Note Title — 2024-04-15

Esponenziale di matrici

$$\exp(A) \approx D(A)^{-1} N(A)$$

$\frac{N(x)}{D(x)}$ = approssimante di Padé di $\exp(x)$

$$D(A)^{-1} N(A) = \exp(A+H) \qquad \frac{\|H\|}{\|A\|} \leq u$$

prendendo un'approssimante di grado $(13,13)$ e $\forall A$ $\|A\| \leq 5.4$

$\Rightarrow$ l'algoritmo è stabile

Questo stima l'errore analitico, dobbiamo preoccuparci anche dell'errore numerico, in particolare quello dovuto al condizionamento nell'inversione di $D(A)$.

Vediamo cosa succede numericamente con approssimanti di grado minore, ad es. 2 oppure 4

Guardiamo per semplicità anziché i valori singolari di $D(A)$ i suoi autovalori

$$D(A) = D\left(V \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} V^{-1}\right) = V \begin{bmatrix} D(\lambda_1) & & \\ & \ddots & \\ & & D(\lambda_n) \end{bmatrix} V^{-1}$$

$$\|A\| \leq 5.4 \Rightarrow |\lambda_i| \leq 5.4$$

$\Rightarrow$ vogliamo vedere $\max\limits_{|\lambda| \leq 5.4} |D(\lambda)|$ e $\min\limits_{|\lambda| \leq 5.4} |D(\lambda)|$

Quando il grado dell'approssimante $\to \infty$, $N(x) \to \exp(x/2)$

$$D(x) \to \exp(-x/2) \qquad D(x) = N(-x)$$

$$\max_{|\lambda| \leq 5.4} |D(\lambda)| \approx \max_{|\lambda| \leq 5.4} \left|\exp\left(-\frac{\lambda}{2}\right)\right| = \exp\left(\max_{|\lambda| \leq 5.4} \mathrm{Re}\left(-\frac{\lambda}{2}\right)\right)$$

$$= \exp\left(\frac{5.4}{2}\right)$$

$$\min_{|A|\le 5.4} |D(A)| \approx \min_{|A|\le 5.4} \left|\exp\left(-\frac{A}{2}\right)\right| = \exp\left(-\frac{5.4}{2}\right)$$

$\Rightarrow$ gli autovalori (ma non per forza i valori singolari) di $D(A)$ non possono essere più di tanto grandi/piccoli

Non c'è un risultato forte di stabilità, ma nella pratica $D(A)^{-1}N(A)$ approssima bene $\exp(A)$ per $\|A\| \le 5.4$

Cosa facciamo per matrici di norma più grande di 5.4?

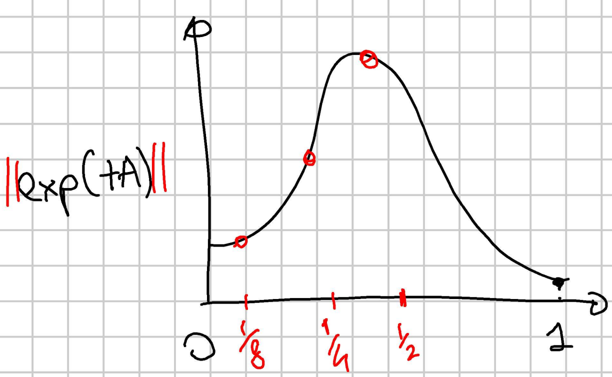
<u>Idea</u>: "scaling and squaring": usiamo l'identità

$$\exp(KA) = \exp(A)^K \qquad \forall K$$

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

Scelgo $s \in \mathbb{N}$ in modo che $\left\|\frac{1}{2^s}A\right\| \le 5.4$

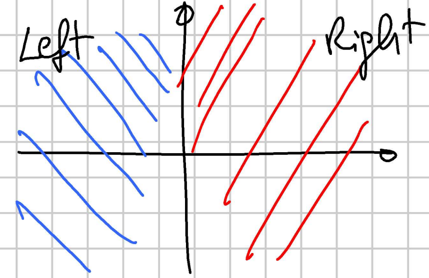
Allora $$\exp(A) \approx \Big(\underbrace{D\left(\tfrac{1}{2^s}A\right)^{-1} N\left(\tfrac{1}{2^s}A\right)}_{\text{approssimante di Padé nella regione in cui è stabile}}\Big)^{\underbrace{2^s}_{s \text{ quadrature successive}}}.$$



Anche con questo algoritmo c'è il rischio di incontrare cancellazione in presenza di "gobbe" nel grafico.

Segno di matrici

$$\text{sign}(x) = \begin{cases} 1 & Re(x) > 0 \\ -1 & Re(x) < 0 \\ \text{indef.} & Re(x) = 0 \end{cases}$$



Se abbiamo A con forma di Jordan $A = VJV^{-1}$, e supponiamo $J = \begin{bmatrix} J_1 & \\ & J_2 \end{bmatrix}$ con $\Lambda(J_1) \subseteq LHP$ (left half-plane), $\Lambda(J_2) \subseteq RHP$ (right " ")

$$\text{sign}(A) = V \left[\begin{array}{c|c} -I & \\ \hline & I \end{array}\right] V^{-1}$$

$$V = [V_1 \ \ V_2]$$

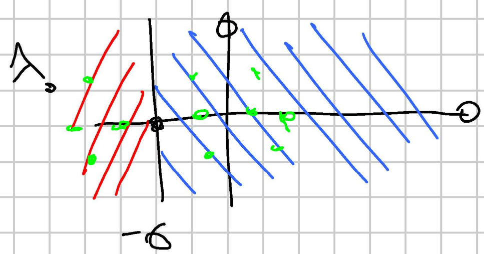
sign(A) ha sempre i soli autovalori ± 1

<u>Oss</u>: $\text{sign}(A) + I = V \left[\begin{array}{c|c} 0 & \\ \hline & 2I \end{array}\right] V^{-1} \Rightarrow \begin{array}{l} Im(\text{sign}(A)+I) = Im\, V_2 \\ Ker(\text{sign}(A)+I) = Im\, V_1 \end{array}$

$Im\, V_1$, $Im\, V_2$ sono i sottosp. invarianti corrispondenti agli autoval. nel LHP e RHP rispettivamente.

<u>Applicazione</u>: in alcuni problemi di fisica, si vogliono calcolare autovalori più a sinistra di alcune matrici (stati elettronici con energia minore) e i proiettori sui sottosp. inv. associati

$\Rightarrow$ data la matrice A, calcolo $\text{sign}(A + \sigma I)$, $\sigma \in \mathbb{R}$



che mi dà i sottosp. invarianti di $A+\sigma I$ associati a LHP e RHP $\Leftrightarrow$ sottosp. inv. di A associati a $\{\lambda : Re(\lambda+\sigma) < 0\}$ e $\{\lambda : Re(\lambda + \sigma) > 0\}$

Più in generale, se calcolo $\text{sign}(\alpha A + \beta I)$ $\alpha, \beta \in \mathbb{C}$ riesco a trovare i sottosp. inv. di A associati ai due semipiani $\mathcal{P}_1 = \{\lambda : \text{Re}(\alpha\lambda + \beta) < 0\}$ e $\mathcal{P}_2 = \{\lambda : \text{Re}(\alpha\lambda + \beta) > 0\}$ e a trovare una matrice Q tale che

$$Q^* A Q = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \qquad \begin{array}{l} \Lambda(A_{11}) \subseteq \mathcal{P}_1 \\ \Lambda(A_{22}) \subseteq \mathcal{P}_2 \end{array}$$

Facendo questo più volte ricorsivamente, si ottiene un algoritmo "divide et impera" per calcolare la forma di Schur di A.

Algoritmo: Schur-Parlett (specializzazione dell'algoritmo a questa funzione)

1) Calcolo $A = UTU^*$ forma di Schur
2) Riordino, per ottenere un'altra forma di Schur della forma

$$A = \hat{U}\hat{T}\hat{U}^*, \quad \hat{T} = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix}$$

$$\Lambda(T_{11}) \subseteq LHP \qquad \Lambda(T_{22}) \subseteq RHP$$

$$f(A) = f(\hat{U}\hat{T}\hat{U}^*) = \hat{U} f(\hat{T}) \hat{U}^* = \hat{U} f\left(\begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix}\right) \hat{U}^*$$

$$= \hat{U} \begin{bmatrix} f(T_{11}) & Z \\ 0 & f(T_{22}) \end{bmatrix} \hat{U}^* = \hat{U} \begin{bmatrix} -I & Z \\ 0 & I \end{bmatrix} \hat{U}^*$$

Per calcolare il blocco off-diagonal Z, uso il fatto che

$$\hat{T} f(\hat{T}) = f(\hat{T}) \hat{T}$$

$$\begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix} \begin{bmatrix} -I & Z \\ 0 & I \end{bmatrix} = \begin{bmatrix} -I & Z \\ 0 & I \end{bmatrix} \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix}$$

$(1,2) \Rightarrow \quad T_{11} Z + T_{12} = -T_{12} + Z T_{22} \quad \Leftrightarrow T_{11} Z - Z T_{22} = -2 T_{12}$

È un'equazione di Sylvester per Z.

$\Lambda(T_{11}) \cap \Lambda(T_{22}) = \emptyset \Rightarrow$ è risolubile.

Condizionamento di $\mathrm{sign}(A)$:

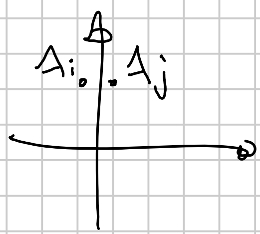
La derivata di Fréchet di $f(A)$ è una matrice $n^2 \times n^2$ che ha autovalori

$$f[\lambda_i, \lambda_j] = \begin{cases} \dfrac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j} & \\ f'(\lambda_i) & i = j \end{cases} \qquad i,j = 1, \dots n$$

Nel caso della funzione segno, $f'(\lambda_i) = 0$ (costante a tratti)

$$f(\lambda_i) - f(\lambda_j) = \begin{cases} 0 & \lambda_i, \lambda_j \text{ nello stesso semipiano} \\ \pm 2 & \lambda_i, \lambda_j \text{ da parti opposte dell'asse immag.} \end{cases}$$

$\Rightarrow f[\lambda_i, \lambda_j]$ grande per



In generale avevamo dimostrato la stima

$$K_{abs, \mathrm{sign}, A} \leq \kappa^2(V) \max |f[\lambda_i, \lambda_j]|$$

$\boxed{\rightarrow \text{viene da } (V \otimes V)(\text{ diagonale })(V^{-1} \otimes V^{-1})}$

<u>Teo</u> ([Byers, Mehrmann, He]) Supponiamo di avere M e una perturbazione E di norma $\|E\| = \varepsilon$ suff. piccola.

Allora,

1) $$\frac{\|\mathrm{sign}(M+E) - \mathrm{sign}(M)\|}{\|\mathrm{sign}(M)\|} = O\left(\frac{\varepsilon}{\delta^2}\right)$$

dove $\delta = \text{sep}(T_{11}, T_{22})$

$M = U \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix} U^*$ $U = [U_1 \ U_2]$

$\Lambda(T_{11}) \subseteq LHP \quad \Lambda(T_{22}) \subseteq RHP$

2) Se questo segno viene usato per calcolare sottospazi invarianti, allora il condizionamento è migliore:
il sottosp. inv. stabile (cioè associato all'LHP) di M
è $\text{Im}\, U \begin{bmatrix} I \\ 0 \end{bmatrix} = U_1$, mentre quello di $M+E$ è
$\text{Im}\, U \begin{bmatrix} I \\ X \end{bmatrix}$ dove X è una matrice t.c. $\|X\| = O\left(\frac{\varepsilon}{\delta}\right)$

<u>Idea della dim</u>: avevamo visto una dimostrazione sul condizionamento dei sottosp. invarianti: data una matrice $M = U \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix} U^*$, con sottosp. invariante $\text{Im}\, U_1 = \text{Im}\, U \begin{bmatrix} I \\ 0 \end{bmatrix}$
una sua perturb. E ha un sottosp. invariante $\text{Im}\, U \begin{bmatrix} I \\ X \end{bmatrix}$,
dove X è la matrice che risolve una certa equazione di matrici proveniente da

$$\begin{bmatrix} I & 0 \\ -X & I \end{bmatrix} \begin{bmatrix} T_{11}+E_{11} & T_{12}+E_{12} \\ E_{21} & T_{22}+E_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ X & I \end{bmatrix} = \begin{bmatrix} * & * \\ 0 & * \end{bmatrix}.$$

e in part. $\|X\| = O\left(\frac{\varepsilon}{\delta}\right)$

Per completare la dimostrazione, usiamo la decomposizione per calcolare $f(M+E)$.
L'altro $\frac{1}{\delta}$ al denominatore deriva dal fatto che il blocco fuori dalla diagonale (quello che avevamo chiamato Z in Schur-Parlett, tale che $f\left(\begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix}\right) = \begin{bmatrix} -I & Z \\ 0 & I \end{bmatrix}$)
ha norma grande quando $\text{sep}(T_{11}, T_{22})$ è piccolo.

$\|Z\| = O\left(\frac{1}{\delta}\right)$. Però perturbazioni di Z non hanno effetto sul sottosp. invariante.

La prossima volta: un algoritmo iterativo per il segno:

$$X_0 = A$$

$$X_{k+1} = \frac{1}{2}\left(X_k + X_k^{-1}\right) \qquad \lim_{k\to\infty} X_k = \operatorname{sign}(A)$$