LECTURE 08 4/2/2026

Scale $A$ such that $\left\|\frac{A}{2^s}\right\| < 1$

Taylor series, Padé

$T_m$ $m$ the degree of approximation

$$T_m\left(\frac{A}{2^s}\right)^{2^s} - \exp(A) = E \quad \text{forward error}$$

$$T_{m,1}\left(\frac{A}{2^s}\right)^{2^s} = \varphi_1(A + \Delta A) = \exp(A + \Delta A) \quad \Delta A \quad \text{backward error}$$

$$\log\left(e^{-x}T_m(x)\right) = h_{m+1}(x)$$ is an error for the approximation $T_m(x) \approx e^x$

$$e^{-x}T_m(x) = e^{h_{m+1}(x)}$$

$$T_m(x) = e^{x + h_{m+1}(x)}$$

$$T_m\left(\frac{A}{2^s}\right)^{2^s} = \exp\left(\frac{A}{2^s} + h_{m+1}\left(\frac{A}{2^s}\right)\right)^{2^s} =$$

$$\exp\Big(A + \underbrace{2^s h_{m+1}\left(\frac{A}{2^s}\right)}_{\Delta A}\Big)$$

$$h_{m+1}(x) = \log\left(e^{-x}\left(e^x - \frac{x^{m+1}}{(m+1)!} - \frac{x^{m+2}}{(m+2)!} - \dots\right)\right) =$$

$$\log\left(1 + \mathcal{O}\left(x^{m+1}\right)\right) =$$

$$\log\left(1 + b_{m+1}x^{m+1} + b_{m+2}x^{m+2} + \dots\right) =$$

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots$$

$$= c_{m+1}x^{m+1} + c_{m+2}x^{m+2} + \dots$$

we can compute the coefficients $c_{m+1}, c_{m+2}, \dots$
by a software in high precision arithmetic

$$h_{m+1}(x) = \sum_{i=m+1}^{\infty \to M \gg m} c_i x^i \qquad \tilde{h}_{m+1}(x) = \sum_{i=m+1}^{\infty} |c_i| x^i$$

We want

relative backward error
$$\frac{\|\Delta A\|}{\|A\|} = \frac{\| 2^s h_{m+1}(2^{-s}A) \|}{\|A\|} =$$

$$\frac{\| h_{m+1}(2^{-s}A) \|}{\| 2^{-s} A \|} \leq \frac{\tilde{h}_{m+1}(2^{-s}\|A\|)}{2^{-s}\|A\|} \leq \text{tolerance}$$

Find $\vartheta$ s.t. $\dfrac{\tilde{h}_{m+1}(\vartheta)}{\vartheta} = \text{tol}$ $\qquad \exists!\ \vartheta_m$

$$\frac{\tilde{h}_{m+1}(\vartheta_m)}{\vartheta_m} = \text{tolerance}$$

if $\quad 2^{-s}\|A\|_1 \leq \vartheta_m \Rightarrow \dfrac{\|\Delta A\|_1}{\|A\|_1} \leq \text{tolerance}$

$$\text{tolerance} = \begin{cases} \text{double precision} & \varepsilon/2 \ \text{unit round off} \\ \text{single precision} & \\ \text{half precision} & (\text{GPU CUDA}) \end{cases}$$

EXAMPLE

$\vartheta_{53} = 9.34 \qquad\qquad \vartheta_{35} = 4.73$

$\vartheta_{54} = 9.60 \qquad\qquad \vartheta_{36} = 4.97$

$$\|A\|_1 = 19$$

$s=1 \quad \frac{19}{2^1} = 9.5 \qquad s=2 \quad \frac{19}{2^2} = 4.75$

$T_{54}$ costs 53 matrix-matrix $\qquad T_{36}$ costs 35 m.-m.

1 squaring $\qquad$ 2 squaring

54 matrix-matrix $\qquad$ 37 m.-m.

With m and s, $T_m\left(\frac{A}{2^s}\right)^{2^s} \approx \exp(A)$

DIRECT METHOD

EARLY TERMINATION

if $\dfrac{\left(\frac{\|A\|_1}{2^s}\right)^{k+1}}{(k+1)!} \leq \left\| \sum_{i=0}^{k} \frac{\left(\frac{A}{2^s}\right)^i}{i!} \right\| \cdot$ tolerance, $k < m$

STOP

BACKWARD ERROR ANALYSIS

It was performed for Padé approximations and other approximations to the exp function

PRECONDITIONING

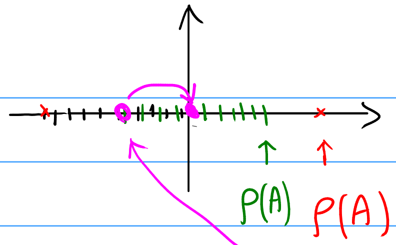
$T_m\left(\frac{A}{2^s}\right) \not\approx \exp\left(\frac{A}{2^s}\right)$

How to immediately reduce $\|A\|$?

$A \to A - \mu I : \|A - \mu I\| < \|A\|$

$e^{\mu} T_m(A - \mu I) \approx \exp(A)$

$A \approx \partial_{xx}$




$\|A\|_2 = \rho(A)$

$A - \mu I$

$A \in \mathbb{C}^{m \times m}$ $\qquad \dfrac{\text{trace}(A)}{m} = \dfrac{\sum \text{ of eigenvalues}}{m} =$ "average eigenvalue"

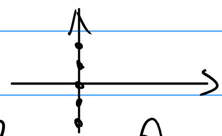
$$\underbrace{A - \overbrace{\frac{\text{trace}(A)}{m}}^{\mu} I}_{B}$$

$$A \to A - \frac{\text{trace}(A)}{m} I \xrightarrow{BEA} (m, s) \to T_m\left(\frac{B}{2^s}\right)^{2^s} e^{\mu}$$

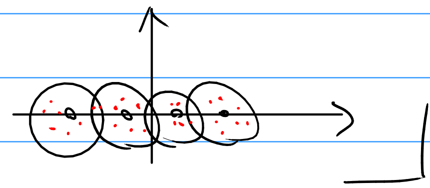
BEA APPLIED TO PADE IS EXACTLY WHAT DONE IN EXPM IN MATLAB (WHEN A IS NOT HERMITIAN).

INTERPOLATION AS A TOOL FOR THE APPROXIMATION OF $\exp(A)$

$L_{m,c} : [-c, c] \to \mathbb{R}$ which interpolates $e^x$ in $[-c, c]$ $\qquad c \in \mathbb{R}$, $c \in i\mathbb{R}$



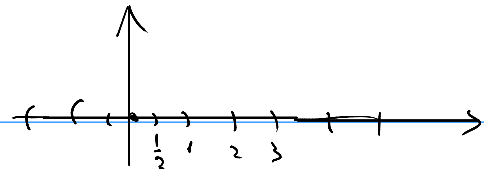
Hint: Gershgorin disks $\qquad A \approx \partial_x \qquad A \approx i\partial_{xx}$



$\vartheta_{m,c}$ : $\dfrac{\tilde{h}(\vartheta_{m,c})}{\vartheta_{m,c}} = \text{tolerance}$ $\qquad$ for given $m, c$

$$\vartheta_m := \max_{0 \le c \le c^*} \{\vartheta_{m,c}\}$$

$m = 20$




$c = 0 \qquad \vartheta_{m,0}$ for $T_m$

$c = \frac{1}{2} \qquad \vartheta_{m,\frac{1}{2}}$

$c = 1 \qquad \vartheta_{m,1}$

$c = 2 \qquad \vartheta_{m,2}$

$\vdots$

$c = c^* \qquad$ NO SOLUTION

$\vartheta_m$ obtained by $L_{m,c}$ is not smaller than $\vartheta_m$ obtained by Taylor

Newton form for $L_{m,c}$

$$d_0 + d_1 (A - \xi_0 I) + d_2 (A - \xi_1 I)(A - \xi_0 I) +$$

+ EARLY TERMINATION

Leja points are good for interpolation (like Chebyshev) for the Newton form (not like Chebyshev) *

* Chebyshev points for degree K are different from the first K of m Chebyshev points !

BACKWARD ERROR ANALYSIS FOR $\varphi_1(A)$

$$\varphi_1(x) \approx T_{m,1}(x) = \frac{e^{x+\Delta x} - 1}{x + \Delta x} = \varphi_1(x + \Delta x)$$

QUASI - BEA

$$x\, T_{m,1}(x) = e^{x+\Delta x} - 1$$

Shifting for $\varphi_1$ is difficult!

$$\varphi_1(A-\mu I) \stackrel{?}{\Rightarrow} \varphi_1(A)$$

$$e^{\mu} e^{A-\mu I} = e^{A}$$

WHY TO COMPUTE $\exp(A)$ or $\varphi_1(A)$ ?

$$u_{m+1} = \underbrace{\exp(\tau A) u_m} + \underbrace{\tau \varphi_1(\tau A) g(t_m, u_m)}$$

IMEX-1

$$u_{m+1} = u_m + \tau A u_{m+1} + \tau g(t_m, u_m)$$

$$\underline{(I - \tau A)} u_{m+1} = \underline{u_m + \tau g(t_m, u_m)}$$

SOLVE A LINEAR SYSTEM!

$$u_{m+1} = (I - \tau A)^{-1} (u_m + \tau g(t_m, u_m))$$

$\mathrm{inv}(I - \tau A)$ in MATLAB ?

How to compute actions of $e^A$ $(e^A v)$

- $e^A = V e^{\Lambda} V^{-1}$

$$e^A v = \underbrace{V \underbrace{e^{\Lambda} \underbrace{V^{-1} v}_{mv}}_{vv}}_{mv}$$

$AV = V\Lambda$ NOT MATRIX FREE

$A(Av)$ much smaller than $A^2 v$

- Taylor : $e^A v = v + Av + \frac{A(Av)}{2} + \frac{A(A(Av))}{6} + \dots$

MATRIX FREE : THE MATRIX ITSELF IS NOT NEEDED

$$F : \mathbb{R}^n \to \mathbb{R}^n$$

$$J_F(x)\, v \approx \frac{F(x+\varepsilon v) - F(x)}{\varepsilon}$$

SPOILER: SOME EXPONENTIAL METHODS REQUIRE
$$\exp\left(J_F(x)\right) v$$

MATRIX FREE ALSO FOR LINEAR SYSTEMS:

GAUSS ELIMINATION IS NOT MATRIX FREE

CONJUGATE GRADIENT IS MATRIX FREE $(Av)$

- PADÉ $$\underbrace{\left(I - \frac{A}{2}\right)^{-1} \underbrace{\left(I + \frac{A}{2}\right) v}}_{\text{solution of a linear system}}$$

$$\underbrace{\left(I - a_1 A + a_2 A^2\right)^{-1}}_{\text{LESS GOOD}} \underbrace{\left(I + a_1 A + a_2 A^2\right) v}_{\text{GOOD}}$$

EXAMPLE: $A$ LARGE AND SPARSE

$Av$ is easy

$A^2$ is difficult!

MISSING: $\varphi_1(A)\,v$

IF $\|A\|$ is large

$$e^{A} v = e^{\frac{A}{s}} e^{\frac{A}{s}} \cdots \underbrace{e^{\frac{A}{s}} v}$$ is not the scaling and squaring

$\varphi_1(A)\,v$ if we can only compute $\varphi_1\left(\frac{A}{s}\right) v$