Math 371 — HW #1
August 30, 2018

Exercises on approximation by series

In order to obtain a bound for the errors when using series to approximate a function, one way is to use the remainder term in the Taylor/Maclaurin formula. If the series is alternating in sign then use the fact that the remainder is bounded in absolute value by the absolute value of the first discarded term.

1. Develop the first two nonzero terms and the error term in the Taylor series for $\ln[1 - (x/2)]$. Aproximate $\ln(.9998)$ and obtain a bound for the error.

2. How many terms are needed to approximate e using e^x with an error not exceeding 10^{-8}? Your answer should read: At least (number) terms.

3. It is known that
$$\pi = 4 - 8\sum_{k=1}^{\infty}(16k^2 - 1)^{-1}.$$
How many terms would be needed to approximate π to within 10^{-16}?

Exercises on floating-point numbers and roundoff errors

Starred exercises are challenging!

1. Show by example that $fl[fl(xy)z] \neq fl[xfl(yz)]$, where x, y, z are normalized machine numbers. Thus multiplication in finite precision is not associative in general. Hint: Decide which type of arithmetic you want to simulate on your calculator, say 2 decimal digit with rounding, and then find three machine numbers for which the above multiplication is not associative.

2. Convert 1/9 into binary and find the machine numbers immediately to its left and to its right in a binary computer with a 43-bit normalized mantissa.

3.$^{(\star)}$ Suppose a real number $x \in (0, 1)$ has two different infinite representations
$$x = 0.a_1a_2a_3\ldots, \qquad \text{and} \qquad x = 0.b_1b_2b_3\ldots,$$
in base 10 and $a_1 \neq 0, b_1 \neq 0$. Show that there is an integer $m \geq 1$ such that

   (i) $a_k = b_k, \quad k = 1, \ldots, m-1.$

   (ii) $a_m = b_m + 1$

   (iii) $a_k = 0$ and $b_k = 9, \quad k = m+1, \ldots.$

   (iv) Show that x is a rational number.

Hint: Use the fact that by $x = 0.a_1a_2a_3\ldots$, we mean the real number that has value $\frac{a_1}{10} + \frac{a_2}{10^2} + \frac{a_3}{10^3} + \ldots$ and facts about geometric series.

4. Perform a roundoff error analysis to estimate the error, in terms of the machine unit roundoff u, committed in computing $3\pi + 7e$.

   In the next few problems we are working with the Hypothetical Marc-32 computer studied in class ($t = 24$).

5. $^{(\star)}$ Show that if x and y are two machine numbers such that $|y| \leq |x| \times 2^{-25}$, then $fl(x+y) = x$.

6. Show that $(64.015625)_{10}$ is a machine number. Determine its bit pattern and hexadecimal representation.

7. Which of these are machine numbers?

$$\text{(a) } 10^{403}, \quad \text{(b) } 1 + 2^{-32}, \quad \text{(c) } 1/5, \quad \text{(d) } 1/256.$$

8. Determine the decimal numbers that have the following machine representations

$$\text{(a) } [\text{CA3F2900}]_{16}, \quad \text{(b) } [\text{4B187ABC}]_{16}.$$

9. What is the significance of the following numbers: $[\text{7F7FFFFF}]_{16}$ and $[\text{00800000}]_{16}$?

10. Show that any positive real number can be expressed as $r \times 2^{2n}$ where n is an integer and $\frac{1}{4} \leq r \leq 1$. The significance of this fact is that computing the square root of any positive real number can be reduced to finding the square root of a number belonging to $[\frac{1}{4}, 1]$. This is described as *range reduction*.

Exercises on loss of significance

1. Determine the first two nonzero terms in the expansion about zero of the function

$$f(x) = \frac{\tan x - \sin x}{x - \sqrt{1 + x^2}}.$$

   Use these to approximate $f(.0125)$. Compare this with the "direct" value of $f(.0125)$. Note: Use/simulate 4-decimal digit arithmetic with rounding for both calculations.

2. Find a way to calculate accurate values for

$$f(x) = \frac{\sqrt{1 + x^2} - 1}{x^2} - \frac{x^s \sin x}{x - \tan x}.$$

   Determine $\lim_{x \to 0} f(x)$.

3. Find a way to calculate

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

   accurately. Determine $f(.008)$ correctly to ten decimal places.