HW-1

1.1

a) i $(110111001.101011101)_2$

$2^7+(2^6\times 0)+2^5+2^4+2^3+(0\times 2^2)+(0\times 2^1)+2^0 = 185$

$2^{-1}+(0\times 2^{-2})+2^{-3}+(0\times 2^{-4})+(2^{-5})+2^{-6}+2^{-7}+(0\times 2^{-8})+2^{-9}$

$= 0.681640625$

Ans $= 185.681640625$

ii) $(1001100101.01101)_2$

$2^0+(0\times 2^1)+2^2+(0\times 2^3)+(0\times 2^4)+(2^5)+2^6+(0\times 2^7)+(0\times 2^8)+2^9$

$= 613$

$(0\times 2^{-1})+(2^{-2})+2^{-3}+(0\times 2^{-4})+2^{-5} = 0.40625$

Ans $= 613.40625$

iii) $(101.01)_2$

$2^0+(0\times 2^1)+2^2 = 5 \mid (2^{-1}\times 0)+(2^{-2}) = .25$

Ans 5.25

6)

i $(100.01)_{10}$

$100/2 \rightarrow 0$ | $50/2 \rightarrow 0$ | $25/2 \rightarrow 1$ | $12/2 \rightarrow 0$ | $6/2 \rightarrow 0$ | $3/2 \rightarrow 1$ | $1/2 \rightarrow 1$ |

~~.01/2 → 0 | .02/2 → 0 | .04/2 → 0 |~~ 1100100

$.01 \times 2 \rightarrow 0$ | $.02 \times 2 \rightarrow 0$ | $1.04 \times 2 \rightarrow 01$ | $.08 \times 2 \rightarrow 0$ | $.16 \times 2 \rightarrow 0$ | $.32 \times 2 \rightarrow 0$ | $.64 \times 2 \rightarrow 1$ | $1.28 \times 2 \rightarrow 0$ | $.56 \times 2 \rightarrow 1$ |

$1.12 \times 2 \rightarrow 0$ |

0.0000001010

Ans = 1100100.0000001010

ii $(64.625)_{10}$

$64 \rightarrow 0$ | $32 \rightarrow 0$ | $16 \rightarrow 0$ | $8 \rightarrow 0$ | $4 \rightarrow 0$ | $2 \rightarrow 0$ | $1 \rightarrow 0$ | $\rightarrow$ 1000 000

$.625 \rightarrow 1$ | $1.25 \rightarrow 0$ | $.5 \rightarrow 1$ | 101

Ans = 1000 000.101

iii $(25)_{10}$

$25 \rightarrow 1$ | $12 \rightarrow 0$ | $6 \rightarrow 0$ | $3 \rightarrow 1$ | $1 \rightarrow 1$ | 11001 = Ans

C) $(64.625)_{10} = (1000000.101)_2$

$= 1.000000101 \times 2^6$

Sign bit	Exponent	$(133)_{10} = (10000101)_2$
0	$2^7 - 1 + 6 = 133$	

Ans

$\underbrace{0}_{\text{sign}} \quad \underbrace{10000101}_{\text{8-bit}} \quad \underbrace{00000010100000000000000}_{\text{23 bits}}$

1.2)

A) $z = xy \qquad fl(x) = x(1+\delta_x) \qquad fl(y) = y(1+\delta_y)$

$fl(z) = fl(fl(x) * fl(y)) \cancel{= z(1+\delta_z)}$

$= (xy(1+\delta_x)(1+\delta_y)(1+\delta_z)) \approx (xy)(\underbrace{\delta_x\delta_y\delta_z}_{=0} + \underbrace{\delta_x\delta_y}_{=0} + \underbrace{\delta_x\delta_z}_{=0} + \underbrace{\delta_y\delta_z}_{=0} + \delta_x + \delta_y + \delta_z + 1)$

$\cancel{\text{Abs Error} = fl(z) - z = xy(1+\delta_x)(1+\delta_y)(1+\delta_z) - xy}$

relative ~~Error~~ $= \cancel{\delta_z} = \cancel{\dfrac{fl(z)-z}{z} = (1+\delta_x)(1+\delta_y)(1+\delta_z) - 1}$

$= xy(1 + \delta_x + \delta_y + \delta_z)$

$fl(z) - z = \cancel{xy} + xy\delta_x + xy\delta_y - \cancel{xy} + xy\delta_z = xy(\delta_x + \delta_y) + xy\delta_z$

$\cancel{\delta_z} = \dfrac{fl(z)-z}{z} = \delta_x + \delta_y + \delta_z - 1 =$ relative error.

B $z = 5x + 7y$ | $fl(x) = x*(1+\delta_x)$ | $fl(y)\ y(1+\delta_y)$

$fl(z) = fl(5fl(x) + 7fl(y))$

$= \left[\left(5x(1+\delta_x)\right) + \left(7y(1+\delta_y)\right)\right](1+\delta_z)$

$fl(z) = (5x + 7y) + [5x(\delta_x + \delta_z)] + [7y(\delta_y + \delta_z)]$

Abs Error $= fl(z) - z$

$= \cancel{5x+} = 5x\delta_x + 7y\delta_y + [\delta_z(5x + 7y)]$

Relative Error

$$= \frac{fl(z) - (5x+7y)}{z} = \frac{5x\delta_x + 7y\delta_y}{5x+7y} + \delta_z$$

1.3

a)

$$f(x) = \sqrt{x^2+2x+2} - x - 1$$

For $x >> 0$ $\quad f(x) = \sqrt{(x+1)^2+1} - x - 1$

$$(x+1)^2 + 1 \simeq (x+1)^2$$

$$x + 1 - x - 1 = 0$$

$1 >$ $f(x)$ will be zero.

Rationalise $\dfrac{\left(\sqrt{x^2+2x+2} - (x+1)\right)\left(\sqrt{x^2+2x+2} + (x+1)\right)}{\sqrt{x^2+2x+2} + (x+1)}$

$$= \frac{\cancel{x^2} + \cancel{2x} + 2 - \cancel{x^2} - \cancel{2x} - 1}{\sqrt{x^2+2x+2} + x + 1} = \frac{1}{\sqrt{x^2+2x+2} + x + 1}$$

We wont loose sig fig due to ~~add~~ substraction.

(b) $$f(x) = \frac{1}{\sqrt{x+2} - \sqrt{x}}$$

Same for $x >> 0$ denominator becomes zero. as we loose sig figs ~~from~~ due to substraction.

Sol^n → Rationalize

$$= \frac{\sqrt{x+2} + \sqrt{x}}{\cancel{x} + 2 - x} = \frac{\sqrt{x+2} + \sqrt{x}}{2}$$

c)

(i) $\frac{1}{3} + \frac{3}{4} = 0.333333 + 0.75 = 1.083333 \ldots\ldots$

(ii) $\frac{1}{3} - \frac{100}{301} = 0.33333 - 0.332226 = 0.001107$

1.4) $f'(x) \approx D_h(x) = \frac{1}{2h}[-3f(x) + 4f(x+h) - f(x+2h)]$

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f'''(x) \dots + Ch^3$$

$$f(x+2h) = f(x) + 2hf'(x) + \frac{4h^2}{2}f''(x) + \frac{8h^3}{6}f'''(x) \dots + Ch^3$$

Approx

$$D_h(x) = \frac{1}{2h}\left[-3f(x) + \left[4f(x) + 4hf'(x) + \cancel{\frac{4h^2}{2}f''(x)} \dots\right] - \left[f(x) + 2hf'(x) + \cancel{\frac{4h^2}{2}f''(x)} \dots\right]\right]$$

$$D_h(x) = \frac{1}{\cancel{2}h}\left[\cancel{-3f(x)} + \cancel{4f(x)} - \cancel{f(x)} + \cancel{2}hf'(x) + Ch^3\right]$$

$$D_h(x) = f'(x) + Ch^2$$

$$D_h(x) - f'(x) \approx Ch^2.$$