Tarea 1 Nicolas Gonzalez · 201620023010

① Valor maximo, en valor absoluto, del error relativo entre un número y su representacion binaria normalizada en punto flotante con k cifras significativas

1.1 $x = 0.d_1 d_2 d_3 \cdots d_k d_{k+1} \cdots * 2^n$

$\bar{x} = 0.d_1 d_2 d_3 \cdots d_k * 2^n$

$|\varepsilon| = \left| \frac{x - \bar{x}}{x} \right|$

$|\varepsilon| = \left| \frac{0.d_1 d_2 d_3 \cdots d_k d_{k+1} \cdots \times 2^n - 0.d_1 d_2 d_3 \cdots d_k \times 2^n}{0.d_1 d_2 d_3 \cdots d_k \times 2^n} \right|$

$(d_{k+1} = 0)$ $|\varepsilon| = \left| \frac{0.d_1 d_2 d_3 \cdots d_k d_{k+1} \cdots - 0.d_1 d_2 d_3 \cdots d_k}{0.d_1 d_2 d_3 \cdots d_k} \right|$

$|\varepsilon| = \left| \frac{0.d_{k+1} d_{k+2} d_{k+3} \cdots \times 2^{-k}}{0.d_1 d_2 d_3 \cdots d_k} \right|$

$|\varepsilon| < \left| \frac{1 \times 2^{-k}}{0.1} \right| = \underbrace{10}_{\text{binario}} \times 2^{-k} = \boxed{2^{-k+1}}$

② Dado $x = 4/5$, escribalo en base binaria y determine el numero maquina que lo representa (para la maquina de 32 bits presentada en clase). Determine el error absoluto y el error relativo en codigo binario

$x = 4/5 = 0.8 \rightarrow$

$\downarrow$

0.8×10^0

$0.8 \times 2 = \boxed{1}.6$

$0.6 \times 2 = \boxed{1}.2$

$0.2 \times 2 = \boxed{0}.4$

$0.4 \times 2 = \boxed{0}.8$

$0.8 \rightarrow 0.\overline{1100}_2$

1 sig. num

24 mantisa

1 sig. exp

6 exponente

1	100 1100 1100 1100 1100 11001	1	000000

$0.1100\,1100\,1100\,1100\,1100\,1100 = 0.7999999523$

$x = 0.8$

$\hat{x} = 0.7999999523$

$|E| = |x - \hat{x}| = 4.77 * 10^{-8} = \boxed{0.477 \times 10^{-7}}$

$\varepsilon = \frac{|E|}{x} = 5.96 \times 10^{-8} = \boxed{0.596 \times 10^{-7}}$

1.2

$$x = 0.d_1d_2d_3 \cdots d_k d_{k+1} \cdots * 2^n$$
$$\bar{x} = 0.d_1d_2d_3 \cdots d_k * 2^n$$

$$|\varepsilon| = \left| \frac{x - \bar{x}}{x} \right|$$

$(d_{k+1} + 1)$

$$= \left| \frac{0.d_1d_2 \cdots d_k d_{k+1} \cdots * 2^n - 0.d_1d_2 \cdots (d_k + 1) * 2^n}{0.d_1d_2 \cdots d_{k-1} d_k d_{k+1} \cdots * 2^n} \right|$$

$$= \left| \frac{0.d_1d_2 \cdots d_k d_{k+1} \cdots - 0.d_1d_2 \cdots (d_k + 1)}{0.d_1d_2 \cdots d_k \cdots} \right|$$

$$= \left| \frac{0.d_1d_2 \cdots d_k d_{k+1} \cdots - 0.d_1d_2d_3 \cdots d_k - 2^{-k}}{0.d_1d_2 \cdots d_k d_{k+1} \cdots} \right|$$

$$= \left| \frac{0.d_{k+1}d_{k+2} \cdots * 2^{-k} - 2^{-k}}{0.d_1d_2 \cdots d_k \cdots} \right|$$

$$= \left| \frac{(0.d_{k+1}d_{k+2} \cdots - 1) * 2^{-k}}{0.d_1d_2d_3 \cdots d_k \cdots} \right|$$

$$< \left| \frac{1 - 0.1}{0.1} 2^{-k} \right| = 1 \times 2^{-k}$$