COMPUTER SCIENCE 349A, SPRING 2018
ASSIGNMENT #2 - 20 MARKS

DUE MONDAY JANUARY 29, 2018 (11:30 p.m. PST)

This is a really large class and the logistics of grading assignments are challenging. Me and the markers require your help in making this process go smoothly. Please ensure that your assignments conform to the following requirements - any violation will result in getting a zero for the particular assignment.

- All assignments should be submitted electronically through the ConneX course website and shoud be **SINGLE PDF FILES**. No other formats will be accepted.. Handwritten answers are ok but they will need to be scanned and merged into a single pdf file together with any code examples and associated plots.
- The assignment number, student name and student number should be clearly visible on the top of every page of your assignment submission.
- **PLEASE DO NOT COPY THE ASSIGNMENT DESCRIPTION IN YOUR SUBMISSION**
- The asnwers to the questions should be in the same order as in the assignment specification.
- Some of the questions of the assignments are recycled from previous years but typically with small changes in either the description or the numbers. Any submission that contains numbers from previous years in any questions will be immediately graded with zero.
- Any assignment related email questions should have a subject line of the form CSC349A Assignment X, where X is the number of the corresponding assignment.
- The total number of points for this assignment is 20.

Question #1 - 6 marks.
If x and y are floating-point numbers, then the evaluation of

$$f(x,y) = \frac{y}{\sqrt{x^2+y^2}+x}$$

may be very inaccurate due to cancellation. For example, with base $b = 10$, precision $k = 4$, idealized chopping arithmetic, if $x = -12.34$ and $y = 0.9555$, then the following results are obtained:

$$\begin{aligned}
fl(x^2) &= fl(152.2756) = 152.2 \text{ or } 0.1522 \times 10^3. \\
fl(y^2) &= fl(0.91298025) = 0.9129 \\
fl(x^2+y^2) &= fl(152.2 + 0.9129) = fl(153.1129) = 153.1 \\
fl(\sqrt{x^2+y^2}) &= fl(\sqrt{153.1}) = fl(12.37335...) = 12.37 \\
fl(\sqrt{x^2+y^2}+x) &= fl(12.37 - 12.34) = fl(0.03) = 0.03 \\
fl(f(x,y)) &= fl(y/\sqrt{x^2+y^2}+x) = fl(0.9555/0.03) = fl(31.85) = 31.85
\end{aligned}$$

However, the correct value of $f(x,y)$ is 25.868066..., so the floating-point approximation has a large relative error (around 23%). Note that $f(x,y)$ can be rewritten as

$$g(x,y) = \frac{\sqrt{x^2+y^2}-x}{y}.$$

(a) Using base $b = 10$, precision $k = 4$, idealized chopping arithmetic, $x = -12.34$ and $y = 0.9555$, evaluate $fl(g(x,y))$ and determine the relative error.

(b) For each of the specified data in the table below, place an X in the appropriate box to indicate which of the formulas $f(x,y)$ or $g(x,y)$ is more accurate in precision $k = 4$ floating-point arithmetic, or if they are both accurate. Put exactly one X in each row of the table. (No justification for your answers is required. It is NOT necessary to do any floating-point computation to answer this question.)

data	$f(x,y)$ more accurate	$g(x,y)$ more accurate	both accurate
$x = 0.1234, y = 12.34$			
$x = 12.34, y = -0.9123$			
$x = -0.1234, y = -0.005678$			

Question #2 - 6 Marks.

For each of the functions below, the floating-point evaluation of the given function is very inaccurate for the specified range of values of the parameters. In each case, find another expression that is mathematically identical to the given function and that is more accurate using floating-point computation than is the given form of the function (for the specified values of the parameters).

Note. Do not use any Taylor polynomial approximations in this question.

(a) $f(x) = \sqrt{x-1} - \sqrt{x}$, when x is positive and large in magnitude.

(b) $g(x) = \ln x - \ln y$, when x is close (but not equal) to y.

(c) $h(x) = \dfrac{x}{x+1} - 1$, when $|x|$ is very large.

Question #3 - 8 Marks
Consider a ternary, normalized floating-point number system that is base 3. Analogous to a bit, a ternary digit is a trit. Assume that a hypothetical ternary computer uses the following floating-point representation:

s_m	s_e	t_1	t_2	t_3	t_4	e_1	e_2

where s_m is the sign of the mantissa and s_e is the sign of the exponent (0 for positive, 1 for negative), t_1, t_2, t_3 and t_4 are the trits of the mantissa, and e_1, e_2 are the trits of the exponent, where each trit is 0,1 or 2. For parts (a) to (b), $X_{(10)}$ is used to indicate that the number provided is in decimal. Show all your work for all parts.

(a) What is the computer representation of $3_{(10)} + \left(\frac{1}{9}\right)_{(10)}$ in this system?

(b) What is the computer representation $-29_{(10)}$ in this system?

(c) What is the smallest positive non-zero number that can be represented in this system? What is it's value in decimal?

(d) What is the size of the gap between any two consecutive numbers in the interval $9_{(10)}$ and $27_{(10)}$ in this ternary floating-point representation system? Your answer should be in decimal.