Assignment 1

Name : Shihab Muhtasim

ID : 21301610

Sec : 3 (MAA)

Course : CSE330

Ans to qr 1

(a)

Given,
$\beta = 2$, $m = 4$, $-4 \leq e \leq 2$.

In this system,

① Lecture note form:

maximum number $= \pm (0.1111)_2 \times 2^2$

② Normalized form:

maximum number $= \pm (1.1111)_2 \times 2^2$

③ Denormalized form:

maximum number $= \pm (0.11111)_2 \times 2^2$

(b)

similarly for the system,

① Lecture note form :

minimum number $= +(0 \cdot 1000)_2 \times 2^{-4}$

② Normalized form :

minimum number $= +(1 \cdot 0000)_2 \times 2^{-4}$

③ Denormalized form :

minimum number $= +(0 \cdot 10000)_2 \times 2^{-4}$

$\times 2^{-4}$

(c)

Using eq 1, for $e = -3$ the numbers generated are,

1) $(0.1000)_2 \times 2^{-3} = 0.5 \times 2^{-3} = 0.0625 = \frac{1}{16}$

2) $(0.1001)_2 \times 2^{-3} = \frac{9}{16} \times 2^{-3} = 0.0703 = \frac{9}{128}$

3) $(0.1010)_2 \times 2^{-3} = 0.625 \times 2^{-3} = 0.078 = \frac{5}{64}$

4) $(0.1011)_2 \times 2^{-3} = 0.6875 \times 2^{-3} = 0.085 = \frac{11}{128}$

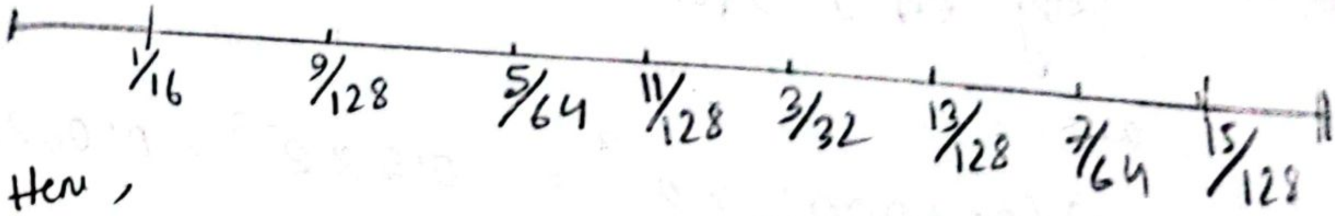
5) $(0.1100)_2 \times 2^{-3} = 0.75 \times 2^{-3} = 0.093 = \frac{3}{32}$

6) $(0.1101)_2 \times 2^{-3} = 0.8125 \times 2^{-3} = 0.101 = \frac{13}{128}$

7) $(0.1110)_2 \times 2^{-3} = 0.875 \times 2^{-3} = 0.109 = \frac{7}{64}$

8) $(0.1111)_2 \times 2^{-3} = 0.9375 \times 2^{-3} = 0.117 = \frac{15}{128}$

Number line is given below:



Here,
The number line is equally spaced as the diffrence between all the numbers to its consecutive one is $\frac{1}{128}$.

However, when considering value of e to be lower / higher, it can differ.

Ans to Q no 2

(a)

Given, $\beta = 2$, $m = 5$, $e_{min} = -2$, $e_{max} = 5$

In normalized form:

minimum number, $|x| = +(1.00000)_2 \times 2^{-2}$

$= +(0.25)_{10}$

In denormalized form:

minimum number, $|x| = +(0.100000)_2 \times 2^{-2}$

$= +(0.125)_{10}$

(b)

Calculating machin epsilon :
for both normalized and denormalized ,

$$\epsilon_m = \frac{1}{2} \times \beta^{-m}$$

Given, $\beta = 2$, $m = 5$,

∴ Machine epsilon for both normalized and denormalized will be,

$$\epsilon_m = \frac{1}{2} \times 2^{-5} = \frac{1}{64} = 0.015625$$

(c)

Maximum delta is the machine epsilon value which will be for the normalized form (eq 2) :

$$\epsilon_m = \frac{1}{2} \times 2^{-5} = \frac{1}{64} = 0.015625$$

CamScanner

lized ,

Ans to qr 3

(a)

i) $(2 \cdot 23)_{10}$ to binary :-

$(10 \cdot 0011101011110000101)_2 \times 2^0$

Representation in normalized form :

$(1 \cdot 00011101011110000101)_2 \times 2^1$

Since the system allows $m = 3$, finding the closest value :

$(2)_{10}$ — $2 \cdot 23$ — $(2 \cdot 25)$

$(1 \cdot 000)_2 \times 2^1$ — $(1 \cdot 000111 \ldots\ldots)_2 \times 2^1$ — $(1 \cdot 001)_2 \times 2^1$

We can see that $(m+1)^{th}$ value is 1 :

$\therefore$ Rounding up $= (1 \cdot 001)_2 \times 2^1$

(ii) $(2.2018)_{10} = (10.0011000111010100100\ldots)_2 \times 2^0$

Normalized : $(1.00011000111010100100 11)_2 \times 2^1$

since $m = 3$

$(2)_{10}$ — $(2.2018)_{10}$ — $(2.25)_{10}$

$(1.000)_2 \times 2^1$ — $(1.00011\ldots)_2 \times 2^1$ — $(1.001)_2 \times 2^1$

Since $(m+1)$th val is 1 :

$\therefore$ Rounded value $= (1.001)_2 \times 2^1$

(b)

i) For $(2.23)_{10}$:

Rounded value, $fl(x) = (2.25)_{10} = (1.001)_2 \times 2^1$

actual value, $x = (2.23)_{10} = (1.0001110101110000101)_2 \times 2^1$

$\therefore$ Rounding error $= |fl(x) - x| = |$

$= |(1.001)_2 \times 2^1 - (1.0001110101110000101)_2 \times 2^1|$

$= (0.0000001010001111)_2 \times 2^1$

ii) For $(2.2018)_{10}$:

Rounded val, $fl(x) = (2.25)_{10} = (1.001)_2 \times 2$

actual val, $x = (2.2018)_{10} = (1.00011001110101 0010011)_2 \times 2^1$

$\therefore$ Rounding error $= |fl(x) - x| = |$

$=$

(i) for ...

9

(C)

... this system and denormalized form,

$max = (0 \cdot 1111)_2 \times 2^2 = (3 \cdot 75)_{10}$

$min = (0 \cdot 1000)_2 \times 2^{-2} = \cancel{(2)_{10}} (0 \cdot 125)_{10}$

∴ The numbers are within the range. So it is representable.

(1) $(2 \cdot 23)_{10}$

$(2 \cdot 23)_{10} = (10 \cdot 0011101011100000101)_2 \times 2^0$

In denormalized form:

$(0 \cdot 10001110101110000101)_2 \times 2^2$

since, m= 3 and (m+1)th val is 1:

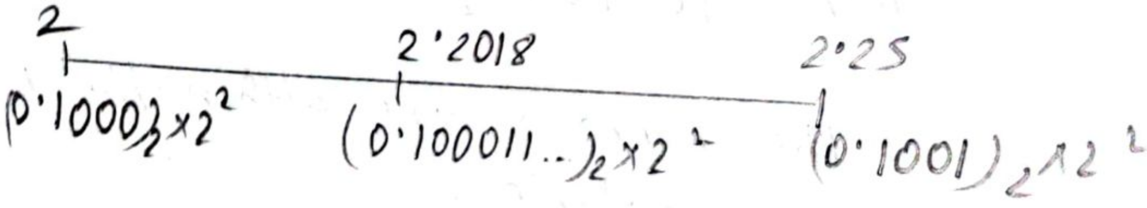
$$\begin{array}{ccc} 2 & 2 \cdot 23 & 2 \cdot 25 \\ (0 \cdot 1000)_2 \times 2^2 & (0 \cdot 1000111\ldots)_2 \times 2^2 & (0 \cdot 1001)_2 \times 2^2 \end{array}$$

∴ Rounded $= (0 \cdot 1001)_2 \times 2^2$

(ii) $(2.2018)_{10} = (10.0011001110101001 0011)_2 \times 2^0$

Denormalized $= (0.10001100111010100 10011)_2 \times 2^2$

Since $m = 3$ and $(m+1)^{th}$ val is 1 :



$\therefore$ Rounded $= (0.1001)_2 \times 2^2$

(b)

(i) For $(2 \cdot 23)_{10}$:

Rounded value, $fl(x) = (1 \cdot 001)_2 \times 2^1$

$= (2 \cdot 25)_{10}$

Actual value, $x = (2 \cdot 23)_{10}$

$\therefore$ Rounding error $= |fl(x) - x|$

$= |2 \cdot 25 - 2 \cdot 23|$

$= (0 \cdot 02)_{10}$

$= (0 \cdot 000001010001111011)_2 \times 2^0$

$= (1 \cdot 010001111011)_2 \times 2^{-6}$

$= (1 \cdot 010)_2 \times 2^{-6}$

(ii) For $(2 \cdot 2018)_{10}$:

Rounded value, $fl(x) = (1 \cdot 001)_2 \times 2^1$

$= (2 \cdot 25)_{10}$

Actual value, $x = (2 \cdot 2018)_{10}$

Rounding error $= |fl(x) - x|$

Rounding error = $|2.25 - 2.2018|$

$= (0.0482)_{10}$

$= (0.000011000101011011 01)_2$

$= (1.1000101011011 01)_2 \times 2^{-5} \times 2^{0}$

$= (1.100)_2 \times 2^{-5}$