Math 310 – Numerical Analysis

Homework Problems due ?????

Pretend that a computer can only represent the floating point numbers 0, $\pm\infty$, and those which in base 2 have the form

$$\pm 1.a_1a_2 \times 2^m,$$

where a_1, a_2 are binary digits (i.e., 0 or 1) and m is an integer with $-2 \leq m \leq 2$. When a real number x is provided as input, it is converted to the machine number $fl(x)$, which is the closest number to x of the above form. When an operation such as addition is performed on inputs x and y, they are first converted to machine numbers, then added exactly, and finally the sum is converted to a machine number, so that the output is $fl(fl(x) + fl(y))$.

1. Give decimal or rational expressions for all 20 of the finite positive machine numbers. Then illustrate them all on a number line.

2. Express 3.5 and 6.5 in base 2. Is either of these a machine number? Find $fl(3.5)$ and $fl(6.5)$.

3. Express $\frac{7}{3}$ in base 2. Find $fl\left(\frac{7}{3}\right)$.

4. What is machine epsilon ϵ_M for this computer? Recall that ϵ_M it is the largest number such that $fl(1 + \epsilon) = 1$.

5. Find $fl(2 + \epsilon_M)$ and $fl\left(\frac{1}{2} + \epsilon_M\right)$.

6. Give examples of machine numbers x and y (other than 0 or $\pm\infty$) such that:

   (a) $fl(x + y) = x$

   (b) $fl(x \cdot y)$ produces an overflow

   (c) $fl(x + y)$ produces an overflow

   (d) $fl(x/y)$ produces an underflow

7. Give examples of real numbers x and y such that:

   (a) $fl(x + y) \neq fl(fl(x) + fl(y))$

   (b) $fl(x \cdot y) \neq fl(fl(x) \cdot fl(y))$