SISTEMAS NÃO-LINEARES DA FÍSICA E DA ENGENHARÍA

Este exemplar corresponde à redação final da tese devidamente corrigida e defendida pelo Sr. DANIEL NORBERTO KOZAKEVICH e aprovada pela Comissão Julgadora.

Campinas, 20 de Junho de 1995

Prof. Dr. .....

José Mario Martínez

Tese apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para a obtenção do título de DOUTOR em MATEMÁTICA APLICADA.

Tese defendida e aprovada em, 20 de junho de 1995

Pela Banca Examinadora composta pelos Profs. Drs.

Prof(a). Dr(a). JOSE MARIO MARTINEZ PEREZ

Prof(a). Dr(a). CLOVIS RAIMUNDO MALISKA

Prof(a). Dr(a). JOSE ALBERTO CUMINATO

Prof(a). Dr(a). MURILO FRANCISCO THOME

Prof(a). Dr(a). MARIO CESAR ZAMBALDI

Departamento de Matemática Aplicada
Instituto de Matemática, Estatística e Ciência da Computação
Universidade Estadual de Campinas

Sistemas Não Lineares da Física e da Engenharia

Daniel Norberto Kozakevich[1]
Julho 1995

Dissertação submetida ao Departamento de Matemática Aplicada
Universidade Estadual de Campinas, como requisito parcial para
a obtenção do título de Doutor em Matemática Aplicada

---

[1]

Dr. José Mario Martínez, DMA–IMECC–UNICAMP
(Orientador)

Dr. Clóvis R. Maliska, FEM–UFSC

Dr. José A. Cuminato, DM, USP–S.Carlos

Dr. Murilo F. Thomé, DM, USP–S.Carlos

Mário C. Zambaldi, DM–CFM–UFSC

1° Suplente: Márcia A. Gomes-Ruggiero, DMA–IMECC–UNICAMP

2° Suplente: Vera L. Rocha Lopes, DMA–IMECC–UNICAMP

3° Suplente: José V. Zago, DMA–IMECC–UNICAMP

4° Suplente: Álvaro De Pierro, DMA–IMECC–UNICAMP

Prefácio

Esta tese contém contribuições teóricas e práticas no campo da resolução de sistemas algébricos não lineares de grande porte. Esse tipo de sistemas aparece com muita frequencia em aplicações de engenharia e física, portanto, é nesse tipo de problemas que nos concentramos.

Nosso aporte comprende quatro áreas:

- A comparação controlada, do ponto de vista computacional, dos métodos de Newton, Newton modificado, Broyden e Column-Updating, com e sem estratégias de globalização, em um conjunto de problemas originados na discretização de equações diferenciais parciais. Procuramos aqui identificar situações problemáticas e fornecer um panorama claro sobre o que é de se esperar de algoritmos mais ou menos clássicos para resolver problemas com variados graus de dificuldade.

- A análise e resolução exaustiva do "problema da cavidade", para altos números de Reynolds, descartando as estratégias de globalização por otimização (de pobre desempenho neste caso) e reivindicando táticas homotopicas muito simples. O desempenho de alguns métodos quase-Newton, neste caso, é muito bom.

- A introdução de um método novo do tipo Newton-inexato, com uma variação que permite uma resolução eficiente de problemas de autovalores não lineares. Esses problemas são, por direito próprio, sistemas não lineares mas, ao mesmo tempo, refletem com bastante fidelidade o grau de dificuldade que pode ser encontrada em outros sistemas dependentes de um parâmetro.

- A resolução de um problema de evolução (petróleo) onde em cada nível temporal deve ser resolvido um sistema não linear. Neste caso, métodos quase-Newton com Jacobiano inicial escolhido como fatoração incompleta provaram ser notavelmente eficientes.

Agradecimentos

Eu gostaria de agradecer:

Ao meu Orientador pela paciência e motivação infindáveis.

Aos Professores do Grupo de Otimização pelo espírito de solidariedade e a boa disposição de construir.

A Sandra e Mário pelo convívio e inestimável colaboração na realização deste trabalho.

Aos Colegas do Depto. de Matemática, CFM–UFSC pela postura de investir e pela confiança depositada.

Aos Colegas da Pós-Graduação, Professores e Funcionários do IMECC pelo excelente ambiente de trabalho criado.

As autoridades do IMECC–UNICAMP pelo suporte oferecido.

A FAPESP, CAPES e CNPq pelo apoio financiero.

A Gaby, Ale e Conce pela compreensão e por compartilharem as dificuldades e alegrias passadas.

... e a todos os que acham ter-me ajudado, sinceramente.

Dedico este trabalho à minha família.

Conteúdo

...

Lista de Tabelas

Lista de Figuras

Capítulo 1

Introdução

Neste trabalho analisaremos o desempenho de um conjunto de algoritmos para resolver sistemas não lineares originados em problemas reais. Selecionamos para isso, diversos problemas da Física e da Engenharia e vários algoritmos cujas implementações computacionais se encontram em diferentes estágios de experimentação.

Descreveremos a seguir, em forma sintética, alguns dos métodos para a resolução de sistemas não lineares de equações, que serão usados nos capítulos posteriores.

1.1 Algoritmos para a Resolução de Sistemas Não Lineares

Dada $F : \mathbb{R}^n \to \mathbb{R}^n$, $F = (f_1, \ldots, f_n)^T$, desejamos achar a solução de

$$F(x) = 0. \tag{1.1}$$

Suporemos que F está bem definida e tem derivadas parciais contínuas em um conjunto aberto de $\mathbb{R}^n$; denotamos com $J(x)$ a matriz das derivadas parciais de F (matriz Jacobiana). Assim

$$J(x) \equiv F'(x) \equiv \begin{bmatrix} f_1'(x) \\ \vdots \\ f_n'(x) \end{bmatrix} \equiv \begin{bmatrix} \nabla f_1(x)^T \\ \vdots \\ \nabla f_n(x)^T \end{bmatrix} \equiv \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1}(x) & \cdots & \frac{\partial f_n}{\partial x_n}(x) \end{bmatrix}.$$

Será de nosso principal interesse, o estudo de problemas de grande porte onde n é grande e $J(x)$ é *estruturalmente esparsa*, o que significa que a maioria dos coeficientes de $J(x)$ são zero para todo x no domínio de F. Esparsidade é um caso particular do conceito mais geral de *estrutura*. As matrizes Jacobianas podem ser simétricas, antisimétricas, positivas definidas, combinação de matrizes com estruturas especiais, etc.. Usualmente é aproveitada a estrutura

particular de $J(x)$ com a finalidade de melhorar as características computacionais do algoritmo usado para resolver (1.1).

Os métodos ordinários para resolver sistemas não lineares são *locais*. Um método local é um procedimento iterativo que converge, se a aproximação inicial está suficientemente perto da solução. Uma caracterização qualitativa do algoritmo é dada pela *taxa de convergência* que indica a velocidade de aproximação assintótica à solução. Na maioria dos casos o domínio de convergência destes métodos é grande, e por este motivo são assíduamente usados. Porém, quando a aproximação inicial não for suficientemente boa, os métodos locais devem ser modificados para incorporar propriedades de *convergência global*.

Disemos que um método para resolver (1.1) é *globalmente convergente* se, ao menos, um ponto limite da sequência gerada pelo método é a solução ou, no mínimo, um *ponto estacionário* onde, $\nabla\|F(x)\|^2 = 0$. A maioria das vezes *todos* os pontos limites são soluções ou pontos estacionários e frequentemente a sequência converge completamente à solução. Em geral, os métodos globais são modificações de métodos locais que tentam preservar as propriedades de convergência do método local original.

1.1.1 O Método de Newton

O Método de Newton é costumeiramente usado para resolver (1.1). Dada uma estimativa da solução como ponto inicial x^0, o método considera a cada iteração a aproximação

$$F(x) \approx L_k(x) \equiv F(x^k) + J(x^k)(x - x^k) \tag{1.2}$$

e calcula x^{k+1} como a solução do sistema linear $L_k(x) = 0$. Assim, uma iteração do método de Newton pode ser descrita por

$$J(x^k)s^k = -F(x^k), \tag{1.3}$$

$$x^{k+1} = x^k + s^k. \tag{1.4}$$

A cada iteração de Newton devemos avaliar o Jacobiano $J(x^k)$ e resolver o sistema linear (1.3). Usando técnicas de diferenciação automática (ver Rall [62] e [63], Griewank [26], etc) é possível calcular $F(x)$ e $J(x)$ de uma forma confiável e com baixo custo computacional.

Se n não for excessivamente grande consegue-se resolver (1.3) usando a fatoração LU com pivotamento parcial ou com a fatoração QR (ver Golub and Van Loan [22]). O custo destes métodos é da ordem de n^3 operações em artitmética de ponto flutuante. Vários algoritmos para fatorações esparsas estão compilados em Duff, Erisman and Reid [16].

Gomes-Ruggiero, Martínez e Moretti [25] descreveram uma primeira versão do pacote computacional Rouxinol onde estão implementados diversos algoritmos para resolver sistemas não lineares esparsos. Os sistemas lineares são resolvidos com a metodologia de George e Ng [21].

O sistema (1.1) tem uma única solução se e somente se $J(x^k)$ é não singular. Um Jacobiano quase singular ou um sistema linear mal condicionado usualmente causam grandes incrementos s^k; logo, a grandeza de $\|s^k\|$ deve ser controlada. O tamanho do passo é comumente normalizado com

$$s^k \leftarrow \min\{1, \frac{\Delta}{\|s^k\|}\}s^k.$$

onde Δ é um parâmetro dado pelo usuário.

O principal resultado relativo à convergência ao método de Newton está dado no seguinte teorema

Teorema 1 *Suponhamos que $F : \Omega \subset \mathbb{R}^n \to \mathbb{R}^n$; Ω um conjunto aberto e conexo; $F \in C^1(\Omega)$, $F(x^*) = 0$, $J(x^*)$ não singular, e que existam $L, p > 0$ tal que para todo $x \in \Omega$*

$$\|J(x) - J(x^*)\| \leq L\|x - x^*\|^p. \tag{1.5}$$

Então existe $\varepsilon > 0$ tal que se $\|x^0 - x^*\| \leq \varepsilon$, a sequência $\{x^k\}$ gerada por (1.3)-(1.4) está bem definida e converge a x^*, e satisfaz

$$\|x^{k+1} - x^*\| \leq c\|x^k - x^*\|^{p+1}. \tag{1.6}$$

(Prova: Ver Ortega e Rheinboldt [61], Dennis e Schnabel [13], etc.). □

A consecução da convergência quadrática ($p = 1$) dependerá da satisfação da condição de Holder (1.5), sem a qual, pode ser provada apenas convergência superlinear para $\{x^k\}$.

1.1.2 Métodos Quase-Newton

Denominamos Métodos Quase-Newton a aqueles que resolvem (1.1) com uma fórmula do tipo

$$x^{k+1} = x^k - B_k^{-1}F(x^k). \tag{1.7}$$

Os métodos Quase-Newton caracterizam-se por evitar o cálculo das derivadas e a necessidade de resolver integralmente os sistemas lineares a cada iteração. Em consequência, o custo de cada iteração diminui sendo que há uma leve perda das propriedades de convergência em relação ao método de Newton.

Uma modificação destes métodos é realizada introduzindo recomeços. Isto significa que $B_k = J(x^k)$ se k é um múltiplo de um inteiro m ou se não há um decréscimo suficiente de $\|F(x^k)\|$; B_k é obtida a partir de B_{k-1} nos outros casos.

O *Método de Newton Estacionário* é o mais simples dos métodos Quase-Newton, onde $B_k = J(x^0)$ para todo $k \in \mathbb{N}$. Neste método as derivadas são avaliadas no ponto inicial sendo necessária somente uma fatoração LU de $J(x^0)$. Há uma paulatina piora nos métodos Newton estacionários, já que, exceto quando $k \equiv 0 \pmod{m}$, B_k não incorpora informação de x^k e $F(x^k)$. Logo, a semelhança do modelo $L_k(x) \equiv F(x^k) + B_k(x - x^k)$ com $F(x)$ pode diminuir com k. Observamos que por (1.7), nos métodos Quase-Newton, x^{k+1} se define como a solução de $L_k(x) = 0$, que existe e é unica se B_k é não singular. Uma maneira de incorporar informação vinda de F sobre o modelo linear consiste em impor condições interpolatórias.

$$L_{k+1}(x^k) = F(x^k), \tag{1.8}$$

$$L_{k+1}(x^{k+1}) = F(x^{k+1}). \tag{1.9}$$

Definindo

$$y^k = F(x^{k+1}) - F(x^k) \tag{1.10}$$

e subtraindo (1.8) de (1.9) obtemos a *Equação Secante*

$$B_{k+1}s^k = y^k. \tag{1.11}$$

Recíprocamente, se B_{k+1} satisfaz (1.11), L_{k+1} interpola F em x^k e x^{k+1}. Designamos *Métodos Secantes* à familia dos métodos baseados em (1.7) e (1.11).

Se $n \geq 2$, existem infinitas possibilidades para escolher B_{k+1} de modo a satisfazer (1.11). Esta versatilidade permite através de uma escolha apropriada, garantir estabilidade numérica. O Método de Broyden "bom" (Primeiro Método de Broyden, [4] e o Método de Atualização da Coluna (COLUM) (Martínez [42]) se aproveitam desta possibilidade. Em ambos métodos

$$B_{k+1} = B_k + \frac{(y^k - B_k s^k)(z^k)^T}{(z^k)^T s^k} \tag{1.12}$$

onde

$$z^k = s^k \tag{1.13}$$

para o método de Broyden, e

$$z^k = e^{j_k}, \tag{1.14}$$

$$|(e^{j_k})^T s^k| = \|s^k\|_\infty \tag{1.15}$$

para COLUM onde $\{e^1, \ldots, e^n\}$ é a base canonica $\mathbb{R}^n$.

Aplicando a fórmula de Sherman-Morrison a (1.12) (Golub and Van Loan [22]) obtemos

$$B_{k+1}^{-1} = B_k^{-1} + \frac{(s^k - B_k^{-1}y^k)(z^k)^T}{(z^k)^T B_k^{-1} y^k} B_k^{-1}. \tag{1.16}$$

Observamos que

$$B_{k+1}^{-1} = (I + u^k(z^k)^T)B_k^{-1}, \tag{1.17}$$

onde $u^k = (s^k - B_k^{-1}y^k)/(z^k)^T B_k^{-1}y^k$, e assim

$$B_k^{-1} = (I + u^{k-1}(z^{k-1})^T)\ldots(I + u^0(z^0)^T)B_0^{-1}, \tag{1.18}$$

para $k = 1, 2, 3\ldots$ Se n for grande a fórmula (1.18) é utilizada.

O Método de Broyden é um caso particular da familia dos Métodos de Atualização Secante com Variação Mínima (Dennis e Schnabel [12],[13], Dennis e Walker [14], Martínez [48], [50]), que inclue vários algoritmos que são úteis para problemas com estrutura particular (ver Hart e Soul [32], Kelley e Sachs [35]), para problemas separáveis com métodos Quase-Newton Particionados (Griewank e Toint [27], [28], [29], [30], Toint [66]), métodos com Atualização Direta na Fatorização (Dennis and Marwil [10], Johnson e Austria [34], Chadee [6], Martínez [47]), algoritmos do tipo BFGS e DFP para minimização irrestrita (ver Dennis e Schnabel [13]), etc.

Os principais resultados sobre a convergência dos algoritmos Quase-Newton são enunciados a seguir:

Assumimos que como no Teorema 1, $F : \Omega \subset \mathbb{R}^n \to \mathbb{R}^n$, Ω é aberto e convexo, $F \in C^1(\Omega)$, $F(x^*) = 0$, $J(x^*)$ é não-singular e que a condição de Holder é satisfeita.

Teorema 2 *Dado $r \in (0, 1)$, existem $\varepsilon, \delta > 0$ tal que se $\|x^0 - x^*\| \leq \varepsilon$ e $\|B_k - J(x^*)\| \leq \delta$ para todo $k \in \mathbb{N}$ então a sequência $\{x^k\}$ gerada por (1.7) está bem definida, converge a x^*, e satisfaz*

$$\|x^{k+1} - x^*\| \leq r\|x^k - x^*\| \tag{1.19}$$

para todo $k \in \mathbb{N}$.

(Prova: ver por exemplo, Dennis e Walker [14].) □.

Usando o teorema anterior podemos provar que o Método de Newton Estacionário com recomeços tem convergência local, com taxa linear.

A ferramenta fundamental para provar convergência superlinear para os métodos Quasi-Newton é o teorema seguinte, devido a Dennis e Moré.

Teorema 3 *Assumamos que $\{x^k\}$ gerada por (1.7) está bem definida e converge para x^*. Então as duas seguintes propriedades são equivalentes*

$$\text{(a)} \quad \lim_{k\to\infty} \frac{\|[B_k - J(x^*)](x^{k+1} - x^k)\|}{\|x^{k+1} - x^k\|} = 0 \tag{1.20}$$

e

$$\text{(b)} \quad \lim_{k\to\infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} = 0. \tag{1.21}$$

(Prova: ver Dennis e Moré [11].) □

A equação (1.20) é chamada a condição de Dennis-Moré. Usando (1.20), podemos provar que o Método de Newton Estacionário com recomeços periódicos (para o qual $\lim_{k\to\infty} B_k = J(x^*)$) tem convergência superlinear. A condição de Dennis-Moré está relacionada com a Equação Secante e permite obter o seguinte resultado (Broyden, Dennis e Moré [5]).

Lema 1 *Se a sequência gerada por um Método Secante converge a x^* e além disso,*

$$\lim_{k\to\infty} \|B_{k+1} - B_k\| = 0. \tag{1.22}$$

então a convergência é superlinear.

□

O Teorema 3 não garante a convergência local de todos os Método Secantes. Em verdade, a hipótese deste teorema requer que *todas as* B_k devam pertencer a uma vizinhança de $J(x^*)$ de radio δ. Entanto devemos observar que, mesmo se a primeira B_0 pertence a esta vizinhança existiria a possibilidade de que $\|B_k - J(x^*)\| \gg \|B_0 - J(x^*)\|$, destruindo a convergência. Afortunadamente, para os métodos LCSU (incluindo o método de Broyden) é possível provar que existe $\delta' > 0$ tal que $\|B_k - J(x^*)\| \leq \delta$ para todo $k \in I\!N$, se $\|B_0 - J(x^*)\| \leq \delta'$.

Teorema 4 *Existem $\varepsilon, \delta > 0$ tal que, se $\|x^0 - x^*\| \leq \varepsilon$ e $\|B_0 - J(x^*)\| \leq \delta$, a sequência gerada pelo método de Broyden está bem definida, converge para x^* e satisfaz (1.21)*

(Prova: Ver Broyden, Dennis Moré [5]. Uma extensão para outros métodos LCSU é mostrada em Martínez [48], [50]. □

O método COLUM não pertence à família dos métodos LCSU, logo a convergência local superlinear não pode ser provada usando as técnicas baseadas em Propriedades de Deterioração Limitada. Para COLUM conseguimos esse resultado mediante o seguinte teorema

Teorema 5 *Suponhamos que a sequência $\{x^k\}$ seja gerada pelo método COLUM, exceto quando $k \equiv 0 \pmod m$, $B_k = J(x^k)$. Então, existe $\varepsilon > 0$ tal que, se $\|x^0 - x^*\| \leq \varepsilon$, a sequência converge superlinearmente a x^*.*

(Prova: ver Martínez [42]). □

Um resultado similar pode ser obtido para o Método de Atualização de uma Coluna da matriz Inversa (ICOLUM), ver Martínez e Zambaldi [56].

Teorema 6 *Suponhamos que $n = 2$. Seja $r \in (0, 1)$. Então existem $\varepsilon, \delta > 0$ tal que, se $\|x^0 - x^*\| \leq \varepsilon$ e $\|B_0 - J(x^*)\| \leq \delta$, a sequência $\{x^k\}$ gerada por COLUM está bem definida, converge para x^*, e satisfaz (1.21).*

(Prova: ver Martínez [52]). □

Teorema 7 *Suponhamos que a sequência $\{x^k\}$ gerada por COLUM esteja bem definida, convirja para x^* e satisfaça (1.21). Então*

$$\lim_{k\to\infty} \frac{\|x^{k+2n} - x^*\|}{\|x^k - x^*\|} = 0 \tag{1.23}$$

e

$$\lim_{k\to\infty} \|x^k - x^*\|^{1/k} = 0. \tag{1.24}$$

(Prova: ver Martínez [52]). A propriedade (1.24) determina uma *convergência R-superlinear.* □

1.2 Métodos Newton Inexatos

Quando o método de Newton é aplicável à resolução de (1.1), é recoméndavel o uso de algoritmos mais eficientes como CUM ou Broyden com recomeços de Newton. No pacote Rouxinol foi incorporado um procedimento automático, no qual uma iteração de Newton é realizada somente quando há expectativas de que sua eficiência melhore a correspondente da Quase-Newton que vem sendo efetuada.

O uso de uma fatoração esparsa LU para resolver (1.3), pode ser altamente inadequada no caso da matriz Jacobiana ter uma estrutura desfavorável. Um excessivo enchimento durante o processo de fatoração impossibilita o uso de tais técnicas, devido a uma grande demanda de memória computacional e um tempo exagerado de computação a cada iteração. Uma alternativa plausível é a introdução de "Jacobianos Falsos". Esta estratégia consiste, no recomeço de uma iteração Quase-Newton, em substituir $B_k = J(x^k)$ por $B_k = \tilde{J}(x^k)$, onde $\tilde{J}(x^k)$ é um "Jacobiano simplificado" de tal forma que a fatoração LU possa ser desenvolvida. Infelizmente, pode acontecer que $\|J(x^k) - \tilde{J}(x^k)\|$ seja tão grande que o método Quase-Newton perca as propriedades de convergência local.

Em tais circunstâncias o uso de um método Newton-Inexato é altamente recomendável. A inconveniência de se usar um método direto (LU) leva a resolver (1.3) com um Método Iterativo Linear. Usualmente, os métodos iterativos lineares preferidos são aqueles definidos sobre espaços de Krylov (Ver Golub e Van Loan [22], Hestenes e Stiefel [33], Saad e Schultz [64], etc.). Essencialmente, a memória exigida é aproximadamente da mesma ordem que para armazenar o sistema inicial.

Quando resolvemos (1.3) usando um método iterativo linear precisamos providenciar um critério de parada para decidir quando terminar o processo de cálculo (correspondente ao laço interno). Um critério que parece razoável (baseado no valor do resíduo do laço externo) é

$$\|J(x^k)s^k + F(x^k)\| \leq \theta_k \|F(x^k)|, \tag{1.25}$$

onde $\theta_k \in (0,1)$. A condição $\theta_k < 1$ é necessária para que eventualmente, um incremento $s^k \equiv 0$ possa ser aceito como solução aproximada de (1.3). Por outro lado se $\theta_k \approx 0$, o número de iterações necessárias pelo método iterativo linear para satisfazer (1.25) poderia ser muito grande. Um valor usualmente adotado é $\theta_k \approx 0.1$.

Dembo, Eisenstat e Steihaug [9], introduziram um algoritmo impondo o critério (1.25) e provaram as principais propriedades de convergência local.

Teorema 8 *Suponhamos que $F(x^*) = 0$, $J(x^*)$ não singular e contínuo em x^*, e $\theta_k \leq \theta_{\max} < \theta < 1$. Então existe $\varepsilon > 0$ tal que, se $\|x^0 - x^*\| \leq \varepsilon$, a sequência $\{x^k\}$ obtida satisfazendo (1.25) com $x_{k+1} = x_k + s_k$ converge a x^* e satisfaz*

$$\|x^{k+1} - x^k\| \leq \theta \|x^k - x^*\|$$

para todo $k \geq 0$, onde $\|y\| = \|J(x^*)y\|$. Se $lim_{k\to\infty}\theta_k = 0$ a convergência é superlinear.

(Prova: Ver Dembo, Eisenstat e Steihaug [9]) □

Os métodos definidos sobre espaços de Krylov são costumeiramente implementados usando um precondicionador. (Ver Axelsson [3]). Basicamente, um precondicionador para o sistema linear $Az = b$ é uma matriz H tal que a resolução do sistema $HAz = Hb$ demande menor esforço que o sistema original. Aplicado sobre (1.3) resulta

$$H_k^{-1} J(x^k)s^k = -H_k^{-1} F(x^k) \tag{1.26}$$

onde H_k^{-1} (ou no mínimo o produto $H_k^{-1}z$) seja fácil de calcular sendo $H_k \approx J(x^k)$.

Diversos precondicionadores para problemas específicos podem ser encontrados em Spedicato [65], em sua grande maioria baseados sobre Fatorizações Incompletas. Uma característica comum aos diferentes esquemas de precondicionamento aplicados ao sistema $Az = b$, é que a primeira iteração do método iterativo linear é $z^1 = \lambda H^{-1}b$, onde H é o precondicionador. Assim, para (1.3), o primeiro incremento deveria ser da forma $-\lambda H_k^{-1}F(x^k)$. Este valor para s^k será aceito se satisfizer (1.25). Entretanto, já que (1.3) não é um sistema linear isolado seria criterioso usar a informação decorrente em iterações futuras. Com efeito, $J(x^k) \approx J(x^{k-1})$, principalmente quando $k \to \infty$. Este fato, motiva o uso de $H_k, F(x^k), F(x^{k+1}), x^{k+1}, x^k$ para construir o precondicionador H_{k+1} de tal modo que satisfaça a Condição Secante. Assim, é razoável introduzir um algoritmo baseado em (1.25) onde a sequência de precondicionadores $H_k \equiv B_k$ são escolhidos de modo a satisfazer (1.11) para todo $k \in I\!N$.

Existem infinitas possibilidades para a escolha B_{k+1} satisfazendo (1.11). Nazareth e Nocedal [59] e Nash [60] sugeriram o uso da fórmula clássica BFGS para precondicionar (1.3) quando se trata de problemas de minimização.

Uma outra opção é definir

$$B_{k+1} = C_{k+1} + D_{k+1} \tag{1.27}$$

onde C_{k+1} é um precondicionador clássico e D_k é escolhida de modo a satisfazer (1.11).

Martínez [51] mostrou que o uso de uma fórmula precondicionadora secante possibilita obter resultados de convergência mais fortes que as enunciadas no Teorema 8. Isto é, a obtenção de convergência superlinear sem a imposição $\theta_k \to 0$. Um Método Newton Inexato Precondicionado foi introduzido por Martínez [51] com convergência superlinear sem impor uma precisão tendendo a infinito na solução de (1.3).

Algoritmo 1 *Seja $\theta_k \in (0, \theta)$ para todo $k \in I\!N, \theta \in (0,1)$ e $\lim_{k\to\infty}\theta_k = 0$. Suponhamos que $x^0 \in I\!R^n$ seja uma aproximação inicial para a soluçào de (1.1) e que $B_0 \in I\!R^{n\times n}$ seja um precondicionador inicial não singular. Dado $x^k \in I\!R^n$ e B_k não singular, os passos para obter x^{k+1}, B_{k+1} são os seguintes:*

- Passo 1
  Calcular
  $$s_Q^k = -B_k^{-1}F(x^k). \tag{1.28}$$

- Passo 2
  Se
  $$\|J(x^k)s_Q^k + F(x^k)\| \leq \theta\|F(x^k)\| \tag{1.29}$$
  definir
  $$s^k = s_Q^k. \tag{1.30}$$
  Senão, obter um incremento s^k tal que satisfaça (1.25) usando um método iterativo.

- Passo 3
  Fazer $x^{k+1} = x^k + s^k$. ⊙

O teorema seguinte estabelece os principais resultados relacionados ao algoritmo anterior.

Teorema 9 *Suponhamos que $F : \Omega \subset I\!R^n \to I\!R^n$, Ω um conjunto aberto e convexo; $F \in C^1(\Omega)$, $J(x^*)$ não singular, $F(x^*) = 0$, e que (1.6) seja satisfeita para algum $L \geq 0, p \geq 1$. Suponhamos que $||B_k||$ e $||B_k^{-1}||$ estejam limitadas e que a condição de Dennis-Moré seja satisteita. Então existe $\varepsilon > 0$ tal que, se $||x^0 - x^*|| \leq \varepsilon$ a sequência $\{x^k\}$ gerada pelo Algoritmo 4.2 converge superlinearmente a x^*. Além disso existe $k_0 \in I\!N$ tal que $s^k = s_Q^k$ para todo $k \geq k_0$.*

(Prova: Ver Martínez [51]). □

O teorema anterior estabelece que se o precondicionador usado satisfizer a condição de Dennis-Moré, a convergência superlinear é obtida sem $\lim_{k\to\infty}\theta_k = 0$. Em verdade, a primeira

iteração s_Q^k satisfará (1.28) e será aceita como o novo incremento preservando a superlinearidade. As Fórmulas de Atualização Secante (LCSU) podem ser usadas, satisfazendo as hipótese do Teorema 4.3.

Recentemente, Abaffy [1] considerou a possibilidade de usar algoritmos iterativos, ponderando as variações das componentes, sem a necessidade de avaliar o resíduo integralmente e introduzindo um novo critério de parada.

1.3 Estratégias de Globalização

Os métodos locais caracterizam-se por apresentar altas taxas de convergência quando o ponto inicial está suficientemente próximo da solução. Entretanto, podem divergir se esta condição não for satisfeita ou se o sistema não linear apresentar fortes não linearidades. Com a finalidade de eliminar ou reduzir esta possibilidade, os algoritmos baseados em métodos locais são usualmente modificados incorporando propriedades de convergência global.

1.3.1 Globalização por Otimização

Uma forma de implementar esta estratégia consiste em transformar (1.1) em um problema de Otimização, através de uma função de mérito como $f(x) = \frac{1}{2}\|F(x)\|^2$, uma vez que qualquer solução de (1.1) será um mínimo da função f. A opção de usar um método que minimize f para resolver (1.1), em geral, pode não ser satisfatória. Os métodos locais convergem rapidamente para a solução, sendo que a seqüência gerada $\{x^k\}$, não é necessariamente monótona. Nestes casos, o método local puro será mais eficiente que a minimização de f. Por outro lado, os métodos de minimização convergem a mínimos locais (não globais) de f, enquanto que o método local converge para a solução de (1.1). Diferentes soluções tem sido propostas para este problema. (Ver Gripo, Lampariello e Lucidi [31]). Descreveremos a estratégia que combina algoritmos locais e métodos de minimização que foi implementada no pacote computacional Rouxinol. Chamamos de *iteração ordinária* a cada iteração realizada pelo método local e *iteração especial* a correspondente do algoritmo de minimização de f. Definimos, para todo $k \in I\!N$,

$$a^k = \text{Argmin}\,\{f(x^0), \ldots, f(x^k)\}. \tag{1.31}$$

As iterações *ordinárias* e *especiais* são combinadas mediante uma estratégia tolerante.

Algoritmo 2 *Inicializar:* $k \leftarrow 0$, $FLAG \leftarrow 1$.
Seja $q \geq 0$ *um inteiro,* $\gamma \in (0,1)$.

- Passo 1
  Se $FLAG = 1$, *obter* x^{k+1} *por meio de uma iteração ordinária.*
  Senão x^{k+1} *será obtido usando uma iteração especial.*

- Passo 2
  Se

$$f(a^{k+1}) \leq \gamma f(a^{k-q}) \tag{1.32}$$

  Tomar $FLAG \leftarrow 1, k \leftarrow k+1$.
  Voltar ao Passo 1.
  Senão, redefinir
  $x^{k+1} \leftarrow a^{k+1}$.
  $FLAG \leftarrow -1, k \leftarrow k+1$.

Voltar ao Passo 1. ⊙

Se a condição (1.32) for satisfeita um número infinito de vezes, então existirá um subsequência $\{x^k\}$ tal que $\lim_{k-\infty}\|F(x^k)\| = 0$. Se a seqüência for limitada será possível achar uma solução de (1.1) que satisfaça uma precisão predeterminada. Contrariamente, se (1.32) não for satisfeita para todo $k \geq k_0$, então todas as iterações que começam em k_0, serão *especiais* e a convergência da seqüência será controlada pelas propriedades de convergência do algoritmo de minimização.

A princípio, qualquer algoritmo de minimização descrito na literatura pode ser usado para definir uma *iteração especial.*(Dennis and Schnabel [13], Fletcher [18], etc.). Em Rouxinol, visando a resolução de problemas de grande porte, foram implementadas estratégias baseadas em Regiões de Confiança combinadas com critérios tipo Newton Inexatos (Friedlander, Gomes-Ruggiero, Martínez and Santos [20]). Esta estratégia está descrita pelo seguinte algoritmo:

Algoritmo 3 *Suponhamos que* $\Delta_{\min} > 0, \alpha \in (0,1)$ *sejam dadas independentemente da iteração* k. *Define-se* $\psi_k(x) = \|F(x^k) + J(x^k)(x - x^k)\|^2$, $\Delta \geq \Delta_{\min}$.

- Passo 1
  Calcular um minimizador aproximado $\overline{x}$ *de* $\psi_k(x)$ *dentro da caixa* $\|x - x^k\|_\infty \leq \Delta$ *tal que* $\psi_k(x) \leq \psi_k(x_Q^k)$,
  x_Q^k *é a projeção de* $x^k - 2J(x^k)^T F(x^k)/M_k$ *na caixa e* $M_k \geq 2\|J(x_k)\|_1 \|J(x_k)\|_\infty$.

- Passo 2
  Se

$$\|F(\overline{x})\|^2 \leq \|F(x^k)\|^2 + \alpha(\psi(x_k) - \psi_k(\overline{x})) \tag{1.33}$$

  definir $x^{k+1} = \overline{x}$.
  Senão
  Escolher $\Delta_{\text{Novo}} \in [0.1\|\overline{x} - x^k\|, 0.9\Delta]$. *Substituir* Δ *by* Δ_{Novo}.

Voltar ao Passo 1. ⊙

O custo computacional no Algortimo 2 é dado principalmente pela resolução de

$$\left.\begin{array}{rl} \text{Minimizar} & \psi_k(x) \\ s.t. & \|x - x^k\|_\infty \leq \Delta \end{array}\right\} \tag{1.34}$$

que consiste em minimizar uma quadrática em uma caixa n-dimensional.

Para estes problemas, há uma preferencia em usar algoritmos que combinam métodos em Subespaços de Krylov com estratégias de Gradientes Projetados. Em Rouxinol, a solução aproximada de (1.34) está definida por $\psi_k(x) \leq \psi_k(x_Q^k)$ e além disso a norma do gradiente projetado de $\psi_k(x)$ é menor que $0.1\|J(x^k)^T F(x^k)\|$. Também é selecionada $\Delta_{\min} = 0.001\times$ (valor típico de $\|x\|$), como valor inicial de $\Delta \equiv \Delta_0 = \|x^0\|$, $\Delta_{\text{novo}} = 0.5\|\overline{x} - x^k\|$, outra escolha é $\Delta = 4 \times \Delta$. As propriedades de convergência do Algoritmo 2 estão dadas em [20]. Cada ponto limite x^* da seqüência $\{x^k\}$ gerada por este algoritmo satisfaz $J(x^*)^T F(x^*) = 0$. Logo, x^* será a solução de (1.1) se $J(x^*)$ for não singular. Infelizmente, se $J(x^*)$ for singular, existirá a possibilidade $F(x^*) \neq 0$. Justamente este é o caso onde qualquer algoritmo de minimização baseado na globalização de (1.1) esbarrará.

Uma característica interessante das *iterações especiais* baseadas em regiões de confiança consiste na facilidade de se adaptar naturalmente a problemas com restrições para resolver (1.1). Métodos desenvolvidos recentemente com uma abordagem Newton Inexato podem ser encontrados en [8] e [17].

1.3.2 Globalização por Homotopias

Uma técnica alternativa para incluir propriedades de globalização, quando não é possível fornecer uma boa aproximação inicial para resolver (1.1) está baseada em métodos homotópicos.

Podemos definir uma homotopia associada para este problema através de uma função $H(x,t) : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ tal que

$$\left.\begin{array}{rcl} H(x,1) & = & F(x) \\ H(x^0,0) & = & 0. \end{array}\right\} \tag{1.35}$$

Se H satisfaz (1.35) é de esperar-se que

$$\Gamma \equiv \{(x,t) \in \mathbb{R}^n \times \mathbb{R} \mid H(x,t) = 0, 0 \leq t \leq 1\} \tag{1.36}$$

seja uma curva que conecta a aproximação inicial x^0 com uma solução x^*. As técnicas homotópicas consistem em traçar Γ desde $t = 0$ a $t = 1$ de maneira confiável e eficiente. A fixação dos extremos é arbitrária. Propostas precursoras para construir homotopias são encontradas em [36] e [8].

O traçado da curva constitui em alguns problemas um objetivo em si mesmo. O caso onde interessa apenas a solução de $H(x,1) = 0$ conduz a uma situação especial, com conseqüências

práticas. Com efeito, a tentativa de abandonar o traçado da curva, quando t está próximo de 1 e passar para um método local, sem cuidar excessivamente das soluções intermediariass, resultará em tornar o processo de resolver $H(x,1) = 0$ mais eficaz.

Alguns resultados clássicos da Geometria Diferencial garantem que o traçado da curva, começando em x^0 conduz à solução de (1.35) (Milnor [58], Ortega e Rheinboldt [61], Chow, Mallet-Paret e Yorke [7], Watson [11] e [12], etc.).

Homotopias *naturais* aparecem com freqüência; esta designação origina-se pelo fato que o parâmetro t passa a representar um valor característico do próprio problema. Em outras ocasiões, é necessário introduzir homotopias artificiais, aplicáveis em princípio a qualquer problema da forma (1.35). A homotopia de Redução do Resíduo está definida como

$$H(x,t) = F(x) + (t-1)F(x^0).$$

A homotopia "regularizante", implementada no pacote computacional HOMPACK (Watson, Billups e Morgan [70]) está definida como

$$H(x,t) = tF(x) + (1-t)(x - x^0).$$

Em geral, a construção da curva demanda o uso de um método numérico. Após escolher H, o procedimento para o traçado da curva se inicia com a parametrização de Γ . Freqüentemente o próprio parâmetro t pode ser usado. Quando para um determinado t_0 temos que $H'_x(x,t_0)$ é singular, x não pode ser explicitado em função t em uma vizinhança de t_0, o que obriga a decrescer t, com o objeto de progredir em Γ. Por isso, usualmente o traçado de Γ é feito usando o comprimento de arco s como parâmetro. Neste caso o procedimento usualmente recomendado para traçar Γ é do tipo Preditor-Corretor.

Independentemente da escolha da homotopia, do parâmetro e da técnica para o traçado da curva, cada ponto solução de (1.35) será obtido aplicando um método local ao sistema não linear $H(x,t) = 0$, constituindo-se na fase corretora, com aproximação inicial fornecida pela etapa preditora. Se neste sistema consideramos t também como uma variável, teremos n equações com $n+1$ incógnitas. Algoritmos especiais locais para sistemas não lineares subdeterminados foram desenvolvidos por Walker e Watson [67], Martínez [49], etc. Uma interessante discussão sobre métodos homotópicos encontra-se em [19].

Bibliografia

[1] Abaffy, J. [1992]: Superlinear convergence theorems for Newton-type methods for nonlinear systems of equations, *JOTA* , 73, pp. 269 - 277.

[2] Abaffy, J.; Galantai, A. ; Spedicato, E. [1987]: The local convergence of ABS method for nonlinear algebraic system, *Numerische Mathematik* 51, pp. 429 - 439.

[3] Axelsson, O., Kaporin, I. E. [1993]: *On computer implementation of Inexact-Newton-Conjugate Gradient-type algorithms.* Preprint.

[4] Broyden, C.G. [1965]: A class of methods for solving nonlinear simultaneous equations, *Mathematics of Computation* 19, pp. 577-593.

[5] Broyden, C.G.; Dennis Jr., J.E.; Moré, J.J. [1973]: On the local and superlinear convergence of quasi-Newton methods, *Journal of the Institute of Mathematics and its Applications* 12, pp. 223-245.

[6] Chadee, F.F. [1985]: Sparse quasi-Newton methods and the continuation problem, T.R. S.O.L. 85-8, Department of Operations Research, Stanford University.

[7] Chow, S.N.; Mallet-Paret, J.; Yorke, J.A. [1978]: Finding zeros of maps: Homotopy methods that are constructive with probability one, Mathematics of Computation 32, pp. 887-899.

[8] Davidenko, D.F. [1953]: On the approximate solution of nonlinear equations, *Ukrain. Mat. Z.* 5, pp. 196 -206.

[9] Dembo, R.S.; Eisenstat, S.C.; Steihaug, T. [1982]: Inexact Newton methods, *SIAM Journal on Numerical Analysis* 19, pp. 400–408.

[10] Dennis Jr., J.E.; Marwil, E.S. [1982]: Direct secant updates of matrix factorizations, *Mathematics of Computation* 38, pp. 459–476.

[11] Dennis Jr., J.E.; Moré, J.J. [1974]: A characterization of superlinear convergence and its application to quasi - Newton methods, *Mathematics of Computation* 28, pp. 549 -560.

[12] Dennis Jr.,J.E.; Schnabel,R.B. [1979]: Least change secant updates for quasi-Newton methods, *SIAM Review* 21, pp. 443-459.

[13] Dennis Jr.,J.E.; Schnabel,R.B. [1983]:*Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs.

[14] Dennis Jr., J.E. ; Walker, H.F. [1981]: Convergence theorems for least-change secant update methods, *SIAM Journal on Numerical Analysis* 18, pp. 949-987.

[15] Deuflhard, P. [1991]: Global inexact Newton methods for very large scale nonlinear problems *Impact of Computing in Science and Engineering* 3, pp. 366–393.

[16] Duff, I.S.; Erisman, A.M.; Reid, J.K. [1989]:*Direct Methods for Sparse Matrices*, Oxford Scientific Publications.

[17] Eisenstat, S.C.; Walker, H.F. [1993]: Globally convergent inexact Newton methods, to appear in *SIAM Journal on Optimization.*

[18] Fletcher, R. [1987]: *Practical Methods of Optimization* (2nd edition), John Wiley and Sons, New York.

[19] Forster, W. [1993]: Homotopy methods, to appear in *Handbook of Global Optimization*, Kluwer.

[20] Friedlander, A.; Gomes-Ruggiero, M.A.; Kozakevich, D. N.; Martínez, J.M.; Santos, S.A. [1993]: A globally convergent method for solving nonlinear systems using a trust - region strategy, Technical Report, Department of Applied Mathematics, University of Campinas.

[21] George, A.; Ng, E. [1987]: Symbolic factorization for sparse Gaussian elimination with partial pivoting, *SIAM Journal on Scientific and Statistical Computing* 8, pp. 877-898.

[22] Golub, G.H.; Van Loan, Ch.F. [1989]: *Matrix Computations*, The Johns Hopkins University Press, Baltimore and London.

[23] Gomes–Ruggiero [1990]: Algoritmos para a resolução de Sistemas Não Lineares., *Tese de Doutorado* FEC - UNICAMP.

[24] Gomes–Ruggiero, M.A.; Martínez, J.M. [1992]: The Column–Updating Method for solving nonlinear equations in Hilbert space, *Mathematical Modelling and Numerical Analysis* 26, pp 309-330.

[25] Gomes–Ruggiero, M.A.; Martínez, J.M.; Moretti, A.C. [1992]: Comparing algorithms for solving sparse nonlinear systems of equations, *SIAM Journal on Scientific and Statistical Computing* 13, pp. 459 - 483.

[26] Griewank, A. [1992]: Achieving Logarithmic Growth of Temporal and Spacial Complexity in Reverse Automatic Differentiation, *Optimization Methods and Software* 1, pp. 35 - 54.

[27] Griewank, A.; Toint, Ph.L. [1982a]: On the unconstrained optimization of partially separable functions, in *Nonlinear Optimization 1981*, edited by M.J.D. Powell, Academic Press, New York.

[28] Griewank, A.; Toint, Ph.L. [1982b]: Partitioned variable metric for large structured optimization problems, *Numerische Mathematik* 39, pp. 119 - 137.

[29] Griewank, A.; Toint, Ph.L. [1982c]: Local convergence analysis for partitioned quasi-Newton updates, *Numerische Mathematik* 39, pp. 429-448.

[30] Griewank, A.; Toint, Ph.L. [1984]: Numerical experiments with partially separable optimization problems, in *Numerical Analysis Proceedings Dundee 1983*, edited by D.F. Griffiths, Lecture Notes in Mathematics vol. 1066, Springer - Verlag, Berlin, pp. 203-220.

[31] Grippo, L.; Lampariello, F.; Lucidi, S. [1986]: A nonmonotone line search technique for Newton's method, *SIAM Journal on Numerical Analysis* 23, pp. 707 - 716.

[32] Hart, W.E.; Soul, S.O.W. [1973]: Quasi-Newton methods for discretized nonlinear boundary value problems, *J. Inst. Math. Applics.* 11, pp. 351 - 359.

[33] Hestenes, M.R.; Stiefel, E. [1952]: Methods of conjugate gradients for solving linear systems, *Journal of Research of the National Bureau of Standards* B49, pp. 409 - 436.

[34] Johnson, G.W.; Austria, N.H. [1983]: A quasi-Newton method employing direct secant updates of matrix factorizations, *SIAM Journal on Numerical Analysis* 20, pp. 315-325.

[35] Kelley, C.T.; Sachs, E.W. [1987]: A quasi-Newton method for elliptic boundary value problems, *SIAM Journal on Numerical Analysis* 24, pp. 516 - 531.

[36] Lahaye, E. [1934]: Une méthode de résolution d'une catégorie d'equations transcendantes, Comptes Rendus Acad. Sci. Paris 198, pp. 1840-1842.

[37] Martínez, J.M. [1979a]: Three new algorithms based on the sequential secant method, *BIT* 19, pp. 236-243.

[38] Martínez, J.M. [1979b]: On the order of convergence of Broyden - Gay - Schnabel's method, *Commentationes Mathematicae Universitatis Carolinae* 19, pp. 107-118.

[39] Martínez, J.M. [1979c]: Generalization of the methods of Brent and Brown for solving nonlinear simultaneous equations, *SIAM Journal on Numerical Analysis* 16, pp. 434 - 448.

[40] Martínez, J.M. [1980]: Solving nonlinear simultaneous equations with a generalization of Brent's method, *BIT* 20, pp. 501 - 510.

[41] Martínez, J.M. [1983]: A quasi–Newton method with a new updating for the LDU factorization of the approximate Jacobian, *Matemática Aplicada e Computacional* 2, pp. 131–142.

[42] Martínez, J.M. [1984]: A quasi–Newton method with modification of one column per iteration, *Computing* 33, pp. 353–362.

[43] Martínez, J.M. [1986a]: The method of Successive Orthogonal Projections for solving nonlinear simultaneous equations, *Calcolo* 23, pp. 93 - 105.

[44] Martínez, J.M. [1986b]: Solving systems of nonlinear simultaneous equations by means of an accelerated Successive Orthogonal Projections Method, *Computational and Applied Mathematics* 165, pp. 169 - 179.

[45] Martínez, J.M. [1986c]: Solution of nonlinear systems of equations by an optimal projection method, *Computing* 37, pp. 59 - 70.

[46] Martínez, J.M. [1987]: Quasi–Newton Methods with Factorization Scaling for Solving Sparse Nonlinear Systems of Equations, *Computing* 38, pp. 133–141.

[47] Martínez, J.M. [1990a]: A family of quasi-Newton methods for nonlinear equations with direct secant updates of matrix factorizations, *SIAM Journal on Numerical Analysis* 27, pp. 1034-1049.

[48] Martínez, J.M. [1990b]: Local convergence theory of inexact Newton methods based on structured least change updates, *Mathematics of Computation* 55, pp. 143-168.

[49] Martínez, J.M. [1991]: Quasi-Newton Methods for Solving Underdetermined Nonlinear Simultaneous Equations, *Journal of Computational and Applied Mathematics* 34, pp. 171–190.

[50] Martínez, J.M. [1992a]: On the relation between two local convergence theories of least change secant update methods, *Mathematics of Computation* 59, pp. 457–481.

[51] Martínez, J.M. [1992b]: A Theory of Secant Preconditioners, to appear in *Mathematics of Computation.*

[52] Martínez, J.M. [1992c]: On the Convergence of the Column-Updating Methods, Technical Report, Department of Applied Mathematics, University of Campinas.

[53] Martínez, J.M. [1992d]: Fixed-Point Quasi-Newton Methods, *SIAM Journal on Numerical Analysis* 29, pp. 1413–1434.

[54] Martínez, J.M. [1992e]: SOR - Secant Methods, to appear in *SIAM Journal on Numerical Analysis.*

[55] Martínez, J.M. [1993]: Algorithms for Solving Nonlinear System of Equations

[56] Martínez, J.M.; Zambaldi, M.C. [1992]: An inverse Column-Updating Method for solving Large-Scale Nonlinear Systems of Equations, to appear in *Optimization Methods and Software.*

[57] Matthies, H.; Strang, G. [1979]: The solution of nonlinear finite element equations, *International Journal of Numerical Methods in Engineering* 14, pp. 1613 - 1626.

[58] Milnor, J.W. [1969]: *Topology from the differential viewpoint*, The University Press of Virginia, Charlottesville, Virginia.

[59] Nazareth, L.; Nocedal, J. [1978]: A study of conjugate gradient methods, Report SOL 78-29, Department of Operations Research, Stanford University.

[60] Nash, S.G. [1985]: Preconditioning of Truncated Newton methods, *SIAM Journal on Scientific and Statistical Computing* 6, pp. 599 -616.

[61] Ortega, J.M.; Rheinboldt, W.G. [1970]: *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, NY.

[62] Rall, L.B. [1984]: Differentiation in PASCAL - SC: Type Gradient, *ACM Transactions on Mathematical Software* 10, pp. 161-184.

[63] Rall, L.B. [1987]: Optimal Implementation of Differentiation Arithmetic, in *Computer Arithmetic, Scientific Computation and Programming Languages*, U. Külisch (ed.), Teubner, Stuttgart.

[64] Saad, Y.; Schultz, M.H. [1986]: GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems, *SIAM Journal on Numerical Analysis* 7, pp. 856-869.

[65] Spedicato, E. [1991] (editor): *Computer Algorithms for Solving Linear Algebraic Equations. The State of Art*, NATO ASI Series, Series F: Computer and Systems Sciences, Vol. 77, Springer Verlag, Berlin.

[66] Toint, Ph.L. [1986]: Numerical solution of large sets of algebraic nonlinear equations, *Mathematics of Computation* 16, pp. 175 - 189.

[67] Walker, H.F.; Watson, L.T. [1989]: Least - Change Update Methods for underdetermined systems, Research Report, Department of Mathematics, Utah State University.

[68] Watson, L.T. [1979]: An algorithm that is globally convergent with probability one for a class of nonlinear two–point boundary value problems, *SIAM Journal on Numerical Analysis* 16, pp. 394–401.

[69] Watson, L.T. [1980]: Solving finite difference approximations to nonlinear two-point boundary value problems by a homotopy method, *SIAM Journal on Scientific and Statistical Computing* 1, pp. 467-480.

[70] Watson, L.T.; Billups, S.C.; Morgan, A.P. [1987]: Algorithm 652: HOMPACK: A suite of codes for globally convergent homotopy algorithms, *ACM Trans. Math. Software* 13, pp. 281–310.

Capítulo 2

Métodos tipo Newton para problemas com valor de contorno

2.1 Introdução

Os sistemas originados pela discretização de problemas de contorno podem ser considerados como uma das aplicações mais importantes dos métodos para resolver sistemas não lineares esparsos e de grande porte.

As modelagens de diversos problemas da mecânica de fluidos. transferência de calor e massa, etc., daõ origem a problemas deste tipo que podem ser representados por operadores da forma

$$G(u) = \nabla^2 u + H(\lambda, u, u_x, u_y, \ldots, u_{yy}) - f(x,y), \quad (x,y) \in \Omega,$$
$$u = \xi(x,y), (x,y) \in \partial\Omega \tag{2.1}$$

sendo $\Omega \subset R^2$, $\lambda \in I\!R$ e H uma função não linear.

Selecionamos para este trabalho as equações de Poisson Não-Linear [10], o Problema de Bratu Modificado [3] e o Problema de Convecção-Difusão Não-Linear [7] que serão aproximadas usando o método das diferenças finitas. Embora a equação de Poisson tenha sido criada artificialmente , podemos considerar esta coleção de equações não lineares como protótipos de problemas reais.

O principal esforço neste trabalho estará concentrado em resolver cada um dos problemas em uma forma padrão identificando valores de λ para os quais o problema apresente características especiais.

Para isto, selecionamos vários algoritmos baseados nas idéias quase-Newton, que foram implementados incorporando-lhes uma estratégia de globalização, com o intuito de estabelecer um marco de referência para a resolução de um conjunto de problemas definidos como em (2.1).

A introdução dos termos originados pela discretização de H produzem uma deterioração das

propriedades do sistema gerado pela discretização do Laplaciano, piorando o condicionamento e nos dois últimos casos causando a perda da simetria. Em geral o parâmetro λ, que pondera os termos não lineares acentua estas características. Valores de λ para os quais o Jacobiano é singular são denominados *autovalores do sistema não linear* (2.1).

O conjunto dos sistemas algébricos não lineares originados pela discretização das correspondentes equações , constitue uma coleção de problemas teste para a validação dos Métodos Especializados na Resolução de Sistemas Não-Lineares Esparsos e de Grande Porte (Ortega e Rheinblodt [9], Schwandt [10], Watson [11], [12], [13], Watson e Scott [14] , Watson e Wang [15], etc.).

Começaremos inicialmente descrevendo a estratégia de globalização implementada; logo a seguir os problemas que foram objeto de estudo e suas respectivas discretizações, salientando as características numéricas que consideramos mais relevantes; na Seção 5 apresentaremos os testes numéricos e análise dos resultados e finalmente na última Seção, conclusões e futuros trabalhos.

O conjunto de métodos e algoritmos básicos de resolução que serão utilizados neste trabalho estão descritos no Capítulo 1.

2.2 Globalização por "backtracking"

Como mencionamos na Introdução, os métodos quase-Newton não possuem a propriedade de decrescimento monótono

$$\|F(x_{k+1})\| \leq \|F(x_k)\| \tag{2.2}$$

e são localmente convergentes, o que significa que conseguem achar a solução do sistema no caso em que a aproximação inicial seja muito boa. Esta última afirmação, na maioria das vezes pode ser considerada pessimista. Ocorre que devido a boas propriedades da matriz Jacobiana, consegue-se convergência em um número finito de iterações e o método passa a exibir propriedades de convergência global.

Habitualmente incorporam-se modificações sobre as iterações locais para satisfazer (2.2) de tal modo que essa imposição aumente a possibilidade de obter convergência global.

Uma das formas de satisfazer (2.2) consiste em introduzir um procedimento denominado estratégia de retrocesso ("backtracking"). Neste caso a iteração básica se transforma em

$$x_{k+1} = x_k - \alpha_k B_k^{-1} F(x_k), \tag{2.3}$$

onde α_k é obtido da sequência $\{2^{-i},\ i = 0, 1, \ldots\}$. A existência de α_k satisfazendo (2.2) está garantida se $d_k = -B_k^{-1}F(x_k)$ for uma direção de descida, isto é

$$[J(x_k)d_k]^T F(x_k) < 0. \tag{2.4}$$

Esta condição é obviamente satisfeita pela iteração de Newton. Nos métodos quase-Newton a condição (2.4) deve ser previamente conferida antes de efetuar o processo de retrocesso. Um procedimento alternativo que evita a necessidade de calcular o Jacobiano consiste em definir uma outra sequência para α_k, como $\{(-1)^{i+1}2^{-i},\ i = 0, 1, \ldots\}$ de tal modo a satisfazer (2.2). Esta estratégia, que poderia ser denominada como retrocesso bidirecionado, será usada nas experiências numéricas. O processo descrito é costumeiramente incorporado aos algoritmos definidos pelos métodos básicos por razões de ordem prática. Analisaremos o desempenho dos métodos quase-Newton com e sem a estratégia de globalização para resolver os sistemas mencionados acima. Não pretendemos mostrar qual é o melhor dos métodos para resolver um determinado problema mas sim, detectar situações onde alguns deles apresentam alguma deficiência ou um comportamento particular.

2.3 Descrição dos problemas

Nesta Secção descreveremos os problemas em que o Laplaciano é combinado com outros termos não lineares. Em todos os casos acharemos as soluções aproximadas das equações discretizadas no quadrado unitário $\Omega : \{[0, 1] \times [0, 1]$. Resulta relativamente fácil encontrar uma solução exata que satisfaça as condições de fronteira, ajustando o termo independente $f(x, y)$. Para todos os casos escolhemos

$$u^*(x,y) = xy(1-x)(1-y)\exp x^{4.5}$$

(ver [7]); em consequência $\xi(x,y) = 0$. f é avaliada em cada nó da malha de tal forma que $u_h{}^*$, a discretização de u^*, é uma solução exata das equações discretizadas.

O fato de ter definido a priori a solução exata, nos permite definir o seguinte critério de parada

$$\mid u_{i,j}{}^k - u_{i,j}{}^* \mid \leq 10^{-4} \mid u^*_{i,j} \mid,$$

assim o processo iterativo será interrompido na $k-$ésima iteração, quando o erro relativo em cada componente da solução u_h for menor que 10^{-4} .

Devido aos distintos efeitos que introduzem diferentes escolhas de H nas propriedades dos sistemas, estes problemas são usados extensamente como problemas teste padrões; tanto para mostrar a eficácia dos algoritmos para a resolução dos sistemas lineares subjacentes, como para a construção de precondicionadores, etc. A nossa abordagem visa mostrar o desempenho de um determinado processo para melhorar a convergência de um método ao resolver o sistema não linear.

Os problemas testes são listados a seguir juntamente com suas respectivas aproximações.

P1 - Problema de Poisson Não Linear

$$-\Delta u + \lambda \frac{u^3}{1+x^2+y^2} - f_1(x,y) = 0$$

cuja discretização pode ser escrita como

$$4\, u_{i,j} - (u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1})$$

$$+ h^2\, \lambda\, u_{i,j}{}^3 \,/\, (1 + x_i{}^2 + y_j{}^2) - h^2 f_1(x_i, y_j) = 0 \qquad (2.5)$$

$$1 \leq i,j \leq (L-1)$$

Se $\lambda > 0$ o problema é fácil; a dificuldade cresce para valores negativos de λ.

P2 - Problema de Bratu

$$-\Delta u + \frac{\partial u}{\partial x} + \lambda e^u - f_2(x,y) = 0$$

sendo discretizado como

$$4\, u_{i,j} - (u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1})$$

$$+ h\, (u_{i+1,j} - u_{i-1,j}) \,/\, 2 + h^2\, \lambda\, e^{u_{i,j}} - h^2 f_2(x_i, y_j) = 0 \qquad (2.6)$$

$$1 \leq i,j \leq (L-1)$$

P3 - Problema de Convecção-Difusão

$$-\Delta u + \lambda u(\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y}) - f_3(x,y) = 0$$

sendo discretizado como

$$4\, u_{i,j} - (u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1})$$

$$+ h/2\, \lambda\, u_{i,j}\, (u_{i+1,j} - u_{i-1,j} + u_{i,j+1} - u_{i,j-1}) - h^2\, f_3(x_i, y_j) = 0 \qquad (2.7)$$

$$1 \leq i,j \leq (L-1)$$

Em todos os casos se conserva o mesmo padrão de esparsidade que gera a aproximação do operador Laplaciano. Nos dois últimos problemas os sistemas são não simétricos.

2.4 Aproximação Numérica

As equações diferenciais serão aproximadas usando o método das diferenças finitas com uma discretização padrão de segunda ordem sobre uma malha uniformemente espaçada de tamanho $h = 1/L$, onde L é o número de divisões. Denotamos o domínio discretizado por Ω_h sendo $x_i = ih, y_j = jh$ as coordenadas dos nós de Ω_h. Assim teremos $(L-1)^2$ nós em Ω_h e L nós sobre cada lado de $\partial\Omega_h$.

Para uma função de malha qualquer $u_{i,j}$ definimos, em cada nó, os seguintes operadores de diferenças que serão utilizados nos diferentes problemas:

$$D_x u_{i,j} = (u_{i+1,j} - u_{i-1,j})/2h$$

$$D_y u_{i,j} = (u_{i,j+1} - u_{i,j-1})/2h$$

$$D_{xx} u_{i,j} = (u_{i+1,j} - 2u_{i,j} + u_{i-1,j})/h^2$$

$$D_{yy} u_{i,j} = (u_{i,j+1} - 2u_{i,j} + u_{i,j-1})/h^2$$

$$\nabla_h^2 u_{i,j} = (D_{xx} + D_{yy}) u_{i,j}.$$

Fixamos em todos os casos, $L = 64$ obtendo assim $N = 3969$ incógnitas.

2.5 Experiências numéricas

Para cada problema foi criada uma sequência de experiências para diferentes valores do parâmetro λ. Esta sequência foi gerada por um procedimento totalmente heurístico, procurando achar os valores de λ com o intuito de criar casos com o maior grau de dificuldade possível, em

relação à sua resolução. Não levamos em consideração valores para λ que tivessem significado físico, nem tivemos qualquer preocupação em obter soluções positivas.

Os métodos utilizados com e sem globalização, para a realização dos testes são: Newton, Newton Modificado, Broyden (Primeiro Método) e Atualização de Coluna (ver Gómes-Ruggiero [1990] [5]).

Os resultados das experiências são apresentados separadamente para cada problema, nas respectivas tabelas. Cada coluna corresponde a um λ diferente e cada parênteses contém o número de iterações e/ou, o número de iterações e o número de avaliações de função, para as versões com e sem globalização, respectivamente.

Declararamos divergência (*div*) quando o número de iteraçõe realizado pelo método ultrapassa $ItMax$ (número máximo de iterações permitidas) ou a norma do resíduo supera $ResMax$ (valor do resíduo máximo permitido) sendo a experiência interrompida em ambos os casos. Em todos os testes fixamos $ResMax = 10^{20}$ e usamos a mesma aproximação inicial $x^0 = 0$.

Devido a fato de que o custo computacional de uma iteração de Newton é, em geral, consideravelmente mais caro que uma iteração quase-Newton, sendo esta relação muito mais drástica quando comparada com a iteração de Newton Modificado, fixamos diferentes valores para $ItMax$ para cada método em particular, com o propósito de colocá-los em uma situação mais equilibrada. A realização de algumas experiências preliminares nos permitiu padronizar uma relação que considera custos equitativos para cada método. Desta forma fixamos, para cada iteração de Newton, 15 de Broyden e Atualização de Coluna e 25 de Newton Modificado. Vale esclarecer que existem pequenas variações destes valores para os distintos problemas e obviamente entre os métodos de Broyden e Atualização de Coluna. Estas experiências também nos possibilitaram estabelecer $ItMax_{Newton} = 10$, considerando: um custo razoável em tempo real e a demanda média do número de iterações para conseguir convergência. O custo médio em tempo real de uma iteração de Newton é de aproximadamente 17 segundos.

Por outro lado, os critérios de parada para a convergência forem estabelecidos quando alguma das seguintes condições foram satisfeitas:

$$|x_i - x_i^*| \leq 10^{-4},$$

ou

$$\| F(x) \|_\infty \leq 10^{-10}.$$

Os casos em que as iterações foram interrompidas por causa desta última condição estão indicadas com um asterisco o que eventualmente indica que a solução obtida é outra diferente da solução exata. Nestes casos para corroborar esta hipótese, calculamos o erro entre as soluções exata e a calculadaem forma aproximada para a componente situada no meio do quadrado como

$$Err_\lambda = |x^{ob} - x^*|.$$

A resposta *(stop)* significa a impossibilidade de obter um decréscimo no resíduo durante a busca linear; também as iterações são interrompidas.

Os testes foram realizados em uma SPARCStation-5, e os programas computacionais implementados em linguagem Fortran77 com precisão dupla.

Analisamos a seguir os resultados obtidos para cada problema.

Método	λ	−200	−100	−35	−10	200	1000
Newton	Sem Global.	(> 10)	(3)*	(10)*	(3)	(6)	(10)
	Globalizado	(9,14)	(3,3)*	(10,10)*	(3,3)	(6,6)	(10,10)
Newton Modif.	Sem Global.	(div)	(7)*	(> 250)	(7)	(> 250)	(> 250)
	Globalizado	(stop)	(7,7)*	(> 250)	(7,7)	(17,66)	(28,143)
Broyden	Sem Global.	(35)	(5)*	(15)*	(4)	(14)	(73)
	Globalizado	(stop)	(5,5)*	(15,15)*	(4,4)	(13,29)	(stop)
Atual. Coluna	Sem Global.	(> 150)	(5)*	(14)*	(4)	(23)	(> 150)
	Globalizado	(stop)	(5,5)*	(14,14)*	(stop)	(stop)	(stop)

Tabela 2.1: Equação de Poisson Não Linear

Equação de Poisson - Na Tabela (2.1) observamos que para $\lambda = -35$ o número de iterações é notoriamente maior que para $\lambda = \pm 100$ o que nos faz suspeitar a proximidade de um autovalor. Para $\lambda = 200$, CUM (Atualização da Coluna) perde a convergência quando é rodado com globalização; o mesmo acontece com Broyden para $\lambda = -200$ e $\lambda = 1000$.

Para $\lambda = -100$ e $\lambda = -35$ obtivemos convergência com a norma do resíduo; para o primeiro caso constatamos a convergência para uma outra solução; temos $Err_{-100} = 0.71762801070788$. Para o outro valor de λ obtivemos $Err_{-35} = 6.9234404057972D - 03$ muito próximo de 10^{-4}. Para $\lambda = -1000$ não se obteve convergência em nenhum caso.

Para $\lambda = 200$ o método de Broyden globalizado mostrou uma pequena margem de vantagem, com um tempo de execução total de 26.38 segundos contra 26.60 segundos sem "backtracking".

O único método favorecido com a globalização foi Newton Modificado para $\lambda = 200$ e $\lambda = 1000$.

Equação de Bratu - Na Tabela (2.2) observamos que para $\lambda = -10$ obtivemos o que poderia ser chamado de resultado padrão em termos do número de iterações realizadas.

Para $\lambda \in [-10, 1000]$ a maioria dos métodos teve um desempenho semelhante. Todos os testes rodaram sem fazer uso da globalização uma única vez.

Para $\lambda = -100$ conseguimos convergência com Newton para uma outra solução; temos

$Err_{-100} = 0.75267972007852$. Idem para $\lambda = -40$ aparecendo uma outra solução diferente, com $Err_{-35} = 1.2259745817774$.

Aparentemente para $\lambda \in [-100, -40]$ existem várias soluções; a análise das soluções neste intervalo está fora do escopo de nosso trabalho.

Um teste realizado com $\lambda = 5000$ produziu resultados similares aos obtidos com $\lambda = 1000$.

Método	λ	-100	-40	-25	-10	100	1000
Newton	Sem Global.	(8)*	(5)*	(6)	(3)	(4)	(4)
	Globalizado	(8,18)*	(5,5)*	(6,6)	(3,3)	(4,4)	(4,4)
Newton Modif.	Sem Global.	(div)	(47,47)*	(66)	(6)	(15)	(43)
	Globalizado	(stop)	(47,47)*	(66,66)	(6,6)	(15,15)	(43)
Broyden	Sem Global.	(> 151)	(9)*	(10)	(4)	(6)	(7)
	Globalizado	(div)	(9,9)*	(10,10)	(4,4)	(6,6)	(7,7)
Atual. Coluna	Sem Global.	(> 151)	(10)*	(23)	(4)	(7)	(7)
	Globalizado	(div)	(10,10)*	(div)	(4,4)	(7,7)	(7,7)

Tabela 2.2: Equação de Bratu

Equação de Convecção–Difusão - Na Tabela (2.3) reportamos os resultados para o Problema de Convecção-Difussão. Este se mostra um problema de difícil resolução. Curiosamente ocorre um grande número de casos onde a globalização prejudica a convergência. Para os valores de $\lambda = \pm 100$ não conseguimos obter convergência com nenhum método. Este problema apresenta resultados "simétricos" em relação aos valores positivos e negativos de λ.

2.6 Conclusões e trabalhos futuros

Neste Capítulo reunimos um conjunto de problemas não lineares originados das discretizações de problemas de contorno de segunda ordem. Os sistemas resultantes foram resolvidos com algoritmos baseados nas idéias dos métodos quase-Newton e implementados com globalização.

Em geral a estratégia de globalização por "backtracking" foi acionada poucas vezes e em vários casos levou à divergência. Podemos concluir que as direções geradas por cada um dos métodos são inadequadas e não permitem obter um decréscimo do resíduo; obviamente, nesta situação, qualquer estratégia de globalização será inútil. Por outro lado, as modificações que introduzem as globalizações na sequência das soluções eventualmente podem ser mal sucedidas.

Método	λ	-50	-20	-10	10	20	50
Newton	Sem Global.	(8)	(4))	(3)	(3)	(4)	(7)
	Globalizado	(> 10)	(4,4))	(3)	(3,3)	(4,4)	(7,13)
Newton Modif.	Sem Global.	(> 250)	(> 250))	(44)	(44)	(> 250)	(> 250)
	Globalizado	(stop)	(stop)	(44)	(44,44)	(stop)	(stop)
Broyden	Sem Global.	(> 150)	(16)	(8)	(9)	(15)	(> 150)
	Globalizado	(stop)	(stop)	(10,14)	(9,9)	(stop)	(stop)
Atual. Coluna	Sem Global.	(> 150)	(16)	(9)	(8)	(15)	(> 150)
	Globalizado	(stop)	(stop)	(8,8)	(9,9)	(18,36)	(stop)

Tabela 2.3: Equação de Convecção-Difussão

Em geral, para uma mesma situação a versão com globalização melhorou ligeiramente o custo computacional.

Em particular, a expectativa em termos de dificuldade para resolver um determinado problema não linear como os apresentados deve ser formada em base às mudanças que produzem os termos originados pela função H, nas propriedades estruturais da matriz Jacobiana.

Problemas de difusão não-linear cuja equação arquetípica pode ser escrita como

$$\frac{\partial u}{\partial t} = \Delta\phi(u) + f(u).$$

têm recentemente suscitado um particular interesse. A solução está definida em um domínio espaço-temporal da forma $\Omega \times [0, T]$. Esta equação modela várias situações reais como:

$$\frac{\partial u}{\partial t} = \Delta u^m$$

conhecida como "a equação em meios porosos" que por sua vez representa outros casos como: a equação de calor com $m = 1$, a teoria de gases ionizados a altas temperaturas com $m > 1$, a teoria de trasferência radiante, a teoria de camada limite, etc. Por outra parte, a equação em meios porosos representa o caso mais simples de uma classe de equações da forma:

$$\frac{\partial u}{\partial t} - \nabla.[\mathrm{K}.\nabla\phi(u)] + \nabla.[\mathrm{v}\phi(u)] + \phi(u) = 0$$

Aqui, o termo de difusão $\Delta\phi$ é acrescido com termos não-lineares convectivos $\nabla\phi$ e tipo fonte/sumidouro ϕ. Equações deste tipo são obtidas pelas modelagens de problemas como: escoamento de águas superficiais, dinâmica populacional, reservatórios de petróleo, etc. O problema matemático correspondente consiste em deteminar de que forma a estrutura do operador influirá no comportamento da solução (ver Peleter e Serrin [8]).

Para a nossa abordagem atual, a seleção apropriada de alguns destes problemas conduziria a definir um outro conjunto de problemas padrão que iriam complementar os escolhidos neste trabalho.

Bibliografia

[1] Deuflhard, P. [1995]: *Newton techniques for highly nonlinear problems. Theory and algorithms*, in preparation.

[2] Deuflhard, P.; Freund, R. and Walter, A. [1990]: Fast secant methods for the iterative solution of large nonsymmetric linear systems, *Impact of Computing in Science and Engineering* 2, pp. 244-276.

[3] Brown, P. N., Saad Y.,[1990] *Hibrid Krylov Methods for Nonlinear Systems of Equations*, SIAM Journal on Scientific and Statistical Computing 11, pp. 450 - 481.

[4] Golub, G.H.; Van Loan, Ch.F. [1989]: *Matrix Computations*, The Johns Hopkins University Press, Baltimore and London. ork.

[5] Gomes-Ruggiero M. A.,[1990]: *Métodos Quase-Newton para Sistemas Não Lineares* Tese de Doutourado, FEE, UNICAMP.

[6] Johnson, G.W.; Austria, N.H. [1983]: *A quasi-Newton method employing direct secant updates of matrix factorizations*, SIAM Journal on Numerical Analysis 20, pp. 315-325.

[7] Kelley, C.T. [1995]: *Iterative methods for linear and nonlinear equations*, SIAM Publications, to appear.

[8] Peleter, L. A., Serrin, J. [1993]: *Nonlinear Difussion Equations and Their Equilibrium States*, Peleter & Seriin (Eds.)

[9] Ortega, J.M.; Rheinboldt, W.G. [1970]:*Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, NY.

[10] Schwandt, H. [1984]: *An interval arithmetic approach for the construction of an almost globally convergent method for the solution of the nonlinear Poisson equation on the unit square*, SIAM Journal on Scientific and Statistical Computing 5, pp. 427 - 452.

[11] Watson, L.T. [1979]: *An algorithm that is globally convergent with probability one for a class of nonlinear two-point boundary value problems*, SIAM Journal on Numerical Analysis 16, pp. 394–401.

[12] Watson, L.T. [1980]: *Solving finite difference approximations to nonlinear two-point boundary value problems by a homotopy method*, SIAM Journal on Scientific and Statistical Computing 1, pp. 467-480.

[13] Watson, L.T. [1983]: *Engineering applications of the Chow–Yorke algorithm*, in Homotopy Methods and Global Convergence (B.C. Eaves and M.J. Todd eds.), Plenum, New York.

[14] Watson, L.T.; Scott, M.R. [1987]: *Solving spline-collocation approximations to nonlinear two-point boundary value problems by a homotopy method*, Applied Mathematics and Computation 24, pp. 333-357.

[15] Watson, L.T.; Wang, C.Y. [1981]: *A homotopy method applied to elastics problems*, International Journal on Solid Structures 17, pp. 29-37.

Capítulo 3

Métodos tipo Newton globalizados para a equação biharmonica não linear

3.1 Introdução

Neste capítulo analisaremos o desempenho de um conjunto de métodos quase-Newton globalizados e Newton-Inexato, para resolver as equações de Navier-Stokes em uma cavidade quadrada para altos números de Reynolds. O fluxo é newtoniano e incompressível em regime estacionário. Formulado em termos da função-corrente define um problema de quarta ordem não linear, com valores de fronteira (Peyret e Taylor [21]).

Uma primeira dificuldade neste problema, que é frequentemente abordado com um enfoque numérico-computacional (Walker [27], Deuflhard [8], Axelsson e Kaporin [1], etc) aparece associada ao termo advectivo. Na medida em que o parâmetro que pondera esse termo (o número de Reynolds), é incrementado se produz correspondentemente, um aumento da não linearidade que por sua vez afeta as estruturas dos sistemas lineares subjacentes, o que se traduz em um paulatino crescimento do mau condicionamento desses sistemas e gradual dissimetria. Em nossos testes o número de Reynolds é incrementado até o aparecimento de soluções espúrias (vizinhança de um ponto limite), (Schreiber e Keller [19]). A possibilidade de continuar construindo a curva requer o uso de técnicas de Continuação mais especializadas (Schreiber e Keller [24], Rheinblodt [18], etc.).

As discontinuidades nas condições de fronteira requerem o uso de técnicas de discretização mais apuradas. Torna-se difícil determinar a influência destas singularidades sobre a precisão da solução. Neste sentido, tem-se realizados significativos esforços, orientados principalmente na direção de criar esquemas de discretização alternativos (Crochet, Davies e Walters [7])

As técnicas de discretização e os métodos utilizados não são novos, tomados individualmente, porém a escolha de uma de tais técnicas que aparece como sendo simples, precisa e robusta combinada com o conjunto de algoritmos para a resolução das equações resultantes, pretende

criar um novo marco de procedimento na resolução numérica deste problema. A introdução de diversas técnicas de globalização pretende ampliar qualitativamente a definição deste marco, que será estabelecido através de um estudo comparativo dos métodos, em relação a sua capacidade e eficiência em termos de esforço computacional. Também, este problema se constitui em mais um problema teste para validar os métodos especializados na Resolução de Sistemas Não-Lineares Esparsos e de Grande Porte.

Consideramos conveniente aclarar que nosso principal objetivo consiste em analisar o desempenho de algoritmos e técnicas complementares para uma estrutura que resulta interessante "per se" e que é originada pela modelagem de um problema real da mecânica dos fluidos, antes que resolver otimamente este problema em particular.

Este Capítulo está organizado como segue: inicialmente descreveremos o problema objeto de estudo e sua discretização, salientando suas características numéricas mais importantes do ponto de vista que nos interessa; logo a seguir são apresentados o testes numéricos, análise dos resultados e as conclusões. O conjunto de algoritmos de resolução utilizados são os descritos nos capítulos anteriores, sendo mencionados quando for necessário, alguns dos parâmetros mais relevantes.

3.2 Formulação Função-Corrente Vorticidade

Um fluxo incompressível é caracterizado por

$$\nabla \cdot \mathbf{V} = 0.$$

Quando esta condição é introduzida na equação de continuidade obtemos

$$\partial \rho / \partial t + \mathbf{V} \cdot \nabla \rho = 0$$

o que significa que a densidade ρ permanece constante ao longo da trajetória das partículas do fluido. Se a viscosidade μ é constante, a equação de momento se reduz a

$$\rho[\partial \mathbf{V}/\partial t + (\mathbf{V} \cdot \nabla)\mathbf{V}] + \nabla p - \mu \nabla^2 \mathbf{V} = \mathbf{f_e} \tag{3.1}$$

que é denominada a forma não-conservativa da equações de Navier-Stokes. Neste caso (formulação nas variáveis primitivas) as incógnitas são o campo de velocidades $\mathbf{V}$ e a pressão p.

Uma outra formulação das equações de Navier-Stokes faz uso do vetor vorticidade

$$\omega = \nabla \times \mathbf{V} \tag{3.2}$$

Pela aplicação do operador rotacional na equação (3.1), o termo que contém a pressão desaparece , resultando

$$\partial \omega / \partial t + (\mathbf{V} \cdot \nabla)\omega - (\omega \cdot \nabla)\mathbf{V} - \nu \nabla^2 \omega = 1/\rho \nabla \times \mathbf{f_e} \tag{3.3}$$

onde $\nu = \mu/\rho$ é viscosidade cinemática. Esta equação é usualmente associada com o vetor função-corrente definido através de

$$\mathbf{V} = \nabla \times \mathbf{\Psi} \tag{3.4}$$

sendo assim automaticamente satisfeita a condição de incompressibilidade. Aplicando o rotacional a (3.4) e usando (3.2) obtemos

$$\nabla^2 \mathbf{\Psi} + \omega = 0. \tag{3.5}$$

Esta formulação se torna interessante quando o vetor $\mathbf{\Psi}$ tem apenas uma componente, o que acontece para um fluxo plano onde

$$\mathbf{V} = \nabla \times (\mathrm{k}\psi)$$

sendo $\mathbf{k}$ é o vetor unitário normal ao plano do fluxo e ψ é uma função escalar.Neste caso, a vorticidade $\omega = \omega\mathbf{k}$ e as equações (3.3) e (3.5) se transformam em equações escalares

$$\partial\omega/\partial t + (\mathbf{V}\cdot\nabla)\omega - \nu\nabla^2\omega = 1/\rho\nabla \times f_e \tag{3.6}$$

$$\nabla^2\Psi + \omega = 0.$$

Como para as equações nas variáveis primitivas, a eq.(3.6) é chamada uma forma não-conservativa Como caracterísitica relevante, na formulação função corrente-vorticidade a pressão não aparece explícitamente.

Em ausência de campos externos e para o caso estacionário a eq.(3.6) se transforma em

$$(\mathbf{V}\cdot\nabla)\omega - \nu\nabla^2\omega = 0 \tag{3.7}$$

$$\nabla^2\Psi + \omega = 0. \tag{3.8}$$

onde $\nu = \mu/\rho$ é a viscosidade cinemática.

Quando estas equações são apropriadamente adimensionalizadas, é possível substituir ν pelo recíproco do número de Reynolds $Re^{-1} = \nu/VL$ sendo V uma velocidade média e L um comprimento característico do modelo físico.

Eliminando ω de (3.7) e (3.8) e tomando para as componentes do campo de velocidades $\mathbf{V}$ como $(u, v) = (\partial\psi/\partial x, -\partial\psi/\partial y)$ obtemos uma equação apenas em termos de ψ, não linear, de quarta ordem

$$F(\psi) = \nabla^4\psi + Re[\psi_y(\nabla^2\psi)_x - \psi_x(\nabla^2\psi)_y] \tag{3.9}$$

Podemos reescrever esta ultima equação na forma

$$F(\psi) = B(\psi) + ReG(\psi), \tag{3.10}$$

sendo $B(\psi)$ o operador biharmonico linear

$$\nabla^4\psi = \psi_{xxxx} + 2\psi_{xxyy} + \psi_{yyyy}$$

e

$$G(\psi) = \psi_y(\nabla^2\psi)_x - \psi_x(\nabla^2\psi)_y$$

um termo não linear em ψ .

3.3 O Problema da Cavidade

Consideramos o problema de resolver o fluxo bidimensional estacionário para um fluido viscoso incompressível. O domínio Ω onde determinaremos a solução em termos da função corrente, é definido como um quadrado de lado L, com seu lado superior aberto em contato com um fluido viscoso que se desloca com velocidade unitária V. Para um fluxo totalmente desenvolvido, indicamos a existência de um vórtice principal que ocupa a região central e uma série de vórtices secundarios próximos aos vértices, girando em sentido contrário. Esta configuração é mostrada na Figura (3.1) conjuntamente com as condições de fronteira

Assim, o problema consiste em achar $\psi(x,y) \in C^4$ tal que

$$F(\psi) = B(\psi) + ReG(\psi) = 0 \text{ em } \Omega, \tag{3.11}$$

onde $\Omega = \{(x.y) : 0 < x < 1,\ 0 < y < 1)\}$ e que satisfaça as seguintes condições de contorno

$$\begin{aligned} \psi &= 0, \qquad (x,y) \in \partial\Omega, \\ \psi_x(0,y) &= 0, \quad 0 \leq y \leq 1, \\ \psi_x(1,y) &= 0, \quad 0 \leq y \leq 1, \\ \psi_y(x,0) &= 0, \quad 0 \leq x \leq 1, \\ \psi_y(x,1) &= 1, \quad 0 \leq x \leq 1. \end{aligned} \tag{3.12}$$

Esta forma de definir as condições na fronteira origina discontinuidades nas derivadas normais nos vértices superiores no lado aberto do quadrado. Uma tentativa para suavizar essas discontinuidades é proposta em [3], mudando esta última condição para

$$\psi_y(x,1) = -16x^2(1-x)^2,\ 0 \leq x \leq 1$$

Como ψ_y é singular nestes cantos, qualquer discretização espalha esta singularidade aos nós vizinhos de tal forma que o esquema considerado deve ser modificado para levar em conta esta singularidade. Um tratamento rigoroso em tal sentido faz uso localmente de uma expressão análitica da singularidade [18]. Em [14] é apresentado um método considerando também uma forma local para a singularidade, introduzindo-a dentro do esquema global em diferenças. Outros métodos alternativos como refinamento da malha, transformação conforme, séries de potências, etc. são listados em [10] ,[6].

3.4 Aproximação Numérica

Obteremos uma solução aproximada usando o método das diferenças finitas. Para isto escolhemos uma discretização padrão de segunda ordem definida sobre uma malha uniformemente

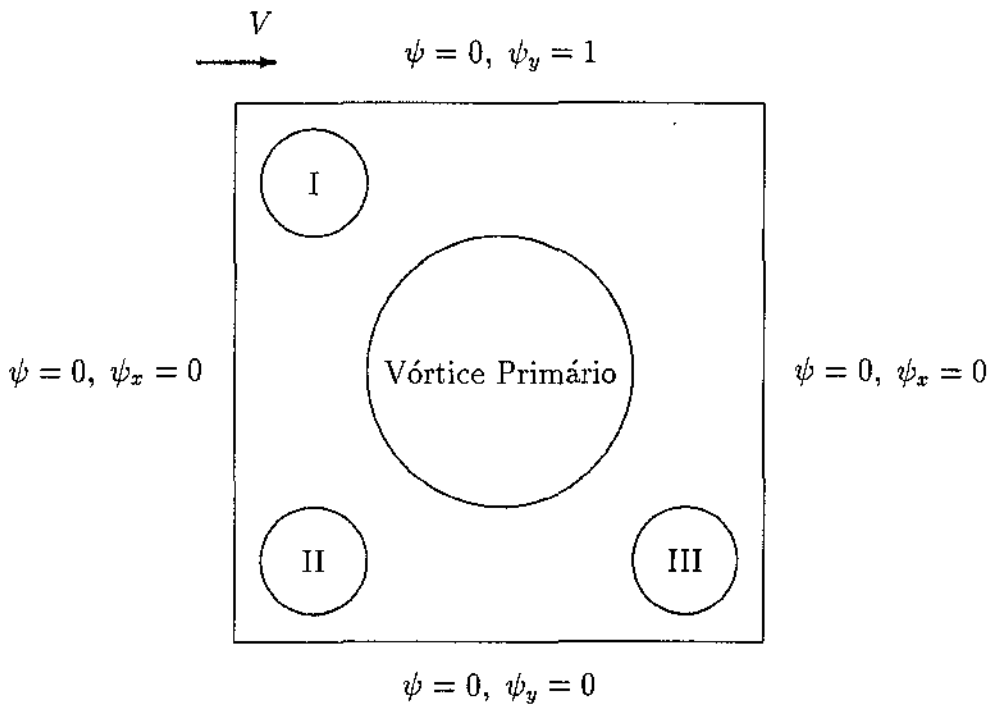

Figura 3.1: Vórtices da Cavidade

espaçada de tamanho $h = 1/L$, onde L é o número de divisões, em ambas direções x e y. Denotamos o domínio discretizado por Ω_h sendo $x_i = ih, y_j = jh$ as coordenadas dos nós de Ω_h. Assim teremos $(L-1)^2$ nós em Ω_h e L nós sobre cada lado de $\partial\Omega_h$.

Para uma função de malha qualquer $u_{i,j}$ definimos em cada nó os seguintes operadores de diferenças

$$D_x u_{i,j} = (u_{i+1,j} - u_{i-1,j})/2h$$

$$D_y u_{i,j} = (u_{i,j+1} - u_{i,j-1})/2h$$

$$D_{xx} u_{i,j} = (u_{i+1,j} - 2u_{i,j} + u_{i-1,j})/h^2$$

$$D_{yy} u_{i,j} = (u_{i,j+1} - 2u_{i,j} + u_{i,j-1})/h^2$$

$$\nabla_h^2 u_{i,j} = (D_{xx} + D_{yy}) u_{i,j}$$

$$\nabla_h^4 u_{ij} = ((D_{xxxx} + 2D_{xxyy} + D_{yyyy}) u_{ij})/h^4$$

$$G_h(u_{ij}) = (((D_y)(D_x \nabla_h^2) - (D_x)(D_y \nabla_h^2)) u_{ij})/h^4.$$

Quando a solução $\psi(x_i, y_j)$ de (3.11) junto com (3.12) é aproximada nos correspondentes nós por uma função de malha ψ_{ij} satisfazendo um esquema em diferenças de acordo ao dado acima, obtemos a seguinte função de resíduo

$$F_h(\psi_{ij}) =$$

$$\nabla_h^4\psi_{ij} + ReG_h(\psi_{ij}) =$$

$$20\psi_{i,j} - 8(\psi_{i-1,j} + \psi_{i+1,j} + \psi_{i,j-1} + \psi_{i,j+1})$$

$$+2(\psi_{i-1,j+1} + \psi_{i+1,j-1} + \psi_{i-1,j-1} + \psi_{i+1,j+1})$$

$$+\psi_{i-2,j} + \psi_{i+2,j} + \psi_{i,j-2} + \psi_{i,j+2}$$

$$+Re/4(\psi_{i,j+1} - \psi_{i,j-1}) \tag{3.13}$$

$$(\psi_{i-2,j} + \psi_{i-1,j-1} + \psi_{i-1,j+1} - 4\psi_{i-1,j} + 4\psi_{i+1,j} - \psi_{i+1,j-1} - \psi_{i+1,j+1} - \psi_{i+2,j})$$

$$-Re/4(\psi_{i+1,j} - \psi_{i-1,j})$$

$$(\psi_{i,j-2} + \psi_{i-1,j-1} + \psi_{i+1,j-1} - 4\psi_{i,j-1} + 4\psi_{i,j+1} - \psi_{i-1,j+1} - \psi_{i+1,j+1} - \psi_{i,j+2}),$$

$$1 \leq i, j \leq (L-1)$$

Desde que $F_h(\psi_{ij}) = 0$ deve ser satisfeita apenas para os nós de Ω_h teremos $N = (L-1)^2$ equações. Porém, os valores de ψ_{ij} sobre o contorno e exteriores vizinhos à fronteira, estarão incluidos nestas equações.

Os valores de contorno são os análogos a (3.12) discretizados

$$\psi_{ij} = 0, (x_i, y_j) \in \partial\Omega_h,$$

Por outro lado, os valores de ψ_{ij} nos nós exteriores a $\overline{\Omega}_h$ podem ser determinados aproximando as derivadas normais especificadas na fronteira mediante um esquema de diferenças centradas

$$\psi_{i,-1=}\psi_{i,1}, 1 \leq i \leq (L-1)$$

$$\psi_{i,(L+1)=}\psi_{i,(L-1)}, 1 \leq i \leq (L-1)$$

$$\psi_{-1,j=}\psi_{1,j}, 1 \leq j \leq (L-1)$$

$$\psi_{(L+1),j=}\psi_{(L-1),j}, 1 \leq j \leq (L-1)$$

Usando estes valores em (3.13) para eliminar os valores externos a $\overline{\Omega}_h$. obtemos N equações para as N incógnitas ψ_{ij} .

Este sistema de equações têm não linearidades quadráticas e é esparso. Cada equação contém 13 incógnitas dispostas em um arranjo molecular em forma de estrela como é mostrada

na Figura (3.2). Esta é a forma padrão para uma aproximação em diferenças centradas de segunda ordem para o operador biharmonico em uma malha retangular.

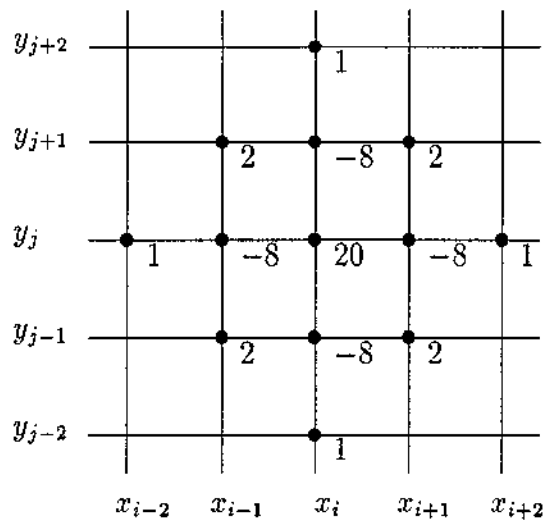

Figura 3.2: Molécula de 13 pontos para o operador biharmonico discretizado

3.5 Procedimento de Resolução

O procedimento escolhido consiste na construção de uma "curva" $(\psi,\ Re)$ satisfazendo $F_h(\psi, Re) = 0$, para uma sequência crescente de números de Reynolds : $Re^{k+1} = Re^k + \Delta Re$ até a vizinhança de um ponto limite ("singular point") com passos fixos ΔRe.

Desta forma, um conjunto de soluções é gerado a partir de $(\psi^0, Re = 0)$ resolvendo o sistema $F_h(\psi, Re + \Delta Re) = 0$ para cada passo ΔRe, utilizando como aproximação inicial a solução de $F_h(\psi, Re) = 0$.

Procedimentos deste tipo, no qual um parâmetro (neste caso o número de Reynolds), variando num intervalo é incrementado gradativamente e onde as soluções intermediárias são usadas como aproximações iniciais para as próximas iterações, são denominadas Técnicas de Continuação.

Com o intuito de obter melhores aproximações iniciais, uma abordagem clássica para a implementaçao destas técnicas modifica o procedimento descrito acima (que pode ser considerado como elementar), diferenciando em relação ao parâmetro (número de Reynolds) e originando uma equação diferencial ordinária com valor inicial

$$\frac{\partial F_h(\psi_h, Re)}{\partial Re}\ \dot{\psi}_h\ (Re) = G_h(\psi_h, Re),$$

com $\psi_h(0) = \psi_h^0$.

Por sua vez, essas aproximações podem ser aprimoradas com a utilização de técnicas mais especializadas de integração. Uma técnica simples, conhecida como Euler-Newton se tem mostrada adequada, sempre que não existam pontos singulares; isto é, pontos onde $D[F_h(\psi, Re)]$ seja singular.

Neste caso torna-se necessário parametrizar a curva pelo cumprimento de arco s em lugar do Re e construir um caminho seguindo a curva. Esta abordagem requer a solução de $F_h(\psi(s), Re(s)) = 0$ mais uma equação não linear para o comprimento de arco s que pode ser aproximada por

$$(\delta s)^2 = \|\delta\psi\|_1^2 + (\delta Re)^2$$

sendo que, para criar o problema com valor inicial, a derivação deve ser feita em relação a s, com valores iniciais definidos por $\psi_h(0) = \psi_h^0$ e $Re(0) = 0$.

A linearização de $F_h(\psi, Re) = 0$, pela aplicação do método de Newton origina uma sequência de problemas lineares,

$$J(\psi^\nu, Re)\delta^\nu = -F_h(\psi^\nu, Re) \tag{3.14}$$

onde $\delta^\nu = \psi^{\nu+1} - \psi^\nu$ e $J_{(N\times N)}$ é a matriz Jacobiana.

Básicamente, a diferença fundamental na capacidade dos métodos para resolver (3.13) está diretamente ligada com a forma com que resolvem os sistemas lineares (3.14) correspondentes; os quais por sua vez, dependem da estrutura e propriedades do Jacobiano

$$J(\psi^\nu, Re) = \partial F_h(\psi^\nu, Re)/\partial\psi = B_h[\psi] + Re\partial G_h[\psi]/\partial\psi.$$

Os coeficientes da matriz Jacobiana estão constituidos por constantes correspondentes às derivadas de B_h e de uma parte antisimétrica $\partial G_h[\psi]/\partial\psi$, ponderada pelo número de Reynolds e que depende da "suavidade " da função de malha ψ.

Na Figura (3.5) mostramos uma linha da matriz Jacobiana onde
$P = (\psi_{i+1,j} - \psi_{i-1,j})$
$Q = (\psi_{i-2,j} + \psi_{i-1,j-1} + \psi_{i-1,j+1} - 4\psi_{i-1,j} + 4\psi_{i+1,j} - \psi_{i+1,j-1} - \psi_{i+1,j+1} - \psi_{i+2,j})$
$R = (\psi_{i,j+1} - \psi_{i,j-1})$
$S = (\psi_{i,j-2} + \psi_{i-1,j-1} + \psi_{i+1,j-1} - 4\psi_{i,j-1} + 4\psi_{i,j+1} - \psi_{i-1,j+1} - \psi_{i+1,j+1} - \psi_{i,j+2})$

3.6 Análise dos Resultados Numéricos

A maioria dos testes e respectivos resultados que serão apresentados foram realizados sobre uma malha uniformemente espaçada, com $L = 64$ divisões em ambas direções x e y, o que origina um problema com $N = 3969$ equações e incógnitas.

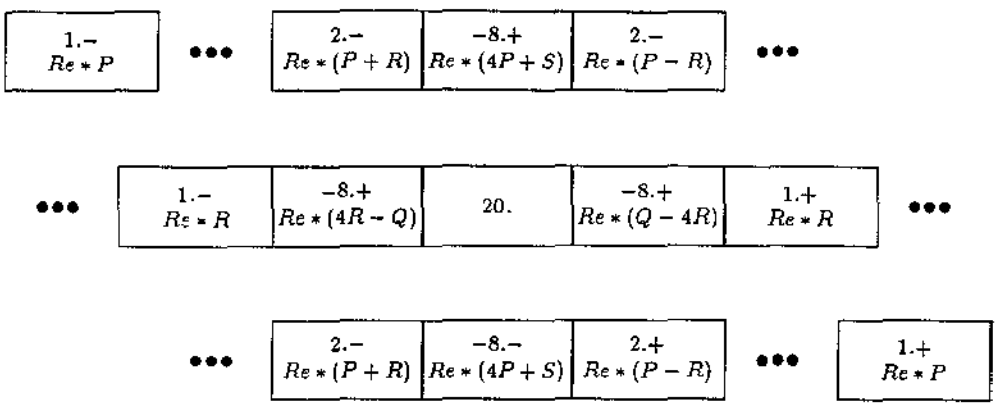

Figura 3.3: Estrutura da matriz jacobiana

Obtivemos várias sequências de soluções para $Re \in [0, 11000]$, usando o procedimento descrito anteriormente, fixando distintos valores para o passos ΔRe. O extremo superior do intervalo foi determinado pela proximidade de um ponto limite, indicado por um súbito aumento do número de iterações do algoritmo usado, com o aparecimento de pronunciadas distorções nos gráficos das soluções.

Uma outra série de testes foi realizado com $Re \in [0, 500]$, com diversos passos ΔRe utilizando as técnicas de globalização como "backtracking" (em forma análoga ao feito no Capítulo 2) e região de confiança (da forma que foi descrito no Capítulo 1).

O esquema de discretização configurado, permite trabalhar com malhas de até aproximadamente 100×100 divisões. Este limitante foi obtido de forma heurística, observando a evolução dos gráficos das curvas de nível das soluções e refinando a malha gradualmente.

A escolha do tamanho da malha responde principalmente à possibilidade de realizar comparações dos resultados obtidos para igual dimensão, com os apresentados por vários outros autores (Axelsson [1993] , Deuflhard [1991],Walker [1992], etc).

Por outro lado, os resultados obtidos com testes sob malhas mais refinadas praticamente não mostraram comportamento muito diferentes aos da malha selecionada.

Os algoritmos testados são os descritos na Capítulo 1, contidos nos pacotes computacionais ROUXINOL, NI-GMRES e BOX-QUACAN .

Considerando que os métodos tipo Newton Inexatos tiveram um desempenho desencorajador, os testes com estes métodos tiveram que ser realizados enfraquecendo as exigências para conseguir convergência, com tempos reais de execução razoáveis. Assim, mostraremos um pequeno conjunto de resultados com NI-GMRES, com a única finalidade de mostrar o efeito que produz o uso dos distintos precondicionadores sobre os sistemas lineares.

Apresentaremos uma análise mais exaustiva sobre um conjunto de resultados obtidos com os métodos Quase-Newton implementados no pacote computacional ROUXINOL, que podem

ser considerados como muito bem sucedidos.

Para este último caso, e com o objetivo de realizar comparações entre alguns dos métodos implementados nesse pacote, geramos inicialmente uma sequência de soluções usando o método de Newton, usando um critério de parada sobre o resíduo : $\| F(\psi, Re) \| \leq 10^{-10}$. Posteriormente, foram gerados os gráficos respectivos, das curvas de nível das soluções, que foram padronizados como representando as "soluções verdadeiras". Alguns destes gráficos para $Re = 0, 1000, 5000$ e 11000, são mostrados no Apêndice.

É conveniente aclarar que as curvas correspondentes a $Re = 0$ devem ser consideradas como uma solução assintótica sem qualquer significado físico. Numa situação real, para este valor do número de Reynolds deveriamos ter considerado simultaneamente $\psi_y(x, 1) = 1,\ 0 \leq x \leq 1$ obtendo-se assim $\psi \equiv 0$.

A validade destas soluções está sustentada pela comparação, para distintos números de Reynolds, com as apresentadas por Ghia, Ghia e Shin [1982], usando uma malha de 256×256 e Benjamin e Denny [1973] para uma malha de 151×151, considerando a precisão da nossa aproximação.

Para garantir que as soluções originadas pelos métodos Quase-Newton fossem, no mínimo, "tão boas" como as obtidas com o método de Newton, foi repetido o mesmo processo descrito acima, confrontando os gráficos de cada uma das soluções que compõem a sequência.

A "não-convergência" de qualquer dos métodos foi estabelecida fixando o número máximo de iterações tipo-Newton.

3.6.1 Métodos Newton e Quase-Newton

Nas Tabelas (3.1) e (3.2) são apresentados um conjunto de resultados para $Re \in [0, 11000]$, com passos $\Delta Re = 250$ e $\Delta Re = 500$ respectivamente.

Cada linha, que representa um experimento para um determinado número de Reynolds, indica os resultados para cada método através de dois ou três números (quando corresponde): iterações de Newton, iterações Quase-Newton e tempo de execução (escalado). Em tempo real, a execução para um experimento particular, por exemplo para $Re = 11000$ usando o método de Newton, demandou pouco mais de 10 minutos.

Para $\Delta Re = 1000$ todos os métodos excederam o número máximo de iterações, e as experiências realizadas com passos menores necessitaram tempos de execução consideravelmente maiores.

Podemos observar que todos os métodos usam somente uma iteração para resolver $F(\psi, 0) = 0$, já que o sistema é linear.

Em todos os casos, a primeira iteração é uma iteração de Newton. Desta maneira, é forçada a realização de uma fatorização LU completa da matriz Jacobiana.

Pode ser observado a demanda de um esforço um pouco maior para atingir $Re = 1500$. Ultrapassado este valor, as iterações continuam com uma demanda do tempo de execução uniforme, até o fim do intervalo.

Em termos de custo computacional, todos os métodos Quase-Newton têm um custo apro-

Número de Reynolds	Newton	Newton Modificado	Broyden	CUM
0	(1 , 188.60)	(1 , 0 , 192.61)	(1 , 0 , 188.32)	(1 , 0 , 189.80)
250	(5 , 942.56)	(1 , 56 , 368.38)	(1 , 23 , 264.70)	(1 , 22 , 258.69)
500	(5 , 756.39)	(1 , 18 , 244.80)	(1 , 13 , 230.61)	(1 , 13 , 230.06)
750	(5 , 754.04)	(1 , 14 , 232.29)	(1 , 8 , 214.96)	(1 , 8 , 213.83)
1000	(5 , 754.02)	(1 , 9 , 217.14)	(1 , 6 , 207.83)	(1 , 7 , 210.76)
1250	(3 , 572.77)	(1 , 7 , 213.18)	(1 , 6 , 210.82)	(1 , 6 , 210.44)
1500	(3 , 585.68)	(1 , 7 , 214.27)	(1 , 5 , 211.31)	(1 , 5 , 210.68)
1750	(3 , 594.76)	(1 , 5 , 214.00)	(1 , 5 , 214.57)	(1 , 5 , 214.14)
2000	(3 , 598.95)	(1 , 5 , 215.51)	(1 , 4 , 212.97)	(1 , 4 , 212.49)
2250	(3 , 602.59)	(1 , 5 , 216.90)	(1 , 5 , 216.72)	(1 , 5 , 216.61)
2500	(3 , 607.52)	(1 , 5 , 219.08)	(1 , 5 , 219.01)	(1 , 5 , 219.47)
2750	(3 , 608.14)	(1 , 5 , 219.10)	(1 , 5 , 218.28)	(1 , 5 , 218.83)
3000	(3 , 613.12)	(1 , 5 , 219.66)	(1 , 5 , 219.74)	(1 , 5 , 219.83)
3250	(3 , 616.41)	(1 , 5 , 220.96)	(1 , 5 , 221.08)	(1 , 5 , 221.17)
3500	(3 , 615.55)	(1 , 4 , 218.14)	(1 , 4 , 218.13)	(1 , 4 , 218.35)
3750	(3 , 619.21)	(1 , 4 , 219.01)	(1 , 4 , 218.88)	(1 , 5 , 222.19)
4000	(3 , 619.82)	(1 , 4 , 218.86)	(1 , 4 , 219.20)	(1 , 4 , 218.74)
4250	(3 , 620.91)	(1 , 4 , 219.88)	(1 , 4 , 219.86)	(1 , 4 , 219.96)
4500	(3 , 621.30)	(1 , 4 , 219.64)	(1 , 4 , 219.82)	(1 , 4 , 220.03)
4750	(3 , 624.95)	(1 , 4 , 220.39)	(1 , 4 , 223.28)	(1 , 4 , 220.47)
5000	(3 , 621.63)	(1 , 4 , 221.01)	(1 , 4 , 223.61)	(1 , 4 , 220.64)
5250	(3 , 625.65)	(1 , 4 , 221.22)	(1 , 4 , 221.28)	(1 , 4 , 220.89)
5500	(3 , 627.07)	(1 , 4 , 221.32)	(1 , 4 , 221.57)	(1 , 4 , 221.15)
5750	(3 , 628.12)	(1 , 4 , 221.77)	(1 , 4 , 222.34)	(1 , 4 , 221.90)
6000	(3 , 627.15)	(1 , 4 , 221.28)	(1 , 4 , 221.64)	(1 , 4 , 220.98)
6250	(3 , 628.21)	(1 , 4 , 222.28)	(1 , 4 , 222.50)	(1 , 4 , 221.65)
6500	(3 , 631.79)	(1 , 4 , 222.58)	(1 , 4 , 222.61)	(1 , 4 , 222.69)
6750	(3 , 631.67)	(1 , 4 , 222.21)	(1 , 4 , 223.25)	(1 , 4 , 222.57)
7000	(3 , 629.86)	(1 , 4 , 222.26)	(1 , 3 , 219.63)	(1 , 3 , 219.23)
7250	(3 , 632.56)	(1 , 4 , 222.23)	(1 , 3 , 219.41)	(1 , 3 , 219.19)
7500	(3 , 630.92)	(1 , 4 , 223.35)	(1 , 3 , 220.57)	(1 , 3 , 220.19)
7750	(3 , 645.42)	(1 , 3 , 220.05)	(1 , 3 , 220.21)	(1 , 3 , 220.86)
8000	(3 , 642.68)	(1 , 3 , 220.33)	(1 , 3 , 219.93)	(1 , 3 , 220.46)
8250	(3 , 633.28)	(1 , 3 , 220.31)	(1 , 3 , 221.29)	(1 , 3 , 220.16)
8500	(3 , 638.68)	(1 , 3 , 220.44)	(1 , 3 , 220.83)	(1 , 3 , 220.29)
8750	(3 , 632.14)	(1 , 3 , 220.54)	(1 , 3 , 219.56)	(1 , 3 , 220.55)
9000	(3 , 642.20)	(1 , 3 , 221.26)	(1 , 3 , 219.56)	(1 , 3 , 220.92)
9250	(3 , 644.13)	(1 , 3 , 220.89)	(1 , 3 , 221.24)	(1 , 3 , 221.05)
9500	(3 , 644.02)	(1 , 3 , 220.90)	(1 , 3 , 221.61)	(1 , 3 , 220.83)
9750	(3 , 636.24)	(1 , 3 , 220.82)	(1 , 3 , 221.19)	(1 , 3 , 221.18)
10000	(3 , 637.02)	(1 , 3 , 220.87)	(1 , 3 , 220.98)	(1 , 3 , 220.73)
10250	(3 , 635.66)	(1 , 4 , 224.31)	(1 , 3 , 221.59)	(1 , 3 , 221.42)
10500	(3 , 636.71)	(1 , 4 , 224.21)	(1 , 4 , 224.97)	(1 , 4 , 224.55)
10750	(3 , 636.46)	(1 , 4 , 224.70)	(1 , 4 , 225.12)	(1 , 4 , 224.85)
11000	(3 , 635.67)	(1 , 4 , 224.37)	(1 , 4 , 224.71)	(1 , 4 , 224.43)

Tabela 3.1: Métodos Newton e Quase-Newton, $\Delta Re = 250$

ximadamente 50 vezes menor que o método de Newton (uma iteração de Newton requer em média 67 segundos). Esta relação é mantida ao longo de todo o intervalo.

Para $\Delta Re = 500$ apenas os métodos de Newton e Broyden conseguem convergência. CUM demandou um pouco mais de 100 iterações para $Re = 250$. Nesta tabela, também são incluidos resultados com a opção de recomeços com um desempenho um pouco pior, quando comparados com os obtidos sem usar esta opção. A opção de recomeços, no método de Newton Modificado, não foi acionada uma única vez, devido a que a condição de decréscimo suficiente foi sempre satisfeita. Em todos os casos, a demanda do custo computacional foi aproximadamente de 60% que para $\Delta Re = 250$.

Número de Reynolds	Newton	Broyden	
		Com Recomeços	Sem Recomeços
0	(1 , 192.89)	(1 , 0 , 188.64)	(1 , 0 , 188.89)
500	(6 , 1150.78)	(3 , 10 , 597.90)	(1 , 79 , 477.47)
1000	(5 , 961.07)	(2 , 7 , 399.75)	(1 , 13 , 231.25)
1500	(4 , 782.63)	(1 , 8 , 221.82)	(1 , 8 , 222.85)
2000	(3 , 599.87)	(1 , 7 , 222.85)	(1 , 7 , 223.33)
2500	(3 , 607.67)	(1 , 7 , 226.00)	(1 , 7 , 227.24)
3000	(3 , 612.56)	(1 , 7 , 227.10)	(1 , 7 , 227.90)
3500	(3 , 617.13)	(1 , 6 , 225.96)	(1 , 6 , 226.38)
4000	(3 , 619.79)	(1 , 6 , 226.67)	(1 , 6 , 228.04)
4500	(3 , 623.49)	(1 , 6 , 227.39)	(1 , 6 , 227.82)
5000	(3 , 623.93)	(1 , 6 , 227.97)	(1 , 6 , 229.29)
5500	(3 , 627.64)	(1 , 6 , 228.46)	(1 , 6 , 228.93)
6000	(3 , 628.09)	(1 , 5 , 225.97)	(1 , 5 , 226.96)
6500	(3 , 629.64)	(1 , 5 , 226.67)	(1 , 5 , 226.55)
7000	(3 , 630.71)	(1 , 5 , 225.97)	(1 , 5 , 227.41)
7500	(3 , 632.49)	(1 , 5 , 227.15)	(1 , 5 , 227.94)
8000	(3 , 634.03)	(1 , 5 , 227.57)	(1 , 5 , 228.02)
8500	(3 , 633.93)	(1 , 4 , 224.19)	(1 , 4 , 224.61)
9000	(3 , 636.06)	(1 , 4 , 224.72)	(1 , 4 , 225.31)
9500	(3 , 635.11)	(1 , 4 , 224.78)	(1 , 4 , 225.14)
10000	(3 , 636.89)	(1 , 5 , 227.78)	(1 , 5 , 228.21)
10500	(3 , 637.49)	(1 , 5 , 228.83)	(1 , 5 , 228.62)
11000	(3 , 638.72)	(1 , 6 , 232.29)	(1 , 6 , 232.74)

Tabela 3.2: Métodos Newton e Quase-Newton, $\Delta Re = 500$

Outros métodos Quase-Newton como: Escalamento na Diagonal e Escalamento na Coluna não conseguiram convergência sequer para $Re = 250$.

Os métodos Quase-Newton com Jacobiano Truncado, mostraram-se muito sensíveis à introdução de um Jacobiano inicial "falso" com resultados negativos.

Outra série de testes foram realizados para avaliar o efeito de introduzir estratégias de globalização. A globalização por "backtracking" não trouxe praticamente vantagens, somente possibilitou a convergência para Newton Modificado entretanto prejudicou a convergência de CUM para $Re = 100$. A introdução da técnica de região de confiança, implementada com Newton Inexato, não conseguiu melhorar o desempenho do algoritmo em nenhum caso.

Estes últimos testes foram realizados em uma SPARCstation-5.

Método		$\Delta Re = 250$	$\Delta Re = 500$
Newton		(141,0,7.881)	(73,0,4.164)
Newton Mod.		(45,260,2.797)	(Não converge)
Broyden	Com Rec.	(46,192,2.795)	(26,129,1.587)
	Seem Rec.	(45,205,2.753)	(23,204,1.511)
CUM	Com Rec.	(47,185,2.835)	(Não Converge)
	Sem Rec.	(45,206,2.750)	

Tabela 3.3: Métodos Newton e Quase-Newton com e sem opções de recomeços

3.6.2 Métodos Newton Inexatos - GMRES Precondicionados.

Os testes realizados com as diferentes opções que podem ser selecionadas usando estes métodos, foram desalentadores quando comparados com o desempenho dos métodos Quase-Newton. Nas condições exigidas para estes últimos, não se obteve convergência em nenhum caso.

Com o objetivo de avaliar a eficiência dos diferentes precondicionadores implementados, os requerimentos sobre distintos parâmetros tiveram que ser relaxados, para obter convergência. Para isso, o tamanho do passo foi reduzido para $\Delta Re = 50$ e a exigência sobre a diminuição no valor do residuo foi aumentada para $\| F(\psi, Re) \| \leq 10^{-5}$. Desta forma, conseguimos obter um conjunto de resultados, que permitiu realizar uma análise mínima. A mudança no valor daquele último parâmetro, impede fazer qualquer tipo de comparação com os resultados obtidos com os métodos Quase-Newton, porque a convergência é obtida para soluções aproximadas diferentes.

Em uma das experiências relativamente "bem sucedidas", obtivemos convergência até $Re = 9100$ usando um precondicionador baseado na Fatoração Incompleta, com $\theta_k = 0.1$.

Uma tentativa para estimar qualitativamente a eficácia do precondicionador ao longo do intervalo de convergência, foi feita calculando o quociente entre os valores acumulados do número de iterações dos laços interno e externo em cada intervalo $Re = Re + 1000$. Os resultados mostram que a eficácia do precondicionador vai decaindo paulatinamente a cada intervalo. Duas causas inter-relacionadas que contribuem no mesmo sentido, podem explicar este comportamento: a perda de diagonal dominância na matriz Jacobiana e a consequente perda na qualidade do precondicionador, o qual está sendo reconstruido sobre uma matriz cada vez pior condicionada [1]. A irregularidade que se produz no intervalo $[1000 - 2000]$, coincide com a apontada anteriormente usando os métodos diretos.

Número de Reynolds	Sem Precondicionamento		Precondicionado	
	$\theta_k = 0.1$	$\theta_k = 0.67$	$\theta_k = 0.1$	$\theta_k = 0.67$
0	(12 , 1114 , 1.0)	(22 , 1208 , 1.1)	(6 , 461 , 0.42)	(15 , 354 , 0.38)
50	(5 , 438 , 0.39)	(9 , 373 , 0.36)	(3 , 230 , 0.21)	(8 , 197 , 0.21)
100	(6 , 544 , 0.49)	(11 , 594 , 0.56)	(3 , 234 , 0.21)	(8 , 182 , 0.20)
150	(7 , 648 , 0.58)	(12 , 567 , 0.54)	(3 , 239 , 0.22)	(8 , 231 , 0.24)
200	(7 , 652 , 0.59)	(12 , 677 , 0.63)	(3 , 245 , 0.22)	(8 , 258 , 0.26)
250	(8 , 754 , 0.68)	(13 , 682 , 0.64)	(3 , 250 , 0.23)	(8 , 262 , 0.26)
300	(10 , 955 , 0.86)	(15 , 977 , 0.90)	(3 , 252 , 0.23)	(8 , 262 , 0.26)
350	(11 , 1053 , 0.94)	(15 , 977 , 0.90)	(3 , 255 , 0.23)	(8 , 275 , 0.28)
400	(12 , 1152 , 1.0)	(15 , 988 , 0.91)	(3 , 258 , 0.23)	(8 , 281 , 0.28)
450	(13 , 1252 , 1.1)	(17 , 1256 , 1.1)	(4 , 364 , 0.33)	(9 , 385 , 0.37)
500	(16 , 1151 , 1.1)	(19 , 1462 , 1.3)	(4 , 361 , 0.32)	(8 , 282 , 0.28)
550	(13 , 1250 , 1.1)	(22 , 1764 , 1.6)	(4 , 366 , 0.33)	(9 , 286 , 0.29)
600	(20 , 1951 , 1.7)	(20 , 1548 , 1.4)	(3 , 272 , 0.24)	(8 , 281 , 0.28)
650	(20 , 1950 , 1.7)	(28 , 2354 , 2.1)	(5 , 498 , 0.44)	(9 , 426 , 0.41)
700	(21 , 2140 , 1.9)	(18 , 1215 , 1.1)	(3 , 283 , 0.25)	(8 , 317 , 0.31)
750	(23 , 2249 , 2.0)	(36 , 3158 , 2.8)	(5 , 500 , 0.45)	(10 , 540 , 0.51)
800	(28 , 2749 , 2.5)	(23 , 1831 , 1.7)	(4 , 400 , 0.36)	(9 , 425 , 0.41)
850	(22 , 2148 , 1.9)	(49 , 4441 , 4.0)	(4 , 400 , 0.36)	(11 , 641 , 0.60)
900	(41 , 4048 , 3.6)	(24 , 1926 , 1.7)	(5 , 500 , 0.45)	(9 , 441 , 0.42)
950	(22 , 2149 , 1.9)	(63 , 5836 , 5.2)	(4 , 400 , 0.36)	(9 , 444 , 0.42)
1000	(68 , 6748 , 6.0)	(28 , 2317 , 2.1)	(4 , 400 , 0.36)	(14 , 949 , 0.87)
Resultados Globais	(385 , 37095 , 1.0)	(471 , 36151 , 0.99)	(79 , 7168 , 0.19)	(192 , 7719 , 0.22)

Tabela 3.4: Método de Newton Inexato, $\Delta Re = 50$

Com a mesma finalidade, uma outra série de testes foi realizado apenas no intervalo $Re \in [0, 1000]$, usando o mesmo precondicionador, com o objeto de avaliar o efeito de se mudar o parâmetro θ_k. Os resultados são mostrados na Tabela (3.4). Para cada valor do número de Reynolds indicamos: número de iterações dos laços externo e interno e o tempo de execução (escalado), apenas para dois diferentes valores de θ_k, que são os mais representativos sobre um conjunto bem mais numeroso de testes. Podemos apreciar uma pequena vantagem para o menor valor de θ_k. Em tempo real, a execução para percorrer completamente o intervalo, demandou mais de 4 horas.

Sem o uso de qualquer precondicionador não se obteve convergência e os outros precondicionadores foram menos eficientes.

3.7 Conclusões e trabalhos futuros.

O estudo realizado neste Capítulo representa um esforço para definir um marco de referência na solução das equações de Navier-Stokes em termos da função-corrente, usando diversos Métodos tipo-Newton, para um problema que modela o fluxo em uma cavidade. para altos números de Reynolds.

Um conjunto numeroso de soluções discretas, em termos da função-corrente, foi obtido para números de Reynolds variando entre $0 - 11000$. Este conjunto foi convalidado por comparação com as soluções obtidas por outros pesquisadores.

Uma estimativa do desempenho e eficiência destes Métodos, orientada para a avaliação do esforço computacional demandado, objeto principal deste trabalho, foi conseguida a partir da realização de inúmeras experiências, usando distintos pacotes computacionais que implementam aqueles métodos . A fixação dos parâmetros próprios de cada pacote, que eventualmente aprimoraram sua eficiência, foram determinados através de extensos e minuciosos testes; alguns dos quais foram explícitamente mostrados.

Os Métodos Quase-Newton se mostraram robustos e os mais eficazes. Não há significativas diferenças entre eles, entretanto exibiram um desempenho marcadamente superior ao Método de Newton.

A seleção de maiores passos, para incrementar o número de Reynolds, tem-se mostrado como a mais indicada em relação a economia do custo computacional global. O aumento do tamanho do passo está associado ao problema de melhorar a aproximação inicial, para a resolução do sistema correspondente a cada número de Reynolds. Isto sugere a introdução de Técnicas de Continuação mais apuradas, implementadas com subrutinas para a determinação automática do tamanho do passo.

Os Métodos tipo-Newton Inexatos com Precondicionamento não são competitivos; a inclusão e comentários acima dos resultados obtidos com este método pretendem apenas mostrar o desempenho de métodos iterativos na resolução de problemas com esta estrutura. Mesmo com uma redução nas exigências estabelecidas como critério para aceitação da solução, requereram, no melhor dos casos, tempos de execução práticamente inaceitáveis.

Os resultados obtidos mediante a formulação definida pela minimização de $f(\mathbf{x}) = \| \mathbf{F}(\mathbf{x}) \|_2^2$, usando o pacote BOX-QUACAN foram definitivamente desencorajadores; não se obteve convergência sequer para $Re = 0$.

Uma tentativa para avaliar a eficácia do uso dos precondicionadores sobre os sistemas originados na linearização de $\mathbf{F}(\mathbf{x}) = \mathbf{0}$, tampouco produziram uma melhoria significativa, porém tendo como fato relevante, a consecução de convergência. O pacote utilizado neste caso, NIGMRES foi testado com Precondicionadores Secantes e com uma Fatoração LU Truncada, sendo esta última, a opção melhor sucedida. Uma modificação na configuração do esquema de discretização do termo não linear, que atenuasse ou evitasse o paulatino crescimento do mau-condicionamento possibilitaria, em princípio, tornar o método mais competitivo.

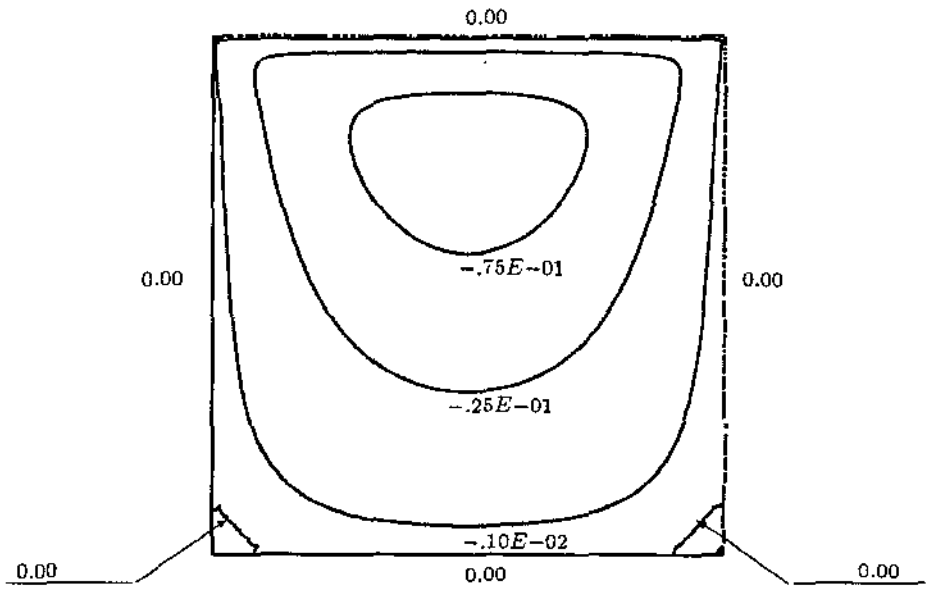

Figura 3.4: Vórtices para Reynolds= 0

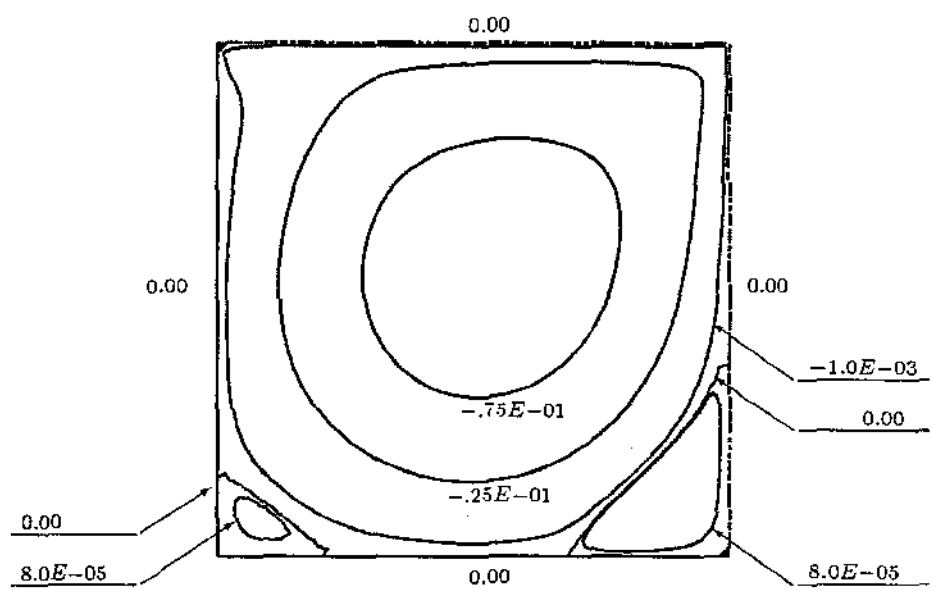

Figura 3.5: Vórtices para Reynolds= 1000

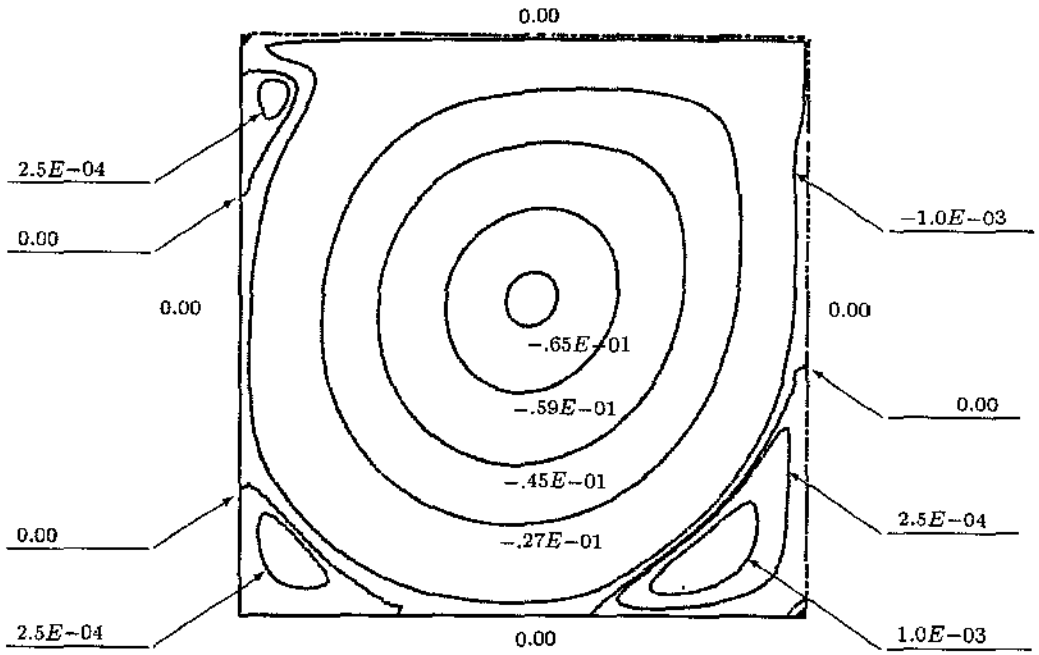

Figura 3.6: Vórtices para Reynolds= 5000

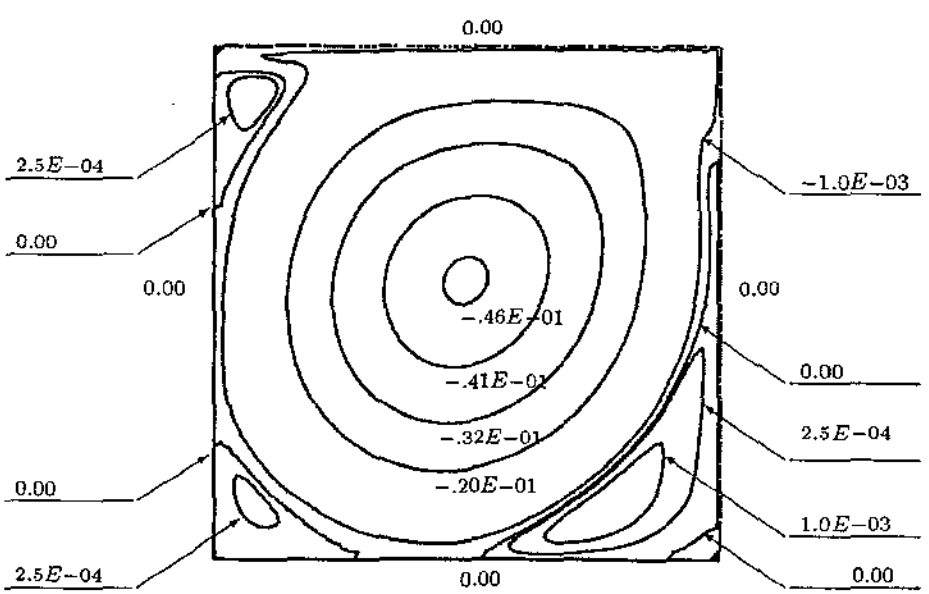

Figura 3.7: Vórtices para Reynolds = 11000

Bibliografia

[1] Axelsson, O., Kaporin, I. E. [1993]: *On computer implementation of Inexact-Newton-Conjugate Gradient-type algorithms.* Preprint.

[2] Benjamin, A. S., Denny, V. E. [1973]: *On the convergence of numerical solutions for 2-D flows in a cavity at hight Re.* J. Comput. Phys **12**, pp. 348-358.

[3] Bercovier, M., Pironneau, O. [1979]: Numerical Math. **33**, pp.211-224.

[4] Brown, P. N., Saad, Y. [1990]: *Hybrid Krilov Methods for Nonlinear Systems of Equations.* SIAM J.Sci. Statist. Comput. **11**, pp. 450-481.

[5] Collatz, L. [1973]: *Numerical Treatment of Differential Equations.* Springer-Verlag. Berlin.

[6] Crank, J., Furzeland, R. M. [1978]: *The numerical solution of elliptic and parabolic partial differential equations with boundary singularities.* J. Comput. Physics **26**, pp. 285-296.

[7] Crochet, M. J., Davies, A. R., Walters,K. [1984]: *Numerical simulation of non-newtonian flow.* Rheology series **1**. Elsevier.

[8] Deuflhard, P. [1991]: *Global Inexact Newton Methods for very large scale nonlinear problems.* Impact of Comp. in Sc. and Eng. **3**, pp. 366-393.

[9] Eisenstat,S. C., Walker, H. F. [1994]: *Globally convergent inexact Newton methods.* To appear in SIAM Journal on Optimization.

[10] Fox, L., Sankar, R.[1969]: *Boundary singularities in linear elliptic differential equations.* J. Inst. Math. Applied **5**, pp. 340-350.

[11] Ghia, U., Ghia, K. N., Shin, C.T.[1982]: *High-Re Solutions for Incompressible Flow Using the Navier-Stokes Equations and a Multigrid Method.* J. Comput. Phys.**48**, pp. 387-411.

[12] Glowinski, R. [1984]: *Numerical Methods for Nonlinear Variational Problems* 2nd ed. Springer-Verlag. N.Y.

[13] Greenspan, D. [1969]: *Numerical solution of prototype cavity flow problems.* Comput. J. **12**.

[14] Holstein, H., Paddon, D.J. [1981]: *A singular finite difference treatment of re-entrant corner flow.* Part I.Newtonian Fluids.J.non-Newtonian Fluid Mech. **8**, pp. 81-93.

[15] Kubicek, M., Hlavacek, V. [1975]: *Solution of nonlinear boundary-value problems.* IX.Chem.Eng.Sci. **30**, pp. 1439-1440.

[16] Matthies, H., Strang, G. [1979]: *The solution of nonlinear finite element equations.* Int. J. Num. Meth. Eng. **14**, pp. 1613-1626.

[17] Mittelmann, H. D., Roose, D. (Eds.) [1990]: *Continuation Techniques and Bifurcation Problems.* Int. Series of Num. Math., Vol **92**.

[18] Moffat, H. K. [1964]: *Viscous and resistive eddies near a sharp corner.* J. Fluid Mech. **18**, pp. 1-18.

[19] Olson, M. D., Tuan, S. -Y. [1981]: Comput. and Fluids bf 7, pp. 123-135.

[20] Oden, J.T. [1972]: *Finite element of nonlinear continua.* New York. McGraw-Hill.

[21] Peyret, R, Taylor, T. [1985]: *Computational methods for fluid flow.* Springer Verlag.

[22] Rheinboldt, W. C. [1986]: *Numerical analysis of parametrized nonlinear equations.* University of Arkansas Lectures notes in the mathematical sciences, **7**

[23] Richtmeyer, R. D., Morton, K. W. [1967]: *Difference methods for initial value problems.* Interscience. Publishers. N.Y.

[24] Schreiber, R., Keller, H. B. [1983]: *Driven cavity flows by efficient numerical techniques.* J. Comput. Phys. **49**, pp. 310-333.

[25] Schreiber, R., Keller, H. B. [1983]: *Spurious Solutions in Driven Cavity Calculations.* J. Comput. Phys. **49**, pp. 165-172.

[26] Smith, G.D. [1987]: *Numerical solutions of partial differential equations: Finite differences methods.* Clarendon.

[27] Walker, H. F. [1992] : *A GMRES-backtracking Newton iterative method.* Proceeding of the Copper Conference on Iterative Methods.

Capítulo 4

Determinação de Pontos Singulares com Métodos Newton-Inexatos

4.1 Introdução

Neste Capítulo, utilizaremos o método Newton-Inexato para a determinação de *pontos singulares* situados sobre uma curva homotópica (ver Cap. 1). Estes pontos estão relacionados intimamente com a estabilidade e multiplicidade das soluções.

Consideremos o seguinte problema não linear de autovalor

$$H(y,t) = 0, \tag{4.1}$$

onde $H : I\!R^{m+1} \to I\!R^m$, $y \in I\!R^m$, $t \in I\!R^1$.

Usualmente $y = y(t)$ é considerada uma solução de (4.1) já que nas aplicações físicas o autovalor t representa um parâmetro de especial interesse (por exemplo a carga sobre uma estrutura, a tensão em um circuito, etc).

Se (y_0, t_0) é uma solução de (4.1) e a matriz Jacobiana $m \times m$, $H_y(y,t)$ é inversível, é possível garantir a existência de uma única curva solução (y,t) que passe por (y_0, t_0) de tal forma a explicitar $y(t_0) = y_0$.

Definimos $\Gamma = \{(y,t) \in I\!R^m \times I\!R | H(y,t) = 0\}$.

Um *ponto singular* é um ponto de Γ onde $H_y(y,t)$ é singular. Quando as linhas de $H'(y,t) = H_y(y,t), H_t(y,t)$ são linearmente independentes, isto é, $H_t(y,t) \notin \mathcal{R}[H_y(y,t)]$ (a imagem de $H_y(y,t)$), o ponto singular é denominado *ponto de retorno.* Existe uma única curva solução que passa por esse ponto, porém a dy/dt é infinita e uma pequena variação de t produz um aumento desproporcionadamente grande em $\|y\|$. Uma situação típica é mostrada na Figura 4.1.

Vários métodos têm sido propostos para a determinação de *pontos de retorno.* A idéia básica consiste em acrescentar uma o mais equações ao sistema (4.1) tal que a solução do

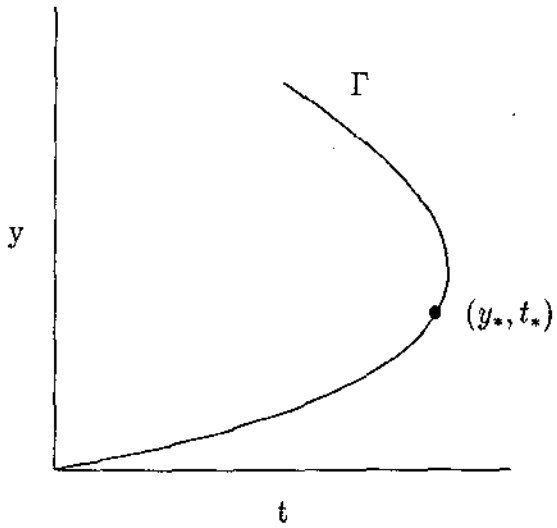

Figura 4.1: Ponto de retorno

sistema aumentado

$$\left.\begin{array}{l} H(y,t) = 0 \\ H_y(y,t)v = 0 \\ l(v) - 1 = 0 \end{array}\right\} \tag{4.2}$$

seja um *ponto de retorno* e de modo a garantir uma matriz Jacobiana não-singular para este novo sistema (ver [15], [14], [1]).

Todos estes métodos usam algum tipo de fatoração de matrizes, o que é inconveniente em problemas de grande porte.

O Método de Newton-Inexato será usado para resolver

$$F(x) = 0, \tag{4.3}$$

onde $F : \mathbb{R}^n \to \mathbb{R}^n$ é a função que aproxima o sistema aumentado 4.2, cuja formulação é chave do presente trabalho.

Na próxima seção introduziremos algoritmos globalmente convergentes; na Seção 3 mostraremos os novos sistemas aumentados, cujas soluções são os *pontos singulares* de (4.1). Na Seção 4 apresentaremos os problemas testes conjuntamente com as experiências numéricas realizadas, utilizando um método de Newton-Inexato globalizado para resolver os sistemas mostrados na Seção 3. A maioria dos problemas foram selecionados da coleção de Melhem e Rheinboldt [14]. As conclusões serão mostradas na Seção 5.

4.2 Algoritmos globalmente convergentes

Introduzimos algoritmos globalmente convergentes para resolver (4.3) cujas direções são geradas pelo método de Newton-Inexato.

Nossa abordagem é similar à de Eisenstat e Walker [3] e de Martínez e Qi [11] porém sendo mais geral, desde que podem ser consideradas estratégias não necessariamente baseadas em buscas lineares.

Consideramos para isso a soma do quadrados de $F(x)$ como a função de mérito

$$f(x) = \frac{1}{2}\|F(x)\|^2, \tag{4.4}$$

e um algoritmo que reduz monotonamente $f(x_k)$. No que se segue, $\|.\|$ representa a norma Euclideana.

Algoritmo 4 Minimização Monótona

Suponhamos que: $\sigma \in (0,1)$, $\gamma \in (0,1]$, $\eta_1, \eta_2 \in (0,1)$, $\eta_1 < \eta_2$ *sejam dados independentemente de* k. $x_0 \in \mathbb{R}^n$ *seja uma aproximação inicial arbitrária e* $\alpha_0 = 1$.

Dado $x_k \in \mathbb{R}^n$, $\alpha_k \in (0,1]$, *os passos para obter* x_{k+1}, α_{k+1} *são:*

- Passo 1.
  Escolher
  $$d_k \in \mathbb{R}^n. \tag{4.5}$$
- Passo 2.
  Se
  $$f(x_k + \alpha_k d_k) < f(x_k) \tag{4.6}$$
  calcular $x_{k+1} = x_k + \alpha_k d_k$. *Se (4.6) não for satisfeita, definir* $x_{k+1} = x_k$.
- Passo 3.
  Se
  $$f(x_{k+1}) \leq (1 - \sigma\gamma\alpha_k) f(x_k) \tag{4.7}$$
  definir $\alpha_{k+1} = 1$. *Senão, escolher*
  $$\alpha_{k+1} \in [\eta_1\alpha_k, \eta_2\alpha_k]. \tag{4.8}$$

O algoritmo acima é muito geral. Nenhuma condição é exigida sobre as direções d_k e até direções nulas $d_k \equiv 0$ podem ser aceitas.

Impondo condições sobre as direções d_k é possível provar interessantes resultados relativos à convergência, que terão implicações de ordem prática.

Teorema 10 Convergência Global

Suponhamos que $\{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}$ *seja limitada. Seja* $\{x_k\}$ *uma sequência gerada pelo algoritmo anterior. Suponhamos que exista* $M > 0$ *tal que para todo* $k = 0, 1, 2, \ldots,$

$$\|d_k\| \leq M \tag{4.9}$$

e

$$\langle J(x_k)d_k, F(x_k)\rangle \leq -\frac{\gamma}{2}\|F(x_k)\|^2 . \tag{4.10}$$

Então

(a) Qualquer ponto limite x_* *de* $\{x_k\}$ *satisfaz* $F(x_*) = 0$.

(b) Se um ponto limite x_* *é uma solução isolada de (4.3) e* $\alpha_k \to 0$, *então* $\{x_k\}$ *converge a* x_*.

(c) Se um ponto limite x_* *é uma solução isolada de (4.3) e existe* $\beta > 0$ *tal que* $\|d_k\| \leq \beta\|F(x_k)\|$ *para todo* $k = 0, 1, 2, \ldots,$ *então* $\{x_k\}$ *converge a* x_*.

Prova: (Ver Kozakevich, Martínez e Santos [9]) □

Essencialmente, o teorema estabelece que se for possível calcular direções de busca d_k tais que (4.9) e (4.10) sejam satisfeitas em cada iteração, então garante-se convergência global para a solução do sistema. Se $J(x_k)$ é não singular, a direção de Newton $d_k^N = -J(x_k)^{-1}F(x_k)$ satisfaz (4.10) com $\gamma = 1$. Em geral, se d_k satisfaz

$$\|J(x_k)d_k + F(x_k)\|^2 \leq t\|F(x_k)\|^2, \tag{4.11}$$

com $t \in [0, 1)$ temos que

$$\langle J(x_k)d_k, J(x_k)d_k\rangle + 2\langle J(x_k)d_k, F(x_k)\rangle \leq (t-1)\|F(x_k)\|^2$$

Assim,

$$\langle J(x_k)d_k, F(x_k)\rangle \leq \frac{t-1}{2}\|F(x_k)\|^2,$$

isto é, a condição (4.10) é satisfeita com $\gamma = 1 - t$. A condição (4.11) é a "versão quadrática" do critério clássico para definir a iteração de Newton-Inexato.

Vemos, em virtude deste teorema, que quando o método de Newton (ou a sua generalização para Newton-Inexato) não converge, usando a globalização dada pelo Algoritmo 4, então a sequência de direções d_k geradas é ilimitada. Neste caso, o método criará uma sequência que tende para um ponto onde o Jacobiano é singular. Este ponto não é necessariamente um minimizador local, ou ainda nem um ponto estacionário de $f(x)$.

4.3 Implementação

Nesta seção descrevemos a implementação do Algoritmo 4. Basicamente, em cada iteração escolhemos, $s_k \equiv \alpha_k d_k$ como sendo um *minimizador aproximado* de

$$\psi(s) \equiv \frac{1}{2}\|J(x_k)s + F(x_k)\|^2$$

sobre uma região de confiança apropriada (ver Fletcher [4]) da forma $\|s\|_\infty \leq \Delta$. Se 0 não for um minimizador de ψ, isto é, $J(x_k)^T F(x_k) \neq 0$, deverá ser possível obter s_k tal que

$$\|J(x_k)s_k + F(x_k)\|^2 < \|F(x_k)\|^2$$

o que implica que

$$(J(x_k)d_k, F(x_k)) < 0.$$

independentemente do valor de $\alpha_k > 0$. Após avaliarmos s_k, são conferidas as desigualdades (4.9) e (4.10). Se alguma delas não for satisfeita, a execução é interrompida. Isto acontece quando o problema não tem solução. A escolha da norma $\|.\|_\infty$ em lugar da Euclideana responde à necessidade de considerar possíveis limitantes para as variáveis x_i.

Algoritmo 5 Minimização em regiões de confiança

Seja $\sigma \in (0,1)$, $\gamma \in (0,1]$, $\eta_1, \eta_2 \in (0,1)$, $\eta_1 < \eta_2$, $M > 0$, $tol \in (0,1)$, $max \in I\!N$ *dados independetemente de* k *e seja* $x_0 \in I\!R^n$ *um ponto inicial arbitrário,* $\Delta_0 = M$ *e* $\alpha_0 = 1$.

Dado $x_k \in I\!R^n$ *tal que* $J(x_k)^T F(x_k) \neq 0$, $\Delta_k > 0$ *e* $\alpha_k \in (0,1]$, *os passos para obter* x_{k+1}, Δ_{k+1} *e* α_{k+1} *são os seguintes:*

- Passo 1.
  Calcular s_k *como uma "solução aproximada" e*

$$\text{Minimizar } \psi(s) \equiv \frac{1}{2}\|J(x_k)s + F(x_k)\|^2 \text{ s.t. } \|s\|_\infty \leq \Delta_k. \tag{4.12}$$

  A solução aproximada de (4.12) é obtida aplicando o método descrito em [7] parando quando

$$\|\nabla_P \psi(s_k)\| \leq tol\ \|\nabla_P \psi(0)\|, \tag{4.13}$$

  (onde $\nabla_P \psi(s)$ *é o gradiente projetado de* ψ *na caixa* $\|s\|_\infty \leq \Delta_k$*) ou quando o número de iterações usado pelo algoritmo [7] ultrapassa max. (Isto garante pelo menos que* $\|J(x_k)s_k + F(x_k)\| < \|F(x_k)\|$*).*

- Passo 2.
  Definir $d_k = s_k/\alpha_k$. *Se (4.10) e (4.9) são satisfeitas, passar para o Passo 3. Senão, parar (o algoritmo falhou, provavelmente pela proximidade de um Jacobiano singular)*

- Passo 3.
  Idem que o Passo 2 do Algoritmo 1.

- Passo 4.
  Idem que o Passo 3 do Algoritmo 1.

- Passo 5.
  Se $\alpha_{k+1} = 1$, *definir* $\Delta_{k+1} = M$.
  Senão, definir $\Delta_{k+1} = \|s_k\|_\infty/2$.

Os parâmetros usados na implementação são $\sigma = 10^{-5}$, $\gamma = 10^{-4}$, $\eta_1 = \frac{1}{2}$, $\eta_2 = \frac{1}{2}$, $M = 10^{10}$, $tol = \frac{1}{10}$, $max = n$.

O código computacional usado para a implementação deste algoritmo é uma adaptação do algoritmo para minimização em caixas realizado por Friedlander, Martínez and Santos [6].

O algoritmo usado para obter as soluções aproximadas de (4.13) combina iterações conjugadas e "chopped" (componentes cortadas) do gradiente, de tal forma que várias restrições ativas podem ser acrescentadas ou eliminadas em uma iteração.

4.3.1 A determinação de pontos singulares

Dado $H : \mathbb{R}^{m+1} \to \mathbb{R}^m$, $H = H(y,t)$, $H \in C^1(\mathbb{R}^{m+1})$, dizemos (conforme a [15]) que (y_*, t_*) é um *ponto singular* de $H(y,t) = 0$ se e somente se $H(y_*, t_*) = 0$ e se $H_y(y_*, t_*)$ é singular. Se o posto de $(H'(y_*, t_*)) = m$ dizemos que (y_*, t_*) é um *ponto de retorno*. Pontos singulares são soluções de

$$\left.\begin{array}{l} H(y,t) = 0 \\ H_y(y,t)v = 0 \\ \|v\|^2 = 1 \end{array}\right\} \tag{4.14}$$

para algum $v \in \mathbb{R}^m$. O sistema (4.14) tem $n = 2m + 1$ equações e incógnitas.

O algoritmo usado para obter as soluções aproximadas de (4.12) combina iterações com direções conjugadas e "chopped" do gradiente, de tal forma que as restrições ativas podem ser acrescentadas ou eliminadas em uma iteração.

A resolução de (4.14) usando um método tipo Newton-Inexato requer o cálculo de derivadas segundas. Entretanto, observamos que

$$H_y(y,t)v = \lim_{h \to 0} \frac{H(y + hv, t) - H(y - hv, t)}{2h}. \tag{4.15}$$

Resulta natural então, substituir (4.14) pelo sistema

$$\left.\begin{array}{l} H(y,t)=0 \\ \frac{H(y+hv,t)-H(y-hv,t)}{2h}=0 \\ \|v\|^2=1 \end{array}\right\} \tag{4.16}$$

para $h>0$. É de se esperar que as soluções de (4.16) para valores pequenos h, sejam boas aproximações das soluções de (4.14). Uma segunda alternativa consiste em considerar em lugar de (4.14), o sistema

$$\left.\begin{array}{l} H(y,t)=0 \\ H_y(y,t)v=0 \\ r^Tv=1 \end{array}\right\} \tag{4.17}$$

onde $r\in\mathbb{R}^m$ não está em $\mathcal{R}(H_y(y,t))$, o que garante a existência da solução das duas últimas equações de (4.17). Este sistema foi usado por Moore e Spence [15] e Seydel [21].

Considerando (4.17), e usando a aproximação (4.15), consideramos o sistema

$$\left.\begin{array}{l} H(y,t)=0 \\ \frac{H(y+hv,t)-H(y-hv,t)}{2h}=0 \\ r^Tv=1 \end{array}\right\} \tag{4.18}$$

A vantagem de (4.18) sobre (4.16) reside em que o termo não-quadrático $(\|v\|^2-1)^2$ na função de mérito (4.4), for substituido por $(r^Tv-1)^2$. Entretanto, se por acaso escolhemos $r\in\mathcal{R}(H_y(y_*,t_*))$, onde (y_*,t_*) é o ponto de retorno que estamos calculando, o sistema (4.17) não terá solução. Se o ângulo entre r e $\mathcal{R}(H_y(y_*,t_*))$ for pequeno, o problema de achar v que satisfaça $H_y(y,t)v=0$ e $r^Tv=1$ pode ser mal condicionado, conduzindo a resultados pouco confiáveis. Em nossas experiências selecionamos $r=v_0=(1,\ldots,1)^T/m^{\frac{1}{2}}$.

Observamos que as matrizes Jacobianas $J_1(y,t)$ e $J_2(y,t)$, dos sistemas (4.16) e (4.18) são respectivamente

$$J_1(x)=\left[\begin{array}{cc} H'(y,t) & 0 \\ \frac{H'(y+hv,t)-H'(y-hv,t)}{2h} & \frac{H_y(y+hv,t)+H_y(y-hv,t)}{2} \\ 0 & 2v^T \end{array}\right] \tag{4.19}$$

e

$$J_2(x) = \begin{bmatrix} H'(y,t) & 0 \\ \frac{H'(y+hv,t)-H'(y-hv,t)}{2h} & \frac{H_y(y+hv,t)+H_y(y-hv,t)}{2} \\ 0 & r^T \end{bmatrix} \tag{4.20}$$

É conveniente frisar, que na implementação do Algoritmo 3.1 não são efetuadas fatorações de matrizes. Em verdade, apenas precisamos de subrotinas que calculem o produto $J_1(y,t)w$ e $J_2(y,t)w$ para vetores arbitrários w. Observando (4.19) e (4.20), notamos que podemos aproveitar a estrutura destas matrizes de forma a facilitar a multiplicação por vetores.

4.4 Descrição dos Problemas e Resultados Numéricos

Para a realização dos testes selecionamos os problemas apresentados em [14] acrescentando-se mais um outro, originado da modelagem de um problema de transferência de energía [2]. Cada problema é descrito sucintamente deixando claro o significado do ponto de retorno, e colocando em relevância alguns dos parâmetros mais importantes, em relação ao problema e ao algoritmo. Com a finalidade de reproduzir os resultados relatados, fomos forçados a introduzir algumas correções em alguns dos dados indicados na referência [14].

Em cada Tabela, apresentada conjuntamente com a descrição do problema, indicamos na primeira coluna o sistema aumentado selecionado: "A" ou "B" que correspondem respectivamente às formulações (4.16) e (4.18); na próxima o ponto inicial, acompanhado com um parâmetro característico (como por exemplo, o tamanho do problema). Os resultados dos testes, são mostrados a partir da terceira coluna, na qual colocamos o valor do ponto de retorno t_* achado em cada caso; a seguir indicamos os valores singulares mínimos e máximos de H_y obtidos na aproximação final; na quinta coluna, a soma dos quadrados de $f(y_*, t_*)$ e finalmente, nas duas últimas colunas, o número de iterações e o número de avaliações da função realizadas pelo algoritmo.

O código computacional para cada problema, foi implementado usando FORTRAN 77, com dupla precisão. Uma adaptação do pacote computacional BOX-QUACAN [6] (para Minimização em Caixas com Canalizações) foi usada como subrotina para resolver o problema de minimização. Os testes correspondentes aos problemas 1-6, foram realizados em um PC486, e o 7 em uma SUN SPARC-Station 2.

4.4.1 *Problema 1 - Estrutura de barras*

Em [16], Oden apresenta um problema, que consiste em determinar os deslocamentos de uma estrutura formada por duas barras construidas com um material isotrópico. A aplicação do método dos elementos finitos nas equações de equilibrio origina o seguinte sistema não linear

$$F(y,t) \equiv A(y)y - tp,\ y \in I\!R^2,\ t \in I\!R^1, \tag{4.21}$$

para o vetor de deslocamento y, onde

$$A(y) = \begin{pmatrix} y_1^2 - 3\mu y_1 + 2\mu^2 & y_1 y_2 - \mu y_2 \\ y_1 y_2 - \mu y_2 & y_2^2 - \mu y_1 + 2 \end{pmatrix},\ \forall\, y \in I\!R^2$$

sendo $p \in I\!R^2$ é um vetor de carga dado e $\mu = 2$.

Os testes foram realizados modificando as relações entre o vetores de carga e o ponto de retorno informados em [14]. Em lugar de $(0.3, 0.91)$ usamos $(\sqrt{0.91}, 0.3)$. Também, observamos que em relação ao primeiro vetor $2 + \frac{2}{\sqrt{3}}$ corresponde a $-\frac{16}{3\sqrt{3}}$ e $\frac{2}{\sqrt{3}}$ corresponde a $\frac{16}{3\sqrt{3}}$.

Vetores de Carga:

$$Q_1 = (1,0),\ Q_2 = (\sqrt{0.9}, 0.3).$$

Pontos iniciais:

$$P_1(y_0,t_0) = [(3,0), -3],\ P_2(y_0,t_0) = [(0,0), 5],$$
$$P_3(y_0,t_0) = [(1,1), 1],\ P_4(y_0,t_0) = [(3,1), -3].$$

Sistema	(y_0,t_0)	t_*	$\sigma_1(H_y)$	$\sigma_n(H_y)$	$f(y_*,t_*)$	Iter.	Aval.
A	Q_1, P_1	-3.079205	4.86e-06	6.66e-01	9.87e-09	9	10
B		-3.079205	1.23e-08	6.66e-01	4.38e-10	8	9
A	Q_1, P_2	3.079201	5.70e-07	6.66e-01	3.72e-10	8	9
B		3.079209	6.27e-05	6.66e-01	7.23e-09	10	14
A	Q_2, P_3	1.933818	3.94e-06	0.47e-01	1.37e-10	4	5
B		1.933836	1.30e-07	0.47e-01	9.97e-14	5	6
A	Q_2, P_4	-2.307797	3.06e-05	0.24e-01	7.94e-10	4	5
B		-2.307831	1.04e-06	0.24e-01	3.73e-10	5	6

Tabela 4.1: Problema 1

4.4.2 *Problema 2 - A função de Freudenstein-Roth*

O seguinte sistema de equações, originalmente formulado em [5], é usado frequentemente como problema teste, por vários autores.

$$\left.\begin{array}{l} y_1 - y_2^3 + 5y_2^2 - 2y_2 + 34t - 47 = 0 \\ y_1 + y_2^3 + y_2^2 - 14y_2 + 10t - 39 = 0 \end{array}\right\} \tag{4.22}$$

Para $0 \leq t \leq 1$ obtemos dois pontos singulares.
Pontos Iniciais:

$$P_1(y_0, t_0) = [(1,1),1],\ P_2(y_0, t_0) = [(50,10),-10]$$

Sistema	(y_0, t_0)	t_*	$\sigma_1(H_y)$	$\sigma_n(H_y)$	$f(y_*, t_*)$	Iter.	Aval.
A	P_1	0.5875923	1.56e-06	1.89e+01	5.18e-09	14	22
B		0.5875873	5.68e-06	1.89e+01	2.87e-11	20	36
A	P_2	-0.6863575	6.65e-05	7.73	8.05e-09	16	33
B		-0.6863527	3.45e-08	7.73	1.00e-14	24	53

Tabela 4.2: Problema 2

4.4.3 *Problema 3 - O problema de estabilidade em aeronavegação*

Uma versão simplificada das equações de equilibrio aerodinâmico em aeronaves, que permitem prever deslocamentos bruscos em resposta a manobras realizadas, envolve cinco equações e oito variáveis ($[y,u]^T$). Três destas, ($u = [u_1, u_2, u_3]^T$) que modelam o elevador, aileron e o deflector respectivamente, funcionam como variáveis de controle. Para o problema de estabilidade de aeronavegação estudado em [13], as equações adimensionalizadas tem a seguinte forma

$$A\,[y,u]^T + \phi(y,u) = 0,\ \ \forall\, y \in I\!R^5,\ u \in I\!R^3, \tag{4.23}$$

onde

$$A = \begin{pmatrix} -3.933 & 0.107 & 0126 & 0 & -9.99 & 0 & -45.83 & -7.64 \\ 0 & -0.987 & 0 & -22.95 & 0 & -28.37 & 0 & 0 \\ 0.002 & 0 & -0.235 & 0 & 5.67 & 0 & -0.921 & 0 \\ 0 & 1.0 & 0 & -1.0 & 0 & -0.168 & 0 & 0 \\ 0 & 0 & -1.0 & 0 & -0.196 & 0 & -0.0071 & 0 \end{pmatrix},$$

e

$$\phi(y,u) = \begin{pmatrix} -0.727y_2y_3 & +8.39y_3y_4 & -684.4y_4y_5 & +63.5y_4y_2 \\ 0.949y_1y_3 & +0.173y_1y_5 & & \\ -.0.176y_1y_2 & -1.578y_1y_4 & +1.132y_4y_2 & \\ -y_1y_5 & & & \\ y_1y_4 & & & \end{pmatrix}.$$

Para realizar os cálculos escolhemos $t = u_2$, e foram fixados $u_1 = \gamma$ e $u_3 = 0$. O número de pontos de retorno varia com os valores de γ.

Neste caso foram necessárias duas correções: o elemento $A_{31} = 0.002$ na matriz A e na função ϕ, o termo $y_1 y_2$ por $y_1 u_2$.

Valores de γ:

$$\gamma_1 = -0.05,\ \gamma_2 = -0.008,\ \gamma_3 = 0.00,\ \gamma_4 = 0.05,\ \gamma_5 = 0.10$$

Pontos iniciais:

$$P_1(y_0, t_0) = [(-3, 1, -.1, .5, -.3), .5],$$
$$P_2(y_0, t_0) = [(-3, -.2, -.1, .02, .1), .2],$$
$$P_3(y_0, t_0) = [(-2.5, -8, .03, -.04), .3],$$
$$P_4(y_0, t_0) = [(-2.5, 1.5, .06, -.08, .06), .7].$$

Sistema	$\gamma_i; P(y_0, t_0)$	t_*	$\sigma_1(H_y)$	$\sigma_n(H_y)$	$f(y_*, t_*)$	Iter.	Aval.
A	$\gamma_1; P_1$	0.5087968	1.26e-05	2.20e+02	1.56e-10	68	139
B		0.5087889	2.64e-06	2.20e+02	1.04e-09	341	56122
A	$\gamma_2; P_2$	0.2063399	5.17e-05	4.99e+01	9.74e-09	33	52
B		0.2065148	2.07e-06	4.99e+01	5.07e-09	49	73
A	$\gamma_3; P_2$	0.872326	1.09e-05	5.86e+01	1.08e-10	37	65
B		0.3887898	1.62e-03	4.71+01	8.62e-04	500	950
A	$\gamma_4; P_3$	0.2929449	4.55e-05	1.84e+02	9.12e-09	22	37
B		0.2929395	7.84e-06	1.84e+02	2.60e-10	61	77
A	$\gamma_5; P_3$	9.227714e-02	3.46e+01	7.96e+01	1.02e-01	66	165
B		2.789284e-02	2.26e-02	1.04e+02	1.59e-01	105	214

Tabela 4.3: Problema 3

4.4.4 *Problema 4 - O circuito gatilho*

A operação de um circuito "gatilho" está descrito em [17]. As equações que descrevem o fluxo de corrente no circuito podem ser escritas na seguinte forma:

$$H(y, t) =$$

$$\begin{cases} (y_1 - y_3)/10000 + (y_1 - y_2)/39 + (y_1 - t)/51 = 0 \\ (y_2 - y_6)/10 + (y_2 - y_1)/39 + I(y_2) = 0 \\ (y_3 - y_1)/10000 + (y_3 - y_4)/25.5 = 0 \\ (y_4 - y_3)/25.5 + y_4/0.62 - y_5 + y_4 = 0 \\ (y_5 - y_6)/13 + y_5 - y_4 + I(y_5) = 0 \\ (y_6 - y_5)/13 + (y_6 - y_2)/10 + (y_6 - U(y_3 - y_1)/0.201 = 0 \end{cases} \tag{4.24}$$

Os dois diodos e o amplificador estão modelados por $I(x) := 5.6 \times 10^{-8}(e^{25x} - 1)$, $U(x) := 7.65tan^{-1}(1962x)$, respectivamente. As quantidades $[(y_1, \ldots, y_6), t]$ representam voltagens, em particular y_6, a voltagem de saída, é de interesse prático, sendo que t é a voltagem de entrada. O comportamento elétrico deste circuito gatilho está caracterizado por dois pontos de retorno.

De acordo com [17], fizemos duas alterações: num coeficiente de H e no valor do diodo $I(x)$.

Pontos iniciais:

$$P_1(y_0, t_0) = [(.05, .5, .05, .05, .15, .13), .5],$$
$$P_2(y_0, t_0) = [(.2, .6, .2, .2, .6, 9.5), .3].$$

	(y_0, t_0)	t_*	$\sigma_1(H_y)$	$\sigma_n(H_y)$	$f(y_*, t_*)$	Iter.	Aval.
A	P_1	0.6020924	1.26e-04	1.03e-01	8.23e-09	98	215
B		0.6013642	7.28e-05	1.03e+01	9.74e-09	118	148
A	P_2	0.3326203	1.82e-05	2.08e+01	7.58e-09	57	93
B		0.3329312	1.94e-05	2.07e+01	8.57e-09	27	43

Tabela 4.4: Problema 4

4.4.5 *Problema 5 - Um problema de reação química*

A equação integral

$$y(s) - t\int_0^1 k(s, \sigma)\, g(y(\sigma))\, d\sigma = 1, \quad 0 \leq s \leq 1, \tag{4.25}$$

com $k(s, \sigma) = s - 1$, $s \geq \sigma$, $k(s, \sigma) = k(\sigma, s)$ e onde

$$g(z) = z\, exp(\frac{\gamma\beta(1 - z)}{1 + \beta(1 - z)}),$$

foi usada por Moore e Spence [15] para testar algoritmos para calcular pontos de retorno. Esta integral representa uma reformulação do problema estudado por Kubicek em [10] que descreve a transferência de calor e massa em uma pastilha de catalizador poroso. A integral (4.25) é

aproximada sobre uma malha regular $s_i = ih$, $i = 1, \ldots, m$ usando a regra do trapézio . Esta discretização conduz ao seguinte sistema de equações

$$y_i - t\, h \sum_{j=0}^{m} w_j\, k(s-i, s_j)\, g(y_j) - 1 = 0, \quad i = 0, \ldots, m \tag{4.26}$$

com $w_1 = 1/2$, e $w_j = 1$ nos outros casos. Com $\gamma = 20$, $\beta = 0.4$ e $m = 32$, são obtidos dois pontos de retorno.

Pontos iniciais:

$$P_1(y_0, t_0) = [(1, \ldots, 1), .2], P_2(y_0, t_0) = [(.5, \ldots, .5), .1].$$

Sistema	(y_0, t_0)	t_*	$\sigma_1(H_y)$	$\sigma_n(H_y)$	$f(y_*, t_*)$	Iter.	Aval.
A	P_1	0.1375316	3.28e-05	1.00	8.30e-09	4	5
	P_2	0.07791575	8.48e-06	1.09	9.50e-11	6	11
B	P_1	0.1375395	2.22e-06	1.00	8.39e-11	4	5
	P_2	0.07791559	4.60e-06	1.09	5.49e-10	6	14

Tabela 4.5: Problema 5

4.4.6 *Problema 6 - A Equação "H" de Chandrasekhar*

Em [2], Chandrasekhar apresenta a seguinte equação integral, no contexto de problemas de transporte de energia radiante.

$$x(t) = 1 + \frac{c}{2} \int_0^1 \frac{t x(t) x(y)}{t + y} dy \tag{4.27}$$

O problema consiste em achar $x(t) \in C[0,1]$, que satisfaça (4.27)

Aproximamos a integral usando quadratura com nós em $\{t_i\}_{i=1}^M$ e pesos $\{w_i\}_{i=1}^M$, resultando assim no sistema não linear

$$f_i(x) = -x_i + 1 + \frac{c}{2} \sum_{j=1}^{M} \frac{t_i x_i x_j}{t_i + t_j} w_j, \quad i = 1, ..., M$$

com $x \in R^n$.

Quando aumentamos o número de pontos usados para aproximar a integral a matriz Jacobiana perde esparsidade,

$$\frac{\partial f_i}{\partial x_j} = \begin{cases} -1 + \frac{c}{2} \sum_{j=1}^{n} \frac{t_i x_j}{t_i + t_j} w_j, & i = j \\ \frac{c}{2} \frac{t_i x_i}{t_i + t_j} w_j, & i \neq j \end{cases}$$

Malha: $M = 8, 16, 32$.
Ponto inicial: $P(y_0, t_0) = [(.5, \ldots, .5), .1]$

Sistema	M	t_*	$\sigma_1(H_y)$	$\sigma_n(H_y)$	$f(y_*, t_*)$	Iter.	Aval.
A	8	1.000003	3.42e-06	8.88e-01	3.67e-10	6	9
B		1.000000	1.80e-08	8.88e-01	4.30e-12	7	10
A	16	1.000000	4.48e-7	9.35e-01	4.66e-11	9	12
B		1.000004	6.43e-05	9.35e-01	3.47e-09	7	12
A	32	1.000000	1.67e-06	9.63e-01	1.29e-09	8	11
B		1.000000	9.12e-08	9.63e-01	2.21e-11	8	13

Tabela 4.6: Problema 6

4.4.7 *Problema 7 - Um problema de valor de contorno*

Consideremos a aproximação do problema com valor de contorno não linear, (ver Simson [22])

$$\Delta u = -t\, g(u), \quad \forall (x, y)^T \in \Omega, \tag{4.28}$$
$$u_{\partial\Omega} = 0$$

sobre uma malha uniforme de tamanho $h = 1/l$ em $\Omega = [(0,1) \times (0,1)]$, na qual definimos a discretização de nove pontos para o operador de Laplace como

$$\square_9 u_{i,j} = \frac{1}{6h^2}[4\,(u_{i,j-1} + u_{i-1,j} + u_{i+1,j} + u_{i,j+1}) +$$
$$(u_{i-1,j-1} + u_{i+1,j-1} + u_{i-1,j+1} + u_{i+1,j+1}) - 20u_{i,j}],$$
$$i, j = 1, \ldots, M-1$$

e para o laplaciano de $g(u)$ uma discretização de cinco pontos

$$\square_5 g(u_{i,j}) = \frac{1}{h^2}[g(u_{i,j-1}) + g(u_{i-1,j}) + g(u_{i+1,j}) + g(u_{i,j+1}) - 4g(u_{i,j})],$$
$$i, j = 1, \ldots, M-1.$$

Assim obtemos

$$\square_9 u_{i,j} + t[g(u_{i,j}) + \frac{1}{12}h^2 \square_5 g(u_{i,j})] = 0,$$

$$i, j = 1, ..., M - 1$$

com as respectivas condições de contorno discretizadas

$$u_{i,j} = 0,\ i = 0,\ i = M,\ 0 \leq j \leq M,\ 0 \leq i \leq M,\ j = 0,\ j = M$$

que representa uma aproximação da ordem de h^4 para (4.28).

Desta forma, originamos um sistema não-linear de $N = (M-1)^2$ equações com N incógnitas u_{ij}, $i, j = 1, \ldots, (l-1)$, além do parâmetro t.

Para $g(u)$ foram escolhidos

a) $g(u) = e^u$

b) $g(u) = 1 + \frac{u+1/2u^2}{1+1/100u^2}$.

e para o cálculo dos pontos de retorno, tamanhos de malha $M = 49, 121, 225$.

Ponto inicial: $P(y_0, t_0) = [(1, \ldots, 1), 8]$

Sistema	$g(u), M$	t_*	$\sigma_1(H_y)$	$\sigma_n(H_y)$	$f(y_*, t_*)$	Iter.	Aval.
A	$a, 49$	6.807507	8.35e-07	3.07e+01	1.93e-09	6	7
B		6.807504	9.44e-07	3.07e+01	1.44e-10	8	9
A	$a, 121$	6.808005	1.25e-06	3.14e+01	3.39e-10	8	9
B		6.808005	1.44e-06	3.14e+01	2.85e-10	8	9
A	$a, 225$	6.808096	3.73244e-06	3.14e+01	1.03e-09	9	10
B		6.808045	1.04e-06	3.16e+01	1.68e-09	7	8
A	$b, 49$	7.980354	5.52e-07	3.07e+01	2.90e-11	15	30
B		7.980359	1.14e-04	3.07e+01	9.15e-09	13	23
A	$b, 121$	7.981427	2.06e-05	3.14e+01	2.76e-10	23	52
B		7.981423	1.16e-06	3.14e+01	1.30e-10	24	54
A	$b, 225$	7.981605	1.07e-04	3.16e+01	6.04e-09	44	76
B		7.981612	8.60e-05	3.16e+01	5.44e-09	23	24

Tabela 4.7: Problema 7

4.5 Análise dos resultados e conclusões

Procedemos a uma análise dos resultados numéricos.

Para calcular os acréscimos em (4.14) e em (4.16) usamos em ambos casos, um incremento $h = 10^{-4}$. Este valor pode ser convalidado a partir dos valores singulares obtido para H_y na aproximação final. Outros valores foram testados sem modificar substancialmente os resultados. Com exceção do Problema 3, o critério de convergência definido pela soma dos quadrados de $f(y_k) = f(y_k, t_k)$ para a aproximação final foi fixada em $\|f(x_k)\| \leq 10^{-8}$.

Para $\gamma = 0$ e usando (4.18), atingiu-se o máximo número de iterações (500). A convergência para $\gamma = 0.1$ foi conseguida, devido a que o critério $\|\nabla_P f(x_k)\| \leq 10^{-5} \max\{|f(x_k)|, 1\} / \max\{\|x_k\|$ em relação ao gradiente projetado sobre a caixa $-10 \leq y_i \leq 10$, $i = 1, \ldots, 5$, $-1 \leq t \leq 1$ foi satisfeito.

O desempenho observado através dos experimentos, dos sistemas ampliados (4.16) e (4.18), não permite concluir sobre a superioridade de quaisquer deles. Em geral, a eficácia de ambas aproximações depende principalmente do ponto inicial escolhido e também da não linearidade do problema.

No Capítulo 3 tratamos de um problema da dinâmica dos fluidos onde o número de Reynolds aparece naturalmente como o autovalor de (4.1). Schreiber e Keller [19] relatam vários pontos de retorno para diferentes tamanhos de malha. Em nenhum caso o nosso método conseguiu convergência, usando ambas formulações e diferentes pontos iniciais. Neste sentido, outras tentativas foram realizadas para obter soluções em diferente pontos não-singulares, redefinindo a função de mérito como $f(x) = \|H(y,t)\|$, com resultados também mal sucedidos. O método de Newton-Inexato se mostra inadequado para resolver este problema devido probavelmente à dispersão no espectro de autovalores da matriz Jacobiana $H_y(y,t)$.

Iniciamos este Capítulo apresentando um método tipo Newton-Inexato por Minimização para resolver sistemas não lineares de equações. O método não usa buscas lineares e está implementado usando estratégias de regiões confiança. Na formulação selecionada. o sistema resultante foi resolvido usando o algoritmo de minimização mencionado. Esta nova metodología de resolução pode ser considerada como uma contribuição para a resolução de problemas não lineares de autovalor.

Diversos tipos de problemas reais foram usados como testes para avaliar a eficiência e precisão deste novo esquema de resolução.

Este método foi testado para calcular pontos singulares de curvas homotópicas, que são soluções de sistemas aumentados aproximados. Salientamos que o ponto chave da nossa formulação, consiste em substituir a equação que estabelece que o espaço nulo do Jacobiano tem um vetor não nulo, por uma equação de diferenças, evitando o cálculo de derivadas.

As experiências realizadas mostram que o método consegue achar as soluções em forma precisa e permitem recomendar um valor para o parâmetro de discretização. [1]

[1] **Agradecimentos.** À Dra. Sandra A. Santos, que acompanhou a implementação e a realização dos testes.

Bibliografia

[1] Abbott, J. P. [1978]: *An efficient algoritm for the determination of certain bifurcation points.* J. of Comp. and Appl. Math.**4**, 19-27.

[2] Chandrasekhar, S. [1960]: *Radiative Transfer.* Dover, New York.

[3] Eisenstat, S. C. , Walker, H. F. [1994]: Globally convergent inexact Newton methods. Para aparecer em *SIAM Journal on Optimization.*

[4] Fletcher, R. [1987]: *Practical Methods of Optimization* (2nd edition), John Wiley and Sons, Chichester, New York, Brisbane, Toronto and Singapore.

[5] Freudestein, F., Roth, B. [1963]: *Numerical solution of systems of nonlinear equations.* J. ACM **10**, 550-556.

[6] Friedlander, A., Martínez, J.M., Santos, S.A.[1992]: *A new trust region algorithm for bound constrained minimization.* Technical Report, Departament of Applied Mathematics, University de Campinas.

[7] A. Friedlander, J. M. Martínez [1994], On the maximization of a concave quadratic function with box constraints. To appear in *SIAM Journal on Optimization.*

[8] Kelley, C. T. [1980]: *Solution of the Chandrasekhar $\mathcal{H}$-equation by Newton's Method.* Journal of Mathematical Physics **21**, pp.1625-1628.

[9] Kozakevich, D. N., Martínez, J.M., Santos, S.A.[1994]: *Inexact-Newton Methods and the Computation fo Singular Points.* Relatório de Pesquisa 14/94. DMA, IMECC, Universidade de Campinas, Brasil.

[10] Kubicek, M., Hlavacek, V. [1975]: *Solution of nonlinear boundary-value problems.* IX.Chem.Eng.Sci.**30**, 1439-1440.

[11] Martínez, J. M., Qi, L. [1993]: *Inexact Newton methods for solving nonsmooth equations.* Relatório de Pesquisa 67/93. DMA, IMECC, Universidade de Campinas, Brasil.

[12] Matthies, H., Strang, G. [1979: *The solution of nonlinear finite element equations.* Int. J. Num. Meth. Eng.**14**, pp.1613-1626.

[13] Mehra, R.K., Kessel, W.C., Caroll, J.V [1977/78/79]: *Global estability and control analysis of aircraft at high angles of attack.* ONR Report-CR-215-248-1,2,3.

[14] Melhem, R.G., Rheinboldt, W.C.[1982]: *A comparison of methods for determining turning points of nonlinear equations.* Computing **29**, 221-226.

[15] Moore, G., Spence, A., [1980]: *The calculation of turning points of nonlinear equations.* SIAM J. Num. Anal. **17**, 567-576.

[16] Oden, J.T. [1972]: *Finite element of nonlinear continua.* New York. McGraw-Hill.

[17] Pönish, G., Schwetlick, H. [1981]: *Computing turning points of curves implicity definded by nonlinear equations depending on a parameter.* Computing **26**, 107-121.

[18] Rheinboldt, W. C. [1986]: *Numerical analysis of parametrized nonlinear equations.* (University of Arkansas Lectures notes in the mathematical sciences; v. 7)

[19] Schreiber, R., Keller, H. B. [1983]: *Spurious Solutions in Driven Cavity Calculations.* J. Comput. Phys. **49**, pp. 165-172.

[20] Schwetlick, H. [1978]: *Numerische Losung nichtilinearer* Gleichungen Deutscher Verlag der Wissenschften. Berlin.

[21] Seydel, R. [1979]: *Numerical computation of branch points in ordinary differential equations.* Numerische Mathematik **32**, pp.51-68.

[22] Simpson, R. B. [1975]: *A method for numerical determination of bifurcation states of nonlinear systems of equations.* SIAM J. Num. Anal. **12**, pp.439-451.

Capítulo 5

Métodos Quase-Newton e Newton Inexato para fluxos em meios porosos

5.1 Introdução

A construção de modelos que representem adequadamente o escoamento dos fluidos num reservatório permite realizar previsões em relação a sua vida útil, capacidade de produção, etc., o que determina em princípio, a viabilidade da exploração da bacia petrolífera. Também, com a finalidade de otimizar a extração do óleo contido, possibilita o controle do regime de produção, a seleção de técnicas e a implementação de métodos alternativos de recuperação secundária.

A escolha de um modelo apropriado para um reservatório, dependerá básicamente das características estruturais da rocha produtora e das propriedades dos fluidos presentes. Os modelos tipo "black-oil" são usados habitualmente na realização de testes orientados a avaliar as características numéricas de um determinado algoritmo.

Consideraremos métodos numéricos para determinar o fluxo de dois fluidos imiscíveis escoando num meio poroso usando o método das diferenças finitas para aproximar as equações.

A maioria dos simuladores comerciais utilizam um tratamento temporal explícito para as não linearidades. Atualmente existe uma tendência direcionada para o desenvolvimento de simuladores que ofereçam soluções totalmente implícitas, cujas principais vantagens são as de fornecer soluções estáveis sem a necessidade de limitar o tamanho dos passos no tempo. Este esquema origina sistemas acoplados de equações algébricas não lineares a cada passo do tempo. A linearização destas equações gera sistemas lineares de grande porte, esparsos e não simétricos. Assim a eficácia do esquema totalmente implícito será dada principalmente pelas virtudes do método para resolver os sistemas envolvidos.

Neste trabalho compararemos métodos Quase-Newton e Newton Inexatos para a resolução dos sistemas de equações que modelam este problema para um escoamento bifásico, bidimensional considerando uma geometria simples do modelo físico. Os testes foram orientados com o duplo objetivo de mostrar algumas das características de cada algoritmo e indicar princi-

palmente qual deles é o mais eficiente para este problema em particular. Nos concentraremos principalmente nos métodos que já têm sua eficiência comprovada na resolução de outros problemas, confrontando métodos diretos e iterativos para os sistemas emergentes. Ambos os métodos têm sido implementados com técnicas complementares baseadas em Fatorações Incompletas que melhoram substancialmente seu desempenho.

Na próxima seção, apresentamos as equações que regem o escoamento num meio poroso e suas correspondentes discretizações espaciais e temporais. As características do problema implementado serão mostradas na Seção 3. Na Seção 4 descreveremos sucintamente o conjunto de métodos para sua resolução. Os resultados numéricos e respectivas análises serão mostrados na Seção 5. Finalmente apresentaremos algumas conclusões e sugestões para futuros trabalhos.

5.2 Descrição do Problema

Os modelos " black-oil" são considerados como protótipos para este tipo de problemas e vêm sendo usados habitualmente na realização de testes para diversos algoritmos na simulação de reservatórios (ver Aziz [2]).

Fazendo uso das hipóteses padrões que são habitualmente aplicadas sobre este modelo para um fluxo multifásico, resultam as seguintes equações:

Equação de conservação de óleo

$$\nabla \cdot (\lambda_o(\nabla p_o - \gamma_o \nabla D) = \partial (\phi_o S_o / B_o) / \partial t + Q_o \tag{5.1}$$

Equação de conservação de água

$$\nabla \cdot (\lambda_w (\nabla p_o - \gamma_w \nabla D)) = \partial (\phi_w S_w / B_w) / \partial t + Q_w \tag{5.2}$$

Equação de conservação de gás

$$\begin{aligned} \nabla \cdot (\lambda_g (\nabla p_g - \gamma_g \nabla D) + R_{go} \lambda_o (\nabla p_o - \gamma_o \nabla D)) = \\ \partial (\phi S_g / B_g + R_{go} S_o / B_o) / \partial t + (Q_g + R_{go} Q_o). \end{aligned} \tag{5.3}$$

Consideraremos para o nosso estudo o escoamento de dois fluidos imiscíveis em um reservatório bidimensional. As equações que regem o fluxo para este caso particular podem ser obtidas a partir das equações de conservação listadas acima, resultando:

$$\partial (\lambda_o \partial p_o / \partial x) / \partial x + \partial (\lambda_o \partial p_o / \partial y) / \partial y = \partial (\phi S_o / B_o) / \partial t + Q_o \tag{5.4}$$

$$\partial (\lambda_w \partial p_w / \partial x) / \partial x + \partial (\lambda_w \partial p_w / \partial y) / \partial y = \partial (\phi S_w / B_w) / \partial t + Q_w, \tag{5.5}$$

em que o e w designam as fases presentes. O termo gravitacional é desprezado, supondo um reservatório horizontal.

As mobilidades das fases estão dadas por

$$\lambda_{pf} = \frac{k_p\, k_{rf}}{B_f\, \mu_f}, \text{ sendo } f = w, o \; ; \; p = x, y.$$

As pressões de ambas as fases estão relacionadas através da pressão capilar

$$P_C = p_o - p_w,$$

e a equação de restrição para a saturação como

$$S_w + S_o = 1.$$

Discretização

Aproximaremos numericamente as equações de fluxo sobre uma malha com espaçamento uniforme. A discretização dos termos de fluxo conduz a :

$$\frac{\partial}{\partial x}\left(\lambda_{xf}\frac{\partial p_f}{\partial x}\right)_{i,j} \approx \frac{1}{\Delta x_i}\left(\lambda_{xp_{i+1/2}}\frac{p_{f_{i+1,j}} - p_{f_{i,j}}}{\Delta x_{i+1/2}} + \lambda_{xp_{i-1/2}}\frac{p_{f_{i-1,j}} - p_{f_{i,j}}}{\Delta x_{i-1/2}}\right), \tag{5.6}$$

$$\frac{\partial}{\partial y}\left(\lambda_{yf}\frac{\partial p_f}{\partial y}\right)_{i,j} \approx \frac{1}{\Delta y_j}\left(\lambda_{yp_{j+1/2}}\frac{p_{f_{i,j+1}} - p_{f_{i,j}}}{\Delta y_{j+1/2}} + \lambda_{yp_{j-1/2}}\frac{p_{f_{i,j-1}} - p_{f_{i,j}}}{\Delta y_{j-1/2}}\right), \tag{5.7}$$

e a aproximação para o termo de acumulação pode ser escrita como

$$\frac{\partial}{\partial t}\left(\frac{\phi S_f}{B_f}\right) \approx \frac{1}{\Delta t}\left[\left(\frac{\phi S_f}{B_f}\right)^{\nu+1} - \left(\frac{\phi S_f}{B_f}\right)^{\nu}\right]. \tag{5.8}$$

Utilizando as aproximações (5.6) a (5.8) nas equações (5.4) e (5.5) e multiplicando pelo volume do bloco da malha $\Delta V_{i,j} = \Delta x_i \Delta y_j \Delta z$, temos

$$\begin{aligned}
&\left[T_{xf_{i-1/2}}\left(p_{f_{i-1,j}} - p_{f_{i,j}}\right) + T_{xf_{i+1/2}}\left(p_{f_{i+1,j}} - p_{f_{i,j}}\right)\right]^{\upsilon} + \\
&\left[T_{yf_{j-1/2}}\left(p_{f_{i,j-1}} - p_{f_{i,j}}\right) + T_{xf_{j+1/2}}\left(p_{f_{i,j+1}} - p_{f_{i,j}}\right)\right]^{\upsilon} = \\
&\frac{\Delta V_{i,j}}{\Delta t}\left[\left(\frac{\phi S_f}{B_f}\right)^{\nu+1} - \left(\frac{\phi S_f}{B_f}\right)^{\nu}\right] + \left(\tilde{Q_c}\right)_{i,j}^{\upsilon}
\end{aligned} \tag{5.9}$$

para $f = w, o$, $i = 1, 2, ..., N_x$, e $j = 1, 2, ..., N_y$.

As transmissibilidades das fases nas direções x e y estão dadas por

$$T_{xf_{i\pm1/2}} = \lambda_{xf_{i\pm1/2}} \frac{\Delta y_j \Delta z}{\Delta x_{i\pm1/2}}$$

e

$$T_{yf_{j\pm1/2}} = \lambda_{yf_{j\pm1/2}} \frac{\Delta x_j \Delta z}{\Delta y_{i\pm1/2}}.$$

A taxa de produção (injeção) do componente c no bloco i, j em condições padrões é

$$Q_{\tilde{c}_{i,j}} = \Delta V_{i,j} q_c.$$

O superscrito υ indica o nível na aproximação temporal: $\upsilon = \nu + 1$ para um esquema totalmente implícito; $\upsilon = \nu$ para um esquema totalmente explícito.

O primeiro termo entre colchetes em (5.9) é a taxa do fluxo da fase f no bloco i, j na direção x. Podemos escrever este termo como se segue:

$$\begin{gathered}\Delta T_{xf} \Delta p_f = T_{xf_{i-1/2}} \left(p_{f_{i-1,j}} - p_{f_{i,j}}\right) + T_{xf_{i+1/2}} \left(p_{f_{i+1,j}} - p_{f_{i,j}}\right) = \\ Q_{xf_{i-1/2}} + Q_{xf_{i+1/2}}\end{gathered} \tag{5.10}$$

Similarmente, o segundo termo entre colchetes expressa a taxa do fluxo na direção y :

$$\begin{gathered}\Delta T_{yf} \Delta p_f = T_{yf_{j-1/2}} \left(p_{f_{i,j-1}} - p_{f_{i,j}}\right) + T_{yf_{j+1/2}} \left(p_{f_{i,j+1}} - p_{f_{i,j}}\right) = \\ Q_{yf_{j-1/2}} + Q_{yf_{j+1/2}}\end{gathered} \tag{5.11}$$

Assim, a equação (5.9) representa o balanço de material da fase f no bloco i, j . Isto significa que a taxa volumétrica de fluxo da fase f deve ser igual a taxa de variação de volume acrescida por uma fonte ou sumidouro da fase f no bloco.

Usando a relação de capilaridade e a de saturação, as equações discretizadas para ambas fases podem ser expressas em termos de p_o e S_w como segue

$$\begin{gathered}(\Delta T_{xw} \Delta p_o)_{i,j} + (\Delta T_{xw} \Delta P_C)_{i,j} + (\Delta T_{yw} \Delta p_o)_{i,j} + (\Delta T_{yw} \Delta P_c)_{i,j} = \\ (\Delta V_{i,j}/\Delta t)\, \Delta_t \left[\phi S_w / B_w\right]_{i,j} - [Q_w]_{i,j}\,,\end{gathered} \tag{5.12}$$

$$\begin{gathered}(\Delta T_{xo} \Delta p_o)_{i,j} + (\Delta T_{yo} \Delta p_o)_{i,j} = \\ (\Delta V_{i,j}/\Delta t)\, \Delta_t \left[\phi(1 - S_w)/B_o\right]_{i,j} - [Q_o]_{i,j}\end{gathered} \tag{5.13}$$

sendo $\Delta_t = [\,]^{\nu+1} - [\,]^{\nu}$.

Estas equações em diferenças aplicadas ao longo da malha, podem ser escritas em forma compacta como

$$F(x) = [F_w(p_o, S_w), F_o(p_o, S_w)]^T = T\, x - C - \frac{\Delta_t [A]}{\Delta t} - Q, \tag{5.14}$$

em que o vetor incógnita x é definido como

$$x = (x_{11}, x_{12}, ..., x_{i,j-1}, x_{i,j}, x_{i,j+1}, ..., x_{NN}),$$

sendo cada elemento de x um subvetor da forma $x_{i,j} = (p_o, S_w)^T_{i,j}$

A matriz de transmissibilidade T é pentadiagonal por blocos cuja estrutura é dada por:

$$T_{i-1/2,j} \quad ... \quad T_{i,j-1/2} \quad T_{i,j} \quad T_{i,j+1/2} \quad ... \quad T_{i+1/2,j},$$

onde cada bloco é 2×2, sendo cada uma das submatrizes

$$T_{i\pm1/2,j} = \begin{bmatrix} T_{xw_{i\pm1/2}} & 0 \\ T_{xo_{i\pm1/2}} & 0 \end{bmatrix},$$

$$T_{i,j\pm1/2} = \begin{bmatrix} T_{yw_{j\pm1/2}} & 0 \\ T_{yo_{j\pm1/2}} & 0 \end{bmatrix},$$

$$T_{i,j} = \begin{bmatrix} -\sum T_w & 0 \\ -\sum T_o & 0 \end{bmatrix},$$

em que

$$\sum T_w = T_{xw_{i-1/2}} + T_{yw_{j-1/2}} + T_{xw_{i+1/2}} + T_{yw_{j+1/2}},$$

$$\sum T_o = T_{xo_{i-1/2}} + T_{yo_{j-1/2}} + T_{xo_{i+1/2}} + T_{yo_{j+1/2}}.$$

A forma em que a matriz T e o vetor x estão definidos implica que a malha é percorrida primeiro na direção y .

O vetores A, C e Q representam os termos de acumulação, capilaridade e injeção/produção cujas componentes são dadas respectivamente por :

$$A_{i,j} = \begin{bmatrix} (\Delta V \phi S_w / B_w)_{i,j} \\ (\Delta V \phi (1 - S_w)/B_o)_{i,j} \end{bmatrix},$$

$$C_{i,j} = \begin{bmatrix} (\Delta T_{xw} \Delta P_c)_{i,j} + (\Delta T_{yw} \Delta P_c)_{i,j} \\ 0 \end{bmatrix},$$

e

$$Q_{i,j} = \begin{bmatrix} Q_{\tilde{w}_{i,j}} \\ Q_{\tilde{o}_{i,j}} \end{bmatrix}.$$

Tratamento das transmissibilidades

As transmissibilidades das fases introduzem não linearidades nas equações aproximadas devido a sua dependência com a pressão e saturação.

As transmissibilidades interbloco na direção x podem ser expressas como

$$T_{xf_{i\pm1/2}} = \left(k_{x_{i\pm1/2}} \frac{\Delta y_j \Delta z}{\Delta x_{i\pm1/2}}\right) \left(\frac{1}{B_f \mu_f}\right)_{i\pm1/2} (k_{rf})_{i\pm1/2},$$

e similarmente na direção y,

$$T_{yf_{j\pm1/2}} = \left(k_{y_{j\pm1/2}} \frac{\Delta x_i \Delta z}{\Delta y_{j\pm1/2}}\right) \left(\frac{1}{B_f \mu_f}\right)_{j\pm1/2} (k_{rf})_{j\pm1/2}.$$

Como se pode observar em ambas equações, as transmissibilidades estão compostas pelos seguintes fatores: no primeiro fator aparece o fator geométrico F_G, que depende da geometria da malha e da distribuição da permeabilidade absoluta (a permeabilidade na interfase do bloco será avaliada utilizando a média harmônica), o fator seguinte F_P contém os parâmetros dependentes exclusivamente da pressão, neste caso usamos uma ponderação centrada; o último fator F_S depende apenas da saturação. A permeabilidade relativa k_{rf} é intrinsecamente relacionada com problemas convectivos com equações de natureza hiperbólica; o uso de uma ponderação centrada pode conduzir a resultados sem significado físico. Por este motivo é usado um esquema "upstream" com a desvantagem de produzir dispersão numérica na solução.

As não linearidades introduzidas pelas transmissibilidades podem ser divididas em dois grupos: não linearidades fracas causadas pela dependência com a pressão, e não linearidades fortes que são os coeficientes dependentes da saturação como k_{rf} e P_c.

Esquema totalmente implícito

Como pode ser observado em (5.14) as equações discretizadas para um fluxo bifásico geram sistemas não lineares de equações em diferenças a cada passo do tempo.

Para alguns problemas que envolvem escoamento multidimensionais, os esquemas explícitos ou parcialmente implícitos para as equações de fluxo discretizadas são inadequados. A estabilidade numérica impõe limitações sérias sobre tais esquemas. Por este motivo, é necessário um tratamento totalmente implícito para a equação (5.14) o que implica em resolver dado o valor inicial x_0,

$$F^{\nu+1}(x) = 0$$

para $\nu = 0, 1, \ldots$.

A primeira etapa de qualquer método de resolução requer uma linearação prévia de (5.14). A escolha na aproximação temporal nos termos de fluxo, acumulação e injeção-produção leva a selecionar diferentes esquemas de linearização que motivam outros tantos métodos. Os coeficientes de transporte (transmissibilidades) introduzem as não linearidades mais significativas nas equações.

As funções dos resíduos para as equações em diferenças podem ser escritas como

$$F_{w_{i,j}} = (\Delta T_{xw}\Delta p_o) + (\Delta T_{xw}\Delta P_C) + (\Delta T_{yw}\Delta p_o) + (\Delta T_{yw}\Delta P_c) + (\Delta V_{i,j}/\Delta t)\,\Delta_t\,[\phi S_w/B_w]_{i,j} - [Q_w]_{i,j} \tag{5.15}$$

$$F_{o_{i,j}} = (\Delta T_{xo}\Delta p_o) + (\Delta T_{yo}\Delta p_o) + (\Delta V_{i,j}/\Delta t)\,\Delta_t\,[\phi(1-S_w)/B_o]_{i,j} - [Q_o]_{i,j}\,. \tag{5.16}$$

A matriz Jacobiana do sistema (5.15) e (5.16) tem a mesma estrutura que a matriz T e é composta pelas seguintes submatrizes

$$J_{i,j} = \begin{bmatrix} \frac{\partial R_{w_{i,j}}}{\partial p_{o_{i,j}}} & \frac{\partial R_{w_{i,j}}}{\partial S_{w_{i,j}}} \\ \frac{\partial R_{o_{i,j}}}{\partial p_{o_{i,j}}} & \frac{\partial R_{o_{i,j}}}{\partial S_{w_{i,j}}} \end{bmatrix},$$

$$J_{i\pm1/2,j} = \begin{bmatrix} \frac{\partial R_{w_{i,j}}}{\partial p_{o_{i\pm1,j}}} & \frac{\partial R_{w_{i,j}}}{\partial S_{w_{i\pm1/2,j}}} \\ \frac{\partial R_{o_{i,j}}}{\partial p_{o_{i\pm1/2,j}}} & \frac{\partial R_{o_{i,j}}}{\partial S_{w_{i\pm1/2,j}}} \end{bmatrix},$$

$$J_{i,j\pm1/2} = \begin{bmatrix} \frac{\partial R_{w_{i,j}}}{\partial p_{o_{i,j\pm1}}} & \frac{\partial R_{w_{i,j}}}{\partial S_{w_{i,j\pm1}}} \\ \frac{\partial R_{o_{i,j}}}{\partial p_{o_{i,j\pm1}}} & \frac{\partial R_{o_{i,j}}}{\partial S_{w_{i,j\pm1}}} \end{bmatrix}$$

Os elementos destas submatrizes contêm as derivadas dos resíduos em relação as variáveis primárias p_o e S_w. Usando (5.15) e (5.16) podemos obter os elementos de cada submatriz de $J_{i,j}$

$$\frac{\partial R_{w_{i,j}}}{\partial p_{o_{i,j}}} = \left(-\sum T_w + \Delta\frac{\partial T_{xw}}{\partial p_{o_{i,j}}}\Delta p_o + \Delta\frac{\partial T_{yw}}{\partial p_{o_{i,j}}}\Delta p_o\right) - \left(\Delta\frac{\partial T_{xw}}{\partial p_{o_{i,j}}}\Delta P_c + \Delta\frac{\partial T_{yw}}{\partial p_{o_{i,j}}}\Delta P_c\right) - \frac{\Delta V_{i,j}}{\Delta t}\left(\frac{\phi'(p)S_w}{B_w}\right)_{i,j} - \frac{\partial Q_{\tilde{w}_{i,j}}}{\partial p_{o_{i,j}}},$$

$$\frac{\partial R_{w_{i,j}}}{\partial S_{w_{i,j}}} = \left(\Delta\frac{\partial T_{xw}}{\partial S_{w_{i,j}}}\Delta p_o + \Delta\frac{\partial T_{yw}}{\partial S_{w_{i,j}}}\Delta p_o\right) - \left[\left(-\sum T_w\right)P_c'(S_w) + \left(\Delta\frac{\partial T_{xw}}{\partial S_{w_{i,j}}}\Delta P_c + \Delta\frac{\partial T_{yw}}{\partial S_{w_{i,j}}}\Delta P_c\right)\right] - \frac{\Delta V_{i,j}}{\Delta t}\left(\frac{\phi}{B_w}\right)_{i,j} - \frac{\partial Q_{\tilde{w}_{i,j}}}{\partial S_{w_{i,j}}},$$

$$\frac{\partial R_{o_{i,j}}}{\partial p_{o_{i,j}}} = \left(-\sum T_o + \Delta\frac{\partial T_{xo}}{\partial p_{o_{i,j}}}\Delta p_o + \Delta\frac{\partial T_{yo}}{\partial p_{o_{i,j}}}\Delta p_o\right) - \frac{\Delta V_{i,j}}{\Delta t}\left(1-S_{w_{i,j}}\right)\left(\frac{\phi'(p)B_w - \phi B_w'(p)}{B_w^2}\right)_{i,j} - \frac{\partial Q_{\tilde{o}_{i,j}}}{\partial p_{o_{i,j}}},$$

$$\frac{\partial R_{o_{i,j}}}{\partial S_{w_{i,j}}} = \left(\Delta\frac{\partial T_{xo}}{\partial S_{w_{i,j}}}\Delta p_o + \Delta\frac{\partial T_{yo}}{\partial S_{w_{i,j}}}\Delta p_o\right) - \frac{\Delta V_{i,j}}{\Delta t}\left(\frac{\phi}{B_w}\right)_{i,j} - \frac{\partial Q_{\tilde{o}_{i,j}}}{\partial S_{w_{i,j}}}$$

sendo

$$\sum T_f = T_{xf_{i-1/2}} + T_{xf_{i+1/2}} + T_{yf_{j-1/2}} + T_{yf_{j+1/2}}.$$

Os elementos das submatrizes $J_{i\pm1/2,j}$ estão dados por:

$$\frac{\partial R_{w_{i,j}}}{\partial p_{o_{i\pm1,j}}} = \left[T_{xw_{i\pm1/2}} + \frac{\partial T_{xw_{i\pm1/2}}}{\partial p_{o_{i\pm1,j}}}\left(p_{o_{i\pm1,j}} - p_{o_{i,j}}\right)\right] + \Delta\frac{\partial T_{xw_{i\pm1/2}}}{\partial p_{o_{i\pm1,j}}}\left(P_{c_{i\pm1,j}} - P_{c_{i,j}}\right),$$

$$\frac{\partial R_{w_{i,j}}}{\partial S_{w_{i\pm1,j}}} = \frac{\partial T_{xw_{i\pm1/2}}}{\partial S_{w_{i\pm1,j}}}\left(p_{o_{i\pm1,j}} - p_{o_{i,j}}\right) + T_{xw_{i\pm1/2}}P'_{c_{i\pm1,j}} + \frac{\partial T_{xw_{i\pm1/2}}}{\partial S_{w_{i\pm1,j}}}\left(P_{c_{i\pm1,j}} - P_{c_{i,j}}\right),$$

$$\frac{\partial R_{o_{i,j}}}{\partial p_{o_{i\pm1,j}}} = T_{xo_{i\pm1/2}} + \frac{\partial T_{xo_{i\pm1/2}}}{\partial p_{o_{i\pm1,j}}}\left(p_{o_{i\pm1,j}} - p_{o_{i,j}}\right),$$

$$\frac{\partial R_{o_{i,j}}}{\partial S_{w_{i\pm1,j}}} = \frac{\partial T_{xo_{i\pm1/2}}}{\partial S_{w_{i\pm1,j}}}\left(p_{o_{i\pm1,j}} - p_{o_{i,j}}\right).$$

Análogamente, para as submatrizes $J_{i,j\pm1/2}$ temos

$$\frac{\partial R_{w_{i,j}}}{\partial p_{o_{i,j\pm1}}} = \left[T_{xw_{j\pm1/2}} + \frac{\partial T_{xw_{j\pm1/2}}}{\partial p_{o_{i,j\pm1}}}\left(p_{o_{i,j\pm1}} - p_{o_{i,j}}\right)\right] + \Delta\frac{\partial T_{xw_{j\pm1/2}}}{\partial p_{o_{i,j\pm1}}}\left(P_{c_{i,j\pm1}} - P_{c_{i,j}}\right),$$

$$\frac{\partial R_{w_{i,j}}}{\partial S_{w_{i,j\pm1}}} = \frac{\partial T_{xw_{j\pm1/2}}}{\partial S_{w_{i,j\pm1}}}\left(p_{o_{i,j\pm1}} - p_{o_{i,j}}\right) + T_{xw_{j\pm1/2}}P'_{c_{i,j\pm1}} + \frac{\partial T_{xw_{j\pm1/2}}}{\partial S_{w_{i,j\pm1}}}\left(P_{c_{i,j\pm1}} - P_{c_{i,j}}\right),$$

$$\frac{\partial R_{o_{i,j}}}{\partial p_{o_{i,j\pm1}}} = T_{xo_{j\pm1/2}} + \frac{\partial T_{xo_{j\pm1/2}}}{\partial p_{o_{i,j\pm1}}}\left(p_{o_{i,j\pm1}} - p_{o_{i,j}}\right),$$

$$\frac{\partial R_{o_{i,j}}}{\partial S_{w_{i,j\pm1}}} = \frac{\partial T_{xo_{j\pm1/2}}}{\partial S_{w_{i,j\pm1}}}\left(p_{o_{i,j\pm1}} - p_{o_{i,j}}\right).$$

A matriz Jacobiana pode ser decomposta como

$$J = T' - T'_c - A' - Q'$$

sendo T' a derivada da matriz de transmissibilidade, T'_c a derivada da matriz capilaridade-transmissilidade, A' a derivada do termo de acumulação e Q' a derivada do termo de injeção/produ Somente as matrizes T' e T'_c introduzem elementos fora da diagonal da matriz Jacobiana devido a que os termos de transmissibilidades dependem das variáveis do bloco i,j e dos nós vizinhos.

5.2.1 Caracterização do Problema

O problema sobre o qual serão realizados os testes é caracterizado pelos seguintes conjuntos de hipóteses e condições:

- Formulação do modelo "black-oil" para simular o fluxo imiscível de óleo e água em reservatórios horizontais.
- Sistema bidimensional em coordenadas cartesianas.
- Admite-se compressibilidade de fluido e rocha. Viscosidade não é dependente da pressão. Meio isotrópico.
- Reservatório heterogêneo com distribuições conhecidas de permeabilidade e porosidade.
- Vazão nula na fronteira, $\nabla p(t) \cdot n = 0$ em Γ.
- Curvas de permeabilidade relativa para as rochas presentes no reservatório cuja dependência com S_w é dada através de dados empíricos, k_{rp} vs. S_w; a expressão que interpola os dados (obtida via regressão quadrática) é dada por $k_{rp} = a_p {S_w}^{b_p}$.
- Processo de injeção de água em reservatórios (para um esquema de um quarto de "five spot").
- Discretização temporal totalmente implícita.

5.3 Descrição dos métodos

Nesta seção comentaremos a incorporação de uma mesma técnica complementar em dois tipos de métodos com concepções diferentes, que foram descritos no primeiro Capítulo. A introdução desta técnica (ver Zambaldi [17]) modifica drasticamente seu desempenho computacional.

A desvantagem na escolha do esquema totalmente implícito está na necessidade de resolver, um sistema não-linear a cada passo do tempo. As matrizes envolvidas caracterizam-se por ser de grande porte, esparsas e não-simétricas. Neste contexto, o ponto crucial do ponto de vista numérico, na aplicabilidade de um determinado método sobre um simulador, reside principalmente na eficiência da resolução dos sistemas.

Embora os métodos diretos sejam robustos e precisos , o tempo de processamento e a demanda de armazenagem são consideravelmente maiores que para os métodos iterativos. Uma família de métodos desta classe, denominados tipo Gradientes Conjugados reúnem as características desejadas desde que incluam o uso de precondicionadores.

5.3.1 Newton Inexato com Precondicionadores de Fatorações Incompletas

Um precondicionador adequado, para um sistema linear $Ax = b$, em que a matriz A tem uma estrutura esparsa arbitrária, pode ser construido usando uma fatoração incompleta. Esta técnica consiste em achar uma fatoração aproximada de $M = \widetilde{L}\widetilde{U}$, em que $\widetilde{L}$ e $\widetilde{U}$ são obtidas por uma fatoração incompleta da matriz A, calculada por eliminação Gaussiana, eliminando os coeficientes que ocasionariam um enchimento na estrutura original de A. Os fatores $\widetilde{L}$ e $\widetilde{U}$ que são matrizes triangulares inferiores e superiores devem reter o mesmo padrão de esparsidade de A. Uma melhor aproximação para a fatoração implicará em uma aceitação na perda desse padrão. Designamos a $ILU(l)$ como sendo uma fatoração incompleta com um nível de preenchimento definido por l .

Na abordagem de Lantangen [10] o nível de preenchimento se estabelece mediante uma matriz de índices $P = [i, j]$ associada diretamente ao parâmetro l, que representa o padrão de esparsidade da fatoração LU . Se $P = \{(i,j)/A_{i,j} \neq 0\}$, onde $A_{i,j}$ representa todos os elementos não nulos de A, conduz a uma matriz LU onde todos os elementos adicionais originados pela decomposição são rejeitados; não admitir qualquer preenchimento será indicado por $l = 0$. Lantangen define a matriz no nível l utilizando o padrão de esparsidade do nível imediato inferior.

Assim, $ILU(l)$ é definida em termos de $ILU(l-1)$ da seguinte forma: seja a matriz P_{l-1} a matriz cujos coeficientes são os indices (i, j) correspondentes ao padrão de esparsidade no nível $(l-1)$. Consideremos a seguir o preenchimento causado ao efetuar o produto $L \times U$. P_l estará dada pelo conjunto de índices de P_{l-1} adicionando-se aqueles correspondentes aos novos coeficientes originados por efetuar o produto LU.

Para resolver $F(x) = 0$, o método Newton Inexato gera uma sequência $\{s^k{}_j\}$, $j = 1, 2, \ldots$ pela aplicação de um método iterativo ao sistema linear

$$J(x^k)s = -F(x^k),$$

de tal forma de obter um $\{s^k{}_j\}$ "suficientemente bom", cuja qualidade é determinada com um critério prefixado (ver Capítulo 1). Neste trabalho utilizamos uma implementação do método Newton Inexato com o Método GMRES para resolver os sistemas lineares, usando precondicionadores baseados em Fatorações Incompletas.

5.3.2 Atualizações Secantes em Fatorações Incompletas

As diferentes estratégias concebidas para definir as fatorações incompletas encontradas na literatura como foi descrito acima, tem como objetivo a construção de precondicionadores para diminuir o trabalho computacional na resolução dos sistemas lineares.

Consideraremos neste trabalho, um conjunto de métodos Quase-Newton onde tanto B_0 como os B_q são aproximações da matriz Jacobiana. Ou seja $\widetilde{B}_0 = \widetilde{J}(x_0)$ e $\widetilde{B}_{k,q} = \widetilde{J}(x_k)$, em que $\widetilde{J} = J + E$ sendo $\widetilde{J}$ obtida a partir de J com algum critério. Espera-se com isso obter alguma economia no produto $B_k^{-1}w$ e uma diminuição na demanda de armazenagem.

Fórmulas secantes

Os métodos quase Newton que obedecem a equação secante, diferenciam-se entre si pelas condições adicionais impostas sobre as matrizes B_k, como preservar alguma estrutura da matriz Jacobiana ou satisfazer algum princípio de variação mínima.

Neste estudo consideraremos quatro fórmulas secantes para gerar as matrizes de iteração dos algoritmos correspondentes aos métodos Quase-Newton usando Fatorações Incompletas. As fórmulas que apresentamos correspondem aos respectivos métodos secantes: Broyden-1 (Primeiro Método de Broyden); Broyden-2 (Segundo Método de Broyden); Atualização de uma Coluna(COLUMN) e Atualização de uma Coluna da Inversa (ICOL).

Consideraremos a seguintes fórmulas para B_k:

Em todos os casos y_k é definido como:

$$y_k = F(x_{k+1}) - F(x_k)$$

- **Fórmula de Broyden-1**

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)\, s_k^T}{s_k^T s_k}.$$

(Ver Broyden [6])

- **Fórmula de Broyden-2**

$$B_{k+1}^{-1} = B_k^{-1} + \frac{\left(s_k - B_k^{-1} y_k\right) y_k^T}{y_k^T y_k}.$$

(Ver Broyden [6])

- **Fórmula de Atualização na Coluna (Column)**

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)\, e_{jk}^T}{e_{jk}^T s_k}, \quad \text{sendo } \mid e_{jk}^T s_k \mid = \| s_k \|_\infty.$$

(Ver Martínez [11])

- **Fórmula de Atualização na Coluna da Matriz Inversa (ICol)**

$$B_{k+1}^{-1} = B_k^{-1} + \frac{\left(s_k - B_k^{-1}y_k\right)e_{jk}^T}{e_{jk}^T y_k}, \quad \text{sendo } \mid e_{jk}^T y_k \mid = \parallel y_k \parallel_\infty.$$

(Ver Martínez e Zambaldi [12])

As fórmulas B-1 e Column satisfazem

$$B_{k+1}^{-1} = \left(I + u_k z_k^T\right)\left(I + u_{k-1} z_{k-1}^T\right) \ldots \left(I + u_0 z_0^T\right) B_0^{-1}$$

em que os vetores u_k, z_k são:

$$u_k = \frac{\left(s_k - B_k^{-1}y_k\right)}{s_k^T B_k^{-1} y_k}, \quad z_k = s_k$$

e

$$u_k = \frac{\left(s_k - B_k^{-1}y_k\right)}{e_{jk}^T B_k^{-1} y_k}, \quad z_k = e_{jk}$$

para cada fórmula, respectivamente.

Para calcular o produto $B_k^{-1}w$ para qualquer vetor arbitrário $w \in R^n$ necessitamos armazenar $2kn$ posições para Broyden-1 e apenas kn para Column.

Similarmente, para B-2 e para Icol temos

$$B_{k+1}^{-1} = u_k z_k^T + u_{k-1} z_{k-1}^T + \ldots + u_0 z_0^T$$

em que os vetores u_k, z_k são:

$$u_k = \frac{\left(s_k - B_k^{-1}y_k\right)}{s_k^T B_k^{-1} y_k}, \; z_k = s_k$$

e

$$u_k = \frac{\left(s_k - B_k^{-1}y_k\right)}{e_{jk}^T B_k^{-1} y_k}, \; z_k = e_{jk}$$

B-1 e B-2 são fórmulas LCSU, enquanto que Column e Icol não satisfazem o princípio de variação mínima.

Já que a demanda de armazenamento de memória para os algoritmos baseados nestas fórmulas cresce com k , devem ser implementados com recomeços.

Habitualmente $B_0 \equiv J(x_0)$, e os recomeços são produzidos utilizando algum critério sobre a taxa de diminuição do resíduo restituindo a matriz $B_k = J(x^k)$ ou, a cada q iterações da forma $B_q = J(x^k)$; isto é, quando k for múltiplo de um inteiro q a fórmula de atualização não é usada.

Zambaldi [17] implementou uma família de métodos quasenewtonianos usando Jacobianos simplificados baseados numa fatoração incompleta do tipo ILU em forma análoga a Lantangen [10] , com recomeços para valores de q prefixados. Desta maneira $\tilde{B}_q$ vem definida pelo valor do parâmetro l e nas q iterações a fatoração da matriz que aproxima o Jacobiano, é atualizada utilizando algum método secante. Basicamente as iterações são dadas por

$$x^{k+1} = x^k - \tilde{B}_k^{-1} F(x^k), \tag{5.17}$$

$$\tilde{B}_{k_q} = \tilde{J}(x_k). \tag{5.18}$$

sendo

$$\tilde{B}_o = \tilde{J(x_0)}$$

em que

$$\tilde{J(x_0)} = J(x_0) + E$$

isto é. $\tilde{J(x_0)}$ vem de uma fatoração incompleta (ILU) de $J(x_0)$, com recomeços a cada q iterações.

Uma outra estratégia mais elementar para conseguir uma matriz com um menor padrão de esparsidade que a matriz Jacobiana, consiste em definir uma matriz aproximada $\tilde{J}$ obtida pela projeção de $J(x_k)$ sobre um subespaço $S \in \mathbb{R}^{n \times n}$. O subespaço S pode ser escolhido dentre aqueles gerados por matrizes de banda. Esta aproximação surge naturalmente na discretização de problemas com valor de fronteira. Os resultados obtidos com diversas aproximações deste tipo foram totalmente desencorajadores.

5.4 Resultados numéricos

Um extenso conjunto de testes foram realizados utilizando diversos algoritmos com os métodos de Newton, Quase-Newton (implementados em Rouxinol) e Newton Inexatos com Precondicionadores Secantes e Fatorações Incompletas (implementados em NIPrec).

Apresentaremos apenas os resultados obtidos utilizando os métodos descritos na seção anterior por apresentar melhor desempenho, sobre um problema cujos parâmetros físicos e geométricos estão listados no fim deste capítulo.

A convergência é aceita quando a condição

$$\|F(x)\|_\infty < 10^{-10}$$

é satisfeita.

A divergência é declarada quando o número de iterações Quase Newtonianas ultrapassam $ItMax_{QN} = 250$ para os métodos QN-ILU, e quando o número de iterações Newton Inexato ultrapassa $ItMax_{NI} = 1000$ e/ou $ItMax_{GMRES} = 25$ para os métodos NI-ILU. Estes limitantes foram determinados como um procedimento totalmente heurístico.

Os resultados apresentam valores acumulados para os parâmetros de interesse, isto é, para o ciclo total da simulação.

Selecionamos várias malhas que foram combinadas com outros tantos parâmetros próprios dos algoritmos e do problema, o que gerou um número muito grande de testes.

5.4.1 Newton Inexatos Precondicionados

Mostramos o primeiro conjunto de resultados na Tabela 5.4.1 para diferentes níveis de preenchimento e distintos valores de θ_k (vários outros valores de θ_k também foram testados).

Em cada parêntese informamos: (Iterações Newton Inexato Acumuladas, Iterações GMRES Acumuladas, Tempo de Execução Acumulado (CPU) [Seg])

Em termos de custo computacional $ILU(4)$ - $\theta_k = 0.1$ mostrou ser o mais eficiente, com tempo de execução: 0.90e+07 enquanto que $ILU(5)$ - $\theta_k = 0.01$ demandou um menor número de iterações. Este fato pode ser explicado como se segue: com $\theta_k = 0.1$ temos mais iterações de GMRES porém em espaços de menor dimensão e, em consequência, com menor custo computacional.

$ILU(l)$	$\theta_k = 0.1$	$\theta_k = 0.01$
$l = 2$	(264 , 20216 , .509e+13)	(214 , 18607 , .497e+13)
$l = 3$	(217 , 3693 , .152e+09)	(159 , 3964 , .181e+10)
$l = 4$	(203 , 2298 , .900e+07)	(161 , 2248 , .142e+08)
$l = 5$	(207 , 2034 , .520e+08)	(162 , 2001 , .659e+08)

Tabela 5.1: Newton Inexato com Fatoração Incompleta . Malha: $M = 50$

5.4.2 Quase-Newton com Jacobianos de Fatorações Incompletas

Apresentamos uma sequência de tabelas para problemas de tamanhos crescentes, em que o número de resultados apresentados serão gradualmente reduzidos, levando em consideração o desempenho computacional observado. Escolhemos várias malhas e realizamos diversas experiências com varios níveis de preenchimento l selecionando valores diferentes para os recomeços q. Os métodos usados foram:

- A : Método de Broyden-1 (Primeiro Método de Broyden)
- B : Método de Broyden-2 (Segundo Método de Broyden)

- C : Método de Atualização de uma Coluna
- D : Método de Atualização de uma Coluna da Inversa

	q	$l=2$	$l=3$	$l=4$	$l=5$
A	10	> 250	> 250	> 250	2844, 289, .348e+03
	25	> 250	2766, 129, .844e+01	1672, 78, .168e+02	1607, 73, .956e+02
	50	> 250	2087, 58, .514e+01	1824, 51, .137e+02	1663, 47, .819e+02
	100	> 250	> 250	> 250	> 250
B	10	> 250	> 250	> 250	> 250
	25	3839, 169, .318e+02	1564, 82, .535e+01	1186, 57, .135e+02	1123, 56, .656e+02
	50	1723, 46, .998e+00	1294, 36, .308e+01	1089, 32, .866e+01	1021, 32, .489e+02
	100	1596, 30, .724e+00	1346, 31, .297E+01	1151, 31, .869E+01	1070, 31, .535e+02
C	10	> 250	> 250	> 250	> 250
	25	> 250	3575, 161, .107e+02	1727, 79, .175e+02	1444, 67, .803e+02
	50	> 250	2258, 57, .442e+01	1773, 49, .132e+02	1593, 46, .680e+02
	100	> 250	2486, 37, .315e+01	2089, 37, .943e+01	2004, 37, .672e+02
D	10	> 250	> 250	> 250	2391, 238 .253e+03
	25	> 250	1777, 86, .561e+01	1306, 61, .137e+02	1175, 57, .777e+02
	50	2049, 55, .994e+00	1445, 36, .327e+01	1181, 35, .879e+01	1086, 33, .484e+02
	100	1757, 30, .676e+00	1457, 31, .285e+01	1206, 31, .873e+01	1129, 31, .562e+02

Tabela 5.2: Quase-Newton com Fatoração Incompleta, Malha: $M = 30$

Com $l = 0$ e $l = 1$ obtivemos convergência somente para esta malha, sendo $l = 1$ o nível que apresentou o melhor resultado.

Os recomeços pioram o tempo de execução e em alguns casos inibem a convergência; para $q = 10$ na maioria dos testes ultrapassou $ItMax_{QN}$. Idealmente, o valor de q (parâmetro que fixa o número de iterações para acionar os recomeços) deveria ser maximizado, porém há uma limitação na capacidade de memória para os vetores que armazenam as atualizações secantes. Este fato que se repetirá para as outras malhas, corrobora que a melhor aproximação ao Jacobiano corresponde aquela que conserva um pouco mais o histórico que vem sendo realizado pelos metodos secantes atualizando a fatoração incompleta. Contrariamente quando recomeçamos com a ILU do Jacobiano no ponto atual, este Jacobiano Incompleto está mais longe do Jacobiano verdadeiro que com as atualizações $ILU + QN$.

Observamos que há um comportamento regular considerando a eficiência do método em relação ao nível de preenchimento com o tamanho da malha. Obviamente convém trabalhar,

	q	$l = 2$	$l = 3$	$l = 4$	$l = 5$
A	50	> 250	3742, 90, .609e+03	2484, 64, .268e+03	2471, 62, .947e+04
	100	> 250	> 250	> 250	> 250
B	50	2768, 68, 141.e+03	1670, 46, .345e+03	1289, 34, .146e+04	1193, 33, .715e+04
	100	2091, 30,.615e+02	1632, 31, .254e+03	1342, 31, .133e+04	1240, 31, .624e+04
C	50	> 250	> 250	2445, 67, .270e+04	2304, 60, .914e+04
	100	> 250	> 250	3104, 47, .176e+04	3096, 48, .724e+04
D	50	3116, 75, .138e+03	1984, 46, .333e+03	1447, 44, .176e+04	1311, 37, .587e+04
	100	2410, 32, .667e+02	1781, 31, .243e+03	1460, 31, .131e+04	1331, 31, .584e+04

Tabela 5.3: Quase-Newton com Fatoração Incompleta, Malha: $M = 40$

	$l = 2$	$l = 3$	$l = 4$	$l = 5$
A	> 250	> 250	> 250	> 250
B	2555, 30, .280E+04	1921, 31, .858E+04	1541, 31, .344+05	1431, 31, .294E+06
C	> 250	> 250	> 250	3556, 52, .405E+06
D	3462, 45, .349E+04	2140, 31, .814e+04	1682, 31, .329e+05	1544, 31, .337e+06

Tabela 5.4: Quase-Newton com Fatoração Incompleta, Malha: $M = 50$

enquanto convergência, com o menor nível de preenchimento. Entretanto, vale observar, que para NI-ILU obtivemos o melhor desempenho com $l = 4$

Em geral, o número de iterações totais diminui e o tempo de execução (CPU) aumenta com o incremento de l. Isto é devido a que por um lado o aumento de preenchimento da fatoração proporciona uma melhor aproximação ao Jacobiano sendo que por outro lado é aumentado o custo das iterações individuais pela perda de esparsidade.

Etapa	Passo de Tempo [Dias]	Iterações Q-N	Tempo de Execução [Segundos]	Erro no Balanço de Água	Erro no Balanço de Óleo
1	0.01	130	0.522E+02	0.000E+00	0.000E+00
2	0.19	47	0.673E+02	0.279E-09	0.919E-08
3	1.64	49	0.823E+02	0.106E-08	0.684E-07
4	1.99	49	0.974E+02	0.267E-08	0.189E-06
5	3.11	49	0.112E+03	0.393E-08	0.283E-06
6	5.55	52	0.128E+03	0.815E-08	0.601E-06
7	7.67	52	0.143E+03	0.167E-07	0.124E-05
8	10.2	54	0.159E+03	0.312E-07	0.228E-05
9	14.4	54	0.174E+03	0.405E-07	0.298E-05
10	19.0	55	0.189E+03	0.530E-07	0.378E-05
11	25.1	63	0.204E+03	0.796E-07	0.536E-05
12	31.7	64	0.219E+03	0.126E-06	0.771E-05
13	40.7	77	0.234E+03	0.165E-06	0.961E-05
14	50.0	92	0.250E+03	0.225E-06	0.120E-04
15	50.0	87	0.265E+03	0.261E-06	0.126E-04
16	50.0	82	0.279E+03	0.293E-06	0.139E-04
17	50.0	80	0.294E+03	0.350E-06	0.150E-04
18	50.0	81	0.309E+03	0.415E-06	0.160E-04
19	50.0	71	0.324E+03	0.465E-06	0.182E-04
20	50.0	70	0.340E+03	0.543E-06	0.205E-04
21	50.0	77	0.355E+03	0.586E-06	0.220E-04
22	50.0	70	0.370E+03	0.655E-06	0.229E-04
23	50.0	66	0.385E+03	0.756E-06	0.240E-04
24	50.0	73	0.400E+03	0.818E-06	0.249E-04
25	50.0	71	0.414E+03	0.924E-06	0.262E-04
26	50.0	68	0.429E+03	0.101E-05	0.290E-04
27	50.0	76	0.445E+03	0.114E-05	0.326E-04
28	50.0	67	0.460E+03	0.134E-05	0.364E-04
29	50.0	76	0.474E+03	0.143E-05	0.393E-04
30	50.0	69	0.489E+03	0.161E-05	0.414E-04
31	50.0	69	0.489E+03	0.165E-05	0.424E-04

Tabela 5.5: ICOL com Fatoraçào Incompleta , Nível: $l = 3$

Na Tabela 5.5 mostramos uma experiência completa da simulação realizada para um tempo total de evolução de pouco mais de 1000 dias (para isto selecionamos o método ICOL para uma malha $M = 50$ e nível $l = 3$). Para levar a cabo completamente a simulação foram necessárias 31 etapas determinadas através de uma rotina que controla automaticamente o tamanho dos passos a partir das variações máximas das variáveis entre dois passos consecutivos. Todas as experiências forma realizadas respeitando esta sequência. O tamanho do passo máximo tolerado é de 50 dias. Na terceira e quarta coluna obervamos o número de iterações e o tempo

de execução requerido para cada etapa. Nas duas últimas colunas são mostrados os erros nos balanços de massa (água e óleo) acumulados durante a simulação.

Todas as experiências foram realizadas em uma Work Station Sparc-5; os códigos computacionais usados estão escritos em linguagem FORTRAN 77, em dupla precisão.

5.5 Conclusões e Trabalhos Futuros

Neste trabalho, através de um vasto número de testes dentre os quais foram extraídos os mais relevantes, mostramos que existe uma clara superioridade computacional dos métodos quase-newtonianos combinados com Fatorações Incompletas sobre métodos iterativos precondicionados que usam esta mesma técnica. Este resultado constitui uma contribuição para a resolução deste tipo de problemas.

Os resultados obtidos com métodos quase-newtonianos com $B_0 = J(x_0)$, como implementados em Rouxinol mostraram-se pouco eficientes em relação aos outros.

Os métodos Newton Inexato com precondicionadores construidos sobre Fatorações Incompletas tiveram um melhor desempenho que quando usados com Precondicionadores Secantes, porém foram menos eficientes que os Quase-Newton com $\tilde{B}_0=\tilde{L}\tilde{U}$ com recomeços, em que $\tilde{L}\tilde{U}$ é obtida a partir de uma Fatoração Incompleta do Jacobiano.

Entre estes últimos Broyden-2 e ICOL mostraram ser os mais eficientes, convergiram em todos os casos e Broyden-2 apresentou uma leve vantagem sobre ICOL. Em condições de convergência nenhum dos métodos mostrou uma notória superioridade em relação aos outros.

Para problemas de evolução, onde um sistema não linear e/ou vários sistemas lineares subjacentes são resolvidos a cada passo do tempo, algoritmos que usam uma fatoração simbólica para as matrizes de iteração aumentam significativamente sua eficácia.

As características do problema proporcionam um vasto e diverso conjunto de possibilidades a serem pesquisadas. Com propósito similar ao do presente trabalho têm-se desenvolvido técnicas como: Decomposição de Domínios, Formulações com Implicitude Variável, etc., que eventualmente poderiam combinar-se com os métodos aqui apresentados.

Por outro lado, o modelo selecionado neste trabalho oferece a possibilidade de modificar a dependência funcional da transmissiblidade com a saturação, permitindo implementar problemas com diferente tipo de não linearidades e analisar sua resposta dos distintos métodos em cada caso. Outras possíveis escolhas, diferentes da selecionada no presente trabalho podem ser:

i) $k_{rw} = \frac{S_w^2}{(\mu_o/\mu_w)(1-S_w^2)+S_w^2}$, $k_{ro} = 1 - k_{rw}$,

ii) $k_{ro} = 1 - S_w$, $k_{rw} = S_w$,

iii) $k_{ro} = 1 - S_w^2$, $k_{rw} = S_w^2$.

Independentemente da relação de dependência funcional, em geral as funções de resíduos contêm não linearidades que podem ser encontradas nos diferentes termos que as compõem,

como foi devidamente exposto.

Os diversos métodos de solução existentes, conforme a abordagem proposta por Rodríguez [15], estão caracterizados por ter uma correspondência direta com um determinado nível de implicitude que podem ser identificados e correlacionados através de uma seleção apropriada dos diferentes termos que compõem a matriz Jacobiana. Respectivamente, existe uma relação explícita entre esse níveis de implicitude e o grau de não linearidade das equações. Esta mesma abordagem poderia ser aplicada como critério para escolher um Jacobiano Simplificado selecionando um nível de implicitude menor ao TI para construir uma matriz que determinaria uma maior esparsidade. Alternativamente sobre esta matriz "falsa" efetivar-se-iam as Atualizações Secantes o que eventualmente serviria para aumentar o padrão de esparsidade na Fatoração LU Incompleta e assim conseguir alguma economia computacional adicional.

Existem diversos problemas na Engenharía de Simulação de Reservatórios, onde o número de incógnitas é significativamente maior. Consequentemente, a necessidade de diminuir os custos computacionais passa a ser um item de importância crucial. Exemplos de problemas deste tipo são: escoamentos multifásicos com multicomponentes, refinamento de malhas em subdomínios específicos, etc.

Parâmetros físicos e geométricos do reservatório

Dimensões laterais l_c [ft]	300.0
Altura [ft]	10.0
ΔT_{min} [Dias]	0.01
ΔT_{max} [Dias]	50.0
Tempo de evolução [Dias]	1000.0
Pressão capilar P_c	0.
Viscosidade da água μ_o e do óleo μ_w [cp]	1.0
Porosidade ϕ	0.10
Permeabilidade absoluta K [mD]	12.5
Compressibilidade relativa da água c_{rw} e do óleo c_{ro} [psi^-1]	1×10^-6
Fator de formação volumétrico da água B_w e do óleo B_o [psi]	1.0
Vazão de injeção de água [bb/d] Q_i	20.
Vazão de produção de água e óleo Q_p [bb/d]	20.
Pressão inicial $p_o[psi]$	3000.
Saturação inicial S_w	.25

Bibliografia

[1] Aziz, K., Settari, A. [1979]: *Petroleum Simulation.* London, Applied Science Pub.

[2] Aziz, K. [1993]: *Notes for Petroleum Reservoir Simulation.* Stanford University.

[3] Behie, A, Forsyth, P. A. [1984]: *Incomplete Factorization Methods for Fully Implicit Simulation of Enchanced Oil Recovery.* SIAM J. Sci. Stat. Comput., [**5**], pp. 543-561.

[4] Behie, A, Vinsome, P. K. W [1982]: *Block Iterative Methods for Fully Implicit Reservoir Simulation* .Soc. Pet. Eng. J., [**22**],pp. 659-668.

[5] Bonet, L. [1990]: *Simulação Numérica de Reservatórios utilizando um Método de Implicitude Auto-adaptável.* Tese de Mestrado. FEM. UNICAMP.

[6] Broyden, C.G. [1965]: A class of methods for solving nonlinear simultaneous equations, *Mathematics of Computation* 19, pp. 577-593.

[7] Collatz, L. [1973]: *Numerical Treatment of Differential Equations.* Springer-Verlag. Berlin.

[8] Ewing, R.E. [1983]: Editor. *The Mathematics of Resevoir Simulation.* SIAM. Philadelphia.

[9] Langtangen, H. P. [1989]: *Conjugate Gradient Methods and ILU Preconditioning of Non-Symetric Matrix with Arbitrary Sparsity Paterns.*Int. J. Num. Meth. Fluids, [**9**], pp. 213-233.

[10] Langtangen, H. P. [1990]: *Implicit Finite Element Methods for Two-phase Flow in Oil Reservoirs.* Int. J. Num. Meth. Fluids, [**10**], pp. 651-681.

[11] Martínez, J.M. [1994]: *Quase-Newton Methods witht Derivatives.* Por aparecer em Calcolo.

[12] Martínez, J.M.; Zambaldi, M.C. [1992]: An inverse Column-Updating Method for solving Large-Scale Nonlinear Systems of Equations, to appear in *Optimization Methods and Software.*

[13] Peaceman, D.W. [1977]: *Fundamentals of numerical reservoir simulation.* Amsterdam.

[14] Pimentel Gomes, H. [1990]: *Modelo composicional de reservatórios com Formulação Totalmente Implícita.* Tese de mestrado. FEM. UNICAMP.

[15] Rodríguez, F. [1988]: *Un enfoque unificado de métodos de simulación numérica de yacimientos". Workshop das aplicações da ciências na engenharía de reservatórios.* Rio de Janeiro.

[16] Smith, G. D. [1987]: *Numerical solutions of partial differential equations: Finite differences methods.* Clarendon.

[17] Zambaldi, M. C. [1995]: *Métodos Quase-Newton com Fatorações Incompletas.* Relatório Técnico. DMA-CFM. UFSC.

Apêndice A

Comentários Finais

Neste trabalho analisamos o desempenho de um conjunto de algoritmos para resolver sistemas não lineares originados em problemas reais. Selecionamos para isso, diversos problemas da Física e da Engenharia e vários algoritmos cujas implementações computacionais se encontram em diferentes estágios de experimentação.

Os problemas foram escolhidos com diferentes critérios; por um lado objetivamos originar casos com diferentes dificuldades numéricas mais ou menos generalizadas em relação às não-linearidades e estruturas, por outro lado consideramos a assididuidade com que estes problemas aparecem na literatura como "problemas testes" para estimar a eficiência dos métodos. Esta última consideração nos levou a tratar com esses problemas e seus respectivos esquemas de aproximação de maneira a reproduzi-los com a maior fidelidade, antes de nos preocupar exessivamente em melhorar sua formulação numérica. Com a finalidade de estabelecer formas padrões de resolução demos um tratamento numérico unificado o que eventualmente implicou na necessidade de "ignorar" em algunas situações, particularidades físicas e numéricas próprias do problema.

Os algoritmos especializados na resolução de sistemas não lineares de grande porte, cujas implementações preexistentes foram adequadas para serem aplicadas aos sistemas, são baseados nas ideias dos métodos de Newton, Quase-Newton e Newton Inexatos.

Em cada uma das respectivas implementações existe um conjunto de parâmetros que devem ser definidos pelo usuário. Em todos os casos a realização de experiências preliminares permitiu fixar valores para alguns desses parâmetros, sendo que os considerados mais relevantes foram deixados expressamente como variáveis constituindo-se em incógnitas adicionais a serem determinadas ao longo das experiências. Devemos esclarecer que a determinação desses valores de forma a otimizar o desempenho dos pacotes computacionais foi uma preocupação secundária; assim mesmo consideramos que a sua determinação constitui uma contribuição para o melhoramento desses pacotes.

Majoritariamente os sistemas foram gerados pela aproximação de problemas de contorno

com um parâmetro que pondera o termo não linear, o qual origina dificuldades na convergência dos algoritmos. Estes problemas possibilitaram uma comparação da eficiência dos algoritmos e uma análise do efeito de introduzir diferentes estratégias de globalização.

Considerando a existência de "problemas padrões" que representam formas simplificadas das criadas pelas modelagens dos problemas reais selecionamos as equações de Poisson Não-Linear, o Problema de Bratu Modificado e o Problema de Convecção-Difusão Não-Linear, estabelecendo um marco de referência ao identificar valores do parâmetro para os quais a resolução das equações apresentam especiais dificuldades. Neste caso a nossa preocupação esteve dirigida a detectar situações problemáticas sem nos interessar exessivamente com a eficácia dos algoritmos utilizados.

Mediante o uso de técnicas elementares de continuação, as equações de Navier-Stokes foram resolvidas numa cavidade quadrada para altos números de Reynolds. O fluxo é newtoniano e incompressível em regime estacionário. A formulação das equações em termos da função-corrente define um problema de quarta ordem não-linear, com valores de fronteira. Outras técnicas de globalização foram introduzidas para resolver este problema, cujos resultados complementam os obtidos com operadores de segunda ordem.

Os métodos de globalização por otimização tiveram um desempenho pouco eficiente tanto em relação a obtenção de convergência quanto em relação ao tempo de execução.

Um novo método para a determinação de *pontos singulares* situados sobre uma curva homotópica foi testado sobre uma coleção de diversos problemas. Estes pontos estão relacionados intimamente com a estabilidade e multiplicidade das soluções. As soluções são obtidas através de sistemas aproximados aumentados. O ponto chave da formulação consiste em substituir a equação que estabelece que o espaço nulo do Jacobiano tem um vetor não nulo por uma equação em diferenças, evitando o cálculo de derivadas.

Algoritmos com atualizações Secantes sobre Fatorizações Incompletas e Newton Inexatos Precondicionados são comparados em termos de eficiência computacional para resolver as equações que regem o escoamento bifásico bidimensional num meio poroso. Os sistemas foram gerados pela discretização das equações originadas por um modelo tipo "black-oil" com uma formulação totalmente implícita.

Para os tamanhos dos problemas definidos os métodos diretos mostraram ser robustos e precisos apresentando vantagens em relação aos métodos iterativos. O número de incógnitas em cada caso foi determinado a partir das experiências realizadas por outros pesquisadores ou por uma limitação na capacidade do computador utilizado.

A eficiência dos métodos teve uma dependência mais marcante em relação às propriedades estruturais das matrizes Jacobianas do que com o tipo ou "grau" da não-linearidade das funções de resíduos.

O Método de Newton teve o melhor desempenho na consecução de convergência sendo que neste caso os métodos quase-newtonianos foram mais eficientes. Em termos gerais o Método de Newton Modificado se mostrou como uma excelente opção e deveria ser testado sempre por quem esteja interessado em obter economia no custo computacional.

Métodos Quase-Newton combinados com Fatorações Incompletas mostraram ser muito eficientes. Isto é, métodos diretos funcionam melhor que os iterativos enquanto exista capacidade de memória computacional.