phi^2(x^k,y^k).
\end{aligned} \tag{4.16}$$

Finalmente substituindo (4.16) em (4.11), obtemos:

$$\begin{aligned}
\| d^k \| &< \| \left(\nabla S(x^k,y^k)\right)^{-1} \| \, \| -S(x^k,y^k) + \rho^k\overline{E} \|. \\
&< \| \left(\nabla S(x^k,y^k)\right)^{-1} \| \sqrt{\frac{\sigma^2+1}{\sigma^2}}\phi(x^k,y^k). \\
&< \overline{\kappa}\phi(x^k,y^k).
\end{aligned}$$

onde:

$$\overline{\kappa} = \kappa\sqrt{\frac{\sigma^2+1}{\sigma^2}} \quad \text{com} \quad \kappa \geq \| \left(\nabla S(x^k, y^k)\right)^{-1} \| .$$

■

Proposição 4.7 *A sequência de direções $\{d^k\}$ gerada pelo algoritmo FDA-MNCP, consiste num campo uniforme de direções viáveis do problema de complementaridade em Ω.*

Prova:

Como [20] e da suposição 4.2:

$$\| \nabla S_i(x_2^k, y_2^k) - \nabla S_i(x_1^k, y_1^k) \| \leq \gamma \| (x_2^k, y_2^k) - (x_1^k, y_1^k) \| .$$

para todo $(x_1^k, y_1^k), (x_2^k, y_2^k) \in \Omega^\star \subset \Omega_c$.

Seja $(x^k, y^k) \in \Omega^\star$ e $\theta > 0$ tal que o seguimento $\left[(x^k, y^k),\ (x^k, y^k) + \tau d^k\right] \subset \Omega$ para todo $\tau \in [0, \theta]$, pelo Teorema do Valor Médio, [18]

$$\begin{aligned} S_i\left((x^k, y^k) + \tau d^k\right) &\geq S_i\left((x^k, y^k)\right) + \tau \nabla S_i^T(x^k, y^k) d^k - \gamma \| \tau d^k \|^2 . \\ &\geq S_i\left((x^k, y^k)\right) + \tau \nabla S_i^T(x^k, y^k) d^k - \tau^2 \gamma \| d^k \|^2 . \end{aligned} \tag{4.17}$$

para $\tau \in [0, \theta]$ e $1 \leq i \leq n+m$, para cada $i = 1, 2, ..., n$ obtemos:

$$\nabla S_i^T d^k = -S_i(x^k, y^k) + \rho^k \overline{E}. \tag{4.18}$$

substituindo (4.18) em (4.17), obtemos:

$$\begin{aligned} S_i\left((x^k, y^k) + \tau d^k\right) &\geq S_i\left((x^k, y^k)\right) + \tau\left(-S_i(x^k, y^k) + \rho^k\right) - \tau^2 \gamma \| d^k \|^2 . \\ &= S_i\left((x^k, y^k)\right) - \tau S_i(x^k, y^k) + \tau \rho^k - \tau^2 \gamma \| d^k \|^2 . \\ &= (1-\tau) S_i(x^k, y^k) + \tau\left(\rho^k - \tau\gamma \| d^k \|^2\right). \end{aligned}$$

para qualquer $\tau \leq \min\left\{1, \frac{\rho^k}{\gamma \|d^k\|}\right\}$, obtemos:

$$S_i\left((x^k, y^k) + \tau d^k\right) \geq 0.$$

então: $(x^k, y^k) + \tau d^k \geq 0$ e $F((x^k, y^k) + \tau d^k) \geq 0$, portanto $(x^k, y^k) + \tau d^k \in \Omega$.

Agora pelo lema anterior e pela definição de ρ^k obtemos:

$$\frac{\rho^k}{\gamma \| d^k \|^2} \geq \frac{\rho_0 \phi^\beta(x^k, y^k)}{n} \frac{1}{\gamma \overline{\kappa}^2 \phi^2(x^k, y^k)} = \frac{\rho_0 \phi^{\beta-2}(x^k, y^k)}{n \gamma \overline{\kappa}^2}, \tag{4.19}$$

então

$$\tau \leq \min\left\{1; \frac{\rho_0 \phi^{\beta-2}(x^k, y^k)}{n\gamma\overline{\kappa}^2}\right\}, \tag{4.20}$$

como $\beta \in [1, 2]$, então o lema é valido para

$$\theta = \min\left\{1; \frac{\rho_0 \phi^{\beta-2}(x^k, y^k)}{n\gamma\overline{\kappa}^2}\right\}.$$

■

Proposição 4.8 *Existe $\xi > 0$ tal que, para $(x^k, y^k) \in \Omega^\star$, a condição de Armijo no algoritmo FDA-MNCP é satisfeita para qualquer $t^k \in [0, \xi]$.*

Prova:

Como [20], seja $t^k \in (0, \theta]$, onde $\theta = \min\left\{1; \frac{\rho_0 \phi^{\beta-2}(x^k, y^k)}{n\gamma\bar{\kappa}^2}\right\}$, aplicando o Teorema do Valor Médio [18], para as "n" primeiras linhas e tomando $(x^{k+1}, y^{k+1}) = (x^k, y^k) + t^k d^k$, temos:

$$\begin{aligned} S_i\left((x^{k+1}, y^{k+1})\right) &= S_i\left((x^k, y^k) + t^k d^k\right). \\ &\leq S_i(x^k, y^k) + t^k \nabla\left[S(x^k, y^k)\right]_i d^k + \frac{(t^k)^2}{2}\left(\nabla S_i(x^k, y^k)\right)' \parallel d^k \parallel . \end{aligned}$$

pela suposição 4.2 obtemos que: $\frac{\left(\nabla S_i(x^k, y^k)\right)'}{2} \leq \gamma$, então

$$\leq S_i(x^k, y^k) + t^k \nabla S_i(x^k, y^k) d^k + (t^k)^2 \gamma \parallel d^k \parallel^2$$

somando primeiro de 1 ate n, obtemos:

$$\begin{aligned} \sum_{i=1}^{n} x_i^{k+1} F_i(x^{k+1}, y^{k+1}) &\leq \sum_{i=1}^{n} x_i^k F_i(x^k, y^k) + t^k \left(\sum_{i=1}^{n} \nabla S_i(x^k, y^k)\right) d^k + \left(t^k\right)^2 \gamma \parallel d^k \parallel^2 . \\ \phi(x^{k+1}, y^{k+1}) &\leq \phi(x^k, y^k) + t^k\left(-\phi(x^k, y^k) + n\rho^k\right) + n\left(t^k\right)^2 \gamma \parallel d^k \parallel^2 . \\ &= \left(1 - t^k\right)\phi(x^k, y^k) + t^k n\rho^k + n\left(t^k\right)^2 \gamma \parallel d^k \parallel^2 . \end{aligned} \tag{4.21}$$

Do Teorema do Valor Médio [18], para $Q_i^2(x^k, y^k); i = (n+1), \cdots, (n+m)$ e como $\nabla\left[S(x^k, y^k)\right]_i d^k = -Q_i(x^k, y^k)$, obtemos a seguinte inequação

$$\begin{aligned} Q_i^2(x^{k+1}, y^{k+1}) &\leq Q_i^2(x^k, y^k) + t^k\left(\nabla\left(Q_i^2(x^k, y^k)\right)\right) + \left(t^k\right)^2 \gamma \parallel d^k \parallel^2 . \\ Q_i^2(x^{k+1}, y^{k+1}) &\leq Q_i^2(x^k, y^k) + t^k\left(2Q_i(x^k, y^k)\underbrace{\nabla Q_i(x^k, y^k)}\right) + \left(t^k\right)^2 \gamma \parallel d^k \parallel^2 . \\ Q_i^2(x^{k+1}, y^{k+1}) &\leq Q_i^2(x^k, y^k) + 2t^k Q_i(x^k, y^k)\left(-Q_i(x^k, y^k)\right) + \left(t^k\right)^2 \gamma \parallel d^k \parallel^2 . \\ Q_i^2(x^{k+1}, y^{k+1}) &\leq Q_i^2(x^k, y^k) - 2t^k Q_i^2(x^k, y^k) + \left(t^k\right)^2 \gamma \parallel d^k \parallel^2 . \\ &< Q_i^2(x^k, y^k) - t^k Q_i^2(x^k, y^k) + \left(t^k\right)^2 \gamma \parallel d^k \parallel^2 . \end{aligned}$$

somando para $i = (n+1), \cdots, (n+m)$, obtemos:

$$\begin{aligned} \parallel Q(x^{k+1}, y^{k+1}) \parallel^2 &\leq \parallel Q(x^k, y^k) \parallel^2 - t^k \parallel Q(x^k, y^k) \parallel^2 + m\left(t^k\right)^2 \gamma \parallel d^k \parallel^2 . \\ &= \left(1 - t^k\right) \parallel Q(x^k, y^k) \parallel^2 + m\left(t^k\right)^2 \gamma \parallel d^k \parallel^2 . \end{aligned} \tag{4.22}$$

finalmente somando (4.21) e (4.22), obtemos:

$$\begin{aligned} \phi(x^{k+1}, y^{k+1}) + \parallel Q(x^{k+1}, y^{k+1}) \parallel^2 \leq\ & \left(1 - t^k\right)\phi(x^k, y^k) + \left(1 - t^k\right) \parallel Q(x^k, y^k) \parallel^2 + \\ & + t^k n\rho^k + n\left(t^k\right)^2 \gamma \parallel d^k \parallel^2 + m\left(t^k\right)^2 \gamma \parallel d^k \parallel^2 . \end{aligned}$$

então:

$$\begin{aligned} f(x^{k+1},y^{k+1}) &\leq \left(1-t^k\right) f(x^k,y^k) + (n+m)\left(t^k\right)^2 \gamma \parallel d^k \parallel^2 + t^k n\rho^k. \\ &= \left(1-t^k\right) f(x^k,y^k) + t^k\left(n\rho^k + (n+m)t^k\gamma \parallel d^k \parallel^2\right). \\ &= \left[1 - t^k + t^k \frac{\left(n\rho^k + (n+m)t^k\gamma \parallel d^k \parallel^2\right)}{f(x^k,y^k)}\right] f(x^k,y^k). \\ &= \left[1 - t^k + \frac{nt^k\rho^k}{f(x^k,y^k)} + \frac{(n+m)t^k t^k\gamma \parallel d^k \parallel^2}{f(x^k,y^k)}\right] f(x^k,y^k). \end{aligned} \quad (4.23)$$

Da definição de ρ^k e da suposição 4.4, segue para $(x^k, y^k \in \Omega^\star)$ as seguintes desigualdades:

√

$$\begin{aligned} \frac{n\rho^k}{f(x^k,y^k)} &= \frac{\rho_0\phi^\beta(x^k,y^k)}{f(x^k,y^k)}. \\ &\leq \frac{\rho_0\phi^\beta(x^k,y^k)}{\phi(x^k,y^k)}. \\ &= \rho_0\phi^{\beta-1}(x^k,y^k). \end{aligned}$$

$$\Longrightarrow \quad \boxed{\frac{n\rho^k}{f(x^k,y^k)} \leq \rho_0\phi^{\beta-1}(x^k,y^k).} \quad (4.24)$$

√

$$\begin{aligned} f(x^k,y^k) &= \phi(x^k,y^k) + \parallel Q(x^k,y^k) \parallel^2 . \\ &= \phi(x^k,y^k) + \underbrace{\parallel Q(x^k,y^k) \parallel} \parallel Q(x^k,y^k) \parallel . \\ &\leq \phi(x^k,y^k) + \sigma\, \underbrace{\phi(x^k,y^k)} \parallel Q(x^k,y^k) \parallel . \\ &\leq \phi(x^k,y^k) + \sigma \quad c \quad \parallel Q(x^k,y^k) \parallel . \\ &\leq \phi(x^k,y^k) + \sigma\ c\ \sigma\ \phi(x^k,y^k). \\ &\leq \phi(x^k,y^k) + c\ \sigma^2\ \phi(x^k,y^k). \end{aligned}$$

$$\Longrightarrow \quad \boxed{f(x^k,y^k) \leq \left(1+c\sigma^2\right)\phi(x^k,y^k).} \quad (4.25)$$

além:

√

$$\begin{aligned} \parallel d^k \parallel &\leq \overline{\kappa}^2\phi^2(x^k,y^k). \\ &\leq \overline{\kappa}^2 c^2. \end{aligned} \quad (4.26)$$

Substituindo (4.24), (4.25) e (4.26) em (4.23), obtemos:

$$f(x^{k+1},y^{k+1}) \leq \left[1 - t^k + t^k\rho_0\phi^{\beta-1}(x^k,y^k) + \frac{(n+m)t^kt^k\gamma\overline{\kappa}^2c^2}{f(x^k,y^k)}\right] f(x^k,y^k). \qquad (4.27)$$

Agora, da busca linear do algoritmo FDA-MNCP aplicado na função $f(x^k,y^k)$ e pela proposição 4.6, obtemos:

$$\begin{aligned} f(x^{k+1},y^{k+1}) &\leq f(x^k,y^k) + t^k\eta\underbrace{\nabla f(x^k,y^k)d^k}. \\ &= f(x^k,y^k) + t^k\eta\left[-\left(1-\rho_0\phi^{\beta-1}(x^k,y^k)\right)f(x^k,y^k)\right]. \\ &= f(x^k,y^k)\left[1 - t^k\eta\left(1-\rho_0\phi^{\beta-1}(x^k,y^k)\right)\right]. \qquad (4.28) \end{aligned}$$

Finalmente, para satisfazer a condição de Armijo, basta que:

$$\begin{aligned} 1 - t^k + t^k\rho_0\phi^{\beta-1}(x^k,y^k) + \frac{(n+m)t^kt^k\gamma\overline{\kappa}^2c^2}{f(x^k,y^k)} &\leq 1 - t^k\eta\left(1-\rho_0\phi^{\beta-1}(x^k,y^k)\right). \\ -t^k + t^k\rho_0\phi^{\beta-1}(x^k,y^k) + \frac{(n+m)t^kt^k\gamma\overline{\kappa}^2c^2}{f(x^k,y^k)} &\leq -t^k\eta\left(1-\rho_0\phi^{\beta-1}(x^k,y^k)\right). \\ \frac{(n+m)t^k\gamma\overline{\kappa}^2c^2}{f(x^k,y^k)} &\leq (1-\eta)\left(1-\rho_0\phi^{\beta-1}(x^k,y^k)\right). \\ t^k &\leq \frac{(1-\eta)\left(1-\rho_0\phi^{\beta-1}(x^k,y^k)\right)}{(n+m)\gamma\overline{\kappa}^2c^2}\underbrace{f(x^k,y^k)}. \\ \text{pelo } 4.25 & \\ t^k &\leq \frac{(1-\eta)\left(1-\rho_0\phi^{\beta-1}(x^k,y^k)\right)}{(n+m)\gamma\overline{\kappa}^2c^2}\left(1+c\sigma^2\right)c. \\ t^k &\leq \frac{(1-\eta)\left(1-\rho_0c^{\beta-1}\right)}{(n+m)\gamma\overline{\kappa}^2c}\left(1+c\sigma^2\right). \qquad (4.29) \end{aligned}$$

então, o lema fica provado para:

$$\xi = \min\left\{\frac{(1-\eta)\left(1-\rho_0c^{\beta-1}\right)}{(n+m)\gamma\overline{\kappa}^2c}\left(1+c\sigma^2\right);\theta\right\}, \quad \text{onde: } \theta = \min\left\{1;\frac{\rho_0\phi^{\beta-2}(x^k,y^k)}{n\gamma\overline{\kappa}}\right\}$$

■

Das proposições 4.7 e 4.8 obtemos que, para todo ponto $(x^k,y^k) \in \Omega^\star$, o ponto $(x^{k+1},y^{k+1}) = (x^k,y^k) + t^kd^k$, com $t^k \in [0,\xi]$ pertence ao conjunto Ω_c.

Agora, vamos mostrar que o passo de Armijo é limitado inferiormente no conjunto $\Omega^\star$.

Proposição 4.9 *Existe $\overline{\xi} > 0$ tal que, para $(x^k,y^k) \in \Omega^\star$, o ponto*

$$(x^{k+1},y^{k+1}) \;=\; (x^k,y^k) \;+\; t^kd^k,$$

pertence ao conjunto $\Omega^\star$ para qualquer $t^k \in [0,\overline{\xi}]$.

Prova:

Considerando [20] e as últimas proposições (4.7) e (4.8), temos $(x^{k+1}, y^{k+1}) = (x^k, y^k) + t^k d^k \in \Omega_c$ para todo $t^k \in (0, \xi)$. Vamos mostrar que $(x^{k+1}, y^{k+1}) \in \Omega^\star$.

Como na proposição (4.7), temos:

$$S_i\left((x^{k+1}, y^{k+1})\right) \geq \left(1 - t^k\right) S_i(x^k, y^k) + t^k \rho^k - \left(t^2\right)^2 \gamma \parallel d^k \parallel^2, \quad \text{para todo } i \in \{1, \cdots, n\}$$

somando todas as inequações acima obtemos:

$$\phi(x^{k+1}, y^{k+1}) \geq \left(1 - t^k\right) \phi(x^k, y^k) + n\rho^k t^k - n\left(t^k\right)^2 \gamma \parallel d^k \parallel^2, \tag{4.30}$$

Agora, usando a fórmula de Newton - Lebnitz [18], segue que:

$$\begin{aligned} \frac{Q(x^{k+1}, y^{k+1}) - Q(x^k, y^k)}{t^k} &= \int_0^1 \nabla Q\left((x^k, y^k) + \theta t^k d^k\right) d\theta. \\ Q(x^{k+1}, y^{k+1}) &= Q(x^k, y^k) + t^k \left[\int_0^1 \nabla Q\left((x^k, y^k) + \theta t^k d^k\right) d\theta\right] d^k. \end{aligned}$$

Somando e subtraindo: $t^k \nabla Q(x^k, y^k) d^k$, na equação acima:

$$\begin{aligned} Q(x^{k+1}, y^{k+1}) &= Q(x^k, y^k) + t^k \nabla Q(x^k, y^k) - t^k \nabla Q(x^k, y^k) d^k + \\ &\quad + t^k \left[\int_0^1 \nabla Q\left((x^k, y^k) + \theta t^k d^k\right) d\theta\right] d^k. \\ &= Q(x^k, y^k) + t^k \left(-Q(x^k, y^k)\right) - t^k \left[\int_0^1 \nabla Q(x^k, y^k) d\theta\right] d^k + \\ &\quad + t^k \left[\int_0^1 \nabla Q\left((x^k, y^k) + \theta t^k d^k\right) d\theta\right] d^k. \\ &= \left(1 - t^k\right) Q(x^k, y^k) + \left\{\int_0^1 \left[\nabla Q\left((x^k, y^k) + \theta t^k d^k\right) - \nabla Q(x^k, y^k)\right] d\theta\right\} d^k. \end{aligned}$$

Da suposição (4.2) sabemos que ∇Q é Lipschitziana, então:

$$\begin{aligned} \parallel Q(x^{k+1}, y^{k+1}) \parallel &= \parallel \left(1 - t^k\right) Q(x^k, y^k) + \\ &\qquad + \left\{\int_0^1 \left[\nabla Q\left((x^k, y^k) + \theta t^k d^k\right) - \nabla Q(x^k, y^k)\right] d\theta\right\} d^k \parallel \\ &\leq \parallel \left(1 - t^k\right) Q(x^k, y^k) \parallel + \\ &\qquad + \parallel \left\{\int_0^1 \left[\nabla Q\left((x^k, y^k) + \theta t^k d^k\right) - \nabla Q(x^k, y^k)\right] d\theta\right\} d^k \parallel \\ &= \left(1 - t^k\right) \parallel Q(x^k, y^k) \parallel + \\ &\quad + \mid t^k \mid \parallel \underbrace{\left\{\int_0^1 \left[\nabla Q\left((x^k, y^k) + \theta t^k d^k\right) - \nabla Q(x^k, y^k)\right] d\theta\right\}}_{} \parallel \mid d^k \mid \\ &\leq \left(1 - t^k\right) \parallel Q(x^k, y^k) \parallel + \mid t^k \mid L \parallel (x^k, y^k) + \theta t^k d^k - (x^k, y^k) \parallel \mid d^k \mid \\ &= \left(1 - t^k\right) \parallel Q(x^k, y^k) \parallel + \mid t^k \mid L\, \theta \mid t^k \mid\mid d^k \mid d^k \mid \\ &\leq \left(1 - t^k\right) \parallel Q(x^k, y^k) \parallel + \mid t^k \mid^2 L \mid d^k \mid^2, \end{aligned}$$

então:

$$\| Q(x^{k+1}, y^{k+1}) \| \leq \left(1 - t^k\right) \| Q(x^k, y^k) \| + \left(t^k\right)^2 L \mid d^k \mid^2,$$

multiplicando a expressão acima por $(-\sigma)$, da suposição (4.4), obtemos:

$$-\sigma \| Q(x^{k+1}, y^{k+1}) \| \geq -\sigma \left(1 - t^k\right) \| Q(x^k, y^k) \| - \sigma \left(t^k\right)^2 L \mid d^k \mid^2, \tag{4.31}$$

finalmente, fazendo $(x^{k+1}, y^{k+1}) = z^{k+1}$ e somando (4.30) e (4.31), temos que:

$$\begin{aligned} \phi(z^{k+1}) - \sigma \| Q(z^{k+1}) \| &\geq \left(1 - t^k\right) \left[\phi(x^k, y^k) - \sigma \| Q(x^k, y^k) \|\right] + n\rho^k t^k \\ &\quad - n \left(t^k\right)^2 \gamma \| d^k \|^2 - L \left(t^k\right)^2 \| d^k \|^2 . \\ &= \left(1 - t^k\right) \underbrace{\left[\phi(x^k, y^k) - \sigma \| Q(x^k, y^k) \|\right]}_{(\star_1)} + \\ &\quad + \underbrace{\left[n\rho^k - nt^k\gamma \| d^k \|^2 - Lt^k \| d^k \|^2\right]}_{(\star_2)} t^k. \end{aligned}$$

como $(\star_1) \geq 0$, basta que $(\star_2) \geq 0$, isto é:

$$n\rho^k - nt^k\gamma \| d^k \|^2 - Lt^k \| d^k \|^2 \geq 0, \tag{4.32}$$

então:

$$t^k \leq \frac{n\rho^k}{(n\gamma + \sigma L) \| d^k \|^2},$$

fazendo $\delta = \frac{n\gamma}{(n\gamma + \sigma L)}$, temos $\delta \in (0, 1]$, assim:

$$t^k \leq \delta \frac{\rho^k}{\gamma \| d^k \|^2}.$$

Portanto, basta tomar $\overline{\xi} = \min\left\{\delta \frac{\rho_{0c}{}^{\beta-2}}{\gamma n \overline{\kappa}^2}, \xi\right\}$ para ter $\phi(z^{k+1}) - \sigma \| Q(z^{k+1}, y^{k+1}) \| \geq 0$, para todo $t^k \in \left(0, \overline{\xi}\right)$ o que nos garante que $z^{k+1} \in \Omega^\star$. ■

Teorema 4.1 *Dado um ponto inicial estritamente viável,* $(x^0, y^0) \in \Omega^\star$, *existe uma subsequência de* $\{(x^k, y^k)\}$ *gerada pelo algoritmo* $FDA - MNCP$ *que converge para* $(x^\star, y^\star)$, *solução do problema de complementaridade mista.*

Prova:

Segue das proposições (4.4) até (4.8) que $\{(x^k, y^k)\} \in \Omega_c$. Como Ω_c é compacto, a sequência de pontos (x^k, y^k) possui um ponto de acumulação em Ω_c.
Seja $(x^\star, y^\star)$ ponto de acumulação da sequência $\{(x^k, y^k)\}$. Já que o tamanho de passo é sempre positivo e limitado inferiormente por $\nu\overline{\xi}$, concluimos que $\| d^k \| \longrightarrow 0$ e, do algoritmo, temos que $f(x^k, y^k) \longrightarrow 0$. Assim, $(x^\star, y^\star)$ é solução do problema de complementaridade mista, [20]. ■

4.4.1 **Análise de Convergência Assintótica**

Nesta seção veremos resultados sobre a taxa de convergência do algoritmo FDA-MNCP. Observaremos que uma iteração realizada no algoritmo é uma iteração do tipo Newton Amortecido com uma perturbação e as suposições pedidas no algoritmo, FDA-MNCP, que satisfazem as suposições do Método de Newton clássico (ver apêndice A.1) e que são utilizadas para o caso de Newton Amortecido com uma perturbação, [20].

Primeiramente veremos o esquema iterativo do método de Newton Amortecido com uma perturbação e compararemos com o esquema iterativo do algoritmo FDA-MNCP. Portanto, para resolver numericamente o sistema de equações (A.5), o método de Newton amortecido realiza a seguinte iteração

$$z^{k+1} \;=\; z^k \;-\; t^k(T'(z^k))^{-1}\left[T(z^k) - \mu^k P\right], \quad k = 0, 1, 2... \tag{4.33}$$

onde $0 < t^k \leq 1, \;\; \mu^k > 0$ e P um ponto fixo em $\mathbb{R}^p$.

Lembremos que a sequência gerada pelo algoritmo FDA-MNCP está baseada na resolução de um sistema de equações mais uma busca linear nas restrições de complementaridade e na função $f(x, y)$ associada ao problema. O que nos remete ao método de Newton Amortecido com uma perturbação (4.33. Observe que, quando tomamos $p = n,\; z^k = x^k,\; T(z^k) = H(x^k),\; (T'(z^k))^{-1} = (\nabla H(x^k))^{-1}, P = E, \mu^k = \rho^k$ e t^k como o passo da busca linear descrita no algoritmo FDA-MNCP, temos que a equação (4.33) é exatamente o esquema de iterações para o algoritmo FDA-MNCP, [20].

$$x^{k+1} \;=\; x^k \;-\; t^k(\nabla H(x^k))^{-1}\left[H(x^k) - \rho^k E\right], \quad k = 0, 1, 2... \tag{4.34}$$

onde $\rho^k \;=\; \rho_0 \frac{\phi^\beta(x^k, y^k)}{n}$ e $\beta \in [1, 2]$. Deste modo, temos o seguinte Teorema [20].

Teorema 4.2 *Considere a sequência $\{(x^k, y^k)\}$ gerada pelo algoritmo $FDA - MNCP$, que converge para uma solução $(x^\star, y^\star)$ do problema de complementaridade mista. Então,*

a) Tomando $\beta \in (1, 2),\; t^k = 1$ para cada k suficientemente grande e a taxa de convergência do algoritmo é superlinear.

b) Se $t^k = 1$ para k suficientemente grande e $\beta = 2$, então a taxa de convergência é quadrática.

Prova:

Fazendo do mesmo modo como [20] para o FDA-NCP e considerando o Teorema 4.1 deduzimos das equações 4.20 e 4.29 que para k suficientemente grande, o comprimento de passo obtido da busca linear de armijo é $t^k = 1$ se $\beta \in (1, 2)$. Tomando da equação 4.33 os seguintes valores:

$z^k = (x^k, y^k)$; $T(z^k) = S(x^k, y^k)$; $T'^{-1}(z^k) = S^{-1}(x^k, y^k)$; $P = \overline{E}$ e t^k que é o passo da busca linear descrita no algoritmo FDA-MNCP, o esquema de iteração para o algoritmo FDA-MNCP é

$$(x^{k+1}, y^{k+1}) = (x^k, y^k) - t^k S^{-1}(x^k, y^k)[S(x^k, y^k) - \rho^k \overline{E}, \quad k = 0, 1, 2, ...$$

onde $\rho^k = \rho_0 \frac{\phi^\beta(x^k, y^k)}{n}$ e $\beta \in [1, 2]$.

A análise que faremos agora segue os mesmos procedimentos do método de Newton usual e pode ser visto em [8].

$$\begin{aligned} \| z^{k+1} - z^\star \| &= \| z^k + t^k d^k - z^\star \| \\ &= \| z^k + t^k \left((\nabla S(z^k))^{-1}[-S(z^k) + \rho^k \overline{E}] \right) - z^\star \| . \end{aligned}$$

Somando e subtraindo $t^k z^k$ e $t^k z^\star$, temos

$$\begin{aligned} \| z^{k+1} - z^\star \| &= \| (1 - t^k)(z^k - z^\star) + t^k \left(z^k - z^\star + (\nabla S(z^k))^{-1}[-S(z^k + \rho^k P)] \right) \| \\ &= \| (1 - t^k)(z^k - z^\star) + t^k (\nabla S(z^k))^{-1} \left((z^k - z^\star)\nabla S(z^k) - S(z^k) + \rho^k P \right) \| . \end{aligned}$$

Pela desigualdade triangular,

$$\begin{aligned} \| z^{k+1} - z^\star \| &\leq \| (1 - t^k)(z^k - z^\star) \| + \\ &\quad + \| t^k (\nabla S(z^k))^{-1}[(z^k - z^\star)\nabla S(z^k) - S(z^k) + \rho^k P] \| \\ &= (1 - t^k) \| z^k - z^\star \| + \\ &\quad + t^k \| (\nabla S(z^k))^{-1} \| \| (z^k - z^\star)\nabla S(z^k) - S(z^k) + \rho P \| . \end{aligned}$$

Pela suposição 4.3, isto é, existe $k_0 > 0$ e $k > 0$ tal que $\| \nabla S(x, y) \| \leq k_0$ e $\| (\nabla S(x, y))^{-1} \| \leq k$, para todo $(x, y) \in \Omega_c$, então

$$\| z^{k+1} - z^\star \| \leq (1 - t^k) \| z^k - z^\star \| + t^k k \rho^k \| P \| + t^k k \| S(z^k) - \nabla S(z^k)(z^k - z^\star) \| .$$

Como $\nabla S(z^k)$ é localmente Lipschitz e da fórmula de Newton-Leibnitz, temos que a última norma do lado direito ficaria da seguinte forma, fazendo $d^k = \| z^k - z^\star \|$,

$$\begin{aligned} \| S(z^k) - \nabla S(z^k) d^k \| &= \| \left(\int_0^1 [\nabla S(z^k + \theta t^k d^k) - \nabla S(z^k)] d\theta \right) d^k \| \\ &= \| \int_0^1 [\nabla S(z^k + \theta t^k d^k) - \nabla S(z^k)] d\theta \| \| d^k \| \\ &\leq \gamma \| z^k + \theta t^k d^k - z^k \| \| d^k \| \\ &= \gamma \| \theta t^k d^k \| \| d^k \| \\ &= \gamma \| \theta \| t^k \| d^k \|^2 . \end{aligned} \tag{4.35}$$

Finalmente, lembrando que $\rho^k = \rho_0 \frac{\phi^\beta(z^k)}{n}$, temos

$$\begin{aligned} \| z^{k+1} - z^\star \| &\leq (1-t^k) \| z^k - z^\star \| + t^k k \rho^k \sqrt{n} + \left(t^k\right)^2 k\gamma \| \theta \| \| z^k - z^\star \|^2 \\ &= (1-t^k) \| z^k - z^\star \| + \frac{k\rho_0 \phi^\beta(z^k)}{\sqrt{n}} + O(\| z^k - z^\star \|^2). \end{aligned} \tag{4.36}$$

Pelo teorema de valor médio segue que:

$$\phi(y) \leq \phi(x) + \beta \phi^{\beta-1}(\overline{x}) \sqrt{n} O(\| y - x \|);$$

onde $\overline{x} = x + \epsilon(y - x)$ para algum $\epsilon \in (0, 1)$, tomando $x = z^\star$ e para todo $y = z^k$ suficientemente próximo de $x^\star$, temos:

$$\phi^\beta(z^k) \leq \beta \phi^{\beta-1}(\overline{x}) \sqrt{n} O(\| z^k - z^\star \|).$$

Então, para provar a parte (i), segue observando que para $\beta \in (1, 2)$

$$\phi(z^k) = O(\| z^k - z^\star \|),$$

e substituindo em 4.36 temos:

$$\lim_{k\to\infty} \frac{\| z^{k+1} - z^\star \|}{\| z^k - z^\star \|} = 0. \tag{4.37}$$

Portanto a convergência é superlinear.

Agora, para provar (ii), de forma similar observe que $\beta = 2$, temos que

$$\phi^2(z^k) = O(\| z^k - z^\star \|^2),$$

neste caso a equação (4.20) não garante que $t^k = 1$ para k suficientemente grande, assim pedimos que $t^k = 1$, para todo $k > k_0$ com k_0 suficientemente grande, logo

$$\lim_{k\to\infty} \frac{\| z^{k+1} - z^\star \|}{\| z^k - z^\star \|^2} < \infty. \tag{4.38}$$

Portanto a convergência é quadrática, pela Definição A.5. ■

5 COMBUSTÃO IN-SITU

Neste capítulo apresentaremos o estudo do método FDA-MNCP e a comparação com o método FDA-NCP [14], tomando como exemplo o modelo simples de combustão in-situ desenvolvido em [5], [23].

5.1 MODELAGEM DO PROBLEMA FÍSICO

O modelo estuda os fluxos unidimensionais, possuindo uma onda de combustão no caso quando o oxidante (ar com oxigênio) é injetado num meio poroso. Inicialmente o meio contém um combustível que é essencialmente imóvel, não vaporiza e a quantidade de oxigêneo é ilimitada. Este é o caso para combustíveis sólidos ou líquidos com baixas saturações.

Como em [23], estudamos o modelo simplificado onde:

- Parte pequena do espaço disponível é ocupado pelo combustível.
- Mudanças de porosidade na reação são desprezíveis.
- Temperatura do sólido e do gás são as mesma (equilíbrio térmico local).
- Velocidade do gás é constante.
- Perda de calor é desprezível.
- As variações de pressão são pequenas em comparação com a pressão prevalecente.

O modelo tem coordenadas temporal "t" e espacial "x" que inclui a equação de equilíbrio do calor, a equação de equilíbrio molar para combustíveis imóveis e a lei dos gases ideais.

$$C_m \frac{\partial T}{\partial t} + \frac{\partial}{\partial x}(C_g \rho u (T - T_{res})) = \lambda \frac{\partial^2 T}{\partial x^2} + Q_r W_r, \tag{5.1}$$

$$\frac{\partial \rho_f}{\partial t} = -\mu_f W_r, \tag{5.2}$$

$$\rho = \frac{P}{TR}, \tag{5.3}$$

onde $T[K]$ é a temperatura, $\rho[\frac{mol}{m^3}]$ é a densidade molar do gás e $\rho_f[\frac{mol}{m^3}]$ é a concentração molar do combustível imóvel. O conjunto de parâmetros junto a seus valores típicos são dados na Tabela 2.

Tabela 2 – Parâmetros dimensionais para combustão in-situ e seus valores típicos [23].

Símbolo	*Quantidade Física*	*Valor*	*Unidade*
T_{res}	Temperatura inicial do reservatório	273	$[K]$
C_m	Capacidade calorífica do meio porosso	$2 \cdot 10^6$	$[J/m^3K]$
c_g	Capacidade calorífica do gás	27.42	$[J/molK]$
λ	Condutividade térmica do meio porosso	0.87	$[J/(msK)]$
Q_r	Entalpia do combustível imóvel em T_{res}	$4 \cdot 10^5$	$[J/mol]$
u_{inj}	Veloc. de Darcy para injeção do gás $(200m/dia)$	0.0023	$[m/s]$
E_r	Energia de ativação	58000	$[J/mol]$
K_p	Parâmetro pré-exponencial	500	$1/s$
R	Constante dos gases ideais	8.314	$[J/molK]$
P	Pressão prevalecente $(1atm)$	101325	$[Pa]$
ρ_f^{res}	Densidade molar inicial do combustível	372	$[mol/m^3]$

Como em [23], se considerarmos por simplicidade $\mu_f = \mu_g = \mu_0 = 1$ e a quantidade de oxigêneo é ilimitada, a razão da reação W_r é tomada como:

$$W_r = k_p \rho_f \exp\left(\frac{-E_r}{RT}\right). \tag{5.4}$$

As variáveis a serem encontradas são a temperatura (T) e a concentração molar do combustível imóvel (ρ_f). Como as equações não estão dimensionadas, fazemos como [23], para obter a forma adimensional:

$$\frac{\partial \theta}{\partial t} + u\frac{\partial(\rho\theta)}{\partial x} = \frac{1}{P_{e_T}}\frac{\partial^2\theta}{\partial x^2} + \Phi(\theta,\eta). \tag{5.5}$$

$$\frac{\partial \eta}{\partial t} = \Phi(\theta,\eta). \tag{5.6}$$

$$\text{onde:} \quad \rho = \frac{\theta_0}{\theta+\theta_0}, \qquad \Phi = \beta(1-\eta)\exp\left(\frac{-E}{\theta+\theta_0}\right).$$

com as constantes adimensionais:

$$P_{E_T} = \frac{x^\star}{\lambda \Delta T^\star}, \quad \beta = \rho_f^\star k_p Q_r, \quad E = \frac{E_r}{R\Delta T^\star}, \quad \theta_0 = \frac{T_{res}}{\Delta T^\star}, \quad u = \frac{u_{inj} t^\star}{x^\star}. \tag{5.7}$$

Aqui P_{e_T} é o número de Peclet para difusão térmica, u torna-se a velocidade da onda térmica adimensional, E é a energia de ativação escalada e θ_0 é a temperatura do reservatório escalado.

Com as condições do reservatório iniciais:

$$t = 0;\ x \geq 0: \quad \theta = 0, \quad \eta = 0$$

e as condições de injeção:

$$t \geq 0;\ x = 0: \quad \theta = 0, \quad \eta = 1.$$

5.2 DESCRIÇÃO DO MÉTODO FDA-MNCP PARA O MODELO SIMPLES DE COMBUSTÃO IN-SITU

Agora descrevemos com detalhe o esquema de diferenças finitas para o modelo de combustão in-situ para o método FDA-MNCP. Para isso usamos a malha descrita no Capítulo 3, aplicando o método de Crank-Nicolson para aproximar as derivadas espaciais em cada caso, isto é:

$$\partial_t\theta(x_m,t_{n+\frac{1}{2}}) = \frac{\theta_m^{n+1}-\theta_m^n}{k}. \tag{5.8}$$

$$\partial_{xx}\theta(x_m,t_{n+\frac{1}{2}}) = \frac{\theta_{m+1}^{n+1}-2\theta_m^{n+1}+\theta_{m-1}^{n+1}}{2h^2}+\frac{\theta_{m+1}^{n}-2\theta_m^{n}+\theta_{m-1}^{n}}{2h^2}. \tag{5.9}$$

$$\partial_x F(\theta(x_m,t_{n+\frac{1}{2}})) = \frac{F_{m+1}^{n+1}-F_{m-1}^{n+1}}{4h}+\frac{F_{m+1}^{n}-F_{m-1}^{n}}{4h}. \tag{5.10}$$

$$\Phi(\theta(x_m,t_{n+\frac{1}{2}})) = \frac{\Phi_m^{n+1}+\Phi_m^n}{2}. \tag{5.11}$$

Considerando a condições de Dirichlet no ponto x_0:

$$\theta(x_0,t)=0, \qquad \eta(x_0,t)=1,$$

e uma condição de Neumann no ponto x_M:

$$\frac{\partial\theta}{\partial x}(x_M,t)=0, \qquad \frac{\partial\eta}{\partial x}(x_M,t)=0,$$

dadas em [23], logo, o valor em x_0 é conhecido em todo tempo mas não em x_M.

$$\theta_0^{n+1}=\theta_0^n, \quad \text{e} \quad \eta_0^{n+1}=\eta_0^n, \quad \text{para todo} \quad n\in\mathbb{N}. \tag{5.12}$$

A condição de fronteira em x_M fornece:

$$\frac{\partial\theta}{\partial x}(x_M,t)=0 \Rightarrow \frac{\theta_{M+1}^n-\theta_{M-1}^n}{2h}=0,$$

portanto

$$\theta_{M+1}^n=\theta_{M-1}^n \Rightarrow F_{M+1}^n=F_{M-1}^n, \text{ para todo } \ n\in\mathbb{N}. \tag{5.13}$$

Dando a forma do método FDA-MNCP feito no Capítulo 4, obtemos:

$$\theta\geq 0;\ \frac{\partial\theta}{\partial t}+u\frac{\partial(\rho\theta)}{\partial x}-\frac{1}{P_{e_T}}\frac{\partial^2\theta}{\partial x^2}-\Phi(\theta,\eta)\geq 0. \tag{5.14}$$

$$\frac{\partial\eta}{\partial t}-\Phi(\theta,\eta)=0. \tag{5.15}$$

Para obter as relações discretas de (5.14), substituimos (5.8-5.11) em (5.14) para obter:

$$\begin{aligned} -2\mu H\theta_{m-1}^{n+1} + (4+4\mu H)\theta_m^{n+1} - 2\mu H\theta_{m+1}^{n+1} + \lambda[F_{m+1}^{n+1} - F_{m-1}^{n+1}] - 2k\Phi_m^{n+1} \quad &\geq \quad (5.16) \\ \geq \quad 2\mu H\theta_{m-1}^{n} + (4-4\mu H)\theta_m^{n} + 2\mu H\theta_{m+1}^{n} + \lambda[F_{m+1}^{n} - F_{m-1}^{n}] + 2k\Phi_m^{n}, & \end{aligned}$$

onde $\lambda = \frac{k}{h}$ e $\mu = \frac{k}{h^2}$. O esquema é válido para $m = 1, 2, ..., M$ dos pontos que não são conhecidos os valores.

Nos pontos de fronteira temos que para $m = 1$, substituindo (5.12) em (5.16) obtemos:

$$\begin{aligned} (4+4\mu H)\theta_1^{n+1} - 2\mu H\theta_2^{n+1} + \lambda[F_2^{n+1} - F_0^{n+1}] - 2k\Phi_1^{n+1} \quad &\geq \quad (5.17) \\ \geq \quad (4-4\mu H)\theta_1^{n} + 2\mu H\theta_2^{n} - \lambda[F_2^{n} - F_0^{n}] + 2k\Phi_1^{n} + 4\mu H\theta_0^{n}, & \end{aligned}$$

para $m = M$, substituindo (5.13) em (5.16) obtemos:

$$-4\mu H\theta_{M-1}^{n+1} + (4+4\mu H)\theta_M^{n+1} - 2k\Phi_M^{n+1} \geq 4\mu H\theta_{M-1}^{n} + (4-4\mu H)\theta_M^{n} + 2k\Phi_M^{n}. \quad (5.18)$$

Então (5.16) é válido para todo $m = 2, ..., M-1$ e unindo as expresões (5.17) e (5.18) obtemos a seguinte desigualdade na variável θ^{n+1}:

$$G^n(\theta^{n+1}, \eta^{n+1}) = A\theta^{n+1} + \lambda P(\theta^{n+1}, \eta^{n+1}) - 2k\Phi(\theta^{n+1}, \eta^{n+1}) - LD(\theta^n, \eta^n) \geq 0, \quad (5.19)$$

onde $LD = B\theta^n - \lambda P(\theta^n, \eta^n) + 2k\Phi(\theta^n, \eta^n) + UR$ é conhecido em cada instante do tempo. Além disso:

$$A = \begin{bmatrix} 4+4\mu H & -2\mu H & 0 & 0 & \cdots & 0 & 0 & 0 \\ -2\mu H & 4+4\mu H & -2\mu H & 0 & \cdots & 0 & 0 & 0 \\ 0 & -2\mu H & 4+4\mu H & -2\mu H & & 0 & 0 & 0 \\ \vdots & & & & \ddots & & & \vdots \\ 0 & 0 & 0 & 0 & & -2\mu H & 4+4\mu H & -2\mu H \\ 0 & 0 & 0 & 0 & & 0 & -4\mu H & 4+4\mu H \end{bmatrix}, \quad (5.20)$$

$$B = \begin{bmatrix} 4-4\mu H & 2\mu H & 0 & 0 & \cdots & 0 & 0 & 0 \\ 2\mu H & 4-4\mu H & 2\mu H & 0 & \cdots & 0 & 0 & 0 \\ 0 & 2\mu H & 4-4\mu H & 2\mu H & & 0 & 0 & 0 \\ \vdots & & & & \ddots & & & \vdots \\ 0 & 0 & 0 & 0 & & 2\mu H & 4-4\mu H & 2\mu H \\ 0 & 0 & 0 & 0 & & 0 & 4\mu H & 4-4\mu H \end{bmatrix}, \quad (5.21)$$

$$P^n = P(\theta^n) = \begin{bmatrix} F_2^n - F_0^n \\ F_3^n - F_1^n \\ F_4^n - F_2^n \\ \vdots \\ F_M^n - F_{M-2}^n \\ 0 \end{bmatrix}, \quad \Phi^n = \Phi(\theta^n) = \begin{bmatrix} \Phi_1^n \\ \Phi_2^n \\ \Phi_3^n \\ \vdots \\ \Phi_{M-1}^n \\ \Phi_M^n \end{bmatrix}, \tag{5.22}$$

$$\theta^n = \begin{bmatrix} \theta_1^n \\ \theta_2^n \\ \theta_3^n \\ \vdots \\ \theta_{M-1}^n \\ \theta_M^n \end{bmatrix}, \quad UR = \begin{bmatrix} 4\mu H\theta_0^n \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \tag{5.23}$$

onde $A, B \in \mathbb{R}^{M\times M}$; $\theta^n, P^n, \Phi^n \in \mathbb{R}^M$.

De forma análoga, para obter a forma discreta de (5.15), substituimos (5.8-5.11) em (5.15) para obter:

$$Diag(2)\eta_m^{n+1} - k\Phi_m^{n+1} = Diag(2)\eta_m^n + k\Phi_m^n. \tag{5.24}$$

O esquema de diferenças é válido para todo $m = 1, 2, ..., M$ dos pontos que não são conhecidos os valores.

Nos pontos da fronteira temos que para $m = 1$, substituindo (5.12) em (5.24) obtemos:

$$2\eta_1^{n+1} - k\Phi_1^{n+1} = 2\eta_1^n + k\Phi_1^n, \tag{5.25}$$

para $m = M$, substituindo (5.13) em (5.24) tem-se:

$$2\eta_M^{n+1} - k\Phi_M^{n+1} = 2\eta_M^n + k\Phi_M^n. \tag{5.26}$$

Portanto, (5.24) é válido para todo $m = 2, ..., M-1$ e unindo as expresões (5.25) e (5.26) obtemos a seguinte desigualdade na variável η^{n+1}:

$$Q(\theta^{n+1}, \eta^{n+1}) = Diag(2)\eta^{n+1} - k\varphi(\theta^{n+1}, \eta^{n+1}) - LDQ(\theta^n, \eta^n), \tag{5.27}$$

onde $LDQ = Diag(2)\eta^n + k\varphi(\theta^n, \eta^n)$ é conhecido em cada instante do tempo.

Além disso:

$$Diag(2) = \begin{bmatrix} 2 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & & 0 & 0 & 0 \\ \vdots & & & & \ddots & & & \vdots \\ 0 & 0 & 0 & 0 & & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & & 0 & 0 & 2 \end{bmatrix}, \tag{5.28}$$

$$\varphi^n = \varphi(\theta, \eta^n) = \begin{bmatrix} \varphi_1^n \\ \varphi_2^n \\ \varphi_3^n \\ \vdots \\ \varphi_{M-1}^n \\ \varphi_M^n \end{bmatrix}, \quad \eta^n = \begin{bmatrix} \eta_1^n \\ \eta_2^n \\ \eta_3^n \\ \vdots \\ \eta_{M-1}^n \\ \eta_M^n \end{bmatrix}. \tag{5.29}$$

Portanto, a forma discreta de (5.14) e (5.15) é dada por (5.19) e (5.27).

$$G^n(\theta^{n+1}, \eta^{n+1}) \bullet \theta^{n+1} = 0, \tag{5.30}$$

$$Q^n(\theta^{n+1}, \eta^{n+1}) = 0, \tag{5.31}$$

e deve-se cumprir:

$$\theta^{n+1} \geq 0. \tag{5.32}$$

Assim, juntando (5.19), (5.27) (5.30) e (5.32) formam um Problema de Complementaridade Mista, que pode ser resolvido pelo algoritmo FDA-MNCP desenvolvido no Capítulo 4.

Algoritmo 5.1 (Implementação FDA-MNCP.)

Passo 1. *Fazer $n = 0$ e $N = 1/\Delta t$.*

Passo 2. *Para obter θ^{n+1} e η^{n+1} aplicamos o método $FDA-MNCP$ descrito no algoritmo para resolver o problema de complementaridade mista.*

$$\begin{aligned} G^n(\theta^{n+1}, \eta^{n+1}) \bullet \theta^{n+1} = 0, \quad \theta^{n+1} \geq 0, \\ Q^n(\theta^{n+1}, \eta^{n+1}) = 0 \end{aligned} \tag{5.33}$$

com

$$\begin{aligned} G^n(\theta^{n+1}, \eta^{n+1}) = A\theta^{n+1} + \lambda P(\theta^{n+1}, \eta^{n+1}) - 2k\Phi(\theta^{n+1}, \eta^{n+1}) - LD(\theta^n, \eta^{n+1}) \geq 0, \\ Q(\theta^{n+1}, \eta^{n+1}) = Diag(2) \cdot \eta^{n+1} - k\varphi(\theta^{n+1}, \eta^{n+1}) - LDQ(\theta^n, \eta^{n+1}) = 0. \end{aligned}$$

onde as matrizes A, B e os vetores P, Φ e φ são dadas em (5.20), (5.21), (5.22), e (5.29).

Passo 3. *Se $n = N$ então* ***FIM.***
Se não $n = n + 1$ e voltar ao Passo ***2***

Os resultados numéricos obtidos da implementação do Algoritmo no Matlab são mostrados na Seção seguinte.

5.3 COMPARAÇÃO DOS MÉTODOS FDA-MNCP COM O MÉTODO FDA-NCP

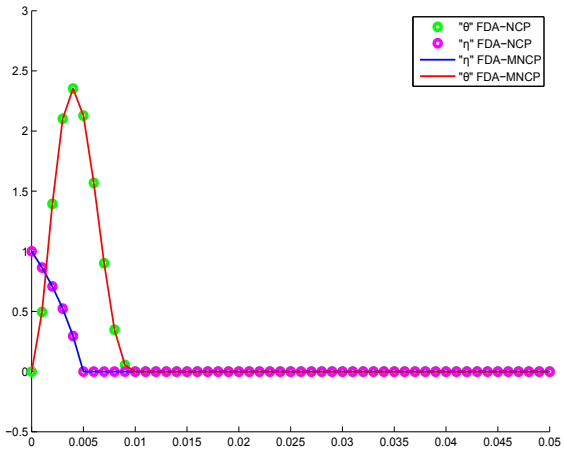
Agora apresentaremos os resultados numéricos das simulações feitas no matlab para o método FDA-MNCP. Para tal simulação, consideramos $[x_0, x_M] = [0,\ 0.05]$ e $[t_0, t_N] = [0, 1]$ como os intervalos de espaço e tempo respectivamente. Mantemos os números de subintervalos em tempo constante e igual a: $N = 10^5$, isto é, $k = \Delta t = 10^{-5}$, enquanto o número de subintervalos em espaço serão iguais a $M = 50, 100, 200, 400$. Para o método FDA-MNCP consideremos uma tolerância de erro de 10^{-8}.

Os valores dos parâmetros adimensionais em 5.7 são:

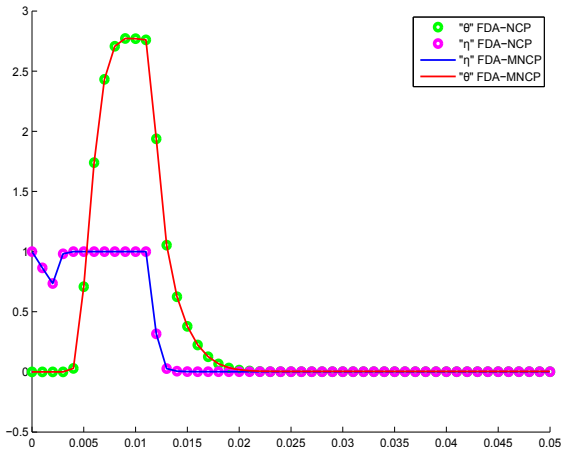
$$\begin{array}{lll} x^\star = 9,1 \times 10^4[m], & t^\star = 1,48 \times 10^8[s], & \Delta T^\star = 74,4[K], \\ u^\star = 6,1 \times 10^{-4}, & P_{e_T} = 1406, & \beta = 7,44 \times 10^{10}, \\ E = 93,8 & \theta_0 = 3.67 & u = 3,76. \end{array}$$

com os dados de entrada anteriores obtemos as Figuras (13), (14), (15), (16), as quais mostram os resultados obtidos pelos algoritmos 5.1 e algoritmo 5 de [23] para o método FDA-MNCP e FDA-NCP [14] , respectivamente.

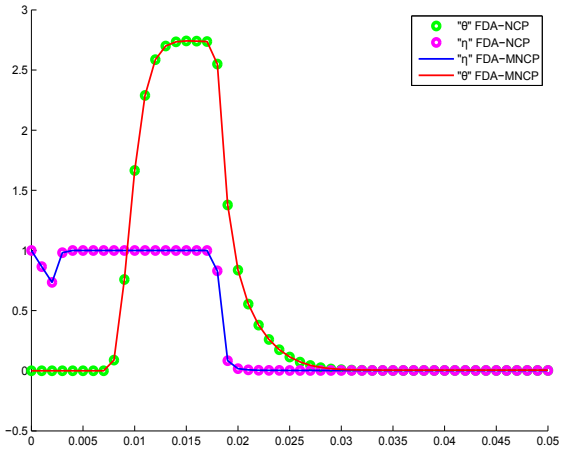
Figura 13 – Comparação dos métodos de FDA-MNCP e FDA-NCP para $M = 50$ nos instantes de tempos $t = 0.000,\ 0.002,\ 0.004,\ 0.006,\ 0.008,\ 0.010$. Os valores de θ são representados por bolinhas verdes e linha contínua vermelha, os valores de η são representados por bolinhas rosas e linha contínua azul.



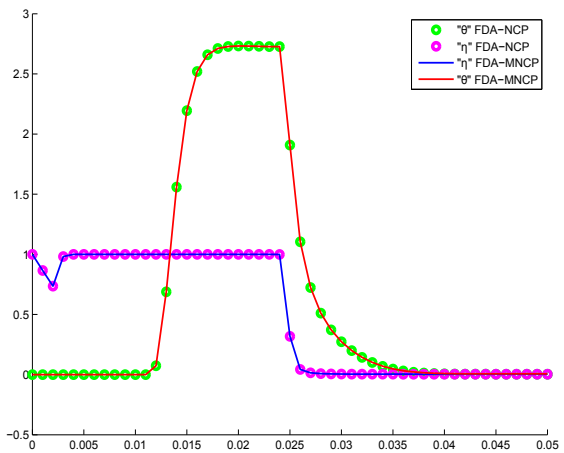
$(a) \quad t = 0.000$



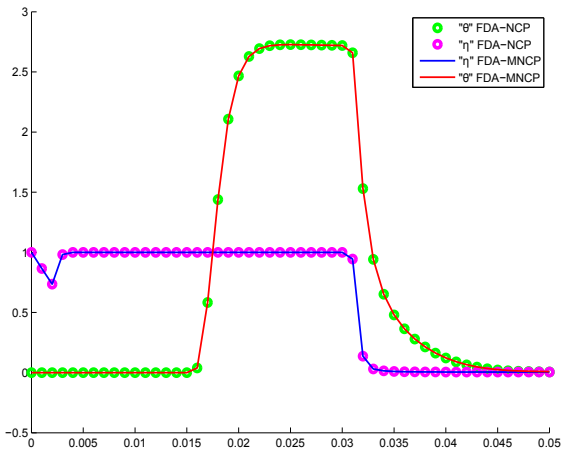
$(b) \quad t = 0.002$



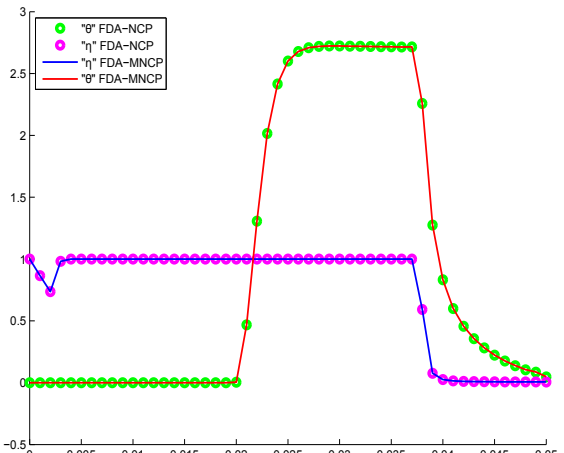
$(c) \quad t = 0.004$



$(d) \quad t = 0.006$

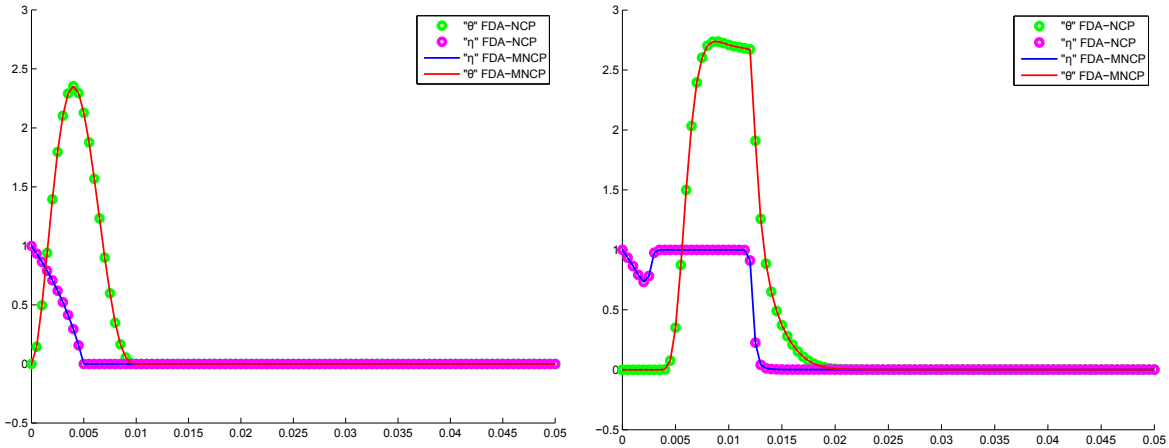


$(e) \quad t = 0.008$



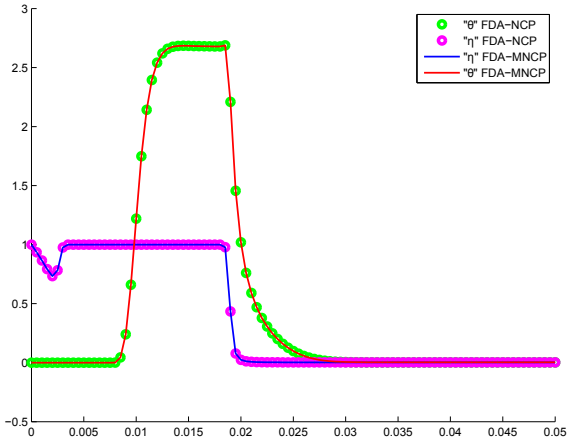
$(f) \quad t = 0.010$

Figura 14 – Comparação dos métodos de FDA-MNCP e FDA-NCP para $M = 100$ nos instantes de tempos $t = 0.000,\ 0.002,\ 0.004,\ 0.006,\ 0.008,\ 0.010$. Os valores de θ são representados por bolinhas verdes e linha contínua vermelha, os valores de η são representados por bolinhas rosas e linha contínua azul.

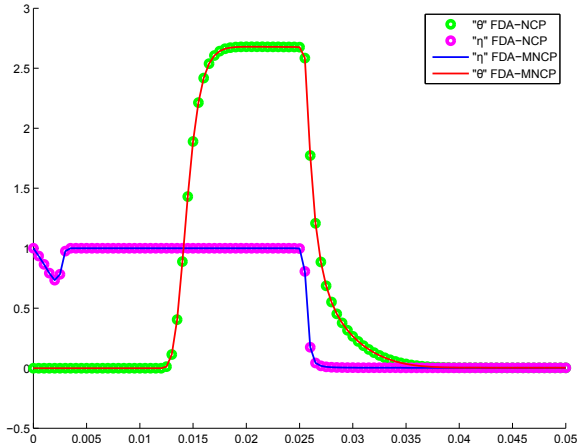


$(a) \quad t = 0.000$

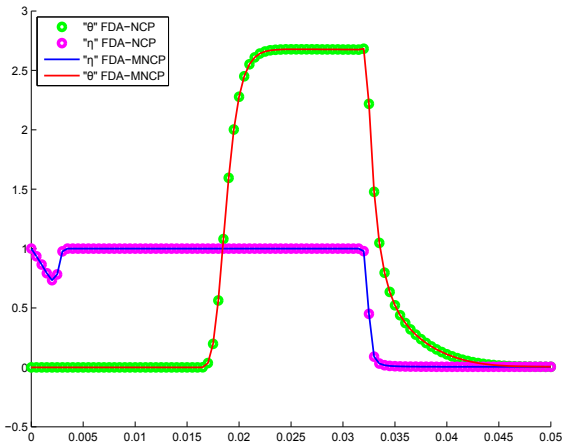
$(b) \quad t = 0.002$



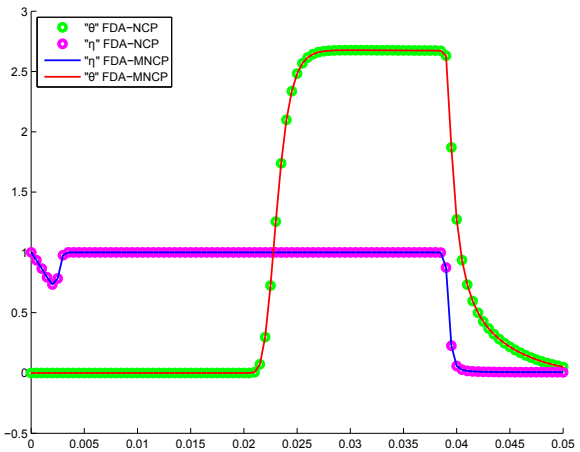
$(c) \quad t = 0.004$



$(d) \quad t = 0.006$

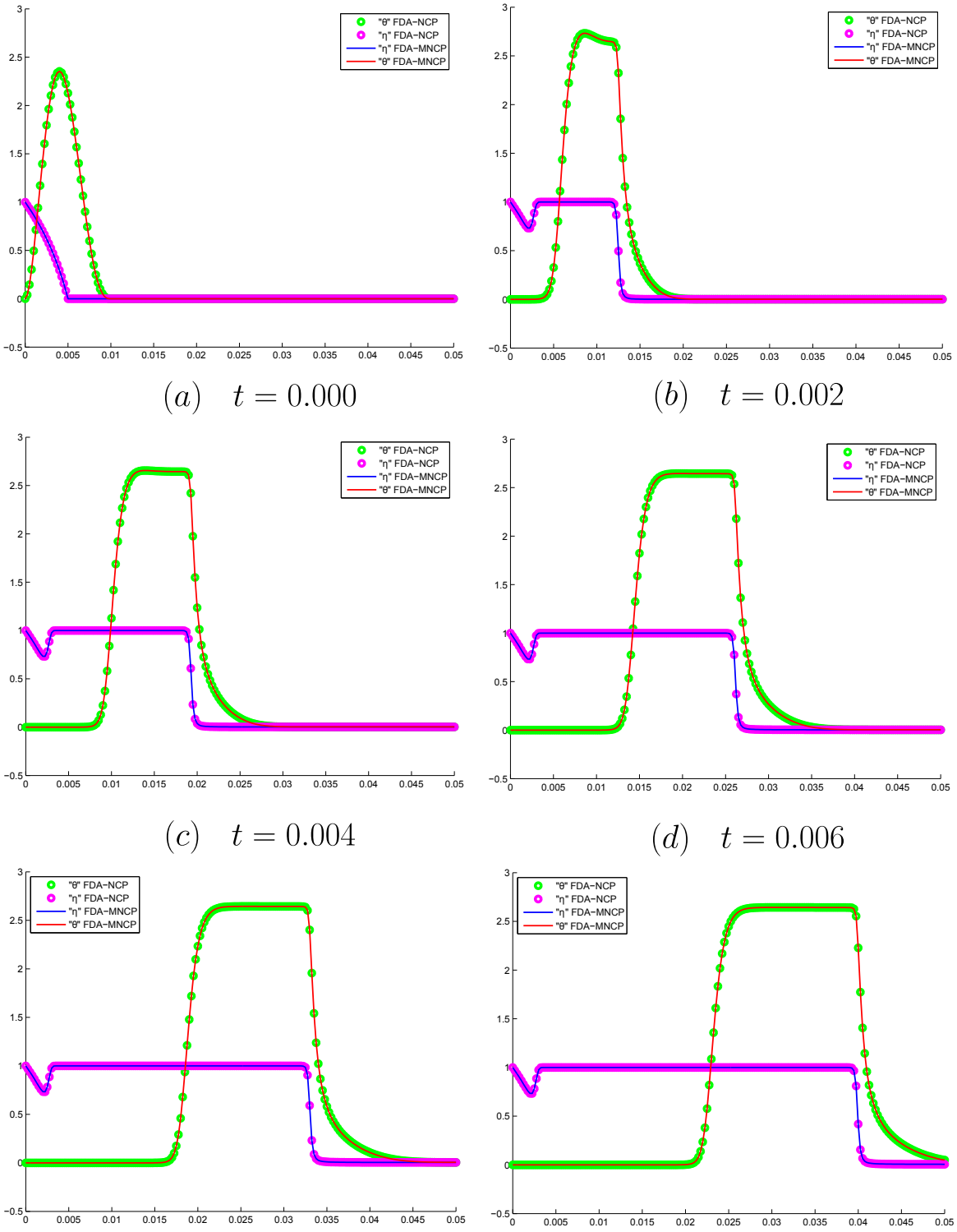


$(e) \quad t = 0.008$



$(f) \quad t = 0.010$

Figura 15 – Comparação dos métodos de FDA-MNCP e FDA-NCP para $M = 200$ nos instantes de tempos $t = 0.000,\ 0.002,\ 0.004,\ 0.006,\ 0.008,\ 0.010$. Os valores de θ são representados por bolinhas verdes e linha contínua vermelha, os valores de η são representados por bolinhas rosas e linha contínua azul.

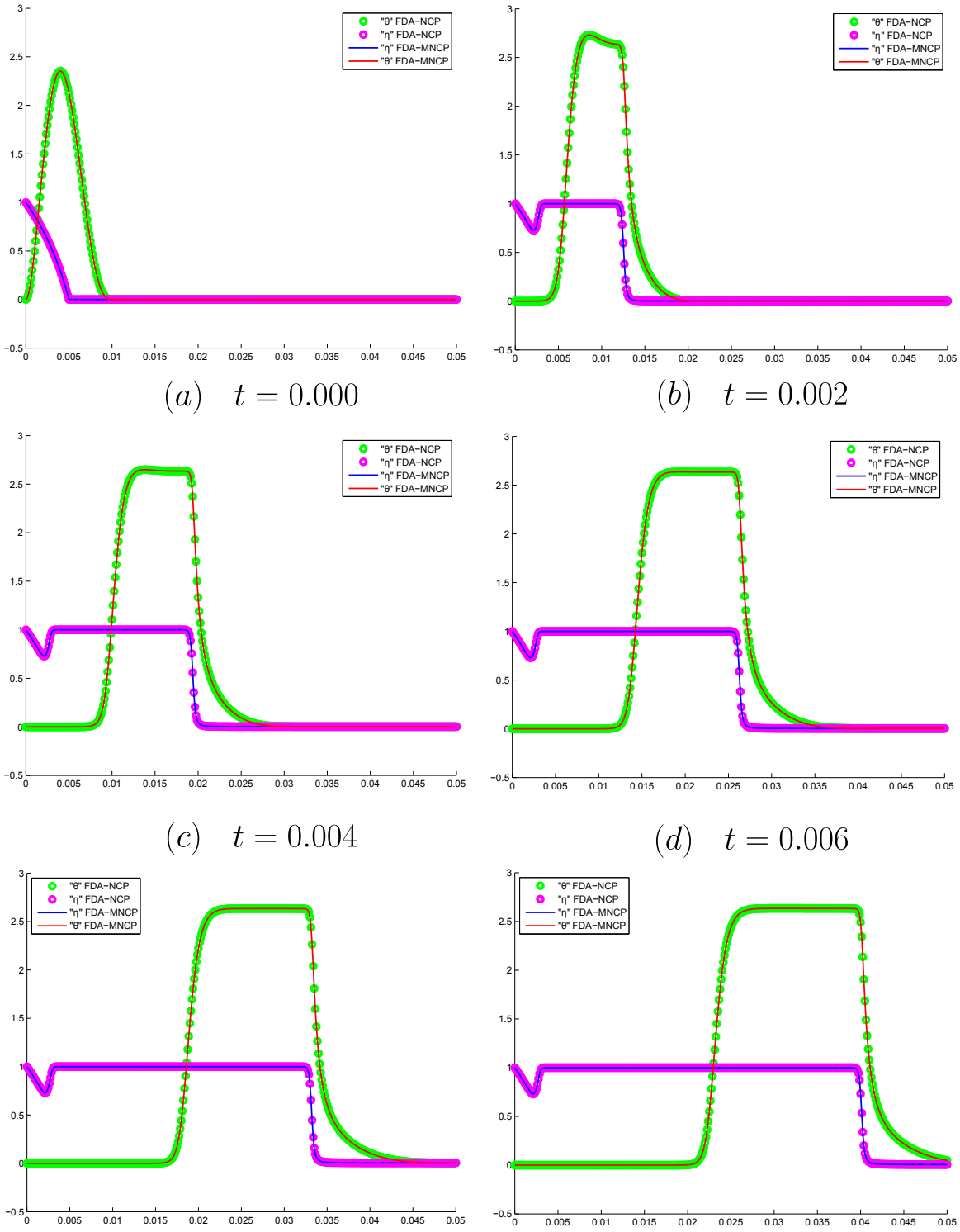


$(a)\quad t = 0.000$ $(b)\quad t = 0.002$

$(c)\quad t = 0.004$ $(d)\quad t = 0.006$

$(e)\quad t = 0.008$ $(f)\quad t = 0.010$

Figura 16 – Comparação dos métodos de FDA-MNCP e FDA-NCP para $M = 400$ nos instantes de tempos $t = 0.000,\ 0.002,\ 0.004,\ 0.006,\ 0.008,\ 0.010$. Os valores de θ são representados por bolinhas verdes e linha contínua vermelha, os valores de η são representados por bolinhas rosas e linha contínua azul.



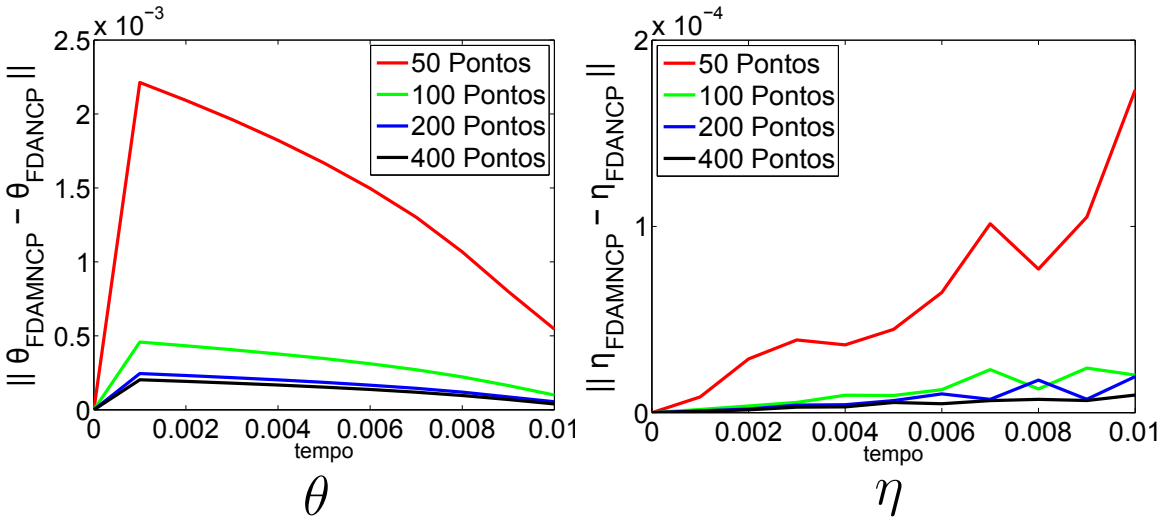
$(a)\quad t = 0.000$ $(b)\quad t = 0.002$

$(c)\quad t = 0.004$ $(d)\quad t = 0.006$

$(e)\quad t = 0.008$ $(f)\quad t = 0.010$

Como podemos observar nas Figuras 13, 14, 15 e 16 obtidas anteriormente nos mostram que os resultados obtidos pelos métodos FDA-MNCP, estudado na Seção 5.2, e FDA-NCP, estudado em [23], coincidem muito bem como pode ser visto nos tempos indicados nas figuras apresentadas anteriormente. Na Figura 17, pode-se observar a diferenças entre os métodos.

Figura 17 – Diferença entre os métodos FDA-MNCP e FDA-NCP.



Onde podemos observar que a diferença das soluções de θ e η são muito pequenas a medida que aumentamos a quantidade de pontos, entre os métodos FDA-MNCP e FDA-NCP [14].

A seguir mostramos quatro tabelas onde fazemos a comparação do tempo de processo computacional para $M = 50,\ 100,\ 20,\ 400$ do método FDA-MNCP estudado na Seção 5.2 e o método FDA-NCP estudado em [23].

Tabela 3 – Comparação do tempo de processo computacional com $M = 50$ para os métodos FDA-MNCP e FDA-NCP. O $t(n)$ é o tempo medido em segundos que demorou o método em encontrar a solução no instante de tempo t.

	FDA-MNCP					FDA-NCP				
t	$t(n)$	iter	BL(t)	$[S]$	$[\nabla S]$	$t(n)$	iter	BL(t)	F	∇F
0.001	0.401	33	1.0	43	34	0.194	20	1.0	43	40
0.002	0.411	34	0.8	44	35	0.194	20	1.0	43	40
0.003	0.406	34	1.0	43	35	0.199	20	1.0	43	40
0.004	0.386	32	0.8	41	33	0.184	20	1.0	43	40
0.005	0.385	32	0.8	41	33	0.178	21	1.0	45	42
0.006	0.406	34	0.8	43	35	0.198	21	1.0	45	42
0.007	0.395	33	0.8	42	34	0.193	21	1.0	45	42
0.008	0.403	33	0.8	43	34	0.229	21	1.0	45	42
0.009	0.390	32	0.8	42	33	0.169	21	1.0	45	42
0.010	0.405	33	0.8	44	34	0.173	21	1.0	45	42

Como podemos observar na Tabela 3, o tempo de processo computacional utilizado pelo método FDA-MNCP duplica o tempo do método FDA-NCP para uma partição do eixo x em cinquenta pontos; o número de iterações do método FDA-MNCP para os tempos indicados também é maior do método FDA-NCP. Além disso a terceira coluna de cada tabela mostra o valor de t^k (visto no algoritmo 4.1 do Capítulo 4) para cada um dos métodos.

Tabela 4 – Comparação do tempo de processo computacional com $M = 100$ para os métodos FDA-MNCP e FDA-NCP. O $t(n)$ é o tempo medido em segundos que demorou o método em encontrar a solução no instante de tempo t.

	FDA-MNCP					FDA-NCP				
t	$t(n)$	iter	BL(t)	$[S]$	$[\nabla S]$	$t(n)$	iter	BL(t)	F	∇F
0.001	0.912	36	0.8	47	37	0.555	21	1.0	45	42
0.002	0.885	36	0.8	46	37	0.519	21	1.0	45	42
0.003	0.870	37	0.8	47	38	0.515	21	1.0	45	42
0.004	0.892	35	0.64	46	36	0.494	21	1.0	45	42
0.005	0.831	34	0.8	44	35	0.486	21	1.0	45	42
0.006	0.793	34	0.8	43	35	0.508	21	1.0	45	42
0.007	0.790	34	0.8	43	35	0.457	21	1.0	45	42
0.008	0.894	36	0.64	49	37	0.515	21	1.0	45	42
0.009	0.869	35	0.8	46	36	0.507	21	1.0	45	42
0.010	0.839	35	0.8	46	36	0.479	21	1.0	45	42

Nesta Tabela 4, podemos observar ainda que o tempo de processo computacional utilizado pelo método FDA-MNCP é um pouco maior que o tempo do método FDA-NCP como o número de iterações.

Tabela 5 – Comparação do tempo de processo computacional com $M = 200$ para os métodos FDA-MNCP e FDA-NCP. O $t(n)$ é o tempo medido em segundos que demorou o método em encontrar a solução no instante de tempo t.

	FDA-MNCP					FDA-NCP				
t	$t(n)$	iter	BL(t)	$[S]$	$[\nabla S]$	$t(n)$	iter	BL(t)	F	∇F
0.001	1.771	36	0.8	47	37	3.223	21	1.0	45	42
0.002	1.852	38	0.8	48	39	3.185	21	1.0	45	42
0.003	1.969	37	0.8	46	38	3.129	21	1.0	45	42
0.004	1.995	38	0.8	47	39	3.245	21	1.0	45	42
0.005	1.771	37	0.8	46	38	3.143	21	1.0	45	42
0.006	1.785	37	1.0	45	38	3.122	21	1.0	45	42
0.007	1.787	37	0.8	47	38	3.205	21	1.0	45	42
0.008	1.778	37	0.8	47	38	3.241	21	1.0	45	42
0.009	2.001	36	0.8	46	37	3.346	22	1.0	47	44
0.010	1.685	35	0.8	45	36	3.396	22	1.0	47	44

Na Tabela 5, verificamos que quando aumentamos o número de partições a 200 o tempo de processo computacional do método FDA-MNCP começa a ser menor que o tempo do método FDA-NCP, embora o número de iterações ainda seja maior.

Tabela 6 – Comparação do tempo de processo computacional com $M = 400$ para os métodos FDA-MNCP e FDA-NCP. O $t(n)$ é o tempo medido em segundos que demorou o método em encontrar a solução no instante de tempo t.

	FDA-MNCP					FDA-NCP				
t	$t(n)$	iter	BL(t)	$[S]$	$[\nabla S]$	$t(n)$	iter	BL(t)	F	∇F
0.001	4.497	37	0.8	48	38	33.127	21	1.0	45	21
0.002	4.337	37	1.0	46	38	33.174	21	1.0	45	21
0.003	4.533	38	0.8	47	39	33.034	21	1.0	45	21
0.004	4.494	39	0.8	48	40	33.398	21	1.0	45	21
0.005	4.429	37	1.0	45	38	34.722	22	1.0	47	22
0.006	4.329	37	1.0	45	38	34.652	22	1.0	47	22
0.007	4.523	39	0.8	49	40	34.777	22	1.0	47	22
0.008	4.530	39	0.8	49	40	34.802	22	1.0	47	22
0.009	4.459	37	0.8	47	38	34.824	22	1.0	47	22
0.010	4.315	37	1.0	48	38	34.571	22	1.0	47	22

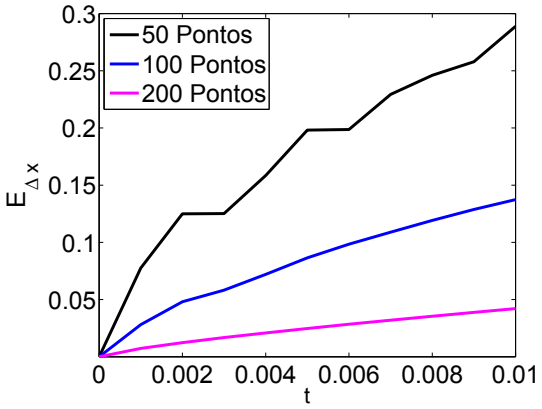
Com a última Tabela 6, podemos observar que aumentando consideravelmente o número de partições o tempo de processo computacional do método FDA-MNCP é muito menor que o tempo do método FDA-NCP [14], isso porque para o cálculo ∇S no método FDA-MNCP, este faz a metade de cálculos que para ∇F do método FDA-NCP [14], embora o número de iterações seja maior.

Também podemos observar que o número de iterações varia pouco em cada uma das tabelas apresentadas anteriormente.

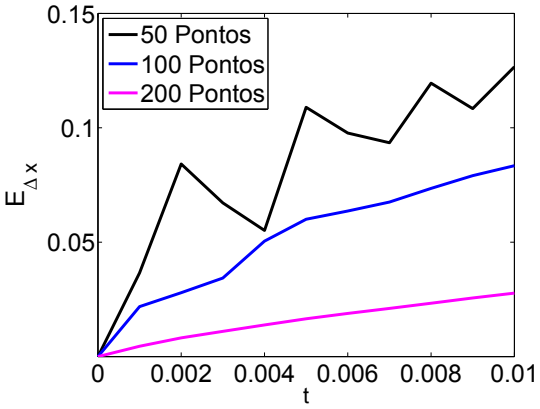
5.4 ANÁLISE DE ERRO

Nesta seção faremos o estudo numérico do erro relativo para o método FDA-MNCP, segundo a Tabela 1 da Secção 3.2 e com a norma dada em (3.7) para cada instante de tempo e $E_{\Delta x}$, onde $\Delta x = h,\ h/2,\ h/4$. O comprimento do subintervalo em tempo será constante e igual a $\Delta t = k = 10^{-5}$ e $h = \frac{1}{50}$, e comparado com o método da FDA-NCP, [23]. As Tabelas 7 e 9 mostram os resultados para os Erros Relativos do método de FDA-MNCP, enquanto as Tabelas 8 e 10 mostram o Erro Relativo para o método de FDA-NCP e são representados nas Figuras 18 e 19.

Figura 18 – Tempo (t)vs$E_{\Delta x}$. Erro Relativo do Método FDA-MNCP. Aqui $\Delta x = \frac{1}{50},\ \frac{1}{100},\ \frac{1}{200}$.

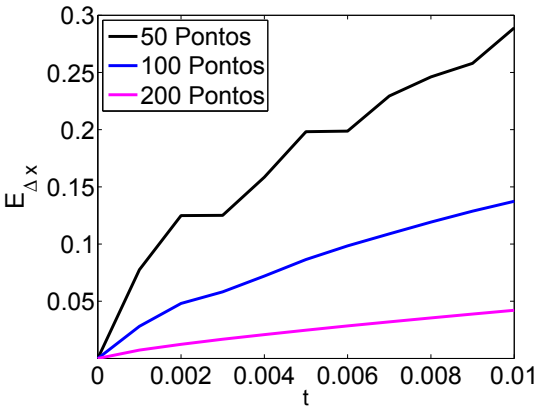


(a) θ vs $E_{\Delta x}$

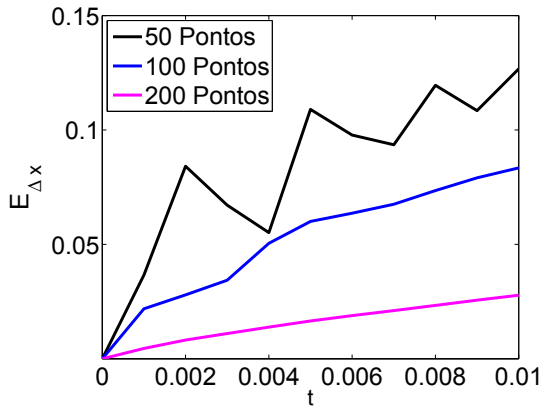


(b) η vs $E_{\Delta x}$

Figura 19 – Tempo (t)vs$E_{\Delta x}$. Erro Relativo do Método FDA-NCP. Aqui $\Delta x = \frac{1}{50},\ \frac{1}{100},\ \frac{1}{200}$.



(a) θ vs $E_{\Delta x}$



(b) η vs $E_{\Delta x}$

Como pode-se observar nas Figuras 18 e 19, os erros relativos correspondentes aos métodos FDA-MNCP e FDA-NCP, respectivamente, são muito similares, como poderemos observar nas seguintes tabelas onde se mostra os erros relativos para θ e η usando $h = 1/50$.

Tabela 7 – Erro Relativo para θ com FDA-MNCP e $h = \frac{1}{50}$ para os instantes de tempo "t" indicados na primeira coluna.

t	E_h	$E_{\frac{h}{2}}$	$E_{\frac{h}{4}}$	$\frac{E_h}{E_{\frac{h}{2}}}$	$\frac{E_{\frac{h}{2}}}{E_{\frac{h}{4}}}$
0.00100000	0.07757130	0.02809022	0.00730451	2.76	3.85
0.00200000	0.12495960	0.04805076	0.01233415	2.60	3.90
0.00300000	0.12512590	0.05824369	0.01679689	2.15	3.47
0.00400000	0.15843617	0.07200921	0.02080162	2.20	3.46
0.00500000	0.19817638	0.08645161	0.02467447	2.29	3.50
0.00600000	0.19866653	0.09841438	0.02840411	2.02	3.46
0.00700000	0.22936049	0.10892940	0.03195340	2.11	3.41
0.00800000	0.24606770	0.11916519	0.03538988	2.06	3.37
0.00900000	0.25789027	0.12872004	0.03877097	2.00	3.32
0.01000000	0.28886846	0.13735367	0.04209700	2.10	3.26

Tabela 8 – Erro Relativo para θ com FDA-NCP e $h = \frac{1}{50}$ para os instantes de tempo "t" indicados na primeira coluna.

t	E_h	$E_{\frac{h}{2}}$	$E_{\frac{h}{4}}$	$\frac{E_h}{E_{\frac{h}{2}}}$	$\frac{E_{\frac{h}{2}}}{E_{\frac{h}{4}}}$
0.001	0.07753953	0.02808885	0.00730328	2.76	3.85
0.002	0.12490382	0.04804832	0.01233527	2.60	3.89
0.003	0.12507648	0.05824052	0.01679940	2.15	3.47
0.004	0.15841526	0.07200834	0.02080540	2.20	3.46
0.005	0.19818512	0.08645353	0.02467944	2.29	3.50
0.006	0.19869586	0.09841770	0.02841001	2.02	3.46
0.007	0.22940100	0.10893437	0.03196011	2.10	3.41
0.008	0.24613384	0.11917200	0.03539747	2.07	3.37
0.009	0.25797041	0.12872819	0.03877948	2.00	3.32
0.010	0.28897837	0.13736335	0.04210642	2.10	3.26

Nas Tabelas 7 e 8 verificamos que os erros relativos E_h, $E_{\frac{h}{2}}$, $E_{\frac{h}{4}}$ para θ tanto do método FDA-MNCP como o método FDA-NCP são muito similares já que nas três primeiras casas decimais são os mesmos. Além disso, como vimos na Tabela 1 o $\frac{E_{\Delta x}}{E_{\frac{\Delta x}{2}}} \approx 4$, para $\Delta x = \frac{h}{2}$.

Tabela 9 – Erro Relativo para η com FDA-MNCP e $h = \frac{1}{50}$ para os instantes de tempo "t" indicados na primeira coluna.

t	E_h	$E_{\frac{h}{2}}$	$E_{\frac{h}{4}}$	$\frac{E_h}{E_{\frac{h}{2}}}$	$\frac{E_{\frac{h}{2}}}{E_{\frac{h}{4}}}$
0.00100000	0.03656879	0.02183277	0.00451877	1.67	4.83
0.00200000	0.08417410	0.02792488	0.00819559	3.01	3.41
0.00300000	0.06717364	0.03432880	0.01103677	1.96	3.11
0.00400000	0.05513563	0.05046208	0.01381861	1.09	3.65
0.00500000	0.10897196	0.06002542	0.01653614	1.82	3.63
0.00600000	0.09771893	0.06361601	0.01887177	1.54	3.37
0.00700000	0.09347666	0.06753455	0.02105732	1.38	3.21
0.00800000	0.11948456	0.07350345	0.02332175	1.63	3.15
0.00900000	0.10838295	0.07908666	0.02560247	1.37	3.09
0.01000000	0.12663557	0.08343716	0.02775604	1.52	3.01

Tabela 10 – Erro Relativo para η com FDA-NCP e $h = \frac{1}{50}$ para os instantes de tempo "t" indicados na primeira coluna.

t	E_h	$E_{\frac{h}{2}}$	$E_{\frac{h}{4}}$	$\frac{E_h}{E_{\frac{h}{2}}}$	$\frac{E_{\frac{h}{2}}}{E_{\frac{h}{4}}}$
0.001	0.03656752	0.02183277	0.00451876	1.67	4.83
0.002	0.08415103	0.02792385	0.00819578	3.01	3.41
0.003	0.06714517	0.03432836	0.01103751	1.96	3.11
0.004	0.05513010	0.05046448	0.01382034	1.09	3.65
0.005	0.10898570	0.06002658	0.01653867	1.82	3.63
0.006	0.09776211	0.06361489	0.01887480	1.54	3.37
0.007	0.09351854	0.06753575	0.02106120	1.38	3.21
0.008	0.11952653	0.07350797	0.02332674	1.63	3.15
0.009	0.10846255	0.07909222	0.02560830	1.37	3.09
0.010	0.12671978	0.08344353	0.02776237	1.52	3.01

De forma análoga, nas Tabelas 9 e 10 observamos que os erros relativos E_h, $E_{\frac{h}{2}}$, $E_{\frac{h}{4}}$ para η tanto do método FDA-MNCP como o método FDA-NCP são muito similares, uma vez que as três primeiras casas decimais são as mesmos. Além disso, como vimos na Tabela 1 o $\frac{E_{\Delta x}}{E_{\frac{\Delta x}{2}}} \approx 4$, para $\Delta x = \frac{h}{2}$.

6 CONCLUSÕES

Propomos a solução do modelo simples de combustão in-situ utilizando um método numérico baseado num esquema de diferenças finitas implícito e um algoritmo de complementaridade mista não linear, que podem ser aplicados a problemas parabólicos podendo ser reescrito na forma de um problema de complementaridade mista. Este método é aplicado ao sistema (5.5).

Neste trabalho, vimos que o algoritmo de ponto interior viável FDA-MNCP, proposto no Capítulo 4, é uma boa técnica para resolver numericamente os problemas de complementaridade mista. Resultados teóricos do algoritmo asseguram a convergência global feita no Capítulo 4.

Resolvemos o modelo de combustão in-situ (5.5) usando o algoritmo de complementaridade mista não linear e comparando ao método FDA-NCP, sendo as duas soluções muito próximas como podemos observar nas Figuras 13, 14, 15 e 16. Isto sugere que é possível aplicar este método a problemas parabólicos e hiperbólicos, que podem ser escritos como problema de complementaridade mista. O método FDA-MNCP mostra a vantagem de poder ser usado com mais pontos na discretização do espaço, como pode-se observar na Tabela 5 e 6, além de ser mais rápido em tempo de processo computacional que o método FDA-NCP quando aumentamos a discretização no espaço. Com respeito aos erros relativos, temos que as Tabelas 7 e 9 mostram uma boa evidência de uma convergência do algoritmo de complementaridade mista, o que é observado na Figura 17, o qual indica uma diminuição do crescimento linear quando refinamos a malha.

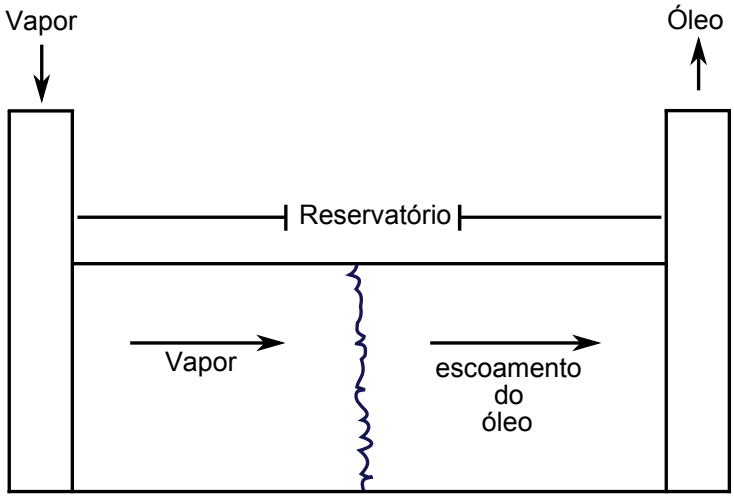
Para trabalhos futuros teremos a possibilidade de obter uma boa aproximação da solução de um sistema como se mostra em 7.1, o qual consta de uma equação Parabólica e duas hiperbólicas pelo método FDA-MNCP apresentado no Capítulo 4.

7 TRABALHOS FUTUROS - MODELO SIMPLES DA INJEÇÃO DE VAPOR NUM MEIO POROSO SATURADO DE ÁGUA

7.1 MODELAGEM DO PROBLEMA FÍSICO

Estudaremos o modelo que descreve a injeção de vapor num meio poroso linear que inicialmente está saturado de água, como mostra a Figura 20.

Figura 20 – Modelo



Fonte: Julio Agustín

Estudaremos um modelo simplificado, da mesma forma que [2]. Vamos negligenciar as perdas de calor, energia cinética e pressão capilar. Além disso negligenciamos o conteúdo do óleo já que estamos mais interesados nos efeitos da evaporação e condensação. Ainda, a velocidade da injeção do vapor, permeabilidade e porosidade são constantes.

O modelo consiste numa equação parabólica de balanço de energia e duas equações hiperbólicas para o equilíbrio do vapor e água.

O modelo tem coordenadas temporal "t" e espacial "x" que inclui a equações de Energia (Parabólica) e as equações de vapor e água(Hiperbólicas).

$$\frac{\partial}{\partial t}[(1-\varphi)C_r T]+\frac{\partial}{\partial x}\left[\rho_w u f_w C_w T+\rho_g u f_g C_g T\right] = \nabla\cdot(k_T\nabla T)+Q_v\cdot q \tag{7.1}$$

$$\frac{\partial(\varphi\rho_w S_w)}{\partial t}+\frac{\partial(\rho_w u f_w)}{\partial x} = \quad q \tag{7.2}$$

$$\frac{\partial(\varphi\rho_g S_g)}{\partial t}+\frac{\partial(\rho_g u f_g)}{\partial x} = -q \tag{7.3}$$

Aqui, $T[K]$ é a temperatura, S_w é a saturação da água e S_g é a saturação do vapor de água, o f_w, f_g são os fluxos fracionado de água e gás [7]. O conjunto de parâmetros

junto a seus valores típicos são dados na tabela 7.1.

Tabela 11 – Parâmetros dimensionais para a injeção de vapor num meio poroso saturado de água e seus valores típicos

Símbolo	*Quantidade Física*	*Valor*	*Unidade*
φ	Porosidade	0.3	[.]
ρ_w	Densidade da água	55000	$[mol/m^3]$
T_{res}	Temperatura inicial do reservatório	300	$[K]$
T_b	Temperatura de ebulição da água	373	$[K]$
C_r	Capacidade de calor do meio poroso	2.10^6	$[J/m^3K]$
c_w	Capacidade calorífica molar da água	29	$[J/molK]$
c_g	Capacidade calorífica molar de vapor	75	$[J/molK]$
k_T	Condutividade térmica	0.87	$[J/(msK)]$
u_{inj}	Veloc. de Darcy para injeção de gás ($100m/dia$)	0.0012	$[m/s]$
R	Constante de gás	8.314	$[J/molK]$
P_g	Pressão prevalecente (atmosférica)	10^5	$[Pa]$
k_v	Parâmetro de transferência empírica	1	$[s^{-1}]$
Q_v	Calor de vaporização da água	4.10^4	$[J/mol]$
q_b	Condensação de água	0.56	$[mol/sKm^3]$
ρ_g	Densidade de vapor a 1 atm.	44.6	$[mol/m^3]$
μ_g	Viscosidade do vapor	0.001	$[Pa\ s]$
μ_w	Viscosidade da água	0.1	$[Pa\ s]$

Agora modelaremos a taxa de evaporação, utilizando a relação de Clausius-Clapeyron, [7] e a lei de Raoult's:

$$s_g^{eq} = \exp\left(\frac{-Q_v}{R}\left(\frac{1}{T} - \frac{1}{T_b}\right)\right), \tag{7.4}$$

onde s_g^{eq} é a saturação de vapor de equilíbrio e T_b é o ponto de ebulição normal da água medida à pressão atmosférica. Consideramos a taxa de evaporização dado por, [2].

$$q = k_v \rho_g \left(s_g^{eq}(T) - S_g\right) S_w. \tag{7.5}$$

onde os valores típicos de Q_v, T_b e R são mostrados na Tabela 7.1. As variáveis a serem encontradas são a temperatura T, saturação da água S_w e saturação do gás S_g.

7.2 EQUAÇÕES ADIMENSIONAIS

As equações não estão dimensionadas, mas introduzindo variáveis independentes e dependentes adimensionais (denotada por $\sim$) como razões das quantidades dimensionais e as quantidades referenciais (denotadas por estrela), [23]

$$\tilde{t} = \frac{t}{t^\star}, \quad \tilde{x} = \frac{x}{x^\star}, \quad \theta = \frac{T - T_{res}}{T^\star}, \quad \omega = S_w, \quad \sigma = S_g. \tag{7.6}$$

Em (7.6) ω representa a saturação de água, σ representa a saturação do vapor. A escolha das variáveis adimensionais é comum na engenharia do petróleo. Aqui é usada com a finalidade de nos ajudar a reescrever o sistema (7.1, 7.2 e 7.3) na forma de um problema de complementaridade mista. Nossa escolha para as quantidades referenciais são:

$$T^\star = \frac{Q_v}{R}, \quad t^\star = \frac{\varphi x^\star}{u}, \quad x^\star = \frac{\varphi k_T}{(1-\varphi)C_r u}, \tag{7.7}$$

onde u_{inj} é a velocidade de Darcy da injeção do gás, Q_v é o calor de evaporização de água, K_T é a condutividade térmica. Em 7.7, $t^\star$ é o tempo característico para a injeção do vapor na temperatura inicial do reservatório T_{res}.

Usando (7.6) e (7.7), na equação

$$\frac{\partial}{\partial t}\left[(1-\varphi)C_r T\right] + \frac{\partial}{\partial x}\left[\rho_w u f_w C_w + \rho_g u f_g C_g T\right] \;=\; \nabla(k_T \nabla T) + Q_v q,$$

temos:

$$\left[\frac{(1-\varphi)C_r T^\star}{t^\star}\right]\frac{\partial \theta}{\partial \tilde{t}} \;+\; \frac{\partial}{\partial \tilde{x}}\left[\left(\frac{\rho_w u C_w f_w}{x^\star} \;+\; \frac{\rho_g u C_g f_g}{x^\star}\right)(\theta T^\star \;+\; T_{res})\right] \;=$$
$$= \;\frac{k_T T^\star}{(x^\star)^2}\cdot\frac{\partial^2\theta}{\partial \tilde{x}^2} \;+\; Q_v k_v \rho_g\left[\exp\left(-\frac{Q_v}{R}\left(\frac{1}{\theta T^\star + T_{res}} - \frac{1}{T_b}\right)\right) - \sigma\right]\omega,$$

multiplicando por: $\left[\frac{t^\star}{(1-\varphi)C_r T^\star}\right]$, obtemos:

$$\frac{\partial\theta}{\partial\tilde{t}} + \frac{\partial}{\partial\tilde{x}}\left[\left(\frac{\rho_w C_w \varphi f_w}{(1-\varphi)C_r} - \frac{\rho_g C_g \varphi f_g}{(1-\varphi)C_r}\right)\left(\theta + \frac{T_{res}}{T^\star}\right)\right] \;=\; \frac{k_T t^\star}{(1-\varphi)C_r (x^\star)^2}\frac{\partial^2\theta}{\partial\tilde{x}^2} \;+$$
$$+\;\frac{Q_v k_v \rho_g t^\star}{(1-\varphi)C_r T^\star}\left[\exp\left(\frac{1}{\frac{T_b}{T^\star}} - \frac{1}{\theta + \frac{T_{res}}{T^\star}}\right) - \sigma\right]\omega,$$

fazendo: $\alpha_\omega = \frac{\rho_w C_w \varphi}{(1-\varphi)C_r}$, $\alpha_\sigma = \frac{\rho_g C_g \varphi}{(1-\varphi)C_r}$, $D = \frac{k_T t^\star}{(1-\varphi)C_r (x^\star)^2}$, $\nu_\theta = \frac{Q_v k_v \rho_g t^\star}{(1-\varphi)C_r T^\star}$, $\theta_0 = \frac{T_{res}}{T^\star}$, $\theta_b = \frac{T_b}{T^\star}$, e omitindo as $\sim$, chegamos a:

$$\boxed{\frac{\partial\theta}{\partial t} \;+\; \frac{\partial}{\partial x}\left[(\alpha_\omega f_\omega + \alpha_\sigma f_\sigma)(\theta + \theta_0)\right] \;=\; D\frac{\partial^2\theta}{\partial x^2} \;+\; \nu_\theta \eta(\theta,\omega,\sigma)} \tag{7.8}$$

onde:

$$\eta(\theta,\omega,\sigma) = \left(\exp\left(\frac{1}{\theta_b} - \frac{1}{\theta_0 + \theta}\right) - \sigma\right)\omega,$$
$$f_\omega(\omega,\sigma) = \frac{1}{1 + \mu\sigma^2/\omega^2}, \quad f_\sigma(\omega,\sigma) = \frac{1}{1 + \omega^2/\mu\sigma^2}.$$

Agora usando (7.6) e (7.7) em

$$\frac{\partial(\varphi\rho_w S_w)}{\partial t} \;+\; \frac{\partial(\rho_w u f_w)}{\partial x} \;=\; q,$$

temos

$$\begin{aligned}
\varphi\rho_w\frac{\partial\omega}{\partial\tilde{t}t^\star} + \rho_w u\frac{\partial f_w}{\partial\tilde{x}x^\star} &= k_v\rho_g\left[\exp\left(-\frac{Q_v}{R}\left(\frac{1}{\theta T^\star + T_{res}} - \frac{1}{T_b}\right)\right) - \sigma\right]\omega, \\
\frac{\varphi\rho_w}{t^\star}\frac{\partial\omega}{\partial\tilde{t}} + \frac{\rho_w u}{x^\star}\frac{\partial f_w}{\partial\tilde{x}} &= k_v\rho_g\left[\exp\left(\frac{1}{\frac{T_b}{T^\star}} - \frac{1}{\theta + \frac{T_{res}}{T^\star}}\right) - \sigma\right]\omega,
\end{aligned}$$

multiplicando por: $\frac{t^\star}{\varphi\rho_w}$, obtemos:

$$\frac{\partial\omega}{\partial\tilde{t}} + \frac{ut^\star}{x^\star\varphi}\frac{\partial f_\omega}{\partial\tilde{x}} = k_v\rho_g\frac{t^\star}{\varphi\rho_w}\left[\exp\left(\frac{1}{\theta_b} - \frac{1}{\theta_0+\theta}\right) - \sigma\right]\omega,$$

fazendo $\nu_\omega = \frac{k_v\rho_g t^\star}{\varphi\rho_w}$ e omitindo $\sim$, chegamos a:

$$\boxed{\frac{\partial\omega}{\partial t} + \frac{\partial f_\omega}{\partial x} = \nu_\omega\eta(\theta,\omega,\sigma).} \tag{7.9}$$

Finalmente usando (7.6) e (7.7) em

$$\frac{\partial(\varphi\rho_g S_g)}{\partial t} + \frac{\partial(\rho_g u f_g)}{\partial x} = -q,$$

temos

$$\begin{aligned}
\varphi\rho_g\frac{\partial\sigma}{\partial\tilde{t}t^\star} + \rho_g u\frac{\partial f_g}{\partial\tilde{x}x^\star} &= -k_v\rho_g\left[\exp\left(-\frac{Q_v}{R}\left(\frac{1}{\theta T^\star + T_{res}} - \frac{1}{T_b}\right)\right) - \sigma\right]\omega, \\
\frac{\varphi\rho_g}{t^\star}\frac{\partial\omega}{\partial\tilde{t}} + \frac{\rho_g u}{x^\star}\frac{\partial f_g}{\partial\tilde{x}} &= -k_v\rho_g\left[\exp\left(\frac{1}{\frac{T_b}{T^\star}} - \frac{1}{\theta + \frac{T_{res}}{T^\star}}\right) - \sigma\right]\omega,
\end{aligned}$$

multiplicando por: $\frac{t^\star}{\varphi\rho_g}$, obtemos:

$$\frac{\partial\sigma}{\partial\tilde{t}} + \frac{ut^\star}{x^\star\varphi}\frac{\partial f_\sigma}{\partial\tilde{x}} = -k_v\rho_g\frac{t^\star}{\varphi\rho_g}\left[\exp\left(\frac{1}{\theta_b} - \frac{1}{\theta_0+\theta}\right) - \sigma\right]\omega,$$

fazendo $\nu_\sigma = \frac{k_v t^\star}{\varphi}$ e omitindo $\sim$, chegamos a:

$$\boxed{\frac{\partial\sigma}{\partial t} + \frac{\partial f_\sigma}{\partial x} = -\nu_\sigma\eta(\theta,\omega,\sigma).} \tag{7.10}$$

portanto o sistema 7.1, 7.2 e 7.3 será escrito em sua forma adimensional como:

$$\frac{\partial\theta}{\partial t} + \frac{\partial}{\partial x}\left[\left(\alpha_\omega f_\omega + \alpha_\sigma f_\sigma\right)(\theta+\theta_0)\right] = D\frac{\partial^2\theta}{\partial x^2} + \nu_\theta\eta(\theta,\omega,\sigma) \tag{7.11}$$

$$\frac{\partial\omega}{\partial t} + \frac{\partial f_\omega}{\partial x} = \nu_\omega\eta(\theta,\omega,\sigma) \tag{7.12}$$

$$\frac{\partial\sigma}{\partial t} + \frac{\partial f_\sigma}{\partial x} = -\nu_\sigma\eta(\theta,\omega,\sigma) \tag{7.13}$$

onde $\eta(\theta,\omega,\sigma) = \left[\exp\left(\frac{1}{\theta_b} - \frac{1}{\theta_0+\theta}\right) - \sigma\right]\omega.$

Com as constantes adimensionais:

$$\alpha_\omega = \frac{\rho_w c_w \varphi}{(1-\varphi)C_r}, \quad \alpha_\sigma = \frac{\rho_g c_g \varphi}{(1-\varphi)C_r}, \quad \theta_0 = \frac{T_{res}}{T^\star}, \quad \theta_b = \frac{T_b}{T^\star},$$
$$\nu_\omega = \frac{k_v \rho_g t^\star}{\varphi \rho_w}, \quad \nu_\sigma = \frac{k_v t^\star}{\varphi}, \quad \nu_\theta = \frac{Q_v k_v \rho_g t^\star}{(1-\varphi)C_r T^\star}, \quad D = \frac{k_T t^\star}{(1-\varphi)C_r (x^\star)^2}. \tag{7.14}$$

aqui θ_0 é a temperatura do reservatório, θ_b é a temperatura da ebulição.

O sistemas (7.11 - 7.13) deve ser resolvido com as condições iniciais do reservatório:

$$t = 0, \quad x \geq 0: \quad \theta = 0, \quad \omega = 0, \quad \sigma = 0 \tag{7.15}$$

e as condições de injeção:

$$t \geq 0, \quad x = 0: \quad \theta = 1, \quad \omega = 0, \quad \sigma = 1 \tag{7.16}$$

Podemos transformar o sistema (7.11 - 7.13) como:

$$\theta \geq 0 \ : \ \frac{\partial \theta}{\partial t} + \frac{\partial}{\partial x}\left[\left(\alpha_\omega f_\omega + \alpha_\sigma f_\sigma\right)(\theta + \theta_0)\right] - D\frac{\partial^2 \theta}{\partial x^2} - \Phi(\theta, \omega, \sigma) \geq 0, \tag{7.17}$$

$$\frac{\partial \omega}{\partial t} + \frac{\partial f_\omega}{\partial x} - \nu_\omega \eta(\theta, \omega, \sigma) = 0, \tag{7.18}$$

$$\frac{\partial \sigma}{\partial t} + \frac{\partial f_\sigma}{\partial x} + \nu_\sigma \eta(\theta, \omega, \sigma) = 0, \tag{7.19}$$

$$\left(\frac{\partial \theta}{\partial t} + \frac{\partial}{\partial x}\left[\left(\alpha_\omega f_\omega + \alpha_\sigma f_\sigma\right)(\theta + \theta_0)\right] - D\frac{\partial^2 \theta}{\partial x^2} - \nu_\theta \eta(\theta, \omega, \sigma)\right)\theta = 0. \tag{7.20}$$

onde $\Phi(\theta, \omega, \sigma) = \nu_\theta \eta(\theta, \omega, \sigma)$.

Nós precisamos que a equação (7.20) juntamente com as equações (7.18 e 7.19) sejam satisfeitas no extremo direito do intervalo, onde a equação (7.11) pode não ser satisfeita. Isto explica o por que escolhemos, as variáveis adimensionais θ, ω e σ em (7.6) que descrevem a temperatura, saturação do água e vapor, respectivamente, dentro do reservatório.

7.3 DESCRIÇÃO DOS ESQUEMAS NUMÉRICOS

Agora descrevemos com detalhe o esquema de diferenças finitas para o modelo de injeção de vapor num meio poroso saturado para o método FDA-MNCP. Para isso usamos a malha descrita no Capítulo 3 aplicando o método de Crank-Nicolson para aproximar as

derivadas espaciais em cada caso, isto é:

$$\partial_t\theta(x_m,t_{n+\frac{1}{2}}) = \frac{\theta_m^{n+1}-\theta_m^n}{k}. \tag{7.21}$$

$$\partial_{xx}\theta(x_m,t_{n+\frac{1}{2}}) = \frac{\theta_{m+1}^{n+1}-2\theta_m^{n+1}+\theta_{m-1}^{n+1}}{2h^2}+\frac{\theta_{m+1}^{n}-2\theta_m^{n}+\theta_{m-1}^{n}}{2h^2}. \tag{7.22}$$

$$\partial_x F(\theta(x_m,t_{n+\frac{1}{2}})) = \frac{F_{m+1}^{n+1}-F_{m-1}^{n+1}}{4h}+\frac{F_{m+1}^{n}-F_{m-1}^{n}}{4h}. \tag{7.23}$$

$$\Phi(\theta(x_m,t_{n+\frac{1}{2}})) = \frac{\Phi_m^{n+1}+\Phi_m^n}{2}. \tag{7.24}$$

Nós temos as seguintes condições de fronteira para $m=0$ e $m=M$.

$$\theta(x_0,t)=1, \qquad \Upsilon(x_0,t)=\begin{pmatrix}\omega(x_0,t)\\ \sigma(x_0,t)\end{pmatrix}=\begin{pmatrix}0\\1\end{pmatrix},$$

e

$$\frac{\partial\theta}{\partial x}(x_M,t)=0, \qquad \frac{\partial\Upsilon}{\partial x}(x_M,t)=\begin{pmatrix}\frac{\partial}{\partial x}\omega(x_M,t)\\ \frac{\partial}{\partial x}\sigma(x_M,t)\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix},$$

isto é, temos uma condição de Dirichlet no ponto x_0 e uma condição de Neumann no ponto x_M, logo, o valor em x_0 é conhecido em todo tempo, o que não acontece em x_M, logo:

$$\theta_0^{n+1}=\theta_0^n, \quad \text{e} \quad \Upsilon_0^{n+1}=\Upsilon_0^n, \qquad \forall n\in\mathbb{N}. \tag{7.25}$$

A condição de fronteira em x_M fornece:

$$\frac{\partial\theta}{\partial x}(x_M,t)=0 \Rightarrow \frac{\theta_{M+1}^n-\theta_{M-1}^n}{2h}=0,$$

por tanto

$$\theta_{M+1}^n=\theta_{M-1}^n \Rightarrow F_{M+1}^n=F_{M-1}^n, \qquad \forall n\in\mathbb{N}. \tag{7.26}$$

Fazendo da mesma forma para Υ, obtemos:

$$\Upsilon_{M+1}^n=\Upsilon_{M-1}^n \Rightarrow F_{M+1}^n=F_{M-1}^n, \qquad \forall n\in\mathbb{N}. \tag{7.27}$$

7.3.1 **Método FDA-MNCP**

Vamos aplicar o FDA-MNCP a uma equação parabólica e duas hiperbólicas não lineares. Consideremos θ, $\Upsilon=\begin{pmatrix}\omega\\ \sigma\end{pmatrix}$. Então as relações (7.17) - (7.20) ficam:

$$\theta\geq 0; \quad G(\theta,\Upsilon)=\theta_t+F(\theta,\Upsilon)_x-D\theta_{xx}-\Phi(\theta,\Upsilon)\geq 0. \tag{7.28}$$

$$\theta\bullet G(\theta,\Upsilon)=0. \tag{7.29}$$

$$Q(\theta,\Upsilon)=0. \tag{7.30}$$

onde:

$$Q(\theta,\Upsilon) \;=\; \left[\; \partial_t\Upsilon + \partial_x F(\Upsilon) - \mathcal{Q} \;\right] = \begin{pmatrix} \frac{\partial\omega}{\partial t} + \frac{\partial f_\omega}{\partial x} - \nu_\omega\eta(\theta,\omega,\sigma) \\ \frac{\partial\sigma}{\partial t} + \frac{\partial f_\sigma}{\partial x} + \nu_\sigma\eta(\theta,\omega,\sigma) \end{pmatrix}, \tag{7.31}$$

$$F(\theta,\Upsilon) = (\alpha_\omega f_\omega + \alpha_\sigma f_\sigma)(\theta+\theta_0), \quad D = \frac{k_T t^\star}{(1-\varphi)C_r(x^\star)^2}, \quad \Phi = \nu_\theta\eta(\theta,\omega,\sigma). \tag{7.32}$$

Para obter as relações discretas de (7.28, 7.29, 7.30) sustituímos (7.21-7.24) em (7.28) para obter:

$$\begin{aligned} -2\mu D\theta_{m-1}^{n+1} + (4+4\mu D)\theta_m^{n+1} - 2\mu D\theta_{m+1}^{n+1} + \lambda[F_{m+1}^{n+1} - F_{m-1}^{n+1}] - 2k\Phi_m^{n+1} \;&\geq \\ \geq\; 2\mu D\theta_{m-1}^{n} + (4-4\mu D)\theta_m^{n} + 2\mu D\theta_{m+1}^{n} + \lambda[F_{m+1}^{n} - F_{m-1}^{n}] + 2k\Phi_m^{n},& \end{aligned} \tag{7.33}$$

onde $\lambda = \frac{k}{h}$ e $\mu = \frac{k}{h^2}$. O esquema é válido para $m = 1, 2, ..., M$ dos pontos que não são conhecidos os valores.

Nos pontos de fronteira temos que para $m = 1$, substituindo (7.25) em (7.33) obtemos:

$$\begin{aligned} (4+4\mu D)\theta_1^{n+1} - 2\mu D\theta_2^{n+1} + \lambda[F_2^{n+1} - F_0^{n+1}] - 2k\Phi_1^{n+1} \;&\geq \\ \geq\; (4-4\mu D)\theta_1^{n} + 2\mu D\theta_2^{n} - \lambda[F_2^{n} - F_0^{n}] + 2k\Phi_1^{n} + 4\mu D\theta_0^{n},& \end{aligned} \tag{7.34}$$

para $m = M$, substituindo (7.26) em (7.33) tem-se:

$$-4\mu D\theta_{M-1}^{n+1} + (4+4\mu D)\theta_M^{n+1} - 2k\Phi_M^{n+1} \geq 4\mu D\theta_{M-1}^{n} + (4-4\mu D)\theta_M^{n} + 2k\Phi_M^{n}. \tag{7.35}$$

Então, (7.33) é válido para todo $m = 2, ..., M-1$ e unindo as expressões (7.34) e (7.35) obtemos a seguinte desigualdade na variável θ^{n+1}:

$$G^n(\theta^{n+1},\Upsilon^{n+1}) = A\theta^{n+1} + \lambda P(\theta^{n+1},\Upsilon^{n+1}) - 2k\Phi(\theta^{n+1},\Upsilon^{n+1}) - LD(\theta^n,\Upsilon^n) \geq 0, \tag{7.36}$$

onde $LD = B\theta^n - \lambda P(\theta^n,\Upsilon^n) + 2k\Phi(\theta^n) + UR$ é conhecido em cada instante do tempo. Além disso:

$$A = \begin{bmatrix} 4+4\mu D & -2\mu D & 0 & 0 & \cdots & 0 & 0 & 0 \\ -2\mu D & 4+4\mu D & -2\mu D & 0 & \cdots & 0 & 0 & 0 \\ 0 & -2\mu D & 4+4\mu D & -2\mu D & & 0 & 0 & 0 \\ \vdots & & & & \ddots & & & \vdots \\ 0 & 0 & 0 & 0 & & -2\mu D & 4+4\mu D & -2\mu D \\ 0 & 0 & 0 & 0 & & 0 & -4\mu D & 4+4\mu D \end{bmatrix},$$

$$B = \begin{bmatrix} 4-4\mu D & 2\mu D & 0 & 0 & \cdots & 0 & 0 & 0 \\ 2\mu D & 4-4\mu D & 2\mu D & 0 & \cdots & 0 & 0 & 0 \\ 0 & 2\mu D & 4-4\mu D & 2\mu D & & 0 & 0 & 0 \\ \vdots & & & & \ddots & & & \vdots \\ 0 & 0 & 0 & 0 & & 2\mu D & 4-4\mu D & 2\mu D \\ 0 & 0 & 0 & 0 & & 0 & 4\mu D & 4-4\mu D \end{bmatrix}, \tag{7.37}$$

$$P^n = P(\theta^n) = \begin{bmatrix} F_2^n - F_0^n \\ F_3^n - F_1^n \\ F_4^n - F_2^n \\ \vdots \\ F_M^n - F_{M-2}^n \\ 0 \end{bmatrix}, \quad \Phi^n = \Phi(\theta^n) = \begin{bmatrix} \Phi_1^n \\ \Phi_2^n \\ \Phi_3^n \\ \vdots \\ \Phi_{M-1}^n \\ \Phi_M^n \end{bmatrix}, \tag{7.38}$$

$$\theta^n = \begin{bmatrix} \theta_1^n \\ \theta_2^n \\ \theta_3^n \\ \vdots \\ \theta_{M-1}^n \\ \theta_M^n \end{bmatrix}, \; UR = \begin{bmatrix} 4\mu D\theta_0^n \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \tag{7.39}$$

onde $A, B \in \mathbb{R}^{M\times M}$; $\theta^n, P^n, \Phi^n \in \mathbb{R}^M$.

Da mesma forma como feito anteriormente, para as equações de $Q(\theta, \Upsilon) = 0$, temos que:

$$\frac{\partial \omega}{\partial t} + \frac{\partial f_\omega}{\partial x} = \nu_\omega \eta(\theta, \omega, \sigma), \tag{7.40}$$

$$\frac{\partial \sigma}{\partial t} + \frac{\partial f_\sigma}{\partial x} = -\nu_\sigma \eta(\theta, \omega, \sigma), \tag{7.41}$$

fazendo $\Upsilon = \begin{pmatrix} \omega \\ \sigma \end{pmatrix}$, temos:

$$\Upsilon_t + F(\Upsilon)_x - \mathcal{Q} = 0,$$

então, usando (7.25) e (7.26), obtemos:

$$Diag(4)\Upsilon_m^{n+1} + \lambda[F_{m+1}^{n+1} - F_{m-1}^{n+1}] - 2k\mathcal{Q}_m^{n+1} = Diag(4)\Upsilon_m^n - \lambda[F_{m+1}^n - F_{m-1}^n] + 2k\mathcal{Q}_m^n, \tag{7.42}$$

onde $\lambda = \frac{k}{h}$. O esquema de diferenças é válido para todo $m = 1, 2, ..., M$ dos pontos que não são conhecidos os valores.

Nos pontos da fronteira temos que para $m = 1$, substituindo (7.25) em (7.42) obtemos:

$$4\Upsilon_1^{n+1} + \lambda[F_2^{n+1} - F_0^{n+1}] - 2k\mathcal{Q}_1^{n+1} = 4\Upsilon_1^n - \lambda[F_2^n - F_0^n] + 2k\mathcal{Q}_1^n, \tag{7.43}$$

para $m = M$, substituindo (7.26) em (7.42) tem-se:

$$4\Upsilon_M^{n+1} - 2k\mathcal{Q}_M^{n+1} = 4\Upsilon_M^n + 2k\mathcal{Q}_M^n. \tag{7.44}$$

Portanto, (7.42) é válido para todo $m = 2, ..., M-1$ e unindo as expresões (7.43) e (7.44) obtemos a seguinte desigualdade na variável Υ^{n+1}:

$$Q(\theta^{n+1}, \Upsilon^{n+1}) = Diag(4)\Upsilon^{n+1} + \lambda N(\Upsilon^{n+1}) - 2k\varphi(\theta^{n+1}, \Upsilon^{n+1}) - LDQ, \tag{7.45}$$

onde $LDQ = Diag(4)\Upsilon^n - \lambda N(\Upsilon^n) + 2k\varphi(\theta^n, \Upsilon^n)$ é conhecido em cada instante do tempo. Além disso:

$$Diag(4) = \begin{bmatrix} 4 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & & 0 & 0 & 0 \\ \vdots & & & & \ddots & & & \vdots \\ 0 & 0 & 0 & 0 & & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & & 0 & 0 & 4 \end{bmatrix}, \tag{7.46}$$

$$N^n = N(\theta^n, \Upsilon^n) = \begin{bmatrix} F_2^n - F_0^n \\ F_3^n - F_1^n \\ F_4^n - F_2^n \\ \vdots \\ F_M^n - F_{M-2}^n \\ 0 \end{bmatrix}, \quad \varphi^n = \varphi(\Upsilon^n) = \begin{bmatrix} \varphi_1^n \\ \varphi_2^n \\ \varphi_3^n \\ \vdots \\ \varphi_{M-1}^n \\ \varphi_M^n \end{bmatrix}, \quad \Upsilon^n = \begin{bmatrix} \Upsilon_1^n \\ \Upsilon_2^n \\ \Upsilon_3^n \\ \vdots \\ \Upsilon_{M-1}^n \\ \Upsilon_M^n \end{bmatrix}. \tag{7.47}$$

Portanto, a forma discreta de (7.28) e (7.30) é dada por (7.36) e (7.45). Do mesmo jeito, sustituímos (7.21) em (7.29) para obter sua forma discreta, dado por:

$$G^n(\theta^{n+1}, \Upsilon^{n+1}) \bullet \theta^{n+1} = 0, \tag{7.48}$$

$$Q^n(\theta^{n+1}, \Upsilon^{n+1}) = 0, \tag{7.49}$$

além disso, deve-se cumprir:

$$\theta^{n+1} \geq 0. \tag{7.50}$$

Assim, juntando (7.36), (7.45) (7.48) e (7.50) formam um Problema de Complementaridade Mista que pode ser resolvido pelo algoritmo FDA-MNCP desenvolvido no Capítulo (4). O que for feito até agora será resumido no seguinte Algoritmo:

Algoritmo 7.1 (Implementação FDA-MNCP.)

Passo 1. *Fazer $n = 0$ e $N = 1/\Delta t$.*

Passo 2. *Para obter θ^{n+1} e Υ^{n+1} aplicamos o método $FDA - MNCP$ descrito no algoritmo para resolver o problema de complementaridade mista.*

$$G^n(\theta^{n+1}, \Upsilon^{n+1}) \bullet \theta^{n+1} = 0, \quad \theta^{n+1} \geq 0, Q^n(\theta^{n+1}, \Upsilon^{n+1}) = 0 \quad e$$

$$G^n(\theta^{n+1}, \Upsilon^{n+1}) = A\theta^{n+1} + \lambda P(\theta^{n+1}, \Upsilon^{n+1}) - 2k\Phi(\theta^{n+1}, \Upsilon^{n+1}) - LD(\theta^n, \Upsilon^{n+1}) \geq 0,$$
$$Q(\theta^{n+1}, \Upsilon^{n+1}) = Diag(4) \cdot \Upsilon^{n+1} + \lambda N(\Upsilon^{n+1}) + 2k\varphi(\theta^{n+1}, \Upsilon^{n+1}) - LDQ(\theta^n, \Upsilon^{n+1}) = 0.$$

onde as matrizes A, B e os vetores P, N e UR são dadas em (7.37), (7.37), (7.38), e (7.47).

Passo 3. *Se $n = N$ então **FIM.** Se não $n = n + 1$ e voltar ao Passo **2***

REFERÊNCIAS

[1] BIEZUNER, R. J. *Notas de Aula: Métodos Numéricos para Equações Diferenciais Parciais Elípticas.* UFMG, Minas Gerais: Instituto de Ciências Exatas, 2007.

[2] BRUINING J., MARCHESIN D., and VAN DUIJN C.J. *Steam injection into water-saturated porous rock.* Computational e Applied Mathematics, 22(3):359-395, 2003.

[3] BURDEN, R. L.; FAIRES, J. D. Numerical Analysis. Ninth edition. Boston, USA: Brooks/Cole - CENGAGE Learning, 2010.

[4] CHAPIRO, G. ; MAZORCHE, S. R. ; HERSKOVITS, J. ; ROCHE, J. R. . Solution of the Non-linear Parabolic Problems using Nonlinear Complementarity Algorithm (FDA-NCP). Mecánica Computacional, v. XXIX, p. 2141-2153, 2010.

[5] CHAPIRO, G. *Gas-Solid Combustion in Insulated Porous Media.* Tese (Doutorado em Matemática) - Instituto Nacional de Matemática Pura e Aplicada, Rio de Janeiro, Brasil, 2009.

[6] CUMINATO J. A. ; MENEGUETTE J. M., *Discretização de Equações Diferenciais Parciais. Técnicas de Diferenças Finitas* Rio de Janeiro, Brasil: Coleção Matemática Aplicada, SBM, 2013.

[7] DAKE L.P. *Fundamentals of reservoir engineering.* Elsevier, 1983.

[8] DENNIS JR J.E. and SCHNABEL R.B. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall, 1983.

[9] DÉBORA DE JESUS BEZERRA *Métodos Numéricos para Leis de Conservação.* (Tese Mestrado em Ciências). Universidade de São Paulo - ICMC-USP, São Paulo, Brasil, 2003.

[10] DUIJN, C. J. V. *An Introduction to Conservation Laws: Theory and Applications to Multi-Phase Flow.* Eindhoven University of Technology, The Netherlands, 2003.

[11] EL-BAKRY A.S. ,TAPIA R.A. , Y. ZHANG, and T. TSUCHIYA. *On the formulation and theory of the newton interior-point method for nonlinear programming.* Journal of Optimization Theory and Applications, 89(3):507-541, 1996.

[12] FERRIS, M. C.; PANG, J. S. *Engineering and economic applications of complementarity problems.* Society for Industrial and Applied Mathematics, SIAM, v. 39, n. 4, 1997.

[13] HERSKOVITS J., LEONTIEV A., DIAS G. and SANTOS G. *Contact Shape Optimization: A Bilevel Programming Approach*, Structural and Multidisciplinary Optimization 20, pp. 214-221.

[14] HERSKOVITS J., MAZORCHE S. *A feasible directions algorithm for nonlinear complementarity problems and applications in mechanics*, Structural and Multidisciplinary Otimization, Feb 2009, vol 37, pp. 435-446.

[15] LAWRENCE C. E. *Partial Differential Equations.* American Mathematical Society, 1997.

[16] LEONTIEV A., HUACASI W. and HERSKOVITS J. *An Optimization Technique For The Solution Of The Signorini Problem Using The Boundary Element Method, Structural and Multidisciplinary Optimization.*, 24, pp. 72-77.

[17] LEVEQUE, R. J. *Numerical methods for conservation laws.* Lectures in mathematics: ETH Zürich, 1992.

[18] LIMA, E. L. *Curso de Análise vol.2.* 11ava. ed. IMPA, Rio de Janeiro: Projeto Euclides, 2014

[19] MAYERS, D.; MORTON, K. *Numerical Solution of Partial Differential Equations.* Cambridge, UK.: Cambridge University Press, 2005.

[20] MAZORCHE, S. R. *Algoritmos para problemas de complementaridade não linear.* Tese (Doutorado em Engenharia Mecânica). Universidade Federal de Rio de Janeiro, Rio de Janeiro, Brasil, 2007.

[21] OLEINIK, O. *Discontinuous Solutions of Nonlinear Differential Equations.* Amer. Math. Soc. Trans. Ser. 2, vol. 26, pp. 95-172, 1957.

[22] PORTO E. M. *Um Novo Aproximante de Riemann para Simulações de Escoamentos com Ondas de Choque.* (Tese de Doutorado em Mecânica). Universidade Federal Fluminense, Rio de Janeiro, Brasil, 2009.

[23] RAMIREZ GUTIERREZ, ANGEL *Aplicação do método de complementaridade não linear para o estudo de combustão de oxigênio in situ* (Tese Mestrado em Matemática). Universidade Federal de Juis de Fora, Juiz de Fora, Brasil, 2013.

[24] SERRE, D. *Systems of Conservation Laws 1. Hyperbolicity, Entropies, Shock Waves.* Cambridge University Press, 1999.

[25] SMOLLER, J. *Shock Waves and Reaction-Diffusion Equations.* 2da. ed. Springer Verlag, p.327, 1994.

[26] STRIKWERDA, J. C. *Finite Difference Schemes and Partial Differential Equations.* Philadelphia, EE. UU.: Society for Industrial and Applied Mathematics, SIAM., 2004.

[27] TANNEHILL, J. C.; ANDERSON, D. A.; PLETCHER, R. H. *Computational Fluid Mechanics and Heat Transfer.* W. J. Minkowycz and E. M. Sparrow, Editors, 1997.

[28] THOMAS, J. *Numerical Partial Differencial Equations. Conservation Laws and Elliptic Equations.* Springer-Verlag New York, Inc., 1999.

APÊNDICE A – Definições e resultados gerais

Neste capítulo iremos a presentar alguns conceitos que foram utilizados ao longo deste trabalho. apresentaremos definicões e resultados que ajudaram a formar uma base de conhecimento para tratarmos do problema em questão que é elaboração do algoritmo $FDA-MNCP$ bem como seus resultados de convergência.

A.1 CONCEITOS BÁSICOS PARA PROBLEMAS DE OTIMIZAÇÃO

Seja, um problema de otmização não linear:

$$\begin{aligned} &\min f(x) \\ \text{S.a: } &g_i(x) \geq 0; \quad i = 1, 2, ..., m \\ &h_j(x) = 0; \quad j = 1, 2, ..., p \end{aligned} \tag{A.1}$$

onde $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, $g : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ e $h : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ são funções suaves em $\mathbb{R}^n$ e pelo menos umas destas funções é não linear. As restrições de desigualdade são chamadas **ativas** se $g_i(x) = 0$ e é chamada **inativa** se $g_i < 0$. Denotemos por:

$$g(x) = \begin{bmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_m(x) \end{bmatrix}, \quad h(x) = \begin{bmatrix} h_1(x) \\ h_2(x) \\ \vdots \\ h_m(x) \end{bmatrix}$$

assim temos:

$$\begin{aligned} &\min f(x) \\ \text{S.a: } &g(x) \geq 0 \\ &h(x) = 0. \end{aligned} \tag{A.2}$$

Veamos algumas definições para o problema de otimização não linear A.2

Definição A.1 *Um ponto $x^\star \in \Omega \subset \mathbb{R}^n$ é chamado **Mínimo Local (ou Mínimo Relativo)** de f sobre Ω se existe uma vizinhança $V \equiv \{x \in \Omega / \parallel x - x^\star \parallel \leq \delta\}$ tal que $f(x) \geq f(x^\star)$ para qualquer $x \in V$. Se $f(x) > f(x^\star)$ para todo $x \in V$, $x \neq x^\star$, então dizemos que $x^\star$ é um **Mínimo Local Estrito.***

Definição A.2 *Um ponto $x^\star \in \Omega \subset \mathbb{R}^n$ é chamado **Mínimo Global (ou Mínimo Absoluto)** de f sobre Ω se $f(x) > f(x^\star)$ para qualquer $x \in \Omega$. Se $f(x) > f(x^\star)$ para todo $x \in \Omega \setminus \{x^\star\}$, então dizemos que $x^\star$ é um **Mínimo Global Estrito.***

Geralmente os métodos de Programação Não-Lineares são (PNL) iterativos. Dado um ponto inicial x^0, uma sequência de pontos, $\{x^k\}$, é obtida por repetidas aplicações de uma regra

algoritmica, onde dita sequência debe convergir a uma solução $x^\star$ do problema. A convergência é dita assintótica quando a solução não é atingida antes de um número finito de iterações.

Definição A.3 *Um algoritmo iterativo é dito ser* ***globalmente convergente*** *se para qualquer ponto inicial* $x^0 \in \mathbb{R}^n (ou x^0 \in \Omega)$ *este gera uma sequência de pontos que converge a uma solução do problema.*

Definição A.4 *Um algoritmo iterativo é dito ser* ***localmente convergente*** *se existe* $\varepsilon > 0$ *tal que para qualquer ponto inicial* $x^0 \in \mathbb{R}^n (ou x^0 \in \Omega)$ *que verifica* $\| x^0 - x^\star \| \leq \varepsilon$, *a sequência gerada de pontos converge para uma solução do problema.*

Agora segue uma definiçao que introduz um critério para medir a velocidade de convergência de métodos iterativos com convergência assintótica.

Definição A.5 *A ordem de convergência de uma sequência* $\{x^k\} \longrightarrow x^\star$ *é o maior numero* p *dos números não negativos* ρ *que satisfzem:*

$$\limsup_{k \longrightarrow \infty} \frac{\| x^{k+1} - x^\star \|}{\| x^k - x^\star \|^\rho} = \beta < \infty$$

Quando $p = 1$ *nos dizemos que a convergência é* $q-$*linear com raio de convergência* $\beta < 1$. *Se* $\beta = 0$ *a convergência é dita Superlinear. A convergência é Quadrática quando* $p = 2$.

A relação entre o limite de $k \longrightarrow 0$, p e β é uma forma de medir a velocidade assintótica da convergência. Uma sequência pode ter uma boa ordem de convergência mas ir muito **"devagar"** para a solução. A convergência é rápida quando p é grande e β é pequeno. Para as provas de convergência, se usara as seguintes notações $O(.)$ e $o(.)$ definidas acontinuação:

Dado uma função $h : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, usamos a expressão $O(h(x))$ para representar as funções $g : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ que satisfazem

$$\lim_{\|x\| \longrightarrow 0} \frac{\| g(x) \|}{\| h(x) \|} < \infty. \tag{A.3}$$

E a expressão $o(h(x))$ para representar as funções $g : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ que satisfazem

$$\lim_{\|x\| \longrightarrow 0} \frac{\| g(x) \|}{\| h(x) \|} = 0. \tag{A.4}$$

A geração de pontos de uma sequência é dada por uma regra algoritmica. Por exemplo, dado um ponto inicial, determinamos uma direção de busca para determinar o próximo ponto e assim sucessivamente. Para isso veremos algumas definições sobre direções.

Definição A.6 *Um vetor* $d \in \mathbb{R}^n$ *é uma* ***direção de descida*** *para uma função real* f *em* $x \in \mathbb{R}^n$ *se existe um* $\delta > 0$ *tal que* $f(x + td) < f(x)$ *para qualquer* $t \in (0, \delta)$.

No caso de f ser diferenciável em x e $d^t \nabla f(x) < 0$ então d é uma direção de descida para f em x.

Definição A.7 *Um vetor $d \in \mathbb{R}^n$ é uma* ***direção viável*** *de um problema PNL A.2, em $x \in \Omega$- $= \{x \in \mathbb{R}^n / g(x) \leq 0\}$, se para algum $\theta > 0$ nos temos $x + td \in \Omega$ para todo $t \in [0, \theta]$.*

É claro que qualquer direçaõ d num ponto interior a Ω é uma direção viável.

Definição A.8 *Um campo vetorial $d(.)$ definido em Ω é um* ***campo uniforme de direções de viáveis*** *do problema PNL A.2, em $\Omega = \{x \in \mathbb{R}^n / g(x) \leq 0\}$, se existe um $\tau > 0$ tal que $x + td(x) \in \Omega$ para todo $t \in [0, \tau]$ e para todo $x \in \Omega$.*

A.2 BUSCA LINEARES INEXATAS

Uma vez definida a direção de busca podemos considerar critérios para determinar o próximo ponto por meio das buscas lineares inexatas. Apresentaremos dois tipos de buscas, Armijo e Wolfe.

Definição A.9 (Busca de Armijo) *Dada uma função f, definimos o tamanho do passo t como sendo o primeiro número inteiro da sequência $\{1, \nu, \nu^2, \nu^3, ...\}$ satisfazendo*

$$f(x + td) \leq f(x) + t\eta \nabla f(x)^T d,$$

onde $\eta \in (0, 1)$ e $\nu \in (0, 1)$ são parâmetros dados.

O critêrio da busca linear inexata de Wolfe tambêm estabelece limites no tamanho do passo, pedindo uma redução na função f e ao mesmo tempo uma redução em sua derivada direcional.

Definição A.10 (Busca de Wolfe) *Dada uma função f, o passo t é aceito se verificar: sendo o primeiro número inteiro da sequência $\{1, \nu, \nu^2, \nu^3, ...\}$, satisfazendo*

$$f(x + td) \leq f(x) + t\eta_1 \nabla f^T(x) d,$$

onde $\nu \in (0, 1)$, $\eta_1 \in (0, 1/2)$, $\eta_2 \in (\eta_1, 1)$ são parâmetros dados e verificando

$$\nabla f^T(x + td) d \geq \eta_2 \nabla f^T(x) d.$$

A.3 O MÉTODO DE NEWTON PARA EQUAÇÕES

O método de Newton clássico é introduzido para resolver sistemas não lineares do tipo

$$T(z) = 0, \tag{A.5}$$

onde $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ é diferenciável. Este método é muito utilizado em algoritmos de Programação Matemática. Para o método de Newton clássico é comum pedir as seguintes suposições:

Suposição A.1 *Existe* $z^\star \in \mathbb{R}^n$ *tal que* $T(z^\star) = 0$.

Suposição A.2 *A matri jacobiana* $T'(z^\star)$ *é nao singular.*

Suposição A.3 *O operador jacobiano* T' *é continuo e localmente Lipschitz em* $x^\star$

Assim, determinar uma sequência $\{z^k\}$ que se aproxima de alguma solução $z^\star$ do sistema (A.5) utilizamos a aproximação linear

$$T(z^k) + T'(z^k)(z - z^k) = 0$$

A relação acima é chamada a equação de iteração do *método de Newton.*
O método de Newton pode ser escrito em forma do esquema iterativo

$$z^{k+1} = z^k - (T'(z^k))^{-1}T(z^k), \quad k = 0, 1, 2, ... \tag{A.6}$$

Desta forma é de se esperar que a sequência de pontos gerada acima se aproxima de uma solução do sistema (A.5). E quando isso ocorre podemos esperar uma convergência rápida. É o diz no próximo resultado. Iremos assumir que $T(z)$ é diferenciável e as suposições A.1 - A.3.

Teorema A.1 *Dado* $z^0 \in \mathbb{R}^n$ *suficientemente próximo de* $z^\star$, *o Algoritmo definido por (A.6) gera uma sequência* z^k *bem definida que converge a* $z^\star$. *A taxa de convergência é quadrática.*

A demonstração deste resultado pode ser vista em [8] e [11]. Portanto este teorema estabelece a convergência local do método de Newton. Dos vários métodos existentes para resolver numericamente problemas de Programação Matemática, muitos utilizam ou se baseiam na iteração de Newton para construir a sequência de pontos que converge a uma solução destes problemas.