ALGORÍTMOS DE BROYDEN COMBINADOS PARA RESOLUÇÃO DE SISTEMAS DE EQUAÇÕES NÃO-LINEARES

LUIZ SATORU OCHI

Orientador:
Prof. Dr. JOSÉ MÁRIO MARTÍNEZ

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

Maio/1981

AGRADECIMENTOS

- Ao Martínez, pela proposta deste trabalho, pelo apoio e dedicação demonstrados.

- Aos amigos que, de uma forma ou de outra, me ajudaram a realizar este trabalho.

- Aos Professores do IPEA de Presidente Prudente, pelo apoio e incentivo.

- Aos meus pais, pelo exemplo vivo que sempre representaram para mim.

- À CAPES e à FAPESP, pelo suporte financeiro.

- À Célia, pelo excelente trabalho de datilografia.

ÍNDICE

CAPÍTULO I

A CONVERGÊNCIA LOCAL E Q-SUPERLINEAR DOS MÉTODOS QUASE-NEWTON

1) INTRODUÇÃO

Neste capítulo damos uma síntese dos principais resultados de convergência dos métodos Quase-Newton.

Introduzida por Broyden (1965), esta família de algorítmos tem por finalidade solucionar sistemas de equações não-lineares; i.é., dado $f: R^n \to R^n$ diferenciável encontrar um vetor - $x^* \in R^n$ tal que $f(x^*) = 0$.

Broyden [3] propôs uma modificação no Método de Newton, substituindo $f'(x)$ por uma aproximação B. Este algorítmo é conhecido como Método Quase-Newton cuja iteração na sua forma geral é: $x_{k+1} = x_k - \lambda_k B_k^{-1} f(x_k)$ (1.1).

O parâmetro λ_k é determinado de tal forma que os resíduos provocados pela sequência $\{B_k^{-1}\}$ seja reduzido e a convergência obtida mais rapidamente.

Particularmente, em problemas de minimização o escalar λ_k muitas vezes é computado para aproximar-se de λ_k^*, o valor que minimiza $f(x_k - tH_k Fx_k)$, com t maior que zero e onde F é o gradiente de um funcional não-linear f.

Quando $\lambda_k = \lambda_k^*$ a cada iteração, o método é conhecido como "Método Perfeito", que se caracteriza pela sua "terminação quadrática"; i.é., quando um algorítmo é aplicado a um funcional quadrático e convexo, a solução "exata" é obtida após um número finito de passos.

Outra família de algorítmos, onde $\lambda_k = 1$ para todo passo k são conhecidos como "Métodos de Predição-Direta".

2) APRESENTAÇÃO DOS ALGORÍTMOS

Considere $f : R^n \to R^n$ onde R^n é o espaço real, linear n-dimensional.

O nosso objetivo é encontrar um vetor $x^* \in D \subset R^n$ tal que $f(x^*) = 0$.

A primeira classe de algorítmos que analisaremos são os métodos adaptados de posto um. Nestes métodos as fórmulas de iteração são

$$x_{k+1} = x_k - B_k^{-1} f(x_k) \tag{2.1}$$

onde B_k é obtida adicionando-se uma matriz de posto um à matriz B_{k-1}.

Precisamente, a sequência B_k é

$$B_{k+1} = B_k + (y_k - B_k s_k) \frac{c_k^T}{c_k^T s_k} \tag{2.2}$$

onde

$$s_k = \Delta x_k = x_{k+1} - x_k \tag{2.3}$$

$$y_k = \Delta f_k = f(x_{k+1}) - f(x_k) \tag{2.4}$$

Do ponto de vista computacional é conveniente trabalharmos com $H_k = B_k^{-1}$, a qual pode ser obtida aplicando a fórmula de Sherman-Morrison [8] a (2.2). Assim

$$H_{k+1} = H_k + (s_k - H_k\, y_k)\, \frac{d_k^T}{d_k^T\, y_k} \qquad (2.5)$$

e $d_k = H_k^T c_k$. Entretanto, para estudos teóricos de algorítmos, é preferível usar (2.2) ao invés de (2.5).

Quando estivermos tratando de um algorítmo específico, o valor de c_k ou d_k será determinado, o que não é necessário para nosso atual propósito. Aqui os $B_{k+1} \in U(x_k, B_k)$ onde $U : R^n \times L(R^n) \to P\{L(R^n)\}$ é uma função atualizadora, $L(R^n)$ denota o conjunto das matrizes reais $(n \times n)$ e $P\{L(R^n)\}$ o conjunto de todos os subespaços não-vazios de $L(R^n)$.

Considere $\bar{B} = B + (y - Bs)\, \dfrac{c^T}{c^T s}$, então (2.1) e (2.2) podem ser escritas como

$$x_{k+1} = x_k - B_k^{-1}\, f(x_k)$$

com $B_{k+1} \in U(x_k; B_k)$ e $U(x,B) = \{\bar{B} : c^T s \neq 0\}$.

Neste caso, $s = \bar{x} - x$ e $y = f(\bar{x}) - f(x)$, onde $\bar{x} = x - B^{-1} f(x)$.

Para uma melhor visualização, consideremos o primeiro método de Broyden (1965) onde $c = s$ e $d = H^T s$. Neste caso a função atualizadora U é dada por

$$U(x,B) = \{\bar{B} : s \neq 0\} \quad \text{onde} \quad \bar{B} = B+(y-Bs)\,\frac{s^T}{s^T s}$$

Note que U como definido acima não leva em conta os erros de arredondamento. Considerando este item, U é definido da forma

$$U(x,B) = \{\bar{B} + E \;:\; \|E\| \leq \|s\| \;;\; s \neq 0\} \;.$$

Nossa atenção estará voltada também a outros algoritmos atualizados de posto-um como o segundo Método de Broyden ($c = B^T y$, $d = y$), o Método de Pearson (1969) ($c = y$, $d = H^T y$) e o Método de McCormick (Pearson 1969) ($c = B^T s$, $d = s$), definindo em cada caso, uma função atualizadora.

McCormick e Pearson são métodos implementados comumente na forma $x_{k+1} = x_k - (B_k^T)^{-1} f(x_k)$, onde $B_{k+1} \in U(x_k, B_k)$ originando neste caso a propriedade da terminação quadrática para certos casos.

DEFINIÇÃO 1: Uma sequência $\{x_k\}$ converge para um vetor x^* Q-superlinearmente se existe uma sequência real $\{\alpha_k\}$ convergindo para zero e $\|x_{k+1} - x^*\| \leq \alpha_k \|x_k - x^*\| \;;\; k \geq 0$.

DEFINIÇÃO 2: Dizemos que $\{x_k\}$ converge para x^* R-superlinearmente, se o $\lim\limits_{k \to \infty} \|x_k - x^*\|^{1/k} = 0$.

Observe que Q-superlinear implica em R-superlinear mas não valendo a recíproca.

De fato, como $\alpha_k \xrightarrow{k \to \infty} 0$, $\exists\, M : 0 < M < 1$ e um k_0 tal que $\forall k > k_0$, $|\alpha_k| < M$.

Como $\|x_{k+1} - x^*\| \leq M^k x \|x_0 - x^*\|$ e $M^k \xrightarrow{k \to \infty} 0$ implica que $\|x_{k+1} - x^*\| \to 0 \Rightarrow \|x_k - x^*\|^{1/k} \to 0$ quando $x \to \infty$.

Será mostrado adiante que todos os métodos citados acima, possuem convergência local e Q-superlinear. Por exemplo, se os $B_{k+1} \in U(x_k, B_k)$ são as aproximações de Broyden ou Pearson - com adaptações de posto-um, uma iteração do tipo

$$x_{k+1} = x_k - \left[t_k B_k + (1-t_k)\ B_k^T\right]^{-1} f(x_k)$$

possui convergência local e superlinear desde que $f'(x^*)$ seja simétrica e positiva definida.

Posteriormente David Gay [6] nos mostrou que tais - restrições não são necessárias para os métodos de Broyden como veremos no próximo capítulo.

Finalmente, como estamos interessados em teoremas de convergência local, as funções adaptadas necessitam apenas serem definidas numa vizinhança de $(x^*, F'(x^*)^{-1})$. Entretanto, o primeiro método de Broyden está definido somente para um $B \in L(R^n)$ não-singular e um vetor x no domínio de f, para o qual

$$s = -B^{-1} f(x) \neq 0 \quad \text{e} \quad \bar{x} = x + s$$

situa-se no domínio de f. Assim é necessário mostrar que existe uma vizinhança de $(x^*, f'(x^*)^{-1})$ onde U esteja definida.

Felizmente o problema é simples, desde que as hipóteses sobre f garantem que existe uma vizinhança N de $(x^*, f'(x^*)^{-1})$ tal que, $(x,B) \in N \Rightarrow B$ é não-singular, x e $\bar{x}$ situam-se no domínio de f e $s = 0 \Leftrightarrow x = x^*$. Portanto, U pode ser definida arbi trariamente para $x \neq x^*$ onde, se $x = x^*$, o algorítmo pára.

Definimos uma função U conveniente (Veja Teor. 3.2), $U(x^*,B) = \{f'(x^*)\}$ para algum $B \in L(R^n)$ com $(x^*,B) \in N$.

Como para o método de Broyden existem aplicações semelhantes para os outros algorítmos.

Observe que, quando falamos numa vizinhanca de $(x^*, F'(x^*))$, provamos que de fato ela existe e supomos a função U definida nesta vizinhança.

Outra classe de algorítmos analisados são os métodos atualizados de posto-dois, sendo mais comumente aplicados aos métodos de Pearson e McCormick para minimizar funções $f : R^n \to R$.

Neste caso, são aplicados para $F(x) = \nabla f(x)$. Se F' é contínua, então o Hessiano $F'(x)$ é necessariamente simétrico e por esta razão as matrizes adaptadas nos métodos de posto-dois são escolhidas simétricas.

Ainda neste capítulo estudaremos estes métodos de posto-dois.

Consideramos $B \in L(R^n)$, façamos $B^{(0)} = B$. Atualizamos $B^{(0)}$ pela fórmula (2.2) e teremos

$$\bar{B}^{(t)} = B^{(0)} + (y-B^{(0)}s)\,\frac{c^T}{c^T s}$$

A matriz $\bar{B}^{(1)}$ pode não ser simétrica; assim façamos

$$B^{(1)} = \left[\bar{B}^{(1)} + \left[\bar{B}^{(1)}\right]^T\right] \Big/ 2$$

Entretanto, $B^{(1)}$ pode não satisfazer as equações - quase-Newton, se repetirmos o processo, é necessário que a sequência $\{B^{(k)}\}$ tenha limite

$$\bar{B} = B + \frac{(y-Bs)c^T + c(y-Bs)^T}{c^Ts} - \frac{s^T(y-Bs)cc^T}{(c^Ts)^2} \qquad (2.7)$$

Modificando (2.7), como visto por Powell (1970) para o caso particular $c = s$, Dennis (1972) calculou o caso geral - concluindo ser este processo obtido de (2.5) modificado, chegando a

$$\bar{H} = H + \frac{(s-Hy)d^T + d(s-Hy)^T}{d^Ty} - \frac{y^T(s-Hy)dd^T}{(d^Ty)^2} \qquad (2.8)$$

Note que (2.7) e (2.8) englobam vários métodos e que geralmente $\bar{B}\bar{H} \neq I$, sempre que $BH = I$ e $d = H^Tc$ de forma que (2.5) e (2.7) sofram as mesmas modificações.

Para $c = s$ resulta no Powell Simétrico, no algorítmo de Broyden (Powell 1970), e $c = y$ corresponde ao algorítmo DFP (Davidon - Fletcher - Powell) [4], $d = y$ em (2.8) obtemos um dos métodos de Greenstadt (1970), embora não o comumente associado ao seu nome, $d = s$ resulta no algorítmo BFGS (Broyden , Fletcher (1970), Goldfarb (1970), Shanno (1970)).

Fica como observação que o subespaço $\frac{n(n+1)}{2}$-dimensional de $L(R^n)$ consiste na família de todas as matrizes simétricas de $L(R^n)$. Assim quando nos referirmos a uma vizinhança de $f'(x^*)$ ou de $f'(x^*)^{-1}$, será vizinhança com respeito a este sub espaço de $L(R^n)$.

3) ALGUNS RESULTADOS DE CONVERGÊNCIA

Apresentamos nesta secção, alguns teoremas de convergência local aplicados aos métodos considerados na secção anterior.

Nesta secção sempre que nos referirmos à função f estaremos admitindo $f : R^n \to R^n$ diferenciável em $D \subset R^n$.

LEMA 3.1: Seja $f : R^n \to R^n$.

Vamos supor que para um $x^* \in D$ e $r > 0$

$$\|f'(x)-f'(x^*)\| \leq k\|x-x^*\|^r \qquad k \geq 0 \tag{3.1}$$

Então para cada y e z em D

$$\|fy - fz-f'(x^*)(y-z)\| \leq k \text{ máx } \{\|y-x^*\|^r, \|z-x^*\|^r\} \|y-z\| \tag{3.2}$$

Além disso se $f'(x^*)$ é não singular existe $\varepsilon > 0$ e $\rho > 0$ tal que $\max\{\|y-x^*\|, \|z-x^*\|\} \leq \varepsilon$ implica que y e z $\in$ D e

$$\left(\frac{1}{\rho}\right) \|y-z\| \leq \|fy-fz\| \leq \rho \|y-z\| \tag{3.3}$$

DEMONSTRAÇÃO: Para mostrar (3.2) considere um vetor $x^* \in D$.

Vamos definir a aplicação $G : D \subset R^n \to R^n$, $z \to fz-f'(x^*)z$ diferenciável em D.

Afirmo que $\forall x,y \in D$

$$\|fy-fx\| \leq \sup_{0\leq t\leq 1} \|f'(x+t(y-x)\| \ \|x-y\| \quad (*)$$

De fato, se $M = \sup_{0\leq t\leq 1} \|f'(x+t(y-x)\| < \infty$, $\forall \varepsilon > 0$ e S o conjunto dos $t \in [0,1]$ satisfazendo

$$\|f(x+t(y-x)-fx\| \leq Mt\ \|y-x\| + \varepsilon t\ \|y-x\| \quad (**)$$

Naturalmente $0 \in S$, assim $\gamma = \sup_{t\in S} t$ está bem definido e da continuidade de $f(x+t(y-x))$ em relação a t, temos:

$$\|f(x+\gamma(y-x))-fx\| \leq M\gamma\ \|y-x\|+\varepsilon\gamma\ \|y-x\| \quad (***)$$

Como ε é arbitrário, resta-nos mostrar que (***) vale somente para $\gamma = 1$.

Suponhamos que $\gamma<1$, como existe f' em $x+\gamma(y-x)$ podemos tomar um $\beta \in (\gamma,1)$ tal que:

$$\|f(x+\beta(y-x)) - f(x+\gamma(y-x) - f'(x+\gamma(y-x))(\beta-\gamma)(y-x)\| \leq \varepsilon\ (\beta-y)\|y-x\|$$

Portanto,

$$\|f(x+\beta(y-x)) - f(x+\gamma(y-x)\| \leq M(\beta-\gamma)\|y-x\|+\varepsilon(\beta-\gamma)\|y-x\|$$

Mas por (***) temos

$$\|f(x+\beta(y-x))-f(x)\| \leq (M\gamma+\varepsilon\gamma)\|y-x\|+(M+\varepsilon)(\beta+\gamma)\|y-x\| = (M+\varepsilon)\beta\|y-x\|$$

Isto é, (**) vale para $1 < \beta < \gamma$, contradição com a definição de γ.

Assim, $\|fy-fx\| \leq \sup\limits_{0\leq t\leq 1} \|f'(x+t(y-x)\| \; \|x-y\|$

Da definição de G, temos $G'(z) = f'(z)-f'(x^*)$, portanto (*) nos dá

$$\|G(y)-G(z)\| \leq \sup_{0\leq t\leq 1} \|G'(z+t(y-z))\| \; \|y-z\| \qquad \text{ou}$$

$$\|fy-fz-f'(x)(y-z)\| \leq \sup_{0\leq t\leq 1} \|f'(z+t(y-z))-f'(x^*)\| \; \|y-z\|$$

Em particular para $t=0$ e $t=1$ e hipótese do Lema temos

$$\|fy-fz-f'(x)(y-t)\| \leq k \text{ máx } \{ \|z-x\|^r, \|y-x\|^r \; \|y-z\|$$

provando (3.2).

(3.3) é uma consequência imediata da continuidade e não-singularidade de f' em x^*.

NOTA: Utilizamos uma norma matricial $\|.\|M$ e uma norma vetorial $\|.\|$, entretanto como as normas num espaço vetorial finito são equivalentes existe uma constante $\eta > 0$, tal que $\|A\| \leq \eta \|A\|_M$.

TEOREMA 3.2: Considere $f : R^n \to R^n$.

Suponha que para algum $x^* \in D$ e $\rho > 0$, a desigualdade (3.1) é verdadeira, $f(x^*) = 0$ e $f'(x^*)$ não singular.

Defina $U : R^n \times L(R^n) \to P\{L(R^n)\}$, com domínio numa vizinhança $N = N_1 \times N_2$ de $(x^*, f'(x^*))$ onde $N_1 \subset D$ e N_2 contém só matrizes não-singulares.

Se existem constantes $\alpha_1, \alpha_2 \geq 0$ tais que $\forall\ (x,B)$ em N e $\bar{x} = x - B^{-1}fx$, a função U satisfaz

$$\|\bar{B}-f'(x^*)\|_M \leq \left[1+\alpha_1 \, \text{máx}\, \{\|\bar{x}-x^*\|^P\}\right] \|B-f'(x^*)\|_M +$$

$$+ \alpha_2 \, \text{máx}\, \{\|\bar{x}-x^*\|^P, \ \|x-x^*\|^P\}; \quad \forall\, \bar{B} \in U(x,B) \tag{3.5}$$

Então para todo $r \in (0,1)$, existem constantes positivas $\varepsilon(r)$ e $\gamma(r)$ tais que para $\|x^0-x^*\| < \varepsilon(r)$, $\|B_0-f'(x^*)\|_M < \gamma(r)$ e todo $B_{k+1} \in U(x_k,B_k)$, a sequência $x_{k+1} = x_k - B_k^{-1}f(x_k)$ está bem definida e converge para x^* .

Além disso,

$$\|x_{k+1}-x^*\| \leq r\|x_k-x^*\| \quad \forall\ k \geq 0 \tag{3.6}$$

e $\{\|B_k\|\}$, $\{\|B_k^{-1}\|\}$ são uniformemente limitadas.

DEMONSTRAÇÃO: Tome $r \in (0,1)$

Considere o conjunto dos γ tais que $\gamma \geq \|f'(x^*)^{-1}\|$.

Escolha $\gamma(r) = \gamma$, $\varepsilon(r) = \varepsilon$ tal que

$$\left[2\ \alpha_1\ \delta + \alpha_2\right] \frac{\varepsilon^P}{1-r^P} \leq \delta \tag{3.7}$$

e para η definido em (3.4);

$$\gamma(1-r)\left[k\ \varepsilon^P + 2\eta\delta\right] \leq r \tag{3.8}$$

Vamos restringir ε e γ tais que $(x,B) \in N$ e $\|B-f'(x^*)\|_M < 2\delta$ e $\|x-x^*\| < \varepsilon$.

Seja $\|B_0-f'(x^*)\|_M < \delta$ e $\|x_0-x^*\| < \varepsilon$, isto implica que $\|B_0-f'(x^*)\| < \eta\delta < 2\eta\delta$ e por (3.8) temos

$$2\gamma(1-r)\eta\delta \leq r \tag{3.9}$$

Ainda $\|B_0^{-1}\| \leq (1+r)\gamma$ (pelo Lema de Perturbação (Banach), ver Ortega e Rheinbolt, 1970, pág. 45) .

O Lema 3.1 nos dá:

$$\|x_1-x^*\| \leq \|B_0^{-1}\| \left[\|f(x_0)-f(x^*)-f'(x^*)(x_0-x^*)\|+\|B_0+f'(x^*)\| \right.\cdot$$

$$\left. \cdot \|x_0-x^*\| \right] \leq \gamma(1+r) \left[k\varepsilon^P+2\eta\delta\right] \|x_0-x^*\| .$$

Por (3.8) segue-se que:

$$\|x_1-x^*\| \leq r\|x_0-x^*\|$$

portanto $\|x_1-x^*\| < \varepsilon$ e $x_1 \in D$.

Completaremos a demonstração por indução; assumimos que $\|B_k-f'(x^*)\|_M \leq 2\delta$ e $\|x_{k+1}-x^*\| \leq r\|x_k-x^*\|$ para $k = 0,1,\dots,m-1$ e provaremos para $k = m$.

De (3.5) temos que:

$$\|B_{k+1}-f'(x^*)\|_M - \|B_k-f'(x^*)\|_M \leq 2\alpha_1\delta\varepsilon^P r^{kP} + \alpha_2\varepsilon^P r^{kP}$$

para $k = 0,1,\dots,m-1$. Assim

$$\|B_m - f'(x^*)\|_M \leq \|B_0 - f'(x^*) + \left[2\alpha_1\delta + \alpha_2\right] \frac{\varepsilon^p}{1-r^p}$$

e por (3.7)

$$\|B_m - f'(x^*)\|_M \leq 2\delta \ .$$

Resta-nos mostrar que $\|x_{m+1} - x^*\| \leq r\|x_m - x^*\|$. Mas isto segue de um argumento semelhante ao utilizado para $m = 1$. De fato, desde que $\|B_m - f'(x^*)\| \leq 2\eta\delta$, o lema de Banach e (3.9) implicam que $\|B_m^{-1}\| \leq (1-r)\gamma$, i.ē., que $\{\|B_m^{-1}\|\}$ e $\{\|B_m\|\}$ são uniformemente unificados.

Pelo Lema 3.1

$$\|x_{m+1} - x^*\| \leq \|B_m^{-1}\| \left[\|f(x_m) - f(x^*) - f'(x^*)(x_m - x^*)\| + \right.$$

$$\left. + \|B_m - f'(x^*)\| \|x_m - x^*\|\right] \leq \gamma(1+r) \left[k\varepsilon^P + 2\eta\delta\right] \|x_m - x^*\| \ .$$

Portanto

$$\|x_{m+1} - x^*\| \leq r\|x_m - x^*\| \qquad \text{cqd//} \ .$$

O teorema 3.2 mostra resultados de convergência da iteração (2.1), mas não dá uma estimativa do seu raio de convergência, mostrado no seguinte corolário.

COROLÁRIO 3.3: Supondo satisfeitas as hipóteses do teorema 3.2 e além disso se alguma subsequência de $\{\|B_k - f'(x^*)\|_M\}$ converge para zero, então $\{x_k\}$ converge Q-superlinearmente para x^*.

DEMONSTRAÇÃO: Devemos mostrar que $\lim_{k\to\infty} \dfrac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|} = 0$.

Seja $r \in (0,1)$.

Pelo teorema 3.2 existem números $\varepsilon(r)$ e $\delta(r)$ tais que se $\|B_0-f'(x^*)\|_M < \delta(r)$ e $\|x_0-x^*\| < \varepsilon(r)$ então

$$\|x_{k+1}-x^*\| \leq r\|x_k-x^*\| \quad \forall\, k \geq 0 .$$

Além disso, pela hipótese do corolário, existe $m > 0$ tal que $\|B_m-f'(x^*)\| < \delta(r)$ e $\|x_m-x^*\| < \varepsilon(r)$.

$$\overset{\text{Teor. } 3.2}{\Longrightarrow} \quad \|x_{k+1}-x^*\| \leq r\|x_k-x^*\| \quad \forall k \geq m \ \text{ e } \ \forall r \in (0,1) \ \text{cqd//} .$$

O corolário acima mostra que é necessária a existência de uma subsequência de $\{\|B_k-f'(x^*)\|_M\}$ convergindo para zero para se obter uma convergência Q-superlinear de $\{x_k\}$.

De fato, recordemos que a iteração (2.1) satisfaz a equação (3.5) para $\alpha_1 = \alpha_2 = 0$, isto é,

$$\|B_{k+1}-f'(x^*)\|_M \leq \|B_k-f'(x^*)\|_M \quad \forall\, k \geq 0 \quad ,$$

mas obtendo em geral uma convergência apenas linear para $\{x_k\}$.

Tomando $H_k = B_k^{-1}$, obtemos um resultado análogo ao do teorema 3.2.

TEOREMA 3.4: Seja $f : R^n \to R^n$ satisfazendo as hipóteses do teorema 3.2.

Vamos definir $G : R^n \times L(R^n) \to P\{L(R^n)\}$ numa vizi nhança $N = N_1 \times N_2$ de $(x^*, f'(x^*)^{-1})$ onde $N_1 \subset D$.

Se existem constantes não negativas ξ_1 e ξ_2 tais que para $\forall (x,B) \in N$ e $\bar{x} = x - Hf(x)$, a função G satisfaz:

$$\|\bar{H}-f'(x^*)^{-1}\|_M \leq \left[1+\xi_1 \max\{\|\bar{x}-x^*\|^P, \|x-x^*\|^P\} \|H-f'(x^*)^{-1}\|_M + \right.$$
$$\left. + \xi_2 \max\{\|\bar{x}-x^*\|^P, \|x-x^*\|^P\}\right] \quad \forall \bar{H} \in G(x,H) \qquad (3.10)$$

Então, $\forall r \in (0,1)$ existem constantes $\varepsilon(r) > 0$ e $\gamma(r) > 0$ tais que $\|x^0-x^*\| < \varepsilon(r)$, $\|H_0-f'(x^*)^{-1}\|_M < \delta(r)$ e $\forall H_{k+1} \in G(x_k,H_k)$, a sequência $x_{k+1} = x_k - H_k f(x_k)$ está bem defi-nida e converge para x^* .

Além disso,

$$\|x_{k+1}-x^*\| \leq r\|x_k-x^*\| \quad \forall k \geq 0 \qquad (3.11)$$

e $\{\|H_k\|\}$ e $\{\|H_k^{-1}\|\}$ são uniformemente limitadas.

DEMONSTRAÇÃO: Considere $r \in (0,1)$; $\sigma \geq \|f'(x^*)\|$ e o conjunto dos γ: $\gamma \geq \|f'(x^*)^{-1}\|$.

Escolha $\delta(r) = \delta$ e $\varepsilon(r) = \varepsilon$, tais que:

$$\left[2\xi_1\delta + \xi_2\right] \frac{\varepsilon^P}{1-r^P} \leq \delta \qquad (3.12)$$

e para η dado em (3.4)

$$2\sigma\delta\eta + (\gamma+2\eta\delta)k\varepsilon^P \leq r \qquad (3.13)$$

Além disso, impomos que ε e δ atendam a condição

$\|H-f'(x^*)^{-1}\|_M < 2\delta$ e $\|x-x^*\| < \varepsilon$ para $(x,H) \in N$.

Assim, se $\|x_0-x^*\| < \varepsilon$ e $\|H_0-f'(x^*)^{-1}\|_M < \delta$ =>
=> $\|H_0-f'(x^*)^{-1}\|_M < \eta\delta < 2\eta\delta$ e como (3.12) está satisfeita,

$$2\,\sigma\,\delta\,\eta \leq r \tag{3.14}$$

Pelo teorema de Banach; $\|H_0^{-1}\| < \frac{\sigma}{1-r}$

Portanto ;

$$x_1-x^* = -H_0\left[f(x_0)-f(x^*)-f'(x^*)(x_0-x^*)\right] + \left[I+H_0f'(x^*)\right](x_0-x^*)$$

e pelo lema 3.1

$$\|x_1-x^*\| \leq \left[\|H_0\|\,k\,\varepsilon^P + 2\sigma\delta\eta\,\|x_0-x^*\|\right]$$

Mas

$$\|H_0\| \leq \|H_0-f'(x^*)^{-1}\|_M + \|f'(x^*)^{-1}\| \leq 2\eta\delta + \gamma$$

$$\overset{(3.13)}{\Rightarrow} \quad \|x_1-x^*\| \leq r\|x_0-x^*\| < \varepsilon \quad \text{e} \quad x_1 \in D$$

Assumindo que $\|H_k-f'(x^*)^{-1}\|_M \leq 2\delta$ e $\|x_{k+1}-x^*\| \leq$
$\leq r\|x_k-x^*\|$ para $k = 0,1,\ldots,m-1$, vamos mostrar que a desigualdade vale para $k = m$.

De (3.10) temos :

$$\|H_{k+1}-f'(x^*)^{-1}\|_M - \|H_k-f'(x^*)^{-1}\|_M \leq 2\xi_1\delta\varepsilon^P r^{kP} + \xi_2\varepsilon^P r^{kP}$$

i.é.,

$$\|H_m-f'(x^*)^{-1}\|_M \leq \|H_0-f'(x^*)^{-1}\|_M + \left[2\xi_1\delta+\xi_2\right]\frac{\varepsilon^P}{1-r^P}$$

e em virtude de (3.12)

$$\|H_m - f'(x^*)^{-1}\|_M \leq 2\delta$$

e de modo semelhante ao caso $m = 1$

$$\|H_m - f'(x^*)^{-1}\| \leq 2\eta\delta \qquad \text{e} \qquad \|H_m^{-1}\| \leq \frac{\sigma}{1-r}$$

pelo lema de Banach e por (3.14).

Com isto mostramos que $\{\|H_m^{-1}\|\}$ é uniformemente limitada.

Pelo Lema 3.1

$$\|x_{m+1} - x^*\| \leq \left[\|H_m\|\, k\varepsilon^p + 2\sigma\delta\eta\right] \|x_m - x^*\|$$

e

$$\|H_m\| \leq \|H_m - f'(x^*)^{-1}\| + \|f'(x^*)^{-1}\| \leq 2\eta\delta + \gamma \quad \Rightarrow$$

$\{\|H_m\|\}$ é também uniformemente limitada e

$$\|x_{m+1} - x^*\| \leq r\|x_m - x^*\| \qquad \text{cqd} \ //$$

COROLÁRIO 3.5: Se as hipóteses do teorema 3.4 estão satisfeitas e se existe uma sequência de $\{\|H_m - f'(x^*)^{-1}\|_M\}$ convergindo para zero, então $\{x_k\}$ converge Q-superlinearmente para x^*.

DEMONSTRAÇÃO: Semelhante à demonstração do Corolário 3.3.

4) OS MÉTODOS DE POSTO-UM

Mostraremos nesta secção que as hipóteses do teorema 3.2 ou 3.4 vistos anteriormente, são satisfeitas nos métodos de posto-um.

De princípio necessitamos dos seguintes resultados abaixo:

LEMA 4.1: Dado $B \in L(R^n)$ e os vetores y, c e $s \in R^n$ com $c^T s \neq 0$, definimos

$$\bar{B} = B + \frac{(y-Bs)\ c^T}{c^T s} \tag{4.1}$$

Então para todo $A \in L(R^n)$ temos que :

$$\bar{E} = E\left[I - \frac{M^{-1}s(Mc)^T}{c^T s}\right] + \frac{M(y-As)(Mc)^T}{c^T s} \tag{4.2}$$

onde $\bar{E} = M(\bar{B}-A)M$ e $E = M(B-A)M$.

Considerando $A = f'(x^*)$ em (4.2) e $\|.\|$ uma norma matricial, vamos definir uma outra norma $\|.\|_M$ pela relação $\|Q\|_M = \|MQM\|$.

A igualdade (4.2) mostra que $\|\bar{B}-f'(x^*)\|_M$ se relaciona com $\|B-f'(x^*)\|_M$.

Para mostrar que esta relação satisfaz a condição (3.5) do teorema 3.2, vamos escrever a Condição (4.2) como

$$\bar{E} = E\left[I - \frac{M^{-1}s(M^{-1}s)^T}{c^T s}\right] + E\left[\frac{M^{-1}s(M^{-1}s-Mc)^T}{c^T s}\right] +$$

$$+ \frac{M(y-As)(Mc)^T}{c^T s} \tag{4.3}$$

LEMA 4.2: Seja $M \in L(R^n)$ uma matriz simétrica, não-singular tal que

$$\|Mc-M^{-1}s\| \leq \beta\|M^{-1}s\| \tag{4.4}$$

para algum $\beta \in [0, 1/3]$ e c, s vetores de R^n com $s \neq 0$.

Então:

a) $(1-\beta)\|M^{-1}s\|^2 \leq c^T s \leq (1+\beta)\|M^{-1}s\|^2$,

b) $\forall$ E não-nulo em $L(R^n)$

$$\left\| E\left[I - \frac{(M^{-1}s)(M^{-1}s)^T}{c^T s}\right]\right\|_F \leq \sqrt{1-\alpha\theta^2}\ \|E\|_F$$

($\|.\|$ é a norma de frobenius).

c) $$\left\| E\left[\frac{I-M^{-1}s(Mc)^T}{c^T s}\right]\right\|_F \leq \left[\sqrt{1-\alpha\theta^2}+(1-\beta)^{-1}\frac{\|Mc-M^{-1}s\|}{\|M^{-1}s\|}\right]\|E\|_F$$

onde

$$\alpha = \frac{1-2\beta}{1-\beta^2} \in |3/8, 1| \tag{4.5}$$

e

$$\theta = \frac{\|EM^{-1}s\|}{\|E\|_F\ \|M^{-1}s\|} \in |0, 1| \tag{4.6}$$

d) $$\left\| \frac{(y-As)(Mc)^T}{c^T s}\right\|_F \leq \frac{2\|y-As\|}{\|M^{-1}s\|}$$

DEMONSTRAÇÃO:

<u>Solução a)</u> Note que $c^T s = (Mc)^T(M^{-1}s) = (Mc-M^{-1}s)^T M^{-1}s + \|M^{-1}s\|^2$

e pela desigualdade de Cauchy-Shwartz e Condição (4.4):

$$\|(Mc-M^{-1}s)^T M^{-1}s\| \leq \beta \|M^{-1}s\|^2 , \qquad (*)$$

Assim ;

$$|c^Ts| \leq \|(Mc-M^{-1}s)^T M^{-1}s\| + \|M^{-1}s\|^2 \leq \beta\|M^{-1}s\|^2 + \|M^{-1}s\|^2 =$$

$$= (1+\beta)\ \|M^{-1}s\|^2$$

$$|c^Ts| \geq \|M^{-1}s\|^2 - \|(Mc-M^{-1}s)^T M^{-1}s\| \overset{(*)}{\geq} \|M^{-1}s\|^2 - \beta\|M^{-1}s\|^2 =$$

$$= (1-\beta)\|M^{-1}s\|^2$$

Portanto ;

$$(1-\beta)\|M^{-1}s\|^2 \leq c^Ts \leq (1+\beta)\|M^{-1}s\|^2$$

<u>Solução b)</u> Note que o item (a) implica $c^Ts > 0$. Como $\|Q\|_F^2$ é o traço de (Q^TQ), através de cálculos diretos (ou pelo lema 4 , (Broyden 1970)), temos:

$$\left\| E\left[I-uv^T\right]\right\|_F^2 = \|E\|_F^2 - 2v^TE^TEu + \|Eu\|^2\ \|v\|^2$$

isto é ,

$$\left\| E\left[I - \frac{(M^{-1}s)(M^{-1}s)^T}{c^Ts}\right]\right\|_F^2 = \|E\|_F^2 + (-2c^Ts + \|M^{-1}s\|^2)\ \frac{\|EM^{-1}s\|^2}{(c^Ts)^2}$$

e em virtude do item (a)

$$\left\| E\left[I - \frac{(M^{-1}s)(M^{-1}s)^T}{c^Ts}\right]\right\|_F^2 \leq \|E\|_F^2 - \left(\frac{1-2\beta}{1-\beta}\right)\frac{\|EM^{-1}s\|^2}{c^Ts} \leq$$

$$\leq \quad \|E\|_F^2 - \alpha\left(\frac{\|EM^{-1}s\|^2}{\|M^{-1}s\|}\right)^2 = \quad \|E\|_F^2 - \alpha\left(\frac{\|E\|_F \quad \|EM^{-1}s\|^2}{\|E\|_F \quad \|M^{-1}s\|}\right)^2$$

Portanto :

$$\left\| E\left[I - \frac{(M^{-1}s)(M^{-1}s)^T}{c^Ts}\right]\right\|_F \quad \leq \quad \|E\|_F \sqrt{1-\alpha\theta^2}$$

<u>Solução c)</u> Para provar (c) basta utilizarmos o item (b) e mostrar que :

$$\left\| E\left[\frac{M^{-1}s(M^{-1}s-Mc)^T}{c^Ts}\right]\right\|_F \quad \leq \quad (1-\beta)^{-1}\left(\frac{\|Mc-M^{-1}s\|}{\|M^{-1}s\|}\right)\|E\|_F$$

Mas

$$\left\|\frac{M^{-1}s(M^{-1}s-Mc)^T}{c^Ts}\right\|_F \quad = \quad \frac{\|M^{-1}s\| \; \|M^{-1}s-Mc\|}{c^Ts}$$

e o resultado segue de (a).

<u>Solução d)</u>: Temos

$$\left\|\frac{(y-As)(Mc)^T}{c^Ts}\right\|_F = \frac{\|y-As\| \; \|Mc\|}{c^Ts} \leq \frac{\|y-As\|}{c^Ts}\left(\|Mc-M^{-1}s\|+\|M^{-1}s\|\right)$$

e em virtude de (4.4)

$$\left\|\frac{(y-As)(Mc)^T}{c^Ts}\right\|_F \quad \leq \quad (1+\beta)\,\|y-As\|\left(\frac{\|M^{-1}s\|}{c^Ts}\right)$$

Usando o fato de $1+\beta \leq 2(1-\beta)$ e o item (a) :

$$\left\| \frac{(y-As)(Mc)^T}{c^Ts} \right\|_F \leq 2\frac{\|y-As\|}{\|M^{-1}s\|} \qquad \text{cqd //}$$

OBSERVAÇÕES:

i) Para aplicar os resultados acima, assumimos que $f : R^n \to R^n$ satisfaz as hipóteses do lema 3.1 e que $f(x^*) = 0$.

ii) Considere agora a adaptação feita em (4.4), onde o vetor c é escolhido de tal forma que, para $\forall(x,B) \in N'$, vizinhança de $(x^*,f'(x^*))$.

$$\frac{\|Mc-M^{-1}s\|}{\|M^{-1}s\|} \leq \mu_1 \|s\|^p , \quad \text{com } \mu_1 \geq 0 , \qquad (4.7)$$

$s \neq 0$ e M é uma matriz simétrica, não-singular $L(R^n)$.

iii) A vizinhança $N' = N_1' \times N_2'$ onde $N_1' \subset D$ e N_2' contêm somente matrizes não-singulares de forma que o vetor $s = -B^{-1}f(x)$ está bem definido.

iv) Nos lemas 4.1 e 4.2, tomo $A = f'(x)$ e uso a norma $\|Q\|_M = \|MQM\|_F$.

v) Seja $E = M[B-f'(x^*)]M$ na parte (c) e (d) do lema 4.2.

Assim, se $\mu_1\|s\|^p \leq 1/3$, $s \neq 0$ e $B \neq f'(x^*)$ então:

$$\|\bar{B} - f'(x^*)\|_M \leq \left[\sqrt{1-\alpha\theta^2} + (1-\beta)^{-1}\mu_1 \|s\|^p\right] \|B-f'(x^*)\|_M +$$

$$+ \frac{2\|M\| \; \|y-f'(x^*)s\|}{\|M^{-1}s\|}$$

onde

$$\theta = \frac{\|M[B-f'(x^*)]s\|}{\|[B-f'(x^*)]\|_M \; \|M^{-1}s\|}$$

vi) Existe $(x,B) \in N'$ tal que

$$\|s\| \leq 2 \text{ máx}\{\|\bar{x}-x^*\|, \|x-x^*\|\}$$

e se $\bar{x} \in D$, pelo lema 3.1 temos

$$\|y-f'(x^*)s\| \leq k \text{ máx } \{\|\bar{x}-x^*\|^P, \|x-x^*\|^P\} \; \|s\|$$

Portanto ,

$$\|\bar{B}-f'(x^*)\|_M \leq \left[\sqrt{1-\alpha\theta^2}+\alpha_1 \text{ máx}\{\|\bar{x}-x^*\|^P, \|x-x^*\|^P\}\right] \|B-f'(x^*)\|_M +$$

$$+ \alpha_2 \text{ máx } \{\|\bar{x}-x^*\|^P, \|x-x^*\|^P\} \tag{4.8}$$

onde $\alpha_1 = (1-\beta)^{-1} \mu_1 \, 2^P$ e $\alpha_2 = 2k\|M\|^2$ se $\bar{x} \in D$.

TEOREMA 4.3: Considere f satisfazendo as hipóteses do teorema 3.2.

Se (4.7) vale para alguma constante $\mu_1 \geq 0$, uma matriz simétrica e não-singular $M \in L(R^n)$ e todo $(x,B) \in N'$.

Então, a função atualizadora $U(x,B) = \{\bar{B}\}$ onde

$$\bar{B} = B + \frac{(y-Bs)\,c^T}{c^Ts}$$

está bem definida numa vizinhança N de $(x^*, f'(x^*))$ e a iteração correspondente

$$x_{k+1} = x_k - B_k^{-1} f(x_k) \tag{4.9}$$

com $B_{k+1} \in U(x_k, B_k)$ é localmente e Q-superlinearmente convergente em x^* .

Além disso, se $f'(x^*)$ é uma matriz simétrica, a mesma conclusão vale para $x_{k+1} = x_k - (B_k^T)^{-1} f(x_k)$.

DEMONSTRAÇÃO:

Vamos definir $N_2 = \{B \in L(R^n) : \|f'(x^*)^{-1}\| \, \|B - f'(x^*)\| < 1/2\}$. Então todo $B \in N_2$ é não-singular e $\|B^{-1}\| \leq 2\|f'(x^*)^{-1}\|$ (pelo lema de Perturbação de Banach) .

Considere agora $\varepsilon > 0$ e $\rho > 0$ como no lema 3.1 tais que

$\max\{\|\bar{x} - x^*\|, \|x - x^*\|\} \leq \varepsilon$ => a desigualdade (3.3) é verdadeira.

Em particular, se $\|x - x^*\| \leq \varepsilon$ e $B \in N_2$ => $x \in D$ e

$$\|s\| = \|B^{-1} f(x)\| \underset{(3.3)}{\leq} 2\rho \|f'(x^*)^{-1}\| \, \|x - x^*\|$$

Assim defino $N_1 = \{x \in R^n : \|x - x^*\| < \varepsilon/2$ e

$$2\rho \, \|f'(x^*)^{-1}\| \, \|x - x^*\| < \min\{\varepsilon/2, (3\mu_1)^{-1/P}\}\}$$

Se $N = (N_1 \times N_2) \cap N'$ e $(x, B) \in N \overset{(4.7)}{=>}$ a desigualdade (4.4) é verdadeira para

$$\beta = 1/3 \quad \text{e} \quad \|\bar{x} - x^*\| \leq \|s\| + \|x - x^*\| \leq \varepsilon$$

Portanto $\bar{x} \in D$ e por (3.3), $s = 0$ se, e só se, $x = x^*$.

Mostramos assim que a função atualizadora associada a iteração (4.9) satisfaz as hipóteses do teorema 3.2.

Portanto a iteração (4.9) é localmente convergente.

Observe que (4.8) pode ser escrita como

$$\|B_{k+1}-f'(x^*)\|_M \leq \sqrt{1-(3/8)\theta_k^2}\ \|B_k-f'(x^*)\|_M + \text{máx}\{\|x_{k+1}-x^*\|^p,\ \|x_k-x^*\|^p\}\left[\alpha_1 \|B_k-f'(x^*)\|_M + \alpha_2\right] \tag{4.10}$$

onde

$$\theta_k = \frac{\|M[B_k-f'(x^*)]s_k\|}{\|B_k-f'(x^*)\|_M\ \|M^{-1}s_k\|} \quad \text{para } B_k \neq f'(x^*)$$

Se existe uma subsequência de $\{B_k\}$ convergindo para $f'(x^*)$, o corolário 3.3 resolve o nosso problema agora, se não existe tal subsequência, isto é, $\{\|B_k-f'(x^*)\|_M\}$ tiver limite diferente de zero e como $\sqrt{1-2} \leq 1-(\alpha/2)$, em virtude de (4.10) temos

$$(3/16)\theta_k^2\ \|B_k-f'(x^*)\|_M \leq \|B_k-f'(x^*)\|_M - \|B_{k+1}-f'(x^*)\|_M + $$
$$+ \text{máx}\{\|x_{k+1}-x^*\|^p,\ \|x_k-x^*\|^p\}\ [\alpha_1\ \|B_k-f'(x^*)\|_M + \alpha_2$$

Portanto

$$\sum_{k=1}^{\infty} \theta_k^2\ \|B_k-f'(x^*)\|_M < +\infty$$

Mas $\{\|B_k - f'(x^*)\|_M\}$ é limitada e $\|B_k - f'(x^*)\|_M \geq 0$, $\forall$ k. Assim :

$$\lim_{k \to +\infty} \frac{\| [B_k - f'(x^*)] s_k \|}{\|s_k\|} = 0 \tag{4.11}$$

Além disso $B_k s_k = -f(x_k) \Rightarrow$

$$\Rightarrow \quad \| [B_k - f'(x^*)] s_k \| \geq \|f(x_{k+1})\| - \|f(x_{k+1}) - f(x_k) - f'(x^*) s_k\|$$

Logo, de $\|x_{k+1} - x^*\| \leq \|x_k - x^*\|$, concluímos por - (3.2) que ,

$$\| [B_k - f'(x^*)] s_k \| \geq \frac{1}{\rho} \|x_{k+1} - x^*\| - k \|x_k - x^*\|^P \|s_k\|$$

por (3.2), (3.3) e pelo lema 3.1 .

Mas

$$\|s_k\| = \|x_{k+1} - x_k\| \leq \|x_{k+1} - x^*\| + \|x_k - x^*\| \leq \|x_k - x^*\| + \|x_k - x^*\| = 2\|x_k - x^*\|$$

decorre que ,

$$\frac{\| [B_k - f'(x^*)] s_k \|}{\|s_k\|} \geq (2\rho)^{-1} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} - k \|x_k - x^*\|^P \tag{4.11a}$$

Assim, de (4.11) e (4.11a) obtemos

$$\lim_{k \to \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0 \qquad \text{cqd } //$$

NOTA: A condição $\frac{\|Mc - M^{-1}s\|}{\|M^{-1}s\|\|} \leq \bar{\mu}_1 \max\{\|\bar{x} - x^*\|^P, \|x - x^*\|^P\}$, (4.7') é equivalente a desigualdade (4.7).

Com efeito, seja N' vizinhança de $(x^*, f'(x^*))$ definida anteriormente.

Para todo $(x,b) \in N'$

$$\|s\| \leq 2 \text{ máx } \{\|\bar{x}-x^*\|, \|x-x^*\|\}$$

Assim

$$\frac{\|Mc-M^{-1}s\|}{\|M^{-1}s\|} \leq \mu_1 \|s\|^p \leq \bar{\mu}_1 \text{máx } \{\|\bar{x}-x^*\|^p, \|x-x^*\|^p\}$$

onde $\bar{\mu}_1 = 2\mu_1$.

Mostramos que a condição (4.7) => (4.7') .

A recíproca será mostrada pelo seguinte corolário:

COROLÁRIO 4.4: Se $f: R^n \to R^n$ satisfaz as hipóteses do teorema 4.3.

Então o primeiro método de Broyden (c=s) é localmente e Q-superlinearmente convergente em x^*. Além disso, se a matriz $f'(x^*)^{-1}$ for simétrica e positiva definida então o método de Pearson (c=y) é também localmente e Q-superlinearmente convergente em x^*.

DEMONSTRAÇÃO:: O resultado para o primeiro método de Broyden (c=s) é imediato pois tomando P=I (matriz Identidade) e $\mu_1 = 0$, a condição (4.7) e, portanto, as hipóteses do teorema 4.3 estão satisfeitas.

Considere agora o caso c=y.

Por hipótese a matriz $f'(x^*)^{-1}$ é simétrica e positiva definida, isto implica que existe uma matriz não-singular , simétrica $P \in L(R^n)$ tal que $f'(x^*)^{-1} = (P^{-1})^2$ ou que $f'(x^*) = (P^{-1})^2$.

Assim ,

$$Py-P^{-1}s = Py-PP^{-1}P^{-1}s = Py-Pf'(x^*)s = P[y-f'(x^*)s]$$

e para $\bar{x} \in D$

$$\|Py-P^{-1}s\| \leq \|P\| \; k \; \max \{ \|\bar{x}-x^*\|^r, \; \|x-x^*\|^r\} \; \|s\|$$

pelo lema 3.1.

Como $f'(x^*)$ é inversível, o lema 3.1 nos garante a existência de $\varepsilon > 0$ e $\rho > 0$ tais que :

$$\text{Se } \max \{ \|\bar{x}-x^*\|, \|x-x^*\|\} \leq \varepsilon \Rightarrow (\frac{1}{\rho}) \|\bar{x}-x\| \leq \|f(\bar{x})-f(x)\| \leq$$
$$\leq \rho \|\bar{x}-x\| , \qquad (*)$$

Considere $N_1 = \{x \in R^n: \|x-x^*\| < \varepsilon/_2$ e $2\rho \|f'(x^*)^{-1}\| \|x-x^*\| < \varepsilon/_2\}$ e $N_2 = \{B \in L(R^n): \|f'(x^*)^{-1}\| \; \|B-f'(x^*)\| < 1/2\}$

Então $N' = N_1 \times N_2$ é uma vizinhança de $(x^*, f'(x^*))$ e se $(x,B) \in N' \Rightarrow B$ é não singular e $\|B^{-1}\| \leq 2\|f'(x^*)^{-1}\|$ (pelo lema de Banach) .

Ainda da condição (*)

$$\|s\| = \|B^{-1}f(x)\| \leq 2\|f'(x^*)^{-1}\| \; \|f(x)\| \leq 2\rho \|x-x^*\| \; \|f'(x^*)^{-1}\| < \varepsilon/2$$

$$\Rightarrow \bar{x} \in D .$$

E como $\|B\| \; \|s\| \geq \|Bs\| \overset{(*)}{\geq} (1/\rho) \; \|x-x^*\|$ concluímos que

$$\text{máx}\{\|\bar{x}-x^*\|, \|x-x^*\|\} \le \|s\|+\|x-x^*\| \le \|s\|+\rho\|B\| \|s\| = (1+\rho\|B\|)\|s\|$$

ou

$$\text{máx}\ \{\|\bar{x}-x^*\|^r, \|x-x^*\|^r\} \le \lambda\|s\|^r \qquad \forall\ (x,B) \in N'$$

e onde $\lambda = (1+\rho\|B\|)$.

A demonstração conclui-se do teorema 4.3. //

Dennis, (1971), demonstrou que os métodos de Broyden são localmente convergentes com as condições do corolário 4.4 para $r = 1$.

Este resultado foi por ele melhorado, com a conclusão de que, se numa k-ésima iteração conservarmos a mesma aproximação B_k durante algumas iterações seguintes, a convergência local é obtida, embora possa ser somente linear.

Mostraremos agora resultados semelhantes aos do corolário 4.4 para o segundo método de Broyden e para o método de McCormick.

TEOREMA 4.5: Considere f satisfazendo as hipóteses do teorema 3.2.

Se para alguma constante $\lambda \ge 0$, alguma matriz simétrica e não-singular $P \in L(R^n)$ e todo (x,G) em N' vizinhança de $(x^*, f'(x^*)^{-1})$ tivermos :

$$\frac{\|Pd-P^{-1}y\|}{\|P^{-1}y\|} \le \lambda\|y\|^r , \qquad y \neq 0 \tag{4.12}$$

Então, a função atualizadora

$$U(x,G) = \{\bar{G}\}$$

onde

$$\bar{G} = G + \frac{(s-Gy)\ d^T}{d^T y} \tag{4.13}$$

está bem definida numa vizinhança N de $(x^*, f'(x^*)^{-1})$ e a iteração correspondente

$$x_{k+1} = x_K - G_k f(x^k) \tag{4.14}$$

é localmente e Q-superlinearmente convergente em x^*.

Além disso, se a matriz $f'(x^*)$ é simétrica, o mesmo resultado vale para $x_{k+1} = x_k - (G_k^T) f(x_k)$.

DEMONSTRAÇÃO: Considere $N_2 = \{G \in L(R^n) : \|f'(x^*)\|\ \|G - f'(x^*)^{-1}\| < 1/2\}$. Pelo lema de perturbação de Banach, toda matriz $G \in N_2$ é não singular e $\|G\| \leq v$ para alguma constante v.

Se tomarmos $\varepsilon > 0$ e $\rho > 0$ como no lema 3.1 tais que $\text{máx}\,\{\|\bar{x}-x^*\|, \|x-x^*\|\} < \varepsilon$ isto implica que

$$(1/\rho)\|\bar{x}-x\| \leq \|f(\bar{x})-f(x)\| \leq \rho\|\bar{x}-x\| \tag{4.14a}$$

Para o caso em que $\|x-x^*\| \leq \varepsilon$ e $G \in N_2 \Rightarrow x \in D$ e $\|s\| \leq \rho\|G\|\ \|x-x^*\| \leq \rho v\|x-x^*\|$.

Seja $N_1 = \{x \in R^n : \|x-x^*\| < \min\{\varepsilon/2;\ \varepsilon/(2\rho v)\}$, onde $\rho\varepsilon < (3\lambda)^{-1/r}\}$.

Assim se $N = (N_1 \times N_2) \cap N'$ e $(x,G) \in N \Rightarrow x$ e x' estão em D.

Pela condição (4.14a) temos :

$$(\frac{1}{\rho})\,\|s\| \;\leq\; \|y\| \;\leq\; \rho\|s\| \tag{4.15}$$

e, em particular,

$$\lambda\,\|y\|^r \;\leq\; \lambda(\rho\varepsilon)^r \;<\; 1/3 \tag{4.16}$$

Portanto $y = 0 \iff s = 0 \iff x = x^*$.

Mostraremos agora que a matriz $\bar{G}$ de (4.13) satisfaz as condições do teorema 3.4.

Com efeito, fazendo $\bar{G} = \bar{B}$, $y = s$ e $d = c$ em (4.13) obtemos a condição (4.1).

Do lema 4.1 obtemos :

$$\bar{E} = E\left[I - \frac{P^{-1}y(Pd)^T}{d^Ty}\right] + \frac{P(s-Ay)(Pd)^T}{d^Ty}$$

com $A \in L(R^n)$ e $\bar{E} = P(\bar{G}-A)P$ e $E = P(G-A)P$.

Fazendo $A = f'(x^*)^{-1}$ e desde que (4.12) e (4.16) implicam na veracidade da condição (4.4) para $\beta = 1/3$, segue-se das partes (c) e (d) do Lema 4.12 e Condição (4.12) que :

$$\|\bar{G}-f'(x^*)^{-1}\|_M \leq \left[\sqrt{1-(3/8)\theta^2} + (3/2)\lambda\|y\|^r\right] \|G-f'(x^*)^{-1}\|_M + \\ + \frac{2\|P\|\;\;\|s-f'(x^*)^{-1}y\|}{\|P^{-1}y\|} \tag{4.17}$$

onde

$$\theta = \frac{\|P\left[G-f'(x^*)^{-1}\right]y\|}{\|G-f'(x^*)^{-1}\|_M\;\|P^{-1}y\|}$$

Mas para $\forall\, x \in N_1$,

$$\|s-f'(x^*)^{-1}y\| \leq k\|f'(x^*)^{-1}\| \max\{\|\bar{x}-x^*\|^r, \|x-x^*\|^r\} \|s\|$$

e por (4.15), (4.16) e (4.17) temos

$$\|\bar{G}-f'(x^*)^{-1}\|_M \leq \sqrt{1-(3/8)\theta^2} \|G-f'(x^*)^{-1}\|_M + \max\{\|\bar{x}-x^*\|^r, \|x-x^*\|^r\} \left[\alpha_1\|G-f'(x^*)^{-1}\|_M+\alpha_2\right] \tag{4.18}$$

onde $\alpha_1 = (2/3)(2\rho)^r\lambda$ e $\alpha_2 = 2\rho k\|P\|^2 \|f'(x^*)^{-1}\|$

Assim as hipóteses do teorema 3.4 estão satisfeitas e consequentemente a iteração (4.14) é localmente convergente em x^*.

Para mostrar que (4.14) é Q-superlinearmente convergente, definimos:

$$\theta_k = \frac{\|P\left[G_k-f'(x^*)^{-1}\right]y_k\|}{\|G_k-f'(x^*)^{-1}\|_M \|P^{-1}y_k\|}$$

Temos dois casos a tratar.

Primeiro; se existe uma subsequência de $\{G_k\}$ convergindo para $f'(x^*)^{-1}$ então pelo corolário 3.5 concluímos nossa demonstração.

Caso contrário; como a sequência $\{\|G_k-f'(x^*)^{-1}\|_M\}$ possui limite diferente de zero, da Condição (4.18) resulta que

$$\sum_{k=1}^{\infty} \theta_k^2\|G_k-f'(x^*)^{-1}\|_M < +\infty \quad \text{e} \quad \lim_{k\to+\infty} \frac{\|\left[G_k-f'(x^*)^{-1}\right]y_k\|}{\|y_k\|} = 0 \tag{4.19}$$

E da igualdade $G_ky_k = G_kf(x_{k+1})+s_k$ obtemos :

$$\left[G_k-f'(x^*)^{-1}\right]y_k = G_kf(x_{k+1})-f'(x^*)^{-1}\left[y_k-f'(x^*)s_k\right] \tag{4.19a}$$

Assim o teorema 3.4 nos mostra que existe um $v' > 0$ tal que $\|G_k^{-1}\| \leq v'$ e por (4.19a), Lema 3.1 e de $\|x_{k+1}-x^*\| \leq$ $\leq \|x_k-x^*\|$ concluímos que :

$$\|f(x_{k+1})\| \leq v'\left[\|G_k-f'(x^*)^{-1}\Big]y_k\| + k\,|f'(x^*)^{-1}\|\ \|x_k-x^*\|_r\ \|s_k\|\right]$$

Desde que $\|s_k\| \geq (1/\rho)\|y_k\|$, pela condição (4.19) e do Lema 3.1 concluímos :

$$\lim_{k\to+\infty} \frac{\|f(x_{k+1})\|}{\|s_k\|} = 0$$

onde $\|f(x_{k+1})\| \geq (1/\rho)\|x_{k+1}-x^*\|$ e $\|s_k\| \leq 2\|x_k-x^*\|$.

Portanto,

$$\lim_{k\to+\infty} \frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|} = 0 \qquad \text{cqd } //$$

OBSERVAÇÃO: Note que as desigualdades,

$$\frac{\|Pd-P^{-1}y\|}{\|P^{-1}y\|} \leq \lambda\|y\|^r, \quad y \neq 0 \tag{4.12}$$

$$\frac{\|Pd-P^{-1}y\|}{\|P^{-1}y\|} \leq \bar{\lambda}\|s\|^r \tag{4.12'}$$

e

$$\frac{\|Pd-P^{-1}y\|}{\|P^{-1}y\|} \leq \bar{\lambda}\max\{\|\bar{x}-x^*\|^r,\ \|x-x^*\|^r\} \tag{4.12''}$$

são equivalentes pelo Lema 3.1 e Teorema 4.3, respectivamente.

COROLÁRIO 4.6: Se $f : R^n \to R^n$ satisfaz as hipóteses do Teorema

4.5, então o segundo método de Broyden (d=y) é localmente e Q-superlinearmente convergente em x^*.

Além disso, se a matriz Jacobiana $f'(x^*)$ é simétrica e positiva definida, então o método de McCormick (d=s) é também localmente e Q-superlinearmente convergente em x^*.

DEMONSTRAÇÃO: A prova é imediata pois no caso (d=y) basta tomarmos $P = I$ (matriz Identidade) e a condição suficiente (4.12) estará satisfeita.

Para (d=s) como $f'(x^*)$ é simétrica e positiva definida, existe uma matriz também simétrica e positiva definida $P \in L(R^n)$ tal que $f'(x^*) = P^2$ e assim $P^{-1}y-Ps = P^{-1}[y-f'(x^*)s]$.

Portanto pelo Lema 3.1, a Condição (4.12") está satisfeita, completando a prova.

Analisaremos agora os métodos simétricos de Posto-um definidos para $c = y-B_1$, onde B é uma matriz simétrica.

Uma ressalva que fazemos de tais métodos, é que sob determinadas condições, esses podem não satisfazer a condição (4.7) para toda vizinhança N' de $(x^*,f'(x^*))$.

Com efeito, da igualdade $c = y-Bs$ obtemos ;

$$\|Pc-P^{-1}s\| \geq \|P^{-1}s\| - \|P[B-f'(x^*)]s\| - \|P[y-f'(x^*)s]\|$$

Isto é ;

$$\frac{\|Pc-P^{-1}s\|}{\|P^{-1}s\|} \geq 1 - \left\{ \|P[B-f'(x^*)]P\| + \frac{\|P[y-f'(x^*)s]\|}{\|P^{-1}s\|} \right\} = 1 - A$$

onde A pode ser escolhido tão pequeno quanto necessário.

Assim para uma determinada vizinhança N_0 de $(x^*, f'(x^*))$ teremos

$$\frac{\|Pc-P^{-1}s\|}{\|P^{-1}s\|} \geq 1/2 , \qquad \forall\ (x,B) \in N_0$$

contradizendo a Condição (4.7) e implicando na não-convergência local dos métodos simétricos de posto-um.

Considere o exemplo ;

$$f : R^2 \to R^2$$
$$x \to x \qquad \forall\ x \in R^n$$

Para $\forall\ \varepsilon \in (0,1)$ considere $H_0 = \begin{bmatrix} 1+\varepsilon & 0 \\ 0 & 1-\varepsilon \end{bmatrix}$ matriz simétrica e $x_0 = (\varepsilon,\ (1+\varepsilon)\varepsilon/(1-\varepsilon))^T$, o vetor inicial.

Assim, $s_0 = -H_0 f(x_0) = -(1+\varepsilon)\ \varepsilon\ (\frac{1}{1})$ e de $y = s$

$$d = s-Hy = \begin{pmatrix} -\varepsilon & 0 \\ 0 & \varepsilon \end{pmatrix} s \quad .$$

Portanto $d^T y = d^T s = 0$, isto é, a matriz H_1 não está definida.

Como x_0 e H_0 foram escolhidas arbitrariamente, isto implica que os métodos simétricos de posto-um não são localmente - convergentes neste caso.

5) A CONVERGÊNCIA DOS MÉTODOS DE POSTO-DOIS

Neste tópico enunciaremos resultados semelhantes aos da secção anterior, mas para métodos de posto-dois.

LEMA 5.1: Considere as matrizes simétricas A, B e $M \in L(R^n)$ tais que;

$$\bar{B} = B + \frac{(y-Bs)c^T + c(y-Bs)^T}{c^T s} - \frac{s^T(y-Bs)cc^T}{(c^T s)^2} \qquad (5.1)$$

onde y e c estão em R^n com $c^T s \neq 0$ e $\bar{T} = M(\bar{B}-A)M$, $T = M(B-A)M$.

Se M é ainda não-singular, teremos :

$$\bar{T} = P^T T P + \frac{M(y-As)(Mc)^T}{c^T s} + \frac{Mc(y-As)^T MP}{c^T s} \qquad (5.2)$$

onde $P = I - \dfrac{(M^{-1}s)(Mc)^T}{c^T s}$.

Mostraremos que a Condição (5.1) satisfaz as hipóteses do Teorema 3.2.

De fato, temos que

$$\|P^T T P\|_F \leq (1 + \frac{(1-\beta)^{-1} \|Mc - M^{-1}s\|}{\|M^{-1}s\|} \|P^T T\|_F$$

Como $\|P^T T\|_F = \|T^T P\|_F$ e T é matriz simétrica, pelo Lema 4.2 chegamos à desigualdade

$$\|T^T P\|_F \leq \left(\sqrt{1-\alpha\theta^2} + (1-\beta)^{-1} \frac{\|Mc - M^{-1}s\|}{\|M^{-1}s\|}\right) \|T\|_F$$

onde $\theta = \dfrac{\|TM^{-1}s\|}{\|T\|_F \; \|M^{-1}s\|}$.

Portanto,

$$\|P^T TP\|_F \overset{(4.4)}{\leq} \left(\sqrt{1-\alpha\theta^2} + \left(\frac{5}{2}\right)(1-\beta)^{-1} \frac{\|Mc-M^{-1}s\|}{\|M^{-1}s\|}\right) \|T\|_F$$

e

$$\left\| \frac{M(y-As)(Mc)^T}{c^T s} \right\|_F \leq \alpha \|M\|_F \frac{\|y-As\|}{\|M^{-1}s\|}$$

Fazendo $T = I$ (matriz Identidade), a parte (c) do Lema 4.2 nos dá :

$$\|P\|_F \leq \left[1 + (1-\beta)^{-1}\beta\right] \sqrt{n} \leq 2\sqrt{n}$$

que por conseguinte implica em

$$\left\| \frac{(Mc)(y-As)^T MP}{c^T s} \right\|_F \leq 4\sqrt{n} \|M\|_F \frac{\|y-As\|}{\|M^{-1}s\|}$$

LEMA 5.2: Considere as matrizes B, $\bar{B}$ de (5.1) e M matriz simétrica, não-singular em $L(R^n)$ tal que a desigualdade (4.4) esteja satisfeita para $\beta \in [0,1/3]$, c e s em R^n com $s \neq 0$.

Então para toda matriz $A \in L(R^n)$ com $A \neq B$ temos:

$$\|\bar{B}-A\|_M \leq \left[\sqrt{1-\alpha\theta^2} + \left(\frac{5}{2}\right)(1-\beta)^{-1} \frac{\|Mc-M^{-1}s\|}{\|M^{-1}s\|}\right] \|B-A\|_M +$$

$$+ \; 2(1+2\sqrt{n}) \; \|M\|_F \frac{\|y-As\|}{\|M^{-1}s\|} \tag{5.3}$$

onde $\alpha = \frac{1-2\beta}{1-\beta^2} \quad [3/8, 1]$ e $\theta = \frac{\|M[B-A]s\|}{\|B-A\|_M \; \|M^{-1}s\|}$

TEOREMA 5.3: Seja f satisfazendo as hipóteses do Teorema 3.2. Se a matriz Jacobiano $f'(x^*)$ é simétrica, (4.7) valer para algum $\mu_1 \geq 0$, alguma matriz simétrica e não-singular $M \in L(R^n)$ e $\forall\ (x,B) \in N'$.

Então; $U(x,B) = \{\bar{B}\}$ onde

$$\bar{B} = B + \frac{(y-Bs)c^T + c(y-Bs)^T}{c^T s} - \frac{s^T(y-Bs)cc^T}{(c^T s)^2}$$

onde B é simétrica está bem definida e além disso, a iteração

$$x_{k+1} = x_k - B_f^{-1} f(x_k)$$

é localmente e Q-superlinearmente convergente em x^*.

DEMONSTRAÇÃO: Considere $0 < \mu_1 \|s\|^p \leq 1/3$ e $A = f'(x^*)$. Então

$$\|\bar{B}-f'(x^*)\|_M \overset{(4.7)}{\leq} \left[\sqrt{1-\alpha\theta^2} + \left(\frac{5}{2}\right)(1-\beta)^{-1}\mu_1 \|s\|^p\right] \|B-f'(x^*)\|_M +$$

$$+ \quad 2(1+2\sqrt{n}) \frac{\|M\| \ \|y-f'(x^*)s\|}{\|M^{-1}s\|}$$

Mas para algum $(x,B) \in N'$

$$\|s\| \leq 2 \text{ máx } \{\|\bar{x}-x^*\|, \|x-x^*\|\} .$$

Portanto, se $\bar{x} \in D$ pelo Lema 3.1 temos que :

$$\|y-f'(x^*)s\| \leq k - \text{máx } \{\|\bar{x}-x^*\|^p, \|x-x^*\|^p\} \ \|s\|$$

e

$$\|\bar{B}-f'(x^*)\|_M \leq \left[\sqrt{1-\alpha\theta^2} + \alpha_1 \text{ máx } \{\|\bar{x}-x^*\|^p, \|x-x^*\|^p\}\right] \|B-f'(x^*)\|_M +$$

$$+ \quad \alpha_2 \text{ máx } \{\|\bar{x}-x^*\|^p, \|x-x^*\|^p\}$$

onde $\alpha_1 = (5/2)(1-\beta)^{-1}\mu_1 2^P$ e $\alpha_2 = 2k(1+2\sqrt{n})\,\|M\|_F\;\|M\|$.

A prova conclui-se de modo análogo a do Teorema 4.3.

COROLÁRIO 5.4: Seja $f : R^n \to R^n$ satisfazendo as hipóteses do Teorema 5.3.

Então o algorítmo de Broyden (c=s) (Powell Simétrico) é localmente e Q-superlinearmente convergente em x*. Além disso,se a matriz Jacobiana f'(x*) é positiva definida, então o método DFP (Davidon - Fletcher - Powell) é também localmente e Q-superlinearmente convergente em x* .

NOTA: Os resultados referentes aos métodos de Broyden obtidos por Dennis e Moré [13], vistos neste capítulo, foram extendidos por Moré e Trangestein [7] que mostraram que a convergência local poderia ser substituída por convergência global através de uma modificação nos métodos de Broyden para sistemas lineares.

CAPÍTULO II

OS RESULTADOS SURPREENDENTES OBTIDOS POR D. GAY [6]

1) INTRODUÇÃO

Mostraremos aqui os principais resultados obtidos por David M. Gay [6] para os métodos de Broyden.

Na secção 2, serão descritos detalhadamente os algorítmos Broyden Bom e Broyden Ruim e mostraremos, na secção 3, o teorema da terminação finita dos métodos de Broyden para sistemas de equações lineares.

Na secção 4, utilizando o resultado da secção anterior, será demonstrada a convergência Q-quadrática do Método de Broyden Bom para sistemas não-lineares, salientando que esta propriedade deve ser verificada também para o Broyden Ruim.

Finalizando o capítulo, analisaremos as consequências decorrentes destes resultados.

2) OS ALGORÍTMOS BOM E RUIM

Considere $f : R^n \to R^n$.

Os métodos de Broyden são descritos na sua forma geral por :

(2.1)

(2.1a) $s_k = -H_k f(x_k)$

(2.1b) $x_{k+1} = x_k + s_k$

(2.1c) $y_k = f(x_{k+1}) - f(x_k)$

(2.1d) Se $y_k = 0$ então $H_{k+1} = H_k$

(2.1e) Escolha; $v_k \in R^n$ tal que $v_k^T y_k = 1$

(2.1f) e $v_k^T H_k^{-1} s_k \neq 0$

(2.1g) $H_{k+1} = H_k + (s_k - H_k y_k) v_k^T$

O primeiro método de Broyden (conhecido como o método de Broyden Bom) resulta da escolha de $v_k = H_k^T s_k / (s_k^T H_k y_k)$ em (2.1e), com $y_k \neq 0$ e $s_k^T H_k y_k \neq 0$.

O segundo método (Broyden Ruim) resulta da escolha de $v_k = y_k/(y_k^T y_k)$ definido para $y_k \neq 0$ e $y_k^T H_k^{-1} s_k \neq 0$.

OBSERVAÇÕES:

i) A sequência de aproximações $\{H_k\}$ acima possui a propriedade de que, se H_k é não-singular, então H_{k+1} é não-singular .

Isto pode ser mostrado através de (2.1f) e da fórmula de Sherman-Morrison para calcular a inversa de $(A + uv^T)$

$$H_{k+1}^{-1} = \frac{H_k^{-1}(s_k - H_k y_k) v_k^T H_k^{-1}}{1 + v_k^T H_k^{-1} (s_k - H_k y_k)}$$

Portanto, de (2.1f) H_k é não-singular para todo $k \geq 0$.

ii) Assim no algorítmo (2.1) observamos que $s_k = 0 \Leftrightarrow f(x_k) = 0$.

Com efeito suponha $s_k = 0$ e $f(x_k) \neq 0$. Se $f(x_k) \neq 0$ e como H_k é não-singular por (i), então $s_k = -H_k f(x_k) \neq 0$, absurdo com a hipótese de $s_k = 0$.

Portanto se $s_k = 0 \Rightarrow f(x_k) = 0$.

Reciprocamente, suponhamos $f(x_k) = 0$ e $s_k \neq 0$.

Como H_k é não-singular, temos que:

$$H_k f(x_k) = 0 \iff f(x_k) = 0$$

Mas como $H_k f(x_k) = s_k \neq 0 \Rightarrow f(x_k) \neq 0$ absurdo com a hipótese de $f(x_k) = 0$.

Assim, se $f(x_k) = 0 \Rightarrow s_k = 0$.

3) OS MÉTODOS DE BROYDEN PARA SISTEMAS LINEARES

Neste tópico, mostraremos os principais resultados obtidos por D.M. Gay para os métodos de Broyden Bom e Ruim, aplicados a sistemas lineares.

A afirmação de que o algorítmo (2.1) converge em no máximo 2n passos para sistemas lineares é uma consequência imediata do seguinte lema abaixo:

LEMA 3.1: Considere o algorítmo (2.1) aplicado a um sistema de equações lineares $f(x) = Ax-b$, onde $A \in R^{nxn}$ é não-singular.

Se $f_k = f(x_k)$ e y_{k-1} são L.I (linearmente independentes) então para $1 \leq j \leq [[(k+1)/2]]$, os vetores

$$(AH_{k-2j+1})^i f_{k-2j+1} ; \quad 0 \leq i \leq j \qquad (3.1)$$

são L.I.

OBSERVAÇÃO: $[[\gamma]]$ denota o maior inteiro menor ou igual a $\gamma \in R$.

DEMONSTRAÇÃO: Como f é linear, resulta de (2.1) que :

$$y_m = f_{m+1}-f_m = Ax_{m+1}-b-Ax_m+b = A(x_{m+1}-x_m) = As_m = -AH_m f_m.$$

Em particular para $m = k-1$

$$y_{k-1} = As_{k-1} = -A\,H_{k-1}\,f_{k-1}$$

Vamos demonstrar o Lema usando indução sobre j.

Para $j=1$, como $0 \leq i \leq j \Rightarrow i = 0$ ou $i = 1$.

Para $i=0$, obtemos f_{k-1} em (3.1).

Para $i=1$, obtemos $AH_{k-1}f_{k-1} = -y_{k-1}$ em (3.1).

Devemos mostrar que os vetores f_{k-1} e y_{k-1} são L.I.

Suponhamos que f_{k-1} e y_{k-1} são L.D (Linearmente dependentes).

Assim pela definição de vetores L.D , existe um escalar $\lambda \neq 0$, $\lambda \neq -1$ tal que $f_{k-1} = \lambda y_{k-1} = \lambda f_k - \lambda f_{k-1}$, i.é.,

$$(1+\lambda)f_{k-1} = \lambda f_k \Rightarrow f_{k-1} = \left(\frac{\lambda}{1+\lambda}\right)f_k \Rightarrow f_{k-1} = \beta f_k$$

onde $\beta = \frac{\lambda}{1+\lambda} \neq 0$ ou seja $y_{k-1} = \frac{1}{1+\lambda}\; f_k$.

Portanto y_{k-1} e f_k são vetores L.D , contradição com a hipótese do Lema.

Logo, f_{k-1} e y_{k-1} são L.I.

Utilizando a hipótese da indução finita, vamos supor o Lema válido para todo $j = m < [[(k+1)/2]]$.

OBSERVAÇÃO:

i) $2m \leq k-1$ (pela definição de $[[.]]$.

$2m \leq k-1 \;\; \Rightarrow \;\; k-2m-1 \geq 0$

ii) Se $y_i = 0 \;\; \Rightarrow \;\; y_{i+1} = 0$

Com efeito, $y_i = 0 \overset{(2.1d)}{\Rightarrow} H_i = H_{i+1}$

$$y_i = 0 \overset{(2.1c)}{\Rightarrow} f_{i+1} = f_i$$

Assim, $0 = y_i = -AH_i f_i = -AH_{i+1} f_{i+1} = y_{i+1}$, ou seja

$$y_{i+1} \neq 0 \;\; \Rightarrow \;\; y_i \neq 0 \; .$$

Em particular para $i = k-1$ obtemos :

$y_{k+1} \neq 0 \;\; \Rightarrow \;\; y_{k-2} \neq 0 \;\; \Rightarrow \; \dots \; \Rightarrow \;\; y_{k-2m-1} \neq 0$, isto é ;

$y_{k-1} \neq 0 \;\; \Rightarrow \;\; y_{k-2m-1} \neq 0$.

iii) $AH_{k-2m-1} = AH_{k-2m}\left[I-(I-AH_{k-2m})f_{k-2m}v_{k-2m}^T\right]$. De fato, como

$s_{k-2m}-H_{k-2m}y_{k-2m} = -H_{k-2m}f_{k-2m}+H_{k-2m}AH_{k-2m}f_{k-2m} =$

$= -H_{k-2m}(I-AH_{k-2m})f_{k-2m}$.

$$\begin{aligned} AH_{k-2m-1} &= A\left[H_{k-2m}+(s_{k-2m}-H_{k-2m}y_{k-2m})v_{k-2m}^T\right] = \\ &= A\left[H_{k-2m}+(-H_{k-2m}(I-AH_{k-2m})f_{k-2m}v_{k-2m}^T\right] = \\ &= AH_{k-2m}\left[I-(I-AH_{k-2m})f_{k-2m}v_{k-2m}^T\right]. \end{aligned}$$

iv) Além disso ;

$$f_{k-2m+1} = y_{k-2m}+f_{k-2m} = -AH_{k-m}f_{k-2m}+f_{k-2m} = (I-AH_{k-2m})f_{k-2m}.$$

Pela hipótese da indução, os vetores

$$(AH_{k-2m+1})^i \ f_{k-2m+1} \ ; \qquad 0 \leq i \leq m$$

são L.I. Assim de (iii), (iv) e da hipótese acima concluímos que existe $\xi\ i,\ell$ (dependendo de k e m) , tal que

$$(*) \quad (AH_{k-2m+1})^i f_{k-2m+1} /\!/ (I-AH_{k-2m}) \left[(AH_{k-2m})^i + \sum_{\ell=1}^{i-1} \xi_{i,\ell} (AH_{k-2m})^\ell \right] f_{k-2m}$$

para $0 \leq i \leq m$; de onde os vetores $(I-AH_{k-2m})(AH_{k-2m})^i \ f_{k-2m}$, $0 \leq i \leq m$ são L.I.

Com efeito, por (iv) obtemos a igualdade

$$f_{k-2m+1} = (I-AH_{k-2m}) \ f_{k-2m}$$

Assim ,

$$f_{k-2m+1} /\!/ (I-AH_{k-2m}) \ f_{k-2m}$$

provando a condição (*) para $i = 0$.

Para $i = 1$ obtemos :

$$AH_{k-2m+1} f_{k-2m+1} = AH_{k-2m}(I-(I-AH_{k-2m}) f_{k-2m} v_{k-2m}^T)(I-AH_{k-2m}) f_{k-2m} =$$

$$= \left[AH_{k-2m} - AH_{k-2m}(I-AH_{k-2m}) f_{k-2m} v_{k-2m}^T \right] (I-AH_{k-2m}) f_{k-2m} =$$

$$= \left[AH_{k-2m} - AH_{k-2m} f_{k-2m} v_{k-2m}^T + (AH_{k-2m})^2 f_{k-2m} v_{k-2m}^T \right] (I-AH_{k-2m}) f_{k-2m} =$$

$$= \Big[AH_{k-2m}(I-AH_{k-2m}) - AH_{k-2m} f_{k-2m} v_{k-2m}^T (I-AH_{k-2m}) \quad +$$

$$+ \ (AH_{k-2m})^2 f_{k-2m} v_{k-2m}^T (I-AH_{k-2m}) \Big] f_{k-2m} =$$

$$= \Big[AH_{k-2m} - (AH_{k-2m})^2 - AH_{k-2m} f_{k-2m} v_{k-2m}^T + AH_{k-2m} f_{k-2m} v_{k-2m}^T AH_{k-2m} \quad +$$

$$+ \left[AH_{k-2m}\right]^2 f_{k-2m} v_{k-2m}^T - (AH_{k-2m})^2 f_{k-2m} v_{k-2m}^T AH_{k-2m}\Big] f_{k-2m} =$$

$$= \Big[(I-AH_{k-2m}) AH_{k-2m} - (I-AH_{k-2m}) AH_{k-2m} f_{k-2m} v_{k-2m}^T +$$

$$+ (I-AH_{k-2m}) AH_{k-2m} f_{k-2m} v_{k-2m}^T AH_{k-2m}\Big] f_{k-2m} =$$

$$= \left[(I-AH_{k-2m}) AH_{k-2m} (I - f_{k-2m} v_{k-2m}^T + f_{k-2m} v_{k-2m}^T AH_{k-2m})\right] f_{k-2m} =$$

$$= \left[(I-AH_{k-2m}) AH_{k-2m} (f_{k-2m} - f_{k-2m} v_{k-2m}^T f_{k-2m} + f_{k-2m} v_{k-2m}^T AH_{k-2m} f_{k-2m})\right]$$

$$= \left[(I-AH_{k-2m}) AH_{k-2m} (f_{k-2m} (I - v_{k-2m}^T f_{k-2m} + v_{k-2m}^T AH_{k-2m} f_{k-2m}))\right] =$$

$$= \left[(I-AH_{k-2m}) AH_{k-2m} f_{k-2m} (1 - v_{k-2m}^T f_{k-2m} - v_{k-2m}^T y_{k-2m})\right] =$$

$$= (I-AH_{k-2m}) AH_{k-2m} f_{k-2m} (1 - v_{k-2m}^T f_{k-2m-1}) =$$

$$= (I-AH_{k-2m}) AH_{k-2m} f_{k-2m} (-v_{k-2m}^T f_{k-2m}) =$$

$$= k (I-AH_{k-2m}) AH_{k-2m} f_{k-2m}, \quad \text{onde} \quad k = -v_{k-2m}^T f_{k-2m} \in R .$$

$$AH_{k-2m+1}\ f_{k-2m+1} \ /\!/\ (I-AH_{k-2m}) AH_{k-2m}\ f_{k-2m}$$

Usando indução sobre i, concluímos que

$$(AH_{k-2m+1})^i f_{k-2m+1} \ /\!/\ (I-AH_{k-2m}) \left[(AH_{k-2m})^i + \sum_{\ell=1}^{i-1} \xi_{i,\ell} (AH_{k-2m})^\ell\right] f_{k-2m}$$

para todo $i \geq 0$, e portanto

$$(I-AH_{k-2m}) (AH_{k-2m})^i f_{k-2m} \ ; \quad 0 \leq i \leq m \quad \text{são L.I}$$

Com efeito, já vimos que os vetores $(AH_{k-2m})^i f_{k-2m}$ $0 \leq i \leq m$ são L.I, assim se

$$\sum_{j=0}^{m} \alpha_j \; (I-AH_{k-2m})(AH_{k-2m})^j \; f_{k-2m} \;=\; 0$$

$$(I-AH_{k-2m}) \sum_{j=0}^{m} \alpha_j (AH_{k-2m})^j \; f_{k-2m} \;=\; 0$$

e de $(I-AH_{k-2m}) \neq [0]$ (matriz nula) temos que ,

$$\sum_{j=0}^{m} \alpha_j (AH_{k-2m})^j \; f_{k-2m} \;=\; 0$$

mas $(AH_{k-2m})^j \; f_{k-2m}$ são vetores L.I para $0 \leq j \leq m$, logo

$$\alpha_0 = \alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$$

concluindo que $(I-AH_{k-2m})(AH_{k-2m})^j f_{k-2m}$ são L.I.

Afirmamos também que $(I-AH_{k-2m})y_{k-2m-1} = 0$.

De fato ,

$$(I-AH_{k-2m})y_{k-2m-1} = y_{k-2m-1} - AH_{k-2m}y_{k-2m-1} \overset{(2.1g)}{=}$$

$$= \quad y_{k-2m-1} - A(H_{k-2m-1} + (s_{k-2m-1} - H_{k-2m-1}y_{k-2m-1})v^T_{k-2m-1})y_{k-2m-1}$$

$$\overset{(2.1e)}{=} \quad y_{k-2m-1} - AH_{k-2m-1}y_{k-2m-1} - A(s_{k-2m-1} - H_{k-2m-1}y_{k-2m-1}) \;=$$

$$= \quad y_{k-2m-1} - As_{k-2m-1} = y_{k-2m-1} - y_{k-2m-1} \;=\; 0 \quad // \; .$$

Assim, $y_{k-2m-1} = -AH_{k-2m-1} \; f_{k-2m-1}$ e $(AH_{k-2m})^i f_{k-2m}$ $0 \leq i \leq m$ são vetores L.I .

Assim de modo semelhante a (*) acima, existe $\delta_{i,\ell}$ (dependendo de k e m) tal que

$$(AH_{k-2m})^i f_{k-2m} \;//\; (I-AH_{k-2m-1}) \; \Big[(AH_{k-2m-1})^i +$$

$$+ \sum_{\ell=1}^{i-1} \delta_{i,\ell} (AH_{k-2m-1})^{\ell} \Big] f_{k-2m-1}$$

para $0 \leq i \leq m$, de onde concluímos que os vetores

$$(AH_{k-2m-1})^{i} f_{k-2m-1}$$

$0 \leq i \leq m+1$ são L.I .

Portanto a Condição (3.1) está satisfeita para $j = m+1$ e pela indução chegamos que

$$(AH_{k-2j+1})^{i} f_{k-2j+1} \quad \text{são L.I}$$

para todo j tal que $0 \leq i \leq j$. cqd //

TEOREMA 3.2: Se $f(x) = Ax-b$, $A \in R^{nxn}$ matriz não-singular, então o algorítmo (2.1) converge em no máximo 2n passos, isto é, $f_{2n} = 0$.

DEMONSTRAÇÃO: Já vimos que as matrizes H_k são não-singulares. Assim valem as seguintes implicações :

i) Se $f_{2n-2} \neq 0 \Rightarrow s_{2n-2} = -H_{2n-2}f_{2n-2} \neq 0$ e
$y_{2n-2} = As_{2n-2} \neq 0$.

ii) Se $f_{2n-1} \neq 0 \Rightarrow f_{2n-2} \neq 0$ e portanto $y_{2n-2} \neq 0$

iii) Pelo Lema 3.1 $f_{2n-1} // y_{2n-2}$
pois em caso contrário o espaço real R^n teria n+1 vetores L.I (por exemplo tome n=1) .

Pelo algorítmo (2.1)

$$H_{2n-1} = H_{2n-2}+(s_{2n-2}-H_{2n-2}y_{2n-2})v_{2n-2}^T$$

multiplicando ambos os lados pelo vetor y_{2n-2} obtemos

$$H_{2n-1}y_{2n-2} = H_{2n-2}y_{2n-2}+(s_{2n-2}-H_{2n-2}y_{2n-2})v_{2n-2}^Ty_{2n-2}$$

$$= H_{2n-2}y_{2n-2}+s_{2n-2}-H_{2n-2}y_{2n-2} = s_{2n-2} \cdot$$

Assim ;

$$s_{2n-2} = H_{2n-1}y_{2n-2} = H_{2n-1}As_{n-2}$$

e

$$y_{2n-2} = As_{2n-2} = AH_{2n-1}y_{2n-2} \quad (0)$$

As condições (iii) e (0) implicam que ;

$$f_{2n-1} = AH_{2n-1}\, f_{2n-1}$$

Portanto ;

$$f_{2n} = f_{2n-1}+As_{2n-1} = f_{2n-1}-AH_{2n-1}f_{2n-1} =$$

$$= f_{2n-1}-f_{2n-1} = 0 \quad \text{cqd } //$$

Do Teorema 3.2 surgem dúvidas como a necessidade ou não dos 2n passos completos para obter a convergência do algorítmo - (2.1). A sugestão para se investigar isto, é iniciar tomando sistema de pequena dimensão, por exemplo n=2, onde ambos os métodos de Broyden (Bom e Ruim) provavelmente necessitam dos 2n passos para convergirem.

Abaixo, no Teorema 3.4, damos as condições necessárias para que o algorítmo (2.1) necessite dos 2n passos completos para convergir.

LEMA 3.3: Se $y_k \neq 0$, $v_k^T y_{k-1} \neq 0$ e posto$(I-AH_k) = n-1$, então, posto$(I-AH_{k+1}) = n-1$.

DEMONSTRAÇÃO: Pelo algorítmo (2.1)

$$I-AH_{k+1} = I-A\left[H_k+(s_k-H_ky_k)v_k^T\right] =$$

$$= I-A\left[H_k+(-H_kf_k+H_kAH_kf_k)v_k^T\right] = I-AH_k\left[I-\left[I-AH_k\right]f_kv_k^T\right] =$$

$$= I-AH_k\left[I-f_kv_k^T+AH_kf_kv_k^T\right] = I-AH_k+AH_kf_kv_k^T-AH_kAH_kf_kv_k^T =$$

$$= (I-AH_k)(I-AH_kf_kv_k^T) = (I-AH_k)(I-y_kv_k^T)$$

isto é, $I-AH_{k+1} = (I-AH_k)(I-y_kv_k^T)$.

Pelo Lema 3.1, $(I-AH_{k+1})y_k = 0 \Rightarrow y_k \in$ Núcleo$(I-AH_{k+1})$. Assim basta mostrarmos que para cada vetor u linearmente independente de y_k ,

$$(I-AH_{k+1})u \neq 0 \qquad (3.3.0)$$

De fato, se (3.3.0) é verdade, então $u \notin$ Núcleo$(I-AH_{k+1})$ isto é, dimensão{Núcleo$(I-AH_{k+1})$} = 1 .

Portanto

$$\text{Posto } (I-AH_{k+1}) = n - \text{dimensão } \{\text{Núcleo } (I-AH_{k+1})\} = n-1$$

Para mostrar (3.3.0) note que

$$v_k^T(I-y_kv_k^T) = 0, \quad (I-AH_k)y_{k-1} = 0 \quad \text{e} \quad (I-AH_k)z \neq 0$$

se z e y_{k-1} são L.I .

Assim para $z = (I-y_k v_k^T)u$ temos que

$$v_k^T z = 0 \quad \text{e} \quad v_k^T y_{k-1} \neq 0$$

isto é, z e y_{k-1} são L.I e $(I-AH_{k+1})u = (I-AH_k)z \neq 0$.

TEOREMA 3.4: Se $(I-AH_0)$ é uma matriz não-singular, $(AH_0)^i f_0$ $0 \leq i \leq n-1$ são linearmente independentes, $v_k^T f_k \neq 0$, $v_k^T H_k^{-1} s_k \neq 0$ $k \geq 0$ e se $v_k^T y_{k-1} \neq 0$ $k \geq 1$.

Então o algorítmo (2.1) necessita dos 2n passos completos para convergir.

DEMONSTRAÇÃO: Vimos na demonstração do Lema 3.1 que

$$f_{k+1} = (I-AH_k)\, f_k \tag{3.2a}$$

$$AH_{k+1} f_{k+1} = -(v_k^T f_k)(I-AH_k)(AH_k)\, f_k \tag{3.2b}$$

$$(AH_{k+1})^i f_{k+1} = -(v_k^T f_k)(I-AH_k)\left[(AH_k)^i + \sum_{\ell=1}^{i-1} \delta_{i,\ell}^{(k)} (AH_k)^\ell\right] f_k \tag{3.2c}$$

Além disso $(AH_{2j-1})^i f_{2j-1}$; $0 \leq i \leq n-j$ (3.3a) são vetores L.I pelo fato de $v_0^T f_0 \neq 0$, $(I-AH_0)$ ser não-singular e $(AH_0)^i f_0$; $0 \leq i \leq n-1$ serem L.I .

Da igualdade $(I-AH_1) = (I-AH_0)(I-y_0 v_0^T)$, utilizando o Lema 3.3 temos que

$$\text{Posto } (I-AH_1) = n-1$$

isto é ,

$$\text{Posto } (I-AH_{2j-1}) = n-1 \quad \text{para } j=1 \tag{3.3b}$$

Supondo que (3.3) vale para $j=k < n$ e de $y_{2k-1} = -AH_{2k-1}f_{2k-1}$, concluímos que os vetores y_{2k-1} e $(AH_{2k})^i f_{2k}$; $0 \leq i \leq n-k-1$ são L.I pelas condições (3.3a) e (3.3b) .

Mas $(I-AH_{2k})y_{2k-1} = 0 \Rightarrow$ Posto $(I-AH_{2k}) = n-1$ pelo Lema 3.3 e a condição (3.2b) .

Assim $\{y_{2k-1}\}$ cobre o núcleo (espaço nulo) de $(I-AH_{2k})$ e portanto, o conjunto de vetores $(I-AH_{2k})(AH_{2k})^i f_k$, $0 \leq i \leq n-k-1$ devem ser L.I, pois em caso contrário y_{2k-1} poderia ser escrito como uma combinação linear dos $(AH_{2k})^i f_{2k}$, $0 \leq i \leq n-k-1$.

Logo de (3.2) segue-se que as condições (3.2a) e (3.2b) são verdadeiras para $j = k+1$ e por indução, valem para todo j tal que $1 \leq j \leq n$.

Em particular $f_{2n-1} \neq 0$, isto é, algorítmo (2.1) necessita dos 2n passos completos para convergir. //

TEOREMA 3.5: Considere a aplicação linear $f(x) = Ax-b$ onde $A \in R^{nxn}$ é uma matriz não-singular.

Se aplicarmos a f, o algorítmo (2.1), substituindo a condição (2.1a) por $s_k = -\lambda_k H_k f_k$ $(\lambda_k \neq 0)$ e $\lambda_{k_i} = 0$ com $k_i \geq k_{i-1}+2$, $1 \leq i \leq n$ então $f_{2n} = 0$.

4) A CONVERGÊNCIA Q-QUADRÁTICA DOS MÉTODOS DE BROYDEN

Neste item, estudaremos a convergência Q-quadrática dos métodos de Broyden predição-direta (isto é, $\lambda_k=1$), mais especificamente, o método de Broyden Bom salientando contudo que os resul tados aqui enunciados devam valer também para o método de Broyden Ruim. Na parte (2.1c) do algorítmo (2.1) tomamos

$$v_k = \frac{H_k^T s_k}{(s_k^T H_k y_k)}$$

A fórmula de Sherman-Morrison (1949) nos dá que

$$H_{k+1}^{-1} = H_k^{-1}+(y_k-H_k^{-1}s_k)s_k^T / (s_k^T s_k)$$

Se $B_k = H_k^{-1}$ podemos reescrever o algorítmo (2.1) na seguinte forma :

(4.1a) $s_k = -B_k^{-1} f_k$

(4.1b) $x_{k+1} = x_k+s_k$

(4.1c) $y_k = f_{k+1}-f_k$

(4.1d) Se $s_k^T B_k^{-1} y_k \neq 0 \Rightarrow B_{k+1} = B_k+(y_k-B_k s_k)s_k^T / (s_k^T s_k)$

(4.1e) Se $y_k = 0 \Rightarrow B_{k+1} = B_k$.

TEOREMA 4.1: Se $f : R^n \to R^n$ é uma função diferenciável,

$$f(x)-f(y) = \int_0^1 f'(y+\tau(x-y))(x-y)d\tau, \quad f(x^*) = 0, \quad A = f'(x^*)$$

é uma matriz não-singular e f' é uma aplicação contínua e Lipschitziana em x^* (isto é, para algum λ e para todo vetor x próximo de x^* valer a desigualdade

$$\|f'(x)-f'(x^*)\| \leq \lambda\|x-x^*\| , \quad (4.2) \quad) .$$

Assim sendo, existem γ, δ, $\varepsilon > 0$ tais que se

$$\|x_0-x^*\| < \delta \quad \text{e} \quad \|B_0-A\| \leq \varepsilon$$

então:

$$\|x_{\ell+2n}-x^*\| \leq \gamma \|x_\ell - x^*\|^2 , \quad \forall\ \ell \geq 0 \tag{4.4}$$

DEMONSTRAÇÃO: Já verificamos que valem as seguintes desigualdades abaixo:

i) $\|x_{k+1}-x^*\| \leq \|x_k-x^*\|$

ii) $\|B_k-A\| \leq 2\delta$

iii) $\|f(x_{k+1})\| \leq \|f(x_k)\|$

iv) $\|B_k^{-1}\| \leq (3/4)\|A^{-1}\|$

v) $s_k^T B_k^{-1} y_k = 0 \iff x_k = x^*$, onde $\delta, \varepsilon > 0$ satisfazem a condição (4.3) e $\delta \geq 1/(4\|A^{-1}\|)$.

Com efeito, as desigualdades (i), (ii), (iii) e (iv) decorrem da Condição (4.2) e dos Teoremas 4.2 e 5.3 do capítulo anterior.

Para provar (v) note que ;

$$s_k^T B_k^{-1} y_k = 0 \overset{(4.1a)}{\iff} f_k^T (B_k^{-1})^T (B_k^{-1}) y_k = 0 \iff$$

$$\iff f_k^T (B_k^{-1})^T (B_k^{-1}) f_{k+1} - f_k^T (B_k^{-1})^T (B_k^{-1}) f_k = 0 \iff$$

$$\iff f_k^T (B_k^{-1})^T (B_k^{-1}) f_{k+1} = f_k^T (B_k^{-1})^T (B_k^{-1}) f_k$$

Como as matrizes B_k são não-singulares, a última igualdade ocorre se, e só se, $x_k = x^*$, provando assim o item (v) .

Devemos mostrar que para um ℓ fixado arbitrariamente , existe um γ independente de ℓ tal que :

$$\|x_{\ell+2n}-x^*\| \leq \gamma \|x_\ell - x^*\|^2 = \gamma h^2 \qquad \text{se} \quad h = \|x_\ell - x^*\|$$

No algorítmo (4.1) considere as sequências $\hat{x}_1, \hat{x}_2, \ldots$ $\ldots, \hat{x}_{2n}$ de vetores e $\hat{B}_1, \hat{B}_2, \ldots, \hat{B}_{2n}$ de matrizes onde $\hat{x}_0 = x_\ell$, $\hat{B}_0 = B_\ell$ e $f(x)$ é substituído por

$$\hat{f}(x) = A(x-x^*) \quad \text{isto é,} \quad \hat{f}_k = f(\hat{x}_k)$$

$$\hat{s}_k = \hat{B}_k^{-1}\hat{f}_k$$

$$\hat{x}_{k+1} = \hat{x}_k + \hat{s}_k$$

$$\hat{y}_k = \hat{f}_{k+1} - \hat{f}_k = A\hat{s}_k \qquad \text{e}$$

$$\hat{B}_{k+1} = \hat{B}_k + (\hat{y}_k - \hat{B}_k\,\hat{s}_k)\hat{s}_k^T / (\hat{s}_k^T\,\hat{s}_k) \quad \text{para} \quad \hat{s}_k \neq 0 \text{ e}$$

$\hat{B}_{k+1} = \hat{B}_k$ se $\hat{s}_k = 0$.

De forma semelhante a (ii) e (iv) obtemos

$$\|\hat{B}_k - A\| \leq 4\delta \tag{4.10}$$

e

$$\|\hat{B}_k^{-1}\| \leq 2\|A^{-1}\| \tag{4.11}$$

Mostraremos usando indução finita sobre j que existem $\gamma_{1,j}$ e $\gamma_{2,j}$ (independentes de ℓ) tais que:

$$\|B_{\ell+j} - \hat{B}_j\|\;\|f_{\ell+j}\| \leq \gamma_{1,j}\,h^2 \tag{4.12a}$$

e

$$\|x_{\ell+j} - \hat{x}_j\| \leq \gamma_{2,j}\,h^2 \quad \text{para} \quad 0 \leq j \leq 2n \tag{4.12b}$$

Desde que $B_\ell = \hat{B}_0$ e $x_\ell = \hat{x}_0$, as condições (4.12a) e (4.12b) estão satisfeitas para $j=0$ com $\gamma_{1,0} = \gamma_{2,0} = 0$.

Supomos agora que as condições (4.12a) e (4.12b) valem para $j=k$. Devemos mostrar que (4.12b) vale para $j = k+1$.

Note que:

$$\|s_{\ell+k}-\hat{s}_k\| = \|B_{\ell+k}^{-1}f_{\ell+k}-\hat{B}_k^{-1}\hat{f}_k\| = \|-B_{\ell+k}^{-1}B_{+k}\hat{B}_k^{-1}f_{\ell+k} +$$

$$+ B_{\ell+k}\hat{B}_k\hat{B}_k^{-1}f_{\ell+k}+\hat{B}_k^{-1}f_{\ell+k}-\hat{B}_k^{-1}\hat{f}_k\| =$$

$$= \|-B_{\ell+k}^{-1}(B_{\ell+k}-\hat{B}_k)\hat{B}_k^{-1}f_{\ell+k} + \hat{B}_k^{-1}(f_{\ell+k}-\hat{f}_k)\| \leq$$

$$\leq \|B_{\ell+k}^{-1}\| \|B_{\ell+k}-\hat{B}_k\| \|\hat{B}_k^{-1}\| \|f_{\ell+k}\| + \|B_k^{-1}\| \|f_{\ell+k}-\hat{f}_k\| =$$

$$= \|B_{\ell+k}^{-1}\| \|\hat{B}_k^{-1}\| \|B_{\ell+k}-\hat{B}_k\| \|f_{\ell+k}\| + \|\hat{B}_k^{-1}\| \|f_{\ell+k}-\hat{f}_k\|$$

Assim

$$f_{\ell+k}-\hat{f}_k = \left[f(x_{\ell+k})-\hat{f}(x_{\ell+k})\right] + \left[\hat{f}(x_{\ell+k})-\hat{f}(x_k)\right]$$

e

$$\|f(x_{\ell+k})-\hat{f}(x_{\ell+k})\| = \|(f(x_{\ell+k})-f(x^*))-\hat{f}(x_{\ell+k})\| =$$

$$= \| \int_0^1 \left[f'(x^*+\tau(x_{\ell+k}-x^*))\right](x_{\ell+k}-x^*)d\tau-f'(x^*)(x_{\ell+k}-x^*\| =$$

$$= \| \int_0^1 \left[f'(x^*+\tau(x_{\ell+k}-x^*))-f'(x^*)\right](x_{+k}-x^*)d\tau\| \overset{(4.2)}{\leq}$$

$$\leq \lambda\|x_{\ell+k}-x^*\| \int_0^1 \|x_{\ell+k}-x^*\|d\tau \overset{(i)}{=} \frac{\lambda}{2}\|x_{\ell+k}-x^*\|^2 \leq \frac{\lambda}{2}h^2$$

Enquanto para $\gamma_{3,k} = \|A\|\gamma_{2,k}$ obtemos

$$\|\hat{f}(x_{+k})-\hat{f}(\hat{x}_k)\| = \|A(x_{\ell+k}-x^*)-A(\hat{x}_k-x^*)\| =$$

$$= \|A(x_{\ell+k}-\hat{x}_k)\| \leq \|A\| \|x_{\ell+k}-\hat{x}_k\| \leq \|A\| \gamma_{2,\ell} = \gamma_{3,k} h^2$$

Portanto por (4.13), (iv), (4.11) e (4.12a) concluímos que :

$$\|s_{\ell+k}-\hat{s}_k\| \leq \frac{4}{3}\|A^{-1}\|2.\|A^{-1}\|\gamma_{1,k}h^2 + 2\|A^{-1}\|\left(\frac{\lambda}{2}h^2 + \gamma_{3,k}h^2\right)$$

$$= \left(\frac{8}{3}\|A^{-1}\|^2\gamma_{1,k}+\|A^{-1}\|(\lambda+2\gamma_{3,k})\right)h^2 = \gamma_{4,k}\cdot h^2 \qquad (4.14)$$

E onde as condições (4.14) e (4.12b) para $j=k$ nos dá a condição (4.12b) para $j = k+1$ e com $\gamma_{2,k+1} = \gamma_{2,k} + \gamma_{4,k}$.

Com efeito, para $j = k+1$

$$\|s_{\ell+k}-\hat{s}_k\| = \|x_{\ell+k+1}-x_{\ell+k}-\hat{x}_{\ell+k}+\hat{x}_k\| \geq \|x_{\ell+k+1}-\hat{x}_{k+1}\| -$$

$$\|x_{\ell+k}-\hat{x}_k\| \Rightarrow \|x_{\ell+k+1}-\hat{x}_{k+1}\| \leq \|x_{\ell+k}-\hat{x}_k\| + \|s_{\ell+k}-\hat{s}_k\| \leq$$

$$\leq (\gamma_{2,k} + \gamma_{4,k})h^2 = \gamma_{2,k+1}\cdot h^2 .$$

Agora nos resta mostrar que as condições (4.14) e (4.12a) para $j=k$ implicam em (4.12a) para $j = k+1$.

Se $s_{\ell+k} = 0$ ou $\hat{s}_k = 0 \Rightarrow f_{\ell+k+1} = 0$ ou $\hat{f}_{k+1} = 0$ e de onde a condição (4.14) implica que

$$\|f_{\ell+k+1}\| \leq \left(\frac{\lambda}{2} + \gamma_{3,k+1}\right) h^2$$

e este juntamente com (ii) e (4.10) nos mostram que

$$\gamma_{5,k+1} = 3\delta(\lambda+2\gamma_{3,k+1})$$

e portanto

$$\|B_{\ell+k+1}-\hat{B}_{k+1}\| \, \|f_{\ell+k+1}\| \leq \gamma_{5,k+1}\, h^2$$

De fato

$$\|B_{\ell+k+1}-\hat{B}_{k+1}\| \ \|f_{\ell+k+1}\| = \|B_{\ell+k+1}-A+A-\hat{B}_{k+1}\| \cdot$$

$$\|f_{\ell+k+1}\| \leq \left[\|B_{\ell+k+1}-A\| + \|\hat{B}_{k+1}-A\|\right] \|f_{\ell+k+1}\| \leq$$

$$\leq 6\delta\left(\frac{\lambda}{2}+\gamma_{3,k+1}\right)h^2 = 3\delta\ (\lambda+2\gamma_{3,k+1})h^2 = \gamma_{5,k+1}\cdot h^2$$

Por outro lado, se ambos $s_{\ell+k} \neq 0$ e $\hat{s}_k \neq 0$, então pelas condições (4.1d), (4.9) e (iii) necessitamos somente mostrar que :

$$\|(y_{\ell+k}-B_{\ell+k}s_{\ell+k})s^T_{\ell+k} \ / \ (s^T_{\ell+k}s_{\ell+k})-(\hat{y}_k-\hat{B}_k\hat{s}_k)\hat{s}^T_k \ / \ (\hat{s}^T_k\ \hat{s}_k)\| \cdot$$

$$\cdot \|f_{\ell+k}\| \leq \gamma_{6,k}\cdot h^2 \qquad (4.15)$$

para que a condição (4.12a) seja verdadeira para $j = k+1$ com $\gamma_{1,k+1} = \text{máx}\ \{\gamma_{5,k+1}, \gamma_{1,k}+\gamma_{6,k}\}$.

Considere agora

$$A_k = \int_0^1 f'(x_{\ell+k}+\tau\ s_{\ell+k})d\tau$$

onde $y_{\ell+k} = A_k s_{\ell+k}$ e desde que $\hat{y}_k = A_k\hat{s}_k$ temos o seguinte limite para a parte esquerda de (4.15) = (PE 4.15)

$$(\text{PE }4.15) = \|(A_k-B_{\ell+k})s_{\ell+k}\ s^T_{\ell+k} \ / \ (s^T_{\ell+k}\ s_{\ell+k})-(A-\hat{B}_k)\hat{s}_k\ s^T_k \ /$$

$$/ \ (\hat{s}^T_k\ \hat{s}_k)\| \cdot \|f_{\ell+k}\| \leq$$

$$\leq \left\| \left[(A_k-A)-(B_{\ell+k}-\hat{B}_k)\right] s_{\ell+k}\ s^T_{\ell+k} \ / \ (s^T_{\ell+k}\ s_{\ell+k})\right\| \ \|f_{\ell+k}\| +$$

$$+ \|(A-\hat{B}_k)\left[s_{\ell+k}\, s_{\ell+k}^T / (s_{\ell+k}^T\, s_{\ell+k}) - s_k^T\, s_k^T / (\hat{s}_k^T\, \hat{s}_k)\right]\| \; \|f_{\ell+k}\| \tag{4.16}$$

Além disso ,

$$\|A_k - A\| \le \int_0^1 \|f'(x_{\ell+k} - \tau s_{\ell+k}) - f'(x^*)\|\, d\tau \le \lambda \int_0^1 \|x_{\ell+k} + \tau s_{\ell+k} - x^*\| d\tau$$

$$\le \lambda \|x_{\ell+k} - x^*\|$$

por (4.2), (i) e de

$$\|f_\ell\| = \|f(x_\ell) - f(x^*)\| \le \|f'(x^*)\| \; \|x_\ell - x^*\| \le (\|A\| + \lambda\delta/2)h = \gamma_7 \cdot h$$

onde $h = \|x_\ell - x^*\|$.

Assim de (i) e (iii) e da igualdade $\|s_{\ell+k}\, s_{\ell+k}^T / / (s_{\ell+k}^T\, s_{\ell+k})\| = 1$ chegamos que

$$\|A_k - A\| \le \lambda h$$

e

$$\|(A_k - A) s_{\ell+k}\, s_{\ell+k}^T / (s_{\ell+k}^T\, s_{\ell+k})\| \; \|f_{\ell+k}\| \le \lambda\, \gamma_7\, h^2$$

Mas por (4.12a)

$$\|B_{\ell+k} - \hat{B}_k\| \; \|f_{\ell+k}\| \le \gamma_{1,k}\, h^2$$

assim o primeiro termo a direita de (4.16) fica limitado por $(\lambda\, \gamma_7 + \gamma_{1,k})\, h^2$.

Ainda mais, por (4.10) e (ii) temos que :

$$\|A - \hat{B}_k\| \; \|f_{\ell+k}\| = \|A - \hat{B}_k\| \; \|B_{\ell+k}\, s_{\ell+k}\| \le 4\delta \|B_{\ell+k}\, s_{\ell+k}\| \le$$

$$\leq 4\delta \|B_{\ell+k}\| \; \|s_{\ell+k}\| = 4\delta \|A+B_{\ell+k}-A\| \; \|s_{\ell+k}\| \leq$$

$$\leq 4\delta (\|A\| + \|B_{\ell+k}-A\|) \; \|s_{\ell+k}\| \leq 4\delta(\|A\|+2\delta)\|s_{\ell+k}\| \leq$$

$$\leq \gamma_8 \|s_{\ell+k}\|$$

onde $\gamma_8 = 4\delta(\|A\| + 2\delta)$ e além disso,

$$\left\| \frac{s_{\ell+k}\; s_{\ell+k}^T}{s_{\ell+k}^T\; s_{\ell+k}} - \frac{\hat{s}_k\; \hat{s}_k^T}{\hat{s}_k^T\; \hat{s}_k} \right\| = \left\| \frac{s_{\ell+k}}{\|s_{\ell+k}\|} \left(\frac{s_{\ell+k}}{\|s_{\ell+k}\|} - \frac{\hat{s}_k}{\|\hat{s}_k\|} \right)^T + \right.$$

$$\left. + \left(\frac{s_{\ell+k}}{\|s_{\ell+k}\|} - \frac{\hat{s}_k}{\|\hat{s}_k\|} \right) \frac{\hat{s}_k^T}{\|\hat{s}_k\|} \right\| \leq 2 \left\| \frac{s_{\ell+k}}{\|s_{\ell+k}\|} - \frac{\hat{s}_k}{\|\hat{s}_k\|} \right\| =$$

$$= \frac{2}{\|s_{\ell+k}\|} \|(s_{\ell+k}-\hat{s}_k) + \|\hat{s}_k\| - (\|s_{\ell+k}\|\hat{s}_k \;/\; \|\hat{s}_k\|)\| \leq$$

$$\leq 4\|s_{\ell+k}-\hat{s}_k\| \;/\; \|s_{\ell+k}\| .$$

Assim de (4.14) concluímos que o segundo termo do lado direito de (4.16) limitado por $4\gamma_8\, \gamma_{4,k}$ e a condição (4.15) está satisfeita, sendo seu limite

$$\gamma_{6,k} = \lambda\, \gamma_7 + \gamma_{1,k} + 4\gamma_8\, \gamma_{4,k} \; .$$

Portanto a Condição (4.12a) vale para $j = k+1$ e por indução finita vemos que (4.12) vale para $j = 2n$.

Mas $\hat{x}_{2n} = x^*$ pelo Teorema 3.2, assim (4.4) está satisfeita para $\gamma = \gamma_{1,2n}$. cqd //

5) CONCLUSÕES DO CAPÍTULO

Os resultados para os métodos de Broyden obtidos por D. Gay (Teorema 2.2 deste Capítulo) foi sem sombra de dúvida uma agradável surpresa por seu resultado de terminação finita para sistemas lineares.

Este resultado, veio a acabar com o descrédito (demonstrado pelo próprio autor (Broyden)), que envolvia os métodos de Broyden (Bom e Ruim) .

Como podemos notar, o Teorema 3.1 deste Capítulo é um melhoramento do Teorema 3.2 do Capítulo anterior.

Enquanto o primeiro fala em convergência local e Q-superlinear, o Teorema 3.2 nos assegura uma convergência local e Q-quadrática.

Vale aqui salientar que a convergência Q-superlinear nos assegura que os progressos obtidos numa iteração k são maiores que aqueles obtidos na iteração anterior k-1.

Enquanto que a convergência Q-quadrática nos assegura o montante de progressos obtidos a cada iteração k .

CAPÍTULO III

EXPERIÊNCIAS NUMÉRICAS REALIZADAS E PROPOSTAS DE NOVOS ALGORÍTMOS

1) INTRODUÇÃO

Objetivando testar os algorítmos de Broyden Bom e Broyden Ruim, elaboramos programas em linguagem FORTRAN IV, cujos resultados serão apresentados neste Capítulo. Os testes foram realizados no computador PDP 1Ø da UNICAMP .

Nos nossos programas usamos :

i) COMO A PRIMEIRA APROXIMAÇÃO DO JACOBIANO INVERSO:
   A matriz identidade I_{nxn} .

ii) PRIMEIRA APROXIMAÇÃO x_0 DA SOLUÇÃO:
   Um vetor arbitrário.

iii) CRITÉRIO DE PARADA:
   Para Sistemas Lineares $Ax = b$ onde A é "pequeno" :
   $\|x^{k+1}-x^k\| \leq 1\emptyset^{-4}$.
   Nos casos restantes: $\|f(x)\| \leq 1\emptyset^{-4}$.

iv) NÚMERO MÁXIMO DE ITERAÇÕES: 5Ø (cinquenta) .

2) FUNÇÕES TESTES

Foram testadas as seguintes funções :

Nº Ø2 $\quad n = 3$

$f_1(x) = 3x_1+x_2+2x_3^2-3$.

$f_2(x) = -3x_1 + 5x_2^3 + 2x_1x_3 - 3$.

$f_3(x) = 25x_1x_2 + 20x_3 + 12$.

Vetor Inicial: $x^0 = (0, 0, 0)$

<u>Nº Ø3</u> $n = 2$

$f_1(x) = x_1^2 - x_2 - 1$

$f_2(x) = (x_1-2)^2 + (x_2-0.5)^2 - 1$

Vetores Iniciais: (0.1, 2)

(2, 0.5)

(1, 0.99)

<u>Nº Ø4</u> $n = 2$

$f_1(x) = -13+x_1+((-x_2+5)x_2-2)x_2)$.

$f_2(x) = -29+x_1+((x_2+1)x_2-14)x_2$.

Vetores Iniciais: (15, -2)

(-5, 0)

(-5, 3)

(0, 2.24)

<u>Nº Ø5</u> $n = 2$

$f_1(x) = x_1^2 - 2x_2 + 1$

$f_2(x) = x_1 + 2x_2^2 - 3$.

Vetores Iniciais: (0, 1) (1, -0.5)

(-0.5, 1) (1, -0.24)

Nº Ø6 $\quad n = 2$

$$f_1(x) = 10.000x_1x_2 - 1$$

$$f_2(x) = \exp(-x_1) + \exp(-x_2) - 10001$$

Vetores Iniciais: (0, 1)

(0, -1)

Nº 1Ø $\quad n = 2$

$$f_1(x) = x_1 - 1$$

$$f_2(x) = x_1x_2 - 1$$

Vetores Iniciais: (-1, 2)

(-1, -2)

(0.01, 0)

Nº 12 $\quad n = 2$

$$f_1(x) = 0.5\ \mathrm{sen}(x_1x_2) - x_2\ /\ (4\pi) - x_1/2$$

$$f_2(x) = (1-1/(4\pi)(\exp(2x_1)-e)+ex_2/-2ex_1$$

Vetores Iniciais: (0.6, 3)

(0.4, 3)

Nº 13 $\quad n = 6$

$$f(x) = \sum_{\substack{j=1 \\ j\neq i}}^{6} \cot\ (b_i x_j), \quad j = 1,2,\ldots,6$$

b = (0.02249; 0.02166; 0.02083; 0.02; 0.01918; 01835)

Vetor Inicial: (75, 75, 75, 75, 75, 75)

Nº 14 $n = 4$

$f_1(x) = 2(x_1+10x_2) + 40(x_1-x_4)^3$

$f_2(x) = 20(x_1+10x_2) + 4(x_2-2x_3)^3$

$f_3(x) = 10(x_3-x_4) - 8(x_2-2x_3)^3$

$f_4(x) = -10(x_3-x_4) - 40(x_1-x_4)^3$

Vetor Inicial: (3, -1, 0, 1)

Nº 18 $n = 9$

CHEBIQUAD, é uma função definida pelo programa fortran publicado em Cosnard, M.Y.: "A comparison of four methods for solving systems of nonlinear equations", TR 75-248, Dept. of Comp. Sci., Cornell Univ.

Vetor Inicial: $(1/(n+1);\ldots;n/(n+1)) = (0.1, 0.2,\ldots,0.9)$.

Nº 21 $n = 30$; $n = 40$

$f_i(x) = (3-0.1x_i)x_i + 1-x_{i-1} - 2x_{i+1}$

$f_1(x) = (3-0.1x_1) + 1-2x_2$.

$f_n(x) = (3-0.1x_n)x_n + 1-x_{n-1}$.

Vetor Inicial: (-1, -1,...,-1)

Nº 34 [Brown e Gearhart, 1971, p. 341] $n = 3$

$f_1(x) = (x_1)^2 + 2(x_2)^2 - 4$

$f_2(x) = (x_1)^2 + (x_2)^2 + x_3 - 8$

$f_3(x) = (x_1-1)^2 + (2x_2-\sqrt{2})^2 + (x_3-5)^2 - 4$

Vetor Inicial: (1, Ø.7, 5)

Vetor Solução: (Ø, $\sqrt{2}$, 6)

SISTEMAS LINEARES DE FUNÇÕES

Nº Ø1

$f_1(x) = 30x_1 + x_2 - 31$

$f_2(x) = x_1 + 10x_2 - 11$

Vetores Iniciais: (0,1), (0,2), (0,3), (0,0), (-1,0), (-2,0), (-3,0), (1,0), (2,0), (0,-1), (0,-2), (-1,-1), (0.5,0.5), (2,2), (3,3), (4,4), (5,5), (1,2), (2,3), (3,4), (-2,-1), (3.1,0), (0.85,3.7), (1.2,1.7), (-2,-2).

Nº Ø2

$f_1(x) = 900x_1 + 30x_2 - 930$

$f_2(x) = x_1 + 10x_2 - 11$

Vetores Iniciais: (-3,-3), (-2,-2), (-1,-1), (0,0), (2,2), (3,3), (4,4), (0,1), (0,2), (0,3), (0,-1), (0,-2) , (0,-3), (-3,0), (-2,0), (-1,0), (1,0), (2,0) .

Nº Ø3

$f_1(x) = 30x_1 + x_2 - 31$

$f_2(x) = 500x_1 + 5000x_2 - 5500$

Vetores Iniciais: (-4,-4), (-3,-3), (-2,-2), (-1,-1), (0,0), (2,2), (-1,0), (1,0), (2,0), (3,0), (4,0), (0,4), (0,3), (0,2), (0,1), (0,-1), (0,-2), (0,-3), (0,-4), (1,-4), (-3,1), (-4,1), (-2,1), (2,1), (3,1), (4,1).

N? 04

$f_1(x) = 900x_1 + 30x_2 - 330$

$f_2(x) = 500x_1 + 5000x_2 - 5500$

Vetores Iniciais: (-4,1), (-3,1), (-2,1), (-1,0), (1,-4), (1,-3), (1,-2), (4,1), (2,1), (1,4), (-4,4), (-4,-4), (4,4), (-3,-3), (3,3), (3,-3), (-3,3), (2,2), (-2,-2), (-2,2), (2,-2), (-1,1), (-1,-1), (1,1), (1,-1), (4,-4), (0,0) .

N? Ø5

$f_1(x) = 1000x_1 - 1$

$f_2(x) = 1000x_2 - 1$

Vetores Iniciais: (2,1), (3,1), (4,1), (2,3), (4,3), (4,-0.001).

N? Ø6

$f_1(x) = 1000x_1 - 1000$

$f_2(x) = 1000x_2 - 1000$

Vetores Iniciais: (1,-4), (2,-4), (3,-4), (-2,-2), (-3,-3), (-4,-4), (2,2), (3,3) .

Nº Ø7

$f_1(x) = 0.001x_1 + 0.002x_2 - 0.003$

$f_2(x) = 0.0001x_1 + 0.0005x_2 - 0.0006$

Vetores Iniciais: (-1,-1), (0,-2), (0,-3), (0,-4), (2,4),(2,-4), (-2,-2), (-3,-3), (0,0), (4,4), (2,2), (3,3) .

Nº Ø8

$f_1(x) = 0.00001x_1 + 0.000008x_2 - 0.000018$

$f_2(x) = 0.000007x_1 + 0.000001x_2 - 0.000008$

Vetores Iniciais: (4,4), (3,3), (2,2), (-1,-1), (0,0), (-2,-2), (-3,-3) .

Nº Ø9

$f_1(x) = 0.00001x_1 - 0.00001$

$f_2(x) = x_2 - 1$

Vetores Iniciais: (1,4), (1,3), (3,3), (2,2), (0,0), (-1,-1), (-2,-2), (-3,-3) .

5) RESULTADOS DOS PRIMEIROS TESTES

Na análise dos métodos de Broyden Bom e Ruim, testados para sistemas lineares e não lineares de equações, o objetivo - principal foi de investigar o porquê do primeiro método de Broyden (Bom) ser considerado superior ao segundo método (Ruim), se ambos possuem as mesmas propriedades matemáticas (tais como teoremas de convergência, etc.) .

Nos testes para sistemas não-lineares, comprovamos que realmente o primeiro método (Bom) é em geral superior ao segundo método (Ruim), como de certa forma já era esperado.

Um resultado que nos deixou muito otimistas foi o fato de que estas diferenças entre os métodos Bom e Ruim, ocorriam até mesmo para sistemas lineares de dimensões pequenas (por exemplo, 2x2). E isto de certa forma nos facilitou em muito na análise do comportamento dos dois métodos.

OBSERVAÇÕES:

Considere $\{H_k^B\}$ e $\{H_k^R\}$ respectivamente as aproximações do Jacobiano inverso para os métodos Bom e Ruim de Broyden [3] .

a) Em sistemas lineares da forma $f(x) = Ax-b$ onde $A \in R^{nxn}$ é uma matriz não-singular, notamos que quando A é "grande", isto é, seus componentes tiverem em módulo valores ≥ 1, $\|H_k^B - A^{-1}\| <<$ $\|H_k^R - A^{-1}\|$, ou seja, a aproximação H_k^B é bem melhor do que H_k^R . Enquanto que se a matriz A é "pequena", isto é, seus componentes em módulo tem valores $<<1$ (em maioria absoluta); a matriz H_k^R se aproxima mais do que H_k^B de $A^{-1} = f'(x)^{-1}$; isto é , $\|H_k^B - A^{-1}\| > \|H_k^R - A^{-1}\|$.

b) Ainda nos sistemas lineares, a supremacia do Broyden Bom ou Ruim ocorre devido a seguinte característica da matriz Jacobiana $A = f'(x)$.

b.1) Quando A é "grande", o método de Broyden Bom é superior ao método de Broyden Ruim.

b.2) Quando A se aproxima da matriz unidade, isto é, seus componentes em módulo tiverem valores próximos de 1, os dois métodos se equivalem.

b.3) Enquanto que se A é "pequena", o método Ruim é superior ao método Bom.

c) Através das observações (a) e (b) ficou-nos evidente que realmente a supremacia do método Bom (Ruim) se deve a uma melhor aproximação de H_k^B (H_k^R) do Jacobiano inverso $A^{-1} = f'(x)^{-1}$.

As conclusões acima de que o "tamanho" da matriz Jacobiana $A = f'(x)$ é fundamental para a superioridade de um ou outro método não nos trouxe de imediato previsões muito otimistas, pois nos métodos quase-Newton, vale recordarmos não nos é requerido o cálculo ou até mesmo a existência da matriz Jacobiana $A = f'(x)$.

Assim para que esta informação nos tornasse útil teríamos que encontrar uma matriz "calculável" pelos algorítmos de Broyden [3] e que fosse uma boa aproximação da matriz $A=f'(x)$.

Pela sua construção, obviamente o candidato a tal aproximação foi H_k^B (H_k^R) do Broyden Bom (Ruim).

E a possibilidade de em sistema linear $(f(x) = Ax-b)$ fazer-se uma comparação entre H_k^B (H_k^R) e $A = f'(x)$ surgiu das seguintes relações :

O algorítmo (2.1) do Capítulo 2 aplicado a $f(x) = Ax-b$

onde $A \in R^{nxn}$ é não singular, nos dá que

$$s_k = x_{k+1} - x_k$$

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

$$s_k = A^{-1} y_k$$

portanto $Y_k^T s_k \sim Y_k^T A^{-1} Y_k$ ($\sim$ denota semelhança) .

Assim ,

$$Y_k^T H_k Y_k \overset{\geq \text{ ou}}{<} Y_k^T s_k \sim Y_k^T A^{-1} Y_k$$

isto é,

$$\lambda_1 = Y_k^T H_k Y_k \overset{\geq \text{ ou}}{<} Y_k^T A^{-1} Y_k = \lambda_2$$

ou seja,

$$H_k \geq A^{-1} \quad \text{quando} \quad \lambda_1 \geq \lambda_2 \qquad \text{e}$$

$$H_k < A^{-1} \quad \text{quando} \quad \lambda_1 < \lambda_2$$

Baseado nestas informações construímos o primeiro método do conjugado (Broyden Bom e Ruim) chamado "ZORRA" o qual estabelece o seguinte critério para calcular a matriz H_k, $k \geq 1$.

Considere $\mu_1 = Y_k^T H_k Y_k$ e $\mu_2 = Y_k^T s_k$.

i) Se $\mu_1 \geq \mu_2$ => A^{-1} é considerado uma matriz "pequena" => A é uma matriz "grande" => H_k é calculada pelo método de Broyden Bom .

ii) Se $\mu_1 < \mu_2$ => A^{-1} é uma matriz "grande" => A é uma matriz "pequena" => H_k é calculada pelo método de Broyden Ruim.

A seguir damos os resultados dos testes efetuados para os métodos de Broyden Bom, Ruim e ZORRA, onde LFUN denota o número da função teste dado anteriormente, e N.C significa que o método não convergiu em 5Ø iterações.

TABELA 1: SISTEMAS LINEARES DE EQUAÇÕES

função	vetor inicial	bom nº iter.	ruim nº iter.	bom $\|H-f'(x)^{-1}\|$	ruim $\|H-f'(x)^{-1}\|$
$f_1=30x_1+x_2-31$	(0;1)	04	04	0.00037417	0.0144474863
$f_2=x_1+10x_2-11$	(0;2)	04	05	0.01100450	0.8926229642
"	(0;3)	04	07	0.02241267	0.4949016142
"	(0;0)	04	05	0.00767346	0.2421245000
"	(-1;0)	04	05	0.00358641	0.1191436406
"	(-2;0)	04	04	0.00218625	0.0737849699
"	(-3;0)	04	04	0.00152645	0.0737849699
"	(1;0)	04	05	0.02846179	0.6029957431
"	(2;0)	04	05	0.01100097	0.2969262253
"	(0;-1)	04	06	0.01834598	0.4139347635
"	(0;-2)	04	07	0.02747174	0.0875213844
"	(-1;-1)	04	05	0.00669729	0.0119216406
"	(0.5;0.5)	04	05	0.00669729	0.2421245000
"	(2;2)	04	05	0.00669729	0.2421245000
"	(3;3)	04	05	0.00669729	0.2421245000
"	(4;4)	04	05	0.00669729	0.2421247000
"	(5;5)	05	05	0.00669729	0.2419194000
"	(1;2)	04	05	0.02846179	0.6029957431
"	(2;3)	04	06	0.01834598	0.4139347635
"	(3;4)	04	07	0.01253693	0.3362363486
"	(3.1;0)	04	05	0.00453422	0.2284008360
"	(0.85;3.7)	05	05	0.00288353	0.1370635747
"	(1.2;1.7)	04	04	0.03071829	0.5596225133
$f_1=900x_1+30x_2-930$	(-3;-3)	05	05	0.01121437	0.8485450215
$f_2=x_1+10x_2-11$	(-2;-2)	05	06	0.01121418	0.8483951331
"	(-1;-1)	05	05	0.01121437	0.8485450215
"	(0;0)	05	05	0.01121437	0.8485450315
"	(0;1)	05	05	0.01130928	0.3034691063
"	(-4;1)	05	05	0.01091921	0.3048897922
"	(-2;1)	05	05	0.09989401	0.8687275188
$f_1=30x_1+x_2-31$					
$f_2=500x_1+5000x_2-5500$	(-4;-4)	05	10	0.00001000	0.0964862041
"	(-3;-3)	05	11	0.00000586	0.9648529762
"	(-2;-2)	05	14	0.00000661	0.9649010413
"	(-1;-1)	05	20	0.00000586	0.9646350439
"	(0;0)	05	24	0.00000586	0.9647515592
"	(2;2)	05	18	0.00000586	0.9640680660

$f_1=30x_1+x_2-31$					
$f_2=500x_1+5000x_2-5500$	(3;3)	05	21	0.0000058	0.0964506
"	(-1;0)	05	13	0.0000200	0.9643120
"	(1;0)	05	09	0.0012179	0.9637823
"	(2;0)	05	27	0.0000364	0.9637124
"	(3;0)	06	14	0.0000721	0.9625832
"	(4;0)	06	N.C	0.0001666	0.9628301
"	(0;4)	05	18	0.0000200	0.0048882
"	(0;3)	05	17	0.0000223	0.0130908
"	(0;2)	05	14	0.0000360	0.0034494
"	(0;1)	06	29	0.0000049	0.9641080
"	(0;-1)	05	09	0.0000003	0.9649221
"	(0;-2)	05	14	0.0000061	0.0044012
"	(0;-3)	05	10	0.0000029	0.9649289
"	(0;-4)	05	N.C	0.0000081	0.9606484
"	(1;-4)	05	13	0.0012173	0.9634890
"	(-3;1)	06	14	0.0000052	0.9648429
"	(-4;1)	06	06	0.0000566	0.9649306
"	(-2;1)	05	N.C	0.0000244	0.9648064
"	(2;1)	06	10	0.0000049	0.9649255
"	(3;1)	06	08	0.0000049	0.9649263
"	(4;1)	05	25	0.0000249	0.9644541
$f_1=900x_1+30x_2-330$					
$f_2=500x_1+5000x_2-5500$	(-4;1)	07	N.C	0.0004408	0.5861905
"	(-3;1)	07	N.C	0.0004343	0.5672171
"	(-2;0)	06	N.C	0.0004212	0.2134191
"	(-1;0)	07	N.C	0.0000252	0.3167966
"	(1;-4)	07	43	0.0000027	0.4800488
"	(1;-3)	06	N.C	0.0000034	0.4812931
"	(1;-2)	07	N.C	0.0000067	0.4823628
"	(2;1)	06	N.C	0.0004849	0.7249230
"	(1;4)	07	N.C	0.0000018	0.4357017
"	(-4;4)	07	N.C	0.0001434	0.5965231
"	(-4;-4)	07	N.C	0.0000034	0.3186566
"	(4;4)	07	N.C	0.0000252	0.3186566
"	(-3;-3)	07	N.C	0.0000104	0.3205960
"	(3;3)	07	N.C	0.0000316	0.3009180
"	(3;-3)	07	N.C	0.0000283	0.5409056
"	(-3;3)	07	N.C	0.0001439	0.6630252
"	(2;2)	06	N.C	0.0000470	0.2564960
"	(-2;-2)	06	N.C	0.0000065	0.3721726
"	(-2;2)	07	N.C	0.0002454	0.7392293
"	(2;-2)	07	N.C	0.0000057	0.4688643

$f_1=900x_1+30x_2-330$					
$f_2=500x_1+5000x_2-5500$	(-1;1)	07	N.C	0.0003879	0.3105751
"	(-1;-1)	07	N.C	0.0000037	0.3856503
"	(1;1)	06	N.C	0.0004736	0.6546360
"	(1;-1)	07	N.C	0.0000141	0.5014634
"	(4;-4)	07	N.C	0.0000332	0.3846757
"	(0;0)	07	N.C	0.0000019	0.6893725
$f_1=900x_1+30x_2-930$					
$f_2=30x_1+300x_2-330$	(-4;-4)	07	09	0.0002030	0.2541848
"	(-3;-3)	07	10	0.0001957	0.2436228
"	(-2;-2)	06	12	0.0002253	0.2540482
"	(-1;-1)	07	10	0.0001960	0.2433814
"	(0;0)	06	10	0.0001963	0.2449734
"	(2;2)	07	10	0.0001956	0.2434308
"	(3;3)	07	10	0.0001958	0.2433313
"	(4;4)	06	10	0.0002253	0.2479888
"	(0;1)	06	07	0.0000143	0.0146833
"	(0;2)	07	21	0.0003184	0.3032594
"	(0;3)	07	10	0.0006736	0.5256924
"	(0;4)	07	42	0.0009750	0.6518023
"	(0;-1)	07	43	0.0005578	0.4298178
"	(0;-2)	07	11	0.0008589	0.5631800
"	(0;-3)	06	12	0.0010368	0.6331006
"	(0;-4)	07	40	0.0011647	0.6474847
$f_1=1000x_1-1$					
$f_2=1000x_2-1$	(2;1)	05	06	0.0000003	0.0005509
"	(3;1)	05	06	0.0000023	0.0087021
"	(4;1)	05	06	0.0000006	0.0032681
"	(2;3)	05	06	0.0000007	0.0044291
"	(4;3)	05	06	0.0000013	0.0045599
"	(10;-0.001)	05	07	0.0001822	0.1344165
$f_1=0.001x_1+0.002x_2-0.003$					
$f_2=0.0001x_1+0.0005x_2-0.0006$					
"	(-1;-1)	08	05	2530959.1	360.62179
"	(0;-2)	08	06	2642231.0	1747.4314
"	(0;-3)	09	05	311468.90	606.16799
"	(0;-4)	08	05	2861618.9	249.67852

$$\begin{cases} f_1=0.001x_1+0.002x_2-0.003 \\ f_2=0.0001x_1+0.0005x_2-0.0006 \end{cases}$$

"	(2;4)	09	06	2345575.1	807.26008
"	(2;-4)	08	06	2442244.2	1389.6233
"	(-2;-2)	08	06	2681860.8	902.62403
"	(-3;-3)	08	05	2564093.2	1002.8328
"	(0;0)	07	05	2802103.8	513.96753
"	(4;4)	09	06	2538503.7	890.60784
"	(2;2)	09	05	2781762.8	523.00085
"	(3;3)	07	05	2571786.8	368.92859

$$\begin{cases} f_1=0.00001x_1+0.000008x_2-0.000018 \\ f_2=0.000007x_1+0.000001x_2-0.000008 \end{cases}$$

"	(4;4)	44	08
"	(3;3)	23	07
"	(2;2)	17	06
"	(-1;-1)	26	07
"	(0;0)	37	06
"	(-2;-2)	32	08
"	(-3;-3)	50	06

$$\begin{cases} f_1=0.00001x_1-0.00001 \\ f_2=x_2-1 \end{cases}$$

	(1;4)	01	01
"	(1;3)	01	01
"	(3;3)	04	04
"	(2;2)	04	04
"	(0;0)	04	04
"	(-1;-1)	04	04
"	(-2;-2)	04	04
"	(-3;-3)	04	04

$$\begin{cases} f_1=10000x_1-10000 \\ f_2=x_2-1 \end{cases}$$

	(-3;-3)	03	03
"	(3;3)	03	03
"	(2;2)	03	03
"	(0;0)	03	03
"	(-1;-1)	03	03

TABELA 2: SISTEMAS NÃO-LINEARES DE EQUAÇÕES

LFUN	VETOR INICIAL	BROYDEN BOM	BROYDEN RUIM	ZORRA
02	(0,0,0)	N.C	N.C	N.C
03	(0.1,2)	11	12	11
03	(2,0.5)	15	14	12
03	(-1,1.5)	19	12	15
03	(1,0.99)	07	07	07
04	(15,-2)	N.C	N.C	N.C
04	(-5,0)	18	N.C	16
04	(-5,3)	41	N.C	17
04	(0,2.24)	N.C	N.C	N.C
04	(2,0.5)	27	N.C	12
05	(0,1)	09	10	09
05	(-0.5,1)	12	08	08
05	(1,-0.5)	N.C	N.C	N.C
05	(1,-0.24)	09	10	09
06	(0,1)	39	N.C	38
06	(0,-1)	37	N.C	N.C
10	(-1,2)	03	03	03
10	(-1,-2)	03	03	03
10	(0.01,0)	01	01	01
12	(0.6,3)	10	N.C	10
12	(0.4,3)	08	N.C	09
13	(75,75,75,75,75,75)	N.C	27	27
14	(3,-1,0,1)	N.C	N.C	N.C
18	(0.1,0.2,...,0.9)	26	N.C	N.C
21	n=30 (-1,...,-1)	N.C	N.C	N.C
21	n=40 (-1,...,-1)	N.C	N.C	N.C

4) NOVOS ALGORÍTMOS COMBINADOS

Tanto no método de Broyden Bom como no Broyden Ruim, sabemos que

$$H_{k+1}\, y_k = s_k$$

onde $s_k = x_{k+1} - x_k$ e $y_k = f(x_{k+1}) - f(x_k)$ para todo $k = 0,1,2...$

Porém, não se satisfaz automaticamente que

$$H_{k+1}\, y_{k-1} = s_{k-1} \tag{1}$$

o qual seria uma propriedade desejável, já que é satisfeita pela inversa do Jacobiano se a função é linear.

Agora

$$H_{k+1}\, y_{k-1} = (H_k + \Delta H_k) y_{k-1} = \Delta H_k y_{k-1} + s_{k-1}$$

para todo $k = 0,1,2,...$ já que a equação $H_k y_{k-1} = s_{k-1}$ é sempre satisfeita.

Logo, achamos que o método de Broyden Bom será melhor que o método de Broyden Ruim quando

$$\|(\Delta H_k)_{BOM}\, y_{k-1}\| < \|(\Delta H_k)_{RUIM}\, y_{k-1}\| \tag{2}$$

porque isto equivale a

$$\|\overset{BOM}{H_{k+1}}\, y_{k-1} - s_{k-1}\| < \|\overset{RUIM}{H_{k+1}}\, y_{k-1} - s_{k-1}\|$$

Desenvolvendo (2), e sabendo que

$$(\Delta H_k)_{BOM} = (s_k - H_k y_k)\, s_k^T H_k \,/\, (s_k^T H_k\, y_k)$$

$$(\Delta H_k)_{RUIM} = (s_k - H_k y_k)\, y_k^T \,/\, (y_k^T\, y_k)$$

Assim o método de Broyden Bom é melhor que o método de Broyden Ruim se e só se ,

$$\left\| \frac{(s_k - H_k y_k) s_k^T H_k y_{k-1}}{s_k^T H_k y_k} \right\| < \left\| \frac{(s_k - H_k y_k) y_k^T y_{k-1}}{y_k^T y_k} \right\| \qquad (2.b)$$

Portanto o Broyden Bom será melhor que o Broyden Ruim se e só se ,

$$\left| \frac{s_k^T H_k y_{k-1}}{s_k^T H_k y_k} \right| < \left| \frac{y_k^T y_k^{-1}}{y_k^T y_k} \right|$$

ou seja, quando

$$\left| \frac{s_k^T s_{k-1}}{s_k^T H_k y_k} \right| < \left| \frac{y_k^T y_{k-1}}{y_k^T y_k} \right| \qquad (3)$$

Tanto (2.b) como (3) nos fornecem critérios para elaborar um algorítmo híbrido, sendo que (2.b) parece ser mais seguro que (3), enquanto que (3) mais barato que (2.b).

Considerando (3) em relação ao caso linear $(f(x)=Ax-b)$ $(y_k = As_k)$, temos que o Broyden Bom é melhor que o Broyden Ruim, provavelmente quando

$$\left| \frac{s_k^T s_{k-1}}{s_k^T H_k A s_k} \right| < \left| \frac{s_k^T A^T A s_{k-1}}{s_k^T A^T A s_k} \right| \qquad (4)$$

desigualdade que nos parece fornecer a explicação do porquê de quando A é "pequena", o Broyden Ruim funciona melhor e vice-versa.

Quando $A \sim 0$, o membro a esquerda de (4) aumenta muito porém o mesmo não ocorre com o membro à direita.

Quando $A \sim \infty$ o membro à esquerda de (4) diminui muito e o da direita não.

Além disso, notamos que quando $A \to \infty$ o membro da esquerda de (4) tende a zero, enquanto o da direita nunca tende a zero.

Isso explicaria porque o método Bom tem possibilidade de ser excelente, e o Ruim não.

Este fato justificaria o porquê do primeiro método de Broyden ser conhecido como Bom e o segundo método de Broyden como Ruim.

Baseado nos critérios (2.b) e (3) acima foram escritos dois algorítmos combinados (Broyden Bom e Ruim) chamados respectivamente de teste 1 e teste 2 .

Damos abaixo os resultados de testes efetuados para ambos em sistemas de equações não-lineares de funções.

OBSERVAÇÕES:

LFUN = Número da função teste.

"NC" = Significa que o método não convergiu em 5Ø iterações efetuadas.

TABELA 3: SISTEMAS NÃO LINEARES

LFUN	VETOR INICIAL	Nº Iter. BOM	Nº Iter. RUIM	Nº Iter. ZORRA	Nº Iter. TESTE 1	Nº Iter. TESTE 2
02	(0,0,0)	N.C	N.C	N.C	36	36
03	(0,0)	12	12	12	12	12
03	(-1,1.5)	19	12	15	14	14
03	(1,0.99)	07	07	07	07	07
03	(2,0.5)	15	14	12	11	11
03	(0.1,2)	11	12	11	10	10
04	(-5,0)	18	N.C	16	14	14
04	(-5,3)	41	N.C	17	19	19
04	(15,-2)	N.C	N.C	N.C	41	41
04	(0,2.24)	N.C	N.C	N.C	34	34
04	(2,-0.5)	27	N.C	12	12	12
05	(0.5,1)	12	08	08	08	08
05	(0,1)	09	10	09	09	09
05	(1,-0.5)	N.C	N.C	N.C	N.C	N.C
05	(1,-0.24)	09	10	09	09	09
06	(0,1)	39	N.C	38	31	31
06	(0,-1)	37	N.C	N.C	33	33
10	(-1,2)	03	03	03	03	03
10	(-1,-2)	03	03	03	03	03
10	(0.01,0)	01	01	01	01	01
12	(0.4,3)	08	N.C	09	08	08
12	(0.6,3)	10	10	10	17	17
13	(75,75,75,75,75,75)	N.C	27	27	25	25
14	(3,-1,0,1)	N.C	N.C	N.C	N.C	N.C
18	(0.1,0.2,...,0.9)	26	N.C	N.C	N.C	N.C
21	n=30 (-1,...,-1)	N.C	N.C	N.C	49	49
21	n=40 (-1,...,-1)	N.C	N.C	N.C	N.C	N.C
34	(1,0.7,5)	37	N.C	19	19	19

CONCLUSÕES

Apesar de não ser necessário nos teoremas de Gay, a aproximação de H_k a A^{-1}, essa aproximação revelou-se a principal explicação das diferenças de performance dos métodos de Broyden Bom e Broyden Ruim.

A razão para isso é que na presença de erro de arredondamento, os valores exatos de x_k não são obtidos; agora, se H_k é uma boa aproximação de $f'(x)^{-1}$, o ponto x_{k+1} será em geral, uma aproximação melhor, devido à sua semelhança com a iteração do método de Newton.

Ou seja, enquanto nos teoremas de Gay, a qualidade da aproximação x_{k+1} depende da "história", se tomarmos a qualidade de H_k como estimativa de $f'(x)^{-1}$ então x^{k+1} será independente dessa "história" .

A hipótese de que as diferenças entre o Broyden Bom e o Broyden Ruim se refletiam ainda para sistemas lineares foi totalmente comprovada.

A observação sobre o comportamento dos métodos em sistemas lineares de dimensões pequenas permitiu-nos chegar às principais conclusões deste trabalho.

Em primeiro lugar, verificamos empiricamente que o Broyden Bom se comportava melhor que o Broyden Ruim (vice-versa) para sistemas com Jacobianos grandes (pequenas) .

A razão para tal comportamento foi revelada na análise do Capítulo III.

Esta análise sugeriu que seria melhor aquele método para o qual a diferença $\|H_{k+1} y_{k-1} - s_{k-1}\|$ fosse menor .

Baseado nesta suposição, elaboramos dois (2) algorítmos combinados que revelaram na prática serem superiores tanto ao Broyden Bom bem como ao Broyden Ruim .

Além disso, esta análise forneceu a explicação do porquê, em geral, o folclore atribui ao Broyden Bom melhores qualidades que ao Broyden Ruim.

Uma palavra para sistemas de grande porte.

As formas mais lógicas de adaptar métodos do tipo quase-Newton a problemas grandes é trabalhar com os chamados "métodos com memória limitada" .

Esses métodos consistem no seguinte: assumimos que a memória (da máquina) disponível NÃO é suficiente para armazenar uma matriz de dimensão (nxn), porém ela é suficiente para armazenar 2P vetores de n posições. Então, a atualização da matriz Jacobiana usada se faz da seguinte forma:

$$H_0 = I$$

$$H_1 = H_0 + v_1 w_1^T$$

$$H_2 = H_0 + v_1 w_1^T + v_2 w_2^T$$

$$\vdots$$

$$H_p = H_0 + v_1 w_1^T + v_2 w_2^T + \quad + v_p w_p^T$$

$$H_{p+1} = H_0 + v_2 w_2^T + v_3 w_3^T + \quad + v_{p+1} w_{p+1}^T$$

$$\vdots$$

$$\vdots$$

$$H_k = H_0 + v_{k-p+1} w_{k-p+1}^T + \quad + v_k w_k^T$$

$$\vdots$$

Essas matrizes não são calculadas explicitamente (o que demandaria n^2 posições). Os vetores $v_k w_k$, $v_{k-1} w_{k-1}, \ldots v_{k-p+1}$, w_{k-p+1} são armazenados usando a memória disponível.

Dessa maneira, podemos estender todas as nossas conclusões sobre os métodos de Broyden a sistemas não lineares de grande porte, e generalizar facilmente os algorítmos combinados .

BIBLIOGRAFIA

[1] ORTEGA, J.M.; RHEINBOLT, W.C. (1970): "Iterative solution of nonlinear equation in several variavles". Academic Press, New York.

[2] BRENT, R. (1973): "Some efficient algorithms for solving systems of nonlinear equations". SIAM J. Number Anal. 10, pp. 327-344.

[3] BROYDEN, C.G. (1965): "A class of methods for solving nonlinear simultaneous". Math. Comp. 19, pp. 577-593.

BROYDEN, C.G. (1970): "The convergence of single-rank quasi-Newton methods". Math. Comp. 24, pp. 365-382.

BROYDEN, C.G. (1971): "The convergence of an algorithm for solving space nonlinear-systems". Math. Comp. 25, pp. 285-294.

BROYDEN, C.G.; DENNIS, J.E. and MORÉ, J.J. (1973): "On the local and superlinear convergence of quasi-newton methods". J. Inst. Math. Appl. 12, pp. 223-245.

[4] DAVIDON, W.C. (1959): "Variable metric method for minimization". Argonne National Lab.,report ANL-5990 REV.

[5] FLETCHER, R. and POWELL, M.J.D. (1963): "A rapidly convergent descent method for minimizations". Comp. J.6, pp.163-168.

[6] GAY; DAVID M.: "Some convergence properties of Broyden's method". SIAM J. Numerical Analysis, vol. 16, nº 4 agosto 1979.

[7] MORÉ, J.J.; TRANGENSTEIN, J.A. (1976): "On the global convergence of Broyden's method". Math. Comp. 30, pp.

523-540.

[8] SHERMAN, J.; MORRISON, W.J. (1949): "Adjustment of an inverse matrix corresponding to changes in the elements of a given colun or a given row of the original matrix". Ann. Math. Sta. 20, pp. 621.

[9] MARTÍNEZ, J.M. (1978): "On the order of convergence of Broyden-Gay-Schnabel's method". Commentationes Mathematicae, Univ. Carolinae, Vol. 16, nº 3.

[10] MARTÍNEZ, J.M.; LOPES, T.L. (1980): "Combination of the sequential secant method and Broyden's method with projected updates". Computing, Vol. 25, nº 4, 1980.

[11] MOLER, CLEVE B.; FORSYTHE, GEORGE E. (1967): "Computer solution of linear algebraic systems". Prentice-Hall, Series in Automatic Computation.

[12] GAY, DAVID M. and SCHNABEL, ROBERT B. (1978): "Solving systems of nonlinear equations by Broyden's methods with projected updates". C. Ac. Press, Inc. Nonlinear Programming-3.

[13] MORÉ, JORGE, J.; DENNIS, J.E.; BROYDEN, C.G. (1973): "On the local and superlinear convergence of quasi-newton methods". Dept. C. Sc.,Cornell University , ITHA CA, N.Y. 15850, U.S.A.