



Departamento de Matemática

Matemática Numérica II

Trabalho Prático nº2

Marta Sofia Pimentel Cavaleiro

Exercício 1

a) Para calcular os possíveis extremantes da função de Rosenbrock,

$$f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

vamos fazer uso do seguinte resultado:

Se $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}$, D aberto, for duas vezes continuamente diferenciável em D , se $x^ \in D$ tal que $\nabla f(x^*) = 0$ e $\nabla^2 f(x^*)$ for definida positiva, então x^* é um minimizante local de f .*

Calculamos então os pontos do domínio onde o vector gradiente se anula, esses vão ser os pontos críticos onde os extremos poderão existir.

$$\nabla f(x_1, x_2) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 400x_1^3 - 400x_1x_2 + 2x_1 - 2 \\ 200(x_2 - x_1^2) \end{bmatrix}$$

$$\nabla f(x_1, x_2) = 0 \Leftrightarrow \begin{bmatrix} 400x_1^3 - 400x_1x_2 + 2x_1 - 2 \\ 200(x_2 - x_1) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Leftrightarrow \begin{cases} x_1 = 1 \\ x_2 = 1 \end{cases}$$

Ora obtemos então um único candidato a extremo e portanto a minimizante: $x^* = (1, 1)$.

Verifiquemos se a matriz Hessiana é definida positiva em $x^* = (1, 1)$. A matriz Hessiana da função f é:

$$\nabla^2 f(x_1, x_2) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}$$

e será definida positiva se para qualquer vector $y \neq 0$ se verificar que $y^T A y > 0$.

$$\nabla^2 f(1,1) = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix}$$

Seja $y = (y_1, y_2) \in \mathbb{R}^2$,

$$\begin{aligned} y^T \nabla^2 f(1,1) y &= [y_1 \ y_2] \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 802y_1^2 - 800y_1y_2 + 200y_2^2 = \\ &= 2y_1^2 + 200(4y_1^2 - 4y_1y_2 + y_2^2) = 2y_1^2 + 200(2y_1 - y_2)^2 > 0 \end{aligned}$$

Verifica-se que a matriz é definida positiva e a função f é duas vezes diferenciável, uma vez que as suas derivadas até à 2ª ordem são contínuas, como facilmente se constata, portanto concluímos que existe um minimizante no ponto (1,1) e sendo que é o único ponto em que o vector gradiente se anula, é o único extremante existente.

b) Para determinar o minimizante de f pelo Método de Newton, aplicou-se o Método de Newton ao sistema de equações não lineares $\nabla f(x)=0$. Como aproximação inicial usou-se o ponto $x_0 = (-1.2, 1)$.

Para determinar o minimizante pelo Método BFGS, o sistema de equações a resolver é o mesmo, e usou-se como aproximação inicial o mesmo ponto, e $B_0 = (\nabla^2 f(x_0))^{-1}$.

Para executar o Método BFGS criou-se o seguinte programa:

```
function z = bfgs_optim (funcao, z_0, z_exacta, e)

syms x y;
gradiente = rot90(jacobian (sym(funcao), [x, y]), -1);
hessiana = jacobian (sym(gradiente), [x, y]);
B = inv(double (subs(hessiana, {x,y}, z_0)));
z = z_0;
F = double (subs(gradiente, {x,y}, z))
iter = 1;

while (norm (F, inf)>e && iter<=40)
    disp (['iteracao: ', num2str(iter)]);
    s = -B*F;
    z = z + s
    G = double (subs(gradiente, {x,y}, z));
    w = G - F;
    F = G;
    p = 1/(w'*s);
    B = (eye (2) -p*s*w')*B*(eye (2) -p*w*s')+p*s*s';
    erro = z_exacta-z;
    norma = norm ((z_exacta-z), inf);
    disp (['norma infinito do erro: ', num2str(norma)]);
    iter = iter +1;
end
```

A condição de paragem ao fim de 40 iterações foi criada propositadamente para este problema uma vez que queríamos a execução de 40 iterações, consequentemente correu-se o programa com $e = 0$, para que a condição de a norma do gradiente ser maior que e não interferisse.

Correram-se então os dois programas e para controlo do erro, e de rapidez de convergência, calculou-se a norma infinito do erro em cada iteração em ambos os métodos. Esses valores estão reunidos na seguinte tabela:

Iteração k	$\ x^* - x_k\ _\infty$	
	Método de Newton	Método BFGS
1	2.17530	2.17530
2	4.1750	2.15070
3	0.41718	4.13030
4	0.05597	2.14620
5	8.60870e-006	2.14330
6	1.85280e-011	8.62770
7	1.11020e-016	2.13900
8	1.11020e-016	2.09530
9	1.11020e-016	2.14270
10	1.11020e-016	2.14210
11	1.11020e-016	2.09880
12	1.11020e-016	1.96310
13	1.11020e-016	2.54250
14	1.11020e-016	1.90190
15	1.11020e-016	1.84940
16	1.11020e-016	1.77190
17	1.11020e-016	1.77910
18	1.11020e-016	1.73620
19	1.11020e-016	1.58920
20	1.11020e-016	1.21950
21	1.11020e-016	1.54260
22	1.11020e-016	1.50710
23	1.11020e-016	3.05200
24	1.11020e-016	1.50560
25	1.11020e-016	1.50410
26	1.11020e-016	5.16750
27	1.11020e-016	1.50380
28	1.11020e-016	1.50350
29	1.11020e-016	1.48250
30	1.11020e-016	1.46480
31	1.11020e-016	1.38820
32	1.11020e-016	1.35120
33	1.11020e-016	1.21440
34	1.11020e-016	1.05420
35	1.11020e-016	1.02070
36	1.11020e-016	1.00970
37	1.11020e-016	0.98745
38	1.11020e-016	0.89493
39	1.11020e-016	0.75011
40	1.11020e-016	0.89277

Verificou-se que ao fim de 7 iterações do Método de Newton o minimizante foi encontrado e a norma do erro manteve-se constante a partir daí, já no Método BFGS ao fim de 40 iterações o minimizante não tinha sido encontrado mas foi-se verificando uma diminuição da norma do erro. De facto, o minimizante só viria a ser encontrado na iteração 81.

c) Para analisar o número de iterações necessárias em cada método para reduzir a norma do gradiente para a ordem de 10^{-5} , adicionou-se aos métodos a condição de pararem quando tal se verificasse e contaram-se as iterações.

Na tabela seguinte esquematizam-se os resultados:

	Método de Newton	Método BFGS
Iteração em que foi atingida a ordem da norma do gradiente $<10^{-5}$	6	77
Valor da norma do gradiente nessa iteração (15 c.d.)	8.608633446552801e-006	2.979402912770157e-006
Valores de x nessa iteração (15 c.d.)	(1.0000000000000000, 0.999999999981472)	(1.000000015631851, 1.000000023893355)

Exercício 2

Sabemos que se uma matriz A é simétrica então é diagonalizável, isto é, existe uma matriz U invertível tal que $UDU^{-1} = A$. Sendo A simétrica sabemos mais: U é ortogonal ($U^{-1} = U^T$) e constituída pelas bases ortonormadas dos subespaços próprios, e D é a matriz cujos elementos diagonais são os valores próprios de A .

Sendo A definida positiva então os seus valores próprios são positivos:

$$\forall p \neq 0 \in \mathbb{R}^n \quad p^T(Ap) > 0 \implies p^T(\lambda p) > 0 \iff \lambda > 0 \text{ uma vez que } \forall p \neq 0 \quad p^T p > 0$$

Queremos então calcular uma raiz quadrada de A , $A^{1/2}$ tal que $A^{1/2}A^{1/2} = A$. Observamos que a raiz quadrada de uma matriz diagonal de elementos positivos é imediata: se $D = [a_{i,i}]$ então $D^{1/2} = [\sqrt{a_{i,i}}]$, $i = 1, \dots, n$.

Então verificamos que,

$$A = UDU^{-1} = U D^{1/2} D^{1/2} U^{-1} = (U D^{1/2} U^{-1})(U D^{1/2} U^{-1})$$

O que nos leva a concluir que uma matriz raiz quadrada de A é $U D^{1/2} U^{-1}$.

Provemos agora que sendo A simétrica e definida positiva, então a raiz quadrada, definida do modo acima, também o é.

$A^{1/2} = U D^{1/2} U^{-1}$ é simétrica, pois resulta do produto de uma matriz com a sua transposta (a matriz diagonal não interfere na simetria uma vez que apenas altera os elementos da diagonal principal).

Seja agora então $p \neq 0 \in \mathbb{R}^n$, vamos provar que $A^{1/2} = U D^{1/2} U^{-1}$ é definida positiva.

A é definida positiva então $p^T(UDU^{-1})p > 0$, isto é, $p^T(U D^{1/2} D^{1/2} U^{-1})p > 0$. $D^{1/2} = [\sqrt{a_{i,i}}]$ é constituída por elementos positivos logo se a retirarmos o restante produto tem de permanecer positivo: $p^T U D^{1/2} U^{-1} p > 0$, logo $A^{1/2}$ é definida positiva.

Exercício 3

Pelo Teorema da Forma Normal de Jordan, sabemos que qualquer matriz B é semelhante a uma outra, chamemos-lhe T , constituída por blocos triangulares superiores cuja diagonal é constituída pelos valores próprios de B . Sabemos que matrizes semelhantes têm igual traço e igual determinante, e como T é triangular superior temos que:

$$\det(B) = \det(T) = \prod_{i=1}^n \lambda_i \quad \text{e} \quad \text{tr}(B) = \text{tr}(T) = \sum_{i=1}^n \lambda_i$$

sendo $\lambda_1, \dots, \lambda_n$ os valores próprios de matriz B .

Assim,

$$\psi(B) = \text{tr}(B) - \ln(\det(B)) = \sum_{i=1}^n \lambda_i - \ln\left(\prod_{i=1}^n \lambda_i\right) = \sum_{i=1}^n (\lambda_i - \ln(\lambda_i))$$

Ora sabendo que B é uma matriz definida positiva sabemos que os seu valores próprios são todos positivos, como já foi demonstrado, e também sabemos que para todo o $x > 0$ se tem que $x > \ln(x)$. Assim $(\lambda_i - \ln(\lambda_i)) > 0$ para todo o $i = 1, \dots, n$ e logo $\psi(B) > 0$.