Correzione del I esonero di Analisi Numerica 1 del 28 Novembre 2013

1 Sia

$$A = \begin{bmatrix} 4\alpha & -2 & 1 & 0 \\ -2 & 4 & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & 4 & -1 \\ 0 & 0 & -1 & 4 \end{bmatrix}, \ \alpha \in \mathbb{R}.$$

(i) Per quali valori di $\alpha \in [-1, 3/4]$ la matrice A può avere un autovalore nullo?
(ii) Per $\alpha = -1$ dare una limitazione superiore e una inferiore per $\mu_2(A)$.
(iii) Per $\alpha = 3/4$ dire se il metodo di Jacobi per la risoluzione del sistema $A\mathbf{x} = \mathbf{b}$ è convergente.
(iv) Per quali valori di α vale la decomposizione di Cholesky $A = LL^H$?

Risoluzione

(i) A prima vista, cioè senza calcolarne il determinante, la matrice A può avere un autovalore nullo solo per $\alpha \in (-3/4, 3/4)$, infatti, per gli altri valori di α, A ha diagonale (almeno) debolmente dominante, dunque, essendo irriducibile, deve essere non singolare (come conseguenza dei teoremi I e III di Gershgorin). Queste osservazioni sono dimostrate qui di seguito.

La matrice A è irriducibile (per ogni valore di α) perché nel grafo orientato G associato ad A ci sono gli archi $ii+1$ e $i+1i$, $\forall\, i = 1, \ldots, n-1$ (perché $a_{i+1,i}a_{i.i+1} \neq 0 \ \forall\, i$), dunque ogni coppia di vertici di G è sicuramente connessa da un cammino, cioè G è fortemente connesso.

Poiché $|4| > |-1| = 1$, $|4| > |1| + |\frac{1}{2}| + |-1| = 5/2$, $|4| > |-2| + |\frac{1}{2}| = 5/2$, la matrice A ha diagonale almeno debolmente dominante se e solo se $|4\alpha| \geq |-2| + |1| = 3$, cioè se e solo se $|\alpha| \geq 3/4$.

Risolvere così l'esercizio basta. Se poi si vogliono trovare precisamente i valori (o il valore) di α per cui A ha un autovalore nullo, allora occorre calcolarne il determinante. Si ha che $\det(A) = 236\alpha - 84$. Dunque A ha un autovalore nullo (o equivalentemente A è singolare) se e solo se $\alpha = 84/236 = 21/59$ (se i conti sono giusti).

(ii) Per $\alpha = -1$ la matrice A diventa:

$$A = \begin{bmatrix} -4 & -2 & 1 & 0 \\ -2 & 4 & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & 4 & -1 \\ 0 & 0 & -1 & 4 \end{bmatrix}.$$

Il primo teorema di Gershgorin e il fatto che A è hermitiana, permettono di concludere che gli autovalori di A sono nel sottoinsieme della retta reale $[-7, -1] \cup [1.5, 6.5]$. Per il terzo teorema di Gershgorin (che si può applicare perché A è irriducibile, vedi sopra), gli estremi di tali intervalli, i numeri $-7, -1, 1.5, 6.5$, non possono essere autovalori di A, perché ognuno di essi sta sulla frontiera dell'unione dei cerchi di Gershgorin, ma non sta sulla frontiera di ogni cerchio di Gershgorin (ad esempio 1.5 non sta sulla frontiera del primo cerchio di Gershgorin).

Dunque, gli autovalori di A sono nel sottoinsieme della retta reale $(-7, -1) \cup (1.5, 6.5)$. Inoltre, poiché il primo cerchio di Gershgorin è disgiunto dagli altri tre cerchi, per il secondo teorema di Gershgorin si ha che un autovalore di A è nell'intervallo $(-7, -1)$ e i rimanenti tre autovalori di A sono nell'intervallo $(1.5, 6.5)$.

Siano $\lambda_i(A)$, $i = 1, 2, 3, 4$, gli autovalori di A. Poiché A è hermitiana,

$$\mu_2(A) := \|A\|_2 \|A^{-1}\|_2 = \frac{\max_i |\lambda_i(A)|}{\min_i |\lambda_i(A)|}.$$

Sia $\lambda_{\max}$ l'autovalore di A per cui $|\lambda_{\max}| = \max_i |\lambda_i(A)|$. Sia $\lambda_{\min}$ l'autovalore di A per cui $|\lambda_{\min}| = \min_i |\lambda_i(A)|$. Nel dare una limitazione superiore per $\mu_2(A)$ vanno considerati due casi.

I caso. Se $\lambda_{\max}$ è in $(-7, -1)$ (cioè nel primo cerchio di Gershgorin), allora $\lambda_{\min}$ deve essere in $(1.5, 6.5)$, perché il primo cerchio di Gershgorin non può contenere più di un autovalore di A. Dunque, in questo I caso, $\mu_2(A) \leq 7/1.5$ (perché $|\lambda_{\max}| \leq 7$ e $|\lambda_{\min}| \geq 1.5$).

II caso. Se $\lambda_{\max}$ è in $(1.5, 6.5)$, allora $\lambda_{\min}$ può essere nell'intervallo $(-7, -1)$ oppure nell'intervallo $(1.5, 6.5)$, non lo si può dire. Dunque, in questo II caso si ha $\mu_2(A) \leq 6.5/1$ (perché $|\lambda_{\max}| \leq 6.5$ e $|\lambda_{\min}| \geq \min\{1.5, 1\} = 1$).

(iii) Sappiamo che la matrice A è irriducibile. Per $\alpha = 3/4$ essa ha la diagonale debolmente dominante. Allora A è non singolare (per i teoremi I e III di Gershgorin) e, dato $\mathbf{b} \in \mathbb{R}^n$, il metodo di Jacobi

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\Big(- \sum_{j \neq i} a_{ij} x_j^{(k)} + b_i \Big), \; i = 1, \ldots, n \; (k = 0, 1, 2, \ldots)$$

genera una successione di vettori $\mathbf{x}^{(k)}$ ben definita e convergente ad $A^{-1}\mathbf{b}$, comunque si scelga il vettore iniziale $\mathbf{x}^{(0)} \in \mathbb{R}^n$.

Infatti, per i teoremi I e III di Gershgorin, la matrice di iterazione del metodo di Jacobi, $J = [-(1 - \delta_{rs}) a_{rs}/a_{rr}]_{r,s=1}^4$, ha gli autovalori tutti contenuti nella parte interna del cerchio di centro l'origine e raggio 1.

(iv) Se fosse A definita positiva, allora esisterebbe una matrice triangolare inferiore L, con elementi diagonali tutti diversi da zero, tale che $A = LL^H$, cioè si potrebbe ottenere la decomposizione di Cholesky di A (vedi lezione).

Calcoliamo allora i valori di α per cui A è definita positiva. Notiamo che A è reale simmetrica per ogni valore di α. Per essere A definita positiva è innanzitutto necessario che i suoi elementi diagonali siano positivi; dunque è necessario che $4\alpha > 0$, cioè $\alpha > 0$. Se fosse $\alpha \geq 3/4$, si avrebbe

$$\begin{aligned}
&4\alpha = a_{11} = |a_{11}| \geq \textstyle\sum_{j \neq 1} |a_{1j}| = |-2| + |1| = 3, \\
&4 = a_{22} = |a_{22}| > \textstyle\sum_{j \neq 2} |a_{2j}| = |-2| + |1/2| = 5/2, \\
&4 = a_{33} = |a_{33}| > \textstyle\sum_{j \neq 3} |a_{3j}| = |1| + |1/2| + |-1| = 5/2, \\
&4 = a_{44} = |a_{44}| > \textstyle\sum_{j \neq 4} |a_{4j}| = |-1| = 1,
\end{aligned}$$

cioè A avrebbe la diagonale (almeno) debolmente dominante e i suoi elementi diagonali sarebbero positivi. Allora, essendo A irriducibile e reale simmetrica, A sarebbe definita positiva (perché avrebbe gli autovalori tutti positivi, per i teoremi I e III di Gershgorin) e, quindi, ammetterebbe la decomposizione di Cholesky.

Risolvere così l'esercizio basta. Se poi si vogliono trovare tutti i valori di $\alpha \in \mathbb{R}$ per cui la matrice (simmetrica reale) A è definita positiva, occorre imporre che i determinanti delle sottomatrici 1×1, 2×2, 3×3, 4×4 in alto a sinistra di A siano tutti positivi, ovvero devono valere contemporaneamente le seguenti quattro disuguaglianze:

$$4\alpha > 0, \quad 16\alpha - 4 > 0, \quad 63\alpha - 22 > 0, \quad 236\alpha - 84 > 0.$$

Dunque, A è definita positiva se e solo se $\alpha > 21/59$ (se i conti non sono sbagliati).

Nota: se $\alpha = 21/59$ esiste ancora L triangolare inferiore tale che $A = LL^H$, ma $L_{44} = 0$.

2 (i) Dimostrare che $n \max_{i,j=1,\dots,n} |a_{ij}|$ è una norma matriciale non indotta.
(ii) Dimostrare che $\rho(A) \leq \|A\|$ per ogni norma matriciale.

Risoluzione
Una norma matriciale $\|\cdot\|$ soddisfa le seguenti quattro proprietà ($A, B \in \mathbb{C}^{n\times n}$, $\alpha \in \mathbb{C}$):
1) $\|A\| \geq O \;\forall\, A$, e $\|A\| = 0$ se e solo se $A = O$;
2) $\|\alpha A\| = |\alpha|\|A\|$;
3) $\|A + B\| \leq \|A\| + \|B\|$;
4) $\|AB\| \leq \|A\|\|B\|$.
(i) Si vuole far vedere che $\varphi(A) := n \max_{ij} |a_{ij}|$ soddisfa le quattro proprietà sopra elencate.
1): È evidente che $\varphi(A)$ è, per ogni matrice A, un numero maggiore o uguale di zero. Inoltre, $\varphi(A) = n \max_{ij} |a_{ij}| = 0 \Leftrightarrow \max_{ij} |a_{ij}| = 0 \Leftrightarrow |a_{ij}| = 0 \;\forall\, i, j \Leftrightarrow a_{ij} = 0 \;\forall\, i, j \Leftrightarrow A = O$.
2): $\varphi(\alpha A) = n \max_{ij} |(\alpha A)_{ij}| = n \max_{ij} |\alpha a_{ij}| = n \max_{ij} |\alpha||a_{ij}| = |\alpha| n \max_{ij} |a_{ij}| = |\alpha|\varphi(A)$.
3): $\varphi(A+B) = n \max_{ij} |(A+B)_{ij}| = n \max_{ij} |a_{ij} + b_{ij}| \leq n \max_{ij}(|a_{ij}| + |b_{ij}|) \leq n(\max_{ij} |a_{ij}| + \max_{ij} |b_{ij}|) = \varphi(A) + \varphi(B)$.
4): $\varphi(AB) = n \max_{ij} |(AB)_{ij}| = n \max_{ij} |\sum_k a_{ik}b_{kj}| \leq n \max_{ij} \sum_k |a_{ik}b_{kj}| \leq n \max_{ij} \max_k |a_{ik}b_{kj}| \sum_k 1 = n^2 \max_{ij} \max_k |a_{ik}||b_{kj}| \leq n^2 \max_{ir} |a_{ir}| \max_{sj} |b_{sj}| = \varphi(A)\varphi(B)$
Nota: invece la funzione $\psi(A) = \max_{ij} |a_{ij}|$ non è una norma matriciale perché non verifica la quarta proprietà:

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix},\; B = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix},\; AB = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix},\; \psi(AB) = 2 > \psi(A)\psi(B) = 1 \cdot 1 = 1.$$

(ii) Sia λ un generico autovalore di A, dunque esiste $\mathbf{x} \neq \mathbf{0}$ tale che $A\mathbf{x} = \lambda\mathbf{x}$. Seguono due possibili risoluzioni dell'esercizio.

a) (Pellico) Allora $A\mathbf{x}\mathbf{x}^H = \lambda\mathbf{x}\mathbf{x}^H$ e, per la seconda e quarta proprietà di $\|\cdot\|$,

$$\|A\|\|\mathbf{x}\mathbf{x}^H\| \geq \|A\mathbf{x}\mathbf{x}^H\| = \|\lambda\mathbf{x}\mathbf{x}^H\| = |\lambda|\|\mathbf{x}\mathbf{x}^H\|.$$

Essendo $\mathbf{x}\mathbf{x}^H \neq O$ (ad es. è diverso da zero almeno un elemento diagonale $|x_i|^2$ di $\mathbf{x}\mathbf{x}^H$), per la prima proprietà di $\|\cdot\|$ si ha $\|\mathbf{x}\mathbf{x}^H\| \neq 0$. Dunque si conclude che $|\lambda| \leq \|A\|$. Data l'arbitrarietà di λ, segue la tesi.

b) Allora $AX = \lambda X$, essendo X la matrice $n \times n$ con prima colonna uguale all'autovettore $\mathbf{x}$ e le rimanenti colonne nulle, $X = [\mathbf{x}|\mathbf{0}|\cdots|\mathbf{0}]$. Dunque

$$\|A\|\|X\| \geq \|AX\| = \|\lambda X\| = |\lambda|\|X\|.$$

Essendo $X \neq O$, si ha $\|X\| \neq 0$. Dunque si conclude che $|\lambda| \leq \|A\|$. Data l'arbitrarietà di λ, segue la tesi.

3 Sia A una matrice definita positiva $n \times n$ reale.
(i) Dimostrare che, se $\mathbf{v}^T A\mathbf{w} = 0$ per $\mathbf{v}, \mathbf{w} \neq \mathbf{0}$, allora $\mathbf{v}$ e $\mathbf{w}$ sono linearmente indipendenti.

(ii) Siano $\mathbf{u}_i$ $(i = 1, \ldots, n)$ vettori non nulli tali che $\mathbf{u}_i^T A \mathbf{u}_j = 0$ per $i \neq j$. Dimostrare che la soluzione del sistema lineare $A\mathbf{x} = \mathbf{b}$ ammette la rappresentazione

$$\mathbf{x} = \sum_{i=1}^{n} \frac{\mathbf{u}_i^T \mathbf{b}}{\mathbf{u}_i^T A \mathbf{u}_i} \mathbf{u}_i \ .$$

<u>Risoluzione</u>

(i) Sia $\alpha\mathbf{v} + \beta\mathbf{w} = \mathbf{0}$. Allora $\mathbf{v}^T A(\alpha\mathbf{v} + \beta\mathbf{w}) = 0 \Rightarrow \alpha\mathbf{v}^T A\mathbf{v} + \beta\mathbf{v}^T A\mathbf{w} = 0 \Rightarrow \alpha\mathbf{v}^T A\mathbf{v} = 0 \Rightarrow \alpha = 0$ ($\mathbf{v}^T A\mathbf{v} > 0$ perché A è definita positiva e $\mathbf{v}$ è non nullo). Analogamente, si ha $\mathbf{w}^T A(\alpha\mathbf{v} + \beta\mathbf{w}) = 0$ $\Rightarrow \alpha\mathbf{w}^T A\mathbf{v} + \beta\mathbf{w}^T A\mathbf{w} = 0 \Rightarrow \beta\mathbf{w}^T A\mathbf{w} = 0 \Rightarrow \beta = 0$ ($\mathbf{w}^T A\mathbf{w} > 0$ perché A è definita positiva e $\mathbf{w}$ è non nullo).

(ii) I vettori $\mathbf{u}_i$ formano una base in $\mathbb{R}^n$, dunque esistono numeri reali α_i tali che $\mathbf{x} := A^{-1}\mathbf{b} = \sum_{i=1}^n \alpha_i \mathbf{u}_i$. Da questa identità segue che

$$\mathbf{b} = A\mathbf{x} = \sum_{i=1}^{n} \alpha_i A\mathbf{u}_i, \quad \mathbf{u}_j^T \mathbf{b} = \sum_{i=1}^{n} \alpha_i \mathbf{u}_j^T A\mathbf{u}_i = \alpha_j \mathbf{u}_j^T A\mathbf{u}_j,$$

dunque per α_j si ottiene l'espressione esplicita $\alpha_j = \mathbf{u}_j^T \mathbf{b}/\mathbf{u}_j^T A\mathbf{u}_j$ ($\mathbf{u}_j^T A\mathbf{u}_j > 0$ perché A è definita positiva e $\mathbf{u}_j \neq \mathbf{0}$). Ne segue che $\mathbf{x} = \sum_{i=1}^n \frac{\mathbf{u}_i^T \mathbf{b}}{\mathbf{u}_i^T A\mathbf{u}_i} \mathbf{u}_i$.

Analisi Numerica 1 - I esonero - 28 Novembre 2013

1 Sia

$$A = \begin{bmatrix} 4\alpha & -2 & 1 & 0 \\ -2 & 4 & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & 4 & -1 \\ 0 & 0 & -1 & 4 \end{bmatrix}, \ \alpha \in \mathbb{R}.$$

(i) Per quali valori di $\alpha \in [-1, 3/4]$ la matrice A può avere un autovalore nullo?
(ii) Per $\alpha = -1$ dare una limitazione superiore e una inferiore per $\mu_2(A)$.
(iii) Per $\alpha = 3/4$ dire se il metodo di Jacobi per la risoluzione del sistema $A\mathbf{x} = \mathbf{b}$ è convergente.
(iv) Per quali valori di α vale la decomposizione di Cholesky $A = LL^H$?

2 (i) Dimostrare che $n \max_{i,=1,\ldots,n} |a_{ij}|$ è una norma matriciale non indotta.
(ii) Dimostrare che $\rho(A) \leq \|A\|$ per ogni norma matriciale.

3 Sia A una matrice definita positiva $n \times n$ reale.
(i) Dimostrare che, se $\mathbf{v}^T A\mathbf{w} = 0$ per $\mathbf{v}, \mathbf{w} \neq \mathbf{0}$, allora $\mathbf{v}$ e $\mathbf{w}$ sono linearmente indipendenti.
(ii) Siano $\mathbf{u}_i$ $(i = 1, \ldots, n)$ vettori non nulli tali che $\mathbf{u}_i^T A\mathbf{u}_j = 0$ per $i \neq j$. Dimostrare che la soluzione del sistema lineare $A\mathbf{x} = \mathbf{b}$ ammette la rappresentazione

$$\mathbf{x} = \sum_{i=1}^{n} \frac{\mathbf{u}_i^T \mathbf{b}}{\mathbf{u}_i^T A\mathbf{u}_i} \mathbf{u}_i \ .$$

Correzione del II esonero di Analisi Numerica 1 del 10 Gennaio 2014

1 (i) Modificando il metodo di Newton per il calcolo degli zeri di $f(x)$, introdurre un metodo $x_{i+1} = g(x_i) = \varphi(x_i, f(x_i))$ che applicato nel caso $f(x) = \alpha x - 1$, $\alpha > 0$, richieda ad ogni passo due sole moltipicazioni senza divisioni e sia tale che

$$x_i = \frac{1}{\alpha} - \varepsilon \quad \Rightarrow \quad x_{i+1} = \frac{1}{\alpha} - \alpha\varepsilon^2.$$

(ii) Verificare che il metodo $x_{i+1} = x_i + x_i(1 - \alpha x_i) + x_i(1 - \alpha x_i)^2$ per approssimare $1/\alpha$ ha ordine di convergenza almeno tre.

Risoluzione (i) Nel metodo di Newton, tramite la formula

$$x_0 \in \mathbb{R}, \; x_{i+1} = x_i - f(x_i)/f'(x_i), \quad i = 0, 1, \ldots,$$

si genera una successione di valori x_i che, sotto opportune ipotesi, è convergente a uno zero di f.

Per $f(x) = \alpha x - 1$, si ha $f'(x_i) = \alpha$. Visto che la successione x_i che vogliamo definire, tramite la formula $x_{i+1} = g(x_i)$, deve convergere ad $1/\alpha$ (altrimenti non potrebbe verificare la richiesta $x_i = 1/\alpha - \varepsilon \Rightarrow x_i = 1/\alpha - \alpha\varepsilon^2$), i valori $1/x_i$ dovranno assomigliare sempre più al valore di α. Quindi è naturale provare a sostituire, nel metodo di Newton, $f'(x_i) = \alpha$ con $1/x_i$:

$$x_0 \in \mathbb{R}, \; x_{i+1} = x_i - x_i f(x_i) = x_i - x_i(\alpha x_i - 1) = x_i(2 - \alpha x_i), \quad i = 0, 1, \ldots. \qquad (*)$$

È evidente che, dato x_i, il nuovo valore x_{i+1} si calcola con due moltiplicazioni ($\alpha \cdot x_i$ e $x_i \cdot (\alpha x_i - 1)$) e nessuna divisione.

Sia ε tale che $x_i = 1/\alpha - \varepsilon$. Allora

$$\begin{aligned} x_{i+1} &= \tfrac{1}{\alpha} - \varepsilon - (\tfrac{1}{\alpha} - \varepsilon)(\alpha(\tfrac{1}{\alpha} - \varepsilon) - 1) = \tfrac{1}{\alpha} - \varepsilon - (\tfrac{1}{\alpha} - \varepsilon)(1 - \varepsilon\alpha - 1) \\ &= \tfrac{1}{\alpha} - \varepsilon + (\tfrac{1}{\alpha} - \varepsilon)\varepsilon\alpha = \tfrac{1}{\alpha} - \varepsilon + \varepsilon - \alpha\varepsilon^2 = \tfrac{1}{\alpha} - \alpha\varepsilon^2, \;\; \Rightarrow \\ |x_{i+1} - \tfrac{1}{\alpha}| &= \alpha|x_i - \tfrac{1}{\alpha}|^2. \end{aligned}$$

Ne segue che (se l'approssimazione iniziale x_0 è scelta abbastanza vicino a $1/\alpha$, allora) la successione definita in (*) converge a $1/\alpha$ con ordine di convergenza quadratico.

(Nota: la successione di matrici $X_{i+1} = X_i(2 - AX_i)$ converge all'inversa di A ...)

(ii) Se nella seguente correzione di (*)

$$x_{i+1} = x_i + x_i(1 - \alpha x_i) + x_i(1 - \alpha x_i)^2 \qquad (**)$$

si pone $x_i = 1/\alpha - \varepsilon$, allora

$$\begin{aligned} x_{i+1} &= \tfrac{1}{\alpha} - \alpha\varepsilon^2 + (\tfrac{1}{\alpha} - \varepsilon)(\alpha(\tfrac{1}{\alpha} - \varepsilon) - 1)^2 = \tfrac{1}{\alpha} - \alpha\varepsilon^2 + (\tfrac{1}{\alpha} - \varepsilon)(-\varepsilon\alpha)^2 \\ &= \tfrac{1}{\alpha} - \alpha\varepsilon^2 + \varepsilon^2\alpha - \alpha^2\varepsilon^3 = \tfrac{1}{\alpha} - \alpha^2\varepsilon^3, \;\; \Rightarrow \\ |x_{i+1} - \tfrac{1}{\alpha}| &= \alpha^2|x_i - \tfrac{1}{\alpha}|^3. \end{aligned}$$

Ne segue che (se l'approssimazione iniziale x_0 è scelta abbastanza vicino a $1/\alpha$, allora) la successione definita in (**) converge a $1/\alpha$ con ordine di convergenza cubico.

Si può risolvere l'esercizio anche nel seguente modo. Si mette in evidenza la funzione di iterazione del metodo, cioè si scrive il metodo nella forma $x_{i+1} = g(x_i)$, dove $g(x) = x + x(1-\alpha x) + x(1-\alpha x)^2$, e si osserva che

$$g(x) = 3x - 3\alpha x^2 + \alpha^2 x^3,\ g'(x) = 3 - 6\alpha x + 3\alpha^2 x^2,\ g''(x) = 6\alpha^2 x - 6\alpha,\ g'''(x) = 6\alpha^2,$$

dunque $g(1/\alpha) = 1/\alpha$, $g'(1/\alpha) = 0$, $g''(1/\alpha) = 0$, $g'''(1/\alpha) = 6\alpha^2 \neq 0$. Quindi $x_{i+1} - 1/\alpha = g(x_i) - g(1/\alpha) = g'''(\xi)(x_i - 1/\alpha)^3/3! = 6\alpha^2(x_i - 1/\alpha)^3/3!$, cioè la successione $x_{i+1} = g(x_i)$, $i = 0, 1, \ldots$, converge a $1/\alpha$ con ordine di convergenza esattamente tre (se x_0 è scelto non troppo lontano da $1/\alpha$).

2 Siano $f \in C^0([a,b])$, $h = (b-a)/n$, $n \in \mathbb{N}$, e $x_i = a + ih$, $i = 0, 1, \ldots, n$.

(i) Introdurre delle funzioni $\varphi_0(x), \varphi_1(x), \ldots, \varphi_{n-1}(x), \varphi_n(x)$ continue in $[a,b]$ che soddisfano le seguenti due condizioni:

a) $\varphi_i(x_j) = \delta_{ij}$, $0 \leq i, j \leq n$;

b) posto $g(x) = \sum_{i=0}^n f(x_i)\varphi_i(x)$, si ha che $g|_{[x_{i-1},x_i]}$ è un polinomio di grado al più uno ($\forall\, i$). (Osservare che $g(x_j) = f(x_j)$, $j = 0, 1, \ldots, n$).

(ii) Nel caso $f \in C^2([a,b])$ trovare una limitazione superiore per $\max_{x\in[a,b]} |f(x) - g(x)|$ in funzione di h.

<u>Risoluzione</u>

La funzione $g = \sum_{j=0}^n f(x_j)\varphi_j$ in $[x_{i-1}, x_i]$ deve essere un polinomio di grado al più uno che in x_{i-1} vale $f(x_{i-1})$ e in x_i vale $f(x_i)$, cioè devono essere verificate le uguaglianze

$$f(x_{i-1})\frac{x - x_i}{x_{i-1} - x_i} + f(x_i)\frac{x - x_{i-1}}{x_i - x_{i-1}} = g|_{[x_{i-1},x_i]}(x) = (\sum_{j=0}^n f(x_j)\varphi_j)|_{[x_{i-1},x_i]}(x) = \sum_{j=0}^n f(x_j)\varphi_j|_{[x_{i-1},x_i]}(x), \quad (*)$$

$i = 1, 2, \ldots, n$. È evidente che le uguaglianze (*) sono verificate se in ogni x dell'intervallo $[x_{i-1}, x_i]$ le funzioni φ_{i-1} e φ_i valgono rispettivamente $(x - x_i)/(x_{i-1} - x_i)$ e $(x - x_{i-1})/(x_i - x_{i-1})$, e tutte le altre funzioni φ_j sono nulle ($\forall\, i$). Ma ciò è come dire

$$\varphi_0(x) = \begin{cases} \frac{x-x_1}{x_0-x_1}, & x \in [x_0, x_1] \\ 0, & \text{altrove} \end{cases}, \quad \varphi_n(x) = \begin{cases} \frac{x-x_{n-1}}{x_n-x_{n-1}}, & x \in [x_{n-1}, x_n] \\ 0, & \text{altrove} \end{cases},$$

$$\varphi_i(x) = \begin{cases} \frac{x-x_{i-1}}{x_i-x_{i-1}}, & x \in [x_{i-1}, x_i], \\ \frac{x_{i+1}-x}{x_{i+1}-x_i}, & x \in [x_i, x_{i+1}], \\ 0, & \text{altrove} \end{cases}, \quad i = 1, \ldots, n-1.$$

Verifichiamo che effettivamente tali funzioni φ_i soddisfano le proprietà richieste. È facile osservare che per costruzione $\varphi_i \in C^0([a,b])$, $\varphi_i(x_i) = 1$ e $\varphi_i(x_j) = 0$ se $j \neq i$. Inoltre, per $i = 1, \ldots, n$ si ha

$$\begin{aligned} &(\textstyle\sum_{j=0}^n f(x_j)\varphi_j)|_{[x_{i-1},x_i]}(x) = \textstyle\sum_{j=0}^n f(x_j)\varphi_j|_{[x_{i-1},x_i]}(x) \\ &= f(x_{i-1})\varphi_{i-1}|_{[x_{i-1},x_i]}(x) + f(x_i)\varphi_i|_{[x_{i-1},x_i]}(x) = f(x_{i-1})\tfrac{x-x_i}{x_{i-1}-x_i} + f(x_i)\tfrac{x-x_{i-1}}{x_i-x_{i-1}} \end{aligned}$$

($\varphi_j|_{[x_{i-1},x_i]} = 0$ se $j \neq i-1, i$).

(ii) Sia $x \in [a,b]$ generico. Allora esiste $i \in \{1,\dots,n\}$ tale che $x \in [x_{i-1},x_i]$. Ma, poiché g in $[x_{i-1},x_i]$ è il polinomio di grado al più uno interpolante f in x_{i-1} e in x_i, per quanto visto a lezione si ha

$$x \in [x_{i-1},x_i] \quad \Rightarrow \quad \exists \xi_i \in (x_{i-1},x_i) \;\Big|\; f(x)-g(x) = \tfrac{f''(\xi_i)}{2!}(x-x_{i-1})(x-x_i),$$
$$|f(x)-g(x)| = \tfrac{|f''(\xi_i)|}{2!}(x-x_{i-1})(x_i-x) \le \max_{x\in[x_{i-1},x_i]} |f''(x)| \tfrac{1}{2}\tfrac{(x_i-x_{i-1})^2}{4}$$

(la parabola $(x-x_{i-1})(x_i-x)$, positiva in $[x_{i-1},x_i]$, assume il suo valore massimo per $x=(x_{i-1}+x_i)/2$). Essendo $x_i - x_{i-1} = h$, $\forall\, i$, da questo risultato segue che $|f(x)-g(x)| \le Mh^2/8$, dove $M = \max_{x\in[a,b]} |f''(x)|$ ($f \in \mathcal{C}^2([a,b]) \Rightarrow M$ ben definito). Data l'arbitrarietà di x, si ha infine la maggiorazione

$$\max_{x\in[a,b]} |f(x)-g(x)| \le M\frac{h^2}{8}.$$

3 (i) Trovare i valori di A_i, $i=0,1,2$, per cui l'identità

$$\int_{-1}^{1} \frac{1}{1+x^2} f(x)\,dx = A_0 f(-1) + A_1 f(0) + A_2 f(1) + E$$

è verificata con $E=0$ quando f è un polinomio di grado minore o uguale a due. Per tali valori di A_i cosa si può dire su E quando $f(x) = x^3 - x^2$?

(ii) Scrivere i primi tre polinomi monici p_0, p_1, p_2 ortogonali rispetto all'intervallo $[-1,1]$ e al peso $\omega(x) = 1/(1+x^2)$.

(iii) Dimostrare che esistono valori di A_0 e A_1 per cui l'identità

$$\int_{-1}^{1} \frac{1}{1+x^2} f(x)\,dx = A_0 f\Big(-\sqrt{\frac{4}{\pi}-1}\Big) + A_1 f\Big(\sqrt{\frac{4}{\pi}-1}\Big) + E$$

è verificata con $E=0$ quando f e un polinomio di grado minore o uguale a tre.

<u>Risoluzione</u> (i) Poiché la funzione peso $\omega(x)$ e i nodi $-1,0,1$ sono simmetrici rispetto al centro dell'intervallo di integrazione, CN affinché gli A_i soddisfino la tesi (ovvero generino una formula di quadratura interpolatoria) è che il coefficiente A_2 sia uguale ad A_0. Dunque il problema si riduce a imporre che valga l'identità

$$\int_{-1}^{1} \frac{1}{1+x^2} f(x)\,dx = A_0(f(-1)+f(1)) + A_1 f(0) + E \tag{$*$}$$

con $E=0$ quando f è un qualsiasi polinomio di grado al più due. Data la linearità dell'integrale e della formula di quadratura, è sufficiente imporre $E=0$ in (*) per $f(x)=1,x,x^2$:

$$\begin{aligned}
&f(x)=1: \quad \tfrac{\pi}{2} = \tfrac{\pi}{4} - (-\tfrac{\pi}{4}) = [\arctan x]_{-1}^{1} = \textstyle\int_{-1}^{1} \tfrac{1}{1+x^2}dx = 2A_0 + A_1,\\
&f(x)=x: \quad 0=0,\\
&f(x)=x^2: \quad 2-\tfrac{\pi}{2} = \textstyle\int_{-1}^{1} \tfrac{x^2+1-1}{1+x^2}dx = \int_{-1}^{1} \tfrac{x^2}{1+x^2}dx = 2A_0.
\end{aligned}$$

Dunque $A_0 = 1-\pi/4$, $A_1 = \pi - 2$. La formula di quadratura trovata,

$$\int_{-1}^{1} \frac{1}{1+x^2} f(x)\,dx = (1-\frac{\pi}{4})(f(-1)+f(1)) + (\pi-2)f(0) + E, \tag{$**$}$$

è in realtà esatta (cioè $E = 0$) anche per $f(x) = x^3$ perché $\int_{-1}^{1} \frac{1}{1+x^2} f(x)\, dx$ e $(1 - \frac{\pi}{4})(f(-1) + f(1)) + (\pi - 2)f(0)$ sono entrambi nulli quando $f(x)$ è una funzione dispari in $[-1, 1]$ e quindi in particolare quando $f(x) = x^3$.

Per la linearità dell'integrale $\int_{-1}^{1} \frac{1}{1+x^2} f(x)\, dx$ e della formula di quadratura $(1 - \frac{\pi}{4})(f(-1) + f(1)) + (\pi - 2)f(0)$, in (**) si ha $E = 0$ ogni volta che f è un polinomio di grado al più tre. Dunque per $f(x) = x^3 - x^2$ l'equazione (**) vale con $E = 0$.

(ii) Il polinomio p_0 deve essere un polinomio costante monico, dunque $p_0(x) = 1$. Il polinomio p_1 deve essere un polinomio di grado uno monico, $p_1(x) = x + \beta$, tale che $\int_{-1}^{1} \frac{1}{1+x^2}(x + \beta)c\, dx = 0$, $\forall, c \in \mathbb{R}$. Questa condizione di ortogonalità implica $0 = \int_{-1}^{1} \frac{1}{1+x^2}(x + \beta)\, dx = \int_{-1}^{1} \frac{1}{1+x^2}\beta\, dx = \beta\pi/2$, $\beta = 0$. Quindi $p_1(x) = x$. Il polinomio p_2 deve essere un polinomio di grado due monico, $p_2(x) = x^2 + \alpha x + \beta$, tale che $\int_{-1}^{1} \frac{1}{1+x^2}(x^2 + \alpha x + \beta)1\, dx = 0$, $\int_{-1}^{1} \frac{1}{1+x^2}(x^2 + \alpha x + \beta)x\, dx = 0$ (soddisfatte queste due condizioni sarà anche vero che $\int_{-1}^{1} \frac{1}{1+x^2}(x^2 + \alpha x + \beta)q_1(x)\, dx = 0$ per ogni polinomio q_1 di grado al più uno). Queste due condizioni di ortogonalità diventano

$$\begin{aligned} 0 &= \textstyle\int_{-1}^{1} \frac{1}{1+x^2}(x^2 + \beta)\, dx = \int_{-1}^{1} \frac{1}{1+x^2}x^2\, dx + \beta \int_{-1}^{1} \frac{1}{1+x^2}\, dx = 2 - \frac{\pi}{2} + \beta\frac{\pi}{2}, \\ 0 &= \textstyle\int_{-1}^{1} \frac{1}{1+x^2}\alpha x^2\, dx = \alpha \int_{-1}^{1} \frac{1}{1+x^2}x^2\, dx = \alpha(2 - \frac{\pi}{2}), \end{aligned}$$

da cui $\alpha = 0$ e $\beta = 1 - 4/\pi$. Quindi $p_2(x) = x^2 + 1 - 4/\pi$.

(iii) Sia $\xi = \sqrt{-1 + 4/\pi}$. Poiché la funzione peso $\omega(x)$ e i nodi $-\xi, \xi$ sono simmetrici rispetto al centro dell'intervallo di integrazione, CN affinché gli A_i soddisfino la tesi (ovvero generino una formula di quadratura più che interpolatoria) è che il coefficiente A_1 sia uguale ad A_0. Dunque il problema si riduce a imporre che l'identità

$$\int_{-1}^{1} \frac{1}{1+x^2} f(x)\, dx = A_0(f(-\xi) + f(\xi)) + E \tag{*}$$

valga con $E = 0$ quando $f(x) = 1, x, x^2, x^3$:

$$\begin{aligned} f(x) &= 1: \quad \tfrac{\pi}{2} = 2A_0, \\ f(x) &= x: \quad 0 = 0, \\ f(x) &= x^2: \quad 2 - \tfrac{\pi}{2} = A_0(\tfrac{4}{\pi} - 1 + \tfrac{4}{\pi} - 1) = A_0(\tfrac{8}{\pi} - 2), \\ f(x) &= x^3: \quad 0 = 0. \end{aligned}$$

La scelta $A_0 = \pi/4$ verifica tutte le uguaglianze, dunque la formula di quadratura cercata è

$$\int_{-1}^{1} \frac{1}{1+x^2} f(x)\, dx = \frac{\pi}{4}(f\Big(-\sqrt{\frac{4}{\pi} - 1}\Big) + f\Big(\sqrt{\frac{4}{\pi} - 1}\Big)) + E. \tag{**}$$

Nota: essendo i nodi della formula di quadratura le radici $-\xi, \xi$ del polinomio di grado due ortogonale rispetto all'intervallo $[-1, 1]$ e al peso $\omega(x) = 1/(1+x^2)$ (vedi il punto (ii)), la formula di quadratura (**) si poteva ottenere imponendo l'identità (*) con $E = 0$ solo per $f(x) = 1, x$ (vedi la teoria sulle formule di quadratura gaussiane). Quindi parte dei calcoli di cui sopra si potevano evitare.

(iv) Scrivere due approssimazioni di $\mathcal{I} = \int_{-1}^{1} \frac{1}{(x^2+1)(x^2+3)} dx = \frac{1}{2}(\frac{1}{2} - \frac{\sqrt{3}}{9})\pi$ usando le due formule di quadratura ottenute nei punti (i) e (iii). (Facoltativo: confrontarle con $\mathcal{I}$).

Risoluzione

$$(i):\ \frac{5\pi-4}{24}=0.48783...,\qquad (iii):\ \frac{\pi^2}{2(4+2\pi)}=0.47989...,\qquad \mathcal{I}=0.483098...$$

$|(i)-\mathcal{I}| = 0.0047.. > 0.0032.. = |(iii)-\mathcal{I}| \Rightarrow$ la formula di quadratura gaussiana fornisce una approssimazione migliore.

Nota sull'Esercizio 1: Il metodo di Newton applicato a $f(x)=\alpha x-1$ converge in un passo.

Analisi Numerica 1 - II esonero - 10 Gennaio 2014

1 (i) Modificando il metodo di Newton per il calcolo degli zeri di $f(x)$, introdurre un metodo $x_{i+1} = g(x_i) = \varphi(x_i, f(x_i))$ che applicato nel caso $f(x) = \alpha x - 1$, $\alpha > 0$, richieda ad ogni passo due sole moltipicazioni senza divisioni e sia tale che

$$x_i = \frac{1}{\alpha} - \varepsilon \quad \Rightarrow \quad x_{i+1} = \frac{1}{\alpha} - \alpha\varepsilon^2.$$

(ii) Verificare che il metodo $x_{i+1} = x_i + x_i(1 - \alpha x_i) + x_i(1 - \alpha x_i)^2$ per approssimare $1/\alpha$ ha ordine di convergenza almeno tre.

2 Siano $f \in C^0([a,b])$, $h = (b-a)/n$, $n \in \mathbb{N}$, e $x_i = a + ih$, $i = 0, 1, \ldots, n$.

(i) Introdurre delle funzioni $\varphi_0(x), \varphi_1(x), \ldots, \varphi_{n-1}(x), \varphi_n(x)$ continue in $[a,b]$ che soddisfano le seguenti due condizioni:

a) $\varphi_i(x_j) = \delta_{ij}$, $0 \leq i, j \leq n$;

b) posto $g(x) = \sum_{i=0}^{n} f(x_i)\varphi_i(x)$, si ha che $g|_{[x_{i-1}, x_i]}$ è un polinomio di grado al più uno ($\forall\, i$). (Osservare che $g(x_j) = f(x_j)$, $j = 0, 1, \ldots, n$).

(ii) Nel caso $f \in C^2([a,b])$ trovare una limitazione superiore per $\max_{x \in [a,b]} |f(x) - g(x)|$ in funzione di h.

3 (i) Trovare i valori di A_i, $i = 0, 1, 2$, per cui l'identità

$$\int_{-1}^{1} \frac{1}{1+x^2} f(x)\, dx = A_0 f(-1) + A_1 f(0) + A_2 f(1) + E$$

è verificata con $E = 0$ quando f è un polinomio di grado minore o uguale a due. Per tali valori di A_i cosa si può dire su E quando $f(x) = x^3 - x^2$?

(ii) Scrivere i primi tre polinomi monici p_0, p_1, p_2 ortogonali rispetto all'intervallo $[-1, 1]$ e al peso $\omega(x) = 1/(1+x^2)$.

(iii) Dimostrare che esistono valori di A_0 e A_1 per cui l'identità

$$\int_{-1}^{1} \frac{1}{1+x^2} f(x)\, dx = A_0 f\Big(-\sqrt{\frac{4}{\pi} - 1}\Big) + A_1 f\Big(\sqrt{\frac{4}{\pi} - 1}\Big) + E$$

è verificata con $E = 0$ quando f e un polinomio di grado minore o uguale a tre.

Esercizio. Siano

$$M_\pm = \begin{bmatrix} \delta & -1 & 0 \\ 0 & -\delta & 1 \\ 1 & 0 & -\delta \end{bmatrix}, \ \delta > 1,$$

e $f(x) = \det(xI - M_\pm)$. (i) Osservare che f ha tre radici reali e distinte $\lambda_1 < \lambda_2 < \lambda_3$ e localizzarle (Nota: gli zeri di f sono gli autovalori di $M_\pm$). (ii) Proporre tre valori di $x_0 \in \mathbb{R}$, $x_0^{(1)}, x_0^{(2)}, x_0^{(3)}$ funzioni di δ, per cui le tre successioni di Newton corrispondenti $x_k^{(i)}$, $i = 1, 2, 3$,

$$x_0 = x_0^{(i)}, \ x_{k+1} = x_k - f(x_k)/f'(x_k), \ k = 0, 1, \ldots,$$

convergono in modo monotono rispettivamente a $\lambda_1, \lambda_2, \lambda_3$.

$$R = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \ M = \delta I - R = \begin{bmatrix} \delta & -1 & 0 \\ 0 & \delta & -1 \\ -1 & 0 & \delta \end{bmatrix}, \delta > 1, \ M_\pm = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} M = \begin{bmatrix} \delta & -1 & 0 \\ 0 & -\delta & 1 \\ 1 & 0 & -\delta \end{bmatrix},$$

$$M_\pm - \lambda I = \begin{bmatrix} \delta - \lambda & -1 & 0 \\ 0 & -\delta - \lambda & 1 \\ 1 & 0 & -\delta - \lambda \end{bmatrix}, \ \det(M_\pm - \lambda I) = (\delta - \lambda)(\delta + \lambda)^2 - 1,$$

quindi $p(\lambda) := p_{M_\pm}(\lambda) = -(\delta - \lambda)(\delta + \lambda)^2 + 1 = 1 - (\delta^2 - \lambda^2)(\delta + \lambda) = \lambda^3 + \lambda^2\delta - \lambda\delta^2 + 1 - \delta^3$. Per i primi due teoremi di Gershgorin, due autovalori di $M_\pm$ sono nel cerchio di centro $-\delta$ e raggio 1 e il terzo autovalore di $M_\pm$ è nel cerchio di centro δ e raggio 1 e deve essere reale. Inoltre si osserva che

$$\begin{aligned}
&p(-(1+\delta)) = -2\delta < -2, \\
&p(-\delta) = 1, \\
&p(-\delta + 1) = 2(1 - \delta) < 0, \\
&p(0) = 1 - \delta^3 < 0, \\
&p(\delta - 1) = 4\delta(1 - \delta) < 0, \\
&p(\delta) = 1, \\
&p(\delta + 1) = 4\delta(\delta + 1) + 2 > 10, \\
&p(\sqrt{\delta^2 - 1}) = -\sqrt{\delta^2 - 1} - \delta + 1 < 0, \\
&p(-\sqrt{\delta^2 - 1}) = \sqrt{\delta^2 - 1} - \delta + 1 > 0, \\
&p'(-\delta) = p'(\delta/3) = 0, \\
&p(\delta/3) = 1 - 32\delta^3/27 < 0, \\
&p''(-\delta/3) = 0, \\
&p(-\delta/3) = 1 - 16\delta^3/27, \ \text{Nota: } p(-\delta/3) = 0 \text{ se } \delta^3 = \tfrac{27}{16}.
\end{aligned}$$

(non è vero dunque in generale che $\pm\sqrt{\delta^2 - 1}$ sono autovalori di $M_\pm$ (era vero per $R = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$)).

Dunque $\lambda_1 \in (-\delta - 1, -\delta)$, $x_0^{(1)} = -\delta - 1$; $\lambda_2 \in (-\delta, -\delta + 1)$, $x_0^{(2)} = -\delta/3$; $\lambda_3 \in (\delta - 1, \delta)$, $x_0^{(3)} = \delta$.

Dall'ultima osservazione su $p(-\delta/3) = 0$ segue che

$$\begin{aligned}
\delta^3 = \tfrac{27}{16} \quad \Rightarrow \quad p(\lambda) &= p_{M_\pm}(\lambda) = \lambda^3 + \lambda^2\delta - \lambda\delta^2 + 1 - \delta^3 \\
&= (\lambda + \tfrac{\delta}{3})(\lambda^2 + \tfrac{2}{3}\delta\lambda - \tfrac{11}{9}\delta^2) = (\lambda + \tfrac{\delta}{3})(\lambda - \tfrac{2\sqrt{3}-1}{3}\delta)(\lambda + \tfrac{2\sqrt{3}+1}{3}\delta) \\
&= \lambda^3 + \tfrac{3}{2(2)^{1/3}}\lambda^2 - \tfrac{9}{4(4)^{1/3}}\lambda - \tfrac{11}{16}.
\end{aligned}$$

Esercizio. Sia $f(x) = \log_e x$ ed $\varepsilon > 0$ fissato. Siano x_i, $i = 0, 1, \dots, n$, punti distinti dell'intervallo $[\varepsilon, 1]$ e p_n il polinomio di grado al più n tale che $p_n(x_i) = f(x_i)$, $i = 0, 1, \dots, n$. (i) Scrivere una espressione esplicita per $f(x) - p_n(x)$, valida per $x > 0$, e ricavare una maggiorazione in termini di ε della quantità $\max_{x\in[\varepsilon,1]} |f(x) - p_n(x)|$ commentando il risultato. (ii) Calcolare la retta minmax di f in $[\frac{1}{e}, 1]$.

(ii) Sia ε, $0 < \varepsilon < 1$, fissato. Cercare il polinomio $p_1(x)$ di grado al più uno tale che

$$\gamma = \max_{x\in[\varepsilon,1]} |\log_e x - p_1(x)| = \min\{\max_{x\in[\varepsilon,1]} |\log_e x - q_1(x)| : \ q_1 \text{ polinomi }, \partial q_1 \le 1\}.$$

Scriverlo nel caso $\varepsilon = 1/e$.

Imporre le identità $f(\varepsilon) - p_1(\varepsilon) = f(1) - p_1(1) = p_1(\xi) - f(\xi)$ e massimizzare $|p_1(\xi) - f(\xi)|$:

$$\log_e \varepsilon - (a\varepsilon + b) = \log_e 1 - (a + b) = (a\xi + b) - \log_e \xi =: \varphi(\xi),$$

$\varphi'(\xi) = a - 1/\xi = 0$ se $\xi = 1/a$. La prima uguaglianza implica $a = -\log_e \varepsilon/(1 - \varepsilon)$, dunque $\xi = -(1 - \varepsilon)/\log_e \varepsilon$. La seconda uguaglianza implica (essendo $a\xi = 1$)

$$-a - b = 1 + b - \log_e \xi, \quad b = \frac{1}{2}(-a - 1 + \log_e \xi) = \frac{1}{2}(\frac{\log_e \varepsilon}{1 - \varepsilon} - 1 + \log_e \xi).$$

Dunque

$$p_1(x) = -\frac{\log_e \varepsilon}{1 - \varepsilon}x + \frac{1}{2}(\frac{\log_e \varepsilon}{1 - \varepsilon} - 1 + \log_e \xi), \ \xi = -\frac{1 - \varepsilon}{\log_e \varepsilon}.$$

Nel caso $\varepsilon = 1/e$ ($\xi = -\frac{e-1}{-e}$, $\log_e \xi = \log_e(e - 1) - 1$) si ha

$$p_1(x) = \frac{e}{e - 1}x + \frac{1}{2}(-\frac{e}{e - 1} - 2 + \log_e(e - 1)),$$

$$\begin{aligned} &\gamma = \max_{x\in[\frac{1}{e},1]} |\log_e x - p_1(x)| \\ &= p_1(\tfrac{1}{e}) - \log_e \tfrac{1}{e} = \log_e \tfrac{e-1}{e} - p_1(\tfrac{e-1}{e}) = p_1(1) - \log_e 1 \\ &= \tfrac{1}{e-1} + \tfrac{1}{2}(\tfrac{-e}{e-1} - 2 + \log_e(e - 1)) - \log_e \tfrac{1}{e} = 0.0616... \end{aligned}$$

Esercizio. Sia $p(x) = a_0x^n + \dots$ un polinomio di grado maggiore o uguale di due con coefficienti reali e $a_0 > 0$. Se tutte le radici λ_i sono reali e $\lambda_1 \ge \lambda_2 \ge \dots \ge \lambda_n$, allora Newton genera una successione monotona decrescente che converge a λ_1 per ogni punto iniziale $x_0 > \lambda_1$.

Risoluzione

Gli $n - 1$ zeri μ_i di p' sono tali che $\lambda_{i+1} \le \mu_i \le \lambda_i$, $i = 1, \dots, n - 1$, dunque sono minori di λ_1. Ne segue che le funzioni p e p' dopo λ_1 sono entrambe positive, cioè, se $x_i > \lambda_1$ allora

$$x_{i+1} = x_i - \frac{p(x_i)}{p'(x_i)} < x_i.$$

Inoltre, posto $g(x) = x - p(x)/p'(x)$, poiché $g'(x) = p(x)p''(x)/(p'(x))^2$ e $p'' > 0$, si ha che $g'(x) > 0$ dopo λ_1, dunque la successione x_i non può essere alternata ($x_{i+1} - \lambda_1 = g(x_i) - g(\lambda_1) = g'(\eta_i)(x_i - \lambda_1)$, $\eta_i \in (\lambda_1, x_i)$). Ciò implica $\lambda_1 \le x_{i+1} < x_i$ per ogni i, quindi x_i converge in modo monotono a un $\beta \ge \lambda_1$ tale che $p(\beta) = 0$ (per il teorema ponte) e tale β deve essere necessariamente λ_1 (in caso contrario p avrebbe più di n zeri).

Risoluzione di un esercizio dato a uno scritto di AN12012-2013 e un Problema

Sia $p_1(x) = ax + b$ il polinomio di grado al più uno tale che

$$\max_{x\in[0,1]} |(-x^2 + 2x) - p_1(x)| = \min_{g\,\partial g\leq 1} \max_{x\in[0,1]} |(-x^2 + 2x) - g(x)|.$$

Troviamolo.

I procedimento (occorre conoscere l'enunciato del teorema di Chebycev):
Sia $\gamma = \max_{x\in[0,1]} |(-x^2 + 2x) - p_1(x)|$. Per il teorema di Chebycev devono esistere almeno tre punti ξ_1, ξ_2, ξ_3 tali che

$$0 \leq \xi_1 < \xi_2 < \xi_3 \leq 1, \quad f(\xi_i) - p_1(\xi_i) = (-1)^i\gamma \text{ (oppure } (-1)^{i-1}\gamma)$$

(si è posto $f(x) = -x^2 + 2x$). Quindi:
ξ_2 deve essere punto di max (min) per $f - p_1$;
se $\xi_1 > 0$, allora ξ_1 deve essere punto di min (max) per $f - p_1$;
se $\xi_3 < 1$, allora ξ_3 deve essere punto di min (max) per $f - p_1$.
Ma $f - p_1$ ha un solo punto estremale, $f'(x) - p_1'(x) = -2x + 2 - a$, cioè $1 - a/2$. Dunque deve essere $\xi_1 = 0$, $\xi_2 = 1 - a/2$, $\xi_3 = 1$.
Inoltre, dalla condizione $f(0) - p_1(0) = f(1) - p_1(1)$, $-b = 1 - a - b$, segue che $a = 1$ e, quindi, $\xi_2 = 1/2$. Dalla condizione $f(0) - p_1(0) = p_1(\xi_2) - f(\xi_2)$, $-b = p_1(\frac{1}{2}) - f(\frac{1}{2})$, segue che $-b = (\frac{1}{2} + b) - \frac{3}{4}$, da cui $b = 1/8$.
Dunque, $p_1(x) = x + 1/8$.

II procedimento
(occorre sapere che $T_n(x)/2^{n-1}$ è il polinomio soluzione del problema $\min_{p \text{ monici } \partial p=n} \max_{x\in[-1,1]} |p(x)|$):

$$** := \max_{x\in[0,1]} | - x^2 + 2x - q_1(x)| = \max_{t\in[-1,1]} | - \tfrac{1}{4}(t^2 + 2t + 1) + t + 1 - \tilde{q}_1(t)|$$
$$= \tfrac{1}{4} \max_{t\in[-1,1]} |t^2 + 2t + 1 - 4t - 4 + 4\tilde{q}_1(t)| = \tfrac{1}{4} \max_{t\in[-1,1]} |t^2 - 2t - 3 + 4\tilde{q}_1(t)| =: *$$

($x = \frac{1}{2}(t + 1)$). Nota: in $[-1, 1]$ si approssima con un polinomio di primo grado un polinomio di secondo grado diverso da $-t^2 + 2t$ (vedi Torti).
Ora, per quanto visto a lezione, * è minimo se $t^2 - 2t - 3 + 4\tilde{q}_1(t) = \frac{1}{2}T_2(t) = \frac{1}{2}(2t^2 - 1)$, cioè se $\tilde{q}_1(t) = \frac{t}{2} + \frac{5}{8}$, e quindi, essendo $t = 2x - 1$, ** è minimo per $q_1(x) = x - \frac{1}{2} + \frac{5}{8} = x + 1/8$.

III procedimento
Si osserva che

$$** := \max_{x\in[0,1]} |(-x^2 + 2x) - q_1(x)| = \max_{t\in[-1,1]} |(-\frac{1}{4}t^2 + \frac{1}{2}t + \frac{3}{4}) - \tilde{q}_1(t)| =: *$$

($x = \frac{1}{2}(t + 1)$). Affinché * sia minimo (vedi pp.54–55 del file AnalisiNumerica1.pdf oppure il Procedimento generale più sotto), occorre scegliere

$$\tilde{q}_1(t) = (-\frac{1}{4}t^2 + \frac{1}{2}t + \frac{3}{4}) - \frac{-1/4}{2}T_2(t) = \frac{1}{2}t + \frac{5}{8}$$

$(T_2(t) = 2t^2 - 1)$. Quindi, affinché ** sia minimo, occorre scegliere $q_1(x) = \frac{1}{2}(2x-1) + \frac{5}{8} = x + 1/8$.

Confronta i grafici di $-x^2 + 2x$ e $x + 1/8$ nell'intervallo $[0, 1]$.

Procedimento generale (in $[-1, 1]$):

Sia $p_n(x) = a_n x^n + \dots$, $a_n \neq 0$. Si vuole minimizzare la quantità $\max_{x\in[-1,1]} |p_n(x) - q(x)|$ al variare di q tra tutti i polinomi di grado al più $n - 1$:

$$\max_{x\in[-1,1]} |p_n(x) - q(x)| = \max_{x\in[-1,1]} |a_n(\frac{1}{a_n}p_n(x) - \frac{1}{a_n}q(x))| = |a_n| \max_{x\in[-1,1]} |\frac{1}{a_n}p_n(x) - \frac{1}{a_n}q(x)|,$$

$$\min_q \max_{x\in[-1,1]} |p_n(x) - q(x)| = |a_n| \min_q \max_{x\in[-1,1]} |\frac{1}{a_n}p_n(x) - \frac{1}{a_n}q(x)| = |a_n|\frac{1}{2^{n-1}}$$

$(\frac{1}{a_n}p_n(x) - \frac{1}{a_n}q(x)$ è un generico polinomio monico di grado n; il minimo è assunto per q tale che $\frac{1}{a_n}p_n(x) - \frac{1}{a_n}q(x) = \frac{1}{2^{n-1}}T_n(x))$.

Dunque, il minimo è assunto per $q(x) = p_n(x) - \frac{a_n}{2^{n-1}}T_n(x)$.

Problema (economizzazione)

Dato un polinomio $p_n(x) = \sum_{i=0}^n a_i x^i$ con $a_n \neq 0$, si vuole determinare il polinomio p_m di grado al più m, $m < n$, per cui

$$\max_{x\in[-1,1]} |\sum_{i=0}^n a_i x^i - p_m(x)| = \min\{\max_{x\in[-1,1]} |\sum_{i=0}^n a_i x^i - q_m(x)| : \ q_m \text{ polinomio}, \ \partial q_m \leq m\}.$$

Risoluzione. È sufficiente trovare i c_i per cui $\sum_{i=0}^n a_i x^i = \sum_{i=0}^n c_i T_i(x)$, dove T_i è l'i-esimo polinomio di Chebycev, e porre $p_m(x) = \sum_{i=0}^m c_i T_i(x)$. Perché?

Equivalentemente, è sufficiente definire successivamente i seguenti polinomi:

$p_{n-1}(x) := p_n(x) - \frac{a_n}{2^{n-1}}T_n(x) = a_{n-1}^{(1)}x^{n-1} + \dots$

$p_{n-2}(x) := p_{n-1}(x) - \frac{a_{n-1}^{(1)}}{2^{n-2}}T_{n-1}(x) = a_{n-2}^{(2)}x^{n-2} + \dots$

$\dots$

All'$(n - m)$-esimo passo, si otterrà $p_m(x)$. Perché?

(1) Sia A $n \times n$ reale definita positiva, e si vuole risolvere il sistema $A\mathbf{x} = \mathbf{b}$. Sia M un'altra matrice definita positiva. Indicare una scelta di M diversa da $M = I$ per cui il metodo: $\mathbf{x}_0 \in \mathbb{R}^n$,

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \omega(\mathbf{b} - A\mathbf{x}_k),$$
$$\omega \text{ tale che } \|\mathbf{b} - A\mathbf{x}_{k+1}\|_M \text{ minimo},$$

$k = 0, 1, 2, \ldots$, sia convergente ($\|\mathbf{v}\|_M = \sqrt{\mathbf{v}^T M \mathbf{v}}$).

(2) Sia A $n \times n$ reale e $A_S = \frac{1}{2}(A + A^T)$. Dimostrare che

$$\lambda \text{ autovalore di } A \;\Rightarrow\; Re(\lambda) \geq \lambda_{\min}(A_S) \;\Rightarrow\; \frac{\lambda_{\min}(A_S)^2}{\lambda_{\max}(A^T A)} \leq 1$$

Ingegneria Civile

1) Utilizzando il metodo di Givens trasformare il vettore 3×1 $\mathbf{w} = [1/\sqrt{2}\ 1/2\ 1/2]^T$ nel vettore $\mathbf{e}_1 = [1\ 0\ 0]^T$

2) Sia A la matrice 4×4 i cui elementi sono tutti uguali a 1 ad eccezione di quelli diagonali per cui $a_{11} = 4$, $a_{i+1,i+1} = a_{i,i} + i$. Dimostrare che A è non singolare e dare una limitazione superiore per $\mu_2(A) = \|A\|_2 \|A^{-1}\|_2$.

3) Dimostrare che per ogni matrice $A \in \mathbb{C}^{n \times n}$ si ha $\|A\|_2^2 \leq \|A\|_1 \|A\|_\infty$

4) Calcolare il numero di valutazioni di $f(x) = 1/(x+1)$ richieste dalla formula del trapezio composta per poter fornire una approssimazione S di $I = \int_0^1 f(x)dx$ tale che $|S - I| < 10^{-3}$

Scritto di Calcolo Numerico 1 (prof. Giuseppe Rodriguez)
25 febbraio 2003

1. Fornire un esempio in cui la cancellazione produca un elevato errore relativo nel risultato di una somma algebrica.

2. Sia data la matrice
$$A = \begin{bmatrix} 2\alpha & -1 & 0 \\ -1 & 2\alpha & -1 \\ 0 & -1 & 2\alpha \end{bmatrix}.$$
Dire per quali valori del parametro α la matrice A risulta non singolare, per quali è definita positiva e per quali il metodo di Jacobi applicato al sistema lineare $A\mathbf{x} = \mathbf{b}$ risulta convergente.

   Fissato $\alpha = 2$ si calcolino le prime due iterate $\mathbf{x}^{(1)}$ e $\mathbf{x}^{(2)}$ del metodo di Jacobi in corrispondenza al termine noto e al vettore iniziale
$$\mathbf{b} = (2, 4, 10)^T, \quad \mathbf{x}^{(0)} = (0, 0, 0)^T.$$

3. Risolvere il sistema lineare
$$\begin{cases} -x_1 + x_3 - \frac{1}{2}x_4 & = 0 \\ x_2 + \frac{1}{2}x_4 & = 4 \\ 2x_1 + x_4 & = 6 \\ 2x_2 + 2x_3 & = 10 \end{cases}$$
mediante l'algoritmo di Gauss con pivoting.

4. Si consideri il problema di Cauchy
$$\{\, y'(t) = -2xy(t)y(0) = 1.$$
Dire se esso ammette una e una sola soluzione e approssimare la sua soluzione nei punti $x_1 = \frac{1}{2}$, $x_2 = 1$ e $x_3 = \frac{3}{2}$ mediante il metodo alle differenze finite di Heun con punto iniziale $x_0 = 0$ e passo $h = \frac{1}{2}$.

Su argomenti di Analisi Numerica 1, prima di Natale 2013

Parabola $p(x)$ che meglio approssima $f(x) = |x|$ in $[-1, 1]$ nel senso minmax.

Essendo f simmetrica rispetto al centro dell'intervallo $[-1, 1]$, anche la parabola p deve essere simmetrica rispetto al centro di tale intervallo, cioè $p(x) = a + bx^2$. Per il teorema di Chebycev esistono almeno quattro punti $x_0 < x_1 < x_2 < x_3$ in $[-1, 1]$ dove l'errore $p(x) - f(x)$ in modulo vale $\gamma = \|p - f\|_{\infty,[-1,1]}$ e $p(x_i) - f(x_i) = (-1)^i\gamma$ (oppure $p(x_i) - f(x_i) = (-1)^{i+1}\gamma$). Da un'osservazione del grafico di $|x|$ in $[-1, 1]$ e di come potrà essere il grafico della parabola p che meglio approssima $|x|$ in $[-1, 1]$ nel senso minmax, si intuisce che i punti in realtà sono cinque, $-1, -\xi, 0, \xi, 1$, con $\xi \in (0, 1)$, e che dunque deve esistere $\xi \in (0, 1)$ tale che

$$\gamma = p(1) - f(1) = f(\xi) - p(\xi) = p(0) - f(0) \quad \text{se e solo se} \quad \gamma = (a + b) - 1 = \xi - (a + b\xi^2) = a.$$

Da tali condizioni segue subito che deve essere $b = 1$, e quindi devono essere verificate contemporaneamente le seguenti tre condizioni: (1) $\xi \in (0, 1)$; (2) $\xi-(a+\xi^2) = a$; (3) $f(\xi)-p(\xi) = \xi-(a+\xi^2)$ è massima. La funzione $\xi - (a + \xi^2)$ è ovviamente massima per $\xi = 1/2$. Andando a sostituire ξ con $1/2$ nella (2) si ottiene il valore di a: $a = 1/8$. Dunque la parabola cercata è $p(x) = x^2 + 1/8$ e $\gamma = 1/8$.

Sia $T_r(x)$ l'r-esimo polinomio di Chebycev. Poiché

$$\int_{-1}^{1} \frac{T_r(x)}{\sqrt{1-x^2}}dx = \int_{-\pi}^{0} \cos(ry)\, dy = \begin{cases} \pi & r = 0 \\ 0 & r = 1, 2, 3, \dots \end{cases}$$

($T_r(x)|_{[-1,1]} = \cos(r \arccos x)$, $y = \arccos x$, $dy = \frac{1}{\sqrt{1-x^2}}dx$), si ha

$$\int_{-1}^{1} \frac{T_i(x)T_j(x)}{\sqrt{1-x^2}}dx = \frac{1}{2}\int_{-1}^{1} \frac{T_{i+j}(x)}{\sqrt{1-x^2}}dx + \frac{1}{2}\int_{-1}^{1} \frac{T_{|i-j|}(x)}{\sqrt{1-x^2}}dx = \begin{cases} 0 & i \neq j \\ \pi/2 & i = j \neq 0 \\ \pi & i = j = 0 \end{cases}.$$

In particolare,

$$\int_{-1}^{1} \frac{T_i(x)^2}{\sqrt{1-x^2}}dx = \begin{cases} \pi/2 & i \neq 0 \\ \pi & i = 0 \end{cases}.$$

Applicazione. Dalla teoria e da quanto appena visto segue che

$$\min_{q_n,\, \partial q_n \leq n} \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}}(f(x) - q_n(x))^2\, dx = \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}}(f(x) - \sum_{i=0}^{n} a_i T_i(x))^2\, dx$$

dove

$$a_i = \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} f(x)T_i(x)\, dx \Big/ \int_{-1}^{1} \frac{T_i(x)^2}{\sqrt{1-x^2}}dx.$$

Esempio. Cerchiamo la parabola $p_2(x)$ che meglio approssima $f(x) = \sqrt{1-x^2}$ nel senso dei minimi quadrati nell'intervallo $[-1, 1]$ e rispetto al peso $\omega(x) = 1/\sqrt{1-x^2}$.

$f(x) = \sqrt{1-x^2} \Rightarrow a_0 = \frac{1}{\pi}\int_{-1}^1 T_0(x)\,dx = 2/\pi$ e $a_i = \frac{2}{\pi}\int_{-1}^1 T_i(x)\,dx,\ i = 1,2,\ldots$. Dunque la parabola cercata è

$$p_2(x) = \frac{2}{\pi}T_0(x) + T_2(x)\frac{2}{\pi}\int_{-1}^1 T_2(x)\,dx = \frac{2}{\pi} + \frac{2}{\pi}(2[\frac{x^3}{3}]_{-1}^1 - [x]_{-1}^1)T_2(x) = \frac{2}{3\pi}(5-4x^2).$$

Dalla teoria segue inoltre che

$$\min_{q_n,\,\partial q_n \le n} \int_{-1}^1 (f(x) - q_n(x))^2\,dx = \int_{-1}^1 (f(x) - \sum_{i=0}^n a_i l_i(x))^2\,dx$$

dove

$$a_i = \int_{-1}^1 f(x) l_i(x)\,dx \Big/ \int_{-1}^1 l_i(x)^2\,dx$$

a condizione che $l_i(x)$ sono ortogonali nell'intervallo $[-1,1]$ rispetto al peso $\omega(x) = 1$.

Esempio. parabola $p_2(x)$ che meglio approssima $f(x) = \sqrt{1-x^2}$ nel senso dei minimi quadrati nell'intervallo $[-1,1]$, rispetto al peso $\omega(x) = 1$.

I primi tre polinomi monici ortogonali nell'intervallo $[-1,1]$, rispetto al peso $\omega(x) = 1$ sono $l_0(x) = 1$, $l_1(x) = x$, $l_2(x) = x^2 - 1/3$ (dimostrarlo!).

$f(x) = \sqrt{1-x^2} \Rightarrow$

$$a_0 = \frac{1}{2}\int_{-1}^1 \sqrt{1-x^2}dx = \frac{\pi}{4}, \quad a_1 = \int_{-1}^1 \sqrt{1-x^2}x\,dx \Big/ \int_{-1}^1 x^2 dx = 0,$$

$$a_2 = \int_{-1}^1 \sqrt{1-x^2}(x^2-1/3)\,dx \Big/ \int_{-1}^1 (x^2-1/3)^2 dx = \frac{\pi/8 - (1/3)(\pi/2)}{8/45} = -\frac{45}{24\cdot 8}\pi.$$

Dunque la parabola cercata è

$$p_2(x) = a_0 l_0(x) + a_1 l_1(x) + a_2 l_2(x) = \frac{\pi}{4} - \frac{1}{24}\frac{45}{8}\pi(x^2 - \frac{1}{3}).$$

Nota: si è usata l'identità

$$\int_{-1}^1 (\sqrt{1-x^2}x)x\,dx = [-\frac{1}{3}(1-x^2)\sqrt{1-x^2}x]_{-1}^1 - \int_{-1}^1 (-\frac{1}{3})(1-x^2)\sqrt{1-x^2}dx = \frac{1}{3}\int_{-1}^1 \sqrt{1-x^2}dx - \frac{1}{3}\int_{-1}^1 x^2\sqrt{1-x^2}dx.$$

Nota: i valori $p_2(0) = 21\pi/64$, $p_2(1) = 3\pi/32$ sono rispettivamente inferiore e superiore rispetto ai valori $10/(3\pi)$ e $2/(3\pi)$ assunti dalla parabola $p_2^{(c)}$ appross. Min. Qu. di f in $[-1,1]$ rispetto al peso $1/\sqrt{1-x^2}$ ($p_2^{(c)}$ è stata trovata sopra). Più precisamente si osserva che, come ci si aspetta, $p_2^{(c)}$ è più (meno) precisa di p_2 agli estremi (al centro) dell'intervallo $[-1,1]$:

$$\begin{array}{ll} y(0) = 1, & y(1) = 0, \\ p_2^{(c)}(0) = 1.0615.., & p_2^{(c)}(1) = 0.2123.., \\ p_2(0) = 1.03.., & p_2(1) = 0.2943..\ . \end{array}$$

Matrice Minimi Quadrati - Caso discreto

Sulla matrice $A_{ij} = \sum_{r=0}^{m} \varphi_i(x_r)\varphi_j(x_r)$, $i, j = 0, \ldots, n$, dove $m \geq n$, x_r sono numeri reali distinti, e φ_j sono funzioni linearmente indipendenti.

Posto $\mathbf{a} = [a_0\ a_1\ \cdots\ a_n]^T$, si ha

$$\mathbf{a}^T A\mathbf{a} = \sum_{i,j=0}^{n} a_i A_{ij} a_j = \sum_{r=0}^{m} \sum_{i} a_i \varphi_i(x_r) \sum_{j} a_j \varphi_j(x_r) = \sum_{r=0}^{m} (\sum_{j} a_j \varphi_j(x_r))^2.$$

È evidente che A è reale simmetrica ($A_{ij} = A_{ji} \in \mathbb{R}$) e semi definita positiva ($\mathbf{a}^T A \geq 0$, $\forall\, \mathbf{a} \in \mathbb{R}^{n+1}$).

Quando A è definita positiva?

La condizione $\mathbf{a}^T A\mathbf{a} = 0$ implica $\sum_{r=0}^{m} (\sum_j a_j \varphi_j(x_r))^2 = 0$, $\sum_j a_j \varphi_j(x_r) = 0$, $r = 0, 1, \ldots, m$. Quindi, ci si chiede se è vera l'implicazione

$$\begin{bmatrix} \varphi_0(x_0) & \varphi_1(x_0) & \cdot & \varphi_n(x_0) \\ \varphi_0(x_1) & \varphi_1(x_1) & \cdot & \varphi_n(x_1) \\ \cdot & \cdot & \cdot & \cdot \\ \varphi_0(x_m) & \varphi_1(x_m) & \cdot & \varphi_n(x_m) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \cdot \\ a_n \end{bmatrix} = \mathbf{0} \quad \Rightarrow \quad a_j = 0\ \forall\, j. \qquad \text{(fatto)}$$

(Nota: la condizione di lineare indipendenza delle φ_j è necessaria affinché il (fatto) possa essere vero).

Risultato 1: se le φ_j sono polinomi di grado al più n che generano lo spazio dei polinomi di grado al più n, allora il (fatto) è vero.

Dimostrazione. Supponiamo $\sum_j a_j \varphi_j(x_r) = 0$, $r = 0, 1, \ldots, m$. Allora $\sum_j a_j \varphi_j$ sarebbe un polinomio di grado al più n che si annulla in $m + 1$ numeri reali distinti (dove, ricordiamo, $m \geq n$). Dunque $\sum_j a_j \varphi_j$ deve essere il polinomio identicamente nullo e, come conseguenza della lineare indipendenza delle φ_j, gli a_j devono essere tutti nulli.

Nota. Come conseguenza del Risultato 1, comunque si scelgano distinti i nodi x_r, $r = 0, 1, \ldots, m$, fissato n con $n \leq m$, è univocamente definito il polinomio di grado minore o uguale a n di migliore approssimazione nel senso dei minimi quadrati della tabella $(x_r, f(x_r))$, $r = 0, 1, \ldots, m$.

Osservando la dimostrazione del Risultato 1, si capisce che vale il seguente Risultato più forte:

Risultato 2: se le φ_j sono polinomi linearmente indipendenti nello spazio dei polinomi di grado al più m, allora il (fatto) è vero.

Dimostrazione. Supponiamo $\sum_j a_j \varphi_j(x_r) = 0$, $r = 0, 1, \ldots, m$. Allora $\sum_j a_j \varphi_j$ sarebbe un polinomio di grado al più m che si annulla in $m + 1$ numeri reali distinti. Dunque $\sum_j a_j \varphi_j$ deve essere il polinomio identicamente nullo e, come conseguenza della lineare indipendenza delle φ_j, gli a_j devono essere tutti nulli.

Il (fatto) non è vero in generale. Ad esempio non è vero se le funzioni φ_i e φ_j, pur essendo indipendenti, nei punti x_k verificano le identità $\varphi_i(x_k) = C\varphi_j(x_k)\ \forall\, k$, per qualche costante C, e in particolare se $\varphi_i(x_k) = 0\ \forall\, k$. In generale, date $n+1$ funzioni φ_j indipendenti, per poter essere vero il (fatto), gli $m + 1$ punti x_r non sempre possono essere scelti a caso. Vediamolo su due esempi.

Esempio ($m = n = 1$): $\varphi_0(x) = x^2$, $\varphi_1(x) = e^{-x^2}$, $x_1 = -x_0$, $x_0 \neq 0$. È evidente che la seconda colonna della matrice

$$G = \begin{bmatrix} \varphi_0(x_0) & \varphi_1(x_0) \\ \varphi_0(x_1) & \varphi_1(x_1) \end{bmatrix}$$

deve essere un multiplo della prima colonna. Invece, se $x_1 \neq -x_0$, la matrice G può essere non singolare.

Esempio ($m = n = 1$): $\varphi_0(x) = x$, $\varphi_1(x) = e^{-x^2}$, $x_1 \neq x_0$. In questo caso la matrice

$$G = \begin{bmatrix} \varphi_0(x_0) & \varphi_1(x_0) \\ \varphi_0(x_1) & \varphi_1(x_1) \end{bmatrix} = \begin{bmatrix} x_0 & e^{-x_0^2} \\ x_1 & e^{-x_1^2} \end{bmatrix}$$

è sempre invertibile. Infatti, poiché $\det G = x_0 e^{-x_1^2} - x_1 e^{-x_0^2}$, la matrice G è sicuramente invertibile se x_1 e x_0 sono opposti in segno. Se invece $x_0 = \alpha x_1$ con $\alpha > 0$, allora la condizione $x_0 e^{-x_1^2} = x_1 e^{-x_0^2}$ è equivalente alla condizione $\alpha = e^{(1-\alpha^2)x_1^2}$ che è verificata solo nel caso $\alpha = 1$.

...

Dunque l'affermazione a pag.30 del file AnalisiNumerica1.pdf va rivista perché non è vera per ogni tipo di spazio di funzioni lineare.

<u>Matrice Minimi Quadrati - Caso continuo</u>

Sulla matrice $A_{ij} = \int_a^b \omega(x)\varphi_i(x)\varphi_j(x)\,dx$, $i,j = 0,\dots,n$, dove $\omega > 0$ in (a,b) e le φ_j sono funzioni linearmente indipendenti.

Posto $\mathbf{a} = [a_0\ a_1\ \cdots\ a_n]^T$, si ha

$$\mathbf{a}^T A \mathbf{a} = \sum_{i,j=0}^{n} a_i A_{ij} a_j = \int_a^b \omega(x) \sum_i a_i \varphi_i(x) \sum_j a_j \varphi_j(x)\,dx = \int_a^b \omega(x) (\sum_j a_j \varphi_j(x))^2\,dx.$$

È evidente che A è reale simmetrica ($A_{ij} = A_{ji} \in \mathbb{R}$) e semi definita positiva ($\mathbf{a}^T A \mathbf{a} \geq 0$, $\forall\,\mathbf{a} \in \mathbb{R}^{n+1}$). Inoltre, A è definita positiva perché la condizione $\mathbf{a}^T A \mathbf{a} = 0$ implica $\int_a^b \omega(x)(\sum_j a_j\varphi_j(x))^2\,dx = 0$, $\sum_j a_j \varphi_j(x) = 0$, $a_j = 0\ \forall\, j$ (per la lineare indipendenza delle φ_j).

Sia $A \in \mathbb{C}^{n\times n}$. Dimostrare che esistono matrici unitarie Q_i tali che

$$\operatorname{diag}(a_{ii},\, i=1,\dots,n) = \frac{1}{2^{n-1}} \sum_{i=1}^{2^{n-1}} Q_i A Q_i^H .$$

Dunque, se $\|\ \|$ è una norma matriciale invariante per trasformazioni unitarie, si ha

$$\|\operatorname{diag}(a_{ii},\, i=1,\dots,n)\| \le \|A\|$$

(risultato utilizzato nell'articolo di Jin e altri).

Caso 4×4. $A_0 := A$,

$$A_1 := \begin{bmatrix} a_{11} & 0 & 0 & 0 \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & a_{32} & a_{33} & a_{34} \\ 0 & a_{42} & a_{43} & a_{44} \end{bmatrix} = \frac{1}{2}(A_0 + \begin{bmatrix} 1 & 0 \\ 0 & -I_3 \end{bmatrix} A_0 \begin{bmatrix} 1 & 0 \\ 0 & -I_3 \end{bmatrix}),$$

$$A_2 := \begin{bmatrix} a_{11} & 0 & 0 & 0 \\ 0 & a_{22} & 0 & 0 \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & a_{43} & a_{44} \end{bmatrix} = \frac{1}{2}(A_1 + \begin{bmatrix} I_2 & 0 \\ 0 & -I_2 \end{bmatrix} A_1 \begin{bmatrix} I_2 & 0 \\ 0 & -I_2 \end{bmatrix}),$$

$$A_3 := \begin{bmatrix} a_{11} & 0 & 0 & 0 \\ 0 & a_{22} & 0 & 0 \\ 0 & 0 & a_{33} & 0 \\ 0 & 0 & 0 & a_{44} \end{bmatrix} = \frac{1}{2}(A_2 + \begin{bmatrix} I_3 & 0 \\ 0 & -1 \end{bmatrix} A_2 \begin{bmatrix} I_3 & 0 \\ 0 & -1 \end{bmatrix}).$$

Caso $n\times n$. $A_0 = A$; per $i = 1,\dots,n-1$

$$\begin{aligned} A_i \ &:= \begin{bmatrix} a_{11} & & & & \\ & \cdot & & & \\ & & a_{ii} & & \\ & & & a_{i+1,i+1} & \cdot & a_{i+1,n} \\ & & & \cdot & & \cdot \\ & & & a_{n,i+1} & \cdot & a_{n,n} \end{bmatrix} \\ &= \frac{1}{2}(A_{i-1} + \begin{bmatrix} I_i & O \\ O & -I_{n-i} \end{bmatrix} A_{i-1} \begin{bmatrix} I_i & O \\ O & -I_{n-i} \end{bmatrix}). \end{aligned}$$

Dunque $A_{n-1} := \operatorname{diag}(a_{11},\dots,a_{nn}) = \frac{1}{2^{n-1}} \sum_{i=1}^{2^{n-1}} Q_i A Q_i^H$, dove $Q_i = \operatorname{diag}(\mathbf{z}_i)$ con

$$\begin{aligned} \mathbf{z}_1 &= \mathbf{e}, \\ \mathbf{z}_2 &= \begin{bmatrix} 1 & O \\ O & -I_{n-1} \end{bmatrix} \mathbf{z}_1, \\ \mathbf{z}_3 &= \begin{bmatrix} I_2 & O \\ O & -I_{n-2} \end{bmatrix} \mathbf{z}_1,\ \mathbf{z}_4 = \begin{bmatrix} I_2 & O \\ O & -I_{n-2} \end{bmatrix} \mathbf{z}_2, \\ \mathbf{z}_5 &= \begin{bmatrix} I_3 & O \\ O & -I_{n-3} \end{bmatrix} \mathbf{z}_1,\ \mathbf{z}_6 = \begin{bmatrix} I_3 & O \\ O & -I_{n-3} \end{bmatrix} \mathbf{z}_2,\ \mathbf{z}_7 = \begin{bmatrix} I_3 & O \\ O & -I_{n-3} \end{bmatrix} \mathbf{z}_3,\ \mathbf{z}_8 = \begin{bmatrix} I_3 & O \\ O & -I_{n-3} \end{bmatrix} \mathbf{z}_4, \\ \dots & \ \dots\ \dots \end{aligned}$$

Risposta alla (5): No, $\begin{bmatrix} -10 & -1 \\ 5 & 1 \end{bmatrix}$ (questo esempio lo ha suggerito uno studente)

A Jalil che mi ha chiesto: perché se

$$A = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

allora $\rho(A) \geq 2$?

Per $\mathbf{x}^T = [1\,1\,1...1]$ si ha $\mathbf{x}^T A\mathbf{x}/\mathbf{x}^T\mathbf{x} = 2$.

Questo basta per provare che $\rho(A) \geq 2$, infatti un teorema dice che se A è reale simmetrica allora al variare di $\mathbf{z}$ in $\mathbb{R}^n$, $\mathbf{z}$ non nullo, il numero reale $\mathbf{z}^T A\mathbf{z}/\mathbf{z}^T\mathbf{z}$ descrive tutto l'intervallo $[\lambda_{\min(A)}, \lambda_{\max(A)}]$, cioè $\max_{\mathbf{z}\, non\, nullo} \mathbf{z}^T A\mathbf{z}/\mathbf{z}^T\mathbf{z} = \lambda_{\max(A)}$, $\min_{z\, non\, nullo} \mathbf{z}^T A\mathbf{z}/\mathbf{z}^T\mathbf{z} = \lambda_{\min(A)}$

Altra dimostrazione (che usa i mezzi che avevate a disposizione), da parte dei tuoi colleghi e, in particolare, da parte di una tua collega:

La tesi e' equivalente a dimostrare che $\rho(1/2A) \geq 1$ ovvero che $(1/2A)^k$ non tende alla matrice nulla.

In aula si è pressoché dimostrato che l'elemento (3,3) di $(1/2A)^k$ tende a $+\infty$.

(prova a vedere come viene per $k = 2, 4, 8, ...$)

Esercizi per i ragazzi di AN1 (prima del primo esonero)

(1) Dati $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ e posto $A = \mathbf{x}\mathbf{y}^H$ scrivere $\|A\|_\infty$, $\|A\|_1$, $\|A\|_2$, $\|A\|_F$ in termini di norme vettoriali di $\mathbf{x}$ e $\mathbf{y}$.

(2) Sia A definita positiva reale e si consideri il metodo di Richardson-Eulero (RE)

$$\mathbf{x}_0 \in \mathbb{R}^n, \quad \mathbf{x}_{k+1} = \mathbf{x}_k + \omega(\mathbf{b} - A\mathbf{x}_k), \ k = 0, 1, 2, \dots,$$

per la risoluzione del sistema $A\mathbf{x} = \mathbf{b}$, $\mathbf{b} \in \mathbb{R}^n$.

(i) Trovare $\omega_* > 0$ tale che RE converge se e solo se $\omega \in (0, \omega_*)$.

(ii) Trovare $\omega_{ott} \in (0, \omega_*)$ tale che RE per $\omega = \omega_{ott}$ ha rapidità di convergenza massima.

(iii) Usare il metoto RE con $\omega = \omega_{ott}$ per risolvere il sistema lineare $A\eta = \mathbf{b}$ ottenuto discretizzando con il metodo delle differenze finite il problema differenziale $-y''(x) = q(x)$, $x \in (a, b)$, $y(a) = \alpha$, $y(b) = \beta$ (caso particolare di quello visto a lezione). Calcolare in questo caso il valore esatto di ω_{ott}.

(3) Sia $A \in \mathbb{C}^{n\times n}$ e si consideri la partizione di A

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \ A_{11} \ k \times k, \quad A_{22} \ n-k \times n-k.$$

Dimostrare che

$$\begin{bmatrix} A_{11} & O \\ O & A_{22} \end{bmatrix} = \frac{1}{2}(A + JAJ), \ J = \begin{bmatrix} I_k & O \\ O & -I_{n-k} \end{bmatrix}.$$

(4) Sia $A \in \mathbb{C}^{n\times n}$ e $\| \ \|$ una norma matriciale invariante per trasformazioni unitarie, cioè tale che $\|QM\| = \|MQ\| = \|M\|$ $(\forall M \in \mathbb{C}^{n\times n})$, se $Q \in \mathbb{C}^{n\times n}$ è unitaria. Dimostrare che $\| \operatorname{diag}(a_{11}, a_{22}, \dots, a_{nn})\| \le \|A\|$. Se $\| \ \|$ non è invariante per trasformazioni unitarie la disuguaglianza è ancora vera?

(5) Sia $A \in \mathbb{C}^{n\times n}$. È vera la disuguaglianza $\rho(A) \ge \max_i |a_{ii}|$?

Per minimizzare

$$\left|\frac{\alpha}{x^3} - (A + \frac{B}{x} + \frac{C}{x^2})\right| \qquad (*)$$

in $[x_{\min}, x_{\max}]$, $0 < x_{\min} < x_{\max}$, si potrebbe:

1) scegliere tre punti in $[x_{\min}, x_{\max}]$ in modo opportuno e imporre che (*) in questi tre punti sia zero

2) scegliere più di tre punti in $[x_{\min}, x_{\max}]$, siano essi $x_0, x_1, \ldots, x_k$, $k \geq 3$, e imporre che la somma $\sum_{i=0}^{k} \omega_i (*)_{x=x_i}$ oppure $\sum_{i=0}^{k} \omega_i (*)^2_{x=x_i}$ (con $\omega_i > 0$ pesi opportuni) sia minima

3) imporre che l'integrale $\int_{x_{\min}}^{x_{\max}} \omega(x)(*)^2 \, dx$ (con $\omega(x) > 0$ funzione peso opportuna, oppure $\omega(x) = 1$) sia minimo

4) si potrebbe cercare di risolvere il problema

$$\min_{A,B,C\in\mathbb{R}} \max_{x\in[x_{\min},x_{\max}]} \left|\frac{\alpha}{x^3} - (A + \frac{B}{X} + \frac{C}{X^2})\right|$$

Inizio risoluzione di quest'ultimo problema:

$$\max_{x\in[x_{\min},x_{\max}]} \left|\frac{\alpha}{x^3} - (A + \frac{B}{x} + \frac{C}{x^2})\right|$$

$$= \max_{y\in[1/x_{\max},1/x_{\min}]} \left|\alpha y^3 - (A + By + Cy^2)\right| \quad (y = m + tn,\ m = \frac{\frac{1}{x_{\max}} + \frac{1}{x_{\min}}}{2},\ n = \frac{\frac{1}{x_{\min}} - \frac{1}{x_{\max}}}{2})$$

$$= \max_{t\in[-1,1]} \left|\alpha(m + tn)^3 - (A + B(m + tn) + C(m + tn)^2)\right|,$$

$$= \max_{t\in[-1,1]} \left|\alpha n^3 t^3 - (A + Bm + Cm^2 - \alpha m^3) - t(Bn + 2Cmn - 3\alpha m^2 n) - t^2(Cn^2 - 3\alpha mn^2)\right|,$$

$$= |\alpha| n^3 \max_{t\in[-1,1]} \left|t^3 - \frac{A + Bm + Cm^2 - \alpha m^3}{\alpha n^3} - t\frac{Bn + 2Cmn - 3\alpha m^2 n}{\alpha n^3} - t^2\frac{Cn^2 - 3\alpha mn^2}{\alpha n^3}\right|.$$

Dalla teoria dei polinomi di Chebycev, quest'ultima quantità è minima quando

$$t^3 - \frac{A + Bm + Cm^2 - \alpha m^3}{\alpha n^3} - t\frac{Bn + 2Cmn - 3\alpha m^2 n}{\alpha n^3} - t^2\frac{Cn^2 - 3\alpha mn^2}{\alpha n^3} = \frac{1}{2^2}T_3(t)$$

dove $T_3(t)$ è il terzo polinomio di Chebycev, cioè $T_3(t) = 4t^3 - 3t$. Dunque, per trovare gli A, B, C "ottimali" basta imporre l'uguaglianza

$$\frac{A + Bm + Cm^2 - \alpha m^3}{\alpha n^3} + t\frac{Bn + 2Cmn - 3\alpha m^2 n}{\alpha n^3} + t^2\frac{Cn^2 - 3\alpha mn^2}{\alpha n^3} = \frac{3}{4}t$$

$\Rightarrow$

$$C = \alpha 3m, \quad B = \alpha(\frac{3}{4}n^2 - 3m^2),\ A = \alpha(m^3 - \frac{3}{4}mn^2)$$

Analisi numerica STM (Di Fiore, Zellini) - Scritto d'esame - 10 Settembre 2013

1) Dimostrare la disuguaglianza
$$\left\| \begin{bmatrix} A & B \\ C & D \end{bmatrix} \right\|_2^2 \geq \|A\|_2^2$$
(per ogni $A \in \mathbb{C}^{n\times n}$, $B \in \mathbb{C}^{n\times m}$, $C \in \mathbb{C}^{m\times n}$, $D \in \mathbb{C}^{m\times m}$).

2) Dimostrare la disuguaglianza
$$\|A\|_F^2 \leq \|A\|_\infty^2 n$$
(per ogni $A \in \mathbb{C}^{n\times n}$).

3) Sia T una generica matrice triangolare inferiore $n \times n$ non singolare.

i) Scrivere la fattorizzazione LU di T.

ii) Trovare l tale che $\mu_2(T) \geq l$, con l funzione di elementi diagonali di T.

4) Siano
$$A = \begin{bmatrix} 1 & 1/12 & 0 & -1/6 \\ 0 & 1 & 3/4 & 0 \\ -3 & 0 & 1 & -1/4 \\ 0 & -1 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix}.$$

i) Dire se A è riducibile o irriducibile, giustificando la risposta.

ii) Usando il metodo di Gauss, scrivere una matrice $A^{(1)}$ e un vettore $\mathbf{b}^{(1)}$ tali che $A^{(1)}_{21} = A^{(1)}_{31} = A^{(1)}_{41} = 0$ e il sistema $A^{(1)}\mathbf{x} = \mathbf{b}^{(1)}$ è equivalente al sistema $A\mathbf{x} = \mathbf{b}$.

iii) Usando il metodo di Givens, scrivere una matrice $A^{(1)}$ e un vettore $\mathbf{b}^{(1)}$ tali che $A^{(1)}_{21} = A^{(1)}_{31} = A^{(1)}_{41} = 0$ e il sistema $A^{(1)}\mathbf{x} = \mathbf{b}^{(1)}$ è equivalente al sistema $A\mathbf{x} = \mathbf{b}$.

iv) Studiare la convergenza del metodo iterativo di Gauss-Seidel, nella risoluzione del sistema $A\mathbf{x} = \mathbf{b}$.

5) Dimostrare che autovettori corrispondenti ad autovalori distinti di una matrice $A \in \mathbb{C}^{n\times n}$ hermitiana sono ortogonali, cioè
$$A\mathbf{x} = \lambda\mathbf{x},\ A\mathbf{y} = \tilde{\lambda}\mathbf{y},\ \lambda \neq \tilde{\lambda},\ A = A^H \quad \Rightarrow \quad \mathbf{x}^H\mathbf{y} = 0.$$