TMA4215 Numerical Mathematics

Autumn 2011

Solution 1

Task 1

a) We would like to show that the error satisfies

$$\lim_{k\to\infty} \frac{|e_{k+1}|}{|e_k|^q} = C.$$

i) The zero $x^\star = \arccos 0.5 \approx 1.0471975512$, and

k	x_k	$\lvert e_k\rvert$	$\lvert e_{k+1}\rvert/\lvert e_k\rvert$	$\lvert e_{k+1}\rvert/\lvert e_k\rvert^2$
0	0.5000000000	$5.47\cdot 10^{-1}$	$4.39\cdot 10^{-1}$	0.803
1	1.2875729002	$2.40\cdot 10^{-1}$	$4.44\cdot 10^{-2}$	0.185
2	1.0578736992	$1.07\cdot 10^{-2}$	$3.03\cdot 10^{-3}$	0.283
3	1.0472298506	$3.23\cdot 10^{-5}$	$9.32\cdot 10^{-6}$	0.287
4	1.0471975514	$3.01\cdot 10^{-10}$	$8.69\cdot 10^{-11}$	0.287
5	1.0471975512	$2.62\cdot 10^{-20}$	$7.56\cdot 10^{-21}$	0.287
6	1.0471975512	$1.98\cdot 10^{-40}$		

As expected, we have quadratic convergence, i.e. $q = 2$, with $C = 0.287$ (in this case, the calculations have been done in Maple with accuracy of over 50 digits).

ii) The zero $x^\star = 0$, and

k	x_k	$\lvert e_k\rvert$	$\lvert e_{k+1}\rvert/\lvert e_k\rvert$	$\lvert e_{k+1}\rvert/\lvert e_k\rvert^2$
1	0.5000000000	$5.00\cdot 10^{-1}$	0.54	3.69
2	0.2707470413	$2.71\cdot 10^{-1}$	0.52	7.07
3	0.1414747338	$1.41\cdot 10^{-1}$	0.51	13.81
4	0.0724047358	$7.24\cdot 10^{-2}$	0.51	27.29
5	0.0366392002	$3.66\cdot 10^{-2}$	0.50	54.26
6	0.0184314669	$1.84\cdot 10^{-2}$	0.50	108.18
7	0.0092440432	$9.24\cdot 10^{-3}$	0.50	216.02
8	0.0046291426	$4.63\cdot 10^{-3}$	0.50	431.71
9	0.0023163571	$2.32\cdot 10^{-3}$	0.50	863.09
10	0.0011586257	$1.16\cdot 10^{-3}$		

In this case the convergence is linear, with constant $C = 0.5$. This is caused by $f'(0)$ being zero, so the condition for quadratic convergence is not satisfied. Instead, using $g(x) = x - f(x)/f'(x)$, we get

$$g'(x) = \frac{f(x)f''(x)}{[f'(x)]^2} \xrightarrow[x\to 0]{} \frac{1}{2},$$

see equation **(5)** p. 105 in K&C. This is in accordance with the measured results.

iii) The zero $x^\star = 0$, and

k	x_k	$\lvert e_k \rvert$	$\lvert e_{k+1}\rvert/\lvert e_k\rvert$	$\lvert e_{k+1}\rvert/\lvert e_k\rvert^2$
1	0.5000000000	$5.00 \cdot 10^{-1}$	0.66	3.02
2	0.3309759368	$3.31 \cdot 10^{-1}$	0.66	4.55
3	0.2199738473	$2.20 \cdot 10^{-1}$	0.67	6.83
4	0.1464514253	$1.46 \cdot 10^{-1}$	0.67	10.25
5	0.0975760249	$9.76 \cdot 10^{-2}$	0.67	15.38
6	0.0650334672	$6.50 \cdot 10^{-2}$	0.67	23.07
7	0.0433505497	$4.34 \cdot 10^{-2}$	0.67	34.60
8	0.0288988576	$2.89 \cdot 10^{-2}$	0.67	51.91
9	0.0192654581	$1.93 \cdot 10^{-2}$	0.67	77.86
10	0.0128435063	$1.28 \cdot 10^{-2}$		

This time the convergence is linear with $C = 0.67$. The reason is the same as in *ii)*.

b) *i)* $x^\star = \arccos(0.5) = \pi/3$, $f'(x^\star) = -\sqrt{3}/2$, so this zero has multiplicity 1.
ii) $x^\star = 0$, and $f'(0) = 0$, $f''(0) = 1$. The zero has multiplicity 2.
iii) $x^\star = 0$, and $f'(0) = f''(0) = 0$, $f'''(0) = 3$. This zero has multiplicity 3.

c) From the definition of multiplicity in the text, we can write

$$\mu(x) = \frac{(x - x^\star)^m q(x)}{m(x - x^\star)^{m-1} q(x) + (x - x^\star)^m q'(x)} = (x - x^\star)\frac{q(x)}{mq(x) - (x - x^\star)q'(x)}.$$

So $x^\star$ is a simple zero of $\mu(x)$ since $q(x^\star) \neq 0$. We find Newton's method applied to $\mu(x)$ as

$$g(x) = x - \frac{\mu(x)}{\mu'(x)} = x - \frac{f(x)f'(x)}{[f'(x)]^2 - f(x)f''(x)},$$

which converges quadratically.

d) You may do this task yourself. Notice that rounding errors can be a problem here, since $f(x)$ and $f'(x)$ both tend to zero when x_k tends to $x^\star$.

e) This task is similar enough to Newton's method that you should be able to do it on your own.

Task 2

a) We rewrite the system of equations as

$$F(X) = \begin{bmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{bmatrix} = \begin{bmatrix} x_1^2 + x_2^2 - 1 \\ x_1^3 - x_2 \end{bmatrix} = 0,$$

where $X = (x_1, x_2)^T$. The Jacobian matrix becomes

$$J(X) = \begin{bmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 \\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 & 2x_2 \\ 3x_1^2 & -1 \end{bmatrix}.$$

We can then write Newton's method as

$$X^{(n+1)} = X^{(n)} + H^{(n)},$$

where $H^{(n)}$ is implicitly given by

$$J(X^{(n)})H^{(n)} = -F(X^{(n)}). \tag{1}$$

In our case we can easily calculate J^{-1} (e.g. in Maple), which leads to the iteration

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mapsto \frac{1}{2x_1(1+3x_1x_2)} \begin{bmatrix} x_1^2 + 4x_1^3x_2 + x_2^2 + 1 \\ x_1^2(3x_2^2 - x_1^2 + 3) \end{bmatrix}.$$

Calculating J^{-1} as we have done will normally be very cumbersome. Instead one usually solves (1) numercally, e.g. with the conjugate gradient method. MATLAB does this for us if we solve (1) using the \ operator.

We must avoid initial values where the Jacobian is singular, i.e. when $\det\big(J(X)\big) = 0$:

$$\det\big(J(X)\big) = -2x_1 - 6x_1^2x_2 = -2x_1(1+3x_1x_2) = 0.$$

Thus, we must keep away from the curves $x_1 = 0$ and $3x_1x_2 = -1$, and choose initial values $x_1 = x_2 = 0.5$. After one iteration we get $x_1 = 1$ and $x_2 = 0.5$. After two iterations we get $x_1 = 0.85$ and $x_2 = 0.55$.

b) See the MATLAB programs on the homepage.

c) As we saw in **a)**, the Jacobian is singular on the x_1 axis. This causes the algorithm to fail, since we don't get a unique solution when solving (1).

Task 3

a) We start with the 2×2 case, and write

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$
$$F = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix}$$

Now,

$$\begin{aligned} \det(A + \varepsilon F) &= (a_{11} + \varepsilon f_{11})(a_{22} + \varepsilon f_{22}) - (a_{12} + \varepsilon f_{12})(a_{21} + \varepsilon f_{21}) \\ &= (a_{11}a_{22} - a_{12}a_{21}) + \varepsilon(a_{11}f_{22} + a_{22}f_{11} - a_{12}f_{21} - a_{21}f_{12}) \\ &\quad + \varepsilon^2(f_{11}f_{22} - f_{12}f_{21}). \end{aligned}$$

We see that $\det(A + \varepsilon F)$ is a polynomial in ε of degree 2. We proceed to prove that if A and F are $N \times N$ matrices, then $\det(A + \varepsilon F)$ is a polynomial of degree N by induction.

Assume that if $\tilde{A}$ and $\tilde{F}$ are $(N-1)\times(N-1)$-matrices, then $\det(\tilde{A}+\varepsilon\tilde{F})$ is a polynomial of degree $N-1$ in ε. We regard the $N \times N$-matrices A and F. Expand the determinant of $B = A + \varepsilon F$ by Laplace' formula. [1]

$$\begin{aligned}\det B &= b_{11}\mathrm{Cof}(b_{11}) - b_{12}\mathrm{Cof}(b_{12}) + \cdots + (-1)^{1+N} b_{1N}\mathrm{Cof}(b_{1N}) \\ &= (a_{11} + \varepsilon f_{11})\mathrm{Cof}(b_{11}) - \cdots + (-1)^{1+N}(a_{1N} + \varepsilon f_{1N})\mathrm{Cof}(b_{1N})\end{aligned}$$

The cofactors $\mathrm{Cof}(b_{ij})$ are the determinants of the matrices which arise from removing row i and coloumn j from B. These matrices are $(N-1)\times(N-1)$-matrices on the form $\tilde{A}+\varepsilon\tilde{F}$, so by the induction hypothesis, they are polynomials in ε of degree $N-1$. Thus each term in the sum above is a polynomial of degree N, and $\det B = \det(A + \varepsilon F)$ is as well.

We also note that if we set $\varepsilon = 0$, $\det B = \det A$. Polynomials are continuous, so if $\det A \neq 0$, there exists a $\delta > 0$ such that $\det A + \varepsilon F \neq 0$ for all $0 < \varepsilon < \delta$.

b) From Cramer's rule,

$$x_i(\varepsilon) = \frac{D_i(\varepsilon)}{D(\varepsilon)}, \qquad i = 1, \ldots, N,$$

where $D(\varepsilon) = \det(A+\varepsilon F)$ and $D_i(\varepsilon)$ is the determinant of the matrix formed by replacing coloumn i of $A + \varepsilon F$ with $b + \varepsilon v$. We see that these matrices are of the form considered in **a)**, and are as such degree N polynomials in ε. In **a)** we also proved that $D(\varepsilon) \neq 0$ for small ε, so $x_i(\varepsilon)$ are continuosu and og differentiable for small ε.

[1] Also known as cofactorexpansion.