

**NATIONAL UNIVERSITY OF SINGAPORE**

PC5215 – NUMERICAL RECIPES WITH APPLICATIONS

(Semester I: AY 2016-17)

Time Allowed: 2 Hours

---

**INSTRUCTIONS TO STUDENTS**

1. Please write your student number only. Do not write your name.
2. This assessment paper contains FIVE questions and comprises THREE printed pages.
3. Students are required to answer ALL questions; questions carry equal marks.
4. Students should write the answers for each question on a new page.
5. This is a CLOSED BOOK examination.
6. Non-programmable calculators are allowed.

1. Floating point representation and calculation on computers are not exact. We define the relative accuracy  $\delta = \Delta x/|x|$  as the maximum possible representation error  $\Delta x$  divided by the magnitude  $|x|$ .
  - a. How is  $\delta$  related to machine epsilon  $\epsilon$ ?
  - b. Based on IEEE 754 single precision representation for real numbers, if a computer gives you a printout of the value as 1.0000000, what is the actual value possible with printout equal to 1?
  - c. If we compute  $y = (1.0000000 - 1.0000010)$  by computer, what is the relative accuracy  $\delta$  for the result  $y$ ?
  - a) *Machine epsilon is defined to be the next representable value after 1. So if the value is  $x$ ,  $\epsilon|x|$  is next representable value for  $x$ . So all values between  $1 - \epsilon/2$  and  $1 + \epsilon/2$  is represented as 1. So  $\delta = \epsilon/2$ .*
  - b) *All real values between  $1 - \delta$  and  $1 + \delta$  are represented as 1. In single precision, this is  $\delta = 2^{-24} = 6 \times 10^{-8}$ .*
  - c) *First number is  $1 \pm 6 \times 10^{-8}$ . The second number is  $1.0000010 \pm 6 \times 10^{-8}$ . The difference is  $0.0000010 \pm 0.00000012$ , so  $\delta = 12\%$ .*

2. A three by three matrix is given as  $A = \begin{bmatrix} 2 & 1 & -5 \\ -2 & 3 & 1 \\ 1 & 0 & 2 \end{bmatrix}$ .
  - a. Perform the  $LU$  decomposition without pivoting.
  - b. Use the result to compute the determinant of the matrix,  $\det(A)$ .

*Write the matrix  $A$  as a product of lower triangle (with diagonals equal 1), multiplied with upper triangle matrix, we can work out the entries, as*

*$\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1/2 & -1/8 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & -5 \\ 0 & 4 & -4 \\ 0 & 0 & 4 \end{bmatrix}$ . The determinant is obtained from the product of diagonals, as 32.*

3. We generate random points on the surface of a sphere according to the probability distribution  $p(\theta, \phi)d\theta d\phi \propto (1 + \cos \theta)d\theta d\phi$ , where  $0 < \theta < \pi$  and  $0 < \phi < 2\pi$  are the angles in a spherical coordinate system.
  - a. Give a method to generate  $(\theta, \phi)$  such that the points are uncorrelated.
  - b. Generate the same distribution of points using Metropolis algorithm.

- a) We first determine the normalization factor (which is  $2\pi^2$ ). Then  $\phi = 2\pi \cdot \xi_1$ . We use inverse method for  $\theta$  determined by  $\xi_2\pi = \theta + \sin \theta$  [although not very practical].
- b) Given an arbitrary starting point of  $(\theta, \phi)$ , we displace them by a small amount to a new location  $(\theta', \phi')$ , subject to the constraint that  $\theta$  is between 0 and  $\pi$  while  $\phi$  is between 0 and  $2\pi$ , we accept the move if  $(\xi < r)$ , where  $r = \frac{1+\cos \theta'}{1+\cos \theta}$ . We do this repeatedly as many steps as necessary.

4. Consider the experimental results  $y_i$  with standard deviation  $\sigma_i$  at point  $x_i$  given by the following table:

$i$	$x_i$	$y_i$	$\sigma_i$
1	0.0	0.05	0.04
2	1.0	1.10	0.08
3	2.0	1.98	0.05

We assume that the data can be fitted by a straight line of the form  $y(x) = a + bx$ .

- Write down the expression of  $\chi^2(a, b)$ .
  - Compute the gradient of  $\chi^2(a, b)$  with respect to the two fitting parameters  $a$  and  $b$  at  $a = 0$  and  $b = 1$ . Based on the gradient information, perform one step move of a steepest descent search for the minimum of  $\chi^2$ .
  - How many steps are required to locate the global minimum of  $\chi^2$  if one uses steepest descent repeatedly? And how many steps are required to locate the global minimum if one uses a conjugate gradient (CG) method? Perform the CG calculation to find  $(a_{min}, b_{min})$ .
- a)  $\chi^2(a, b) = \left(\frac{0.05-a}{0.04}\right)^2 + \left(\frac{1.10-a-b}{0.08}\right)^2 + \left(\frac{1.98-a-2b}{0.05}\right)^2$ .
- b) We take the derivatives and evaluate at  $a=0, b=1$ , find  $\frac{\partial \chi^2}{\partial a} = -77.75, \frac{\partial \chi^2}{\partial b} = 0.75$ . We make one-step move according to  $a \rightarrow a = -77.75 \lambda, b = 1 + 0.75 \lambda$ . Put into  $\chi^2$  and minimize with respect to  $\lambda$ , we find a quadratic equation for  $\lambda$ . The solution is  $\lambda = -0.00043$ . This gives  $a = 0.0334306, b = 0.999678$ .
- c) If we use steepest descent, we need infinite number of steps. If we use conjugate gradient method, we take two steps. The first step is the same as in b), the second step following the GC algorithm, we find  $a = 0.0591582, b = 0.967576$ .

5. Consider the diffusion equation for the variable  $u(x, t)$  with a decay term

$$\frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial x^2} - \frac{u}{\tau}$$

where  $D$  is the diffusion constant and  $\tau$  the relaxation time. A numerical procedure can be devised to solve the equation by the fast Fourier transform (FFT).

- First, the formal solution can be written as  $u(x, t + \Delta t) = \exp[\Delta t(\hat{A} + \hat{B})]u(x, t)$ . Determine the operators  $\hat{A}$  and  $\hat{B}$ .
- Derive the operator-splitting formula (Trotter-Suzuki formula) and indicate the order of error in  $\Delta t$ .
- Describe a procedure to solve the equation using FFT.

a)  $\hat{A} = D \frac{\partial^2}{\partial x^2}, \hat{B} = -\frac{1}{\tau}$ .

b) Expand  $e^{\Delta t(\hat{A}+\hat{B})}$ , and  $e^{\Delta t\hat{A}}e^{\Delta t\hat{B}}$ , we find they differ at  $\Delta t^2$ .

c) We use discretized form for the Fourier transform  $\tilde{u}(k) = \int_{-\infty}^{+\infty} e^{-ikx} u(x) dx$ , and the inverse  $u(x) = \int_{-\infty}^{+\infty} e^{ikx} \tilde{u}(k) \frac{dk}{2\pi}$ . First step we multiply  $u(x)$  by  $e^{-\frac{\Delta t}{\tau}}$ , Fourier transform into  $k$  space, then second step multiple  $\tilde{u}(k)$  by  $e^{-\Delta t D k^2}$ , then Inverse Fourier transform back to  $x$  space. Note that  $\frac{\partial}{\partial x} \rightarrow ik$ .

- The End -

[WJS]