Scientific Computing | Fitchburg State University, Department of Mathematics, HW #3 | September 21, 2016

The following problems should be done in julia. Type up the code you used to find your results and answer all questions. Submit your homework via Blackboard by using any program you want and submitting the PDF version of it. This assignment should be submitted no later than 11:59pm on September 28, 2016.

1. Recall that in class, we took the original form of the quadratic formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

and derived an equivalent one of the form:

$$x = \frac{-2c}{b \pm \sqrt{b^2 - 4ac}}$$

and if there is roundoff in one form then there isn't with the other one.

   (a) Write a julia function that will use the form of the quadratic with the least amount of roundoff. You should return the 2 roots and no need to check if the result is real. (Hint: make sure that you are always adding $b$ and $\sqrt{b^2 - 4ac}$, because the roundoff will be worse when subtracting).

   (b) Find the solution to $12.242x^2 + 42.382x + 0.0012 = 0$ using the function in (a) in both 64-bit and 16-bit floating point numbers.

   (c) Assume that the 64-bit numbers are exact. Find the absolute and relative errors in the 16-bit values. Discuss your results compared to what we found in class (see the notes).

   (d) Find the solution to $100.0x^2 - 400.0x + 399.9 = 0$ using both 64-bit and 16-bit floating point numbers. If anything unusual occurs, discuss what happened.

   (e) Write (or adapt) a function to use Newton's method to find the solution to the quadratic $ax^2 + bx + c = 0$. The input should be $a$, $b$ and $c$ and the output should be the two values of the roots. (Again, don't worry about the case where the solution is not real.) Hint: to get two values, you need to start Newton's method with two values. Recall that for a quadratic function, one root is left of the vertex and the other is to the right. Pick two values relative to the vertex as your starting points.

   (f) Find the solution to $12.242x^2 + 42.382x + 0.0012 = 0$ using your method in (e) for both 16- and 64-bit numbers. Find the absolute and relative error in the 16-bit answers. Use the 64-bit result as exact.

   (g) Find the solution to $100.0x^2 - 400.0x + 399.9 = 0$ using your method in (e) for both 16- and 64-bit numbers. Find the absolute and relative error in the 16-bit answers. Use the 64-bit result as exact.

   (h) Which is the most accurate method to solving quadratic equations? Why? Explain using the examples above.

2. Newton's method is an example of a iterative formula that generally converges. However, not all iterative formulas converge. The following is another example:

$$x_{n+1} = ax_n(1 - x_n)$$

   (a) If $a = 2$ and $x_1 = 0.6$, find the first 125 iterations, but print out (or show) only the last 25 terms. Does it appear that the sequence converges? If so, to what? If there any anything else interesting about the sequence, discuss it.

   (b) Repeat part (a) with $a = 3$ and $x_1 = 0.6$.

   (c) Repeat part (a) with $a = 3.5$ and $x_1 = 0.6$.

(d) Repeat part (a) with $a = 4$ and $x_1 = 0.6$.

Note: this iterative formula produces a chaotic sequence in which doesn't have any apparent pattern and this one will cover equally the interval $[0, 1]$. (I'll explain this more when we get to random numbers)

3. A matrix is an array of numbers and are often studied in linear algebra. Consider a 2 by 2 matrix (that is with 2 rows and 2 column) that looks like

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

The determinant of a 2 by 2 matrix is given by $\det(A) = ad - bc$. In this case we will examine the the matrix

$$A = \begin{bmatrix} 5.99 & 3.99 \\ 3.01 & 2.01 \end{bmatrix}$$

(a) Find the exact value of $\det(A)$. (You may do this by hand or some other method, but explain how you arrived at this.)

(b) Create a function to produce the determinant. The first line should be `function det(a,b,c,d)`.

(c) Use the function to find the determinant of $A$ using both 16-bit and 64-bit floating point numbers. Find the absolute and relative errors of the answer in (c).

(d) Try to find a way to get the exact value of $\det(A)$ using Julia. (Hint: consider different number types).

Note: although this problem talks about arrays, you don't need to use them in your code at all. You may use them, however, if you'd like to.