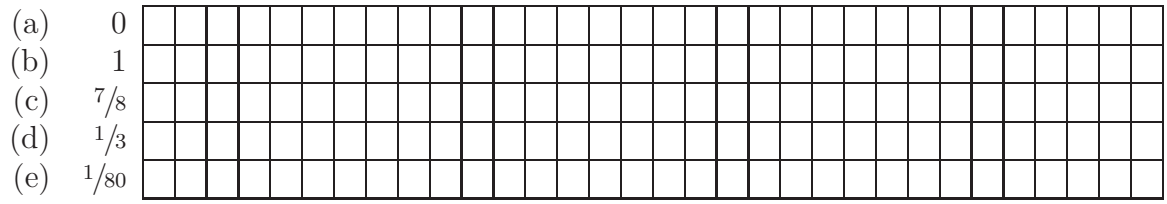
Tufts University
Department of Mathematics
Fall 2018

MA 126: Numerical Analysis

Homework 3 (v1.1) [1]

Assigned Friday 21 September 2018
Due Friday 28 September 2018 at 3 pm

1. Write the IEEE single-precision (32-bit) floating-point representation for the following numbers by filling in the boxes with zeros and ones:



   Remember that the first (leftmost) bit is the sign bit, followed by eight bits of exponent and 23 bits of mantissa (significand).

2. A 64-bit *signed integer* has a sign bit, followed by 63 bits that represent the magnitude of the integer as a base-two number.

   (a) How many distinct 64-bit signed integers are there?

   (b) What is the smallest possible 64-bit signed integer? The largest?

   (c) Suppose that you have a list of n 64-bit numbers. Show that a sort of these numbers interpreted as signed integers will yield the same result as a sort of these numbers interpreted as double-precision floating-point numbers.

3. Find the smallest nonzero root of the equation $\tan x = x$, to within an accuracy of 10^{-15}, using your own implementation of

   (a) the bisection method,

   (b) Newton's method,

   in a computer language of your choice. Note that you will have to find an interval containing the root in the first case, and a good initial guess in the second case.

4. In class it was shown that Newton's method can be used to find the root $x = 0$ of the equation $\arctan x = 0$, as long as the initial guess satisfies $x_0 \in (-a, +a)$ for a sufficiently small. Find the largest possible value of a for which this is true to within an accuracy of 10^{-15}.