Seminar 1 (exercises)

(16/10/2019)

$\boxed{1}$

32-bit floating point representation.
(1 bit sign; 8 bits exponent; 23 bits mantissa)

a) Largest positive number: (normalized) $\left(S_n = \frac{r^{k_0} - r^{n+1}}{1-r} \right)$

0|11111111|11111111111111111111111

$s = 0$

$c = 2^0 + 2^1 + 2^2 + 2^3 + 2^4 + 2^5 + 2^6 + 2^7 = \frac{1-2^8}{1-2} = 255$

$f = 2^{-1} + 2^{-2} + \dots + 2^{-23} = \frac{2^{-1} - 2^{-24}}{1 - 2^{-1}} = \frac{2(2^{23}-1)}{2^{24}(2-1)}$

$$\Rightarrow (-1)^0 \cdot 2^{255 - \lfloor \cancel{\frac{255}{2}} \rfloor 127} \cdot \left(1 + \frac{2(2^{23}-1)}{2^{24}(2-1)}\right) = 3'4028235 \cdot 10^{38}$$

b) Smallest number greater than one. $(x \neq 1)$

$\underbrace{(-1)^0 \cdot 2^0 \cdot (1 + 2^{-23})}_{\boxed{1+2^{-23}}} \Rightarrow \begin{cases} s = 0 \\ c = \lfloor 255/2 \rfloor = 127 = 2^7 + 1 \\ f = 2^{-23} \end{cases}$

(epsilon $= 2^{-23}$)

0|10000001|00000000000000000000001

c) Largest value smaller than 1.

$$(-1)^0 \cdot 2^{-1} \cdot (1+f) \Rightarrow (-1)^0 \cdot 2^{-1} \cdot \left(1 + \frac{2(2^{23}-1)}{2^{24}(2-1)}\right) =$$

$$= 0.9999999404$$

$S = 0$ =

$C = -1 = 128 - 127 = 2^7 - 127$

$f = 2^{-1} + 2^{-2} + \dots + 2^{-23} = \frac{2(2^{23}-1)}{2^{24}(2-1)}$

$\Rightarrow 0 | 10000000 | 11111111111111111111111$

d) Smallest positive ~~denormal~~ number.

$0 | 00000001 | 00000000000000000000000$

$$(-1)^0 \cdot 2^{1-127} \cdot 1 = \boxed{2^{-126}}$$

$$\begin{cases} S = 0 \\ C = 1 \\ f = 0 \end{cases}$$

e) Largest possible denormal number.

$(1 - 1/2^{23})$

$0 | 00000000 | 11111111111111111111111$

$$(-1)^0 \cdot 2^{-127} \cdot (2^{-1} + 2^{-2} + \dots + 2^{-23}) = 2^{-127} \cdot \frac{(2^{23}-1)}{\cancel{2^{23}}(2-1)}$$

$$= 2^{-127} - 2^{-150} = 5.87747 \cdot 10^{-39}$$

$$\begin{cases} S = 0 \\ C = 0 \\ f = 2^{-1} + 2^{-2} + \dots + 2^{-23} = (1 - 1/2^{23}) \end{cases}$$

f) Smallest denormal (positive).

0|00000000|00000000000000000000001

$(-1)^0 \cdot 2^{-127} \cdot (2^{-23}) = \boxed{2^{-150}}$

$$\begin{cases} s = 0 \\ c = 0 \\ f = 2^{-23} \end{cases}$$

g) Zero.

0|00000000|00000000000000000000000

$(-1)^0 \cdot 2^{-127} \cdot 1 = 2^{-127} \equiv 0$ //

$$\begin{cases} s = 0 \\ c = 0 \\ f = 0 \end{cases}$$

h) Infinity

0|11111111|00000000000000000000000

$(-1)^0 \cdot 2^{255-127} \cdot 1 = 2^{128} \equiv \infty$ //

$$\begin{cases} s = 0 \\ c = 255 \\ f = 0 \end{cases}$$

i) NAN

$$\begin{cases} s = 0 \\ c = 255 \\ f \neq 0 \end{cases}$$

(if $x > 2^{128} \Rightarrow$ NAN)

I.E.

0|11111111|00000000000000000000001| $= 2^{128} \cdot (1 + 2^{-23}) \geq$ NAN

$\boxed{2}$ $A = \begin{pmatrix} 9 & -6+3i & 3-9i \\ -6-3i & 6 & -2+7i \\ 3+9i & -2-7i & 25 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 0 & 3 \\ 3 & 2 & 2 \\ 2 & 3 & 2 \end{pmatrix}$

a) Cholesky of A

$A = L \cdot L^*$

$L_i \rightarrow$ i-th row of L

$L^j \rightarrow$ j-th row of L

- $L_1 \cdot L^{*1}$

  $l_{11} \cdot \overline{l_{11}} = |l_{11}|^2 = 9 \Rightarrow l_{11} = +\sqrt{9} = 3$

- $L_1 \cdot L^{*2}$

  $3 \cdot \overline{l_{21}} = -6+3i \Rightarrow \overline{l_{21}} = -2+i$

- $L_1 \cdot L^{*3}$

  $3 \cdot \overline{l_{31}} = 3-9i \Rightarrow \overline{l_{31}} = 1-3i$

- $L_2 \cdot L^{*2}$

  $5 + |l_{22}|^2 = 6 \Rightarrow l_{22} = 1$

- $L_2 \cdot L^{*3}$

  $-2+6i-i-3+\overline{l_{32}} = -2+7i \Rightarrow$

  $\Rightarrow \overline{l_{32}} = 3+2i$

- $L_3 \cdot L^{*3}$

  $10+13+|l_{33}|^2 = 25 \Rightarrow l_{33} = \sqrt{2}$

$$L = \begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \qquad L^* = \begin{pmatrix} \overline{l_{11}} & \overline{l_{21}} & \overline{l_{31}} \\ 0 & \overline{l_{22}} & \overline{l_{32}} \\ 0 & 0 & \overline{l_{33}} \end{pmatrix}$$

$$\downarrow$$

$$\begin{pmatrix} 3 & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} 3 & \overline{l_{21}} & \overline{l_{31}} \\ 0 & \overline{l_{22}} & \overline{l_{32}} \\ 0 & 0 & \overline{l_{33}} \end{pmatrix}$$

$$\downarrow$$

$$\begin{pmatrix} 3 & 0 & 0 \\ -2-i & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} 3 & -2+i & \overline{l_{31}} \\ 0 & \overline{l_{22}} & \overline{l_{32}} \\ 0 & 0 & \overline{l_{33}} \end{pmatrix}$$

$$\downarrow$$

$$\begin{pmatrix} 3 & 0 & 0 \\ -2-i & l_{22} & 0 \\ 1+3i & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} 3 & -2+i & 1-3i \\ 0 & \overline{l_{22}} & \overline{l_{32}} \\ 0 & 0 & \overline{l_{33}} \end{pmatrix}$$

$$\downarrow$$

$$\begin{pmatrix} 3 & 0 & 0 \\ -2-i & 1 & 0 \\ 1+3i & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} 3 & -2+i & 1-3i \\ 0 & 1 & \overline{l_{32}} \\ 0 & 0 & \overline{l_{33}} \end{pmatrix}$$

$$\downarrow$$

$$\begin{pmatrix} 3 & 0 & 0 \\ -2-i & 1 & 0 \\ 1+3i & 3-2i & l_{33} \end{pmatrix} \begin{pmatrix} 3 & -2+i & 1-3i \\ 0 & 1 & 3+2i \\ 0 & 0 & \overline{l_{33}} \end{pmatrix}$$

$$\downarrow$$

$$\begin{pmatrix} 3 & 0 & 0 \\ -2-i & 1 & 0 \\ 1+3i & 3-2i & \sqrt{2} \end{pmatrix} \begin{pmatrix} 3 & -2+i & 1-3i \\ 0 & 1 & 3+2i \\ 0 & 0 & \sqrt{2} \end{pmatrix}$$

b) LDL^* of A

$$L = \begin{pmatrix} 1 & 0 & 0 \\ \ell_{21} & 1 & 0 \\ \ell_{31} & \ell_{32} & 1 \end{pmatrix}; \quad L^* = \begin{pmatrix} 1 & \overline{\ell_{21}} & \overline{\ell_{31}} \\ 0 & 1 & \overline{\ell_{32}} \\ 0 & 0 & 1 \end{pmatrix}; \quad D = \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix}$$

$$A = LDL^*; \quad LD = \begin{pmatrix} d_1 & 0 & 0 \\ \ell_{21}d_1 & d_2 & 0 \\ \ell_{31}d_1 & \ell_{32}d_2 & d_3 \end{pmatrix}$$

- $LD_1 \cdot (L^*)^1$

  $d_1 = 9$

- $LD_1 \cdot (L^*)^2$

  $9 \cdot \overline{\ell_{21}} = -6+3i \Rightarrow \overline{\ell_{21}} = -\frac{2}{3} + \frac{i}{3}$

- $LD_1 \cdot (L^*)^3$

  $9 \cdot \overline{\ell_{31}} = 3 - 9i \Rightarrow \overline{\ell_{31}} = \frac{1}{3} - i$

- $LD_2 \cdot (L^*)^2$

  $9 \cdot |\ell_{21}|^2 + d_2 = 6 \Rightarrow d_2 = 6 - 9 \cdot \left(\frac{4}{9} + \frac{1}{9}\right)$

  $\Rightarrow d_2 = 1$

- $LD_2 \cdot (L^*)^3$

  $9 \cdot \left(-\frac{2}{3} - \frac{i}{3}\right)\left(\frac{1}{3} - i\right) + \overline{\ell_{32}} = -2+7i$

  $9\left(-\frac{2}{9} + \frac{2i}{3} - \frac{i}{9} - \frac{1}{3}\right) + \overline{\ell_{32}} = -2+7i$

  $-5 + 5i + \overline{\ell_{32}} = -2+7i$

  $\overline{\ell_{32}} = 3+2i$

- $LD_3 \cdot (L^*)^3$

  $9 \cdot |\ell_{31}|^2 + 1 \cdot |\ell_{32}|^2 + d_3 = 25$

  $10 + 13 + d_3 = 25 \Rightarrow d_3 = 2$

$$\begin{pmatrix} 1 & 0 & 0 \\ \ell_{21} & 1 & 0 \\ \ell_{31} & \ell_{32} & 1 \end{pmatrix} \begin{pmatrix} 1 & \overline{\ell_{21}} & \overline{\ell_{31}} \\ 0 & 1 & \overline{\ell_{32}} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix}$$

$$\downarrow$$

$$\begin{pmatrix} 1 & 0 & 0 \\ -\frac{2}{3}-\frac{i}{3} & 1 & 0 \\ \ell_{31} & \ell_{32} & 1 \end{pmatrix} \begin{pmatrix} 1 & -\frac{2}{3}+\frac{i}{3} & \overline{\ell_{31}} \\ 0 & 1 & \overline{\ell_{32}} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 9 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix}$$

$$\downarrow$$

$$\begin{pmatrix} 1 & 0 & 0 \\ -\frac{2}{3}-\frac{i}{3} & 1 & 0 \\ \frac{1}{3}+i & \ell_{32} & 1 \end{pmatrix} \begin{pmatrix} 1 & -\frac{2}{3}+\frac{i}{3} & \frac{1}{3}-i \\ 0 & 1 & \overline{\ell_{32}} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 9 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & d_3 \end{pmatrix}$$

$$\downarrow$$

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 \\ -\frac{2}{3}-\frac{i}{3} & 1 & 0 \\ \frac{1}{3}+i & 3-2i & 1 \end{pmatrix}}_{L} \underbrace{\begin{pmatrix} 1 & -\frac{2}{3}+\frac{i}{3} & \frac{1}{3}-i \\ 0 & 1 & 3+2i \\ 0 & 0 & 1 \end{pmatrix}}_{L^*} \underbrace{\begin{pmatrix} 9 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}}_{D}$$

c) Doolittle factorization of B

$B = LU$

- $L_1 \cdot U^1$

  $u_{11} = 1$

- $L_1 U^2$

  $u_{12} = 0$

- $L_1 U^3$

  $u_{13} = 3$

- $L_2 \cdot U^1$

  $\ell_{21} = 3$

- $L_3 \cdot U^1$

  $\ell_{31} = 2$

- $L_2 \cdot U^2$

  $u_{22} = 2$

- $L_2 \cdot U^3$

  $9 + u_{13} = 2 \Rightarrow u_{13} = -7$

- $L_3 \cdot U^2$

  $\ell_{32} \cdot 2 = 3 \Rightarrow \ell_{32} = 3/2$

- $L_3 \cdot U^3$

  $6 - \frac{21}{2} + u_{23} = 2 \Rightarrow u_{23} = 13/2$

$$L = \begin{pmatrix} 1 & 0 & 0 \\ \ell_{21} & 1 & 0 \\ \ell_{31} & \ell_{32} & 1 \end{pmatrix} \qquad U = \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{pmatrix}$$

$$\downarrow \qquad\qquad \downarrow$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & \ell_{32} & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 3 \\ 0 & u_{22} & u_{13} \\ 0 & 0 & u_{23} \end{pmatrix}$$

$$\downarrow \qquad\qquad \downarrow$$

$$\boxed{\begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & 3/2 & 1 \end{pmatrix} = L} \qquad \begin{pmatrix} 1 & 0 & 3 \\ 0 & 2 & -7 \\ 0 & 0 & u_{23} \end{pmatrix}$$

$$\downarrow$$

$$\boxed{\begin{pmatrix} 1 & 0 & 3 \\ 0 & 2 & -7 \\ 0 & 0 & 13/2 \end{pmatrix} = U}$$

We check:

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & 3/2 & 1 \end{pmatrix}}_{L} \underbrace{\begin{pmatrix} 1 & 0 & 3 \\ 0 & 2 & -7 \\ 0 & 0 & 13/2 \end{pmatrix}}_{U} = \underbrace{\begin{pmatrix} 1 & 0 & 3 \\ 3 & 2 & 2 \\ 2 & 3 & 2 \end{pmatrix}}_{B} \checkmark$$

d) Crout factorization of B

$B = LU$

- $L_1 \cdot U^1$: $l_{11} = 1$
- $L_2 \cdot U^1$: $l_{21} = 3$
- $L_3 \cdot U^1$: $l_{31} = 2$

- $L_1 \cdot U^2$: $u_{12} = 0$
- $L_1 \cdot U^3$: $u_{13} = 3$

- $L_2 \cdot U^2$: $l_{22} = 2$
- $L_3 \cdot U^2$: $l_{32} = 3$

- $L_2 \cdot U^3$

$$9 + 2u_{23} = 2 \Rightarrow u_{23} = -\frac{7}{2}$$

- $L_3 \cdot U^3$

$$6 - \frac{21}{2} + l_{33} = 2 \Rightarrow l_{33} = \frac{13}{2}$$

$$L = \begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 3 & l_{22} & 0 \\ 2 & l_{32} & l_{33} \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 3 & 2 & 0 \\ 2 & 3 & l_{33} \end{pmatrix} \to \boxed{\begin{pmatrix} 1 & 0 & 0 \\ 3 & 2 & 0 \\ 2 & 3 & \frac{13}{2} \end{pmatrix} = L}$$

$$U = \begin{pmatrix} 1 & u_{12} & u_{13} \\ 0 & 1 & u_{23} \\ 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & u_{23} \\ 0 & 0 & 1 \end{pmatrix} \to \boxed{\begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & -\frac{7}{2} \\ 0 & 0 & 1 \end{pmatrix} = U}$$

- We double check:

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 3 & 2 & 0 \\ 2 & 3 & \frac{13}{2} \end{pmatrix}}_{=L} \underbrace{\begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & -\frac{7}{2} \\ 0 & 0 & 1 \end{pmatrix}}_{=U} = \underbrace{\begin{pmatrix} 1 & 0 & 3 \\ 3 & 2 & 2 \\ 2 & 3 & 2 \end{pmatrix}}_{=B} \checkmark$$

e) LU factorization with partial pivoting of B.

$$PB = LU$$

- $L_1 \leftrightarrow L_2$
- $F_2' = F_2 - \frac{1}{3}F_1 \Rightarrow l_{21} = 1/3$

  $F_3' = F_3 - \frac{2}{3}F_1 \Rightarrow l_{31} = 2/3$

  $\longrightarrow l_{11} = 1$
- $L_3 \leftrightarrow L_2$
- $F_2' = F_2 \rightarrow l_{22} = 1$

  $F_3' = F_3 - \frac{-2/3}{5/3}F_2 \rightarrow l_{32} = -\frac{2}{5}$
- $l_{33} = 1$

We double check:

$$PB = \begin{pmatrix} 0&1&0\\0&0&1\\1&0&0 \end{pmatrix}\begin{pmatrix} 1&0&3\\3&2&2\\2&3&2 \end{pmatrix} = \begin{pmatrix} 3&2&2\\2&3&2\\1&0&3 \end{pmatrix}$$

$$LU = \begin{pmatrix} 1&0&0\\2/3&1&0\\1/3&-2/5&1 \end{pmatrix}\begin{pmatrix} 3&2&2\\0&5/3&2/3\\0&0&13/5 \end{pmatrix} =$$

$$= \begin{pmatrix} 3&2&2\\2&3&2\\1&0&3 \end{pmatrix}$$

$\Rightarrow PB = LU$ //

$$P = \begin{pmatrix} 1&0&0\\0&1&0\\0&0&1 \end{pmatrix}, \quad L = \begin{pmatrix} l_{11}&0&0\\l_{21}&l_{22}&0\\l_{31}&l_{32}&l_{33} \end{pmatrix}, \quad U = \begin{pmatrix} 1&0&3\\3&2&2\\2&3&2 \end{pmatrix}$$

$$\downarrow$$

$$\begin{pmatrix} 0&1&0\\1&0&0\\0&0&1 \end{pmatrix} \quad \begin{pmatrix} 1&0&0\\1/3&l_{22}&0\\2/3&l_{32}&l_{33} \end{pmatrix} \quad \begin{pmatrix} 3&2&2\\1&0&3\\2&3&2 \end{pmatrix}$$

$$\downarrow$$

$$\begin{pmatrix} 3&2&2\\0&-2/3&7/3\\0&5/3&2/3 \end{pmatrix}$$

$$\downarrow$$

$$\boxed{\begin{pmatrix} 0&1&0\\0&0&1\\1&0&0 \end{pmatrix} = P} \quad \begin{pmatrix} 1&0&0\\2/3&l_{22}&0\\1/3&l_{32}&l_{33} \end{pmatrix} \quad \begin{pmatrix} 3&2&2\\0&5/3&2/3\\0&-2/3&7/3 \end{pmatrix}$$

$$\downarrow$$

$$\begin{pmatrix} 1&0&0\\2/3&1&0\\1/3&-2/5&l_{33} \end{pmatrix} \quad \boxed{\begin{pmatrix} 3&2&2\\0&5/3&2/3\\0&0&13/5 \end{pmatrix} = U}$$

$$\downarrow$$

$$\boxed{\begin{pmatrix} 1&0&0\\2/3&1&0\\1/3&-2/5&1 \end{pmatrix} = L}$$