Midterm 01

Oct 27, 2020.

1. (10 pts) The (fictional) one-and-half precision format uses 48 bits to store a binary floating point number of the form $\pm 1.a_1 a_2 \cdots a_t \times 2^e$ where $a_j \in \{0,1\}$, $-510 \le e \le 511$. Find t and derive an upper bound for relative error caused by rounding. Express your final answer as a real number, but need not convert it to decimal expression.

2. (10 pts) Given p_0, p_1 and p_2, the general solution to the recursion formula $p_n = \frac{10}{3}p_{n-1} - 3p_{n-2} + \frac{2}{3}p_{n-3}$ is $p_n = c_1 1^n + c_2 2^n + c_3(\frac{1}{3})^3$ (need not show this). Find all $(c_1, c_2, c_3) \ne (0,0,0)$ such that the above iteration is unstable in relative error. Explain.

3. (10 pts) The first few iteration $(p_i, f(p_i))$, $i = 0,1,2,3,4$ of method of false position for some equation $f(x) = 0$ is given by

$$(0,-2), \quad (3,1), \quad (*,2), \quad (*,1), \quad (*,\frac{2}{9})$$

   Find p_5 (4 digits will do). Explain.

4. (15 pts) Use any method to find a solution of $\sqrt{1+0.9x} - \sqrt{1-0.8x} = 1.0 \times 10^{-10}$ to 15 correct digits. You need to prevent loss of accuracy. Standard methods only gives you about 5 correct digits (and 1/3 partial credits).

5. (10+5 pts) It is known that the unique solution to $f(x) = x + 3\sin(x) - 0.01 = 0$ is located near $x = 0$.

   (a) Find a fixed point iteration that will converge for any $x_0 \in [-\frac{1}{2}, \frac{1}{2}]$. Show that your method satisfies the assumptions of a relevant Theorem, but need not prove the Theorem again. You can use the numerical values of $\sin(\frac{1}{2})$, $\cos(\frac{1}{2})$, $\exp(\frac{1}{2})$, etc. in your proof.

   (b) Find an N (need not be optimal) such that $|x_n - x^*| < 10^{-30}$ for all $n \ge N$ with $x_0 = 0$ (assuming a higher precision floating point arithmetic is used).

6. (15 pts) Give a (at least) cubically convergent method to solve for $e^x - 1 = 0$. Give the formula and prove that it is at least cubically convergent (locally). If you cannot do it, do the same for a locally (at least) quadratically convergent method for 1/3 partial credit.

7. (15 pts) Another way of computing π is given by the Wallis formula

$$\frac{\pi}{2} = \prod_{n=1}^{\infty} \left(\frac{(2n)^2}{(2n-1)(2n+1)} \right).$$

The N-term approximation is therefore given by

$$\frac{\pi_N}{2} = \left(\frac{2\cdot 2}{1\cdot 3}\right)\left(\frac{4\cdot 4}{3\cdot 5}\right)\left(\frac{6\cdot 6}{5\cdot 7}\right)\cdots\left(\frac{2N\cdot 2N}{(2N-1)\cdot(2N+1)}\right)$$

To prevent overflow too quickly, it is better to evaluate the last multiplication as $*(2N)/(2N-1)*2N/(2N+1)$. Find the rate of convergence of $\lim_{n\to\infty}\pi_n = \pi$ numerically. Extra points without using the limit π explicitly.

Hint: The convergence is slow, try not to produce all the data points. For example, $\pi_{100}, \pi_{200}, \cdots, \pi_{10000}$ should be enough to analyze.

To check if an integer j is a multiple of 100 or not, you can use mod or check round$(j/100)*100$.

8. (15 pts) Use any method to solve the nonlinear system of equations

$$\sin(x) + \frac{2y}{1+x} = 0.01, \qquad 5x + \sin(\frac{6y}{1+y^2}) = 0.02.$$

Write your answer in the format of 'format long e'.

Hint: the solution is near $(0,0)$ where $\sin x \approx x$, $\frac{y}{1+x} \approx y$, etc. to leading orders.