COMPUTER SCIENCE 349A, SPRING 2019
ASSIGNMENT #2 - 20 MARKS

DUE MONDAY FEBRUARY 4, 2019 (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 - 8 Marks
Consider a quaternary, *normalized* floating-point number system that is base 4. Analogous to a bit, a quaternary digit is a quit. Assume that a hypothetical quaternary computer uses the following floating-point representation:

s_m	s_e	e_1	e_0	q_1	q_2	q_3	q_4

where s_m is the sign of the mantissa and s_e is the sign of the exponent (0 for positive, 1 for negative), q_1, q_2, q_3 and q_4 are the quits of the mantissa, and e_1, e_0 are the quits of the exponent, where each quit is 0,1, 2 or 3. For parts (a) and (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 -30_{10} in this system?

(b) What is the computer representation of 7.1875_{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 $16_{(10)}$ and $64_{(10)}$ in this quaternary floating-point representation system? Your answer should be in decimal.

Question #2 - 6 marks.
If x and y are floating-point numbers, then the evaluation of

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

in a floating point system may be very inaccurate due to cancellation. To illustrate, use base $b = 10$, precision $k = 4$, idealized, chopping floating-point arithmetic for parts (a) and (b) below.

(a) Let $x = 123.4 = 0.1234 \times 10^3$ and $y = -1.234 = -0.1234 \times 10^1$, evaluate $fl(f(x, y))$ and determine the relative error.

(b) Now let $x = -123.4 = -0.1234 \times 10^3$ and $y = 1.234 = 0.1234 \times 10^1$, evaluate $fl(f(x, y))$ and determine the relative error.

(c) Based on the above results and similar kinds of computations (you might do some more computations, but it is not necessary to hand them in), specify which ranges of values of x and y are such that the computation of $fl(f(x, y))$ will always be very inaccurate? (Your ranges can be loosely deifned, that is, you do not have to give specific numeric bounds. You may say things like "large x", or "large y", or "small positive x", etc.)

Question #3 - 6 Marks.

(a) Determine the second order ($n = 2$) Taylor polynomial approximation for $f(x) = \sqrt{x+3}$ expanded about $a = 1$ and its remainder term. Leave your answer in terms of factors $(x-1)$ (that is, do not simplify). Show all your work.

(b) Use the polynomial approximation in (a) (without the remainder term) to approximate $f(1.14) = \sqrt{4.14}$. Use either hand computation, your calculator or MATLAB. Give an exact answer to 6 significant digits. Calculate the absolute error of your approximation.

(d) Determine a good upper bound for the truncation error of the Taylor polynomial approximation in (a) for all values of x such that $1 \leq x \leq 1.2$ by bounding the error term found in (a). (Note: this upper bound should be bigger the the absolute error you calculate in (b).)