Math 441 (2020)
Homework 1 - max 20
Due: Thur. September 3, 2020

. NAME: _____

1. [5pts] The real number $x = \frac{11}{8}$ has decimal representation $x_d = 1.375$ and binary representation $x_b = 1.011$. Compute each of the following relative errors in decimal.

$$E_d = \frac{|x_d - chop(x_d, 3)|}{x_d} \quad , \quad E_b = \frac{|x_b - chop(x_b, 3)|}{x_b}$$

Here, for example, $chop(x_d, 3) = 1.37$. For, E_b compute $x_b - chop(x_b, 3)$ in binary, then re-express the numerator and denominator in decimal before computing the ratio in decimal.

2. [7pts] Suppose one can compute $\sqrt{x}$ exactly but an error of $\delta > 0$ is incurred by some finite representation $\hat{x}$ of x. In particular:

$$\hat{x} = x + \delta$$

a) For $\delta > 0$ find a uniform upper bound on the absolute error

$$E_a = |\sqrt{x} - \sqrt{\hat{x}}\ |$$

valid for all $x \in [0, 1]$, i.e., an $E(\delta)$ such that

$$E_a(x, \delta) \leq E(\delta) \qquad \forall x \in [0, 1]$$

Hint: multiply top and bottom of E_a by $\sqrt{x} + \sqrt{\hat{x}}$

b) If $\delta = 10^{-6}$ what does a) imply the upper bound on E_a is on $[0, 1]$?

3. [8pts] The Taylor series for $f(x) = \ln(1+x)$ is

$$\ln(1+x) = \sum_{k=1}^{n} (-1)^{k-1} \frac{x^k}{k} + E_n(\zeta, x) = P_n(x) + E_n(\zeta, x)$$

and converges for $x \in (-1, 1]$.

a) Use the Alternating Series Test to bound the error $|E_n|$ by $\hat{E}_n$. Use $\hat{E}_n$ to find an n sufficiently large so that

$$|\ln(2) - P_n(1)| \leq \hat{E}_n \leq 10^{-6}$$

Here $x = 1$.

b) One can accelerate the series convergence rate using the following identity

$$\ln(2) = ln(e \cdot 2/e) = 1 + \ln(2/e) = 1 + \ln\left(1 + (\frac{2}{e} - 1)\right) = 1 + \ln(1+x)$$

Use the series above for this (new) x

$$x = \frac{2}{e} - 1 \simeq -0.2642411176$$

with $n \leq 10$ to show the accelerated convergence. Specifically, make a table with $n = 1, 2, \ldots 10$ having columns

$$n \qquad 1 + P_n(x) \qquad E_a = |1 + P_n(x) - \ln 2|$$

You may use the exact value $\ln 2 = 0.6931471806\ldots$ and ANY software you desire to compute the absolute error E_a.