COMP 350 Numerical Computing

Assignment #1: Floating point in C, overflow and underflow, numerical cancellation
Date Given: Monday, September 21. Date Due: 5pm, Wednesday, September 30, 2015

Submit your assignment including your code through myCourses.

1. (5 points) Write a C program to find the smallest positive integer x such that the floating point expression

$$(1 \oslash x) \otimes x$$

is not equal to 1, using single precision. Make sure that the variable x has type float, and assign the value of the expression $1 \oslash x$ to a float variable before doing the multiplication operation. Repeat with double precision.

2. (5 points) A calculus student was asked to determine $\lim_{n \to n} x_n$, where $x_n = (100^n)/n!$. He wrote a C program in single precision to evaluate x_n by using

$$x_1 = 100, \quad x_n = 100x_{n-1}/n, \quad n = 2, 3, \ldots, 60.$$

The numbers printed became ever larger and finally became ∞. So the student concluded that $\lim_{n \to n} x_n = \infty$. Please write a C program in single precision to verify the student's observation. The student's conclusion is actually wrong. What is the problem with his program?

Bonus (3 points): Can you rewrite a C program to evaluate x_n so that you can make a right conclusion about $\lim_{n \to n} x_n$?

3. For any $x_0 > -1$, the sequence defined recursively by

$$x_{n+1} = 2^{n+1}(\sqrt{1 + 2^{-n}x_n} - 1), \qquad (n \geq 0)$$

converges to $\ln(x_0 + 1)$.

   (a) (4 points) Let $x_0 = 1$. Use the formula to compute $x_n - \ln(x_0 + 1)$ for $n = 1, 2, \ldots, 60$ in double precision. Explain your results.

   (b) (6 points) Improve the formula to avoid the difficulty you encountered in 3(a). Again compute $x_n - \ln(x_0 + 1)$ for $n = 1, 2, \ldots, 60$ in double precision.

Note: You should make your code efficient, i.e., avoid unnecessary operations.