Note Title 2020-04-23

Funzioni di matrici larghe e sparse

[Güttel '13]

Di solito, vogliamo $f(A)b$ ($f(A)$ è piena)

(stessa situazione di $f(x)=x^{-1}$: uno non vuole A^{-1}, ma $A^{-1}b$)

Tecniche:

1) Rimpiazzo f con un suo polinomio approssimante, e.g.

$$f(x)=\exp(x) \approx p(x) = 1+x+\frac{x^2}{2}+\frac{x^3}{3!}$$

$$f(A)b = \left(I + A + \frac{1}{2}A^2 + \frac{1}{6}A^3\right)b = b + Ab + \frac{1}{2}A^2 b + \frac{1}{6}A^3 b$$

calcolo Ab, $A(Ab)$ $A(A^2b)$, 3 prodotti mat.-vec.

2) Rimpiazzo f con una funz. raz. approssimante

$$\exp(x) \approx \frac{x^2+6x+12}{x^2-6x+12} = r(x) \quad \text{(approssimante di Padé (2,2))}$$

e calcolo $r(A)b = (A^2-6A+12I)^{-1}(A^2+6A+12I)b$

non vogliamo formarla esplicitamente (servono metodi iterativi)

Oppure: frazioni parziali:

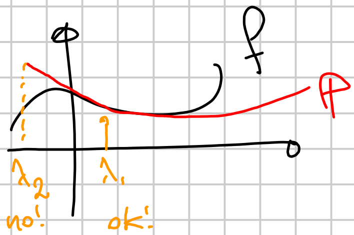
$$r(x) = \frac{a_1}{x-\xi_1} + \frac{a_2}{x-\xi_2}$$

(ξ_1, ξ_2 soluzioni (eventualmente complesse) di $x^2-6x+12=0$, denominatore $=0$)

$$r(A)b = a_1(A-\xi_1 I)^{-1}b + a_2(A-\xi_2 I)^{-1}b$$

(solo inversioni di matrici del tipo $A-\xi I$, A shiftata; si possono fare anche con un metodo diretto)

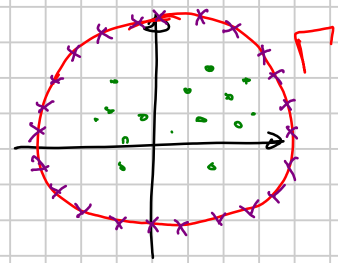
(Tipicamente, $p(x)$ non sarà un polinomio interpolatore di grado n, si sceglie un polinomio o funzione razionale di grado basso; servirà un polinomio/fun. raz. che approssimi bene f sullo spettro di A)



Es: se $\|A\| \leq 5.4$, l'appr. di Padé $(13,13)$ è t.c. $\|\exp(A) - r(A)\| \leq u\|A\|$

Trade-off: funzioni razionali approssimano meglio, ma sono più costose
(→ sist. lineari shiftati vs. → prodotti mat-vec per i polinomi)

3)



$$f(A)b = \left(\frac{1}{2\pi i}\int_\Gamma f(z)(zI-A)^{-1}\,dz\right) b$$

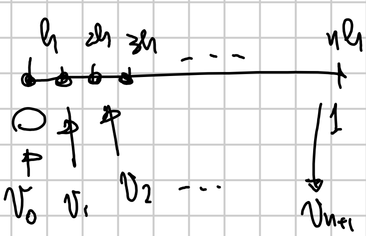
$$\approx \sum_{k=1}^{n} w_k f(\xi_k)(\xi_i I - A)^{-1} b$$

Di nuovo, serve risolvere sist. della forma $(A - \xi I)^{-1} b$
Tipicamente, serve risolverne tanti, e con ξ_i molto diversi: meglio usare un'approssimante raz. con poli scelti bene.

4) Metodi ad-hoc per la funzione: per esempio $f(x) = \exp(x)$

$$\exp(A)b = v(1) \qquad \begin{cases} \dot{v}(t) = A v(t) \\ v(0) = b \end{cases} \rightarrow v(t) = \exp(tA)\cdot b$$

Posso usare un metodo numerico per eq. differenziali, ad esempio



$$v_{k+1} \approx v(kh) = v_k + hAv_k = (I + hA)v_k$$

$$v_{n+1} \approx v(1) \qquad \left(I + \frac{1}{n}A\right)^n v_0 \approx \exp(A) v_0$$

5) Metodi che usano Arnoldi

<u>Krylov space</u>

$$A \in \mathbb{C}^{m \times m}, \quad m \geq n$$

$$K_n(A,b) = \text{span}(b, Ab, A^2b, \dots A^{n-1}b)$$

dim. spazio

$$= \{(\alpha_0 + \alpha_1 A + \alpha_2 A^2 + \dots + \alpha_{n-1}A^{n-1})b : \alpha_0, \alpha_1, \dots \alpha_{n-1} \in \mathbb{C}\}$$

$$= \{ p(A)b : p \text{ di grado } < n \}$$

Supponiamo di aver calcolato $Ab, A^2b, \dots A^n b$: questo ci dice come calcolare Av per ogni $v \in K_n(A,b)$:

$$v = \alpha_0 b + \alpha_1 Ab + \dots + \alpha_{n-1}A^{n-1}b = p(A)b$$

$$\Rightarrow Av = \alpha_0 Ab + \alpha_1 A^2 b + \dots + \alpha_{n-1}A^n b = Ap(A)b$$

Spesso, questo non è ~~soddisfacente~~: $A^k b \to$ autovettore dominante di A (in un senso opportuno), quindi $Ab, A^2b, A^3b, \dots$ "tendono a diventare lo stesso vettore".

Alcune operazioni, ad es. calcolare gli α_i a partire da v, sono mal condizionate.

$$\underset{\text{base di } K_n(A,b)}{V_n} = [b \; Ab \; A^2b \; \underbrace{\dots \; A^{n-1}b}_{\text{colonne circa parallele}}] \qquad \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{n-1} \end{bmatrix} = \underbrace{(V_n^* V_n)^{-1}}_{\text{mal condizionata}} V_n^* b$$

Sarebbe molto meglio avere una ortonormale ~~ortogonale~~, invece:

$V_n = [v_1 \; v_2 \; \dots \; v_n]$ ortonormale ~~ortogonale~~: in questo caso,

$$v_i^* v_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

$$v = \beta_1 v_1 + \beta_2 v_2 + \dots + \beta_n v_n, \qquad \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} = \underbrace{(V_n^* V_n)^{-1}}_{= I} V_n^* b$$

"Ricetta" per calcolare Av:

$$A\,[\underbrace{b \;\; Ab \;\; \cdots \;\; A^{n-1}b}_{V_n}] = [b \;\; \underbrace{Ab \;\; A^2b \;\; \cdots}_{V_{n+1}} \;\; A^n b] \underbrace{\begin{bmatrix} 0 & & & \\ 1 & \ddots & & 0 \\ & \ddots & \ddots & \\ 0 & & \ddots & 0 \\ & & & 1 \end{bmatrix}}_{\underline{H}_n}$$

mi dice come agisce A sulle colonne di V_n

Vogliamo una relazione dello stesso tipo con V_n ortogonale:

$$AV_n = V_{n+1} \cdot \begin{bmatrix} \ddots & & \\ & \underline{H}_n & \\ & & \ddots \end{bmatrix}, \qquad \underline{H}_n \in \mathbb{C}^{(n+1)\times n}$$

Possiamo calcolare insieme i v_i e gli elementi di $\underline{H}_n$ iterativamente:

supponiamo di aver già calcolato $[v_1, v_2, \ldots v_j]$ (base di $K_j(A,b)$)

(*)

$$\underbrace{w = Av_j}_{\text{noto}} = \underbrace{v_1}_{\text{noto}}\,\alpha_{1,j} + v_2\alpha_{2,j} + \cdots + v_j\alpha_{j,j} + \underbrace{v_{j+1}\alpha_{j+1,j}}_{\text{ignoto}} \in K_{j+1}(A,b)$$

vett. scalare

(se $v_j = p(A)b$ con p di grado $<j$, allora $Av_j = Ap(A)b$ cioè $\tilde{p}(A)b$ con $\tilde{p} = xp$ di grado $<j+1$)

Utilizzando ortogonalità, posso calcolare

$$\alpha_{i,j} = v_i^* A v_j$$

e sottrarli.

rimane

$$z = Av_j - v_1\alpha_{1j} - \cdots - v_j\alpha_{jj} = \underbrace{v_{j+1}}_{=\frac{z}{\|z\|}}\underbrace{\alpha_{j+1,j}}_{\|z\|}$$

$V = [v_1 \; v_2 \; \cdots \; v_n \; v_{n+1}]$ $\quad v_1 =$ base o.n. di span(b)

$$v_1 = \frac{b}{\|b\|}$$

$$\underline{H} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & & \\ 0 & \alpha_{32} & \alpha_{33} & \ddots & \\ 0 & 0 & \alpha_{43} & \ddots & \end{bmatrix}$$

$$\begin{bmatrix} \vdots & & & \\ 0 & 0 & 0 & \ddots \\ & \vdots & \vdots & \ddots \\ 0 & 0 & 0 & \alpha_{n+1,n} \end{bmatrix}$$

$$AV_n = V_{n+1} \underline{H}_n$$

$$\shortparallel$$

$$A\begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} = \begin{bmatrix} v_1 & v_2 & \cdots & v_{n+1} \end{bmatrix} \underline{H}_n$$

Perché siamo partiti da $w = Av_j$ e non da $w = Av$ qualunque altro vettore $v \in K_j(A,b)$?

Perché mi serviva che Av avesse una componente $\neq 0$ lungo v_{j+1}:

se fossi partito ad es. da $w = Av_1$,

allora $v_1 = \frac{b}{\|b\|}$ si scrive come $p(A)b$ per p di grado 0

$Av_1 = \frac{Ab}{\|b\|}$ si scrive come $\tilde{p}(A)b$ per $\tilde{p}(x) = x\,p(x)$ di grado 1,

cioè come comb. lin. di v_0 e v_1,

quindi avrei avuto $w = Av_1 = \underbrace{v_1\alpha_{1j}}_{\neq 0} + \underbrace{v_2\alpha_{2j}}_{\neq 0} + \underset{=0}{\cdots} + \underset{=0}{v_j\alpha_{jj}} + \boxed{\underset{=0}{v_{j+1}\alpha_{j+1,j}}}$

Mi serve invece partire da un vettore $v = p(A)b$ con p di grado esattamente $j-1$, non meno. In questo modo, (v = "vettore di continuazione")

$w = Av = \underbrace{Ap(A)}_{\text{poly. di grado esattamente } j} v = v_1\alpha_{1j} + \cdots + v_j\alpha_{jj} + \underbrace{v_{j+1}\alpha_{j+1,j}}_{\neq 0}$

Cioè, voglio partire da $v \in K_j(A,b) \setminus K_{j-1}(A,b)$

v_j è un buon vettore di continuazione, perché Av_j non ha una componente automaticamente uguale a zero lungo v_{j+1}

(Anche con queste precauzioni, è tuttavia possibile che $\alpha_{j+1,j} = 0$: succede quando ho "breakdown", cioè quando $K_j(A,b) = K_{j-1}(A,b)$, ovvero $A^j b = p(A) b$ per un qualche polinomio di grado $< j$

$w = Av$ è sempre comb. lineare di $(v_1, \dots v_j)$ per ogni $v \in K_j(A,b)$

Per un certo indice k, prima o poi,

$$\Big[\underbrace{b, Ab, A^2b, A^3b, \dots, A^k b}_{\text{lin. indipendenti}}, \underbrace{A^{k+1} b}_{\substack{\text{comb. lineare}\\ \text{dei precedenti}}}\Big]$$

Succederà al massimo dopo m passi per $A \in \mathbb{C}^{m \times m}$, perché esiste p di grado m t.c. $p(A) = 0$ (pol. caratteristico)

Al minimo, potrebbe succedere già al primo passo:

Ab è un multiplo di b (quando sono partito da b autovett. di A)

Quando raggiungo breakdown, vuol dire che ho trovato un polinomio $d(x)$ tale che $d(A) b = 0$. Allora, quando questo succede, so calcolare i prodotti $p(A) b$ anche per b di grado maggiore:

Scrivo $p(x) = q(x) d(x) + r(x)$, e ho $p(A) b = \cancel{q(A) d(A) b} + r(A) b$

Quando ho breakdown, $A K_j(A,b) \subseteq K_j(A,b)$

$\Rightarrow K_j(A,b)$ è un sottosp. invariante di A.

Arnoldi

$$\underbrace{A [v_1\ v_2\ \dots\ v_n]}_{V_n} = \underbrace{[v_1\ v_2\ v_3\ \dots\ v_n\ v_{n+1}]}_{V_{n+1}} \underbrace{\begin{bmatrix} \times & \times & \cdots & \times \\ \times & \times & & \vdots \\ 0 & \times & & \\ 0 & 0 & \times & \\ \vdots & 0 & \times & \\ & \vdots & & \ddots \\ 0 & 0 & & \times \end{bmatrix}}_{\underline{H}_n}$$

H_n

Quando ho breakdown, $H_{n+1,n} = 0$ $\quad A V_n = V_n H_n$.

Come usiamo Arnoldi per calcolare funzioni di matrici?

Idea: $f(A) \approx V_n f(H_n) \underbrace{V_n^* b}$ $\qquad V_n^* A V_n = V_n^* V_{n+1} \underline{H}_n = H_n$

$= V_n f(H_n) e_1 \cdot \|b\|$

V_n ha colonne ortogonali, e la prima è $\frac{b}{\|b\|}$ $\Rightarrow$ $V_n^* b = \begin{bmatrix} \|b\| \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \|b\| \cdot e_1$

Lemma: per tutti i polinomi di grado $< n$,

$$p(A) b = V_n p(H_n) V_n^* b \qquad (*)$$

Dim: dimostriamo che $A^j b = V_n H_n^j V_n^* b \quad \forall\, j = 0, 1, \dots n-1$

poi $(*)$ segue prendendo comb. lineari.

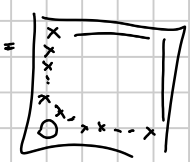
Passo base:
$$A^0 b = V_n H_n^0 \underbrace{V_n^* b}_{= e_1 \cdot \|b\|} = V_n \cdot e_1 \cdot \|b\| = \frac{b}{\|b\|} \cdot \|b\| = b$$

Passo induttivo:
$$A^{j+1} b = A \underbrace{A^j b} = A V_n \underbrace{H_n^j V_n^* b} = \underbrace{V_{n+1} \underline{H}_n} H_n^j \underbrace{V_n^* b}_{e_1 \cdot \|b\|}$$

$$= V_n H_n H_n^* e_1 \cdot \|b\|$$

l'ultima colonna di V_{n+1} e l'ultima riga di $\underline{H}_n$ non servono, perché $H_n^j e_1$ = prima colonna di H_n^j

H_n^j ha j sottodiagonali diverse da 0

$H_n^j =$ 

$\Rightarrow$ se $j < n$, $H_n^j e_1$ ha uno 0 come ultimo elemento

$\Rightarrow$ quando faccio i prodotti, l'ultima riga di $\underline{H}_n$ non influisce sul prodotto.