High Performance Computing
Homework 2

Francisco Jose Castillo Carrasco

September 17, 2018

Problem 1

(a) What is the normalized IEEE single-precision representation of the number 5.5?

Solution: We can find the binary representation of the number as follows. First, the integer part

$$5/2 = 2 + 1;$$
$$2/2 = 1 + 0;$$
$$1/2 = 0 + 1;$$

which implies

$$5_{10} = 101_2.$$

Second, the decimal part

$$0.5 \times 2 = 1 + 0;$$

which gives

$$0.5_{10} = 0.1_2.$$

Finally, we have that $5.5_{10} = 101.1_2$. We express that in the normalized format and obtain

$$5.5_{10} = 1.011000\ldots000 \times 2^2.$$

This means that $a_2 = a_3 = 1$ and the rest of a_j are zero.

(b) If we change the significand of 5.5 by one ulp, by how much does the value of the floating point representation change? Express your answer as a power of 2.

Solution: We change the *unit of last place*, a_{23}, from 0 to 1. This produces a change in the value of $\Delta x = 2^{-23+2} = 2^{-21}$

Problem 2

Let $x = 1/3$. (a) Find the binary representation of x.

Solution: Similarly as in the previous problem, we obtain

$$x = \frac{1}{3}_{10} = 0.0101\overline{01}\ldots$$

(b) Find the IEEE-754 single precision representation $\hat{x}$ of x when rounding to nearest.

Solution: To find the normalized representation we must make $a_0 = 1$ such as

$$1.01\overline{01}\cdots \times 2^2,$$

and cut at the $23rd$ digit,

$$\hat{x} = 1.01010101010101010101011 \times 2^2.$$

Note that we have rounded to nearest in a_{23}.

(c) What is the absolute error due to rounding? In other words, what is $|\hat{x} - x|$?

Solution: We begin using geometric series to represent both quantities as

$$x = S = \frac{a}{1-r},$$

and

$$\hat{x} = S_N + 2^{-23} = \frac{a(1 - r^{N+1})}{1-r} + 2^{-23}$$

with $a = r = 1/4$ and $N = 11$. Therefore,

$$|\hat{x} - x| = |S_N + 2^{-23} - S| = \left|2^{-23} - \frac{r^{N+1}}{1-r}\right| = \left|2^{-23} - \frac{(1/4)^{N+1}}{3/4}\right| = 9.934 \cdot 10^{-9}.$$

Problem 3

How does the spacing depend on e?

Solution: We obtain the spacing by focusing in the last digit a_{23} that is multiplied by 2^e. Hence,

$$1\ ulp = 2^{-23}2^e = 2^{e-23}.$$

Problem 4

How many possible nonnegative normalized IEE single precitions floating point numbers are there?

Solution: Since we have 24 bits of mantisa and one of them is fixed, we have 2^{23} possibilities due to the mantissa. The exponent multiplies those possibilities by 127×2. Hence, the number of possible nonnegative normalized IEE single precitions floating point numbers is $N = 254 \times 2^{23}$

Problem 5

Consider IEEE single-precision representations. (a) Is $1,000,000.0$ exactly representable in IEEE single precision?

Solution: Yes, we find the representation doing the same as in Problem 1, although this is a much longer case. We obtained

$$1,000,000.0 = 1.11101000010010000000000 \times 2^{19}$$

(b) What is the smallest positive integer M that does not have an exact IEEE single-precision representation?

Solution: Since we have 24 bits of mantisa, the largest value that those bits can represent is $2^{24} = 16777216$. Therfore the next integer will not be representable. Therefore the smallest positive integer M that does not have an exact IEEE single-precision representation is

$$M = 2^{24} + 1 = 16,777,217.$$

Problem 6

True or False: If x has a terminating base-2 expansion, then x has a terminatiing base-10 expansion.

Solution: Let $x = p/q$. Since x has base-2 expansion, q divides some power of 2,

$$\frac{2^e}{q} = K \in \mathbb{Z}.$$

Now we have some power e of 10 divided by q,

$$\frac{10^e}{q} = \frac{2^e 5^e}{q} = K5^e = C \in \mathbb{Z}.$$

Hence, if q divides some power of 2, it also divides some that same power of 10 (since 10 is multiple of 2). Therefore, since q divides some power of 10, $x = p/q$ has a terminating base-10 expansion.

The assertion is TRUE.

Problem 7

Solution:

Problem 8

Solution: