MATH 2089 JUNE 2010 Numerical Methods

1. a) i) `x = 1e+200;` gives $x = 10^{200}$

`ans1 = exp(x^3)` gives ans1 = Inf as $e^{600} >$ realmax $\approx 10^{308}$

ii) `v = [-2 : 2 : 2]` gives $v = [-2 \quad 0 \quad 2]$

`ans2 = v./sqrt(v)` gives ans2 $= [-2 \quad 0 \quad 2] ./ [\sqrt{-2} \quad 0 \quad \sqrt{2}]$

$= [\sqrt{2}\,i \quad \text{NaN} \quad \sqrt{2}]$

iii) `x = 0.1;`

`h = 1e-18;` gives $h = 10^{-18}$

`ans3 = x+h <= x` gives ans 3 = 1 (for logical true)

as $h = 10^{-18} < |x|\,\varepsilon = 0.1 \times 2.2 \times 10^{-16} = 2.2 \times 10^{-17}$

$\therefore$ $x+h$ is stored as x on the computer and $x \le x$ is true.

b) i) One minute on a 3 GHz quad core computer where each core can do two floating point operations per clock cycle

$\Rightarrow$ Speed of computer $= 3 \times 10^9 \times 4 \times 2$

$= 2.4 \times 10^{10}$ flops / sec

$\Rightarrow$ in one minute can do $60 \times 2.4 \times 10^{10} = 1.44 \times 10^{12}$ flops

If coefficient matrix has no special structure $\Rightarrow$ need to use LU factorization to solve linear system

$$\Rightarrow \quad \frac{2n^3}{3} = 1.44 \times 10^{12}$$

$$\Rightarrow \quad n^3 = 2.16 \times 10^{12} \Rightarrow n \approx 13{,}000 \quad (\text{integer})$$

ii) Multiplying two $n \times n$ matrices takes 12 secs $= 2n^3$ flops

Solving a symmetric positive definite linear system is dominated by time for Cholesky factorization $= \frac{n^3}{3}$

$$\therefore \quad \frac{n^3}{3} = \frac{1}{6}(2n^3) = \frac{1}{6}(12) = 2 \text{ secs.}$$

1 c) $A, \underset{\sim}{b}$ known to 8 significant figures

$\Rightarrow$ rel err $(A) \leq \frac{1}{2} \times 10^{-8}$, rel err $(\underset{\sim}{b}) \leq \frac{1}{2} \times 10^{-8}$

i) $\kappa(A) = \|A\| \, \|A^{-1}\| = 1.9 \times 10^{1} \times 2.2 \times 10^{3} = 4.18 \times 10^{4}$

<u>N.B</u> For any norm, matrix A, $\kappa(A) \geq 1$.

ii) $$\text{Rel error}(\underset{\sim}{x}) \approx \kappa(A)\left(\text{rel error}(A) + \text{rel error}(\underset{\sim}{b})\right)$$
$$= 4.18 \times 10^{4}\left(\tfrac{1}{2} \times 10^{-8} + \tfrac{1}{2} \times 10^{-8}\right)$$
$$= 4.18 \times 10^{-4} = 0.418 \times 10^{-3}$$

$\therefore$ can expect at least 3 significant figures in compute $\underset{\sim}{x}$

d) Intersection of $f_1(x) = \frac{1}{1+x^2}$ and $f_2(x) = \log(|x|)$

i) Need to solve $$f(x) = f_1(x) - f_2(x)$$
$$= \frac{1}{1+x^2} - \log|x| = 0$$

ii) myfun = @(x) 1./(1+x.^2) - log(abs(x))

iii) Errors $e_k = |x^* - x_k|$. Look at behaviour as k increases:

As $e_k \to 0$ (2nd column) $x_k \to x^*$ so method is converging.

As $\frac{e_{k+1}}{e_k} \to 0 \Rightarrow$ order of convergence $\nu > 1$

As $\frac{e_{k+1}}{(e_k)^2} \to \infty \Rightarrow$ order of convergence $\nu < 2$

$\therefore$ Have order of convergence $\nu \in (1, 2)$ as expected for the Secant method when $f(x^*) = 0$, $f'(x^*) \neq 0$.

MATH 2089 JUNE 2010 Numerical Methods

2. a) Data

i	0	1	2	3	4
t_i	0	0.5	1	1.5	2
c_i	1	0.4283	0.5297	0.1344	0.0549

2 i)

A) Interpolating polynomial of degree n has $n+1$ parameters $a_0, a_1, \ldots, a_n$ $\Rightarrow$ degree $n = 4$ polynomial will interpolate 5 data values.

B) As $t \to \infty$ a quartic (degree 4) polynomial $\to +\infty$ if $a_4 > 0$, $-\infty$ if $a_4 < 0$, which is unrealistic for chemical concentration.

From plot the interpolating polynomial is < 0 for $t \approx 1.8$ which is not realistic for a chemical concentration.

ii) Approximation $c(t) = \alpha e^{-\beta t}$

Take logs (base e) $\Rightarrow \log(\alpha e^{-\beta t_i}) \approx \log(c_i)$

$\Rightarrow \log \alpha - \beta t_i \approx \log(c_i)$

Let $\bar{\alpha} = \log \alpha$ $\Rightarrow \bar{\alpha} - \beta t_i \approx \log(c_i)$

Linear least squares:

```
A = [ ones(size(tdat))  -tdat ];
x = A \ log(cdat)            % \ solves least squares
alpha = exp(x(1))
beta = x(2)
```

iii) Using simpson's rule and data in Table $h = 0.5$

$$\int_0^2 c(t)\,dt \approx \frac{h}{3}\left[c_0 + 4c_1 + 2c_2 + 4c_3 + c_4\right]$$

$$= \frac{0.5}{3}\left[1 + 4\times 0.4283 + 2\times 0.5297 + 4\times 0.1344 + 0.0549\right]$$

$$= \frac{0.5}{3} \times \cancel{2.1473}\ 4.3651 = \cancel{0.3579}\ 0.7275$$

[Do not quote answer to more figures than input data]

Additional data values at $t = 0.25, 0.75, 1.25, 1.75 \Rightarrow h_{new} = \frac{h}{2}$

As Simpson has error $O(h^4) \Rightarrow$ expect error to decrease by $\frac{1}{2^4} = \frac{1}{16}$.

2. iii) c) The Gauss-Legendre rule chooses optimal nodes x_i and weights w_i. To use it you need to be able to evaluate $c(x_i)$ – as the Gauss-Legendre nodes are <u>not</u> equally spaced this would require data values other than those in the Table.

2 b) IVP: $y''' + 2y'' - (\pi^2+1)y = \pi(\pi^2+1)e^{-t}\sin(\pi t)$

$y(0) = 1, \quad y'(0) = -1, \quad y''(0) = 1-\pi^2$

i) Third order ODE (highest derivative is y''')

$\Rightarrow$ use $\underset{\sim}{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y \\ y' \\ y'' \end{bmatrix}$

$$\Rightarrow \underset{\sim}{x}' = \frac{d}{dt}\underset{\sim}{x} = \begin{bmatrix} x_1' \\ x_2' \\ x_3' \end{bmatrix} = \begin{bmatrix} y' \\ y'' \\ y''' \end{bmatrix} = \begin{bmatrix} x_2 \\ x_3 \\ \pi(\pi^2+1)e^{-t}\sin(\pi t) - 2x_3 + (\pi^2+1)x_1 \end{bmatrix}$$

$$\therefore \underset{\sim}{f}(t, \underset{\sim}{x}) = \begin{bmatrix} x_2 \\ x_3 \\ \pi(\pi^2+1)e^{-t}\sin(\pi t) - 2x_3 + (\pi^2+1)x_1 \end{bmatrix}$$

Initial conditions $\underset{\sim}{x}(0) = \begin{bmatrix} x_1(0) \\ x_2(0) \\ x_3(0) \end{bmatrix} = \begin{bmatrix} y(0) \\ y'(0) \\ y''(0) \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1-\pi^2 \end{bmatrix}$

ii) Anonymous function myode to specify $\underset{\sim}{f}(t, \underset{\sim}{x})$

```
myode = @(t, x) = [ x(2) ;
                    x(3) ;
                    pi*(pi^2+1)*exp(-t).*sin(pi*t) ...
                        - 2*x(3) + (pi^2+1)*x(1) ];
```

(Last line is a continuation of second to last line).

MATH 2089 JUNE 2010 Numerical Methods

3. Fick's second law : concentration $c(x, y, t)$

PDE
$$\frac{\partial c}{\partial t} = D\left(\frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2}\right)$$

Space domain $\Omega = \{(x,y) : 0 \le x \le 2,\ 0 \le y \le 1\}$

a) Additional information:

Time domain $[0, T]$

Initial conditions $c(x,y,0)$ for $(x,y) \in \Omega$

Boundary conditions $c(x,y,t)$ for $(x,y) \in \partial\Omega$ $t > 0$

$\partial\Omega = \{(x,y) \in \Omega : x = 0$ or $x = 2$ or $y = 0$ or $y = 1\}$

b) Steady state $\Rightarrow$ no change w.r.t. time $t \Rightarrow \frac{\partial c}{\partial t} = 0$

$\Rightarrow \frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2} = 0$ 2-D Laplaces equation.

c) Using $c_{i,j}^{\ell} \approx c(x_i, y_j, t_\ell)$ at (x_i, y_j), $t_{\ell+1}$

$$\left.\frac{\partial c}{\partial t}\right|_{\substack{(x_i,y_j)\\ t_{\ell+1}}} = \frac{c_{i,j}^{\ell} - c_{i,j}^{\ell+1}}{(-\Delta t)} + O(\Delta t)$$

$$= \frac{c_{i,j}^{\ell+1} - c_{i,j}^{\ell}}{\Delta t} + O(\Delta t)$$

d) Central difference approximations to space derivatives at (x_i, y_j) and $t_{\ell+1}$ (h is spacing for both x and y discretizations)

$$\left.\frac{\partial^2 c}{\partial x^2}\right|_{\substack{(x_i,y_j)\\ t_{\ell+1}}} = \frac{c_{i-1,j}^{\ell+1} - 2c_{i,j}^{\ell+1} + c_{i+1,j}^{\ell+1}}{h^2} + O(h^2)$$

$$\left.\frac{\partial^2 c}{\partial y^2}\right|_{\substack{(x_i,y_j)\\ t_{\ell+1}}} = \frac{c_{i,j-1}^{\ell+1} - 2c_{i,j}^{\ell+1} + c_{i,j+1}^{\ell+1}}{h^2} + O(h^2)$$

3 e) Ignoring $O(\Delta t), O(h^2)$ in finite difference approximations gives

$$\frac{c_{i,j}^{l+1} - c_{i,j}^{l}}{\Delta t} = D\left[\frac{c_{i-1,j}^{l+1} - 2c_{i,j}^{l+1} + c_{i+1,j}^{l+1}}{h^2} + \frac{c_{i,j-1}^{l+1} - 2c_{i,j}^{l+1} + c_{i,j+1}^{l+1}}{h^2}\right]$$

Multiply through by Δt, use $s = \frac{D\Delta t}{h^2}$, unknowns on LHS

$$(1+4s)\,c_{i,j}^{l+1} - s\,c_{i-1,j}^{l+1} - s\,c_{i+1,j}^{l+1} - s\,c_{i,j-1}^{l+1} - s\,c_{i,j+1}^{l+1} = c_{i,j}^{l}$$

So $\alpha = 1+4s$ and $\beta = -s$.

f) $c(0,y,t) = 6$ for $0<y<1$ Boundary condition on $x=0$
$c(x,y,0) = 3$ for $(x,y)\in\Omega$ Initial condition at $t=0$

At (x_1, y_3) and t_1, so $i=1$, $j=3$, $l+1=1 \Rightarrow l=0$

$$\alpha\,c_{1,3}^{1} + \beta\,c_{0,3}^{1} + \beta\,c_{2,3}^{1} + \beta\,c_{1,2}^{1} + \beta\,c_{1,4}^{1} = c_{1,3}^{0}$$

But $c_{0,3}^{l} = 6$ from BC, $c_{i,j}^{0} = 3$ from IC, so

$$\alpha\,c_{1,3}^{1} + \beta\,c_{2,3}^{1} + \beta\,c_{1,2}^{1} + \beta\,c_{1,4}^{1} = 3+6\beta = 3-6s$$

g). Implicit method (this method) : need to solve (structured) linear system to get unknowns $c_{i,j}^{l+1}$; unconditionally stable, so no restrictions on $\Delta t, h$.

Explicit method : explicit formula for unknowns $c_{i,j}^{l+1}$ in terms of known values $c_{i,j}^{l}$. Condition on s, hence $\Delta t, h$, for method to be stable.

h) Sparsity $= \frac{154}{36^2} \times 100 = 11.9\ \%$.

Even though A is sparse (high percentage of values $=0$), A^{-1} is typically dense - all values non-zero $\Rightarrow$ increased storage and calculation

A is banded (lower bandwidth $m_\ell = 9$, upper bandwidth $m_u = 9$)
$\Rightarrow$ fill-in when calculating factorization only occurs within bands.

Symmetric as coefficients of all non-zero off diagonal elements are β

Positive definite as diagonally dominant: $A_{ii} = 1+4s > \sum_{j\neq i} |A_{ij}| \le 4s$

Not Toeplitz (see diagonals -1 & 1 which contain both zero and non-zero values, so are not constant).