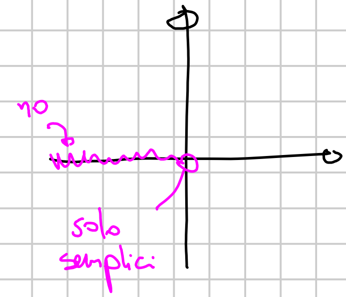
Note Title 2020-04-09

$f(A) = A^{1/2}$ (principal square root)

(autovalori nel semipiano destro)

$f(A)$ ben definita a meno che A abbia:

- autovalori reali negativi
- blocchi di Jordan in 0



Derivata di $g(Y) = Y^2$: è $L_{g,Y}(E) = EY + YE$

Derivata della funzione inversa $f(X) = X^{1/2}$: operatore inverso di $L_{f,X^{1/2}}(E) = EX^{1/2} + X^{1/2}E$

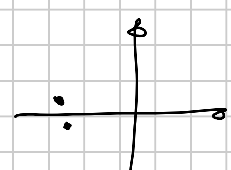
$$\hat{L}_{g,Y} = Y^T \otimes I + I \otimes Y$$

$$\hat{L}_{f,X} = \left(X^{\frac{1}{2}T} \otimes I + I \otimes X^{\frac{1}{2}}\right)^{-1}$$ (corrisponde al caso scalare $\left(2x^{\frac{1}{2}}\right)^{-1}$)

Autovalori: $\dfrac{1}{\lambda_i^{1/2} + \lambda_j^{1/2}}$ $\left(f[\lambda_i, \lambda_j] = \dfrac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j} = \dfrac{\lambda_i^{1/2} - \lambda_j^{1/2}}{\lambda_i - \lambda_j} = \dfrac{\lambda_i^{1/2} - \lambda_j^{1/2}}{(\lambda_i^{1/2} - \lambda_j^{1/2})(\lambda_i^{1/2} + \lambda_j^{1/2})} = \dfrac{1}{\lambda_i^{1/2} + \lambda_j^{1/2}} \right)$

Autovalori grandi $\Leftrightarrow$ mal condizionamento se:

- c'è un autovalore di X con $|\lambda_i|$ piccolo
- due autovalori vicini da lati opposti di $\mathbb{R}_-$



Metodi di calcolo:

Schur method

1. Si riduce A in forma triangolare $A = QTQ^*$

2. Si calcola la diagonale di $S = f(T)$ tramite $s_{ii} = f(t_{ii})$

$$\begin{bmatrix} s_{11} & s_{12} & s_{13} \\ 0 & s_{22} & s_{23} \\ 0 & 0 & s_{33} \end{bmatrix} = f\left(\begin{bmatrix} t_{11} & t_{12} & t_{13} \\ 0 & t_{22} & t_{23} \\ 0 & 0 & t_{33} \end{bmatrix}\right)$$

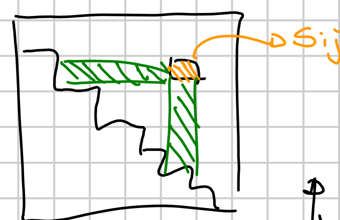
3. Calcola elementi off-diagonal di S tramite una ricorrenza ottenuta da $ST = TS$. Per calcolare s_{ij}, compare $t_{ii} - t_{jj}$ al denominatore

4. $f(A) = Q f(T) Q^* = Q S Q^*$

Per $f(x) = x^{1/2}$, possibile miglioramento al passo 3: posso ricavare un'altra ricorrenza da $S^2 = T$

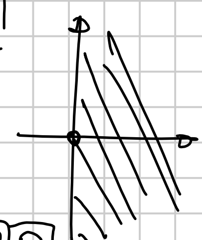
Elemento (i,j) di $S^2 = T$: $s_{ii}\boxed{s_{ij}} + \underbrace{s_{i,i+1}s_{i+1,j} + \dots s_{i,j-1}s_{j-1,j}}_{\text{già calcolati se procedo una diag. per volta}} + \boxed{s_{ij}}s_{jj} = t_{ij}$

$$s_{ij} = \frac{t_{ij} - \sum_{k=i+1}^{j-1} s_{ik}s_{kj}}{s_{ii} + s_{jj}}$$



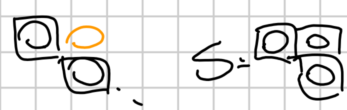
$s_{ii} + s_{jj}$ è sempre $\neq 0$:

$s_{ii} = t_{ii}^{1/2}$ è preso con parte reale positiva



(e ha al più un autovalore 0)

(Potrei avere autoval. 0, ma "semisemplici", $T =$, $S =$)



È più o meno quello che fa Matlab.

(Matlab in realtà fa "divide-and-conquer": $T = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix}$, calcola $f(T_{11})$, $f(T_{22})$ ricorsivamente, poi risolve un'eq. di Sylvester per S_{12})

In più possiamo dire qualcosa sulla stabilità.

$$s_{ij} = \frac{t_{ij} - \sum_{k=i+1}^{j-1} s_{ik} s_{kj}}{s_{ii} + s_{jj}}$$

$$\hat{s}_{ij} = \frac{t_{ij} - \sum \hat{s}_{ik}\hat{s}_{kj}}{\hat{s}_{ii} + \hat{s}_{jj}} + e_{ij}, \quad \text{con}$$

$$|e_{ij}| \leq \frac{|\hat{t}_{ij}| + \sum |\hat{s}_{ik}| \cdot |\hat{s}_{kj}|}{|\hat{s}_{ii} + \hat{s}_{jj}|} \cdot u$$

Si riesce a scrivere l'errore come

$$|\hat{t}_{ij} - \hat{s}_{ii}\hat{s}_{ij} - \hat{s}_{ij}\hat{s}_{jj} - \sum \hat{s}_{ik}\hat{s}_{kj}| = f_{ij},$$

con $|f_{ij}| \leq \left(\sum_{k=i}^{j} |\hat{s}_{ik}| \cdot |\hat{s}_{kj}|\right) \cdot u$

$$|T - \hat{S}^2| = |F| \leq u \cdot |\hat{S}|^2 \quad (|\cdot| \text{ elementwise})$$

Passando a norme

$$\|T - \hat{S}^2\|_F \leq u \cdot \|\hat{S}\|_F^2$$

$$\|A - \hat{X}^2\|_F \leq u \cdot \|\hat{X}\|_F^2, \text{ dove } \hat{X} = fl(A^{1/2})$$

In alternativa, abbiamo metodi iterativi che non richiedono di calcolare forme di Schur.

Metodo di Newton scalare per $\sqrt{x}$: $f(x) = x^2 - a \quad f'(x) = 2x$

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} = x_k - \underbrace{\frac{x_k^2 - a}{2x_k}} = \frac{1}{2}\left(x_k + \frac{a}{x_k}\right)$$

Due modi diversi di estenderlo a matrici:

(1) Metodo di Newton multivariato su $\mathbb{R}^{n^2}$ $(\mathbb{C}^{n^2})$:

$$f(X) = X^2 - A \qquad L_{f,X}[E] = XE + EX \qquad \hat{L}_{f,X} = X^T \otimes I + I \otimes X$$

$$\operatorname{vec} X_{k+1} = \operatorname{vec} X_k - \underbrace{\left(X_k^T \otimes I + I \otimes X_k\right)^{-1} \cdot \operatorname{vec}\left(X_k^2 - A\right)}_{\operatorname{vec} E_k}$$

Formulato in termini di matrici:

$$X_{k+1} = X_k - E_k \quad \text{dove } E_k \text{ risolve} \quad X_k E + E X_k = -\left(X_k^2 - A\right)$$

(equazione di Sylvester)

(2) Prendo il metodo di Newton scalare, e sostituisco $x \to X$

$$X_{k+1} = \frac{1}{2}\left(X_k + X_k^{-1} A\right)$$

Costo di (1): un'eq. di Sylvester per passo $\Rightarrow$ una forma di Schur per passo

Costo di (2): Un'inversa per passo

Convergenza di (1): è un metodo di Newton multivariato, quindi (se $f'(X)$ è invertibile nella soluzione) converge localmente di ordine 2

Convergenza di (2): dobbiamo studiarla!

In realtà, questi due metodi sono sostanzialmente lo stesso metodo se iniziamo l'iterazione da X_0 che commuti con A: per esempio, $X_0 = A$, $X_0 = I$, $X_0 = \alpha I$ $\alpha \in \mathbb{C}$ ...

Metodo (1) richiede di risolvere $E_1 X_k + X_k E_1 = A - X_k^2$

Metodo (2) usa la versione scalare, $E_2 = X_{k+1} - X_k = (2X_k)^{-1}(A - X_k^2)$

quindi in pratica assume che E e X_k commutino:

$$E X_k + X_k E = 2 X_k E = A - X_k^2$$

Ma, effettivamente, se parto da X_0 che commuta con A, ad ogni passo le iterate sono funzioni razionali di A e X_0, quindi commutano con A. In particolare,

$E_2 = 2X_k^{-1}(A - X_k^2)$ commuta con A, quindi risolve

l'eq. di Sylvester $X_k E_2 + E_2 X_k = 2 X_k E_2 = A - X_k^2$,

quindi $E_1 = E_2$!

Quindi (1) e (2) sono lo stesso metodo se parto da X_0 che commuta con A!

Altro modo (indipendente) di stabilire convergenza di (2):

$X_{k+1} = \frac{1}{2}\left(X_k + X_k^{-1}A\right)$ $\quad X_0 = \alpha I$ prendo $\boxed{\alpha > 0}$

Premoltiplico per $A^{-1/2}$ e uso commutatività:

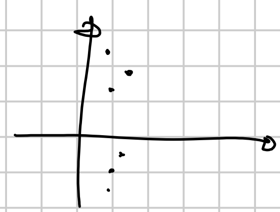
$$A^{-1/2}X_{k+1} = \frac{1}{2}\left(A^{-1/2}X_k + A^{-1/2}X_k^{-1}A\right) = \frac{1}{2}\left(A^{-1/2}X_k + \left(A^{-1/2}X_k\right)^{-1}\right)$$

$A^{-1/2}X_0 = \alpha A^{-1/2}$

Questa è esattamente l'iterazione segno su $Z_k = A^{-1/2}X_k$!

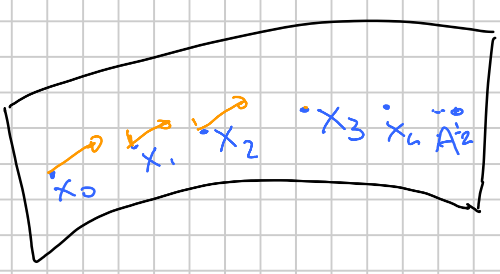
Converge quadraticamente a $\lim Z_k = \operatorname{sign}(Z_0) = \operatorname{sign}(\alpha A^{-1/2}) = I$

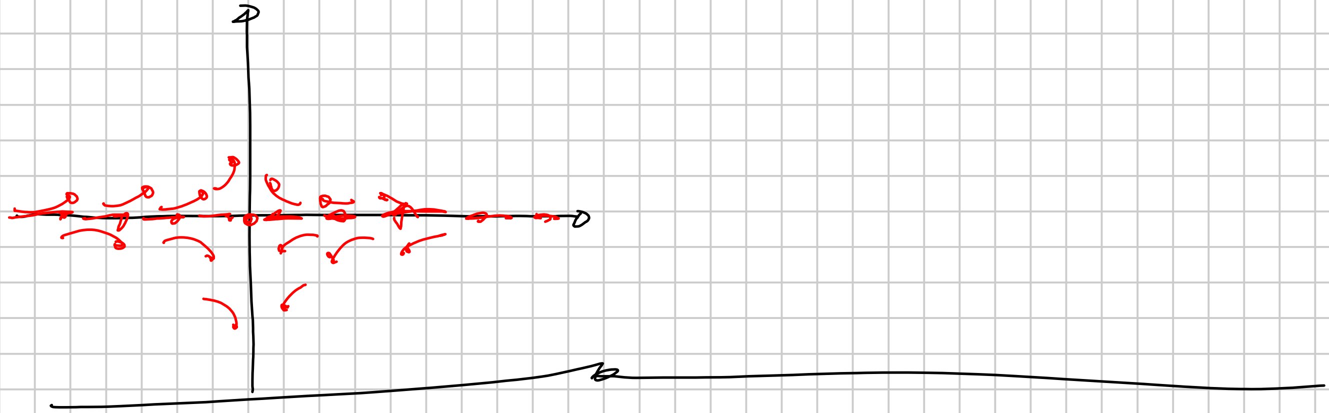
(perché $A^{1/2}$, e quindi anche $\alpha A^{-1/2}$, ha autoval. tutti nel semipiano dx)



Cosa sta succedendo numericamente?

← matrici che commutano con A





Possiamo studiare il comportamento di questa iterazione come stabilità di un sistema dinamico.

È vero o no che $X = A^{1/2}$ è un punto fisso stabile del sist. dinamico $h(X) = \frac{1}{2}(X + X^{-1}A)$?

Da teoria di sistemi dinamici, bisogna prendere gli autovalori dello Jacobiano $J = \nabla h$ (come funzione $\mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$ $\mathbb{C}^{n^2} \to \mathbb{C}^{n^2}$)

Se le potenze di J sono limitate, l'iterazione è stabile ($J, J^2, J^3, J^4, \dots$ non diverge) $\Leftrightarrow$ autovalori di J sono tutti in val. ass. ≤ 1, e quelli con val. ess. 1 non hanno blocchi di Jordan.

$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \to$ sì

$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \to$ no

$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^k = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$

Se il punto fisso non è stabile, l'iterazione $\{X_k\}$ <u>non</u> convergerà ad $A^{\frac{1}{2}}$ in modo stabile per piccole perturbazioni

Studiamo questi autovalori:

$$h(X) = \frac{1}{2}(X + X^{-1}A)$$

$$h(X+E) = \frac{1}{2}\left(X + E + (X+E)^{-1}A\right)$$

$$= \frac{1}{2}\left(X + E + (X^{-1} - X^{-1}EX^{-1})A\right) + o(\|E\|)$$

sviluppo perturbativo di $(X+E)^{-1}$:

$$(X+E)^{-1} - X^{-1} = (X+E)^{-1}\left(X - (X+E)\right)X^{-1}$$

$$= -(X+E)^{-1}EX^{-1}$$

$$= -X^{-1}EX^{-1} + o(\|E\|)$$

$$= \underbrace{\frac{1}{2}\left(X + X^{-1}A\right)}_{h(X)} + \underbrace{\frac{1}{2}\left(E - X^{-1}EX^{-1}A\right)}_{L_{h,X}(E)} + o(\|E\|)$$

$\mathrm{vec}(MEN) = (N^T \otimes M)\,\mathrm{vec}(E)$

$$\hat{L}_{h,X} = \frac{1}{2}\left(I \otimes I - (X^{-1}A)^T \otimes X^{-1}\right)$$

$$\hat{L}_{h,A^{1/2}} = \frac{1}{2}\left(I \otimes I - (A^{1/2})^T \otimes A^{-1/2}\right)$$

$(A^{\frac{1}{2}})^T = Q_1 T_1 Q_1^*$

$A^{-1/2} = Q_2 T_2 Q_2^*$

$$= \frac{1}{2}(Q_1 \otimes Q_2)\underbrace{\left(I \otimes I - T_1 \otimes T_2\right)}_{\text{triang. sup.}}(Q_1 \otimes Q_2)^*$$

sulla diagonale: $1 - \lambda_i^{1/2}\lambda_j^{-1/2}$ ($\lambda_1, \dots \lambda_n$ autoval. di A)

$\Rightarrow$ Gli autoval. di $\hat{L}_{h,A^{1/2}}$ sono $1 - \lambda_i^{1/2}\lambda_j^{-1/2}$, $i,j = 1,\dots n$

In generale, $|1 - \lambda_i^{1/2}\lambda_j^{-1/2}| \leq 1$ è falso

---

La stabilità asintotica è diversa per altre varianti: $X_{k+1} = \frac{1}{2}(X_k + X_k^{-1}A)$

Scriviamo $Y_k = A^{-1}X_k$

$$\begin{cases} X_{k+1} = \frac{1}{2}(X_k + Y_k^{-1}) \\ Y_{k+1} = \frac{1}{2}(Y_k + X_k^{-1}) \end{cases}$$

Iterazione di Denman-Beavers (DB)

$X_0 = \alpha A$

$Y_0 = \alpha I$

$$Y_{k+1} = A^{-1}X_{k+1} = \frac{1}{2}A^{-1}\left(X_k + X_k^{-1}A\right) \underset{\text{commutatività}}{=} \frac{1}{2}\left(A^{-1}X_k + X_k^{-1}\right) = \frac{1}{2}\left(Y_k + X_k^{-1}\right)$$

In aritm. esatta, $X_k \to A^{1/2}$ $Y_k \to A^{-1/2}$

DB costa 2 inverse per passo e calcola contemporaneamente $A^{\frac{1}{2}}$, $A^{-1/2}$

A differenza dell'altra, DB è asintoticamente stabile!

Calcoliamo $L_{DB,(X,Y)}$, che manda (E,F) in una certa coppia

$$\begin{bmatrix} \operatorname{vec} E \\ \operatorname{vec} F \end{bmatrix} \longrightarrow \begin{bmatrix} \vdots \\ \vdots \end{bmatrix} \quad \text{lungo } 2n^2$$

$$g_{DB}(X+E, Y+F) = \left(\frac{1}{2}\left((X+E)+(Y+F)^{-1}\right), \; \frac{1}{2}\left((Y+F)+(X+E)^{-1}\right) \right)$$

$$= \left(\frac{1}{2}\left(X+E+Y^{-1}-Y^{-1}FY^{-1}\right), \; \frac{1}{2}\left(Y+F+X^{-1}-X^{-1}EX^{-1}\right) \right) + o(\|E\|+\|F\|)$$

$$L_{DB,(X,Y)} = \left(\frac{1}{2}\left(E-Y^{-1}FY^{-1}\right), \; \frac{1}{2}\left(F-X^{-1}EX^{-1}\right) \right)$$

$$L_{DB,(A^{\frac{1}{2}},A^{-\frac{1}{2}})} = \left(\frac{1}{2}\left(E-A^{\frac{1}{2}}FA^{\frac{1}{2}}\right), \; \frac{1}{2}\left(F-A^{-\frac{1}{2}}EA^{-\frac{1}{2}}\right) \right)$$

$L_{DB,(A^{\frac{1}{2}},A^{-\frac{1}{2}})}^2 = (E, F)$ = identità. $\Rightarrow L_{DB,(A^{\frac{1}{2}},A^{-\frac{1}{2}})}$ è idempotente

$\Rightarrow$ autovalori sono $1, -1$ e sono semplici

Le potenze $L_{DB}^2, L_{DB}^3, L_{DB}^4, \ldots = I, L_{DB}, I, L_{DB}, \ldots$ sono limitate $\Rightarrow$ asintot. stabile.

(Altre varianti su [Higham, cap. 6])