## Método de Atualização Multi-Coluna Inverso para Sistemas Não-Lineares

Luziane Ferreira de Mendonça

Profa. Dra. Véra Lucia da Rocha Lopes Orientadora

> Prof. Dr. José Mario Martínez Co-Orientador

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica, UNICAMP, como requisito parcial para a obtenção do título de Mestre em Matemática Aplicada.

IMECC - UNICAMP Março de 2002

## Método de Atualização Multi-Coluna Inverso para Sistemas Não-Lineares

Luziane Ferreira de Mendonça

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Luziane Ferreira de Mendonça e aprovada pela Comissão Julgadora.

#### Banca Examinadora

Profa. Dra. Véra Lucia da Rocha Lopes Prof. Dr. Lúcio Tunes dos Santos Prof. Dr. Mario Cezar Zambaldi

Campinas, 01 de Março de 2002

Hodralopes Profa. Dra. Vera Lucia da Rocha Lopes Ban-

Prof. Dr. José Mario Martínez

Dissertação apresentada ao Instituto de Matemática. Estatística e Computação Científica. UNICAMP. como requisito parcial para a obtenção do título de Mestre em Matemática Aplicada.

> UNICAMP BIBLIOTECA CENTRAL

UNIDADE
M523m
V EX
TOMBO BC/ 49126
PROC 16-83710 0
СОХ
PRECO 125-1-1,00
DATA
Nº CPD

CM00167658-8

BIBID 255870

#### FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO IMECC DA UNICAMP

Mendonça, Luziane Ferreira de
M523m Método de atualização multi-coluna inverso para sistemas não lineares / Luziane Ferreira de Mendonça -- Campinas, [S.P. is.n.], 2002.
Orientadora : Véra Lucia da Rocha Lopes Co-orientador: José Mario Martínez Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatistica e Computação Científica.
1. Sistemas não-lineares, 2. Métodos numéricos, 3. Métodos iterativos (Matemática). I. Lopes, Véra Lucia da Rocha. II. Martínez, José Mario. III. Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica. IV. Título. Dissertação de Mestrado defendida em 01 de março de 2002 e aprovada pela Banca Examinadora composta pelos Profs. Drs.

Prof (a). Dr (a). VÉRA LŮČĨA DA ROCHA LOPES

Prof (a). Dr (a). MÁRIO ČÉSAR ZAMBALDI

Prof (a). Dr (a). LÚCIO TUNES DOS SANTOS

"Se não houver frutos valeu a beleza das flores; se não houver flores, valeu a sombra das folhas; se não houver folhas, valeu a intenção da semente."

Henfil

## Agradecimentos

Aos meus pais, Antonina e José, por serem o princípio de tudo; a você, minha mãe, pela presença e força no meu passado, presente, e futuro;

Ao meu irmão, Helder, pelas conversas, companheirismo e cumplicidade que existem apenas entre bons irmãos;

A professora Véra, pela sensibilidade, apoio e paciência; pela orientação segura e, principalmente, por sua amizade; ao professor Martínez, pela proposta de trabalho e por sua co-orientação;

Aos amigos adquiridos nesta instituição, pelo especial sentimento fraterno que fornece combustível para a vida; a todos vocês que me doaram um pouco de sua atenção e bons momentos; em particular, agradeço a Heleno, Irene, Lilian, Lucelina e Rodrigo;

À Rosana Pérez, pelos ensinamentos, experiência e amizade;

Aos professores Lúcio e Márcia, pela grande paciência e ótimas sugestões para a dissertação;

Aos demais membros da Banca Examinadora, pelos comentários que em muito contribuiram para a redação final desta dissertação.

À FAPESP, pelo suporte financeiro imprescindível para a realização deste trabalho.

### Resumo

Este trabalho propõe um método de atualização de q colunas por iteração de maneira a satisfazer (quando possível) as q últimas equações secantes; este método nem sempre está definido, pois é possível que duas equações secantes sejam imcompatíveis; mais ainda, é possível que sua compatibilidade seja tão tênue conforme q assume valores maiores que 2, que sua implementação pode ser muito mal-condiconada. É proposta uma implementação correta do ponto de vista da álgebra linear e da estabilidade numérica, a análise teórica do método (convergência local) e a determinação do número ótimo de equações secantes que devem ser usadas. Foram realizados vários testes numéricos com problemas de pequeno e grande porte presentes na literatura, fazendo uma comparação entre os métodos de Newton, Broyden, CUM, ICUM, o método multi-coluna com q igual a 2 e q igual a 3. O último capítulo é composto pela análise dos resultados obtidos por esses métodos na resolução de um problema prático geofísico (traçamento de raios em sísmica).

### Abstract

In this work it is introduced new quasi-Newton methods for solving largescale nonlinear systems of equations. In these methods q (> 1) columns of the approximation of the inverse Jacobian matrix are updated, in such a way that the q last secant equations are satisfied (when it is possible) at every iteration. The new methods obtained are called a q-Columns Inverse Updating Method. It is also shown an optimal maximum value for q, that makes the method competitive. It is proposed a right implementation from the point of view of linear algebra and numerical stability. It is presented a local convergence analysis for the case n = 2 and several numerical comparative tests with other quasi-Newton methods, in particular the ICUM (Inverse Column Updating Methods) are presented.

# Índice

	Introdução	1
1	Revisão Teórica e Algumas Conclusões Computacionais	5
	1.1 Introdução	5
	1.2 O Método de Newton	6
	1.3 Os Métodos Quase-Newton	7
	1.3.1 Método de Broyden [6]	8
	1.3.2 Método de Atualização de um Coluna por Iteração (CUM)	9
	1.3.3 Método de Atualização de uma Coluna da Jacobiana Inversa	10
	1.3.4 Método de Atualização de uma Coluna da Jacobiana Inversa Anterior	11
	1.4 Problemas-Teste Utilizados	13
	1.5 Comparação Numérica entre ICUM e ICUMA	15
2	Método de Atualização de Duas Colunas da Jacobiana	
	Inversa	19
	2.1 Descrição do Método	19
	2.2 Convergência	24

	2.2 Implementação Computacional do ITCUM e Testes			
	Numéricos	30		
3	Método de Atualização de $q \geq 3$ Colunas da Jacobiana			
	Inversa	37		
	3.1 Descrição do Método	37		
	3.2 A Não-Singularidade da Matriz $A \in I\!\!R^{qn \times qn}$	39		
	3.3 Resolvendo o Sistema $A  u^k =  v^k$	45		
	3.4 Testes Numéricos	46		
4	Traçamento de Raios em Sísmica	51		
	4.1 Descrição do Problema	51		
	4.2 Testes Numéricos	55		
	4.2.1 Problema de Grande Porte	60		
5	Conclusão	63		
Α	Algumas Passagens do Teorema 2.1	65		
	Referências	71		

# Índice de Tabelas

Tabela	1.1	Problemas de pequeno porte	17
Tabela	1.2	Problema 10 - $n = 50$	17
Tabela	1.3	Problemas de grande porte	18
Tabela	2.1	Problemas de pequeno porte	33
Tabela	2.2	Problema 10	33
Tabela	2.3	Problemas de grande porte	34
Tabela	2.4	Problemas de pequeno porte - Versões de ITCUM	35
Tabela	<b>2.5</b>	Problema 10 - Versões de ITCUM	36
Tabela	2.6	Problemas de grande porte - Versões de ITCUM	36
Tabela	3.1	Método multi-coluna, $q=1,2,3$ - Escolha 1	49
Tabela	3.2	Método multi-coluna, $q=1,2,3$ - Escolha 2	50
Tabela	4.1	Traçamento de raios em sísmica - $n = 9$	56
Tabela	4.2	Traçamento de raios - trajetórias elementares	58
Tabela	4.3	Traçamento de raios em sísmica - $n = 1001$	61

# Índice de Figuras

Figura	4.1	(a) Raio com assinatura (2, 1)	53
		(b) Raio com assinatura (2, 2)	53
Figura	4.2	Lei de Snell	53
Figura	4.3	Traçamento de raios - estrutura 1	55
Figura	4.4	Traçamento de raios em sísmica - $n = 9$	57
Figura	4.5	Traçamento de raios - estrutura 2	57
Figura	4.6	Traçamento de raios - $a^1 = (1, 1)$	59
Figura	4.7	Traçamento de raios - $a^2 = (2, 1)$	59
Figura	4.8	Traçamento de raios - $a^3 = (1, 2)$	59
Figura	4.9	Traçamento de raios - estrutura 3	60
Figura	4.1	<b>0</b> Traçamento de raios em sísmica; $n = 1001$	62
Figura	4.1	1 Traçamento de raios em sísmica; n = 1001 - Ampliação.	62

### Introdução

A resolução de um sistema de equações não lineares é uma tarefa necessária durante a resolução de problemas das mais diversas áreas (física, engenharia, economia e outras ciências).

Dada a função não-linear  $F : \mathbb{R}^n \to \mathbb{R}^n$  que possui derivadas contínuas,  $F = (f_1, \ldots, f_n)^T$ , consideremos o problema de resolver

$$F(x) = 0. \tag{1}$$

Denotaremos a matriz de derivadas parciais de F (a matriz jacobiana ou simplesmente o Jacobiano) por J(x). Na maioria dos problemas reais, o sistema não-linear resultante é de grande porte e a matriz jacobiana é estruturalmente esparsa (Dennis e Schnabel [6]). Isto significa que boa parte dos elementos de J(x) são nulos.

Existem vários métodos propostos para a resolução de sistemas não lineares. O método de Newton é o mais conhecido, e serve de base para obtenção de outros métodos eficientes; nesse método iterativo, a seqüência de aproximações  $x^k$  é gerada por:

$$x^{k+1} = x^k - J(x^k)^{-1} F(x^k);$$
(2)

o método de Newton tem convergência quadrática local (Ortega e Rheinboldt [21]).

Os métodos quase-Newton buscam, através de uma aproximação  $B_k$  para  $J(x^k)$ , evitar o cálculo da matriz jacobiana em cada iteração. A seqüência de aproximações é gerada por

$$x^{k+1} = x^k - B_k^{-1} F(x^k), (3)$$

onde a matriz  $B_{k+1}$  é obtida a partir de  $B_k$  utilizando fórmulas de recorrência baseadas em iterações anteriores. Sob certas condições, os métodos quase-Newton têm convergência local superlinear [5].

Uma classe de métodos quase-Newton das mais bem sucedidas é a dos métodos secantes. Nestes, utiliza-se como critério de escolha para a aproximação da matriz jacobiana a matriz que satisfaça a equação secante:

$$B_{k+1}s^{k} = F(x^{k+1}) - F(x^{k}) = y^{k}$$
(4)

onde  $s^k = x^{k+1} - x^k$ .

O método proposto por Broyden em 1965 é um dos métodos secantes mais conhecidos, denominado primeiro método de Broyden; a fórmula para  $B_{k+1}$  consiste numa correção de posto um da matriz  $B_k$ :

$$B_{k+1} = B_k + \frac{(y^k - B_k s^k) s^{k^T}}{s^{k^T} s^k}.$$
 (5)

Outro método secante, proposto por Martínez em 1984 [16], é o método de atualização de uma coluna por iteração, onde a matriz  $B_{k+1}$  é obtida através da atualização de uma coluna de  $B_k$ . Escolhendo o índice  $i_k$  da coluna de  $B_k$  que será modificada,  $B_{k+1}$  pode ser escrita como

$$B_{k+1} = B_k + \frac{(y^k - B_k s^k) e_{i_k}{}^T}{e_{i_k}{}^T s^k},$$
(6)

onde  $e_{i_k}$  é um vetor da base canônica do  $\mathbb{R}^n$ .

O ICUM (Inverse Column-Updating Method), proposto por Martínez e Zambaldi (1992) [18], também é um método de atualização de uma coluna por iteração onde a coluna  $j_k$  (tal que  $|y_{j_k}^k| = ||y^k||_{\infty}$ ) da matriz de aproximação para o inverso do Jacobiano  $(H_k)$  é modificada por iteração de modo a satisfazer a equação secante. Neste caso,  $H_{k+1}$  é definida por

$$H_{k+1} = H_k + \frac{(s^k - H_k y^k) e_{j_k}{}^T}{e_{j_k}{}^T y^k}.$$
(7)

Em um estudo numérico recente, Lukšan e Vlček [13] indicaram que o método inverso de atualização de uma coluna por iteração seria o método quase-Newton

mais eficiente, na prática, para resolver sistemas não-lineares de grande porte. A teoria de convergência de ICUM foi desenvolvida em Martínez e Zambaldi(1992) e em Lopes e Martínez (1995).

Em outros trabalhos (Martínez e Ochi (1982)[17]) tem sido questionada a importância da "equação secante anterior" no intuito de determinar a eficiência relativa de diferentes métodos quase-Newton.

A eficiência de ICUM e as considerações acima induzem à introdução de um método de atualização inverso de q colunas. Nesse método,  $H_k$  seria igual a  $H_{k+1}$  exceto em q colunas, as quais seriam atualizadas de maneira que as últimas q equações secantes fossem satisfeitas.

A definição desse método suscita diversas questões. Em primeiro lugar, devese observar que o método nem sempre está definido, pois é possível que duas equações secantes sejam incompatíveis. Mais ainda, é possível que sua compatibilidade torne-se tão tênue conforme q assume valores maiores que 2, que sua implementação pode ser muito mal-condicionada. Por isso, é necessário analisar cuidadosamente a álgebra linear estável que deve ser utilizada para a implementação desse método, nos casos em que existe.

As propriedades teóricas do método também precisam ser estudadas. Nos dois extremos temos o método ICUM e o método secante sequencial [15], cujas propriedades são bem conhecidas. Mas as propriedades dos métodos intermediários não são perfeitamente claras.

Para problemas de grande porte, é claro que o método ICUM é muito mais eficiente que o método secante sequencial que, de fato, nem pode ser implementado razoavelmente para esse tipo de problema. Mas o método baseado nas equações secantes anteriores deverá ser mais eficiente que ICUM. Portanto, a determinação do número ótimo de equações secantes que devem ser satisfeitas é um problema prático interessante.

Neste trabalho propomos o método de atualização de q colunas por iteração, de maneira a satisfazer (quando possível) as q últimas equações secantes. Propomos uma implementação correta do ponto de vista da álgebra linear e da estabilidade numérica; fazemos a análise teórica do método (convergência local) e a determinação do número ótimo de equações secantes que devem ser usadas. Nesse sentido, este trabalho está organizado da forma como se segue. No Capítulo 1, descrevemos os vários métodos utilizados nos testes numéricos, com ênfase maior em ICUM (método de atualização de uma coluna da matriz jacobiana inversa) e ICUMA (método análogo ao anterior, onde a matriz de aproximação da matriz jacobiana respeita a equação secante anterior). No Capítulo 2, apresentamos o método ITCUM, que corresponde ao método de atualização de 2 colunas por iteração, um algoritmo para o método, fazemos uma análise de convergência do mesmo e testes numéricos, onde há uma comparação dos resultados obtidos por este e os outros métodos citados anteriormente quando aplicados a problemas variados. No Capítulo 3, descrevemos o método de atualização de q colunas por iteração, com  $q \ge 3$ . Uma aplicação a um problema prático compõe o Capítulo 4. Por último, o Capítulo 5 contém as conclusões e os comentários finais. Antes das referências bibliográficas, incluimos um Apêndice no qual apresentamos algumas passagens que auxiliam na compreensão da demonstração do teorema de convergência de ITCUM.

## Capítulo 1

# Revisão Teórica e Algumas Conclusões Computacionais

### 1.1 Introdução

Neste capítulo, descrevemos alguns métodos, escolhidos para implementação dos testes numéricos da Seção (2.4) e para o problema prático do Capítulo 4, com exceção dos métodos gerados pela troca de mais de uma coluna da aproximação da matriz jacobiana inversa, que estão expostos nos Capítulos 2 e 3 deste trabalho. São estes os métodos aqui descritos:

- Newton [6],
- Broyden [6],
- Atualização de Uma Coluna por Iteração (CUM)[16],
- Atualização de Uma Coluna da Jacobiana Inversa (ICUM)[18],
- Atualização de Uma Coluna da Jacobiana Inversa Satisfazendo a Equação Secante Anterior (ICUMA)[17].

O problema típico que estamos interessados em resolver, está descrito na introdução deste trabalho (equação (1)):

$$F(x) = 0, \quad F: I\!\!R^n \to I\!\!R^n, \quad F \in C^1(I\!\!R^n),$$

$$F(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{pmatrix} e J(x) = F'(x) = \begin{pmatrix} f'_1(x) \\ \vdots \\ f'_n(x) \end{pmatrix} = \begin{pmatrix} \nabla f_1^T(x) \\ \vdots \\ \nabla f_n^T(x) \end{pmatrix}.$$
 (1.1)

A resolução dos sistemas lineares gerados pelos algoritmos é feita através da fatoração LU com estratégia de pivoteamento parcial.

### 1.2 O método de Newton

Como todo método iterativo, a solução  $x^*$  é aproximada por uma sequência de pontos  $\{x^k\}$ , os quais são obtidos na forma como segue.

Considere a aproximação de Taylor de primeira ordem de F, numa vizinhança do ponto  $x^k$ :

$$F(x) \approx L_k(x) = F(x^k) + J(x^k)(x - x^k).$$
 (1.2)

O ponto  $x^{k+1}$  é solução do modelo local linear para F construído em torno de  $x^k$  (equação (1.2)), ou seja,  $L_k(x^{k+1}) = 0$ ; logo, uma iteração Newton consiste em resolver:

$$J(x^k)s^k = -F(x^k) \quad \text{e} \quad x^{k+1} = x^k + s^k.$$
(1.3)

A implementação de (1.3) pressupõe, a cada iteração, o cálculo de  $J(x^k)$  e  $F(x^k)$ , mais  $\mathcal{O}(n^3/3)$  operações necessárias para resolver o sistema (via fatoração LU). Dessa forma, para a maioria dos problemas de grande porte (n grande), o método de Newton envolve um alto custo computacional. Esta desvantagem é compensada pela taxa quadrática de convergência se trabalharmos em uma certa

vizinhança de  $x^*$  (condição (1.4)), que é dada pelo seguinte teorema (Dennis e Schanabel [6]):

**Teorema 1.1** Seja  $F : D \subset \mathbb{R}^n \to \mathbb{R}^n$  uma função continuamente diferenciável no conjunto aberto D,  $F(x^*) = 0$ ,  $J(x^*)$  não singular; seja  $x^k$  a seqüência gerada pelo método de Newton; suponhamos que existam L, p > 0 tais que para todo  $x \in D$ ,

$$||J(x) - J(x^*)|| \le L ||x - x^*||^p.$$
(1.4)

Então existe  $\varepsilon > 0$  tal que se  $||x^0 - x^*|| \le \varepsilon$ , a sequência  $x^k$  está bem definida, converge para  $x^*$  e satisfaz

$$\|x^{k+1} - x^*\| \le c \|x^k - x^*\|^{p+1}.$$
(1.5)

A convergência quadrática é obtida quando p = 1.

Na sua forma básica, este é um método iterativo local, onde apenas podemos garantir a convergência a uma solução se o ponto inicial é suficientemente bom, embora na prática possa haver convergência mesmo que a aproximação inicial não seja boa.

### 1.3 Os Métodos Quase-Newton

Os métodos quase-Newton buscam reduzir o número de operações realizadas por iteração, tentando ser "tão bons" quanto o método de Newton, no sentido de manter ao máximo as propriedades de convergência deste.

Para resolver o sistema linear (1.3) da maneira mais adequada possível, esses métodos estabelecem uma aproximação  $B_k$  para a matriz jacobiana  $J(x^k)$ ; porém, essa redução nos custos acarreta uma redução na taxa de convergência, que no melhor dos casos é superlinear.

Tais métodos são caracterizados pelo processo iterativo:

$$B_k s^k = -F(x^k) \ e \ x^{k+1} = x^k + s^k.$$
(1.6)

Uma família de métodos quase-Newton bem sucedida é a dos métodos secantes. Em analogia com o que foi feito para o método de Newton, a função F é aproximada em torno de  $x^{k+1}$  por

$$F(x) \approx L_{k+1}(x) = F(x^{k+1}) + B_{k+1}(x - x^{k+1}).$$
(1.7)

Obrigando a função linear  $L_{k+1}$  a interpolar F no ponto  $x^k$  (observe que a igualdade  $L_{k+1}(x^{k+1}) = F(x^{k+1})$  é satisfeita para qualquer matriz  $B_{k+1}$ ), obtemos

$$y^{k} = F(x^{k}) - F(x^{k+1}) = B_{k+1}(x^{k} - x^{k+1}) = B_{k+1}s^{k}$$

Logo, as matrizes de aproximação para a matriz jacobiana são escolhidas de modo a satisfazer a "equação secante":

$$B_{k+1}s^k = y^k. (1.8)$$

A equação secante não determina univocamente a matriz de aproximação (basta interpretar a equação (1.6) como sendo um sistema cujas variáveis são as componentes de  $B_{k+1}$ , ou seja,  $n^2$  variáveis, e com somente *n* equações). São as condições adicionais impostas para  $B_{k+1}$  que definem diferentes métodos secantes.

Na maioria dos casos, consegue-se provar resultados de convergência superlinear para os métodos secantes.

#### 1.3.1 Método de Broyden [6]

Impondo que a matriz de aproximação  $B_{k+1}$  possa ser calculada a partir de  $B_k$  com a adição de uma matriz de posto unitário,  $B_{k+1} = B_k + \Delta B_k$ , temos:

$$B_{k+1}s^k = y^k \Rightarrow B_{k+1}s^k - B_ks^k = y^k - B_ks^k \Rightarrow \triangle B_ks^k = y^k - B_ks^k.$$

Podemos então tomar

$$\triangle B_k = \frac{(y^k - B_k s^k) w_k^T}{w_k^T s^k} \tag{1.9}$$

com  $w_k \in \mathbb{R}^n$  arbitrário não ortogonal a  $s^k$ .

O primeiro método de Broyden é obtido ao fazermos a escolha  $w_k = s^k$ , ou seja,

$$\Delta B_k = \frac{(y^k - B_k s^k) s^{k^1}}{s^{k^T} s^k}.$$
 (1.10)

Tal método é equivalente que a matriz  $B_{k+1}$  deve ser aquela que, dentre as matrizes que satisfazem a equação secante, é a mais próxima de  $B_k$ , considerando a norma 2 (ou na norma de Frobenius). Geometricamente, isto significa que  $B_{k+1}$  é a projeção ortogonal de  $B_k$  no conjunto das matrizes que satisfazem a equação secante.

Neste tipo de correção de posto um,  $B_{k+1}^{-1}$  também pode ser obtida a partir de  $B_k^{-1}$  por meio de uma correção de posto um, fazendo uso da fórmula de Sherman-Morrison [8].

#### 1.3.2 Método de Atualização de uma Coluna por iteração (CUM)

Proposto por Martínez [16], este é um método quase-Newton em que a cada iteração, apenas uma coluna de  $B_k$  é modificada para obter  $B_{k+1}$ , que deve satisfazer a equação secante.

O cálculo de  $B_{k+1}$  começa com a escolha do índice  $i_k$  da coluna de  $B_k$  que deverá ser modificada. Essa escolha deverá ser tal que

$$|s_{i_k}^k| > \alpha ||s^k||_{\infty}, \tag{1.11}$$

 $\operatorname{com} \alpha \in (0, 1].$ 

Dessa forma, é possível escrever  $B_{k+1}$  como

$$B_{k+1} = B_k + d^k e_{i_k}^T, (1.12)$$

onde  $e_{i_k}$  é um vetor da base canônica do  $\mathbb{R}^n$ . Então:

$$B_{k+1}s^{k} = y^{k} \Rightarrow (B_{k} + d^{k}e_{i_{k}}^{T})s^{k} = y^{k} \Rightarrow d^{k} = \frac{y^{k} - B_{k}s^{k}}{e_{i_{k}}^{T}s^{k}}.$$
 (1.13)

Portanto, a equação (1.12) pode ser escrita como

$$B_{k+1} = B_k + \left(\frac{y^k - B_k s^k}{e_{i_k}^T s^k}\right) e_{i_k}^T \tag{1.14}$$

A matriz  $B_{k+1}$  é uma matriz da forma  $B_k + ab^T$  e, pela mesma razão do método de Broyden, o cálculo de  $B_{k+1}^{-1}$  pode ser feito com o uso da fórmula de Sherman-Morrison.

Este método possui convergência local linear, obtida para uma versão com reinício (a taxa superlinear é obtida, em particular, se  $B_k = J(x^k)$  sempre que  $k \equiv 0 \mod m$ ). Gomes-Ruggiero, Martínez e Moretti [9] mostraram que o CUM é muito eficiente quando comparado com a implementação do primeiro método de Broyden com memória limitada na resolução de sistemas não lineares de grande porte.

#### 1.3.3 Método de Atualização de uma Coluna da Jacobiana Inversa (ICUM)

A concepção desse método proposto por Martínez e Zambaldi [18] é inspirada pelo método CUM. No Inverse Column-Updating Method (ICUM), apenas uma coluna da matriz de aproximação da inversa da matriz jacobiana é atualizada a cada iteração, de modo que a equação secante sempre seja satisfeita.

De forma análoga à seção anterior, o cálculo da aproximação  $H_{k+1}$  da inversa da matriz jacobiana  $J(x^{k+1})$  se inicia com a escolha do índice  $j_k$  da coluna de  $H_k$  que deverá ser modificada. Esse índice deverá satisfazer à desigualdade

$$|y_{j_k}^k| \ge \alpha ||y^k||_{\infty}, \quad \alpha > 0; \tag{1.15}$$

em geral, é utilizado o valor  $\alpha = 1$ .

Assim,  $H_{k+1}$  pode ser escrita como

$$H_{k+1} = H_k + d^k e_{j_k}^T, (1.16)$$

onde  $e_{j_k}$  é um vetor da base canônica do  $\mathbb{R}^n$ .

Portanto,

$$H_{k+1}y^k = s^k \Rightarrow (H_k + d^k e_{j_k}^T)y^k = s^k \Rightarrow d^k = \frac{s^k - H_k y^k}{e_{j_k}^T y^k}.$$
 (1.17)

A nova matriz  $H_{k+1}$  pode ser obtida através da equação

$$H_{k+1} = H_k + \left(\frac{s^k - H_k y^k}{e_{j_k}^T y^k}\right) e_{j_k}^T, \qquad (1.18)$$

ou seja,  $H_{k+1}$  difere de  $H_k$  exceto na coluna  $j_k$ , que é determinada por

$$h_{k+1}^{ij_k} = \frac{s_i^k - \sum_{l \neq j_k} h_k^{il} y_l^k}{y_{j_k}^k}.$$
(1.19)

Este método tem as mesmas propriedades de convergência de CUM e, se o método é reiniciado periodicamente, obtemos convergência *R*-superlinear. Não necessariamente o reínicio precisa ser realizado com a "matriz jacobiana verdadeira"; em nossos testes numéricos, utilizamos a projeção ortogonal da matriz jacobiana no espaço das matrizes tridiagonais ou no espaço das matrizes diagonais.

Seu desempenho prático é muito satisfatório quando comparado ao CUM, ao método de Newton e ao primeiro método de Broyden.

#### 1.3.4 Método de Atualização de uma Coluna da Jacobiana Inversa Anterior (ICUMA)

É possível compreender este método como sendo o próprio ICUM, com uma pequena alteração.

A equação secante (4) significa que o modelo linear  $L_{k+1}(x) = F(x^{k+1}) + B_{k+1}(x - x^{k+1})$  interpola F em  $x^k$  e  $x^{k+1}$ . Por analogia com (4), definimos a "equação secante anterior" como sendo

$$B_{k+1}s^{k-1} = y^{k-1} (1.20)$$

onde  $y^{k-1} = F(x^k) - F(x^{k-1})$ .

No método de atualização de uma coluna da matriz jacobiana inversa anterior, uma coluna da matriz de aproximação da inversa da matriz jacobiana  $H_{k+1}$  é atualizada a cada iteração, de modo que a equação secante anterior seja satisfeita, ou seja,

$$H_{k+1}y^{k-1} = s^{k-1}. (1.21)$$

Como em ICUM, a matriz  $H_{k+1}$  pode ser escrita por

$$H_{k+1} = H_k + r^k e_{p_k}^T, (1.22)$$

onde  $e_{p_k}^T$  é um vetor pertencente à base canônica do  $\mathbb{R}^n$ , e  $p_k$  representa o índice da coluna de  $H_k$  que deverá ser alterada.

De forma semelhante ao descrito na seção anterior, esse índice deverá ser escolhido de tal sorte que

$$|y_{p_k}^{k-1}| \ge \alpha ||y^{k-1}||_{\infty}, \quad \alpha > 0.$$
(1.23)

De (1.21) e (1.22), obtemos:

$$H_{k+1}y^{k-1} = s^{k-1} \Rightarrow (H_k + r^k e_{p_k}^T)y^{k-1} = s^{k-1} \Rightarrow r^k = \frac{s^{k-1} - H_k y^{k-1}}{e_{p_k}^T y^{k-1}}, \quad (1.24)$$

e então

$$H_{k+1} = H_k + \left(\frac{s^{k-1} - H_k y^{k-1}}{e_{p_k}^T y^{k-1}}\right) e_{p_k}^T.$$
(1.25)

Em nossos testes numéricos, de forma análoga à implementação de ICUM, escolhemos o índice  $p_k$ tal que

$$|y_{p_k}^{k-1}| = ||y^{k-1}||_{\infty}.$$
(1.26)

### 1.4 Problemas-Teste Utilizados

Para todos os métodos descritos e introduzidos neste capítulo, foram utilizados os problemas a seguir, extraídos da literatura.

Os problemas teste estão separados em duas classes, pequeno e grande porte de acordo com o número de variáveis que possuem. Eles são utilizados em todos os testes efetuados neste trabalho.

#### 1) Problemas de pequeno porte: [12] e [20]

**Problema 1:** Rosenbrock (n=2). Função 1 de Moré, Garbow and Hillstrom [20].  $x_0 = (-1.2, 1)^T$ .

**Problema 2:** Freudenstein-Roth (n=2). Função 2 de Moré, Garbow e Hillstrom [20].  $x_0 = (0.5, -2)^T$ .

**Problema 3:** Powell badly scaled function (n=2). Função 3 de Moré, Garbow e Hillstrom [20].  $x_0 = (0.5, -2)^T$ .

**Problema 4:** Powell singular function (n=4). Função 13 de Moré, Garbow e Hillstrom [20].  $x_0 = (3, -1, 01)^T$ .

**Problema 5:** Extended Rosenbrock (n=50). Função 21 de Moré, Garbow e Hillstrom [20].  $x_0 = (-1.2, 1, -1.2, 1, ...)^T$ .

**Problema 6:** Trigonometric function (n=2). Função 26 de Moré, Garbow e Hillstrom [20].  $x_0 = (1/n, ..., 1/n)^T$ .

**Problema 7:** Discrete boundary value function (n=2). Função 28 de Moré, Garbow e Hillstrom [20].  $x_0 = (\xi_j)$ , onde  $\xi_j = t_j(t_j - 1)$ , h = 1/(n+1) e  $t_j = jh$ .

**Problema 8:** Broyden banded function (n=2). Função 30 de Moré, Garbow e Hillstrom [20].  $x_0 = (-1, ..., -1)^T$ .

**Problema 9:** Linear System (n=50). Função 4 de Lopes e Martínez [12].  $x_0 = (1, -1, 1, -1, ...)^T$ . **Problema 10:** Chandrasekhar H-equation (n=50). Função 8 de Lopes e Martínez [12].  $x_0 = (0, ..., 0)^T$ 

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

discretizando a integral por meio de retângulos, considerando a imagem do ponto médio de cada intervalo (regra do ponto médio), obtemos

$$f_i(x) = -x_i + 1 + \frac{c}{2n} \sum_{j=1}^n \frac{t_i x_i x_j}{t_i + t_j}$$

#### 2) Problemas de grande porte: [18]

 $A_0$ ,  $A_2$ ,  $A_4$ ,  $B \in C$ : Discretizações da equação de Poisson no quadrado [0, 1] × [0, 1], utilizando diferenças finitas (N=32 e 50). Correspondem aos problemas  $A_0$ ,  $A_2$ ,  $A_4$ ,  $B \in C$  de Martínez e Zambaldi [18]. Em todos os casos, o ponto inicial é  $x_0 = (-1, -1, ..., -1)^T$  e o número de variáveis é igual a  $n = (N-1)^2$ .

$$\mathbf{A}_{i}: \quad \Delta u = \frac{10^{i}u^{3}}{1+s^{2}+t^{2}}, \qquad i = 0, 2, 4.$$

$$u(s,t) = \begin{cases} 1. & s = 0, \quad t \in [0, 1] \\ 1. & t = 0, \quad s \in [0, 1] \\ 2-e^{s}. & t = 1, \quad s \in [0, 1] \\ 2-e^{t}. & s = 1, \quad t \in [0, 1] \end{cases}$$

$$\mathbf{B}: \quad \Delta u = u^{3}, \qquad u(s,t) = 0 \text{ na fronteira.}$$

$$\mathbf{C}: \quad \Delta u = 0, \qquad u(s,t) = 0 \text{ na fronteira.}$$

As matrizes jacobianas dos problemas de grande porte são esparsas com estrutura pentadiagonal; logo, são bem representadas por sua parte tridiagonal. Motivados por este fato, em nossos testes numéricos fazemos

$$H_k = [\mathcal{P}_{\tau}(J(x^k))]^{-1}, \text{ se } k = 0 \text{ ou } k \equiv 1 \pmod{m+1},$$
 (1.27)

onde  $\mathcal{P}_{\tau}(J(x^k))$  é a projeção da matriz  $J(x^k)$  no espaço das matrizes tridiagonais. Os algoritmos são reiniciados utilizando m = 30.

Para os demais problemas, fazemos:

$$H_0 = [\mathcal{P}_{\mathcal{D}}(J(x^0))]^{-1}, \qquad (1.28)$$

onde  $\mathcal{P}_{\mathcal{D}}(J(x^k))$  é a projeção da matriz  $J(x^k)$  no espaço das matrizes diagonais. Caso algum elemento da diagonal da matriz jacobiana tenha valor nulo, o substituimos por 1. Os dois métodos são implementados sem reinício, e assim como nos problemas de grande porte, a forma de implementação do ICUM é baseada em [18].

A convergência é obtida quando

$$||F(x^k)||_{\infty} \le 10^{-5}.$$

A execução do teste numérico será interrompida se o número de iterações exceder 300, ou se

$$||F(x^k)||_{\infty} \ge 10^{20} ||F(x^0)||_{\infty}.$$
(1.29)

Neste último caso, dizemos que o método diverge (isto é indicado nas tabelas com o termo Div).

Os testes numéricos foram realizados em um computador AMD Athlon - 800MHz (financiado pela FAPESP, com os recursos da reserva técnica destinada ao desenvolvimento desta Dissertação), e os algoritmos foram programados no software Matlab 6.0.

### 1.5 Comparação Numérica entre ICUM e ICUMA

Com o intuito de verificar o desempenho do ICUMA, descrito na seção 1.3.4, apresentamos no que segue a idéia central da sua implementação numérica e

alguns testes numéricos comparando-o com o ICUM na resolução de sistemas não lineares.

Substituindo k pelo valor de k-1 em (1.25), e utilizando essa igualdade de forma sucessiva, obtemos

$$H_k = H_0 + \sum_{l=0}^{k-1} \left( \frac{s^{l-1} - H_l y^{l-1}}{e_{p_l}^T y^{l-1}} \right) e_{p_l}^T,$$
(1.30)

que é equivalente a

$$H_k = H_0 + \sum_{l=0}^{k-1} \left(\frac{w_l}{e_{p_l}^T y^{l-1}}\right) e_{p_l}^T,$$
(1.31)

onde  $w_l = s^{l-1} - H_l y^{l-1}$ .

A igualdade (1.31) é a base da implementação de ICUMA; observe que a cada iteração é necessário armazenar um vetor e um índice, limitando assim o número de iterações do tipo ICUMA que podem ser realizadas de forma consecutiva. Considerando que há memória suficiente para armazenar m vetores, o algoritmo será reiniciado a cada m iterações.

Na **Tabela 1.3**, os resultados em cada experimento são representados pelo par (KON; TIME), onde KON é o número de iterações e TIME é o tempo de CPU em segundos. Nas **Tabelas 1.1** e **1.2**, é apresentado apenas o número de iterações <sup>1</sup>.

<sup>&</sup>lt;sup>1</sup>A justificativa para essa escolha consiste no fato de que o tempo necessário para a convergência de problemas de pequeno porte é bem reduzido.

Prob	ICUM	ICUMA
01	13	Div
02	19	Div
03	83	101
04	Div	Div
05	08	42
06	09	08
07	05	05
08	05	06
09	04	05

Tabela 1.1: Problemas de pequeno porte.

С	ICUM	ICUMA
0.1	04	04
0.5	06	06
0.9	09	09
0.99	12	13
0.999	13	24
$1 - 10^{-4}$	15	23
$1 - 10^{-5}$	16	23
$1 - 10^{-6}$	17	24
$1 - 10^{-7}$	17	24
$1 - 10^{-8}$	17	24
1	17	24

**Tabela 1.2:** Problema 10 - n = 50.

Prob	N	n	ICUM	ICUMA
$A_0$	32	961	(86; 7.79)	(109; 10.49)
	50	2401	(155; 53.17)	(174; 60.37)
$A_2$	32	961	(70; 6.48)	(88; 8.57)
	50	2401	(104; 35.87)	(125; 45.31)
$A_4$	32	961	(77; 7.08)	(82; 7.97)
	50	2401	(103; 35.37)	(124; 46.36)
B	32	961	(62; 5.54)	(90; 8.07)
	50	2401	(92; 29.55)	(149; 48,33)
C	32	961	(61; 5.00)	(73; 6.15)
	50	2401	(100; 30.04)	(211; 67.37)

Tabela 1.3: Problemas de grande porte.

Como pode ser observado, ICUM apresenta um melhor desempenho em todos os problemas selecionados, embora o método ICUMA comporte-se de forma razoável. Para os problemas de grande porte, ainda foram realizados testes para outros valores de N (N = 45 e N = 55), onde foi constatado o mesmo comportamento.

## Capítulo 2

# Método de Atualização de Duas Colunas da Jacobiana Inversa

#### 2.1 Descrição do Método

Consideremos o problema de resolver o sistema linear dado por (1), onde a função  $F : \mathbb{R}^n \to \mathbb{R}^n$  tem derivadas contínuas e seja m um inteiro positivo. Suponhamos que  $x^0 \in \mathbb{R}^n$  seja a uma aproximação inicial para a solução de (1) e que  $H_0 \in \mathbb{R}^{n \times n}$ .

O método de atualização de duas colunas da matriz jacobiana inversa (ITCUM) para este problema consiste na iteração: dado  $x^k \in \mathbb{R}^n$ ,  $H_k \in \mathbb{R}^{n \times n}$ ,  $F(x^k) \neq 0$ ,

$$x^{k+1} = x^k - H_k F(x^k), (2.1)$$

onde  $H_k$  é atualizada de tal forma que  $H_{k+1}$  difira da matriz anterior em duas colunas, e satisfaça as duas últimas equações secantes <sup>1</sup>:

$$H_{k+1}y^k = s^k \tag{2.2}$$

$$H_{k+1}y^{k-1} = s^{k-1}, (2.3)$$

onde  $s^k = x^{k+1} - x^k$  e  $y^k = F(x^{k+1}) - F(x^k)$ .

 $<sup>$$^{1}</sup>Se$$ a última equação secante (2.2) e a equação secante anterior (2.3) são satisfeitas, o modelo linear cuja solução é  $x^{k+1}$  interpola a função F no pontos  $x^{k+1}$ ,  $x^{k}$  e  $x^{k-1}$ .

Portanto, a matriz  $H_{k+1}$  pode ser uma correção de posto 2 da matriz  $H_k$ , isto é:

$$H_{k+1} = H_k + u_{i_1}^k e_{i_1}^T + u_{i_2}^k e_{i_2}^T, (2.4)$$

onde  $e_{i_1} e e_{i_2}$  pertencem à base canônica do  $\mathbb{R}^n$ , e  $u_{i_1}^k e u_{i_2}^k$  são, respectivamente, os vetores de ordem n responsáveis pelas correções a serem realizadas nas colunas<sup>2</sup>  $i_1 e i_2$  de  $H_k$ , e devem ser tais que as equações dadas em (2.2) e (2.3) sejam satisfeitas.

Observe que as equações (2.2) e (2.3) podem ser incompatíveis e, portanto, o método pode não estar definido. Com o propósito de fazer uma análise do método e determinar condições para uma boa definição de ITCUM, consideremos as equações dadas em (2.2) e (2.3) com  $H_{k+1}$  definida em (2.4):

$$\begin{cases} (H_k + u_{i_1}^k e_{i_1}^T + u_{i_2}^k e_{i_2}^T) y^k = s^k \\ (H_k + u_{i_1}^k e_{i_1}^T + u_{i_2}^k e_{i_2}^T) y^{k-1} = s^{k-1} \end{cases}$$
(2.5)

de forma equivalente,

$$\begin{cases} u_{i_1}^k(e_{i_1}^T y^k) + u_{i_2}^k(e_{i_2}^T y^k) &= s^k - H_k y^k \\ u_{i_1}^k(e_{i_1}^T y^{k-1}) + u_{i_2}^k(e_{i_2}^T y^{k-1}) &= s^{k-1} - H_k y^{k-1}. \end{cases}$$
(2.6)

As equações (2.6) formam um sistema linear de 2n equações com 2n incógnitas, onde as variáveis e os escalares são os componentes dos vetores  $u_{i_1}^k e u_{i_2}^k$  e dos vetores  $y^k e y^{k-1}$ , respectivamente.

Para facilitar a notação, definimos

Logo, usando (2.7), o sistema (2.6) pode ser visto na forma matricial como

$$Au^{k} = \left(\frac{\alpha^{k}\mathbf{I} \mid \beta^{k}\mathbf{I}}{\gamma^{k}\mathbf{I} \mid \delta^{k}\mathbf{I}}\right) \left(\frac{u_{i_{1}}^{k}}{u_{i_{2}}^{k}}\right) = \left(\frac{s^{k} - H_{k}y^{k}}{s^{k-1} - H_{k}y^{k-1}}\right) = \left(\frac{v_{1}^{k}}{v_{2}^{k}}\right) = v^{k}, \quad (2.8)$$

<sup>&</sup>lt;sup>2</sup>Para facilitar a notação ao longo deste capítulo, escreveremos sempre  $i_1^k e i_2^k$  como, respectivamente,  $i_1 e i_2$ .

onde  $A \in \mathbb{R}^{2n \times 2n}$ , I é a matriz identidade de ordem  $n, u^k \in \mathbb{R}^{2n}$  e  $v^k \in \mathbb{R}^{2n}$ .

Portanto, a existência dos vetores  $u_{i_1}^k \in u_{i_2}^k$  satisfazendo (2.2) e (2.3) vai estar determinada pela não singularidade da matriz A. É fácil ver que o determinante de A é dado por:

$$\det(A) = \left[\det\left(\begin{array}{cc} \alpha^k & \beta^k \\ \gamma^k & \delta^k \end{array}\right)\right]^n = \det(\bar{A})^n.$$
(2.9)

Isto mostra um fato interessante: para analisar a não singularidade da matriz A de ordem 2n, basta analisar a não singularidade da matriz  $\overline{A}$  de ordem 2.

Supondo  $\sigma^k = \det(\bar{A}) \neq 0$ , a matriz A será não singular. Assim, para encontrar uma expressão geral para o vetor  $u^k \text{ em } (2.8)$ , deve-se resolver um sistema linear. Isto pode ser feito via fatoração LU, por exemplo, que é a estratégia que usamos a seguir.

Caso 1- Se  $|\alpha^k| \ge |\gamma^k|$ :

$$LU = \begin{pmatrix} \mathbf{I} & | & \mathbf{O} \\ \frac{--}{\alpha^{k}} & -- & -- \\ \frac{\gamma^{k}}{\alpha^{k}} \mathbf{I} & | & \mathbf{I} \end{pmatrix} \begin{pmatrix} \alpha^{k} \mathbf{I} & | & \beta^{k} \mathbf{I} \\ \frac{--}{\alpha^{k}} & -- & -- \\ \mathbf{O} & | & \frac{\alpha^{k} \delta^{k} - \beta^{k} \gamma^{k}}{\alpha^{k}} \mathbf{I} \end{pmatrix}$$
$$= \begin{pmatrix} \alpha^{k} \mathbf{I} & | & \beta^{k} \mathbf{I} \\ \frac{--}{\gamma^{k}} & -- & -- \\ \frac{--}{\gamma^{k}} & \frac{--}{\alpha^{k}} \end{pmatrix} = A.$$

Caso 2- Se  $|\alpha^k| < |\gamma^k|$ :

$$LU = \begin{pmatrix} \mathbf{I} & | & \mathbf{O} \\ \frac{--}{\gamma^{k}} \mathbf{I} & | & \mathbf{I} \end{pmatrix} \begin{pmatrix} \gamma^{k} \mathbf{I} & | & \delta^{k} \mathbf{I} \\ \frac{--}{\gamma^{k}} - \frac{--}{\gamma^{k}} \mathbf{I} \\ \mathbf{O} & | & \frac{\gamma^{k} \beta^{k} - \alpha^{k} \delta^{k}}{\gamma^{k}} \mathbf{I} \end{pmatrix}$$
$$= \begin{pmatrix} \mathbf{O} & | & \mathbf{I} \\ \mathbf{I} & | & \mathbf{O} \end{pmatrix} \begin{pmatrix} \alpha^{k} \mathbf{I} & | & \beta^{k} \mathbf{I} \\ \frac{--}{\gamma^{k}} - \frac{--}{\gamma^{k}} \mathbf{I} \end{pmatrix} = PA.$$

Utilizando a fatoração LU obtida acima, resolvemos o sistema (2.8):

$$LUu^k = Pv^k,$$

obtendo, em ambos os casos,

$$u^{k} = \begin{pmatrix} \frac{\delta^{k} v_{1}^{k} - \beta^{k} v_{2}^{k}}{\sigma^{k}} \\ \frac{\alpha^{k} v_{2}^{k} - \gamma^{k} v_{1}^{k}}{\sigma^{k}} \end{pmatrix} = \begin{pmatrix} u_{i_{1}}^{k} \\ u_{i_{2}}^{k} \end{pmatrix}, \qquad (2.10)$$

onde  $\sigma^k = \alpha^k \delta^k - \beta^k \gamma^k$ .

Substituindo (2.10) em (2.4), obtemos:

$$H_{k+1} = H_k + \left(\frac{\delta^k v_1^k - \beta^k v_2^k}{\sigma^k}\right) e_{i_1}^T + \left(\frac{\alpha^k v_2^k - \gamma^k v_1^k}{\sigma^k}\right) e_{i_2}^T.$$
 (2.11)

Como visto anteriormente, a matriz  $H_{k+1}$  difere da matriz  $H_k$  em apenas duas colunas  $(i_1 \in i_2)$ ; a partir da igualdade (2.11), é possível escrever essas colunas da seguinte forma:

Substituindo os vetores  $v_1^k \in v_2^k$  definidos em (2.8) nas igualdades acima, obtemos:

$$h_{i_{1}}^{k+1} = h_{i_{1}}^{k} + \frac{\delta^{k}}{\sigma^{k}} \left( s^{k} - H_{k} y^{k} \right) - \frac{\beta^{k}}{\sigma^{k}} \left( s^{k-1} - H_{k} y^{k-1} \right),$$
  

$$h_{i_{2}}^{k+1} = h_{i_{2}}^{k} + \frac{\alpha^{k}}{\sigma^{k}} \left( s^{k-1} - H_{k} y^{k-1} \right) - \frac{\gamma^{k}}{\sigma^{k}} \left( s^{k} - H_{k} y^{k} \right).$$
(2.13)

A partir de (2.13), cada componente *j* das colunas a serem modificadas será atualizada da forma como segue:

$$h_{j\,i_{1}}^{k+1} = \frac{\delta^{k}}{\sigma^{k}} \left( s_{j}^{k} - \sum_{p \neq i_{1}} h_{j\,p}^{k} y_{p}^{k} \right) - \frac{\beta^{k}}{\sigma^{k}} \left( s_{j}^{k-1} - \sum_{p \neq i_{1}} h_{j\,p}^{k} y_{p}^{k-1} \right)$$

$$h_{j\,i_{2}}^{k+1} = \frac{\alpha^{k}}{\sigma^{k}} \left( s_{j}^{k-1} - \sum_{p \neq i_{2}} h_{j\,p}^{k} y_{p}^{k-1} \right) - \frac{\gamma^{k}}{\sigma^{k}} \left( s_{j}^{k} - \sum_{p \neq i_{2}} h_{j\,p}^{k} y_{p}^{k} \right),$$
(2.14)

para j = 1, ..., n.

É interessante observar que as igualdades (2.14) podem ser reescritas com o índice p de todos os somatórios diferentes de  $i_1$  e  $i_2$ .

Como visto anteriormente, a escolha dos índices  $i_1$  e  $i_2$  das colunas a serem modificadas está sujeita à hipótese de

$$\sigma^{k} = det(\bar{A}) = \alpha^{k} \delta^{k} - \beta^{k} \gamma^{k} \neq 0.$$
(2.15)

Note que, caso  $y^k$  seja múltiplo de  $y^{k-1}$ ,  $\sigma^k$  assume valor nulo, impossibilitando escolher as colunas que deverão ser alteradas.

A escolha de  $i_1$  e  $i_2$  adotada, em geral, em nossos testes numéricos é tal que

$$|\alpha^k| = |y_{i_1}^k| = ||y^k||_{\infty} \qquad |\delta^k| = |y_{i_2}^{k-1}| = ||y^{k-1}||_{\infty}.$$

Ocorre uma alteração de índice quando  $\sigma^k$  assume valor (em módulo) menor que uma tolerância  $(tol_{\sigma})$ . Neste caso, alteramos o índice  $i_2$ , de modo que ele passe a ser o índice que fornece o maior valor (em módulo) para  $\sigma^k$ , com  $i_1$ fixado inicialmente, isto é,

$$|(\alpha^{k}y^{k-1} - \gamma^{k}y^{k})_{i_{2}}| = ||\alpha^{k}y^{k-1} - \gamma^{k}y^{k}||_{\infty}.$$

### 2.2 Convergência

Nesta seção omitiremos detalhes de algumas provas. A maioria delas já é conhecida de trabalhos anteriores. No entanto, com o intuito de fazer um trabalho de mais fácil leitura, elas serão apresentadas num apêndice, no final do trabalho.

Denotaremos por ||. || como a norma euclidiana de vetores e sua norma subordinada de matrizes. Quando outra norma for utilizada, esta será indicada.

Suponhamos  $F:\Omega\subset I\!\!R^n\to I\!\!R^n,\;F\in C^1(\Omega),\;\Omega$ um conjunto convexo,  $x^*\in \;\Omega,\;F(x^*)=0$ e

$$||J(x) - J(x^*)|| \le L ||x - x^*||^p, \quad L, \ p > 0$$
(2.16)

para todo  $x \in \Omega$ ; a equação (2.16) implica que para todo  $u, v \in \Omega$ 

$$||F(v) - F(u) - J(x^*)(v - u)|| \le L ||v - u||\sigma(u, v)^p,$$
(2.17)

onde  $\sigma(u,v) = \max\{\|u-x^*\|, \|v-x^*\|\}$  (ver apêndice -  $\mathbf{P1}).$ 

Suponhamos que  $J(x^*)$  seja não singular e seja  $M = ||J(x^*)^{-1}||$ . Por (2.17), deduzimos que (ver apêndice - **P2**), para todo  $u, v \in \Omega$ ,

$$||v - u - J(x^*)^{-1}[F(v) - F(u)]|| \le ML||v - u||\sigma(u, v)^p.$$
(2.18)

**Lema 2.1** Existe  $\varepsilon_1 > 0$  tal que  $F(v) \neq F(u)$  sempre que  $v \neq u$ ,  $||v-x^*|| \leq \varepsilon_1$ ,  $||u-x^*|| \leq \varepsilon_1$ .

**Prova:** Ver Lema 3.1 de Martínez e Zambaldi [18] (apêndice - **P7**).

O resultado de convergência local é estabelecido no Teorema 2.1 a seguir.

**Teorema 2.1** Sejam  $\{x^k\}$  e  $\{H_k\}$  seqüências geradas pelo método ITCUM. Suponhamos que  $F(x_k) \neq 0$  e  $|\sigma^k| > tol_{\sigma} > 0$  para todo k = 0, 1, 2, ... e seja  $r \in (0, 1)$ . Então existem  $\varepsilon = \varepsilon(r), \eta = \eta(r)$  tais que, se  $||x^1 - x^*|| \leq \varepsilon$  e
$||H_k - J(x^*)^{-1}|| \leq \eta$ , sempre que  $(k-1) \equiv 0 \pmod{m}$ , as seqüências  $\{x^k\} \in \{H_k\}$  estão bem definidas,  $\{x^k\}$  converge para  $x^* \in \{H_k\}$ 

$$||x^{k+1} - x^*|| \le r||x^k - x^*||$$
(2.19)

para todo k = 1, 2, ....

#### Prova:

Sejam  $c_1 = 2n^2 M^2 L$  e  $c_2 = n^{5/2}$ . Dados  $\varepsilon, \eta > 0$ , definimos  $b_i(\varepsilon, \eta), i = 0, 1, \cdots, m-1$  por

$$b_0(\varepsilon, \eta) = \eta$$
  

$$b_i(\varepsilon, \eta) = R_i \left( c_2 b_{i-1}(\varepsilon, \eta) + c_1 \varepsilon^p \right), \qquad i = 1, \cdots, m-1,$$
(2.20)

onde  $R_i = \sup R_{i, k-1} \in R_{i, k-1} = \frac{2 ||y^{k-1}||_{\infty} ||y^{k-2}||_{\infty}}{|\sigma^{k-1}|}, \quad (k-1) \equiv i \pmod{m}.$ 

Podemos ver que  $1 \leq R_i < \infty$ ,  $i = 1, \dots, m-1$ , (ver apêndice - **P3** e **P4**) e então para quaisquer  $\varepsilon$ ,  $\eta > 0$ ,

$$0 < b_0(\varepsilon, \eta) < b_1(\varepsilon, \eta) < \dots < b_{m-1}(\varepsilon, \eta) \quad \text{e} \quad \lim_{\varepsilon, \eta \to 0} b_i(\varepsilon, \eta) = 0, \quad (2.21)$$

para  $i = 0, 1, \cdots, m - 1.$ 

Por (2.21), podemos escolher  $\varepsilon=\varepsilon_r>0$  e  $\eta=\eta_r>0$ tais que<sup>3</sup>  $\varepsilon\leq\varepsilon_1$  e

$$b_i(\varepsilon, \eta) + L\varepsilon^p < \frac{r}{M_1},$$
 (2.22)

para  $i = 0, 1, \dots, m - 1$ , onde  $M_1 = \max\{||J(x^*)||, 2M\}$ .

Suponhamos que  $||x^1 - x^*|| \leq \varepsilon$  e  $||H_k - J(x^*)^{-1}|| \leq \eta$  sempre que  $(k-1) \equiv 0 \pmod{m}$ . Provaremos por indução em k que se  $(k-1) \equiv q \pmod{m}$  então  $H_k$  é não singular,

$$||H_k - J(x^*)^{-1}|| \leq b_q(\varepsilon, \eta)$$
 (2.23)

$$||H_k|| \leq 2M \qquad e \qquad (2.24)$$

$$|x^{k+1} - x^*|| \leq r ||x^k - x^*||$$
(2.25)

<sup>&</sup>lt;sup>3</sup>O parâmetro  $\varepsilon_1$  é o mesmo utilizado no Lema 3.1 de Martínez e Zambaldi [18], cujo enunciado e demonstração encontram-se no apêndice - **P8**.

para  $q = 0, 1, \dots, m - 1$ .

Para k = 1, por hipótese,

$$||H_1 - J(x^*)^{-1}|| \le \eta = b_0(\varepsilon, \eta),$$
 (2.26)

portanto (ver apêndice - P5), por (2.22) e (2.26),

$$||H_1|| \leq ||J(x^*)^{-1}|| + ||H_1 - J(x^*)^{-1}||$$
  
$$\leq ||J(x^*)^{-1}|| + \eta$$
  
$$\leq ||J(x^*)^{-1}|| + \frac{1}{||J(x^*)||} \leq 2||J(x^*)^{-1}|| = 2M.$$

Logo,

$$\|H_1\| \le 2M. \tag{2.27}$$

De (2.17), (2.27) e por (2.25) com k = 1 (ver apêndice - P6)

$$\begin{aligned} \|x^{2} - x^{*}\| &= \|x^{1} - x^{*} - H_{1}F(x^{1})\| \\ &= \|x^{1} - x^{*} - H_{1}J(x^{*})(x^{1} - x^{*})\| \\ &+ \|H_{1}\left[F(x^{1}) - F(x^{*}) - J(x^{*})(x^{1} - x^{*})\right]\| \\ &\leq \|[I - H_{1}J(x^{*})](x^{1} - x^{*})\| + 2ML\|x^{1} - x^{*}\|^{p+1} \\ &\leq \left(\|J(x^{*})^{-1} - H_{1}\|\|J(x^{*})\| + 2ML\|x^{1} - x^{*}\|^{p}\right)\|x^{1} - x^{*}\|, \end{aligned}$$

Usando a definição de  $M_1$ , as hipóteses  $||x^1 - x^*|| \le \varepsilon$ ,  $||H_1 - J(x^*)^{-1}|| \le \eta$  e (2.22) na expressão acima, temos

$$\begin{aligned} \|x^{2} - x^{*}\| &\leq M_{1} \left( \|J(x^{*})^{-1} - H_{1}\| + L\|x^{1} - x^{*}\|^{p} \right) \|x^{1} - x^{*}\| \\ &\leq M_{1} \left( \eta + L\varepsilon^{p} \right) \|x^{1} - x^{*}\| \\ &= M_{1} \left( b_{0}(\varepsilon, \eta) + L\varepsilon^{p} \right) \|x^{1} - x^{*}\| \leq r \|x^{1} - x^{*}\|. \end{aligned}$$

Logo,  $||x^2 - x^*|| \le r ||x^1 - x^*||$ , e, portanto, o teorema é válido para o caso k = 1.

Consideremos agora k > 1,  $(k - 1) \equiv q \pmod{m}$ . Se q = 0, a prova de (2.25), (2.23) e (2.24) é análoga à prova para k = 1.

26

Suponhamos q > 0. Provaremos primeiro que  $H_k$  está bem definida e que (2.23) se verifica.

Pela hipótese de indução  $H_{k-1}$  é não singular. Sejam  $i_1 e i_2$  os índices das trocas de colunas tais que  $\sigma^{k-1} \neq 0$ .

Cada componente  $j,\,j=1,\,2,\,\cdots$ ,nda colun<br/>a $i_1$ está dada por:

$$h_{j\,i_1}^k = \frac{\delta^{k-1}}{\sigma^{k-1}} \left( s_j^{k-1} - \sum_{p \neq i_1} h_{j\,p}^{k-1} y_p^{k-1} \right) - \frac{\beta^{k-1}}{\sigma^{k-1}} \left( s_j^{k-2} - \sum_{p \neq i_1} h_{j\,p}^{k-1} y_p^{k-2} \right) \quad (2.28)$$

e, assim<br/>, $H_k$ está bem definida. Definamos  $J(x^\ast)=H^\ast=(h_{i\,j}^\ast).$ Adicionando e subtraindo os termos

$$\frac{\delta^{k-1}}{\sigma^{k-1}} \left( \sum_{p \neq i_1} h_{jp}^* y_p^{k-1} \right) \ \mathbf{e} \ \frac{\beta^{k-1}}{\sigma^{k-1}} \left( \sum_{p \neq i_1} h_{jp}^* y_p^{k-2} \right)$$

respectivamente, em (2.28),<br/>temos que (ver apêndice - P7), para todo  $j=1,\,2,\,\cdots,\,n,$ 

$$\begin{aligned} |h_{j\,i_{1}}^{k} - h_{j\,i_{1}}^{*}| &\leq \left|\frac{\delta^{k-1}}{\sigma^{k-1}}\right| \left|s_{j}^{k-1} - \sum_{p=1}^{n} h_{j\,p}^{*}y_{p}^{k-1}\right| + \left|\frac{\beta^{k-1}}{\sigma^{k-1}}\right| \left|s_{j}^{k-2} - \sum_{p=1}^{n} h_{j\,p}^{*}y_{p}^{k-2}\right| + \\ &\left|\frac{\delta^{k-1}}{\sigma^{k-1}}\right| \sum_{p\neq i_{1}} |h_{j\,p}^{*} - h_{j\,p}^{k-1}| |y_{p}^{k-1}| + \left|\frac{\beta^{k-1}}{\sigma^{k-1}}\right| \sum_{p\neq i_{1}} |h_{j\,p}^{*} - h_{j\,p}^{k-1}| |y_{p}^{k-2}|.\end{aligned}$$

Usando (2.18) e as desigualdades  $|y_p^{k-1}| \leq ||y^{k-1}||_{\infty} e |y_p^{k-2}| \leq ||y^{k-2}||_{\infty}$ , obtemos

$$\begin{aligned} \left|h_{j\,i_{1}}^{k}-h_{j\,i_{1}}^{*}\right| &\leq \left|\frac{\delta^{k-1}}{\sigma^{k-1}}\right| \|s^{k-1}-J(x^{*})^{-1}y^{k-1}\| + \left|\frac{\beta^{k-1}}{\sigma^{k-1}}\right| \|s^{k-2}-J(x^{*})^{-1}y^{k-2}\| \\ &+ \left|\frac{\delta^{k-1}}{\sigma^{k-1}}\right| \|y^{k-1}\|_{\infty} \sum_{p=1}^{n} \left|h_{j\,p}^{*}-h_{j\,p}^{k-1}\right| + \left|\frac{\beta^{k-1}}{\sigma^{k-1}}\right| \|y^{k-2}\|_{\infty} \sum_{p=1}^{n} \left|h_{j\,p}^{*}-h_{j\,p}^{k-1}\right| \\ &\leq \left|\frac{\delta^{k-1}}{\sigma^{k-1}}\right| ML \|s^{k-1}\|\varepsilon^{p} + \left|\frac{\beta^{k-1}}{\sigma^{k-1}}\right| ML \|s^{k-2}\|\varepsilon^{p} + R_{q} \sum_{p=1}^{n} \left|h_{j\,p}^{*}-h_{j\,p}^{k-1}\right| \\ &\leq \left|\frac{\delta^{k-1}}{\sigma^{k-1}}\right| 2M^{2}L \|y^{k-1}\|\varepsilon^{p} + \left|\frac{\beta^{k-1}}{\sigma^{k-1}}\right| 2M^{2}L \|y^{k-2}\|\varepsilon^{p} \\ &+ R_{q} n \|H_{k-1} - J(x^{*})^{-1}\|. \end{aligned}$$

$$(2.29)$$

Nas duas últimas desigualdades usamos os seguintes limitantes (ver apêndice -  ${\bf P8},$  desigualdade (A.16))

$$||s^{k-1}|| \le 2M ||y^{k-1}||$$
 e  $||s^{k-2}|| \le 2M ||y^{k-2}||.$ 

Mas

$$|\delta^{k-1}| \leq ||y^{k-2}|| \leq \sqrt{n} ||y^{k-2}||_{\infty} \qquad |\beta^{k-1}| \leq ||y^{k-1}|| \leq \sqrt{n} ||y^{k-1}||_{\infty};$$

então, usando essas desigualdades e a definição de R em (2.29), temos:

$$\begin{aligned} |h_{j\,i_{1}}^{k} - h_{j\,i_{1}}^{*}| &\leq \frac{\|y^{k-2}\|_{\infty}}{|\sigma^{k-1}|} \|y^{k-1}\|_{\infty} \sqrt{n} \ 2M^{2}L \ \varepsilon^{p} + \\ &\qquad \frac{\|y^{k-1}\|_{\infty}}{|\sigma^{k-1}|} \|y^{k-2}\|_{\infty} \sqrt{n} \ 2M^{2}L \ \varepsilon^{p} + R_{q} \ n \|H_{k-1} - J(x^{*})^{-1}\| \\ &= R_{q} \sqrt{n} \ 2M^{2}L \ \varepsilon^{p} + R_{q} \ n \|H_{k-1} - J(x^{*})^{-1}\|. \end{aligned}$$
(2.30)

A prova de que  $\left|h_{j\,i_2}^k - h_{j\,i_1}^*\right| \leq R_q \sqrt{n} \, 2M^2 L \, \varepsilon^p + R_q \, n \, ||H_{k-1} - J(x^*)^{-1}||$  é completamente análoga.

Pela hipótese de indução e, como  $R_q\geq 1$  <br/>e $R_q\sqrt{n}\,2M^2L\,\varepsilon^p>0,$ temos, para todo  $s\neq i_1$  <br/>e $s\neq i_2,$ 

$$\begin{aligned} |h_{js}^{k} - h_{js}^{*}| &= |h_{js}^{k-1} - h_{js}^{*}| \\ &\leq ||H_{k-1} - J(x^{*})^{-1}|| \\ &\leq R_{q} \sqrt{n} \, 2M^{2}L \, \varepsilon^{p} + R_{q} \, n \, ||H_{k-1} - J(x^{*})^{-1}||. \end{aligned}$$
(2.31)

Logo, de (2.30) e (2.31), temos que, para todo  $p=1,\,2,\,\cdots,\,n,$ 

$$|h_{jp}^{k} - h_{jp}^{*}| \leq R_{q} \sqrt{n} 2M^{2}L \varepsilon^{p} + R_{q} n ||H_{k-1} - J(x^{*})^{-1}||;$$

assim,

$$||H_k - J(x^*)^{-1}||_{\infty} \leq R_q n^{3/2} 2M^2 L \varepsilon^p + R_q n^2 ||H_{k-1} - J(x^*)^{-1}||.$$
(2.32)

Portanto, de (2.32), (2.20) e da hipótese de indução

$$|H_{k} - J(x^{*})^{-1}|| \leq \sqrt{n} ||H_{k} - J(x^{*})^{-1}||_{\infty}$$
  

$$\leq R_{q} \left( n^{2} 2 M^{2} L \varepsilon^{p} + n^{5/2} ||H_{k-1} - J(x^{*})^{-1}|| \right)$$
  

$$\leq R_{q} \left( n^{2} 2 M^{2} L \varepsilon^{p} + n^{5/2} b_{q-1}(\varepsilon, \eta) \right)$$
  

$$= R_{q} \left( c_{2} b_{q-1}(\varepsilon, \eta) + c_{1} \varepsilon^{p} \right)$$
  

$$= b_{q}(\varepsilon, \eta). \qquad (2.33)$$

Logo,  $||H_k - J(x^*)^{-1}|| \le b_q(\varepsilon, \eta)$ . Assim, por (2.22),

$$||H_k - J(x^*)^{-1}|| \le \frac{r}{M_1} \le \frac{1}{2M};$$

portanto, pelo Lema de Banach [7],  $H_k$  é não singular e usando a hipótese  $\sigma^k \neq 0$ , nós podemos concluir que para todo k, as seqüências  $\{x_k\} \in \{H_k\}$  estão bem definidos, e ainda

$$\begin{aligned} \|H_k\| &\leq \|J(x^*)^{-1}\| + \|H_k - J(x^*)^{-1}\| \\ &\leq \|J(x^*)^{-1}\| + \frac{r}{M_1} \\ &\leq \|J(x^*)^{-1}\| + \frac{1}{\|J(x^*)\|} \\ &\leq 2 \|J(x^*)^{-1}\| = 2 M. \end{aligned}$$
(2.34)

Logo,  $||H_k|| \leq 2 M$ , e finalmente, por (2.17), (2.22) e (2.34),

$$||x^{k+1} - x^*|| = ||x^k - x^* - H_k F(x^k)||$$
  

$$= ||x^k - x^* - H_k [F(x^k) - F(x^*) - J(x^*)(x^k - x^*)]|$$
  

$$- H_k J(x^*)(x^k - x^*)||$$
  

$$\leq ||[I - H_k J(x^*)] (x^k - x^*)|| + 2 M L ||x^k - x^*||^{p+1}$$
  

$$\leq [||J(x^*)|| ||J(x^*)^{-1} - H_k|| + 2 M L ||x^k - x^*||^p] ||x^k - x^*||$$
  

$$\leq M_1 (b_q(\varepsilon, \eta) + L \varepsilon^p) ||x^k - x^*||$$
  

$$\leq r ||x^k - x^*||. \qquad (2.35)$$

Logo,  $||x^{k+1}-x^*||\,\leq\,r\,||x^k-x^*||,$ o que completa a demonstração.

## 2.3 Implementação Computacional do ITCUM e Testes Numéricos

Nesta seção, descrevemos uma implementação computacional de ITCUM direcionada para problemas de grande porte. Com este propósito, foram utilizados alguns dos problemas testes de Martínez e Zambaldi [18], Lopes e Martínez [12] e Moré, Garbow e Hillstrom [20], os mesmos da **Seção 1.4**.

Desenvolvendo a equação (2.11) de forma retroativa, obtemos:

$$H_{k} = H_{1} + \sum_{p=1}^{k-1} \left( \frac{\delta^{p} v_{1}^{p} - \beta^{p} v_{2}^{p}}{\sigma^{p}} \right) (e_{i_{1}}^{p})^{T} + \sum_{p=1}^{k-1} \left( \frac{\alpha^{p} v_{2}^{p} - \gamma^{p} v_{1}^{p}}{\sigma^{p}} \right) (e_{i_{2}}^{p})^{T}, \quad (2.36)$$

onde

$$v_1^p = s^p - H_p y^p$$
 e  $v_2^p = s^{p-1} - H_p y^{p-1}$ ,

que é equivalente a

$$H_k = H_1 + \sum_{p=1}^{k-1} u_{i_1}^p (e_{i_1}^p)^T + \sum_{p=1}^{k-1} u_{i_2}^p (e_{i_2}^p)^T, \qquad (2.37)$$

onde

$$u_{i_1}^p = rac{\delta^p v_1^p - eta^p v_2^p}{\sigma^p} \quad \mathrm{e} \quad u_{i_2}^p = rac{lpha^p v_2^p - \gamma^p v_1^p}{\sigma^p}.$$

A implementação de ITCUM é baseada na equação (2.37) e, conforme pode ser observado, o cálculo de  $H_k$  implica no armazenamento de dois vetores  $(u_{i_1}^k e u_{i_2}^k)$ e dois índices adicionais  $(i_1 e i_2)$  a cada iteração k. Por esta razão, o número de iterações consecutivas do método é limitado pela disponibilidade de memória da máquina.

Considerando que há espaço suficiente para armazenar m pares de vetores, é então possível efetuar uma iteração "Newton"<sup>4</sup> e m iterações ITCUM consecutivas.

<sup>4</sup>Nos reinícios, não utilizamos a "verdadeira" matriz jacobiana.

Portanto, se  $\rho \equiv 1 \mod (m+1), \theta \in \{2, ..., m+1\}$ , temos

$$H_{\rho+\theta} = H_{\rho} + \sum_{l=0}^{\theta-1} u_{i_1}^{\rho+l} (e_{i_1}^{\rho+l})^T + \sum_{l=0}^{\theta-1} u_{i_2}^{\rho+l} (e_{i_2}^{\rho+l})^T.$$
(2.38)

Então, o parâmetro m determina o número de iterações do tipo ITCUM que poderão ser realizadas entre um reinício e outro. Neste caso, usamos m=30 quando trabalhamos com os problemas de grande porte (para os problemas de pequeno porte não é necessário realizar reinícios).

A matriz de reinício é determinada da mesma forma que na **Seção 1.4**. A implementação de ICUM é baseada na forma descrita em [18] para todos os problemas.

Em ITCUM, por hipótese, temos que  $F(x^k) \neq 0$  e  $H_k$  é não singular; portanto  $s^k \neq 0$  e, como conseqüência,  $x^k \neq x^{k+1}$ . Assim, pelo **Lema 2.1** apresentado na seção anterior, perto de uma solução isolada não é possível que  $y^k = 0$ . Talvez isto ocorra quando  $x^k$  situa-se "longe" de uma solução  $x^*$ , fazendo com que  $\sigma^k$  assuma, de forma definitiva, valor nulo; ao longo da implementação, esta situação será evitada através da verificação da desigualdade

$$||y^{k}|| \le tol ||F(x^{k})||;$$
(2.39)

onde tol é um número positivo e pequeno. Caso (2.39) seja satisfeita, fazemos  $H_{k+1} = H_k$  (nos experimentos desta seção, tomamos tol = 10<sup>-6</sup>).

A escolha dos índices  $i_1 e i_2$  das colunas a serem alteradas em cada iteração k foi feita conforme descrito na **Seção (2.1)**. Inicialmente, foi estabelecido um parâmetro para a troca de sigma  $(tol_{\sigma})$ ; assim, o índice  $i_2$  terá sua forma de escolha alterada quando

$$|\sigma^k| \le tol_\sigma \tag{2.40}$$

se verifica com a escolha original de  $i_2$ . Para os nossos problemas, optamos por

$$tol_{\sigma} = 10^{-6}$$

Uma outra forma de determinar o parâmetro para a troca do índice  $i_2$  é estabelecer uma tolerância para o seno do ângulo formado entre os vetores  $(\alpha^k, \beta^k)^T$  e  $(\gamma^k, \delta^k)^T$ , que correspondem ás linhas da matriz A. Quando esse seno assumir valor (em módulo) suficientemente próximo de 0, esses vetores podem ser considerados linearmente dependentes. Logo, alteramos a escolha do índice  $i_2$  quando a seguinte desigualdade se verifica com a escolha original de  $i_2$ , onde  $\theta$  é o ângulo formado por esses dois vetores acima mencionados:

$$|\sin(\theta)| \le tol_{\theta};\tag{2.41}$$

para nossos problemas, optamos por

$$tol_{\theta} = 10^{-3};$$

é interessante observar que esse critério para a alteração do índice pode ser reescrito como

$$|\sigma^k| \le tol_{\theta} \| (\alpha^k, \beta^k)^T \|_{\infty} \| (\gamma^k, \delta^k)^T \|_{\infty};$$

$$(2.42)$$

Nas tabelas seguintes, apresentamos os resultados obtidos com as duas formas de tolerância para a singularidade da matriz A. Nessas tabelas utilizaremos  $ITCUM_{\sigma}$  para denotar a implementação de ITCUM com a tolerância para  $\sigma$ , e  $ITCUM_{\theta}$  para denotar a implementação com a tolerância para  $\sin(\theta)$ .

A forma de implementação para os problemas de grande porte acima citados é a mesma descrita em [18] (parâmetros de tolerância, pontos iniciais, etc.) para os demais métodos quase-Newton.

O critério de parada é dado por:

$$||F(x^k)||_{\infty} \le 10^{-5} ||F(x^0)||_{\infty} \tag{2.43}$$

A execução do teste numérico é interrompida quando o número de iterações excede 300 (neste caso, dizemos que o algoritmo não converge; nas tabelas isto é representado pelo termo NC), ou quando

$$||F(x^{k})||_{\infty} \ge 10^{20} ||F(x^{0})||_{\infty};$$
(2.44)

neste último caso, dizemos que o método diverge (representamos este caso pelo termo Div).

Em todos os problemas, o desempenho do ITCUM é avaliado através de uma comparação com os resultados obtidos com o método de Newton e com alguns métodos quase-Newton (o primeiro método de Broyden, Column-Updating Method e Inverse Column-Updating Method), descritos no capítulo anterior.

Nas **Tabelas 2.3** e **2.6** os resultados em cada experimento são representados pelo par (KON; TIME), onde KON é o número de iterações e TIME é o tempo de CPU em segundos. Nas **Tabelas 2.1, 2.2, 2.4** e **2.5**, é apresentado apenas o número de iterações, por se tratar de problemas de pequeno porte ( $n \leq 50$ ).

Prob	Newton	Broyden	CUM	ICUM	$ITCUM_{\sigma}$	$ITCUM_{\theta}$
01	02	12	13	08	05	05
02	41	NC	NC	19	$\mathbf{NC}$	NC
03	11	33	40	83	22	22
04	04	60	67	Div	56	56
05	02	12	13	08	05	05
06	09	08	08	09	08	08
07	02	05	05	05	04	04
08	04	06	05	05	06	06
09	01	04	05	04	04	05

Tabela 2.1: Problemas de pequeno porte.

с	Newton	Broyden	CUM	ICUM	$ITCUM_{\sigma}$	$ITCUM_{\theta}$
0.1	03	03	04	04	03	03
0.5	03	06	06	06	05	05
0.9	05	10	10	09	07	07
0.99	06	12	33	12	11	11
0.999	07	14	39	13	13	13
$1 - 10^{-4}$	08	17	32	15	13	13
$1 - 10^{-5}$	09	24	38	16	15	15
$1 - 10^{-6}$	10	27	43	17	16	16
$1 - 10^{-7}$	10	31	39	17	16	16
$1 - 10^{-8}$	10	28	33	17	16	16
1	10	33	33	17	16	16

Tabela 2.2: Problema 10.

Pr	N	Newton	Broyden	CUM	ICUM	$ITCUM_{\sigma}$	$ITCUM_{ heta}$
$A_0$	32	(03; 01.04)	(150; 13.29)	(117; 10.16)	(86; 07.79)	(80; 08.29)	(84; 08.90)
	50	(03; 04.05)	(172; 59.98)	(238; 78.27)	(155; 53.17)	(142; 52.46)	(147; 53.45)
$A_2$	32	(06; 01.92)	(100; 09.39)	(162; 14.39)	(70; 06.48)	(73; 07.64)	(73; 07.64)
	50	(05; 07.36)	(141; 50.36)	(138; 47.24)	(104;  35.87)	(110; 41.25)	(110; 41.25)
$A_4$	32	(10; 03.13)	(75; 07.30)	(83; 07.42)	(77; 07.08)	(76; 08.13)	(76; 08.13)
	50	(10; 14.39)	(188; 67.12)	(114; 38.28)	(103; 35.37)	(94; 37.57)	(94; 37.57)
B	32	(02; 00.71)	(68; 05.99)	(95; 07.96)	(62; 05.54)	(54; 05.36)	(55; 05.22)
	50	(02; 03.08)	(155; 59.65)	(176; 55.42)	(92; 29.55)	(105; 38.77)	(167; 57.01)
C	32	(01; 00.39)	(62; 04.72)	(82; 05.82)	(61; 05.00)	(71; 06.77)	(82; 07.19)
	50	(01; 01.59)	(132; 43.17)	(141; 41.36)	(115; 34.38)	(112; 38.78)	(139; 45.10)

Tabela 2.3: Problemas de grande porte.

Como é natural em problemas não lineares, é praticamente impossível encontrar um melhor método para resolver todos os problemas. Este fato se reflete no comportamento dos métodos nas tabelas acima. Em vários casos o desempenho de ITCUM é superior ao de ICUM, o que era esperado pois a correção da matriz  $H_k$ a cada iteração usando ITCUM possui posto maior do que quando usando ICUM.

Para determinar as novas colunas, enquanto para ICUM há apenas a manipulação de uma equação, o ITCUM precisa resolver um sistema linear, onde existe a possibilidade do determinante da matriz do sistema ser nulo.

Quando a dimensão dos problemas aumenta, o desempenho de ITCUM tornase, em média, inferior ao de ICUM (**Tabela 2.3**); este fato decorre da escolha dos índices, que torna-se mais complicada devido ao tamanho dos vetores.

Além dos métodos acima citados, os teste numéricos foram também realizados com variações de ITCUM, geradas por diferentes formas de escolha para os índices  $i_1 \, e \, i_2$  de troca de coluna e por critérios variados para a alteração desses mesmos índices (quando necessário) ao longo da implementação; dentre as versões utilizadas, o critério de escolha descrito acima mostrou ser o de melhor desempenho em nossos problemas.

As próximas tabelas ilustram o comportamento de algumas destas versões, comparando-as com os resultados obtidos pelo ITCUM escolhido.

As versões aqui apresentadas são as seguintes:

**ITCUM2:** Define os índices  $i_1 e i_2$  tais que

$$|y_{i_1}^k| = \|y^k\|_\infty$$
 e  $|y_{i_2}^{k-1}| = \|y^{k-1}\|_\infty$ 

Se  $\sigma^k = 0$ , faz uma nova escolha para o segundo índice, de tal sorte que  $i_2$  será o índice da maior coordenada (em módulo) de  $y^{k-1}$  que gera  $\sigma^k \neq 0$ .

**ITCUM3:** Define os índices  $i_1 e i_2$  como sendo os argumentos de

$$\max_{j, p} |y_j^k y_p^{k-1} - y_p^k y_j^{k-1}|$$

ITCUM4: Define os índices  $i_1 \in i_2$  tais que

$$y_{i_1}^k| = ||y^k||_{\infty}$$
  $i_2 = \arg(\max_p |y_p^{k-1}y^k - y_p^k y^{k-1}|)$ 

**ITCUM5**: Define os índices  $i_1 \in i_2$  tais que

$$|y_{i_1}^k| = ||y^k||_{\infty} \qquad |y_{i_2}^{k-1}| = ||y^{k-1}||_{\infty}$$

Se  $\sigma^k = 0$ , é realizada uma iteração ICUM.

Prob	$ITCUM_{\sigma}$	$ITCUM_{\theta}$	ITCUM2	ITCUM3	ITCUM4	ITCUM5
01	05	05	05	05	05	30
02	NC	NC	Div	NC	Div	Div
03	22	22	22	22	22	44
04	56	59	21	57	57	Div
05	05	05	05	05	05	28
06	08	08	08	08	08	08
07	04	04	04	04	04	04
08	06	06	06	06	06	06
09	04	05	04	04	04	04

Tabela 2.4: Problemas de pequeno porte - Versões de ITCUM.

с	$ITCUM_{\sigma}$	$ITCUM_{\theta}$	ITCUM2	ITCUM3	ITCUM4	ITCUM5
0.1	03	03	03	03	03	03
0.5	05	05	05	05	05	05
0.9	07	07	09	07	07	07
0.99	11	11	12	12	16	25
0.999	13	13	19	19	16	Div
$1 - 10^{-4}$	13	13	22	27	16	25
$1 - 10^{-5}$	15	15	23	22	18	31
$1 - 10^{-6}$	16	16	23	21	20	42
$1 - 10^{-7}$	16	16	23	21	20	41
$1 - 10^{-8}$	16	16	23	21	20	41
1	16	16	23	21	20	41

Tabela 2.5: Problema 10 - Versões de ITCUM.

Pr	Ň	$ITCUM_{\sigma}$	$ITCUM_{\theta}$	ITCUM2	ITCUM3	ITCUM4	ITCUM5
$A_0$	32	(80; 08.29)	(84; 08.90)	Div	(100; 11.10)	(79; 08.18)	(87; 08.41)
	50	(142; 52.46)	(147; 53.45)	(215; 81.45)	(144; 57.12)	(106; 40.64)	(159; 58.55)
$A_2$	32	(73; 07.64)	(86; 08.89)	(95; 10.55)	(71; 08.40)	(69; 07.46)	(77; 07.90)
	50	(110; 41.25)	(119; 43.39)	Div	(130; 53.50)	(99; 37.79)	(94; 36.53)
$A_4$	32	(76; 08.13)	(86; 09.01)	Div	(76; 08.63)	(78; 08.13)	(77; 07.69)
	50	(94; 37.57)	(114; 42.62)	Div	(108; 44.22)	(91; 34.06)	(175; 62.72)
B	32	(54; 05.36)	(55; 05.22)	(87; 08.96)	(59; 06.30)	(62; 06.83)	(62; 06.39)
	50	(105; 38.77)	(167; 57.01)	(172; 60.86)	(89; 34.27)	(94; 34.76)	(150; 56.08)
	32	(71; 06.77)	(82; 07.19)	(79; 07.63)	(61; 06.20)	(54; 05.06)	(74; 07.03)
-	50	(112; 38.78)	(139; 45.10)	(213; 71.51)	(110; 40.04)	(87; 29.22)	(154; 54.82)

 Tabela 2.6: Problemas de grande porte - Versões de ITCUM.

Como pode ser observado nas **Tabelas 2.4, 2.5** e **2.6**, a versão de ITCUM escolhida é a que apresenta, em média, o melhor desempenho.

# Capítulo 3

# Método de Atualização de $q \ge 3$ Colunas da Jacobiana Inversa

#### 3.1 Descrição do Método

O método de atualização de q colunas da matriz de aproximação da matriz jacobiana inversa, método multi-coluna inverso, para o problema (1.1) consiste na iteração

$$x^{k+1} = x^k - H_k F(x^k), (3.1)$$

onde a matriz  $H_k$  é atualizada de tal forma que  $H_{k+1}$  difira da matriz anterior em q colunas, e satisfaça as q últimas equações secantes:

$$\begin{array}{rcl}
H_{k+1}y_{k} &=& s_{k} \\
H_{k+1}y_{k-1} &=& s_{k-1} \\
& & \vdots \\
H_{k+1}y_{k-q+1} &=& s_{k-q+1}.
\end{array}$$
(3.2)

A matriz  $H_{k+1}$  pode ser obtida através de uma correção de posto q de  $H_k$ , isto é:

$$H_{k+1} = H_k + u_{i_1}^k \mathbf{e}_{i_1}^T + u_{i_2}^k \mathbf{e}_{i_2}^T + \dots + u_{i_q}^k \mathbf{e}_{i_q}^T,$$
(3.3)

onde os vetores  $\mathbf{e}_{i_1}, \cdots, \mathbf{e}_{i_q}$  pertencem à base canônica do  $\mathbb{R}^n$ . Assim, os vetores  $u_{i_r}^k, r = 1, \cdots, q$ , devem ser tais que as últimas q equações secantes, (3.3), sejam satisfeitas; esses vetores são responsáveis pelas correções a serem realizadas nas colunas  $i_1, i_2, \cdots, i_q$  de  $H_k$ .

Com o propósito de fazer uma análise do método multi-coluna inverso que nos permita encontrar expressões gerais para definir os vetores  $u_{i_r}^k$ ,  $r = 1, \dots, q$ , consideremos as equações (3.2) com  $H_{k+1}$  definida por (3.3):

$$(H_{k} + \sum_{r=1}^{q} u_{i_{r}}^{k} \mathbf{e}_{i_{r}}^{T}) y^{k} = s^{k},$$

$$(H_{k} + \sum_{r=1}^{q} u_{i_{r}}^{k} \mathbf{e}_{i_{r}}^{T}) y^{k-1} = s^{k-1},$$

$$\vdots$$

$$(H_{k} + \sum_{r=1}^{q} u_{i_{r}}^{k} \mathbf{e}_{i_{r}}^{T}) y^{k-q+1} = s^{k-q+1},$$

$$(3.4)$$

equivalentemente,

$$\begin{pmatrix}
\left(\sum_{r=1}^{q} u_{i_{r}}^{k} \mathbf{e}_{i_{r}}^{T}\right) y^{k} = s^{k} - H_{k} y^{k}, \\
\left(\sum_{r=1}^{q} u_{i_{r}}^{k} \mathbf{e}_{i_{r}}^{T}\right) y^{k-1} = s^{k-1} - H_{k} y^{k-1}, \\
\vdots \\
\left(\sum_{r=1}^{q} u_{i_{r}}^{k} \mathbf{e}_{i_{r}}^{T}\right) y^{k-q+1} = s^{k-q+1} - H_{k} y^{k-q+1}.
\end{cases}$$
(3.5)

Cada uma das equações de (3.5) representa uma combinação linear dos vetores  $u_{i_1}, u_{i_2}, \dots, u_{i_q}$  com escalares sendo algumas q componentes dos vetores  $y^k, y^{k-1}, \dots, y^{k-q+1}$ , respectivamente:

$$\begin{array}{rclcrcrcrcrcrc}
 u_{i_{1}}^{k}(\mathbf{e}_{i_{1}}^{T}y^{k}) &+ \cdots &+ & u_{i_{q}}^{k}(\mathbf{e}_{i_{q}}^{T}y^{k}) &= & s^{k} - H_{k}y^{k}, \\
 u_{i_{1}}^{k}(\mathbf{e}_{i_{1}}^{T}y^{k-1}) &+ &\cdots &+ & u_{i_{q}}^{k}(\mathbf{e}_{i_{q}}^{T}y^{k-1}) &= & s^{k-1} - H_{k}y^{k-1}, \\
 & & \vdots & & \vdots & & \\
 u_{i_{1}}^{k}(\mathbf{e}_{i_{1}}^{T}y^{k-q+1}) &+ &\cdots &+ & u_{i_{q}}^{k}(\mathbf{e}_{i_{q}}^{T}y^{k-q+1}) &= & s^{k-q+1} - H_{k}y^{k-q+1}.
\end{array}$$
(3.6)

Para simplificar a notação, definimos para todo  $s, r = 1, \cdots, q$ 

$$a_{sr}^{k} = \mathbf{e}_{i_{r}}^{T} \ y^{k-s+1}.$$
(3.7)

Usando (3.7), o sistema linear (3.6) pode ser escrito na forma matricial como

$$\underbrace{\begin{pmatrix} a_{11}^{k} \mathbf{I} & | & \cdots & | & a_{1q}^{k} \mathbf{I} \\ \hline - & - & & - & - \\ \hline a_{q1}^{k} \mathbf{I} & | & \cdots & | & a_{qq}^{k} \mathbf{I} \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} u_{i_{1}}^{k} \\ \vdots \\ \hline - & \vdots \\ u_{i_{q}}^{k} \end{pmatrix}}_{u^{k}} = \underbrace{\begin{pmatrix} s^{k} - H_{k} y^{k} \\ \hline - & - & - & - \\ \vdots \\ \hline s^{k-q+1} - H_{k} y^{k-q+1} \end{pmatrix}}_{v^{k}} (3.8)$$

Portanto, a existência e unicidade dos vetores  $u_{i_1}^k, u_{i_2}^k, \dots, u_{i_q}^k$  satisfazendo (3.3) está determinada pela não singularidade da matriz  $A \in I\!\!R^{qn \times qn}$ . Isto implica que devemos fazer uma análise das características da matriz por blocos A, o que nos conduzirá a caracterizar, quando possível, os vetores  $u_{i_r}^k$ , com  $r = 1, \dots, q$ , que é o objetivo do capítulo.

### **3.2** A não singularidade da matriz $A \in \mathbb{R}^{qn \times qn}$

Seja  $A \in \mathbb{R}^{qn \times qn}$  a matriz de coeficientes do sistema (3.8). O objetivo desta seção é caracterizar o determinante desta matriz, o qual chamaremos de  $\sigma^k$ , e supondo que dito determinante é não nulo, encontrar a forma geral da matriz inversa de A.

No caso do determinante faremos uso de um resultado para matrizes por blocos dado em [4], o qual incluímos no que segue para maior clareza na leitura do texto:

Considere as matrizes  $E \in \mathbb{R}^{n \times n}$ ,  $H \in \mathbb{R}^{m \times m}$ ,  $F \in \mathbb{R}^{m \times n} e G \in \mathbb{R}^{n \times m}$ . Se E é não singular, então é válida a igualdade:

$$\det \begin{bmatrix} E & F \\ G & H \end{bmatrix} = \det(E) \det(H - GE^{-1}F).$$
(3.9)

Suponha que as matrizes

$$\left(\begin{array}{cc} a_{11}^k \mathbf{I}\end{array}\right) \quad \mathrm{e} \quad \left(\begin{array}{cc} a_{11}^k \mathbf{I} & \mid & a_{12}^k \mathbf{I} \\ -\frac{a_{11}^k}{a_{21}^k} \mathbf{I} & \mid & a_{22}^k \mathbf{I} \end{array}\right)$$

são não singulares. Provaremos por indução em q que a seguinte igualdade se satisfaz:

$$\left(\sigma^{k}\right)^{n} = det \underbrace{\begin{pmatrix} a_{11}^{k} \mathbf{I} & | & \cdots & | & a_{1q}^{k} \mathbf{I} \\ \vdots & \vdots & \vdots & \vdots \\ \hline a_{q1}^{k} \mathbf{I} & | & \cdots & | & \overline{a_{qq}^{k}} \mathbf{I} \end{pmatrix}}_{A} = \begin{bmatrix} det \begin{pmatrix} a_{11}^{k} & \cdots & a_{1q}^{k} \\ \vdots & & \\ a_{q1}^{k} & \cdots & a_{qq}^{k} \end{pmatrix}}_{\bar{A}} \end{bmatrix}^{n} (3.10)$$

i) Verifiquemos que a igualdade (3.10), com  $a_{11}^k \neq 0$ , se satisfaz para q = 2. De (3.9) e usando as propriedades dos determinantes temos:

$$det \left( \begin{array}{ccc} a_{11}^{k} \mathbf{I} & | & a_{12}^{k} \mathbf{I} \\ -\frac{----}{a_{21}^{k} \mathbf{I}} & | & a_{22}^{k} \mathbf{I} \end{array} \right) = det \left( a_{11}^{k} \mathbf{I} \right) det \left( a_{22}^{k} \mathbf{I} - a_{21}^{k} \mathbf{I} (a_{11}^{k} \mathbf{I})^{-1} a_{12}^{k} \mathbf{I} \right)$$
$$= \left( a_{11}^{k} \right)^{n} det \left[ \left( a_{22}^{k} - \frac{a_{12}^{k} a_{21}^{k}}{a_{11}^{k}} \right) \mathbf{I} \right]$$
$$= \left( a_{11}^{k} \right)^{n} \left( a_{22}^{k} - \frac{a_{12}^{k} a_{21}^{k}}{a_{11}^{k}} \right)^{n}$$
$$= \left( a_{11}^{k} a_{22}^{k} - a_{12}^{k} a_{21}^{k} \right)^{n}$$
$$= \left[ det \left( \begin{array}{c} a_{11}^{k} & a_{12}^{k} \\ a_{21} & a_{22}^{k} \end{array} \right) \right]^{n}. \tag{3.11}$$

ii) Suponhamos que a igualdade (3.10) é verdadeira para q = r - 1, isto é,

$$det \begin{pmatrix} a_{11}^{k} \mathbf{I} & \cdots & | & a_{1r-1}^{k} \mathbf{I} \\ \vdots & \vdots & \vdots & \vdots \\ a_{r-11}^{k} \mathbf{I} & | & \cdots & | & a_{r-1r-1}^{k} \mathbf{I} \end{pmatrix} = \begin{bmatrix} det \begin{pmatrix} a_{11}^{k} & \cdots & a_{1r-1}^{k} \\ \vdots & & \\ a_{r-11}^{k} & \cdots & a_{r-1r-1}^{k} \end{pmatrix} \end{bmatrix}^{n} (3.12)$$

Provemos que (3.10) se verifica para q = r. Seja

$$A = \begin{pmatrix} a_{11}^{k} \mathbf{I} & | & \cdots & | & a_{1r}^{k} \mathbf{I} \\ \hline \vdots & \vdots & \vdots & \vdots \\ \hline a_{r1}^{k} \mathbf{I} & | & \cdots & | & \overline{a_{rr}^{k}} \mathbf{I} \end{pmatrix} = \begin{pmatrix} a_{11}^{k} \mathbf{I} & | & -\mathbf{B} \\ \hline \mathbf{C} & | & -\mathbf{D} \end{pmatrix}, \quad (3.13)$$

onde

$$\mathbf{B} = \left(a_{12}^{k}\mathbf{I} \mid \cdots \mid a_{1r}^{k}\mathbf{I}\right) \in \mathbb{R}^{n \times (r-1)n}$$

$$\mathbf{C} = \begin{pmatrix} a_{2_1}^k \mathbf{I} \\ \hline \\ \vdots \\ \hline \\ a_{r_1}^k \mathbf{I} \end{pmatrix} \in \mathbb{R}^{(r-1)n \times n} \quad \mathbf{D} = \begin{pmatrix} a_{2_2}^k \mathbf{I} & | & \cdots & | & a_{2_r}^k \mathbf{I} \\ \hline \\ \hline \\ a_{r_2}^k \mathbf{I} & | & \cdots & | & a_{r_r}^k \mathbf{I} \end{pmatrix} \in \mathbb{R}^{(r-1)n \times (r-1)n}.$$

Utilizando (3.12) e (3.13), obtemos:

$$det(A) = det \left( \begin{array}{ccc} a_{11}^{k} \mathbf{I} & | & \mathbf{B} \\ -\frac{a_{11}}{\mathbf{C}} & | & -\frac{\mathbf{D}}{\mathbf{D}} \end{array} \right)$$
  
=  $det(a_{11}^{k} \mathbf{I}) det(\mathbf{D} - \mathbf{C} (a_{11}^{k} \mathbf{I})^{-1} \mathbf{B})$   
=  $(a_{11}^{k})^{n} det(\mathbf{D} - (a_{11}^{k})^{-1} \mathbf{C} \mathbf{B}).$  (3.14)

A matriz por blocos  $\mathbf{D} - (a_{11}^k)^{-1} \mathbf{CB}$  é de ordem (r-1)n, com o primeiro bloco não singular. Logo, podemos aplicar a hipótese de indução; para isto, definimos as matrizes

$$\bar{\mathbf{B}} = \begin{pmatrix} a_{1\,2}^k & \cdots & a_{1\,r}^k \end{pmatrix} \in I\!\!R^{1 \times (r-1)}$$

$$\bar{\mathbf{C}} = \begin{pmatrix} a_{21}^k \\ \vdots \\ a_{r1}^k \end{pmatrix} \in \mathbb{R}^{(r-1)\times 1} \quad \bar{\mathbf{D}} = \begin{pmatrix} a_{22}^k & \cdots & a_{2r}^k \\ \vdots & & \vdots \\ a_{r2}^k & \cdots & a_{rr}^k \end{pmatrix} \in \mathbb{R}^{(r-1)\times(r-1)}.$$

Assim, de (3.14), temos

$$det(A) = (a_{11}^{k})^{n} \left[det(\bar{\mathbf{D}} - \bar{\mathbf{C}} (a_{11}^{k})^{-1} \bar{\mathbf{B}})\right]^{n}$$
  
=  $\left[a_{11}^{k} det(\bar{\mathbf{D}} - \bar{\mathbf{C}} (a_{11}^{k})^{-1} \bar{\mathbf{B}})\right]^{n}$   
=  $\left[det(a_{11}^{k}) det(\bar{\mathbf{D}} - \bar{\mathbf{C}} (a_{11}^{k})^{-1} \bar{\mathbf{B}})\right]^{n}$   
=  $\left[det\left(-\frac{a_{11}^{k}}{\bar{\mathbf{C}}} + \frac{\bar{\mathbf{B}}}{\bar{\mathbf{D}}} - \right)\right]^{n} = \left[det(\bar{A})\right]^{n}$ .

Portanto, (3.10) está provado.

Logo, de (3.10) concluímos que para garantir a não singularidade da matriz por blocos A de ordem qn, basta garantir a não singularidade da matriz  $\overline{A}$  de ordem q. Suponhamos que a matriz  $A \in I\!\!R^{qn \times qn}$  é não singular. Nosso interesse agora é encontrar a forma geral para a matriz inversa de A. Apresentaremos em detalhe os casos q = 2 e q = 3, para os quais usamos o processo de eliminação gaussiana. Este processo permite vislumbrar a expressão geral de  $A^{-1}$ . A notação  $l_p$  denota a linha p de blocos da manipulação de matriz ampliada  $[A \mid I]$  em questão. A cada passo da eliminação, operamos toda uma linha de blocos, ao invés de apenas uma linha da matriz. Assim, para q = 2, temos:

$$\left(\begin{array}{ccccccccc} a_{11}^{k}\mathbf{I} & | & a_{12}\mathbf{I} & || & \mathbf{I} & | & \mathbf{O} \\ \hline --- & --- & || & --- & --- \\ a_{21}^{k}\mathbf{I} & | & a_{22}^{k}\mathbf{I} & || & \mathbf{O} & | & \mathbf{I} \end{array}\right);$$

Suponhamos que  $a_{11}^k \neq 0$  (caso contrário, fazemos uma permutação e teremos que  $a_{21}^k \neq 0$ , já que  $\sigma^k \neq 0$ ).

Ao final do primeiro passo da eliminação gaussiana, o novo  $l_2$ será

-

$$l_{2} \rightarrow -\frac{a_{21}^{k}}{a_{11}^{k}} l_{1} + l_{2} \qquad \begin{pmatrix} a_{11}^{k} \mathbf{I} & | & a_{12} \mathbf{I} & || & \mathbf{I} & | & \mathbf{O} \\ --- & -- & || & --- & || & --- \\ \mathbf{O} & | & \frac{\sigma^{k}}{a_{11}^{k}} \mathbf{I} & || & -\frac{a_{21}^{k}}{a_{11}^{k}} \mathbf{I} & || & \mathbf{I} \end{pmatrix};$$

ao final do segundo passo da eliminação gaussiana, o novo  $l_1$ será

$$l_{1} \rightarrow -\frac{a_{11}^{k} a_{12}^{k}}{\sigma^{k}} l_{2} + l_{1} \qquad \begin{pmatrix} a_{11}^{k} \mathbf{I} & | & \mathbf{O} & || & \frac{a_{11}^{k} a_{2,2}^{k}}{\sigma^{k}} \mathbf{I} & | & -\frac{a_{11}^{k} a_{12}^{k}}{\sigma^{k}} \mathbf{I} \\ ---- & || & ---- & || & -\frac{a_{11}^{k}}{\sigma^{k}} \mathbf{I} & || & -\frac{a_{11}^{k} a_{12}^{k}}{\sigma^{k}} \mathbf{I} & || & -\frac{a_{11}^{k} a_{12}^{k}}{\sigma^{k}} \mathbf{I} & || & 1 \end{pmatrix};$$

ao final do terceiro passo da eliminação gaussiana, os novos  $l_1$  e  $l_2$ serão

$$\begin{split} l_1 &\to \frac{1}{a_{11}^k} \ l_1 &= \left( \begin{array}{cccc} \mathbf{I} &\mid \mathbf{O} & \mid\mid & \frac{a_{22}^k}{\sigma^k} \mathbf{I} \mid -\frac{a_{12}^k}{\sigma^k} \mathbf{I} \\ -- &-- &\mid\mid & --- & -- \\ \mathbf{O} \mid &\mathbf{I} \mid\mid & -\frac{a_{21}^k}{\sigma^k} \mathbf{I} \mid & \frac{a_{11}^k}{\sigma^k} \mathbf{I} \end{array} \right). \end{split}$$

Finalmente, a inversa da matriz A é dada por

$$A^{-1} = \begin{pmatrix} \frac{a_{22}^k}{\sigma^k} \mathbf{I} & | & -\frac{a_{12}^k}{\sigma^k} \mathbf{I} \\ -\frac{a_{21}^k}{\sigma^k} \mathbf{I} & | & \frac{a_{11}^k}{\sigma^k} \mathbf{I} \end{pmatrix} = \frac{1}{\sigma^k} \begin{pmatrix} a_{22}^k \mathbf{I} & | & -a_{12}^k \mathbf{I} \\ -\frac{a_{21}^k}{\sigma^k} \mathbf{I} & | & a_{11}^k \mathbf{I} \end{pmatrix}.$$

Os coeficientes dos blocos de  $A^{-1}$  correspondem, respectivamente, às componentes da inversa da matriz de ordem 2,

$$\bar{A} = \begin{pmatrix} a_{11}^k & a_{12}^k \\ a_{21}^k & a_{22}^k \end{pmatrix}.$$

Assim, para determinar a matriz  $A^{-1} \in \mathbb{R}^{2n \times 2n}$ , basta calcular a matriz inversa de  $\bar{A} \in \mathbb{R}^{2 \times 2}$ .

Para o caso em que q = 3, também utilizando a eliminação gaussiana, obtemos a inversa da matriz A:

$$A^{-1} = \frac{1}{\sigma^{k}} \begin{pmatrix} (a_{22}^{k} a_{33}^{k} - a_{23}^{k} a_{32}^{k})\mathbf{I} \mid (a_{13}^{k} a_{32}^{k} - a_{12}^{k} a_{33}^{k})\mathbf{I} \mid (a_{12}^{k} a_{23}^{k} - a_{13}^{k} a_{22}^{k})\mathbf{I} \\ (a_{23}^{k} a_{31}^{k} - a_{21}^{k} a_{33}^{k})\mathbf{I} \mid (a_{11}^{k} a_{33}^{k} - a_{13}^{k} a_{31}^{k})\mathbf{I} \mid (a_{13}^{k} a_{21}^{k} - a_{11}^{k} a_{23}^{k})\mathbf{I} \\ (a_{31}^{k} a_{22}^{k} - a_{21}^{k} a_{32}^{k})\mathbf{I} \mid (a_{13}^{k} a_{21}^{k} - a_{11}^{k} a_{32}^{k})\mathbf{I} \mid (a_{11}^{k} a_{22}^{k} - a_{12}^{k} a_{21}^{k})\mathbf{I} \end{pmatrix}$$
(3.15)

de forma equivalente,

$$A^{-1} = \frac{1}{\sigma^{k}} \begin{pmatrix} c_{11}^{k} \mathbf{I} \mid c_{21}^{k} \mathbf{I} \mid c_{31}^{k} \mathbf{I} \\ \frac{--}{c_{12}^{k} \mathbf{I}} \mid c_{22}^{k} \mathbf{I} \mid c_{32}^{k} \mathbf{I} \\ \frac{--}{c_{13}^{k} \mathbf{I}} \mid c_{23}^{k} \mathbf{I} \mid c_{33}^{k} \mathbf{I} \end{pmatrix} \in \mathbb{R}^{3n \times 3n}$$

onde para  $r, s = 1, 2, 3, c_{rs}$  denota o cofator rs da matriz  $\overline{A}$ .

De forma análoga para  $A \in I\!\!R^{qn \times qn}$ , a matriz inversa de A é dada por:

$$A^{-1} = \frac{1}{\sigma^{k}} \begin{pmatrix} \frac{c_{11}^{k}\mathbf{I}}{-} & \frac{c_{21}^{k}\mathbf{I}}{-} & \frac{c_{21}^{k}\mathbf{I}}{-} & \frac{c_{q1}^{k}\mathbf{I}}{-} \\ \frac{c_{12}^{k}\mathbf{I}}{-} & \frac{c_{22}^{k}\mathbf{I}}{-} & \frac{c_{q2}^{k}\mathbf{I}}{-} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{c_{1q}^{k}\mathbf{I}}{-} & \frac{c_{2q}^{k}\mathbf{I}}{-} & \frac{c_{q1}^{k}\mathbf{I}}{-} \\ \end{pmatrix} \in \mathbb{R}^{qn \times qn} \quad (3.16)$$

onde para  $r, s = 1, 2, \cdots, q, c_{rs}$  representa o cofator rs de  $\overline{A} \in \mathbb{R}^{q \times q}$ .

Novamente, os coeficientes dos blocos de  $A^{-1}$  correspondem aos cofatores da matriz  $\overline{A}$  divididos por  $\sigma^k$ . Esta análise nos permite concluir que, para calcular a inversa de A de ordem qn, deve-se calcular a inversa de  $\overline{A}$ , que é uma matriz de ordem q; as componentes desta matriz serão os correspondentes múltiplos da identidade.

Antes de finalizar esta seção, vale a pena ressaltar que nosso interesse no cálculo da matriz inversa de  $A \in I\!\!R^{qn \times qn}$  é puramente teórico. No entanto, a relação

existente entre A e  $\overline{A}$  será explorada sob o ponto de vista prático, e nos permitirá concluir sobre as vantagens e desvantagens do método multi-coluna inverso à medida que q aumenta.

É interessante observar que para problemas de grande porte (n grande), o número de operações e custo do algoritmo está fortemente ligado ao valor de q, isto é, ao número de colunas a serem trocadas.

### **3.3** Resolvendo o sistema $Au^k = v^k$

Nesta seção, supondo que  $\sigma^k = det(\bar{A}) \neq 0$ , resolveremos o sistema linear (3.16). Isto nos conduzirá a encontrar uma expressão geral para o vetor  $u^k$ , isto é, encontrar a forma geral dos vetores  $u_{i_r}^k$ ,  $r = 1, \dots, q$ .

Para encontrar uma expressão geral para o vetor  $u_{i_r}^k$  de (3.8), deve-se resolver um sistema linear, o que pode ser feito teoricamente usando a inversa de A.

Suponhamos que  $\sigma^k = det(\bar{A}) \neq 0$ . Logo,

$$u^k = A^{-1}v^k.$$

De (3.16), temos:

$$u_{i_{1}}^{k} = \frac{1}{\sigma_{k}} \left( c_{1 \ 1}^{k} v_{1}^{k} + c_{2 \ 1}^{k} v_{2}^{k} + \dots + c_{q \ 1}^{k} v_{q}^{k} \right)$$
  

$$u_{i_{2}}^{k} = \frac{1}{\sigma_{k}} \left( c_{1 \ 2}^{k} v_{1}^{k} + c_{2 \ 2}^{k} v_{2}^{k} + \dots + c_{q \ 2}^{k} v_{q}^{k} \right)$$
  

$$\vdots \qquad (3.17)$$

$$u_{i_{q}}^{k} = \frac{1}{\sigma_{k}} \left( c_{1\,q}^{k} v_{1}^{k} + c_{2\,q}^{k} v_{2}^{k} + \dots + c_{q\,q}^{k} v_{q}^{k} \right),$$

de forma equivalente

$$u_{i_{r}}^{k} = \frac{1}{\sigma_{k}} \sum_{p=1}^{q} c_{p\tau}^{k} v_{p}^{k}, \quad r = 1, 2, \cdots, q,$$
(3.18)

isto é, para cada  $r, p = 1, 2, \dots, q$ , os vetores  $u_{i_r}^k$  são combinações lineares dos vetores  $v_p^k$  onde os escalares da dita combinação são, respectivamente, as componentes da coluna p da matriz adjunta de  $\overline{A}$  vezes  $\det(\overline{A})^{-1}$ .

Substituindo o resultado obtido em (3.18) na igualdade (3.3), obtemos a fórmula para atualização da matriz  $H_k$ , dada por

$$H_{k+1} = H_k + \frac{1}{\sigma_k} \sum_{r=1}^q \sum_{p=1}^q c_{pr}^k v_p^k \mathbf{e}_{i_r}^T.$$
(3.19)

### 3.4 Testes Numéricos

Nesta seção, descrevemos os primeiros resultados obtidos com a implementação do método multi-coluna inverso, para q = 3. Com o intuito de fazer uma comparação com os resultados obtidos na seção 2.3, utilizaremos os problemas teste  $A_0$  a C dessa seção.

Tomando q = 3, a equação (3.19) pode ser reescrita como:

$$H_{k+1} = H_k + \frac{1}{\sigma_k} \sum_{r=1}^3 \sum_{p=1}^3 c_{pr}^k v_p^k \mathbf{e}_{i_r}^T, \qquad (3.20)$$

onde, substituindo os cofatores obtidos em (3.15), e o determinante obtido em (3.10),

$$\begin{aligned} c_{11}^{k} &= a_{22}^{k} a_{33}^{k} - a_{23}^{k} a_{32}^{k} & c_{12}^{k} &= a_{23}^{k} a_{31}^{k} - a_{21}^{k} a_{33}^{k} & c_{13}^{k} &= a_{21}^{k} a_{32}^{k} - a_{22}^{k} a_{31}^{k} \\ c_{21}^{k} &= a_{13}^{k} a_{32}^{k} - a_{12}^{k} a_{33}^{k} & c_{22}^{k} &= a_{11}^{k} a_{33}^{k} - a_{13}^{k} a_{31}^{k} & c_{23}^{k} &= a_{12}^{k} a_{31}^{k} - a_{11}^{k} a_{32}^{k} \\ c_{31}^{k} &= a_{12}^{k} a_{23}^{k} - a_{13}^{k} a_{22}^{k} & c_{32}^{k} &= a_{13}^{k} a_{21}^{k} - a_{11}^{k} a_{23}^{k} & c_{33}^{k} &= a_{11}^{k} a_{22}^{k} - a_{12}^{k} a_{21}^{k} \\ \sigma^{k} &= a_{11}^{k} a_{22}^{k} a_{23}^{k} + a_{13}^{k} a_{23}^{k} a_{32}^{k} + a_{13}^{k} a_{21}^{k} a_{32}^{k} - a_{11}^{k} a_{23}^{k} a_{31}^{k} - a_{11}^{k} a_{23}^{k} a_{32}^{k} - a_{12}^{k} a_{21}^{k} \\ \sigma^{k} &= a_{11}^{k} a_{22}^{k} a_{23}^{k} + a_{12}^{k} a_{23}^{k} a_{32}^{k} + a_{13}^{k} a_{21}^{k} a_{32}^{k} - a_{11}^{k} a_{23}^{k} a_{31}^{k} - a_{11}^{k} a_{23}^{k} a_{32}^{k} - a_{12}^{k} a_{21}^{k} a_{33}^{k} \\ \sigma^{k} &= a_{11}^{k} a_{22}^{k} a_{32}^{k} + a_{12}^{k} a_{23}^{k} a_{32}^{k} + a_{13}^{k} a_{21}^{k} a_{32}^{k} - a_{11}^{k} a_{23}^{k} a_{32}^{k} - a_{11}^{k} a_{23}^{k} a_{32}^{k} - a_{12}^{k} a_{21}^{k} a_{33}^{k} \\ \sigma^{k} &= a_{11}^{k} a_{22}^{k} a_{23}^{k} + a_{12}^{k} a_{23}^{k} a_{32}^{k} + a_{13}^{k} a_{22}^{k} a_{32}^{k} - a_{13}^{k} a_{22}^{k} a_{31}^{k} - a_{11}^{k} a_{23}^{k} a_{32}^{k} - a_{12}^{k} a_{21}^{k} a_{33}^{k} \\ \sigma^{k} &= a_{11}^{k} a_{22}^{k} a_{23}^{k} + a_{12}^{k} a_{23}^{k} a_{32}^{k} + a_{13}^{k} a_{22}^{k} a_{33}^{k} - a_{11}^{k} a_{23}^{k} a_{32}^{k} - a_{12}^{k} a_{21}^{k} a_{33}^{k} \\ \sigma^{k} &= a_{11}^{k} a_{22}^{k} a_{23}^{k} + a_{12}^{k} a_{23}^{k} a_{32}^{k} - a_{11}^{k} a_{23}^{k} a_{32}^{k} - a_{11}^{k} a_{23}^{k} a_{32}^{k} - a_{12}^{k} a_{21}^{k} a_{33}^{k} \\ \sigma^{k} &= a_{12}^{k} a_{23}^{k} a_{23}^{k} + a_{13}^{k} a_{23}^{k} a_{33}^{k} - a_{11}^{k} a_{23}^{k} a_{33}^{k} - a_{11}^{k} a_{23}^{k} a_{33}^{k} - a_{12}^{k} a_{23}^{k} a_{33}^{k} - a_{12}^{k} a_{33}^{k} - a_{12}^{k} a_{33}^{k} - a_{12}^{k} a_{33}^{$$

$$= u_{11}u_{22}u_{33} + u_{12}u_{23}u_{31} + u_{13}u_{21}u_{32} - u_{13}u_{22}u_{31} - u_{11}u_{23}u_{32} - u_{12}u_{21}u_{33}$$

$$v_1^k = s^k - H_k y^k$$
  $v_2^k = s^{k-1} - H_{k-1} y^{k-1}$   $v_3^k = s^{k-2} - H_{k-2} y^{k-2}$ .

Desenvolvendo a equação (3.20) de forma retroativa, obtemos:

$$H_{k+1} = H_2 + \sum_{j=2}^{k} \left[ \frac{1}{\sigma^j} \sum_{r=1}^{3} \sum_{p=1}^{3} c_{pr}^j v_p^j \mathbf{e}_{i_r}^j^T \right], \qquad (3.21)$$

e por (3.18), é equivalente a

$$H_{k+1} = H_2 + \sum_{j=2}^{k} \left[ \sum_{r=1}^{3} u_{i_r}^j \mathbf{e}_{i_r}^{j^T} \right].$$
(3.22)

A implementação computacional é baseada na equação (3.22) e conforme pode ser observado, o cálculo de  $H_k$  implica no armazenamento de três vetores  $(u_{i_1}^k, u_{i_2}^k e u_{i_3}^k)$  e três índices adicionais  $(i_1, i_2 e i_3)$  a cada iteração. Por esta razão, assim como em ITCUM, o número de iterações consecutivas do método é limitado pela disponibilidade de memória da máquina.

Novamente, considerando que é possível armazenar 3m vetores, o parâmetro m determina o número de iterações consecutivas do método multi-coluna inverso com q = 3 que poderão ser realizadas entre um reinício e outro.

Em nossos testes, realizamos dois tipos de escolha dos índices  $i_1$ ,  $i_2 e i_3$  das colunas a serem alteradas:

• Primeira escolha:

$$|y_{i_1}^k| = \|y^k\|_{\infty} \qquad |y_{i_2}^{k-1}| = \|y^{k-1}\|_{\infty} \qquad |y_{i_3}^{k-2}| = \|y^{k-2}\|_{\infty} \quad (3.23)$$

caso  $\sigma^k$  assuma valor nulo com essa escolha, ou caso a matriz  $\bar{A}$  torne-se mal-condicionada, alteramos o índice  $i_3$ , de modo que

$$\left| \left( (a_{21}^k a_{32}^k - a_{22}^k a_{31}^k) y^k + (a_{12}^k a_{31}^k - a_{11}^k a_{32}^k) y^{k-1} + (a_{11}^k a_{22}^k - a_{12}^k a_{21}^k) y^{k-2} \right)_{i_3} \right| = \\ \left\| (a_{21}^k a_{32}^k - a_{22}^k a_{31}^k) y^k + (a_{12}^k a_{31}^k - a_{11}^k a_{32}^k) y^{k-1} + (a_{11}^k a_{22}^k - a_{12}^k a_{21}^k) y^{k-2} \right\|_{\infty}.$$

• Segunda escolha

$$|y_{i_1}^k| = ||y^k||_{\infty} \qquad |y_{i_2}^{k-1}| = ||y^{k-1}||_{\infty} \qquad |y_{i_3}^{k-2}| = ||y^{k-2}||_{\infty} \quad (3.24)$$

caso  $\sigma^k$  assuma valor nulo com essa escolha, ou caso a matriz  $\overline{A}$  torne-se mal-condicionada, alteramos os índices  $i_2 \in i_3$ , de modo que eles passam a ser os respectivos argumentos de:

 $\max_{p,\,s} \left| a_{11}^k \left( y_p^{k-1} y_s^{k-2} - y_s^{k-1} y_p^{k-2} \right) + a_{21}^k \left( y_s^k y_p^{k-2} - y_p^k y_s^{k-2} \right) + a_{31}^k \left( y_p^k y_s^{k-1} - y_s^k y_p^{k-1} \right) \right|$ 

Foram impostos dois tipos de tolerância para que ocorra a troca de escolha dos índices  $i_2$  e/ou  $i_3$ . O primeiro estabelece uma tolerância para det $(A) = \sigma^k$ , denotada por  $tol_{\sigma}$ ; assim, alteramos a escolha dos índices caso  $\sigma^k$  tenha valor próximo de zero:

$$|\sigma^k| \le tol_{\sigma}.$$

O segundo tipo de tolerância  $(tol_r)$  implementado refere-se ao número de condição da matriz A. Definimos rcond(A) como sendo

$$\operatorname{rcond}(A) = \frac{1}{\operatorname{cond}(A)}.$$

Logo, caso rcond(A) esteja próximo de zero, concluímos que a matriz é malcondiconada; assim, alteramos a escolha dos índices caso a seguinte desigualdade seja verificada:

$$rcond(A) \leq tol_r;$$

em todos os nossos problemas, optamos por<sup>1</sup>

$$tol_r = 10^{-16}.$$

Todos os demais parâmetros, assim como critérios de convergência e divergência, são os mesmos descritos na seção 2.3

<sup>&</sup>lt;sup>1</sup>Lembrando que  $A \in \mathbb{R}^{3 \times 3}$ , o fato de possuir um número de condição maior que  $10^{16}$  nos permite declará-la mal-condicionada.

Nas tabelas seguintes, apresentamos os resultados obtidos com as duas formas de tolerância descritas acima. Nessas tabelas, denotaremos por ICUM3 o método multi-coluna inverso com q = 3; analogamente ao que foi feito na implementação de ITCUM, usaremos a notação  $ICUM3_{\sigma}$  para representar a implementação de ICUM3 com a tolerância para  $\sigma$ , e  $ICUM3_r$  para denotar a implementação com a tolerância para rcond(A).

Em todos os problemas, o desempenho de ICUM3 é avaliado através de uma comparação com os resultados obtidos com ICUM e ITCUM, numa tentativa de observar o comportamento do método multi-coluna conforme q aumenta.

A tabela seguinte apresenta o comportamento de ICUM3 quando utilizamos a primeira escolha para os índices, descrita em (3.23).

Pr	N	ICUM	$ITCUM_{\sigma}$	$ITCUM_{\theta}$	$ICUM3_{\sigma}$	$ICUM3_{\tau}$
$A_0$	32	(86; 07.79)	(80; 08.29)	(84; 08.90)	$\sigma^{60}=0$	$\sigma^{37} = 0$
	50	(155; 53.17)	(142; 52.46)	(147; 53.45)	$\sigma^{42}=0$	$\sigma^{30}=0$
$A_2$	32	(70; 06.48)	(73; 07.68)	(86; 08.89)	$\sigma^{21}=0$	$\sigma^{23}=0$
	50	(104; 35.87)	(110; 41.25)	(119; 43.39)	$\sigma^{49}=0$	$\sigma^{32} = 0$
$A_4$	32	(77; 07.08)	(76; 08.13)	(86; 09.01)	$\sigma^5 = 0$	$\sigma^5 = 0$
	50	(103; 35.37)	(94; 37.57)	(114; 42.62)	$\sigma^4 = 0$	$\sigma^4 = 0$
В	32	(62; 05.54)	(54; 05.36)	(55; 05.22)	(53;06.20)	$\sigma^{39} = 0$
	50	(92; 29.55)	(105; 38.77)	(167; 57.01)	$\sigma^{37} = 0$	$\sigma^{45} = 0$
C	32	(61; 05.00)	(71; 06.77)	(82; 07.19)	$\sigma^{59} = 0$	$\sigma^{36} = 0$
	50	(115; 34.38)	(112; 38.78)	(139; 45.10)	$\sigma^{45}=0$	$\sigma^{43} = 0$

Tabela 3.1: Método multi-coluna, q = 1, 2, 3 - Escolha 1.

A tabela seguinte apresenta o comportamento de ICUM3 quando utilizamos a segunda escolha para os índices, descrita em (3.24).

Pr	N	ICUM	$ITCUM_{\sigma}$	$ITCUM_{\theta}$	$ICUM3_{\sigma}$	ICUM3 <sub>r</sub>
$A_0$	32	(86; 07.79)	(80; 08.29)	(84; 08.90)	(68; 31.64)	(64; 14.72)
	50	(115; 53.17)	(142; 52.46)	(147; 53.45)	(161; 370.91)	(220; 260.95)
$A_2$	32	(70; 06.48)	(73; 07.64)	(86; 08.89)	(87; 37.73)	(97; 30.48)
	50	(104; 35.87)	(110; 41.25)	(119; 43.39)	(120; 272.92)	$\sigma^{57} = 0$
$A_4$	32	(77; 07.08)	(76; 08.13)	(86; 09.01)	(57; 30.48)	(61; 21.58)
	50	(103; 35.37)	(94; 37.57)	(114; 42.62)	(94; 224.54)	(104; 152.97)
B	32	(62; 05.54)	(54; 05.36)	(55; 05.22)	(55; 24.83)	$\sigma^{39} = 0$
	50	(92; 29.55)	(105; 38.77)	(167; 57.01)	(101; 210.04)	$\sigma^{64} = 0$
C	32	(61; 05.00)	(71; 06.77)	(82; 07.19)	(62; 29.27)	(64; 12.25)
	50	(115; 34.38)	(112; 38.78)	(139; 45.10)	(82; 158.29)	(94; 99.98)

**Tabela 3.2:** Método multi-coluna, q = 1, 2, 3 - Escolha 2.

Como pode ser observado, a primeira escolha não é muito eficaz, no sentido de que a troca de escolha de índices que se realiza quando a tolerância para  $\sigma^k$  ou rcond(A) é satisfeita, não é suficiente para evitar que o determinante da matriz deixe de ser próximo de zero. Tal escolha obteve convergência em apenas um teste; neste, o tempo necessário para atingir o critério de convergência é menor que o obtido com a segunda escolha.

A segunda escolha foi implementada numa tentativa de evitar a singularidade da matriz A de forma mais eficiente. Porém, em todas as iterações em que é necessário alterar a escolha dos índices  $i_2 e i_3$ , é preciso realizar uma busca em todos os possíveis índices "candidatos" a  $i_2 e i_3$ ; isso faz com que esta escolha demore um tempo maior por iteração.

Um fato que pode ser observado nas **Tabelas 3.1** e **3.2** é que, na maioria dos problemas, *ICUM*3 consegue convergir com um menor número de iterações quando comparado à ITCUM e a ICUM; porém, ele necessita de um tempo maior para realizar cada iteração. Assim, embora convirja em poucas iterações, o tempo necessário para essa convergência é maior.

# Capítulo 4

# Traçamento de raios em sísmica

Existem muitos problemas práticos em diversas áreas (Economia, Física, Engenharia, entre outros) que são modelados de maneira muito conveniente por sistemas não lineares [22]. Uma boa opção para a resolução desses problemas é utilizar os métodos quase-Newton.

Motivados por este fato, e numa tentativa de realizar uma comparação dos resultados obtidos pelo método de Newton e pelos métodos quase-Newton presentes nesta dissertação (Broyden, CUM, ICUM, ITCUM e ICUM3), neste capítulo descrevemos e resolvemos um problema que ocorre frequentemente na área da Geofísica. O problema é encontrar a trajetória de um raio refletido/transmitido em um meio heterogêneo, gerado por uma fonte S e captado por um receptor G.

### 4.1 Descrição do problema

Inicialmente, vamos assumir que a subsuperfície da Terra pode ser representada em duas dimensões, e que a sua estrutura pode ser dividida (modelada) por regiões, as quais possuem velocidade de propagação da onda constante.

As interfaces entre as regiões consecutivas são definidas em função da coordenada horizontal x; elas se localizam onde ocorre um contraste de velocidade, e são representadas por

$$z = f_i(x), \quad i = 1, \cdots, m,$$

onde m corresponde ao número de interfaces, z denota a coordenada vertical ("apontando para baixo") e as funções  $f_i$  são contínuas e, possivelmente, suaves. A região localizada entre duas interfaces sucessivas é assumida como homogênea (isto é, a velocidade de propagação é constante) e as interfaces não se interceptam.

Em pontos distintos da superfície da Terra, posicionam-se uma fonte de ondas sonoras e um receptor (geofone); a fonte emite uma onda sonora que atravessará um número determinado de regiões subterrâneas, e ao retornar à superfície, será captada pelo geofone. A frente de onda de uma onda elementar pode ser descrita pela superfície t = T(x, z), onde T(x, z) é conhecida como a função de tempo de trânsito. Ela satisfaz a chamada equação iconal, uma equação diferencial parcial de primeira ordem de fundamental importância, que leva diretamente para a concepção da Teoria de Raios [1]. Os raios são trajetórias ortogonais ao movimento das frentes de onda e podem ser obtidos como pontos estacionários do funcional de Fermat, o qual representa o tempo de trânsito entre dois pontos dados.

Denotaremos por  $(a_i)$ ,  $i = 1, \dots, m-1$ , o número de vezes que o raio cruza a região situada entre as interfaces  $f_i \in f_{i+1}$ , fazendo um movimento de "zigzag" (para baixo e para cima). O vetor  $(a_1, \dots, a_{m-1})$  é chamado de assinatura do raio. A assinatura não é suficiente para determinar de forma única a trajetória do raio; assumiremos que a trajetória do raio é tal que haja  $2a_i - 1$ movimentos em cada região *i* antes de passar para a região i+1 e, somente após essas reflexões, retornar diretamente ao geofone (veja **Figura 4.1**).



Figura 4.1: (a) - Raio com assinatura (2, 1); (b) - Raio com assinatura (2, 2)

Quando a Terra é modelada por uma seqüência de camadas ou blocos homogêneos, a Teoria dos Raios requer apenas que duas condições simples sejam satisfeitas. A primeira é que os raios devem ter uma trajetória em linha reta em cada camada. A outra é que eles devem obedecer a lei de Snell nas interfaces.

A primeira das duas condições é uma conseqüência necessária da lei de Snell que pode ser expressada da seguinte forma

$$\frac{\operatorname{sen}\alpha_I}{v_I} = \frac{\operatorname{sen}\alpha_T}{v_T} = \frac{\operatorname{sen}\alpha_R}{v_R},\tag{4.1}$$

onde  $v_I \in v_R$  são as velocidades no lado incidente  $(v_I = v_R) \in v_T$  é a velocidade no lado transmitido (veja **Figura 4.2**).



Figura 4.2: Lei de Snell

Como pode ser observado na **Figura 4.2**, os ângulos  $\alpha_I \in \theta_I$ ,  $\alpha_T \in \theta_T$ ,  $\alpha_R \in \theta_R$  são complementares, o que nos permite reescrever a lei de Snell (4.1) da seguinte forma

$$\frac{\cos \theta_I}{v_I} = \frac{\cos \theta_T}{v_T} = \frac{\cos \theta_R}{v_R}.$$
(4.2)

Logo, dados os pontos inicial **S** (fonte) e final **G** (geofone) na superfície da Terra, a velocidade da j-ésima região  $(v_j)$ ,  $j = 1, \dots, m$  e a assinatura  $(a_j)$ ,  $j = 1, \dots, m-1$ , nosso problema é encontrar os pontos  $\mathbf{X}_{\mathbf{k}} = (x_k, f_{i_k}(x_k))$ , para  $k = 1, 2, \dots, n$ , que satisfazem a lei de Snell em cada refletor  $i_k$ . A dimensão n é obtida por

$$n = 2\sum_{j=1}^{m-1} a_j + 1.$$

Para simplificar, utilizamos  $X_0 = S \in X_{n+1} = G$ . O raio entre cada dois pontos consecutivos pode ser descrito pelo segmento de reta entre  $X_{k-1} \in X_k$ .

Definindo  $\tau_k = (1, f'_{i_k}(x_k))$  como o vetor tangente à k-ésima interface no ponto  $\mathbf{X}_k$  e lembrando da fórmula para o cosseno entre vetores, podemos reescrever a lei de Snell (4.2) como

$$\frac{1}{v_{i_k}} \frac{\tau_k^T (\mathbf{X}_k - \mathbf{X}_{k-1})}{\|\tau_k\| \|\mathbf{X}_k - \mathbf{X}_{k-1}\|} = \frac{1}{v_{i_{k+1}}} \frac{\tau_k^T (\mathbf{X}_{k+1} - \mathbf{X}_k)}{\|\tau_k\| \|\mathbf{X}_{k+1} - \mathbf{X}_k\|},$$
(4.3)

onde  $\|\cdot\|$  denota a norma euclidiana. Assim, utilizando a equação (4.3) para  $k = 1, \dots, n$ , obtemos um sistema linear de n equações e n incógnitas. Definindo as funções  $\phi_k : \mathbb{R}^n \to \mathbb{R}, \ k = 1, \dots, n$  por

$$\phi_{k}(\mathbf{x}) = v_{i_{k+1}} \frac{(x_{k} - x_{k-1}) + f'_{i_{k}}(x_{k}) \left(f_{i_{k}}(x_{k}) - f_{i_{k-1}}(x_{k-1})\right)}{\left[\left(x_{k} - x_{k-1}\right)^{2} + \left(f_{i_{k}}(x_{k}) - f_{i_{k-1}}(x_{k-1})\right)^{2}\right]^{1/2}} - v_{i_{k}} \frac{(x_{k+1} - x_{k}) + f'_{i_{k}}(x_{k}) \left(f_{i_{k+1}}(x_{k+1}) - f_{i_{k}}(x_{k})\right)}{\left[\left(x_{k+1} - x_{k}\right)^{2} + \left(f_{i_{k}}(x_{k+1}) - f_{i_{k}}(x_{k})\right)^{2}\right]^{1/2}},$$

$$(4.4)$$

resolver o problema de traçamento de raios é equivalente a resolver o sistema de equações não lineares

$$\Phi(\mathbf{x}) = 0, \tag{4.5}$$

onde  $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ ,  $\Phi(\mathbf{x}) = (\phi_1(\mathbf{x}), \phi_2(\mathbf{x}), \cdots, \phi_n(\mathbf{x}))^T$ .

#### 4.2 Testes numéricos

O primeiro teste foi realizado no problema de traçamento de raios com a particular estrutura retratada na Figura 4.3, cujas interfaces são dadas por

 $z = f_1(x) = 2 - \exp(-10x^2)$  e  $z = f_2(x) = 3.5 - \exp(-10x^2)$ ,

onde x representa a distância (em km), e as velocidades das regiões são

$$v_1 = 2 \ km/s$$
 e  $v_2 = 2.5 \ km/s$ .

A fonte e o geofone estão posicionados na superfície de forma simétrica, a  $1.5 \ km$  da origem.



Figura 4.3: Traçamento de raios - estrutura 1.

O raio possui a assinatura a = (4); portanto, para determinarmos a trajetória do raio precisamos resolver o sistema  $\Phi(\mathbf{x}) = 0$ ,  $\Phi : \mathbb{R}^9 \to \mathbb{R}^9$ , definido pela equação (4.5).

Utilizamos o ponto inicial

 $\mathbf{x}^{0} = (-1.35; -0.8; -0.4; -0.1; 0; 0.1; 0.4; 0.8; 1.35)^{T};$ 

este problema é muito sensível ao ponto inicial. É necessário que este ponto esteja suficientemente próximo da solução para que os métodos convirjam.

A implementação dos métodos quase-Newton foi realizada conforme descrito nos capítulos anteriores. O método multi-coluna inverso com q = 3 foi implementado com o segundo tipo de escolha para os índices, pois esta mostrou-se mais eficiente no sentido de evitar que  $\sigma^k$  assuma valor nulo.

O critério de convergência é dado por

$$\|\mathbf{\Phi}(\mathbf{x}^*)\|_{\infty} \le 10^{-5},$$

A execução do algoritmo é interrompida quando o número de iterações excede 300 ou quando  $\|\mathbf{\Phi}(\mathbf{x})\|_{\infty} \geq 10^{20}$ ; quando este último caso ocorre, dizemos que o método diverge.

Na **Tabela 4.1** apresentamos o número de iterações obtidas pelo método de Newton e por cada método quase-Newton (não mencionamos o tempo devido à dimensão do problema ser muito pequena).

O vetor solução  $\mathbf{x}^*$  encontrado em todos os métodos é exatamente o mesmo:

 $\mathbf{x}^* = (-1.1669; -0.8523; -0.5464; -0.0375; 0; 0.0378; 0.5500; 0.8906; 1.2305)^T$ .

Método	Iterações
Newton	06
Broyden	11
CUM	13
ICUM	12
$ITCUM_{\sigma}$	12
$ITCUM_{\theta}$	12
$ICUM3_{\sigma}$	09
$ICUM3_{\theta}$	09
	i

Tabela 4.1: Traçamento de raios em sísmica - n = 9.

Na Figura 4.4, os segmentos de reta em vermelho representam a trajetória do raio - partindo do ponto S e chegando ao ponto G - determinados pela solução  $x^*$ .



Figura 4.4: Traçamento de raios em sísmica; n = 9.

O segundo teste foi realizado com a estrutura representada na Figura 4.5. As interfaces são dadas por

$$egin{aligned} z &= f_1(x) = 0.75 + 0.125 \, x, \ z &= f_2(x) = 2 - 0.5 \, \exp\left(-2 \, x^2
ight) \, \mathrm{e} \ z &= f_3(x) = 3 - x^2/8, \end{aligned}$$

e as velocidades das regiões são

$$v_1 = 1.5 \ km/s, \quad v_2 = 2 \ km/s \quad e \quad v_3 = 2.5 \ km/s.$$

A fonte e o geofone estão dispostos da mesma forma que na estrutura do exemplo anterior.



Figura 4.5: Traçamento de raios - estrutura 2.

Para esta estrutura, foram utilizadas três tipos de assinatura

 $a^1 = (1, 1),$   $a^2 = (2, 1)$  e  $a^3 = (1, 2),$ 

cujos respectivos pontos iniciais são dados por

 $\begin{aligned} \mathbf{x^0} &= (-1; \ -0.5; \ 0; \ 0.5; \ 1) \,, \\ \mathbf{x^0} &= (-1.3; \ -0.2; \ 0; \ 1.2; \ 2; \ 1.8; \ 1.8) \ e \\ \mathbf{x^0} &= (-1.125; \ -0.75; \ -0.375; \ 0; \ 0.375; \ 0.75; \ 1.125) \end{aligned}$ 

O vetores soluções  $\mathbf{x}^*$  encontrados são, respectivamente,

 $\begin{aligned} \mathbf{x}^* &= (-1.2001; \ -0.3409; \ 0.8963; \ 1.2049; \ 1.5057); \\ \mathbf{x}^* &= (-1.1115; \ -0.0125; \ 0.7244; \ 1.5003; \ 2.0053; \ 1.9341; \ 1.8181); \\ \mathbf{x}^* &= (-1.4612; \ -1.2713; \ -1.1490; \ -0.0215; \ 1.2788; \ 1.4157; \ 1.5930). \end{aligned}$ 

Na **Tabela 4.2** apresentamos o número de iterações obtidas pelo método de Newton e por cada método quase-Newton para cada assinatura utilizada.

Método	$a^1 = (1, 1)$	$a^2 = (2, 1)$	$a^3 = (1, 2)$
Newton	05	04	04
Broyden	07	06	04
CUM	07	06	04
ICUM	07	06	05
$ITCUM_{\sigma}$	06	04	04
$ITCUM_{\theta}$	06	04	04
$ICUM3_{\sigma}$	05	04	04
$ICUM3_{\theta}$	05	04	04
	,	r	

Tabela 4.2: Traçamento de raios - trajetórias elementares

Nas figuras a seguir estão representadas as trajetórias do raio para cada assinatura, obtidas através da solução  $\mathbf{x}^*$ .



Figura 4.6: Traçamento de raios -  $a^1 = (1, 1)$ .



**Figura 4.7:** Traçamento de raios -  $a^2 = (2, 1)$ .



Figura 4.8: Traçamento de raios -  $a^3 = (1, 2)$ .
#### 4.2.1 Problema de grande porte

Para a resolução de um sistema de grande porte, tomamos o problema de tracamento de raios com a as seguintes interfaces

$$z = f_1(x) = 2.5$$
 e  $z = f_2(x) = 4$ ,

isto é, linhas retas para facilitar a obtenção de um bom ponto inicial para a execução dos métodos numéricos (Figura 4.9). As velocidades nas regiões são

$$v_1 = 2.8 \ km/s$$
 e  $v_2 = 1.2 \ km/s;$ 

a fonte e o geofone estão posicionados na superfície da Terra de forma simétrica, a  $2\,km$  da origem.



Figura 4.9: Traçamento de raios - estrutura 3.

Os métodos utilizados necessitam de um bom ponto inicial para obterem convergência. Em [10], há a descrição de técnicas para encontrar um ponto inicial (é utilizado o método da continuação ou homotópico); para configurações geométricas diferentes, podemos empregar outras técnicas para gerar  $\mathbf{x}^0$ .

Devido à forma particular das interfaces utilizadas (linhas retas), as componentes do ponto inicial são aquelas que dividem o intervalo  $[x_S, x_G] \text{ cm } n+1$  partes iguais, ou seja,

$$\mathbf{x_k^0} = x_S + k\left(\frac{x_G - x_S}{n+1}\right), \quad k = 1, 1, \cdots, n.$$

O raio possui a assinatura a = (500); portanto, queremos encontrar um vetor  $\mathbf{x}^* = (x_1^*, x_2^*, \cdots, x_{1001}^*)^T$  tal que  $\Phi(\mathbf{x}^*) = 0$ , onde  $\Phi : \mathbb{R}^{1001} \to \mathbb{R}^{1001}$ .

O critério de convergência e o de divergência são os mesmos adotados na seção anterior.

Na **Tabela 4.3** apresentamos o número de iterações obtidas pelo método de Newton e por cada método quase-Newton, assim como o tempo necessário para obter a solução.

Método	Iterações	Tempo
Newton	04	103.54
Broyden	05	2.08
CUM	05	1.98
ICUM	05	1.95
$ITCUM_{\sigma}$	04	2.05
$ITCUM_{\theta}$	04	2.03
$ICUM3_{\sigma}$	04	2.31
$ICUM3_{\theta}$	04	2.28
	ł	

**Tabela 4.3:** Traçamento de raios em sísmica - n = 1001.

Como pode ser observado, o método de Newton foi aquele que apresentou o menor número de iterações, porém o tempo (medido em segundos) foi o maior de todos os métodos implementados. Os métodos quase-Newton (com exceção do método multi-coluna com q = 3) conseguiram obter convergência de forma rápida (ver [23]); dentre eles, o ITCUM foi o que apresentou o menor número de iterações e o ICUM o menor tempo. O método multi-coluna com q = 3 demorou a convergir devido à dificuldade de escolha dos índices que gerem a matriz  $\overline{A}$  inversível.

Na figura a seguir, os segmentos de reta que representam a trajetória do raio estão traçados em vermelho. Por realizar um movimento de "zig-zag" repetidas vezes entre as interfaces, em um intervalo pequeno do eixo x, é necessário uma ampliação para que o raio possa ser visto. Na **Figura 4.11** apresentamos a

ampliação da região selecionada com um retângulo da Figura 4.10, que permite uma melhor visualização da trajetória do raio.



Figura 4.10: Traçamento de raios em sísmica; = 1001



Figura 4.11: Traçamento de raios em sísmica; = 1001 - Ampliação.

# Capítulo 5

### Conclusão

O ICUM [18] foi considerado recentemente o mais eficiente método quase-Newton para resolver sistemas não lineares de grande porte [13]; essa eficiência pode ser verificada na utilização desse método para resolução de problemas práticos de diversas áreas, como feito em [22] e [23].

Sob o ponto de vista do número de equações secantes satisfeitas por iteração, em posições extremas temos ICUM e o método secante sequencial [15] (que, de fato, nem pode ser implementado razoavelmente para problemas de grande porte) cujas propriedades são bem conhecidas. Mas as propriedades dos métodos intermediários não estavam perfeitamente claras.

Neste trabalho, fizemos o desenvolvimento do método baseado nas q equações secantes anteriores, assim como a determinação do número ótimo dessas equações que devem ser satisfeitas (método multi-coluna inverso).

Inicialmente, desenvolvemos esse método alterando apenas duas colunas por iteração (q = 2); neste caso, provamos que ITCUM obtém convergência local linear. Em nossos testes numéricos, o desempenho médio de ITCUM (em termos de iterações efetuadas) é superior ao de ICUM.

Porém, enquanto em ICUM há apenas a manipulação de uma equação, o IT-CUM necessita de resolução teórica de um sistema linear para poder determinar as colunas a serem alteradas, onde existe a possibilidade do determinante da matriz ser nulo. A busca por índices que tornem essa matriz inversível e o fato de atualizar duas colunas (e não apenas uma como em ICUM), faz com que ITCUM gaste um maior tempo de CPU por iteração que ICUM, resultando em um tempo final de CPU para ITCUM maior do que o de ICUM.

Quando a dimensão dos problemas aumenta, o desempenho de ITCUM torna-se, em média, inferior ao de ICUM (a escolha dos índices torna-se mais complicada devido ao tamanho dos vetores).

O método multi-coluna com q = 3 (ICUM3) não obteve um bom desempenho em termos de convergência quando aplicado em nossos testes numéricos, obtendo o pior desempenho em relação ao tempo de execução dentre os métodos quase-Newton utilizados para comparação. Isto se deve ao fato de que, neste caso, evitar a singularidade da matriz A requer uma busca mais complexa dos índices das colunas a serem alteradas e também um número maior de operações por iteração, o que o torna computacionalmente bem mais caro.

Essas observações nos permitem concluir que o método multi-coluna com  $q \ge 3$  possui desempenho inferior aos demais métodos quase-Newton presentes neste trabalho; assim, concluímos que o número ótimo de colunas no método de atualização multi-coluna inverso para sistemas não lineares é no máximo 2.

Tais conclusões foram confirmadas nos resultados obtidos com a implementação de todos os métodos no problema prático em Geofísica, de traçamento de raios em sísmica, que foi estudado no capítulo 4.

Nossos testes numéricos também revelaram que ICUM, juntamente com IT-CUM, foram os métodos que obtiveram melhor performance para os problemasteste de pequeno porte; esse resultado vai além das conclusões obtidas por Lukšan e Vlček para problemas de grande porte.

# Apêndice A

# Algumas Passagens do Teorema 2.1

**P1** - Desenvolvimento entre as equações (2.16) e (2.17):

Sejam  $u, v \in \Omega$ ; então é válida a identidade:

$$F(v) - F(u) = \int_0^1 \left[ J \left( u + t(v - u) \right) \right] (v - u) dt,$$

e portanto

$$F(v) - F(u) - J(x^*)(v - u) = \int_0^1 \left[ J\left(u + t(v - u)\right) - J(x^*) \right] (v - u) dt.$$

Assim, por (2.16),

$$\begin{aligned} \|F(v) - F(u) - J(x^{*})(v - u)\| &= \left\| \int_{0}^{1} \left[ J\left(u + t(v - u)\right) - J(x^{*}) \right](v - u) dt \right\| \\ &\leq \|v - u\| \int_{0}^{1} \|J\left(u + t(v - u)\right) - J(x^{*})\| dt \\ &\leq \|v - u\| \int_{0}^{1} L \|x - x^{*}\|^{p} dt, \end{aligned}$$
(A.1)

onde x = u + t(v - u).

Como

$$x - x^*| \le \max\{|u - x^*|, |v - x^*|\} = \sigma(u, v),$$

é possível concluir de (A.1) que:

.

$$||F(v) - F(u) - J(x^{*})(v - u)|| \leq ||v - u|| L \int_{0}^{1} ||x - x^{*}||^{p} dt$$
  
$$\leq L \sigma(u, v)^{p} ||v - u||$$
(A.2)

**P2** - Desenvolvimento entre as equações (2.17) e (2.18):

Utilizando propriedade da norma e a desigualdade (2.17), temos:

$$\begin{aligned} \|F(v) - F(u) - J(x^{*})(v - u)\| &= \|J(x^{*})^{-1} [F(v) - F(u) - J(x^{*})(v - u)]\| \\ &\leq \|J(x^{*})^{-1}\| \|F(v) - F(u) - J(x^{*})(v - u)\| \\ &\leq ML \|v - u\| \sigma(u, v)^{p} \end{aligned}$$
(A.3)

**P3** - Propriedades das bolas  $b_i(\varepsilon, \eta), i = 0, 1, \cdots, m$ , dadas em (2.21):

Seja 
$$R_{i, k-1} = 2 \frac{\|y^{k-1}\|_{\infty} \|y^{k-2}\|_{\infty}}{|\sigma^{k-1}|};$$
 então  
 $|\sigma^{k-1}| = |\alpha^{k-1}\delta^{k-1} - \beta^{k-1}\gamma^{k-1}|$   
 $\leq \|y^{k-1}\|_{\infty} \|y^{k-2}\|_{\infty} + \|y^{k-1}\|_{\infty} \|y^{k-2}\|_{\infty}$   
 $= 2 \|y^{k-1}\|_{\infty} \|y^{k-2}\|_{\infty};$  (A.4)

portanto,  $R_i \ge R_{i, k-1} \ge 1$ .

Então, para todo  $i = 1, \dots, m-1$ , temos a seguinte desigualdade:

$$b_{i}(\varepsilon, \eta) = R_{i} c_{2} b_{i-1}(\varepsilon, \eta) + R c_{1} \varepsilon^{p}$$
  
$$= R_{i} (c_{2} b_{i-1}(\varepsilon, \eta) + c_{1} \varepsilon^{p})$$
  
$$\geq c_{2} b_{i-1}(\varepsilon, \eta) + c_{1} \varepsilon^{p}. \qquad (A.5)$$

Como  $c_1 > 0$  e  $c_2 > 1$ , de (A.5) temos:

$$b_i(\varepsilon, \eta) > c_2 b_{i-1}(\varepsilon, \eta) > b_{i-1}(\varepsilon, \eta).$$

P4 - Verificação da limitação  $R_i < \infty$ , para todo  $i = 1, \cdots, m-1$ : Seja  $k \in I\!\!N$ ; então, como  $r \in (0, 1)$ ,

$$\begin{split} \|x^{k}\|_{\infty} &\leq \|x^{k} - x^{*}\|_{\infty} + \|x^{*}\|_{\infty} \\ &\leq r \|x^{k-1} - x^{*}\|_{\infty} + \|x^{*}\|_{\infty} \\ &\leq r^{2} \|x^{k-2} - x^{*}\|_{\infty} + \|x^{*}\|_{\infty} \\ &\vdots \\ &\vdots \\ &\leq r^{k} \|x^{1} - x^{*}\|_{\infty} + \|x^{*}\|_{\infty} \\ &< \|x^{1} - x^{*}\|_{\infty} + \|x^{*}\|_{\infty} \\ &\leq \varepsilon + \|x^{*}\|_{\infty} \end{split}$$

Portanto,  $||x^k||_{\infty} \leq \varepsilon + ||x^*||_{\infty}$ , para todo  $k \in \mathbb{N}$ . Assim, para todo  $k, x^k \in \mathcal{B}[0, \varepsilon + ||x^*||_{\infty}]$ .

Com<br/>o $\mathcal{B}\left[0, \varepsilon + \|x^*\|_{\infty}\right] \subseteq \mathbb{R}^n$ é um conjunto compacto em <br/>  $\mathbb{R}^n$  e  $F: \mathbb{R}^n \to \mathbb{R}^n$ é uma função contínua, então<br/>  $F\left(\mathcal{B}\left[0, \varepsilon + \|x^*\|_{\infty}\right]\right) \subseteq \mathbb{R}^n$ é compacto, e em particular, limitado. Logo,

 $||F(x^k)||_{\infty} \leq W < \infty$ , para todo  $k \in \mathbb{N}$ , W constante.

Então

$$||y^{k}||_{\infty} \le ||F(x^{k})||_{\infty} + ||F(x^{k-1})|_{\infty} \le W + W = 2W < \infty.$$

Como  $|\sigma^{k-1}| \ge tol_{\sigma} > 0$ , temos que

$$R_{i, k-1} = 2 \frac{\|y^{k-1}\|_{\infty} \|y^{k-2}\|_{\infty}}{|\sigma^{k-1}|}$$

$$\leq 2 \frac{\|y^{k-1}\|_{\infty} \|y^{k-2}\|_{\infty}}{tol_{\sigma}}$$

$$\leq \frac{8W^2}{tol_{\sigma}} < \infty.$$

Logo,

$$R_i = \sup R_{i, k-1} < \infty,$$

para todo  $i = 1, \cdots, m - 1$ .

**P5** - Desenvolvimento utilizado para a obtenção da desigualdade (2.27):  
Temos de (2.22) e da definição de 
$$M_1$$
 que

$$b_0 + L \varepsilon^p < \frac{r}{M_1} \le \frac{r}{\|J(x^*)\|} < \frac{1}{\|J(x^*)\|};$$
 (A.6)

então, como  $L \varepsilon^p > 0$ ,

$$b_0 < b_0 + L \varepsilon^p < \frac{1}{\|J(x^*)\|}.$$
 (A.7)

 ${\bf P6}$  - Desenvolvimento utilizado para a obtenção da desigual<br/>dade (2.25), para k=1;

Temos que

$$||x^{2} - x^{1}|| = ||x^{1} - x^{*} - H_{1}F(x^{1})(x^{1} - x^{*})|| + ||H_{1}[F(x^{1}) - F(x^{*}) - J(x^{*})(x^{1} - x^{*})]||$$
  

$$\leq ||(I - H_{1}J(x^{*}))(x^{1} - x^{*})|| + ||H_{1}|| ||F(x^{1}) - F(x^{*}) - J(x^{*})(x^{1} - x^{*})||.$$
(A.8)

Usando (2.17) com  $v = x^1$  e  $u = x^*$ , temos que

$$||F(x^{1}) - F(x^{*}) - J(x^{*})(x^{1} - x^{*})|| \leq L ||x^{1} - x^{*}||\sigma(x^{1}, x^{*})^{p}$$
  
$$\leq L ||x^{1} - x^{*}||^{p+1}, \qquad (A.9)$$

pois

$$\sigma(x^{1}, x^{*}) = \max\{\|x^{1} - x^{*}\|, \|x^{*} - x^{*}\|\} = \|x^{1} - x^{*}\|.$$

Substituindo (A.9) em (A.8) e como  $H_1 \leq 2M$ :

$$||x^{2} - x^{1}|| \leq ||(I - H_{1}J(x^{*}))(x^{1} - x^{*})|| + ||H_{1}|| ||F(x^{1}) - F(x^{*}) - J(x^{*})(x^{1} - x^{*})|| \leq ||(I - H_{1}J(x^{*}))(x^{1} - x^{*})|| + 2ML ||x^{1} - x^{*}||^{p+1}.$$
(A.10)

**P7** - Desenvolvimento entre as equações (2.28) e (2.29):

De (2.28), temos:

$$\begin{split} |h_{j\,i_{1}}^{k} - h_{j\,i_{1}}^{*}| &= \left| \frac{\delta^{k-1}}{\sigma^{k-1}} \left( s_{j}^{k-1} - \sum_{p \neq i_{1}} h_{j\,p}^{*} y_{p}^{k-1} + \sum_{p \neq i_{1}} h_{j\,p}^{*} y_{p}^{k-1} - \sum_{p \neq i_{1}} h_{j\,p}^{k-1} y_{p}^{k-1} \right) \right. \\ &- \left. \frac{\beta^{k-1}}{\sigma^{k-1}} \left( s_{j}^{k-2} - \sum_{p \neq i_{1}} h_{j\,p}^{*} y_{p}^{k-2} + \sum_{p \neq i_{1}} h_{j\,p}^{*} y_{p}^{k-2} - \sum_{p \neq i_{1}} h_{j\,p}^{k-1} y_{p}^{k-2} \right) \right. \\ &- \left. h_{j\,i_{1}}^{*} \left( \frac{y_{i_{1}}^{k-1} y_{i_{2}}^{k-2} - y_{i_{2}}^{k-1} y_{i_{1}}^{k-2}}{\sigma^{k-1}} \right) \right| \\ &= \left| \frac{\delta^{k-1}}{\sigma^{k-1}} \left( s_{j}^{k-1} - \sum_{p \neq i_{1}} h_{j\,p}^{*} y_{p}^{k-1} \right) - \frac{\beta^{k-1}}{\sigma^{k-1}} \left( s_{j}^{k-2} - \sum_{p \neq i_{1}} h_{j\,p}^{*} y_{p}^{k-2} \right) \right. \\ &+ \left. \frac{\delta^{k-1}}{\sigma^{k-1}} \left[ \sum_{p \neq i_{1}} (h_{j\,p}^{*} - h_{j\,p}^{k-1}) y_{p}^{k-1} \right] - \frac{\beta^{k-1}}{\sigma^{k-1}} \left[ \sum_{p \neq i_{1}} (h_{j\,p}^{*} - h_{j\,p}^{k-1}) y_{p}^{k-2} \right] \right| . \end{split}$$

Portanto,

$$\begin{aligned} |h_{j\,i_{1}}^{k} - h_{j\,i_{1}}^{*}| &\leq \left| \frac{\delta^{k-1}}{\sigma^{k-1}} \right| \left| s_{j}^{k-1} - \sum h_{j\,p}^{*} y_{p}^{k-1} \right| + \left| \frac{\beta^{k-1}}{\sigma^{k-1}} \right| \left| s_{j}^{k-2} - \sum h_{j\,p}^{*} y_{p}^{k-2} \right| \\ &\left| \frac{\delta^{k-1}}{\sigma^{k-1}} \right| \sum_{p \neq i_{1}} \left| h_{j\,p}^{*} - h_{j\,p}^{k-1} \right| \left| y_{p}^{k-1} \right| + \left| \frac{\beta^{k-1}}{\sigma^{k-1}} \right| \sum_{p \neq i_{1}} \left| h_{j\,p}^{*} - h_{j\,p}^{k-1} \right| \left| y_{p}^{k-2} \right|. \end{aligned}$$

**P8 - Lema:** Existe  $\varepsilon_1 > 0$  tal que  $F(v) \neq F(u)$  quando  $v \neq u$ ,  $||v - x^*|| \leq \varepsilon_1 e$  $||u - x^*|| \leq \varepsilon_1$ 

Por (2.18) nós temos que, para todo 
$$u, v \in D$$
,  
 $\|v - u\| - \|J(x^*)^{-1} [F(v) - F(u)]\| \le M L \|v - u\| \sigma(u, v)^p$ . (A.11)

Então,

$$||F(v) - F(u)|| \le ||v - u|| \left(\frac{1}{M} - L\sigma(u, v)^p\right).$$
 (A.12)

Seja  $\varepsilon_1 > 0$  tal que

$$\varepsilon_1^p < \frac{1}{2 \, M \, L}.\tag{A.13}$$

Por (A.13), se  $||u - x^*|| \le \varepsilon_1$  e  $||v - x^*|| \le \varepsilon_1$ , temos que

$$L \sigma(u, v)^p \le L \varepsilon^p < \frac{1}{2M}.$$
 (A.14)

Então

$$\frac{1}{M} - L\,\sigma(u,\,v)^p > \frac{1}{M} - \frac{1}{2M} = \frac{1}{2M}.$$
(A.15)

Portanto, por (A.12) e (A.14),

$$||F(v) - F(u)|| \ge \frac{1}{2M} ||v - u||$$
(A.16)

De (A.16) o resultado dasejado é obtido diretamente.

UNICAMP
BIBLIOTECA CENTRAL
SEÇÃO CIRCUMANTE
and the second

## Bibliografia

- [1] Bleistein, N. (1984). *Mathematical methods for wave phenomena*, Academic Press.
- [2] Broyden, C. G.; Dennis, J. E. Jr; Moré, J. J. (1973). On the local and superlinear convergence of quasi-Newton methods, J. Inst. Math. Appl. 12, pp 223-245.
- [3] Cunha, M. C. C. (2000). Métodos Numéricos, Editora da UNICAMP, 2.ed., Campinas, SP.
- [4] Daniel, J. W.; Noble, Benjamin (1977). Applied linear algebra, Prentice-Hall, 2. ed., New Jersey.
- [5] Dennis, J. E. Jr; Moré, J. J.(1997). Quase-Newton methods, motivation and theory, SIAM Review 19, pp 46-89.
- [6] Dennis, J. E. Jr; Schnabel, R. B.(1983). Numerical methods for unconstrained optimization and nonlinear equations, Prentice Hall, Englewood Cliffs, N.J.
- [7] Golub, G. H.; Van Loan, Ch. F.(1995). Matrix Computations, The Johns Hopkins University Press, 3nd. edition, Baltimore and London.
- [8] Gomes-Ruggiero, M. A. (1990). Método quase-Newton para resolução de sistemas não lineares esparsos e de grande porte, Tese de Doutorado, FEE-Unicamp, Campinas, Brasil.
- [9] Gomes-Ruggiero, M. A.; Martínez, J. M.; Moretti, A. C.(1992). Comparing algorithms for solving sparse nonlinear systems of equations, SIAM J. Sci. Stat. Comput. 13, pp 459-483.

- [10] Keller, H. B.; Perozzi, D. J. (1983). Fast seismic ray tracing, SIAM Journal of Applied Mathematics 43, N.4, pp 981-992.
- [11] Kelley, C. T. (1995). Iterative Methods for Linear and Nonlinear Equations, Frontiers in Applieda Mathematics vol. 16, SIAM, Philadelphia.
- [12] Lopes, V. L. R.; Martínez, J. M.(1995). Convergence properties of the inverse column-updating method, *Optimization Methods and Software 6*, pp 127-144.
- [13] Lukšan,L.; Vlček, J.(1998). Computational experience with globally convergent descent methods for large sparse systems of nonlinear equations, *Optimization Methods and Software 8*, pp 185-199.
- [14] Marchand, P. (1999). Graphics and GUIs with MATLAB, CRC Press LCC, 2nd. edition, London.
- [15] Martínez, J. M.(1979). Three new algorithms based on the sequential secant method, BIT 19, pp 236-243.
- [16] Martínez, J. M.(1984). A quasi-Newton method with modification of one column per iteration, *Computing 33*, pp 353-362.
- [17] Martínez, J. M.; Ochi, L. S.(1982). Sobre dois métodos de Broyden, Matemática Aplicada e Computacional 1, pp 135-141.
- [18] Martínez, J. M.; Zambaldi, M. C.(1992). An inverse column-updating method for solving large-scale nonlinear systems of equations, *Optimiza*tion Methods and Software 1, pp 129-140.
- [19] Meyer, C. C. (2000). Matrix Analysis and Applied Linear Algebra, SIAM, Philadelphia.
- [20] Moré, J. J.; Garbow, B. S.; Hillstrom, K. E. (1981). Testing unconstrained optimization software, ACM Transactions on Mathematical Software 7, pp 17-41.
- [21] Ortega, J. M.; Rheinboldt, W. G.(1970). Iterative solution of nonlinear equations in several variables, Academic Press, NY.

- [22] Pérez, R.; Lopes, V. L. R. (2001). Recent applications of quasi-Newton methods for solving nonlinear systems of equations. *Technical Report* 26/01. Departamento de Matemática Aplicada, IMECC, UNICAMP, Campinas, Brasil.
- [23] Pérez, R.; Lopes, V. L. R. (2001). Solving recent applications by quasi-Newton methods. *Technical Report 44/01*. Departamento de Matemática Aplicada, IMECC, UNICAMP, Campinas, Brasil.
- [24] Rheinboldt, W.C. (1998). Methods for Solving Systems of Nonlinear Equations, SIAM, 2nd. edition, Philadelphia.
- [25] Zambaldi, M. C.(1993). Novos resultados sobre fórmulas secantes e aplicações, Tese de Doutorado, Departamento de Matemática Aplicada, UNI-CAMP, Campinas, Brasil.