Howework 2 – Introduction to Computational Science

Claudio Maggioni

March 19, 2020

Question 1

The solutions assume that the sign bit 1 is negative and 0 is positive.

Point a

- 13 is equal to 0101000000001011;
- 42.125 is equal to 0010100010001101;
- 0.8 is equal to 01100110011000011. 0.78 is approximated to 0.7998046875;

Point b

1011010111001101 is $(-1)*(0.25+0.125+0.03125+0.0078125+0.00320625+0.001953125)*2^5$, which is equal to -13.4151.

Point c

x_{max} is 0111111111111111, equal to 255.9375. Since denormalized numbers do not belong to this representation (since the exponent 0000 cannot be used for valid numbers other than 0) x_{min} is 0000000000000001, equal to 0.0078125.

Question 2

Point a

$$\frac{(x+\Delta x)+(y+\Delta y)-(x+y)}{x+y} = \frac{\Delta x}{x+y} + \frac{\Delta y}{x+y} = \frac{x}{x+y}\frac{\Delta x}{x} + \frac{y}{x+y}\frac{\Delta y}{y}$$

Point b

$$\frac{(x+\Delta x)-(y+\Delta y)-(x-y)}{x-y} = \frac{\Delta x}{x-y} - \frac{\Delta y}{x-y} = \frac{x}{x-y}\frac{\Delta x}{x} - \frac{y}{x-y}\frac{\Delta y}{y}$$

Point c

$$\frac{((x+\Delta x)(y+\Delta y))-(xy)}{xy} = \frac{y\Delta x + x\Delta y + \Delta x\Delta y}{xy} \approx \frac{y\Delta x + x\Delta y}{xy} = \frac{\Delta x}{y} + \frac{\Delta y}{x}$$

Point d

$$\frac{((x+\Delta x)/(y+\Delta y))-(x/y)}{x/y} = \frac{\frac{(x+\Delta x)y}{(y+\Delta y)y} - \frac{(y+\Delta y)x}{(y+\Delta y)y}}{x/y} = \frac{y\Delta x - x\Delta y}{x(y+\Delta y)} = \frac{y\Delta x}{x(y+\Delta y)} - \frac{\Delta y}{y+\Delta y} =$$

$$= \left(\frac{x(y+\Delta y)}{y\Delta x}\right)^{-1} - \left(\frac{y+\Delta y}{\Delta y}\right)^{-1} = \left(\frac{x}{\Delta x} - \frac{y\Delta x}{x\Delta y}\right)^{-1} - \left(\frac{y}{\Delta y}+1\right)^{-1} =$$

$$= \left(\frac{x}{\Delta x}\left(1-\frac{\Delta y}{y}\right)\right)^{-1} - \left(\frac{y}{\Delta y}+1\right)^{-1} \approx \left(\frac{x}{\Delta x}\right)^{-1} - \left(\frac{y}{\Delta y}\right)^{-1} = \frac{\Delta x}{x} - \frac{\Delta y}{y}$$

Point e

Division and multiplication may suffer from cancellation.

Exercise 3

Point d

The error at first keeps getting exponentially smaller due to a better approximation of h when computing the derivative (i.e. h is exponentially nearer to 0), but at 10^{-9} this trend almost becomes the opposite due to loss of significant digits when subtracting from e^{x+h} e^x and amplifiying this error by effectively multiplying that with exponentially increasing powers of 10.