Indian Institute of Science Education and Research, Mohali

Computational Methods in Physics (PHY422)

(Jan.-Apr. 2015)

Assignment 1

Feb.28, 2015

---

1. Describe the representation of integers using two's complement.
   (a) What is the largest and the smallest integer that can be represented using 8 bits?
   (b) What is the largest and the smallest integer that can be represented using 4 bytes?
   (c) What is the largest and the smallest integer that can be represented using 8 bytes?
2. Integer operations.
   (a) Describe the addition operation for integers when using two's complement.
   (b) What happens when we add one to the largest integer?
   (c) Describe multiplication of integers in this representation.
   (d) Write a program to carry out integer division and calculate the divisor and remainder.
3. Representation of floating point numbers.
   (a) Describe the representation of floating point numbers in digital computers.
   (b) What is the smallest number that can be represented in the standard 4 byte float/real variable?
   (c) What happens if you divide this number by 2? Try this in a program.
   (d) What happens if you divide this number by 10^7?
   (e) What is the largest number that can be represented in the standard 4 byte float/real variable?
   (f) What happens if you multiply this number by 2? Try this in a program.
   (g) Describe how floating point numbers can be added in a digital computer? Describe each step of the process in detail.
   (h) Given $1.0e0$, what is the smallest number that you can add to it such that the resulting answer is different from $1.0e0$? Write a program to calculate this number.
   (i) Given $1.0e0$, what is the largest number that we can add to it such that the resulting answer is $1.0e0$? Write a program to calculate this number.
   (j) Describe how floating point numbers can be multiplied in a digital computer? Calculate the number of bits of storage that is required to carry out the operation for 4 byte float/real variable.
   (k) Describe how, given a variable x, we may compute $1.0/x$ in a digital computer? Calculate the number of bits of storage that is required to carry out the operation for 4 byte float/real variable.

   Describe round-off error in floating point operations in digital computer. Estimate the round-off error in the following operations and discuss situations where relative error in the result due to round-off can be significant.
   (a)

$$s = x_1 + x_2$$

(b)

$$s = (x_1 + x_2) + x_3$$

(c)

$$s = x_1 * x_2$$

(d)

$$s = (x_1 + x_2) * x3$$

(e)

$$s = x_1 * x_1 - x_2 * x_2$$

4. Given an integer n, how many operations are required to test whether this number is a prime or not? Describe the most optimal method that you can think of for carrying our such a test. Estimate the number of operations in this approach.

5. Calculate the number of operations in the Gauss elimination method for finding solutions to n linear algebraic equations in n variables. If you had the freedom to assign parts of the calculation to different processors, how will you divide the workload in this algorithm? If you are given a computer with $m < n$ processors, how well can you divide the workload such that each processor is occupied most of the time? Can you ensure, by calculating the number of operations for each processor in your scheme for dividing work, that the workload on each processor is the same? Done one processor have to wait for another to complete calculations in this scheme?

6. Write a program to do interpolation for the function $\tanh(x)$ in the range $-100 \leq x \leq 100$ using Newton's divided interpolation scheme. If you use a single polynomial, and you sample the function at 10^4 equally spaced points for testing the interpolating polynomial, and require better than 1% accuracy, then how many points do you need to construct the polynomial?

7. Repeat the above exercise by using rational function interpolation instead of polynomial fitting.

8. Write down an expression for coefficients of a cubic spline for a given data set.

9. Derive the formula for numerical differentiation that is accurate to fourth order in the step size.

10. Use the Runge-Kutta method to derive the family of second order accurate formulae for solving coupled ordinary differential equation of first order. Write a program to solve the equation of motion of two particles joined by a spring of spring constant k and natural length a, moving in one dimension. Compare the solution with the analytic solution in this case and test your program.