CS 450
Assignment 1

Howard Yang

Sept. 29, 2023

Question-1: Answer the following question:

(a) What three properties characterize a wellposed problem?
(b) List three sources of error in scientific computation
(c) Explain the distinction between truncation (or discretization) and rounding.

Question-2: What is an inverse problem? How are the conditioning of a problem and its inverse related?

Question-3: Which of the following two mathematically equivalent expressions The number e can be defined by $e = \sum_{n=0}^{\infty}(1/n!)$, where $n! = n(n-1)\cdots 2 \cdot 1$ for $n \neq 0$ and $0! = 1$. Compute the absolute error and relative error in the following approximations of e :

(a)

$$\sum_{n=0}^{5} \frac{1}{n!}$$

(b)

$$\sum_{n=0}^{10} \frac{1}{n!}$$

Question-4: Perform the following computations (i) exactly, (ii) using three-digit chopping arithmetic, and (iii) using three-digit rounding arithmetic. (iv) Compute the relative errors in parts (ii) and (iii) :

(a)

$$\frac{4}{5} + \frac{1}{3}$$

(b)

$$\frac{4}{5} \cdot \frac{1}{3}$$

(c)

$$\left(\frac{1}{3} - \frac{3}{11}\right) + \frac{3}{20}$$

(d)

$$\left(\frac{1}{3} + \frac{3}{11}\right) - \frac{3}{20}$$

Question-5: Let

$$f(x) = \frac{e^x - e^{-x}}{x}$$

(a) Find $\lim_{x\to 0}\left(e^x - e^{-x}\right)/x$.
(b) Use three-digit rounding arithmetic to evaluate $f(0.1)$.
(c) Replace each exponential function with its third Maclaurin polynomial, and repeat part (b).
(d) The actual value is $f(0.1) = 2.003335000$. Find the relative error for the values obtained in parts (b) and (c).

Question-6: Use the 64-bit long real format to find the decimal equivalent of the following floating-point machine numbers.
(a) 0 10000001010 1001001100000000000000000000000000000000000000000000
(b) 1 10000001010 1001001100000000000000000000000000000000000000000000

Question-7: The sine function is given by the infinite series

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$$

(a) What are the forward and backward errors if we approximate the sine function by using only the first term in the series, i.e., $\sin(x) \approx x$, for $x = 0.1, 0.5$, and 1.0?
(b) What are the forward and backward errors if we approximate the sine function by using the first two terms in the series, i.e., $\sin(x) \approx x - x^3/6$, for $x = 0.1, 0.5$, and 1.0?

Question-8:
(a) Which of the two mathematically equivalent expressions

$$x^2 - y^2 \quad \text{and} \quad (x - y)(x + y)$$

can be evaluated more accurately in floating-point arithmetic? Why?
(b) For what values of x and y, relative to each other, is there a substantial difference in the accuracy of the two expressions?

Question-9: (Programming Questions) Write a program to compute the absolute and relative errors in Stirling's approximation

$$n! \approx \sqrt{2\pi n}(n/e)^n$$

for $n = 1, \ldots, 10$. Does the absolute error grow or shrink as n increases? Does the relative error grow or shrink as n increases?

Question-10: (Programming Questions) In most floating-point systems, a quick approximation to the unit roundoff can be obtained by evaluating the expression

$$\epsilon_{\text{mach}} \approx |3 * (4/3 - 1) - 1|.$$

(a) Explain why this trick works.

(b) Try it on a variety of computers (in both single and double precision) and calculators to confirm that it works.

(c) Would this trick work in a floating-point system with base $\beta = 3$?

Question-11: (Programming Questions) Write a program to compute the mathematical constant e, the base of natural logarithms, from the definition

$$e = \lim_{n \to \infty} (1 + 1/n)^n.$$

Specifically, compute $(1 + 1/n)^n$ for $n = 10^k$, $k = 1, 2, \ldots, 20$. If the programming language you use does not have an operator for exponentiation, you may use the equivalent formula

$$(1 + 1/n)^n = \exp(n \log(1 + 1/n)),$$

where exp and log are built-in functions. Determine the error in your successive approximations by comparing them with the value of exp(1). Does the error always decrease as n increases? Explain your results.