CSE 330

Numerical Methods

Name : Udoy Saha

ID : 21301095

Section : 11

Ans to the ques no:- 1

Given that,

$\beta = 2$

$m = 4$

$-4 \leq e \leq 2$

a

Maximum number in,

Lecture Note Form : $(0 \cdot 1 1 1 1)_2 \times 2^2$

$= (3 \cdot 75)_{10}$

Normalized form $= (1 \cdot 1 1 1 1)_2 \times 2^2$

$= (7 \cdot 75)_{10}$

Denormalized form $= (0 \cdot 1 1 1 1 1)_2 \times 2^2$

$= (3 \cdot 875)_{10}$

(Ans)

b Non negative ~~ae~~ minimum number in,

Lecture Note form : $(0 \cdot 1 0 0 0)_2 \times 2^{-4}$

$= (0 \cdot 03125)_{10}$

Normalized form : $(1.0000)_2 \times 2^{-4}$

$= (0.0625)_{10}$

Denormalized form : $(0.10000)_2 \times 2^{-4}$

$= (0.03125)_{10}$

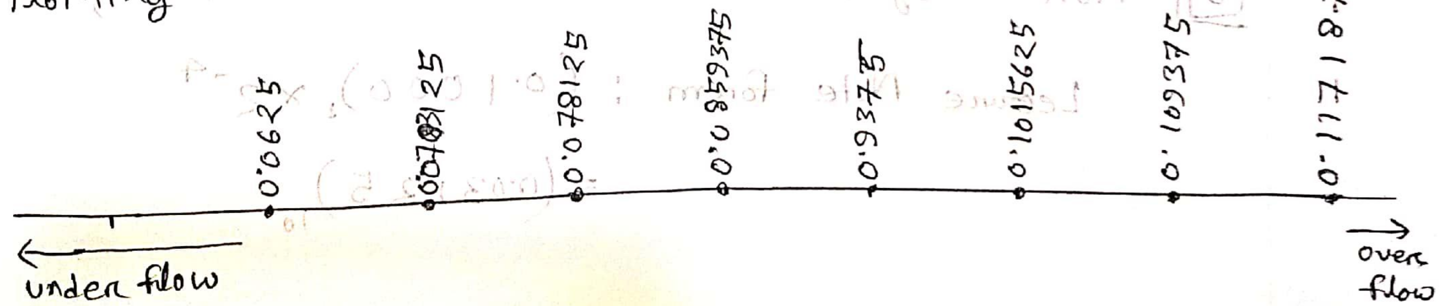
(Ans)

$\boxed{c}$ For Eq.(1), if $e = -3$, the numbers will be in the form $\rightarrow (0.1\_\_\_\_)_2 \times 2^{-3}$

Finding combinations,

$(0.1000)_2 \times 2^{-3} = 0.0625$	$(0.1100)_2 \times 2^{-3} = 0.9375$
$(0.1001)_2 \times 2^{-3} = 0.0703125$	$(0.1101)_2 \times 2^{-3} = 0.1015625$
$(0.1010)_2 \times 2^{-3} = 0.078125$	$(0.1110)_2 \times 2^{-3} = 0.109375$
$(0.1011)_2 \times 2^{-3} = 0.0859375$	$(0.1111)_2 \times 2^{-3} = 0.1171875$

Plotting them on real line,



The number line will be equally spaced. Because, the difference of every number is $\rightarrow$ $(0\cdot0001)_2 \times 2^{-3}$

$= (0\cdot0078125)_{10}$

(Ans)

Ans to the ques no:- 2

Given that,

$\beta = 2$

$m = 5$

$-2 \leq e \leq 5$

$\boxed{a}$ Minimum $|x|$ in the forms,

Normalized : $(1\cdot00000)_2 \times 2^{-2}$

$= (0\cdot25)_{10}$

Denormalized : $(0.100000)_2 \times 2^{-2}$

$= (0.125)_{10}$

(Ans)

[b] For the Normalized form, lets take two values (Adjacent), which are $(1.00000)_2 \times 2^e$ and, $(1.00001)_2 \times 2^e$.

$\therefore$ Machine epsilon, $\varepsilon_m = \frac{1}{2}\left[(1.00001)_2 \times 2^e - (1.00000)_2 \times 2^e\right]$

$= \frac{1}{2} \cdot (0.00001)_2 \times 2^e$

$= \frac{1}{2} \cdot 2^{-5} \cdot 2^e$

$= \frac{1}{2} \cdot 2^{e-5}$

$= \frac{1}{2} \cdot 2^{1-5}$ $\left[\because L(x) = \beta^{-1}\right]$

$= 0.03125$

Similarly for the denormalized form, the machine epsilon will be,

$$\varepsilon_m = \frac{1}{2}\left[(0.100001)_2 \times 2^e - (0.100000)_2 \times 2^e\right]$$

$$= \frac{1}{2} \times (0.000001)_2 \times 2^e$$

$$= \frac{1}{2} \times 2^{-6} \times 2^e$$

$$= \frac{1}{2} \times 2^{e-6}$$

$$= \frac{1}{2} \times 2^{1-6} \qquad [\because \lfloor |x| \rfloor \equiv \beta^{-1}]$$

$$= 0.015625$$

(Ans)

c) We know, $|\delta| \leq \varepsilon_m$

Now, Eq.(2) is the Normalized form of floating point presentation.

From 'b', the ε_m for this system = 0.03125

$\therefore$ Maximum $|\delta| = 0.03125$

(Ans)

— o — x — o —

Ans. to the ques: no:- 3

Given that,

$$\beta = 2$$

$$m = 3$$

$$-2 \le e \le 2$$

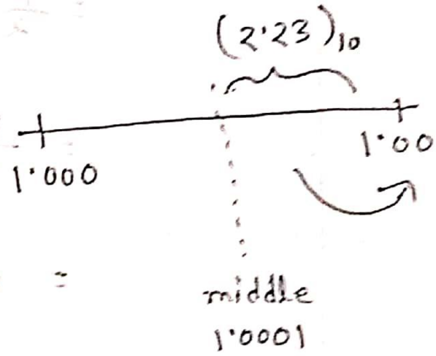
(a)

$$Fl\left((2.23)_{10}\right)$$

$$= Fl\left((10.00111)_2\right)$$

$$= Fl\left((1.000111)_2 \times 2^1\right)$$

$$= (1.001)_2 \times 2^1$$

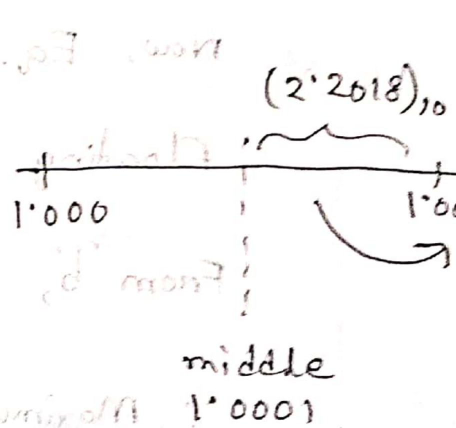


$$Fl\left((2.2018)_{10}\right)$$

$$= Fl\left((10.0011001 1)_2\right)$$

$$= Fl\left((1.00011 0011)_2 \times 2^1\right)$$

$$= (1.001)_2 \times 2^1$$



(Ans)

(b) For $(2.23)_{10}$,

$$|\delta| = \frac{|Fl(x) - x|}{|x|}$$

$$= \frac{|(1.001)_2 \times 2^1 - (2.23)_{10}|}{|(2.23)_{10}|}$$

$$= \frac{|2.25 - 2.23|}{|2.23|}$$

$$= 0.00896860986$$

For $(2.2018)_{10}$,

$$|\delta| = \frac{|(1.001)_2 \times 2^1 - (2.2018)_{10}|}{|(2.2018)_{10}|}$$

$$= \frac{|2.25 - 2.2018|}{|2.2018|}$$

$$= 0.0218911 7924$$

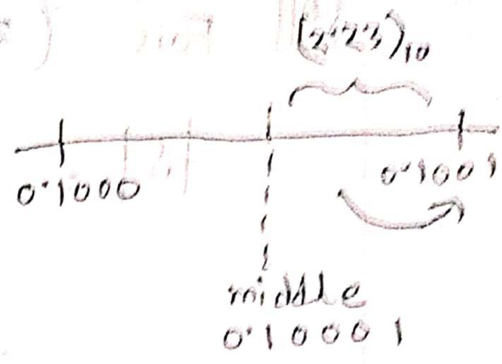
(Ans)

c) $Fl\left((2.23)_{10}\right)$

$= Fl\left((10.0011101)_2\right)$

$= Fl\left((0.100011101)_2 \times 2^2\right)$

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



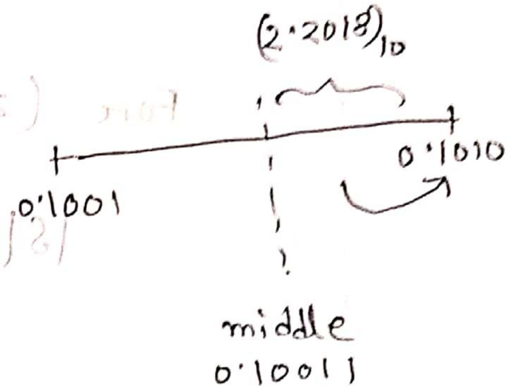
$\therefore$ $(2.23)_{10}$ or $(10.0011101)_2$ is not exactly representable in the system, but can be represented as $(0.1001)_2 \times 2^2$

$Fl\left((2.2018)_{10}\right)$

$= Fl\left((10.001100111)_2\right)$

$= Fl\left((0.1001100111)_2 \times 2^2\right)$

$= (0.1010)_2 \times 2^2$



$\therefore$ $(2.2018)_{10}$ or $(10.001100111)_2$ is not exactly representable in the system, but can be represented as $(0.1010)_2 \times 2^2$

— o — x — o —