

CPSC 303

— Remarks on why MATLAB has (as in Homework 1,2):

largest positive number $\approx 2^{1024} \approx 1.8 \times 10^{308}$

smallest " " $\approx 2^{-1074} \approx 4.9 \times 10^{-324}$

MATLAB's **realmin** $\approx 2^{-1022}$, 2^{-1023} , 2.2×10^{-308}

— More on divided differences

Main Thm: $f[x_0, \dots, x_n] = \frac{1}{n!} f^{(n)}(\xi)$, $\xi \in (x_0, x_n)$

Pf: --- accidentally, come across interpolation ---

Handout (Note) About Normal vs Subnormal in Double Precision

[Not covered in Ch 2 of [A & G]]

Idea:

"Normal" number $\pm 1.b_1 b_2 \dots b_{52} \times 2^m$
 $\uparrow$
 S $b_1, \dots, b_{52} \in \{0, 1\}$ m integer,
base 2 arithmetic $-1022 \leq m \leq 1023$

Normal Number

$S, b_1, \dots, b_{52}$,

11 bits (binary digits)

$\underbrace{\hspace{1cm}}_{1 + 52}$

$\underbrace{\hspace{1cm}}_{11}$

for m

= 64 bits

Intuitively think of

base 10
scientific notation 6.27195×10^{2020}

$$m = -1022, \dots, -1, 0, 1, \dots, 1023$$

(11 bits is
2048
possibilities)

2046 values

2 leftover

2 special values of
the 11-bit "m":

① Inf, -Inf, NaN, ...

② Revert to "Subnormal form"

Form:

$$\pm 0.b_1 b_2 \dots b_{52} \times 2^{-1022}$$

52 bits fixed

e.g. $\pm 0.00010101110\dots0 \times 2^{-1022}$

drop
down to
small power of 2
 < 52 bits

base 10: $0.d_1 d_2 d_3 d_4 d_5$

$\left\{ \begin{array}{l} 0.12345 \\ 0.27193 \\ 0.00021 \end{array} \right.$

Variant $\pm 0.0000b_1 \dots b_{48} \times 2^{-1022 - \underbrace{b_{49} \dots b_{52}}}$

Largest number in double precision:

$$+ 1.b_1 \dots b_{52} \times 2^{1023}$$

↓

$$1.11111 \dots 1$$
$$\left(2 - \frac{1}{2^{52}}\right)$$

Smallest pos number

$$+ \underbrace{0.000 \dots 01}_{2^{-52}} \times 2^{-1022}$$
$$2^{-52} \cdot 2^{-1022} = 2^{-1074}$$

HW 1, 2

$$X_{n+2} = (1+r)X_{n+1} - rX_n$$

general sol is

$$C_1(1^n) + C_2(r^n)$$

$$r = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots$$

$$\begin{bmatrix} X_{n+2} \\ X_{n+1} \end{bmatrix} = \begin{bmatrix} 1+r & -r \\ 1 & 0 \end{bmatrix} \begin{bmatrix} X_{n+1} \\ X_n \end{bmatrix}$$

$$x_0 = 1, x_1 = r$$

$$\begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix} = \begin{bmatrix} 1+r & -r \\ 1 & 0 \end{bmatrix}^n \begin{bmatrix} r \\ 1 \end{bmatrix}$$

$$r = \frac{1}{\text{power of 2}} \quad \text{😊}, \quad r = \frac{1}{3}, \frac{1}{5}, \frac{1-\sqrt{5}}{2}, \frac{1}{\sqrt{7}}$$

$$x_0 = 1, x_1 = r \quad \left\{ \begin{array}{l} \text{theoretically} \\ \text{exact arithmetic} \end{array} \right\} \quad x_n = r^n$$

$$\text{numerical} \quad : \quad x_n = C_1 + C_2 r^n$$

$$\swarrow \quad C_2 = 1$$

C_1 small, but
not 0

Another 3-term:

$$\left(r - \frac{1}{2}\right)\left(r - \frac{1}{3}\right) = 0$$

general solution

$$x_n = C_1 \left(\frac{1}{2}\right)^n + C_2 \left(\frac{1}{3}\right)^n$$

recur

$$r^2 - \frac{5}{6}r + \frac{1}{6} = 0 \quad x_{n+2} - \frac{5}{6}x_{n+1} + \frac{1}{6}x_n = 0$$

What about

$$X_{n+2} - 2X_{n+1} + X_n = 0 \quad ??$$

our method:

$$r^2 - 2r + 1 = 0$$

$$\boxed{(r-1)^2 = 0}, \quad r = 1, 1$$

general solution: $C_1 1^n + C_2 1^n \quad ??$

$$X_n = \begin{cases} (C_1 + C_2) \cdot 1 \\ C \cdot 1 \end{cases}$$

linear polynomial

Recall:

$$X_n = C_1 + C_2 n$$

$$\begin{array}{ccccccc} & & 3 & 3 & 3 & 3 & 3 & \dots \\ \text{1st Diff} & & 0 & 0 & 0 & 0 & 0 & \dots \end{array}$$

$$\begin{array}{ccccccc} & & 4 & 8 & 12 & 16 & 20 & \dots \\ \text{1st Diff} & & 4 & 4 & 4 & 4 & 4 & \dots \\ \text{2nd Diff} & & 0 & 0 & 0 & 0 & 0 & \dots \end{array}$$

And $(r-1)^3 = 0$

$$X_{n+3} - 3X_{n+2} + 3X_{n+1} - X_n = 0$$

Gen sol. $C_1 + C_2 n + C_3 n^2$

Differences, special case Divide Diff

Added after class on Whiteboard:

$$X_{n+1} - X_n = 0 \quad \text{or} \quad X_{n+1} = X_n \quad \text{gen sol. } X_n = C_1$$

$$X_{n+2} - 2X_{n+1} + X_n = 0$$

$$\text{gen sol } X_n = C_1 + C_2 n$$

$$= C_1(1^n) + C_2(1^n n)$$

(Div Diff:

$$\dots, X_1, X_2, X_3, X_4, \dots$$

1st Diff

$$X_2 - X_1, X_3 - X_2, X_4 - X_3$$

2nd Diff

$$\downarrow \downarrow \quad \searrow \downarrow$$

$$\dots, X_3 - 2X_2 + X_1, X_4 - 2X_3 + X_2, \dots$$

$$\swarrow \quad \downarrow \quad \swarrow \quad \swarrow$$

$$X_{n+2} - 2X_{n+1} + X_n$$