COMP 350 Numerical Computing

Assignment #1: Floating Point Computing

Date Given: Wednesday, September 9. Date Due: 5pm, Wednesday, September 23, 2015

You can submit either an electronic copy (in PDF format) of your assignment through myCourses or a hard copy, which should be placed in the marked COMP 350 boxes in a cabinet located in Trottier building on the 2nd foor near the elevator.

1. (2 point) Is there a real number which has finite binary representation but infinite (or non-terminating) decimal representation? Give reasons.

2. (2 points) Suppose the 2's complement representation of a number is (a 32-bit word is used)

   $$\boxed{11111111111111111111111111101001}$$

   What is the number? Give your answer in decimal representation.

3. Suppose in IEEE single precision, the width of the exponent field is 4, not 8, and the width of the fraction field is 5, not 23.

   (a) (1 point) What should the exponent bias be?

   (b) (2 points) What are the largest and smallest nonnegative normalized floating point numbers in this system?

   (c) (2 points) What are the largest and smallest nonnegative subnormal floating point numbers in this system?

   (d) (1 point) What is the machine epsilon of this system.

   (e) (2 points) What are the two floating point numbers (neither is equal to 11) closest to 11?

   (f) (2 points) Given number $-(1.0110101)_2$. Round it using the four rounding modes. Give the answers as normalized floating point numbers, in the form **binary-significand**$\times 2^E$, where E is decimal.

4. Are the following statements true or false? If a statement is true, give a proof and if it's false, give a counter example. We assume no overflow occurs in the calculations and the rounding mode used can be any of the four rounding modes.

   (a) (2 points) If x is a nonzero finite floating point number, then $x \oplus x = 2x$.

   (b) (2 points) If x and y are two finite floating point number, then $x \ominus y = -(y \ominus x)$.

5. (2 points) What are the values of the expressions $\infty/0$, $\infty/(-\infty)$, 1^{NaN}, and $-0/NaN$?