Gerasimos Angelatos

APC 523 Pset 1

1 prove $\left|\frac{x - rd(x)}{x}\right| \leq 2^{-p}$ $\rightarrow$ $x = \pm\left(\sum_{l=1}^{\infty} b_{-l} 2^{-l}\right) \cdot 2^{e}$

now $rd(x) = \begin{cases} \left(\pm\sum_{l=1}^{p} b_{-l} 2^{-l}\right) \cdot 2^{e} : b_{-(p+1)} = 0 \\ \pm\left(\left(\sum_{l=1}^{p} b_{-l} 2^{-l}\right) + \underset{\uparrow \text{ from rounding}}{2^{-p}}\right) 2^{e} : b_{-(p+1)} = 0 \end{cases}$

$\rightarrow$ First consider $b_{-(p+1)} = 0$ (discarded bit is zero)

Then $x - rd(x) = \pm\left(\sum_{l=1}^{\infty} b_{-l} 2^{-l} - \sum_{l=1}^{p} b_{-l} 2^{-l}\right) \cdot 2^{e}$

$= \pm\left(b_{-(p+1)} \cdot 2^{-(p+1)} + \sum_{l=p+2}^{\infty} b_{-l} 2^{-l}\right) \cdot 2^{e}$, $0 = b_{-(p+1)}$

① so $x - rd(x) = \pm\sum_{l=p+2}^{\infty} b_{-l} 2^{-l}\, 2^{e}$ for $b_{-(p+1)} = 0$

similarly, if $b_{-(p+1)} = 1$ then

$x - rd(x) = \pm\left(\sum_{l=1}^{\infty} b_{-l} 2^{-l} - \sum_{l=1}^{p} b_{-l} 2^{-l} - 2^{-p}\right) 2^{e}$

$2^{-(p+1)} - 2^{-p} = -2^{-(p+1)}$

$= \pm\left(b_{-(p+1)} \cdot 2^{-(p+1)} - 2^{-p} + \sum_{l=p+2}^{\infty} b_{-l} 2^{-l}\right) 2^{e}$, $1 = b_{-(p+1)}$, $-2^{-(p+1)}$

$x - rd(x) = \pm\left(\left(\sum_{l=p+2}^{\infty} b_{-l} 2^{-l}\right) - 2^{-(p+1)}\right) 2^{e}$ for $b_{-(p+1)} = 1$

$\rightarrow$ returning to $b_{-(p+1)} = 0$, $\left|\frac{x - rd(x)}{x}\right|$ is maximized if all b_{-l} from $p+2 \rightarrow \infty = 1$

~~max $|x|$ is 2^{e-1}~~

min floating point x is 2^{e-1} (see class notes)

so $\left|\frac{x - rd(x)}{x}\right| \leq \frac{\left|\sum_{l=p+2}^{\infty} 2^{-l}\right| 2^{e}}{2^{e-1}}$

$\downarrow$

$$= 1 + 2^{-1} + 2^{-2} \dots = \underline{2}$$

$$\left|\frac{x - rd(x)}{x}\right| \leq \left|2^{-p-2}\left(\sum_{l=0}^{\infty} 2^{-l}\right)\right| \cdot 2$$

$$\leq 2^{-p-1} \cdot 2$$

$$\therefore \left|\frac{x - rd(x)}{x}\right| \leq 2^{-p} \quad \text{for } b_{-(p+1)} = 0$$

now for $b_{-(p+1)} = 1$, $|x - rd(x)|$ is maximized if $\underline{b_{-(p+1)} = 1}$ and $\underline{b_{-(p+2)} \to b_{-\infty} = 0}$

Since then are just barely at the threshold to round up and x is thus farthest from from $rd(x)$

$$\therefore \text{ for } b_{-(p+1)} = 1 \quad \left|\frac{x - rd(x)}{x}\right| \leq \frac{\left|\sum_{l=p+2}^{\infty} 0 \cdot 2^{-l} - 2^{-(p+1)}\right| 2^{e}}{2^{e-1}}$$

$$\leq \frac{2^{-(p+1)} 2^{e}}{2^{e-1}}$$

$$\therefore \left|\frac{x - rd(x)}{x}\right| \leq 2^{-p} \quad \text{for } b_{-(p+1)} = 1$$

$\therefore$ have shown $\left|\frac{x - rd(x)}{x}\right| \leq 2^{-p}$ for all cases as required