ALGORITMOS PARA RESOLUÇÃO DE SISTEMAS NÃO LINEARES ESPARSOS TARCÍSIO LATERZA LOPES

Orientador

Prof. Dr. José Mário Martinez

Dissertação apresentada no Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do título / de Mestre em Matemática Aplicada.

Maio, 1982

UNICAMP
BIBLICISCA CENTRAL

Aos meus pais.

ACRADECIMENTOS

Ao Martinez, pela proposta desse trabalho, pe lo apoio e orientação constantes.

A todos os amigos que, de uma forma ou de outra, me ajudaram a realizar esse trabalho.

 $\tilde{\Lambda}$ Maria do Carmo, por sua ajuda, carinho e apoio.

Ao Laboratório de Matemática Aplicada, em cujo contexto este trabalho foi realizado.

ÍDICE

INTRODUÇÃO	0.
CAPÍTULO I - MÉTODO DE NEWTON E MÉTODOS (JUASE-NEWTON PARA	
RESOLUÇÃO DE SISTEMAS NÃO LINEARES	02
Método de Newton	02
Métodos Quase-Newton	03
CAPÍTULO II - RESOLUÇÃO DE SISTEMAS LINEARES ATRAVÉS DE DE-	
COMPOSIÇÃO III	07
Introdução	07
0 Método	07
CAPÍTULO III - ARMAZENAMENTO DA DECOMPOSIÇÃO LU DE UMA MA -	
TRIZ ESPARSA	13
Matrizes a serem armazenadas	13
Armazenamento e operação das matrizes L, U, P e R ⁻¹	13
CAPÍTULO IV - ALGORITMOS PROPOSTOS	18
CAPÍTULO V - EXPERIÊNCIAS NUMÉRICAS E CONCLUSÕES	23
Experiências realizadas	23
Conclusões	25
CAPÍTULO VI - REFERÊNCIAS	27
APÉNDICE - DOCUMENTAÇÃO DA SUBROTINA SNLESP	29
Propósito	29
Parâmetros	29
Subrotinas providas pelo usuário	32
Alternativas de métodos	34
Tamanho do programa	35
Precisão	35
Acesso à subrotina SNLESP	35
Exemplo de programa principal	36
Da f a	วด

INTRODUÇÃO

Neste trabalho são apresentados dois algoritmos para resolução de sistemas não lineares esparsos.

No capítulo I os métodos mais conhecidos são descritos e discutidos.

No capítulo II um método para fatorização IV de matrizes esparsas é apresentado.

No capítulo III é explicado o armazenamento na memória do computador de matrizes esparsas e de sua decomposição LU.

O capítulo IV apresenta os dois algoritmos propostos para resolução de sistemas não lineares esparsos.

No capítulo V são apresentadas as experiências numéricas e conclusões. As experiências foram realizadas no computador PDP-10 da UNICAMP, e o programa foi feito em FORTRAN-10 e sua documentação, disponível no laboratório de Matemática Aplicada encontra-se no apêndice deste trabalho.

02

I.1 Método de Newton

O problema consiste en achar $x^* \in \Omega \subset R^n$,

tal que $F(x^*)=0$, orde:

$$\begin{cases}
F : \Omega \longrightarrow \mathbb{R}^{n} \\
F = (f_{1} \dots f_{n})^{T} \\
f_{1} : \Omega \longrightarrow \mathbb{R}
\end{cases}$$

$$\Omega \text{ aborto}$$

$$F \in C^{1}(\Omega)$$
(1.1)

Dado $x^k \in \Omega$, uma aproximação da solução x^* de l.l e desenvolvendo—se por Taylor, obtemos :

$$\text{onde} \begin{cases} F(x) \simeq F(x^k) + F'(x^k) (x - x^k) = L(x) \\ F'(x^k) = \left(\frac{\partial f_i}{\partial x_j} (x^k) \right) = \text{matriz jacobiana de F.no} \\ \text{ponto } x^k \end{cases}$$

0 método de Newton consiste em tomar como a - proximação \boldsymbol{x}^{k+1} , a solução de $L(\boldsymbol{x}){=}0$, ou seja :

$$\begin{cases} x^{k+1} = x^k + \triangle x^k \\ F'(x^k) \triangle x^k = -F(x^k) \end{cases}$$
 (1.2)

Se n $\hat{\mathbf{e}}$ muito grande mas o jacobiano $\hat{\mathbf{e}}$ esparso, podemos permutar linhas e colunas de $\mathbf{F}'(\mathbf{x}^k)$ a cada iteração, antes de resolver 1.2. Estas permutações objetivam reduzir o aparecimento de elementos não nulos durante o processo de eliminação de Gauss, necessário $\hat{\mathbf{a}}$ resolução de 1.2. Uma técnica de permutação $\hat{\mathbf{e}}$ explicada em [8].

Uma segunda opção seria permutar a matriz de estrutura esparsa E do jacobiano, antes de se iniciarem as iterações. A matriz E se define assim:

$$\begin{cases} E = (E_{ij}) \\ F'(x) = (F'_{ij}) \end{cases}$$

$$\begin{cases} E_{ij} = 0 \iff F'_{ij} = 0, \forall x \in \Omega \\ E_{ij} \neq 0 \text{ caso contrario} \end{cases}$$

As permutações aplicadas em E seriam aplicadas ao jacobiano calculado a cada iteração, antes de resolver 1.2. Uma terceira opção seria a utilização de métodos iterativos para resolver 1.2.

Cada iteração do método de Newton exige o cálculo do jacobiano e a resolução de um sistema linear. Uma tentativa de baratear cada iteração dá origem ao método de Newton com Refinamentos, que consiste em utilizar o mesmo jacobiano em p+1 iterações consecutivas, ou seja, a cada iteração do método de Newton seguem-se p subiterações onde o jacobiano não muda, logo não precisa ser calculado. O número ótimo de subiterações pode ser calculado usando noções teóricas [1,10,14] ou práticas [1].

I.2 MÉTODOS QUASE-NEWTON [2,4,5]

Estes métodos surgiram da tentativa de baratear o cálculo do jacobiano e de sua inversa, mas conservando as propriedades es senciais do método de Newton. Consiste em substituir o jacobiano por uma matriz B_{k+1} , não singular, que satisfaça :

$$B_{k+1} \triangle x^k = \triangle F^k = F(x^{k+1}) - F(x^k)$$
 (1.3)

propriedade esta que seria satisfeita se F fosse linear e B_{k+1} o jacobiano verdadeiro. Nesse caso, l.2 é substituído por :

$$\text{onde} \begin{cases} x^{k+1} = x^k + \triangle x^k \\ B_k \triangle x^k = -F(x^k) \end{cases}$$

O que diferencia um método quase-Newton de outro \tilde{e} a fórmula utilizada para calcular B_{k} , pois existem infinitas matrizes que satisfazem 1.3. As fórmulas mais conhecidas são :

$$\mathbf{B}_{k+1} = \mathbf{B}_{k} + (\triangle \mathbf{F}^{k} - \mathbf{B}_{k} \triangle \mathbf{x}^{k}) (\triangle \mathbf{x}^{k})^{\mathrm{T}} / (\triangle \mathbf{x}^{k})^{\mathrm{T}} \triangle \mathbf{x}^{k}$$
 (1.4)

- a fórmula de Broyden Ruim [2,7,12,13]

$$\mathbf{B}_{k+1} = \mathbf{B}_k + (\triangle \mathbf{F}_k - \mathbf{B}_k \triangle \mathbf{x}^k) (\triangle \mathbf{F}^k)^T \mathbf{B}_k / (\triangle \mathbf{F}^k)^T \mathbf{B}_k \triangle \mathbf{x}^k$$

As principais características destas fórmulas são:

a) as matrizes B_{k+1} se obtém através de atualiçãoes de posto um à matriz B_{k} , ou seja ,

$$B_{k+1} = B_k + \Delta B_k$$
, $\Delta B_k = vw^T$ (v,w $\in \mathbb{R}^n$)

- b) a estrutura esparsa de B_knão é conservada
- c) a fórmula 1.4 é a única solução do problema de minimiza ção:

$$\min \left\{ \left\| B - B_{k} \right\|_{F} : B \triangle x^{k} = \triangle F^{k} \right\}$$

onde ||.|| $_{F}$ é a norma de Frobenius ||A || = $\sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j}^{2}}$

$$||A|| = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^{2}}$$

Outra fóngula desenvolvida por Broyden [3,6] serva a estrutura esparsa do jacobiano e é apropriada para sistemas esparsos de grande porte. A fórmula é:

$$B_{k+1} = B_k + \Delta B_k$$

$$\Delta B_k = \sum_{k=1}^{N} \left[(S_i \triangle x^k)^T (S_i \triangle x^k) \right]^{+} e_i^T (\triangle F^k - B_k \triangle x^k) e_i (S_i \triangle x^k)^T$$

onde S, são operadores de projeção "esparsa", que projeta Δx^{K} ortogonalmente (norma de Frobenius) nos subespa $\cos\ Z_{\underline{i}}$ (i=1,...,n) , onde

$$Z_{i} = \begin{cases} v \in \mathbb{R}^{n} : e_{j}^{T}v = 0 \text{ para todo } j \text{ tal que} \\ F_{i,j} = 0, \text{ para todo } x \in \Omega \end{cases}$$

Z_i C Rⁿ é o subespaço que identifica a estrutura es parsa da i-ésima linha do jacobiano.

Então o subespaço Z \in R que identifica a estrutura esparsa do jacobiano é definido por :

$$Z = \left\{ A \in R^{n \times n} : A^{T} e_{i} \in Z_{i} \text{ para } i=1,...,n \right\}$$

Na fórmula 1.5, (.) $^+$ é a inversa generalizada. Para um escalar a, $a^+=0 \iff a=0$ e $a^+=a^{-1} \iff a\neq 0$ Características desta fórmula :

- a) as matrizes B_{k+1} são obtidas de B_k com um custo proporcional ao número de elementos distintos de zero de um membro de z
- b) as matrizes $B_{k+1}^{}$ são obtidas por atualizações de posto n_{\star}
- c) a estrutura esparsa do jacobiano é mantida.
- d) a fórmula 1.5 é a única solução de :

$$\begin{aligned} & \text{Min} \left\{ \left\| \, \mathbf{B} \, - \, \mathbf{B}_{k} \, \right\|_{\,F} \, : \, \mathbf{B} \, \in \, \mathbf{Y} \, \cap \, \mathbf{z} \, \right\} \\ & \text{onde } \mathbf{Y} \, = \, \left\{ \, \mathbf{A} \, \in \, \mathbf{R}^{n \, \, \mathbf{X} \, \, n} \, : \, \mathbf{A} \triangle \mathbf{x}^{k} \, = \, \triangle \mathbf{F}^{k} \, \right\} \end{aligned}$$

A desvantagem desta fórmula é que não se obtém diretamente a inversa de B_{k+1} a partir de B_k . Sendo assim, a fórmula 1.3 é resolvida a cada iteração, fatorizando-se a matriz B_{k+1} .

J. E. Dennis e Earl S. Marwil [6] desenvolveram outra fórmula baseada na decomposição $L_k U_k$ da matriz $P_k B_k$, sendo P_k a matriz de permutação de linhas resultante da estratégia de pivotamen to parcial ; $L_k \in \mathcal{K}_k$ sendo este último um subespaço afim de matrizes n x n triangulares inferiores ; $U_k \in \mathcal{U}_k$ um sub-

espaço n x n de matrizes triangulares superiores.

Assume—se que \mathcal{L}_k e \mathcal{U}_k refletem a estrutu—ra esparsa de P_k A = LU para A \in Z . P_k pode ser escolhida para afetar—a esparsidade de \mathcal{L}_k e \mathcal{U}_k

A cada iteração, a fórmula mantém L_k constante e atualiza U_k pela atualização esparsa de Broyden. Ou seja,

$$\begin{split} & P_{k+1} = P_k \\ & I_{k+1} = I_k \\ & U_{k+1} = U_k + \triangle U_k \\ & \triangle U_k = \sum_{i=1}^n \left[\left((S_i \triangle x^k)^T (S_i \triangle x^k) \right) \right] + e_i^T w e_i (S_i \triangle x^k)^T \right] \\ & \text{Onde} \left\{ S_i \text{ são operadores de projeção "esparsa" em \mathcal{U}_k} \\ & w = v_k - U_k \triangle x^k \\ & v_k = I_k^{-1} \triangle F^k \\ \end{split}$$

Neste caso
$$P_{k+1}B_{k+1} = L_{k+1}U_{k+1}$$
.

Caracteristicas desta fórmula:

- a) as matrizes \mathbf{U}_{k+1} são obtidas de \mathbf{U}_k com um custo proporcional ao número de elementos não nulos de um membro de \mathbf{U}_k
- b) a estrutura esparsa da decomposição III do jacobiano é mantida
- c) a formula 1.6 resolve :

$$\min \left\{ \begin{array}{l} \| \, \mathbf{U} - \mathbf{U}_{\mathbf{k}} \|_{\mathrm{F}} : \, \mathbf{U} \in \mathbf{U}_{\mathbf{k}} \,\, \tilde{\mathbf{e}} \,\, \mathbf{o} \,\, \text{ponto mais proximo} \\ \mathrm{em} \,\, \mathbf{U}_{\mathbf{k}} \,\, \mathrm{de} \,\, \mathbf{Y} \end{array} \right.$$

d) a fórmula 1.3 \tilde{e} resolvida a cada iteração sem a necessidade de se fatorizar a matriz B_{k+1} , apenas resolvendo dois sistemas triangulares.

II. RESOLUÇÃO DE SISTEMAS LINEARES ATRAVÉS DE DECOMPOSIÇÃO LU

II.1 Introdução

A decomposição IV de uma matriz quadrada não singular é uma técnica baseada na eliminação de Gauss e consiste em decom por a matriz em fatores triangulares, L (triang. inferior) e V (triang. su perior), tendo um deles diagonal unitária. A decomposição é obtida através de operações entre linhas (fatoração por colunas) ou colunas (fatoração por linhas). No caso de sistemas não lineares, a matriz a ser decomposta é o jacobiano do sistema, que é mais simples de ser calculado por linhas, motivo pelo qual estudaremos a decomposição IV por linhas.

II.2 O Método

Aplicam-se operações e permutações entre colunas da matriz de tal forma a reduzí-la a uma matriz triangular inferior. As permutações e operações entre colunas são realizadas através de pós-multiplicação por matrizes de permutação simples e por matrizes elementares, ∞ mo será visto a seguir.

Uma matriz de permutação simples Q é uma matriz identidade com a p-ésima linha (ou coluna) permutada com alguma linha (ou coluna) j. Quando pré-multiplicada por uma matriz A, o resultado é a matriz A com as colunas p e j permutadas.

Exemplo:
$$Q_{2} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$
 (columns 2 e 4 permutadas)
$$p = 2 \quad j = 4$$

$$A = \begin{pmatrix} 11 & 12 & 13 & 14 \\ 21 & 22 & 23 & 24 \\ 31 & 32 & 33 & 34 \\ 41 & 42 & 43 & 44 \end{pmatrix}$$

$$AQ_2 = \begin{pmatrix} 11 & 14 & 13 & 12 \\ 21 & 24 & 23 & 22 \\ 31 & 34 & 33 & 32 \\ 41 & 44 & 43 & 42 \end{pmatrix}$$

Uma matriz elementar T é uma matriz identidade com a parte à direita da diagonal da p-ésima linha preenchida com ele
mentos escolhidos de tal maneira a zerar a parte à direita da p-ésima linha de uma matriz A.

Estes elementos são calculados pela fórmula:

$$T_{p,ij} = (-a_{ij} / a_{ii})$$
 para $i = p$
 $j = i+1,...,n$

A matriz T deve ser pré-multiplicada pela ma

triz A. Exemplos:

$$A = \begin{pmatrix} 8 & 5 & 0 & 3 \\ 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 \\ 1 & 7 & 6 & 2 \end{pmatrix}$$

$$T_{1} = \begin{pmatrix} 1 & -5/8 & 0 & -3/8 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$T_{3} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -9/8 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$AT_{1} = \begin{pmatrix} 8 & 0 & 0 & 0 \\ 2 & 7/4 & 4 & 17/4 \\ 6 & 13/4 & 8 & 27/4 \\ 1 & 51/8 & 6 & 13/8 \end{pmatrix}$$

$$AT_{3} = \begin{pmatrix} 8 & 5 & 0 & 3 \\ 2 & 3 & 4 & 1/2 \\ 6 & 7 & 8 & 0 \\ 1 & 7 & 6 & -19 \end{pmatrix}$$

As matrizes de permutações simples servem para escolher o elemento pivô. A escolha do pivô é feita através de dois parâmetros, TOLPIV e TOLABS, e pode ser explicada através do seguinte algoritmo:

a)
$$\ell = \text{Arg max} \left| \begin{array}{c} a \\ pk \end{array} \right|$$

Se $\left| \begin{array}{c} a \\ p\ell \end{array} \right| < \text{TOLABS}, pare (matriz singular)$

- q = m (d
- c) m = m + 1

d) Se
$$\left| \frac{a}{pn} \right| < \frac{TOLABS}{a} \frac{Ou}{a} \left| \frac{a}{a} \frac{pn}{pq} \right| < TOLPIV, var para c)$$

- e) PIVO = a
- f) $Q_{\rm p}$ = matriz identidade com as colunas p e m permutadas

A fatoração realiza-se através da sequinte

fórmula:

$$\mathbf{AQ_1T_1Q_2T_2} \ \dots \ \mathbf{Q_{n-1}T_{n-1}} = \mathbf{L}$$

onde n = ordem da matriz A

L = matriz triangular inferior

O = matrizes de permutação simples, determinadas pela
escolha do pivo

T = matrizes elementares, utilizadas para zerar a parte
à direita da p-ésima linha transformada de A

(TOLABS = 1.E-10 ; TOLPIV = 0.5)Exemplo:

$$A = \begin{pmatrix} 3 & 6 & 0 & 12 \\ 5/2 & 5 & 8 & 34 \\ 1 & 0 & 31/3 & 1 \\ 18/5 & 3 & 8/3 & 11 \end{pmatrix}$$

$$Q_{3} = I = Matriz Identidade 4x4$$

$$AQ_{1}T_{1}Q_{2}T_{2}Q_{3} = \begin{pmatrix} 6 & 0 & 0 & 0 \\ 5 & 24 & 0 & 0 \\ 0 & 1 & 10 & 1 \\ 3 & 5 & 1 & 21/10 \end{pmatrix}$$

$$T_{3} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1/10 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$AQ_{1}T_{1}Q_{2}T_{2}Q_{3}T_{3} = \begin{pmatrix} 6 & 0 & 0 & 0 \\ 5 & 24 & 0 & 0 \\ 0 & 1 & 10 & 0 \\ 3 & 5 & 1 & 2 \end{pmatrix} = L$$

$$\begin{array}{lll} {\rm AQ_1T_1Q_2T_2Q_3T_3} &=& {\rm L} \\ \\ {\rm A} &=& {\rm LT_3^{-1}Q_3^{-1}T_2^{-1}Q_2^{-1}T_1^{-1}Q_1^{-1}} \\ \\ {\rm Mas} && {\rm Q_p^{-1}} &=& {\rm Q_p} \quad (~{\rm fācil~verificação}~) \end{array}$$

$$T_{p}^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & \vdots \\ \vdots & \ddots & & \vdots \\ 0 & \cdots & 1 & -T_{p,p,p} & \cdots & -T_{p,p,n} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ 0 & \cdots & & 1 \end{pmatrix}$$
 \triangleleft - linha p

$$\text{Logo A} = \text{L (} \text{T}_{3}^{-1} \text{Q}_{3} \text{T}_{2}^{-1} \text{Q}_{2} \text{T}_{1}^{-1} \text{Q}_{1} \text{)}$$

Pós-multiplicando os dois termos da igualdade por $Q_1Q_2Q_3$,

obtemos:

$$AQ_{1}Q_{2}Q_{3} = L (T_{3}^{-1}Q_{3}T_{2}^{-1}Q_{2}T_{1}^{-1}Q_{1})Q_{1}Q_{2}Q_{3} = L U$$

Fazendo as contas:

$$U = \begin{pmatrix} 1 & 2 & 0 & 1/2 \\ 0 & 1 & 1/3 & 0 \\ 0 & 0 & 1 & 1/10 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Assim, conseguimos encontrar L (triangular inferior) e U (triangular superior unitária) tais que $AQ_1Q_2Q_3=IU$.

Generalizando, temos:

$$AQ_1...Q_{n-1} = W = L(T_{n-1}^{-1}Q_{n-1}...T_1^{-1}Q_1Q_1...Q_{n-1})$$

Se matriz A, antes de ser fatorizada, for submetida a trocas de linhas e colunas, a fórmula fica :

$$A = P^{-1}IUQ_{n-1}...Q_1Q^{-1}$$

onde P é a matriz de permutação das linhas de A (matriz identidade com linhas permutadas)
Q é a matriz de permutação das colunas de A (matriz identidade com colunas permutadas)
PAQ é a matriz apos a troca de linhas e colunas, antes de ser fatorizada

Para resolvermos o sistema linear Ax=b , realizamos as seguintes etapas :

- a) resolva Ly=Pb
- b) resolva Uz=y
- c) faça $x=x_1, x_2, \dots, x_{n-1}, x_n$

III- ARMAZENAMENTO DA DECOMPOSIÇÃO LU DE UMA MATRIZ ESPARSA

III.l Matrizes a serem armazenadas

Inicialmente temos a aproximação do jacobiano permutado PB $_{k}$ Q. Após a decomposição temos B $_{k}$ =P $^{-1}$ IUQ $_{n-1}$...Q $_{1}$ Q $^{-1}$. Para a resolução de um sistema linear, precisamos de P e R (R = Ω_{1} ...Q $_{n-1}$). Para a decomposição, precisamos de P e Q.

 $\label{eq:portanto} \text{Portanto, precisamos armazenar P^{-1} (ou P) , $L , U , Q e R (ou R^{-1}).}$

As matrizes de permutação têm uma inversa fácil de se calcular (é igual à transposta) implicitamente. Não é necessário armazenar a matriz $\mathbf{B}_{\mathbf{k}}$, pois a cada linha calculada desta se obtêm a linha correspondente de Le U.

III.2 Armazenamento e operação das matrizes L , U , P e ${ m R}^{-1}$

a) Armazenamento de L e U

- Vetor real VETIU de dimensão NMEDZ (nº de elementos dis tintos de zero após a fatorização)

Armazena os valores das matrizes L e U, por linha, alternadamente (primeiro a U).

Exemplo:

$$U = \begin{pmatrix} 1 & 3 & 2 \\ & 1 & 5 \\ & & 1 \end{pmatrix} \qquad L = \begin{pmatrix} 4 & \\ & 1 & 3 \\ & 5 & 4 & 1 \end{pmatrix}$$

VETLU: (
$$3$$
, 2 , 1 , 4 , 5 , 1 , 1 , 3 , 1 , 5 , 4 , 1)

$$1^{\underline{a}} \qquad 1^{\underline{a}} \qquad 2^{\underline{a}} \qquad 2^{\underline{a}} \qquad 3^{\underline{a}} \qquad 3^{\underline{a}}$$
linha linha linha linha linha linha da U da L da U da L

- Obs: Os elementos de uma linha da L estarão ordenados por coluna de maneira crescente. Os elementos de uma linha da U virão em qualquer ordem, mas o último elemento sempre pertencerã à diagonal.
- Vetor inteiro APONIL de dimensão N Aponta para o início de cada linha da L, no vetor VEIIU Exemplo:

- Vetor inteiro APONIU de dimensão N Aponta para o início de cada linha da U, no vetor VETIU Exemplo:

- Vetor inteiro COLUNS de dimensão NMEDZ Indica a coluna a que pertence cada elemento do vetor VETLU

Exemplo:

b) Armazenamento das matrizes P, Q, R e R⁻¹

A matriz P é fornecida pelo usuário através do vetor inteiro P, de dimensão n. Neste caso, P(I)=J indica que a I-ésima linha de PB_kQ corresponde à J-ésima linha da matriz B_k.

Exemplo:
$$B_{k} = \begin{pmatrix} 11 & 12 & 13 \\ 21 & 22 & 23 \\ 31 & 32 & 33 \end{pmatrix} P = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} e_{3}^{T} \\ e_{1}^{T} \\ e_{2}^{T} \end{pmatrix}$$

$$PB_{k} = \begin{pmatrix} 31 & 32 & 33 \\ 11 & 12 & 13 \\ 21 & 22 & 23 \end{pmatrix}$$

Então o vetor P será: (3,1,2)

A pré-multiplicação de um vetor b pela matriz P resultando no vetor c se fará pelo seguinte algoritmo :

Para I=l até N faça começo

IND =
$$P(I) / DESLOC$$

 $C(I) = B(IND)$

fim

Obs: Por motivos de maior compactação de dados, o con teúdo que especifica a matriz P está armazenado na metade esquerda de cada posição do vetor P. Para recuperá-lo, basta dividir a posição do vetor P pelo valor DESLOC (no caso do programa implementado, DESLOC = 2^{17}).

Example:
$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \qquad Pb = \begin{pmatrix} b_2 \\ b_3 \\ b_1 \end{pmatrix}$$

Vetor P: (2, 3, 1)

Seguindo o algoritmo:

$$I = 1$$
; $IND = 2$; $C(1) = B(2)$

$$I = 2$$
; $IND = 3$; $C(2) = B(3)$

$$I = 3$$
; $IND = 1$; $C(3) = B(1)$

O usuário também fornece a matriz Q, através do vetor Ω de dimensão n. Neste caso, $\Omega(I)=J$ indica que a $I-\acute{e}si$ ma columa de PB,Q corresponde à $J-\acute{e}si$ ma columa da matriz B_k .

Exemplo:

$$B_{k} = \begin{pmatrix} 11 & 12 & 13 \\ 21 & 22 & 23 \\ 31 & 32 & 33 \end{pmatrix} \quad Q = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = (e_{2} \quad e_{3} \quad e_{1})$$

$$p_{B_{k}} = \begin{pmatrix} 12 & 13 & 11 \\ 22 & 23 & 21 \\ 32 & 33 & 31 \end{pmatrix}$$

Então o vetor Q será: (2,3,1)

Após a fatoração LU da matriz ${\tt B}_{\rm k}$, o armazenamento das matrizes de permutação estará na seguinte situação :

A pos-multiplicação da matriz R por um vetor x resultando o vetor y se fará através de um dos seguintes algoritmos:

10) Para I=l até N faça começo
$$IND = Q(I) \cdot AND \cdot MASKRA$$

$$Y(IND) = X(I)$$
 fim

29) Para I=l até N faça começo
$$IID = P(I) .AND. MASKRA$$
$$Y(IND) = X(I)$$

fin

Obs : A expressão P(I) .AND. MASKRA serve para pegar o lado direito da I-ésima posição do vetor P (no caso do programa imple mentado, MASKRA = 2^{17} -1).

Exemplo:

$$R = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = (e_{2} e_{3} e_{1})$$

$$R^{-1} = R^{T} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} = (e_{3} e_{1} e_{2})$$

$$Rx = \begin{pmatrix} x_{3} \\ x_{1} \\ x_{2} \end{pmatrix}$$

Vetor P: (2,3,1) - metade direita Vetor Q: (2,3,1) - metade direita

Seguindo os algoritmos:

$$I = 1$$
; $IND = 3$; $Y(2) = X(1)$
 $I = 2$; $IND = 1$; $Y(3) = X(2)$
 $I = 3$; $IND = 2$; $Y(1) = X(3)$

IV ALGORITMOS PROPOSTOS:

A fórmula 1.6 proposta por Dennis & Marwil se caracteriza pela decomposição IV da matriz B_k e pela atualização do fator que não tem diagonal unitária, a cada iteração do método. Se considerarmos que a fatorização é feita por linhas, o fator com diagonal unitária será a matriz triangular superior V. Neste caso, a fórmula 1.6 modificada para o caso de fatorização por linhas e generalizada para o caso da matriz B_k ser multiplicada à esquerda e à direita por matrizes de permutação, ficaria :

$$\begin{split} & P_k = P_k^{-1} L_k U_k P_k^{-1} \\ & P_{k+1} = P_k \\ & P_{k+1} = R_k \\ & U_{k+1} = U_k \\ & U_{k+1} = L_k + \Delta L_k \\ & \Delta L_k = \sum_{i=1}^n \{ \left((S_i U_k y_k)^T (S_i U_k y_k) \right)^+ e_i^T w_k e_i \left(S_i U_k y_k \right)^T \} \\ & \text{onde} \left\{ S_i \text{ são operadores de projeção "esparsa" em } \right\}_k \\ & V_k = R_k^{-1} \Delta x^k \\ & W_k = P_k \Delta F^k - L_k U_k y_k \end{split}$$

O primeiro algoritmo proposto é a atualização da matriz triangular que possui diagonal unitária (no caso da fatorização por linhas, é a U), fazendo com que esta matriz perca a unitariedade da sua di agonal nas iterações onde se realiza a atualização. Neste caso, o sub-es paço afim \mathcal{U}_k é formado pelas matrizes triangulares superiores nxn que refletem a estrutura esparsa de \mathbf{U}_k , sem a restrição de diagonal unitária. A fórmula, então, seria idêntica à fórmula 1.6, generalizada para o caso de \mathbf{B}_k ter linhas e columas permutadas :

$$\begin{split} \mathbf{B}_{k} &= \mathbf{P}_{k}^{-1} \mathbf{L}_{k} \mathbf{U}_{k} \mathbf{R}_{k}^{-1} \\ \mathbf{P}_{k+1} &= \mathbf{P}_{k} \\ \mathbf{P}_{k+1} &= \mathbf{R}_{k} \\ \mathbf{U}_{k+1} &= \mathbf{U}_{k} + \Delta \mathbf{U}_{k} \\ \Delta \mathbf{U}_{k} &= \sum_{i=1}^{n} \{ \left((\mathbf{S}_{i} \mathbf{Y}_{k})^{T} (\mathbf{S}_{i} \mathbf{Y}_{k}) \right)^{+} \mathbf{e}_{i}^{T} \mathbf{V}_{k} \mathbf{e}_{i} (\mathbf{S}_{i} \mathbf{Y}_{k})^{T} \} \\ &\text{onde} \begin{cases} \mathbf{S}_{i} \quad \mathbf{\tilde{s}\tilde{a}o} \quad \mathbf{operadores} \quad \mathbf{de} \quad \mathbf{proje} \mathbf{\tilde{s}\tilde{a}o} \quad \mathbf{"esparsa"} \quad \mathbf{em} \quad \mathbf{\tilde{M}}_{k} \\ \mathbf{Y}_{k} &= \mathbf{R}_{k}^{-1} \Delta \mathbf{x}^{k} \\ \mathbf{V}_{k} &= \mathbf{L}_{k}^{-1} \mathbf{P}_{k} \Delta \mathbf{F}^{k} - \mathbf{U}_{k} \mathbf{Y}_{k} \end{split}$$

O segundo algoritmo proposto é a alternância na atualização, ou seja, em algumas iterações se utiliza a fórmula 4.1 e, em ou tras, a fórmula 4.2. O critério para escolha entre modificar a L ou a U pode ser o mesmo adotado por L.S.Ochi e J.M.Martínez em $\begin{bmatrix} 12,13 \end{bmatrix}$, ou seja realiza-se a modificação que minimize $\|B_{k+1}\Delta x^{k-1} - \Delta F^{k-1}\|$.

Para a fórmula 4.1, temos :

$$\begin{aligned} & || B_{k+1} \Delta x^{k-1} - \Delta F^{k-1} || = \\ & = || P_k^{-1} L_{k+1} U_k R_k^{-1} x^{k-1} - F^{k-1} || \\ & = || P_k^{-1} (L_k + \Delta L_k) U_k R^{-1} \Delta x^{k-1} - \Delta F^{k-1} || \\ & = || P_k^{-1} L_k U_k R_k^{-1} \Delta x^{k-1} + P_k^{-1} \Delta L_k U_k R_k^{-1} \Delta x^{k-1} - \Delta F^{k-1} || \\ & = || B_k x^{k-1} - P_k^{-1} \Delta L_k U_k R_k^{-1} \Delta x^{k-1} - \Delta F^{k-1} || \\ & = || \Delta F^{k-1} - P_k^{-1} \Delta L_k U_k R_k^{-1} \Delta x^{k-1} - \Delta F^{k-1} || \\ & = || P_k^{-1} \Delta L_k U_k R_k^{-1} \Delta x^{k-1} || \\ & = || \Delta L_k U_k R_k^{-1} \Delta x^{k-1} || \end{aligned}$$

17

Para a fórmula 4.2, temos:

$$\left|\left|\left|\left|\mathbf{B}_{k+1}\mathbf{\Delta}\mathbf{x}^{k-1}\right|-\left|\mathbf{\Delta}\mathbf{F}^{k-1}\right|\right|\right|=\left|\left|\left|\mathbf{L}_{k}\mathbf{\Delta}\mathbf{U}_{k}\mathbf{R}_{k}^{-1}\mathbf{\Delta}\mathbf{x}^{k-1}\right|\right|\right|$$

Como este critério é muito trabalhoso, optou-se pela alternância simples, ou seja, se na iteração k houve atualização da L, então na iteração k+l haverá atualização da U e vice-versa.

Foi implementado um programa escrito em FORTRAN-10, no computador PDP-10 da UNICAMP, e sua documentação encontra-se no apêndice. O algoritmo deste programa é explicado a seguir :

- a) Inicializações
- b) Impressão da iteração
- c) Teste de convergência ou fracasso
- d) Decide : 1- fatorização
 2- refatorização
 3- atualização da L
 4- atualização da U
 5- conservação da fatorização
- e) Se a decisão foi 1 ou 2, acha fatorização LU do jacobiano. Se foi 3 ou 4, atualiza o fator correspondente através da formula 4.1 ou 4.2.
- f) Acha Δx^k , x^{k+1} , $F(x^{k+1})$, ΔF^k e $\|F(x^{k+1})\|$
- g) k + k + 1
- h) Atualiza as variaveis XMIN, FXMIN e NITMIN
- i) Vá para b.

Explicações sobre cada passo:

a) Inicializações

Entre outras variaveis, inicializa :

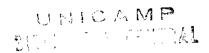
NFATLU=NTT=NTTMIN=0

XK≕X

FXK=FXMIN=F (XK)

NORMFK=NORMIN= || FXK ||

DIST=DISTFX * NORMIN



b) Impressão da iteração

Imprime-se, a cada IMPR iterações, o número da iteração, o tipo da iteração e a norma de F(XK). Se IMPR=0, nada é impresso.IMPR. é um parâme tro de entrada, descrito no apendice.

c) Teste de convergência ou fracasso:

Declara-se convergência (CODERR=0) quando:

NORMEK < EPSLON

OH:

NORMER
$$< \sqrt{\text{EPSLON}}$$
 e $||\Delta x^{k}|| < \text{EPSLON} * (1 + ||XK||)$

Declara-se divergência ou fracasso (CODERR=5) quando NORMFK > DIST O programa também para (CODERR=4) se foi excedido o número máximo de iterações (NIT > NITMAX). CODERR é um parâmetro de saída, descrito no apêndice. EPSLON e NITMAX são parâmetros de entrada, também descritos no apêndice.

d) Decisão sobre a matriz B_{k+1}

Depende dos parâmetros de entrada QUAL, FREQ e REFAT, exceto se NIT=0, neste caso a opção é a fatorização. Nas outras iterações, os seguintes testes são feitos :

Se MOD (NIT, RUFAT)=0 , refatorização

Se MOD (NIT+1,FRFQ)=0, atualização da L (se QUAL=1 ou se QUAL=3 e a última atualização foi da U) ou da U (se QUAL=2 ou se QUAL=3 e a última atualização foi da L).

Se nenhum dos testes foi satisfeito: conservação da fatorização No caso de atualização da L ou da U, pode acontecer um crescimento excessivo de um dos elementos atualizados, ou a equação da fórmula 1.3 não ficar satisfeita. Neste caso, haverá uma refatorização.

e)

f) Cálculo diversos vetores

O cálculo de DELTAX (Δx^k) se faz através do seguinte algoritmo:

1) DELTAX =
$$-B_k^{-1}$$
. FXK

2) NORMDX =
$$||DELTAX||$$

3) Se NORMOX > DISIX então faça DELTAX =
$$\frac{\text{DELTAX}}{\text{NORMOX}}$$
 * DISIX NORMOX = DISIX

O terceiro passo do algoritmo caracteriza o controle do passo, feito através do parâmetro de entrada DISTX.

Os outros cálculos são:

$$XK = x^{k+1} = XK + DELTAX$$

Acha F (XK)
 $DELTAF = FXK - F (XK)$
 $FXK = F (XK)$
 $NORMFK = || FXK||$

g)

h) Atualizações

Se
$$\|FXK\|$$
 < NORMIN então faça $\{NTIMIN = NIT XMIN = XK FXMIN = FXK \}$

V. EXPERIÊNCIAS NUMÉRICAS E CONCLUSÕES

V.1 Experiências realizadas

Foram testadas as duas funções utilizadas

em [3,6] , que são :

$$\begin{array}{lll} 1 - & f_{i}(x) = & (3-k_{1}x_{i})x_{i}+1-x_{i-1}-2x_{i+1}, & 1 < i < n \\ & = & (3-k_{1}x_{i})x_{i}+1-x_{i-1}, & i = n \\ & = & (3-k_{1}x_{i})x_{i}+1-2x_{i+1}, & i = 1 \end{array}$$

$$2 - f_{i}(x) = & (k_{1}-k_{2}x_{i}^{2})x_{i}+1-k_{3} \sum_{j \in B_{i}} (x_{j}+x_{j}^{2}), & i = 1, \dots, n$$

onde $B_i=\{i_1,\ldots,i_2\}-\{i\}$, sendo $i_1=\max(1,i-r_1)$ e $i_2=\min(n,i+r_2)$. Os parâmetros k_1,k_2,k_3 variam para aumentar a não-linearidade do problema, en quanto que r_1 e r_2 determinam a largura da faixa da matriz jacobiano. O chute inicial foi $x_i=-1$, $i=1,\ldots,n$. O critério de convergência utilizado foi $\|F\|_{\infty}<10^{-6}$ (observe que não foi incluído o segundo teste de convergência, especificado no capítulo anterior). Foram rodados os 23 casos especificados em [3,6], e aqui reproduzidos na tabela I. Observe a existência de mais 3 casos, dois deles (24 e 25) se referindo ã função 1, apenas variando o tamanho do problema. O caso 26 se refere à função abaixo :

$$3 - f_i(x) = x_i^2 + \sum_{j \in N_i} x_j - NI - I$$

onde N_i é um donjunto de N_i indices entre l e n, e diferentes de i gerados aleatoriamente para cada linha. Neste caso, o chute inicial foi $x_i=0.5$, $i=1,\ldots,n$. O jacobiano inicial e de cada iteração de refatorização foi o analítico, fornecido pela rotina JACOB.

Na tabela II encontram-se os resultados das rodadas, sendo que a coluna ROYDEN ESPARSO e DENNIS-MARWIL foram tiradas de [6]. A coluna NEWION endica o método de Newton, que se obtêm fazendo

REFAT=1; a coluna NEWTON C/ REF. indica o método de Newton com refina - mentos, que se obtém fazendo REFAT=∞ e FREQ=∞; a coluna MUDA SO L indica o método de Dennis-Marwil modificado (fórmula 4.1), que se obtém fazen do QUAL=1, FREQ=1 e REFAT=∞; a coluna MUDA SO U indica o primeiro algo ritmo proposto (fórmula 4.2), que se obtem fazendo QUAL=2 e FREQ e REFAT iguais ao caso anterior; a coluna MUDA L E U indica o segundo algoritmo proposto, que é obtido fazendo QUAL=3 e FREQ e REFAT iguais ao caso anterior.

Os casos assinalados com * foram aqueles que não convergiram.

Por não se dispor de uma rotina de pré-assinalamento de pivôs, adotou-se P(I)=I e Q(I)=I (matrizes de permutação iguais à identidade).

As colunas NIT,NFATIU,MUTILL,MUTILU e EME indicam o número de iterações, número de fatorizações IU, número de elementos diferentes de zero na L e na U no pior caso (soma máxima) e o valor do parâmetro de entrada EME, respectivamente. Em nenhum dos casos rodados houve controle do passo e o valor do parâmetro TOLPIV foi 10⁻³. Finalmente, o valor do parâmetro DISTEX foi 10³. Houve permutações de colunas apenas no caso 26 e as refatorizações foram devidas a crescimento excessivo da L ou da U.

O programa teve quatro versões. A versão zero equivale ao primeiro programa desenvolvido. As versões seguintes apa receram devido à necessidade de diminuir o tempo de CPU no caso 25, principalmente, que demorava mais de 27 minutos de CPU para executar, quando se utilizava o segundo algoritmo proposto. A partir deste momento, ficou claro que um programa para problemas de grande porte não poderia ser codificado como se fosse um programa comum. Na versão zero, um vetor auxiliar era zerado a cada vez que se chamava a rotina JACOB, o que equivale a ter um código do tipo abaixo (considerando o caso 25):

DO 10 I=1,6000 DO 10 J=1,6000 10 A(J)=0

Este conjunto de instruções leva 4 minutos de CPU para executar. No caso de se rodar o método de Newton, isto
seria executado 4 vezes, o que daria 16 minutos só para zerar um vetor.

Como este vetor era utilizado para armazenar uma linha do jacobiano, nas
versões posteriores, o vetor era zerado uma vez a cada decomposição LU,
a partir daí, apenas as posições preenchidas eram zeradas quando não eram
mais necessárias. Algumas outras melhorias foram incorporadas, o que levou
a uma economia de 7 minutos de CPU no caso da utilização do segundo método proposto, como pode ser visto na tabela III.

Nas versões zero e um, o teste para escolher o elemento pivô era feito em todos os elementos da linha. Nas versões posteriores, o teste foi feito apenas para os elementos contidos entre o primeiros e último diferente de zero da linha. Outra melhoria feita simulta — neamente foi a eliminação da chamada das rotinas que realizam permutações no caso em que estas não acontecem, ou seja, no caso em que P(I)=I ou R(I)=I. Na última versão outras melhorias foram feitas, o que levou a um tempo final de 7 minutos de CPU, uma economia de 75%.

V.2 Conclusões

Como pode ser visto na tabela III, o método de Newton foi o mais demorado de todos. Como pode ser visto na tabela II, o método de Broyden esparso levaria mais tempo ainda, pois também envolve uma decomposição III a cada iteração e, nos casos rodados, o cálculo do jacobiano é barato de se calcular. Resta-nos analisar os outros métodos.Os métodos DENNIS-MARWIL, MUDA SO L e MUDA SO U são praticamente iguais (nos casos rodados), com ligeira vantagem para MUDA SO U, que sempre ganha ou empata com os outros dois, exceto no caso 9, em que a L tem mais elementos diferentes de zero que a U.

O método MUDA L E U pareceu o mais instável de todos, mas é o que ganha na maior parte dos casos, o que leva a crer que o desenvolvimento de um critério mais barato para se escolher entre mudar a L ou a U numa certa iteração deve compensar.

O método de Newton com refinamento levou um tempo de excução parecido com os 4 últimos citados anteriormente, mas falhou no caso 26, exatamente por não prever refatorização nas primeiras iterações.

TABELA I

CASO	FUNÇÃO	N	K _I	К2	К ₃	R _I	R ₂	N _I
1	ŧ	5	0.1		: <u> </u>	- <u> </u>	_	
2	1	5	Q5		_	_		· —
3	I	Ю	0.5			<u> </u>	<u></u>	
4	ı	50	0.5	<u></u>	****	, -	-	-
5	1	600	ο5	4-16-	1994	_	_	_
6	1	600	2	- -			<u>-</u>	-
7	2	100	I	1	. 1	3	3	_
8	2	100	1	1	<u> </u>	2	4	-
9	2	100	<u> </u>	1	1	5	ı	
10	2	50	1	1	1	5	5	_
11	2	50	2	I	 I	5	5	
12	2	50	1	2	I	5	5	_
13	2	50	3	2	1	5	5	
14	2	50	2	3	l	5	5	_
15	2	50	3	3	i	5	5	
16	2	50	2	2	ı	5	5	_
17	2	50	i	2	2	5	5	_
18	2	50	2	2	5	5	5	
19	2	50	2	3	2	5	5	_
20	2	50	2	4	l	5	5	-
21	2	50	2	5		5	5	_
55	2	50	3	4	L .	5	5	<u> </u>
23	2	50	3	5	1	5	5	
24	ı	2000	0.5	<u>-</u>		_		
25	ı	60 00	0.5		_	<u> </u>		
26	3	100	_	-	-	-	-	9

TABELA II

CASO	NEWTON NIT=NFATLU	NEWTON	NEWTON C/ REF(NEATLU = 1)	MUD	MUDA SÓL		να sού υ	MUD	ALeU	BROYDEN ESPARSO	DENNIS- - MARWIL		EME	MUTILL	MUTILLI
		NIT	NIT	NFATLU	NIT	NFATLU	NIT	NFATLU	NIT=NFATLU	NIT	NFATLU	EWE	MOLILL	MICTAL	
ı	3	8	6	i	5	1	6	ı	5	5	ı	103	9	9	
2	3	7	5	ı	5	ı	5	l	4	5	1	103	9	9	
3	3	51	6	I	6	ı	5	1	5	6	1	³	19	19	
4	4	14	7	1	6	ı	6	1	5	6	1	Ю ³	39	39	
5	4	15	7	1	6	1	6	† ———— 	5	6	!	10 ³	1199	1199	
6	4	15	9	; I	7	l	6	1	7	8	1	iO ³	1199	1199	
7	4	10	7	ı	6	1	6		6	7		ю³	394	394	
8	4	10	7	1	6	ı	6	1	6	7		103	297	490	
9	4	12	6	ı	7	1	5		6	7	1	103	585	199	
10	3	9	7	1	6	-	6	ı	6 /	7		103	285	285	
11	4	13	8	l	8	l	6	1	7	8	I	103	285	285	
15	4	13	9	ţ ţ	8	: I	7	1	8	8	1	103	285	285	
13	4	21	6	2	11	l	6	2	9	12	1	103	285	285	
13	4	21	12	· I	l1	!	9	· ·	9	12	;	104	285	285	
14	4	21	6	2	7	2	9	ı	11		1	10 ³	285	285	
14	4	21	11	1	7	2	9	1	11	11	<u> </u>	104	285	285	
14	4	21	11	1	11	•	9	ı	11	11	1	105	285	285	
15	5	25	15	t	12		- II	1	11	13		103	285	285	
16	4	17	8	2	10	1	9		9	10		103	285	285	
16	4	17	Ю	ı	10	1	9	ı	9	ю	·	104	285	285	

TABELA II cont.

CASO	CASO	NEWTON	NEWTON C/ REF(NFATLU=1)	MUD	A SÓL	Mul	DA SÓU	MUD	ALeU	BROYDEN ESPARSO		ENNIS - ARWIL	EME	MUTILL	MUTILU
	NIT= NFATLU	NIT	NIT	NFATLU	NIT	NFATLU	NIT	NFATLU	NIT=NFATLU	NIT	NFATLU	_ (VI_	I WOTILL	MOTILO	
17	4	15	12	2	10	1	!7	2	В	ю	1	103	285	285	
17	4	15	25	l	ю	ι	18	1 #	8	Ю	1	Ю ³	265	285	
18	4	14	11	1	9	•	11	. 1	8	9	1	103	285	285	
19	4	16	10	l	9	1	8	1	9	10	i.	103	285	285	
20	4	25	8	2	12	ı	8	2	12	13	1	ю ³	285	285	
20	4	25	13	2	12	1	10	l	15	13	i i	10 ⁴	285	285	
20	4	25	13	i	12	1	Ю	l	12	13	i	105	265	285	
21	5	28	15	2	7	2	10	2	13	14	. 1	103	285	285	
21	5	28	15	· I	l3	1	25	: 2	13 -	14	1	iO ⁴	285	285	
55	5	29	- 11	2	11	5	8	5	13	14	l	103	285	285	
22	5	29	15	1	14		Н	1	13	14	1	10 ⁴	285	285	
23	5	33	17		9	2	9	2	13	16	1	10 ³	285	285	
23	5	33	17	ı	16		18	2 🌞	13	16	. 1	IO ⁴	285	265	
23	5	33	17	1	16		21	i *	13	16	<u> </u>	Ю ⁵	285	285	
24	4	15	7	l	6	l	6	:		-	-	103	3999	3999	
25	4	15	7	1	6	1	. 6	1	-	-	-	Ю ³	11999	11999	
26	3	13 *	4	2	4	2	4	2	_	-		εOd	3822	3633	

TABELA III

VERSAO	CASO	NEWTON	NEWTON C/ REFINAMENTO	MUDA SÓ L	MUDA SÓ U	MUDA L. e U
0	5	85 .3	24.8	23.6	23.3	23.5
3	5	20 .5	8.5	8 .9	8.6	8.5
0	6	86.2	25.4	26.5	25.5	25.4
3	6	21.5	8.9	9,3	8.8	8.1
0	24	_	-	224.3	232.4	225.3
2	24	-	-	79	76.6	77.3
3	24	170	53.8	54.4	52.7	56,8
0	25	-		-		1663
ı	25	_				1220
3	25	1422	330	381	394	376
3	26	31	15	23	23	23

VI REFERÊNCIAS

- 1. R. P. Brent, "Some efficient algorithms for solving systems of non-linear equations", SIAM J. Numer. Anal., 10, 327-344 (1973)
- 2. C. G. Broyden, "A class of methods for solving nonlinear simultaneous equations", Math. Comp. 19, 577-593 (1965)
- 3. _____, "The convergence of an algorithm for solving sparse nonlinear systems", Math. Comp. 25, 285-294 (1971)
- 4. J. E. Dennis, "A brief introduction to Quasi-Newton methods", Cornell University, Computer Science Department, TR 77-327 (1977)
- 5. _____e J. J. Moré, "Quasi-Newton methods, motivation and theory", SIAM Review 19, 46-89 (1977)
- e E. S. Marwil, "Direct secant updates of Matrix factorizations", Technical Report Of Rice University, Department of Mathematical Sciences (1981)
- 7. D. M. Gay, "Some convergence properties of Broyden's method", SIAM J. Numer. Anal. 16, 623-630 (1979)
- 8. E. Hellerman and D. C. Rarick, "Reinversion with the preassigned pivot procedure", Math. Prog. 1, 195-216 (1971)
- 9. J. M. Martinez, "Algoritmos para resolução numérica de sistemas não lineares", Atas do XIII Colóquio Brasileiro de Matemática, 1981
- 10. ______, "Generalization of the methods of Brent and Brown for solving nonlinear simultaneous equations", SIAM J. Numer. Anal. 16, 434-448 (1979)

- 11. J. M. Martinez, "Solving nonlinear simultaneous equations with a generalization of Brent's method", BIT 20, 501-510 (1980)
- 12. e L. S. Ochi, "Sobre dois métodos de Eroyden", Mat. Apl. e Comput., Vol. 1, Nº 2, 1982
- 13. L. S. Ochi, "Algoritmos de Broyden combinados para resolução de sistemas não lineares", Tese de Mestrado, Departamento de Matemática Aplicada, UNICAMP, 1981
- 14. A. N. Ostrowski, Solution of equations in Euclidean and Banach spaces, Academic Press, New York and London, 1973
- 15. M. J. D. Powell, "A hybrid method for nonlinear equations", Numerical methods for nonlinear equations, editado por P. Rabinovitz, Gordon and Breach, 1970

APĒNDICE

"SNLESP"

(Subrotina; Fortran-10)

l.-Proposito

Resolver um sistema não linear de equações algébricas :

$$f_1(x_1,...,x_n) = 0$$

$$\vdots$$

$$f_n(x_1,...,x_n) = 0$$

de preferência com jacobiano esparso.

Utiliza-se métodos Quase-Newton com atualizações esparsas. De acordo com os parâmetros QUAL, FREQ e REFAT, pode-se utilizar o método de Newton, o método de Newton com refinamento, o método de Dennis & Marwil (ligeiramente modificado) ou um dos métodos introduzidos por Lopes, através dos parâmetros acima. Pode-se utilizar também uma combinação destes métodos dando a estes parâmetros valores apropriados (ver Item 5).

2.-Parâmetros

O comando de chamada de SNLESP deve ser do seguinte tipo:

CALL SMLESP (N, METOR, DISTEX, DISTEX, EPSLOY, TOLDTV,

- * TOLABS, E4E, NITHAX, IMPR, SAIDA, QUAL, FREQ, REFAT,
- * P,Q,F,JACOB,X,TRAB,FXMIN,NIT,NITMIN,MUTULL,
- * MUTILU, NFATLU, CODERR)

onde:

N - Número de funções e de variáveis do sistema

MEMOR - Dimensão do vetor TRAB. Deve ser pelo menos 8N + 2(NL + NU) + 1, orde NL(NU) = NP estimado de elementos diferentes de zero da L(U), incluindo a diagonal.

- DISTX Limite superior estimativo para a distância entre o ponto inicial e a solução (norma suprema da diferença dos dois pontos).

 Serve para a realização de controle do passo. Ver referência no item 4.
- DISTFX Limite superior para a razão entre as normas de $F(x^{n+1})$ e $F(x^0)$. Serve para determinação de divergência. Ver referência no item 4.
- EPSLON Precisão desejada na solução. O programa declarará convergência quando a norma suprema de F(x) for menor do que EPSLON ou quando, simultaneamente, a norma suprema de F(x) for menor que $\sqrt{\text{EPSLON}}$ e a razão entre as normas supremas de Δx^K e x^K acrescida de um for menor que EPSLON. Ver referência no îtem 4.
- TOLPIV Tolerância relativa de pivô, utilizada durante a decomposição LU do jacobiano. Ver referência no ítem 4.
- TOLABS Tolerância absoluto de pivô, utilizada durante a decomposição LU do jacobiano. Ver referência no îtem 4.
 - EME Crescimento relativo máximo de cada elemento da L ou da U, duran te uma atualização da fatorização do jacobiano.
- NITIVAX Número máximo de iterações permitido.
 - IMPR Parâmetro de controle de impressão : se =0 , não imprime nada ; se $\neq 0$, imprime a norma de $F(x_0)$ e diagnóstico de cada iteração multipla de \mid IMPR \mid .
 - SAIDA Unidade de saída para impressão dos diagnósticos.
 - QUAL Controle de atualização da fatorização da aproximação do jacobia no. Se =1, atualiza só a L; se =2, atualiza só a U; se =3, atualiza ambas, alternadamente.
 - FREQ Frequência de atualização da fatorização. Atualiza a fatorização nas iterações múltiplas de FREQ. O fator a ser atualizado depende do parâmetro QUAL.O parâmetro REFAT tem prioridade sobre o parâmetro FREQ.

- REFAT Frequência de refatorização. Se MOD(K+1,REFAT)=0, então a fatorização anterior $\tilde{\mathbf{e}}$ esquecida e realiza-se nova decomposição IU do jacobiano, antes de se calcular $\Delta \mathbf{x}$.
 - P Vetor especificando a matriz de permutação das linhas do jacobiano. Ver referência no îtem 4.
 - Q Vetor especificando a matriz de permutação das colunas do jacobiano. Ver referência no item 4.
 - F Subrotina externa fornecida pelo usuário, utilizada para calcu lar F(x).
- JACOB Subrotina externa fornecida pelo usuário, utilizada para calcu lar uma linha da matriz F'(x).
 - X Ponto inicial (na entrada) e melhor aproximação obtida pelo algoritmo (na saída).
- TRAB Vetor de trabalho de dimensão MEMOR.
- FXMIN Contêm os valores de F(X), onde X é a melhor aproximação obtida pelo algoritmo.
 - NIT Número de iterações realizadas.
- NTIMIN Número da iteração onde se conseguiu a melhor aproximação da solução.
- MUTILL Número de elementos armazenados na L.
- MUTILU Número de elementos armazenados na U.
- NFATLU Número de fatorizações LU realizadas.
- CODERR Código de erro :
 - 0 → significa convergência
 - 1 + indica erro nos parâmetros
 - 2 → indica singularidade no jacobiano
 - 3 → faltou memória aumentar MEMOR
 - 4 → excedeu o nº máximo de iterações NITMAX
 - 5 → significa divergência

Parametros rotinas - F , JACOB

Parâmetros inteiros - N,MEMOR,NTIMAX,IMPR,SAIDA,QUAL,FREQ,REFAT,P,Q,NIT,
NTIMIN,MUTILL,MUTILL,NFATILL,CODERR

Parâmetros reais - DISTX,DISTFX,EPSLON,TOLPIV,TOLABS,EME,TRAB,FXMIN,X
Parâmetros de entrada - N, ..., JACOB
Parâmetros de saída - FXMIN,NIT,NITMIN,MUTILL,MUTILU,NFATIU,CODERR
Parâmetros de entrada e saída - X
Parâmetros de trabalho - TRAB
Dimensões de parâmetros vetores - P(N),Q(N),X(N),TRAB(MEMOR),FXMIN(N)

3.-Subrotinas providas pelo usuário

A primeira subrotina é aquela especificada pelo parâmetro F , e tem como objetivo calcular as componentes ${\rm f}_{\frac{1}{2}}(x)$, para um ponto dado. Sua forma deve ser :

Ela será chamada em SNLESP que, portanto, fornecerá seus parametros de entrada N e X. Deve ser programada de tal maneira que a componente $f_i(x)$ seja colocada na posição FX(I).

A segunda subrotina especificada pelo parâmetro JACOB, tem como objetivo calcular uma linha da matriz jacobiana nun ponto X dado. Sua forma deve ser :

SUBROUTINE JACOB (N, LINHA, X, VALORS, COLS, NELEM)
INTEGER COLS (N)
REAL X (N), VALORS (N)
:
:
:
:
:
:
:
:
:

END

onde:

N - Dimensão do sistema

LINHA - Nº da linha (equação) do jacobiano desejada

X - Ponto dado

VALORS - Valores do jacobiano calculados nesta equação. Devem ser colocados apenas aqueles em que f_i(x_j) dependa de x_j (i=LINHA), ou seja, apenas as posições que possam assumir valores não nulos para algum ponto X dado. Estes valores fornecidos devem estar compactados dentro do vetor VALORS, em ordem crescente de coluna.

COLS - Índices das variáveis (colunas) dos valores de VALORS.

NELEM - Nº de elementos preenchidos em VALORS ou COLS.

Exemplo:

Seja o sistema (n>l) :
$$f_i(x) = (3-0.5x_i)x_i-x_{i-1}-2x_{i+1}+1 \text{ para } 1 < i < n$$

$$= (3-0.5x_i)x_i-x_{i-1}+1 \text{ para } i=n$$

$$= (3-0.5x_i)x_i-2x_{i+1}+1 \text{ para } i=1$$

neste caso o jacobiano (para n=3) fica:
$$\begin{pmatrix}
3-x_1 & -2 & 0 \\
-1 & 3-x_2 & -2 \\
0 & -1 & 3-x_3
\end{pmatrix}$$

as posições (i,j) para j>i+l e j<i-l serão sempre nulas.

Neste caso as subrotinas F e JACOB serão :

SUBROUTINE F (N,X,FX)

DIMENSION X (N), FX (N)

$$FX(I) = (3. - X(I)*0.5)*X(I) + 1.0$$
 $IF(I.NE.N) FX(I) = FX(I) - 2. * X(I+1)$
 $IF(I.NE.1) FX(I) = FX(I) - X(I-1)$

10 CONTINUE

RETURN

END)

```
SUBROUTINE JACOB (N, LINHA, X, VALORS, COLS, NELEM)
      INTEGER COLS (N), C(3)
      REAL X(N), VALORS(N), D(3)
      DATA D / -1. , 0. , -2. /
      D(2) = 3. - X(LINHA)
      C(1) = LINHA - 1
      C(2) = LINHA
      C(3) = LINHA + 1
      NELEM = 0
      DO 10 K=1,3
        IF (LINHA.EQ.1.AND.K.EQ.1.OR.LINHA.EQ.N.AND.K.EQ.3) GOTO 10
        NELEM = NELEM + 1
        COLS(NELEM) = C(K)
        VALORS(NELEM) = D(K)
10
      CONTINUE
      RETURN
      END
```

4.-Origem

Ver

T.L.Lopes - "Algoritmos para resolução de sistemas não lineares esparsos " - Tese de Mestrado - IMEXX - UNICAMP Orientador : J. M. Martinez

5.-Alternativas de métodos

São formecidos algumas combinações dos parâmetros QUAL, FREQ e REFAT e o método utilizado pelo programa em cada caso:

			· · · · · · · · · · · · · · · · · · ·	<u>}</u>
QUAL	FREQ	REFAT	METOXO	
qualquer	qualquer	1	Newton	
qualquer	∞	р	Newton c/ refina/	→ (p-1) subitera ções
1	1	α,	Dennis & Marwil (modificado)	
2	1	∞	Lopes (muda só a U)	
3	1	80	Lopes(muda L e U, alternada/)	
1	1	р	Dennis & Marwil com refatorizações a cada p iterações	

6.-Tamanho do programa

As rotinas utilizadas somam 799 linhas FORTRAN, das quais 197 são comentários. Ocupam 15K de memória no PDP-10 da UNICAMP, sem contar as dimensões de P,Q,X,FXMIN e TRAB.

7.-Precisão

Simples.

8.-Acesso à subrotina SNLESP

O código FORTRAN da SNLESP e das rotinas por ela utilizadas estão gravadas na fita MA127. Para jogá-las na sua área, os usuários do PDP-10 da UNICAMP devem dar o sequinte comando:

FILE R MA127 SNLESP.FOR 4

Este comando ordena ao operador que monte a fita MA127 e jogue o arquivo SMLESP.FOR na area do usuario. Assim que este arquivo apareça no diretório, o usuário pode dar o comando abaixo, para executar seu programa principal que está gravado no arquivo PRINC.FOR, por exemplo:

EXEC/F10 PRINC.FOR, SNLESP.FOR

Usuários externos à UNICAMP podem se comunicar com os responsáveis no endereço :

Laboratório de Matemática Aplicada IMECC - UNICAMP CP 1170 13100 CAMPINAS SP BRASIL

9.-Exemplo de programa principal

Seja o sistema especificado no îtem 3, para n=10. As subrotinas F e JACOB serão as mesmas descritas no îtem 3. Supondo-se $\mathbf{x_0} = (-1, \dots, -1)$, EPSION= 10^{-6} , QUAL=3, FREQ=1 e REFAT=1000 (método de Lopes, mudando a L e a U, alternadamente). Então, o programa principal poderia ser :

```
\mathbb{C}
                                                                                        FROM A RESERVED BY THE STATE OF STATE O
Ċ
                                                                                          INTERIOR STATE OF A STATE OF A STATE OF A STATE OF CASE
                                                                                        Sens A(t3), by Eller, by Coll. In
                                                                                        EXICCORD FY BELLER
                                                                                          * ≠ } ;
                                                                                        91513#157g
                                                                                        ijott/#1 ...
                                                                                        SPARSE TELL . The
                                                                                          100t 1V#3.5 %
                                                                                          CONSESSED ON
                                                                                        MM1 4 1 233
                                                                                         MITSHARITE
                                                                                        1MPB=1
                                                                                        Salibers
                                                                                        WULL#1
                                                                                        £R&@≇1
                                                                                        出展图 A T # 1 0 3
                                                                                        00 10 1=1, a
                                                                                        2(1)#1
                                                                                        U(l)=l
                                                                                        A(I)=-1.3
 13
                                                                                        CONTINUE
                                                                                      CRALL SHOULD HER $57, HEREN, HOTEN, HOLDEN, HOLDEN, FORTEV, FRANCISC, EIGH.

**EFTEN, ESTE, L. E. L., H. N., FRENCH, H. E., FRENCH, H. E., FRENCH, H. E. L., H. C. H., FRENCH, H. E., FREN
                                                                                                                                                                                                                               April Wage data Caralla and Ober at Cherry de Caleda
                                                                                         2
                                                                                                                                                                                                                                31 x 1 / (1) (1)
                                                                                         j
                                                                                        ARITE (20, 20) (18), 1000 to 1800 to 1
                                                                                      FORMAT('15,10); or similar : 1,7/,1 off = 1,11,/,1 off to 1,11,/,1 off to 1,11,/,1 off to 2,11,/,1 off to 2,11,//)
  2.4
                                                                                        48111 (3445) 1478 L.
                                                                                         更要被6百里(* 5. 高州) (6. 人) (2. 4. 5. 5) (6. 4. 4. 5. 5) (6. 4. 7. 5) (6. 4. 7. 5)
  36
                                                                                                                                                                       * F(A) = *,/,1087.5)
                                                                                          ï
                                                                                         STOR
                                                                                         \triangle N \, \mathcal{Q}
```

A salda (arquivo 20) seria:

```
7FXE 637 = 3.5 0000
IPERACAC T
FATORILACIA
/FX( 1)/ = 9.1010001
IPERACAU . ¿
MUDARCA DA L
//X( 2)/ = 0.115333304 1
ITERACAD - 3
MUDANIA DA G
             Complete (Superior of
78%( 3)/ = -
ITERACAD -
MODANIA DA L
/fX( 4)/ = 0.15 (/) (% (mile))
ITERAÇAD -
MIDANJA DA U
             - 6. t + 1 1273 - 53
78X( 5)/ = -
ITEPACAD: -
MUDANCA DA E
/FX( 6)/ = 0.545195967 /
SATOR OF SMILE :
is IT =
NITHIA =
            O.
            19
MUTILL =
ming Takk 🐃
            19
NIATEU =
            î
CJDLBB ≖
MELITOR & ENCHAPTER
-1.630 -1.310 -1.300 -1.000 -1.000 -1.000 -1.000 -1.000 -1.000
```

magnification against the second section of

€ (χ) ≠

, អាក្ស បុន្តសម ១១, ២០០០ 🔭

10.-Data

13 de maio de 1982-

X.X.X.X.X.X.X.