Howework 2 – Introduction to Computational Science

Claudio Maggioni

March 20, 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.

Exercise 4

Point a

$$||A||_\infty = \max_{1\le j\le 2} \sum_{j=1}^{2} |a_{i,j}| = \max\{5_{j=1}, 2.5_{j=2}\} = 5$$

$$||A||_1 = \max_{1\le i\le 2} \sum_{i=1}^{2} |a_{i,j}| = \max\{3.5_{i=1}, 4_{i=2}\} = 4$$

$$||A||_F = \left(\sum_{i=1}^{2}\sum_{j=1}^{2}(a_{i,j})^2\right)^{\frac{1}{2}} = (4+9+2.25+1)^{\frac{1}{2}} \approx 4.031$$

Point b

$$||B||_\infty = \max_{1\le j\le 3} \sum_{j=1}^{3} |b_{i,j}| = \max\{12_{j=1}, 6_{j=2}, 3_{j=3}\} = 12$$

$$||B||_1 = \max_{1\le i\le 3} \sum_{i=1}^{3} |b_{i,j}| = \max\{9_{i=1}, 7_{i=2}, 5_{i=3}\} = 9$$

$$||B||_F = \left(\sum_{i=1}^{3}\sum_{j=1}^{3}(b_{i,j})^2\right)^{\frac{1}{2}} = (36+16+4+1+4+9+4+1+0)^{\frac{1}{2}} \approx 8.660$$

Point c

$$C = \begin{bmatrix} 2 & 3 \\ 1.5 & -1 \end{bmatrix} \begin{bmatrix} 2 & 1.5 \\ 3 & -1 \end{bmatrix} = \begin{bmatrix} 4+9 & 3-3 \\ 3-3 & 2.25+1 \end{bmatrix} = \begin{bmatrix} 13 & 0 \\ 0 & 3.25 \end{bmatrix}$$

$$det(C - \lambda i) = det\left(\begin{bmatrix} 13-\lambda & 0 \\ 0 & 3.25-\lambda \end{bmatrix}\right) = (42.25 - 13\lambda - 3.25\lambda + \lambda^2) = \lambda^2 - 16.25\lambda + 42.25$$

$$det(C - \lambda i) = 0 \Leftrightarrow = \lambda^2 - 16.25\lambda + 42.25 = 0 \Leftrightarrow \lambda = 3.25 \vee \lambda = 13$$

$$||A||_2 = \sqrt{\lambda_{max}(AA^T)} = \sqrt{13} \approx 3.606$$

Exercise 5

Point a

$$A^{-1} = \frac{1}{(-2) - 4.5} \begin{bmatrix} -1 & -3 \\ -1.5 & 2 \end{bmatrix} = \begin{bmatrix} 2/13 & 6/13 \\ 3/13 & -4/13 \end{bmatrix}$$

$$||A^{-1}||_\infty = \max_{1 \le j \le 2} \sum_{j=1}^{2} |a_{i,j}| = \max\{\frac{8}{13}.\frac{7}{13}\} = \frac{8}{13}$$

$$||A^{-1}||_1 = \max_{1 \le i \le 2} \sum_{i=1}^{2} |a_{i,j}| = \max\{\frac{5}{13}, \frac{10}{13}\} = \frac{10}{13}$$

$$||A^{-1}||_F = \left(\sum_{i=1}^{2}\sum_{j=1}^{2}(a_{i,j})^2\right)^{\frac{1}{2}} = \left(\frac{65}{13^2}\right)^{\frac{1}{2}} = \sqrt{\frac{5}{13}}$$

$$k_1(A) = \frac{4 * 10}{13} \approx 3.077$$

$$k_\infty(A) = \frac{5 * 8}{13} \approx 3.077$$

$$k_F(A) \approx 8.660 * \sqrt{\frac{5}{13}} \approx 5.371$$

Point b

$$||A^{-1}||_2 = \sigma_{max}(A^{-1}) = \frac{1}{\sigma_{min}(A)} = \frac{1}{\sqrt{\lambda_{min}}} = \frac{1}{\sqrt{3.25}}$$

$$k_2(A) = \frac{\sqrt{13}}{\sqrt{3.25}} = 2$$

Point c

$$k_\infty \begin{bmatrix} -4 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{bmatrix} = \left\|\begin{bmatrix} -4 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{bmatrix}\right\|_\infty \left\|\begin{bmatrix} -1/4 & 0 & 0 \\ 0 & 1/5 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}\right\|_\infty = 5 * \frac{1}{3} = \frac{5}{3}$$

Point d

The condition of that matrix cannot be computed since that matrix has no inverse since it is not full rank ($M_{2,:} = M_{1,:} * -2$).