Scritto d'esame Analisi Numerica, Matematica - 19 Febbraio 2013

Esercizio 1

Siano x_i, $i = 0, 1, \dots, n$, numeri reali distinti ed f una funzione C^1 in un aperto contenente gli x_i. Sia $p(x)$ il polinomio di grado al più n tale che $p(x_i) = f(x_i)$, $i = 0, 1, \dots, n$. Trovare un polinomio $r(x)$ tale che $p(x)+r(x)$ ha grado minimo, $p(x_i)+r(x_i) = f(x_i)$, $i = 0, 1, \dots, n$, e $p'(x_n) + r'(x_n) = f'(x_n)$.

Esercizio 2

Sia $\mathcal{L}$ un sottospazio di $\mathbb{C}^{n\times n}$. Sia $A \in \mathbb{C}^{n\times n}$ e $\mathcal{L}_A$ la matrice di $\mathcal{L}$ che meglio approssima A in norma di Frobenius, cioè tale che $\|A - \mathcal{L}_A\|_F = \min_{X\in\mathcal{L}} \|A - X\|_F$ (Nota: $\mathcal{L}_A$ esiste ed è unica). Nell'ipotesi che $\mathcal{L}$ sia chiuso per trasposizione coniugata (cioè $M \in \mathcal{L} \Rightarrow M^H \in \mathcal{L}$), mostrare che

i) $A = A^H \Rightarrow \mathcal{L}_A = (\mathcal{L}_A)^H$

ii) $\mathcal{L} = \{UDU^H : D \text{ diagonali}\}$ e A definita positiva $\Rightarrow \mathcal{L}_A$ definita positiva. (U unitaria $n \times n$ fissata).

Esercizio 3

Supponendo che esista la seguente decomposizione

$$\begin{bmatrix} d_1 & f_1 & & \\ e_2 & d_2 & \cdot & \\ & \cdot & \cdot & f_{n-1} \\ & & e_n & d_n \end{bmatrix} = LU, \;\; L_{ii} = 1,$$

L triangolare inferiore, U triangolare superiore,

i) osservare che $L_{ij} = 0 = U_{ji}$ per $i > j + 1$.

ii) posto $m_{i+1} = L_{i+1,i}$, $i = 1, \dots, n-1$, e $u_i = U_{ii}$, $i = 1, \dots, n$, mostrare che $U_{i,i+1} = f_i$, $i = 1, \dots, n-1$, $u_1 = d_1$, ed $m_i = e_i/u_{i-1}$, $u_i = d_i - m_i f_{i-1}$, $i = 2, \dots, n$.

iii) posto $q_0 = 1$, $q_1 = d_1$, $q_i = d_i q_{i-1} - e_i f_{i-1} q_{i-2}$, $i = 2, \dots, n$, mostrare che $u_i = q_i/q_{i-1}$, $i = 1, 2, \dots, n$.

Esercizio 4

Determinare il polinomio $p(x)$ di grado minore o uguale a tre che meglio approssima $f(x) = x^4$ in $[-1, 1]$ nel senso minmax. Osservare che $p(x)$ interpola $f(x)$ in quattro punti.

Esercizio 1

È evidente che il polinomio $r(x)$ deve essere tale che $r(x_i) = 0$, $i = 0, 1, \dots, n$, e quindi $r(x) = q(x)\prod_{j=0}^n (x - x_j)$, dove $q(x)$ è un generico polinomio. L'uguaglianza $p'(x_n)+r'(x_n) = f'(x_n)$ diventa quindi $p'(x_n)+q'(x_n)\prod_{j=0}^n(x_n - x_j)+q(x_n)\sum_{k=0}^n \prod_{j=0,\, j\neq k}^n (x_n - x_j) = p'(x_n)+ q(x_n)\prod_{j=0,\, j\neq n}^n (x_n - x_j) = f'(x_n)$, da cui la seguente condizione sul polinomio q

$$q(x_n) = (f'(x_n) - p'(x_n))/\prod_{j=0,\, j\neq n}^n (x_n - x_j). \qquad (*)$$

Infine, è richiesto che $p(x) + r(x) = p(x) + q(x)\prod_{j=0}^{n}(x - x_j)$ abbia grado minimo. Poiché p ha un grado fissato, e il grado di $\prod_{j=0}^{n}(x - x_j)$ e esattamente $n + 1$, non possiamo far altro che minimizzare il grado di q. Il polinomio q di grado minimo che soddisfa (*) è ovviamente il polinomio costante

$$q(x) = (f'(x_n) - p'(x_n))/\prod_{j=0,\, j\neq n}^{n}(x_n - x_j).$$

Quindi il polinomio richiesto è

$$p(x) + r(x) = p(x) + \Big(\ (f'(x_n) - p'(x_n))/\prod_{j=0,\, j\neq n}^{n}(x_n - x_j)\ \Big)\prod_{j=0}^{n}(x - x_j)$$

e ha sempre grado esattamente $n+1$ a meno che fortuitamente la condizione $f'(x_n)-p'(x_n) = 0$ è già soddisfatta, in tal caso il polinomio cercato $p(x) + r(x)$ coincide con $p(x)$ e, quindi, ha il grado di p.

Esercizio 2

i) Sia $\mathcal{L}_A$ la matrice di $\mathcal{L}$ per cui $\|A - \mathcal{L}_A\|_F = \min_{X\in\mathcal{L}} \|A - X\|_F$. Poiché la norma di Frobenius di una matrice M è uguale alla norma di Frobenius della trasposta coniugata M^H, si ha che $\|A - \mathcal{L}_A\|_F = \|(A - \mathcal{L}_A)^H\|_F = \|A^H - (\mathcal{L}_A)^H\|_F$. Ma, per ipotesi, A è hermitiana, dunque $\|A - \mathcal{L}_A\|_F = \|A - (\mathcal{L}_A)^H\|_F$, e abbiamo l'identità:

$$\|A - (\mathcal{L}_A)^H\|_F = \|A - \mathcal{L}_A\|_F = \min_{X\in\mathcal{L}} \|A - X\|_F.$$

Inoltre, per l'ipotesi su $\mathcal{L}$, essendo $\mathcal{L}_A$ un elemento di $\mathcal{L}$ anche $(\mathcal{L}_A)^H$ deve essere un elemento di $\mathcal{L}$. Quindi, se la matrice $(\mathcal{L}_A)^H$ fosse diversa da $\mathcal{L}_A$, avremmo due diverse matrici di $\mathcal{L}$ che rendono minima $\|A - X\|_F$, al variare di $X \in \mathcal{L}$, mentre sappiamo che ve n'è solo una. Ne segue che $(\mathcal{L}_A)^H = \mathcal{L}_A$.

ii) Siano $\mathcal{L} = \{UDU^H :\ D \text{ diagonali}\}$ e A definita positiva (cioè $A = A^H$ e $\mathbf{z}^H A\mathbf{z} > 0\ \forall\, \mathbf{z} \in \mathbb{C}^n$, $\mathbf{z} \neq \mathbf{0}$). Si vuole mostrare che allora anche $\mathcal{L}_A$ è definita positiva. Si osserva innanzitutto che $\mathcal{L}$ è chiuso per trasposizione coniugata. Infatti, se $M \in \mathcal{L}$ allora $M^H = (UDU^H)^H = UD^HU^H = U\overline{D}U^H$ dove $\overline{D}$ è una matrice diagonale, dunque $M^H \in \mathcal{L}$. Quindi, per il punto i) possiamo dire che $\mathcal{L}_A$ deve essere hermitiana. Resta da dimostrare che $\mathbf{z}^H\mathcal{L}_A\mathbf{z} > 0$ per ogni $\mathbf{z} \in \mathbb{C}^n$, $\mathbf{z} \neq \mathbf{0}$, o, equivalentemente, che gli autovalori di $\mathcal{L}_A$ sono positivi. Per farlo, proviamo a cercare una rappresentazione di $\mathcal{L}_A$, usando il fatto che la norma di Frobenius è invariante per trasformazioni unitarie. Dobbiamo minimizzare

$$\|A - UDU^H\|_F = \|U^HAU - D\|_F$$

al variare di D tra tutte le matrici diagonali, ovvero minimizzare

$$\|U^HAU - D\|_F = \sqrt{\sum_{i\neq j} |(U^HAU)_{ij}|^2 + \sum_i |(U^HAU)_{ii} - D_{ii}|^2}$$

al variare di $D_{ii} \in \mathbb{C}$, $i = 1, 2, \ldots, n$. È evidente che il minimo si raggiunge quando si sceglie $D_{ii} = (U^H AU)_{ii}$, $\forall\, i$. Dunque si ha la rappresentazione di $\mathcal{L}_A$ cercata: $\mathcal{L}_A = U \operatorname{diag}((U^H AU)_{ii})U^H$. Ora,

$$\mathbf{z}^H \mathcal{L}_A \mathbf{z} = \mathbf{z}^H U \operatorname{diag}((U^H AU)_{ii})U^H \mathbf{z} = \sum_i |(U^H \mathbf{z})_i|^2 (U^H AU)_{ii}$$

dove $(U^H AU)_{ii} = (U\mathbf{e}_i)^H A(U\mathbf{e}_i) > 0$, $\forall\, i$, perché A è definita positiva. Dunque $\mathbf{z}^H \mathcal{L}_A \mathbf{z} \geq 0$. Notiamo infine che se fosse $\mathbf{z}^H \mathcal{L}_A \mathbf{z} = 0$ si dovrebbe avere $\sum_i |(U^H \mathbf{z})_i|^2 (U^H AU)_{ii} = 0$, e, di conseguenza, $(U^H \mathbf{z})_i = 0\ \forall\, i$, $U^H \mathbf{z} = \mathbf{0}$, $\mathbf{z} = \mathbf{0}$. Quindi, $\mathbf{z}^H \mathcal{L}_A \mathbf{z} > 0\ \forall\, \mathbf{z} \in \mathbb{C}^n$, $\mathbf{z} \neq \mathbf{0}$.

Esercizio 3

Sia A una matrice $n \times n$ tridiagonale e tale che esistono una matrice L triangolare inferiore con $L_{ii} = 1$ ed una matrice U triangolare superiore tali che

$$A = \begin{bmatrix} d_1 & f_1 & & \\ e_2 & d_2 & \cdot & \\ & \cdot & \cdot & f_{n-1} \\ & & e_n & d_n \end{bmatrix} = LU.$$

Siano $[A]_{i \times i}$ le sottomatrici $i \times i$ in alto a sinistra di A. Dal fatto che la rappresentazione LU di A sia possibile segue che $\det([A]_{i \times i}) \neq 0$, per $i = 1, 2, \ldots, n-1$ (perché?). Inoltre, per $i = 1, 2, \ldots, n$, deve essere $\det([A]_{i \times i}) = \prod_{j=1}^{i} U_{jj}$ (perché?).

i) Dimostriamo ora che $L_{ij} = U_{ji} = 0$ se $i > j + 1$ $(j = 1, \ldots, n-2)$, per induzione su j. Si ha $A_{1i} = U_{1i}$ e $A_{i1} = U_{11} L_{i1}$. Dunque, $U_{1i} = 0 = L_{i1}$, $i = 3, \ldots, n$ (perché $U_{11} \neq 0$) e la tesi è dimostrata per $j = 1$. Supponiamo di aver già dimostrato che per $k \leq j - 1$ si ha $U_{ki} = 0 = L_{ik}$, $i = k + 2, \ldots, n$. Allora per $i > j + 1$ si ha $A_{ji} = L_{j,j-1} * 0 + 1 * U_{ji}$ e $A_{ij} = 0 * U_{j-1,j} + L_{ij} * U_{jj}$, da cui, essendo $U_{jj} \neq 0$, si ha $U_{ji} = 0 = L_{ij}$, $i = j + 2, \ldots, n$.

Nota: Più in generale si dimostra che se A $n \times n$ generica è decomponibile nel prodotto LU e tale che $A_{ij} = 0$ per $i > j + r$ e $j > i + s$ (A "matrice a bande"), allora $L_{ij} = 0$, $i > j + r$, e $U_{ij} = 0$, $j > i + s$ (L ed U hanno le stesse bande di A).

ii) Imponendo l'identità

$$\begin{bmatrix} d_1 & f_1 & & \\ e_2 & d_2 & \cdot & \\ & \cdot & \cdot & f_{n-1} \\ & & e_n & d_n \end{bmatrix} = \begin{bmatrix} 1 & & & \\ m_2 & 1 & & \\ & \cdot & \cdot & \\ & & m_n & 1 \end{bmatrix} \begin{bmatrix} u_1 & f_1 & & \\ & u_2 & \cdot & \\ & & \cdot & f_{n-1} \\ & & & u_n \end{bmatrix}$$

(ovvero che gli elementi $(1, 1)$, $(i, i-1)$ e (i, i), $i = 2, \ldots, n$, del primo membro siano uguali ai corrispondenti del secondo membro), si ottengono le seguenti condizioni

$u_1 = d_1$,

$e_i = m_i u_{i-1}$, $e_i / u_{i-1} = m_i$, $i = 2, \ldots, n$,

$d_i = m_i f_{i-1} + u_i$, $d_i - m_i f_{i-1} = u_i$, $i = 2, \ldots, n$, dalle quali si ricava il seguente algoritmo per il calcolo della decomposizione LU della matrice tridiagonale A:

$$u_1 = d_1; \quad \text{per} \;\; i = 2, \ldots, n: \quad m_i = e_i/u_{i-1}, \; u_i = d_i - m_i f_{i-1}.$$

iii) Poniamo $q_0 = 1$, $q_1 = d_1$, $q_i = d_i q_{i-1} - e_i f_{i-1} q_{i-2}$, $i = 2, \ldots, n$, osservando che q_i non è altro che il determinante della sottomatrice $i \times i$ in alto a sinistra di A. Dimostriamo ora per induzione su i l'identità $u_i = q_i/q_{i-1}$, $i = 1, 2, \ldots, n$. Per $i = 1$ essa è vera. Supponiamo ora che sia vero che $u_{i-1} = q_{i-1}/q_{i-2}$. Allora, per le identità di cui al punto ii), si ha

$$\begin{aligned} u_i &= d_i - m_i f_{i-1} = d_i - f_{i-1} e_i / u_{i-1} = d_i - f_{i-1} e_i q_{i-2}/q_{i-1} \\ &= (d_i q_{i-1} - f_{i-1} e_i q_{i-2})/q_{i-1} = q_i/q_{i-1}. \end{aligned}$$

Nota: L'identità $u_i = q_i/q_{i-1}$ ovvero l'identità $\det([A]_{i\times i}) = U_{ii} \det([A]_{i-1\times i-1})$ è vera più in generale per ogni matrice A decomponibile nel prodotto LU (cioè A non deve essere necessariamente tridiagonale).

Esercizio 4

Si deve calcolare $p(x)$, di grado minore o uguale a 3, polinomio di migliore approssimazione minmax di x^4 in $[-1, 1]$. Esso deve essere tale che $x^4 - p(x) = \frac{1}{2^3} T_4(x)$, dove T_4 è il polinomio di Chebycev di grado 4. Ricaviamoci T_4 esplicitamente: $T_0(x) = 1$, $T_1(x) = x$, $T_2(x) = 2xT_1(x) - T_0(x) = 2x^2 - 1$, $T_3(x) = 2xT_2(x) - T_1(x) = 4x^3 - 3x$, $T_4(x) = 2xT_3(x) - T_2(x) = 8x^4 - 8x^2 + 1$. Dunque deve essere $x^4 - p(x) = \frac{1}{2^3}(8x^4 - 8x^2 + 1) = x^4 - x^2 + 1/8$, da cui $p(x) = x^2 - 1/8$.

Nota: Calcoliamo il polinomio $p(x)$ di migliore aprossimazione nel senso dei minimi quadrati (caso continuo) di x^4 in $[-1, 1]$. I polinomi ortogonali rispetto all'intervallo $[-1, 1]$ e al peso $\omega(x) = 1$, ovvero rispetto al prodotto scalare $(h, g) = \int_{-1}^{1} h(x)g(x)\,dx$, sono: $p_0(x) = 1$, $p_1(x) = x$, $p_2(x) = x^2 - 1/3$. Dunque, $(p_0, p_0) = 2$, $(p_1, p_1) = 2/3$, $(p_2, p_2) = 8/45$, $(f, p_0) = 2/5$, $(f, p_1) = 0$, $(f, p_2) = 16/105$, quindi

$$\begin{aligned} p(x) &= \textstyle\sum_{i=0}^{2} ((f, p_i)/(p_i, p_i)) p_i(x) \\ &= ((2/5)/2) + ((16/105)/(8/45))(x^2 - 1/3) = 1/5 + (6/7)(x^2 - 1/3) = 6x^2/7 - 3/35 \end{aligned}$$

È interessante confrontare graficamente il modo in cui i polinomi $6x^2/7 - 3/35$ e $x^2 - 1/8$ approssimano x^4.

Scritto d'esame Analisi Numerica, Matematica - 29 Gennaio 2013

Esercizio 1

Si consideri l'equazione non lineare $x = c/(1+x^2) =: g(x)$, $c > 2$, e sia α_c la sua unica radice reale.

i) Dimostrare che $\exists\, \xi \in \mathbb{R}^+$ tale che la successione $x_0 = \xi$, $x_{k+1} = g(x_k)$, $k = 0, 1, 2, \ldots$, è costituita dai due soli elementi ξ e $g(\xi)$ ed osservare che $\xi, g(\xi) \in (0, c)$.

ii) Dimostrare la disuguaglianza $\alpha_c > 1$. Provare l'inclusione $\{x : x > 1,\ g'(x) \geq -1\} \subset \{x : g(x) < x\}$. Dedurre che $g'(\alpha_c) < -1$, cioè la successione $x_{k+1} = g(x_k)$ non può convergere ad α_c a meno che $x_0 = \alpha_c$.

iii) Sia $g(x)$ una funzione generica, regolare in un intorno di un suo punto fisso α. Supponiamo $g'(\alpha) \neq 1$.

Posto $h(x) = (g(g(x)) - 2g(x) + x)/(g(x) - x)$, dimostrare che $h(\alpha) \neq 0$.

Posto $\psi(x) = x - (g(x) - x)/h(x)$, dimostrare che α è un punto fisso di ψ e la successione $x_{k+1} = \psi(x_k)$, $k = 0, 1, 2, \ldots$, converge ad α con ordine di convergenza almeno quadratico (se $x_0 \approx \alpha$).

Esercizio 2

Sia $\mathcal{I} = \int_0^1 \frac{1}{1+x^2}\, dx$.

i) Calcolare le approssimazioni di $\mathcal{I}$ fornite dalle formule di quadratura di Newton-Cotes a quattro nodi e Gauss-Legendre a tre nodi e confrontarle con $\mathcal{I}$.

ii) Posto $x_i = i\frac{1}{n}$, $i = 0, 1, \ldots, n$, scrivere precisamente l'approssimazione t_{n+1} di $\mathcal{I}$ ottenuta con la formula composta del trapezio, definita in termini dei $n+1$ valori $f(x_i)$, $i = 0, 1, \ldots, n$, e $\forall\, \varepsilon > 0$ determinare ν_ε t.c. $\forall\, n > \nu_\varepsilon$ si ha $|t_{n+1} - \mathcal{I}| < \varepsilon$

(Nota: $\int_a^b f(x)\, dx = \frac{b-a}{2}(f(a) + f(b)) - \frac{(b-a)^3}{12} f''(\eta)$)

Esercizio 3

Sia A $n \times n$ non singolare con autovalori reali positivi.

i) Sotto quali ipotesi su $\alpha \in \mathbb{R}$ vale la seguente formula per A^{-1}, $A^{-1} \doteq \alpha \sum_{k=0}^{+\infty} (I - \alpha A)^k$?

ii) Dare una limitazione superiore per gli α per cui vale $\cdot$ quando

$$A_{ij} = \begin{cases} i & i = j \\ 1/2^{j-1} & i < j \\ 1/2^{i-1} & i > j \end{cases}, \quad 1 \leq i, j \leq n.$$

ESERCIZIO 1

L'equazione $x = c/(1+x^2)$ ha come radici reali le ascisse dei punti di intersezione tra la bisettrice $y = x$ e la funzione $y = c/(1+x^2)$. Quest'ultima funzione è pari, in zero vale c, e, per valori positivi di c, è positiva, decrescente in $[0, +\infty)$ e tende a 0^+ quando $x \to +\infty$. È quindi evidente che, per $c > 0$, l'equazione $x = c/(1+x^2)$ ha una unica radice reale e che essa è contenuta nell'intervallo $(0, c)$; chiamiamo α_c tale radice. Supponiamo da ora in poi c positivo.

i) La successione $x_0 = \xi$, $x_{k+1} = g(x_k)$, $k = 0, 1, 2, \ldots$, è formata da al più due elementi se e solo se è verificata l'uguaglianza

$$g(g(\xi)) = \xi \quad \Leftrightarrow \quad \frac{c}{1 + \frac{c^2}{(1+\xi^2)^2}} = \xi \quad \Leftrightarrow \quad c^2\xi/((1+\xi^2)^2) - c + \xi = 0,$$

e quest'ultima equazione è verificata se e solo se vale almeno una delle due seguenti identità

$$c = \frac{1+\xi^2}{\xi}, \quad c = \xi(1+\xi^2).$$

La seconda identità ha come unica soluzione $\xi = \alpha_c$ e le soluzioni della prima identità sono le radici di $\xi^2 - c\xi + 1 = 0$. Quindi ξ è tale che $g(g(\xi)) = \xi$ se e solo se ξ assume uno dei seguenti tre valori

$$\xi = \alpha_c, \quad \xi = \frac{1}{2}(c - \sqrt{c^2-4}) =: \xi^-, \quad \xi = \frac{1}{2}(c + \sqrt{c^2-4}) =: \xi^+ .$$

Osserviamo che per $c = 2$ tali tre valori coincidono tra loro e con $\alpha_2 = 1$, la soluzione dell'equazione $x = 2/(1+x^2)$. Per $c > 2$ tali tre valori sono distinti e α_c, la soluzione dell'equazione $x = c/(1+x^2)$, appartiene all'intervallo (ξ^-, ξ^+).

È quindi evidente che, per $c > 2$, la successione $x_0 = \xi^\pm$, $x_{k+1} = c/(1+x_k^2)$, $k = 0, 1, 2, \ldots$, assume alternatamente i soli due valori distinti ξ^- e ξ^+, entrambi positivi ed entrambi più piccoli di c, "ballando" attorno ad α_c. Con un disegno si capisce bene ciò che accade.

ii) Sia $c > 2$. Mostriamo che $\alpha_c > 1$. È sufficiente osservare che, essendo $g(1) = c/2$, quando $c > 2$ vale la disuguaglianza $g(1) > 1$, cioè il grafico di g in 1 sta sopra il grafico della bisettrice. Allora la tesi segue dal fatto che per $c > 2$ la funzione g è decrescente in $(1, +\infty)$.

Dimostriamo l'inclusione $\{x : x > 1, g'(x) \geq -1\} \subset \{x : g(x) < x\}$. Sia $x \in \{x : x > 1, g'(x) \geq -1\}$. Allora

$$-2c\frac{x}{(1+x^2)^2} \geq -1 \quad \Leftrightarrow \quad 2c\frac{x}{(1+x^2)^2} \leq 1 \quad \Rightarrow \quad g(x) = \frac{c}{1+x^2} \leq \frac{1+x^2}{2x} < \frac{x^2+x^2}{2x} = x.$$

Proviamo infine che $g'(\alpha_c) < -1$. Se fosse $g'(\alpha_c) \geq -1$, allora, visto che $\alpha_c > 1$, per il risultato precedente si dovrebbe avere $g(\alpha_c) < \alpha_c$ e ciò è assurdo.

iii) Sappiamo che g è una funzione generica tale che l'equazione $x = g(x)$ ha un punto fisso $\alpha \in \mathbb{R}$ ($\alpha = g(\alpha)$), g è regolare in un intorno di α, e in α la derivata di g non è uguale a 1, cioè $g'(\alpha) \neq 1$.

Posto $h(x) = (g(g(x)) - 2g(x) + x)/(g(x) - x)$, occorre verificare che $h(\alpha) \neq 0$. Si osserva che il limite

$$\lim_{x\to\alpha} h(x) = \frac{g(g(\alpha)) - 2g(\alpha) + \alpha}{g(\alpha) - \alpha} = \frac{g(\alpha) - 2\alpha + \alpha}{g(\alpha) - \alpha} = \frac{\alpha - 2\alpha + \alpha}{\alpha - \alpha} = \frac{0}{0}$$

è una forma indeterminata. Allora, si applica de l'Hospital e si ottiene

$$\lim_{x\to\alpha} h(x) = \lim_{x\to\alpha} \frac{g'(g(x))g'(x) - 2g'(x) + 1}{g'(x) - 1} = \frac{g'(g(\alpha))g'(\alpha) - 2g'(\alpha) + 1}{g'(\alpha) - 1} = \frac{g'(\alpha)g'(\alpha) - 2g'(\alpha) + 1}{g'(\alpha) - 1} = g'(\alpha) - 1.$$

Quindi il valore $h(\alpha)$ è ben definito ed uguale a $g'(\alpha) - 1$, e siccome per ipotesi $g'(\alpha) - 1 \neq 0$, si ha che $h(\alpha) \neq 0$.

Posto $\psi(x) = x - (g(x) - x)/h(x)$, usando il fatto che $h(\alpha) \neq 0$ si conclude subito che $\psi(\alpha) = \alpha - 0/h(\alpha) = \alpha$, ovvero che α è un punto fisso per ψ. Dobbiamo ora dimostrare che $\psi'(\alpha) = 0$. Essendo

$$\psi'(x) = 1 - ((g'(x) - 1)h(x) - h'(x)(g(x) - x))\frac{1}{h(x)^2},$$

$$\psi'(\alpha) = 1 - (g'(\alpha) - 1)h(\alpha)\frac{1}{h(\alpha)^2} + \lim_{x\to\alpha} h'(x)(g(x) - x))\frac{1}{h(x)^2} = \lim_{x\to\alpha} h'(x)(g(x) - x))\frac{1}{h(x)^2},$$

occorre investigare se $h'(x)(g(x) - x)$ è ben definito in α e in caso affermativo che valore assume:

$$\begin{aligned}
\lim_{x\to\alpha} h'(x)(g(x) - x) &= \lim_{x\to\alpha} \frac{(g'(g(x))g'(x) - 2g'(x) + 1)(g(x) - x) - (g'(x) - 1)(g(g(x)) - 2g(x) + x)}{g(x) - x} \\
&= \lim_{x\to\alpha} \frac{[g''(g(x))g'(x)^2 + g'(g(x))g''(x) - 2g''(x)](g(x) - x) - g''(x)[g(g(x)) - 2g(x) + x]}{g'(x) - 1} \\
&= \frac{[g''(g(\alpha))g'(\alpha)^2 + g'(g(\alpha))g''(\alpha) - 2g''(\alpha)](g(\alpha) - \alpha) - g''(\alpha)[g(g(\alpha)) - 2g(\alpha) + \alpha]}{g'(\alpha) - 1} \\
&= 0.
\end{aligned}$$

Dunque $\lim_{x\to\alpha} h'(x)(g(x) - x) = 0$, e, di conseguenza,

$$\psi'(\alpha) = \lim_{x\to\alpha} h'(x)(g(x) - x))\frac{1}{h(x)^2} = \frac{0}{h(\alpha)^2} = 0.$$

Le condizioni $\alpha = \psi(\alpha)$, ψ regolare in un intorno di α, e $\psi'(\alpha) = 0$, ci dicono che esiste $\delta > 0$ tale che, se $x_0 \in (\alpha - \delta, \alpha + \delta)$, allora la successione $x_{k+1} = \psi(x_k)$, $k = 0, 1, 2, \ldots$, converge ad α con ordine di convergenza almeno quadratico.

ESERCIZIO 2 i) Sia $\mathcal{I} = \int_0^1 \frac{1}{1+x^2}\, dx = \frac{\pi}{4}$. Occorre calcolare le approssimazioni di $\mathcal{I}$ fornite dalle formule di quadratura di Newton-Cotes a quattro nodi e di Gauss-Legendre a tre nodi. La formula di quadratura di Newton-Cotes a quattro nodi in $[-1, 1]$ è del tipo

$$\begin{aligned}
\int_{-1}^{1} f(x)\, dx &= A_0 f(-1) + A_1 f(-\frac{1}{3}) + A_2 f(\frac{1}{3}) + A_3 f(1) + E \\
&= A_0(f(-1) + f(1)) + A_1(f(-\frac{1}{3}) + f(\frac{1}{3})) + E
\end{aligned}$$

(si ha $A_i = A_{3-i}$ perché i nodi e il peso sono simmetrici rispetto al centro dell'intervallo di integrazione). Per determinare i coefficienti A_0 e A_1 imponiamo che E sia zero per $f = 1, x^2$ (per $f = x, x^3$ l'errore E sarà automaticamente zero perché in tal caso sia l'integrale che la formula di quadratura hanno valore zero). Otteniamo così le condizioni: $2 = 2A_0 + 2A_1$, $2/3 = 2A_0 + A_1 2/9$, cioè $A_1 = 3/4$, $A_0 = 1/4$. Quindi la formula di quadratura cercata è

$$\int_{-1}^{1} f(x)\,dx = \frac{1}{4}(f(-1) + f(1)) + \frac{3}{4}(f(-\frac{1}{3}) + f(\frac{1}{3})) + E,$$

e, in $[a, b]$,

$$\begin{aligned}
\int_a^b f(x)\,dx &= \frac{b-a}{2}\int_{-1}^{1} f(\frac{a+b}{2} + t\frac{b-a}{2})\,dt \\
&= \frac{b-a}{2}\int_{-1}^{1} \tilde{f}(t)\,dt \\
&= \frac{b-a}{2}\Big(\frac{1}{4}(\tilde{f}(-1) + \tilde{f}(1)) + \frac{3}{4}(\tilde{f}(-\frac{1}{3}) + \tilde{f}(\frac{1}{3})) + \tilde{E}\Big) \\
&= \frac{b-a}{2}\Big(\frac{1}{4}(f(a) + f(b)) + \frac{3}{4}(f(\frac{a+b}{2} - \frac{1}{3}\frac{b-a}{2}) + f(\frac{a+b}{2} + \frac{1}{3}\frac{b-a}{2}))\Big) + E.
\end{aligned}$$

Nel nostro caso in cui $a = 0$, $b = 1$, $f(x) = 1/(1 + x^2)$:

$$\int_0^1 \frac{1}{1+x^2}\,dx = \frac{1}{2}\Big(\frac{1}{4}(f(0)+f(1)) + \frac{3}{4}(f(\frac{1}{3}) + f(\frac{2}{3}))\Big) + E = \frac{1}{2}\Big(\frac{1}{4}(1 + \frac{1}{2}) + \frac{3}{4}(\frac{9}{10} + \frac{9}{13})\Big) + E,$$

$$\int_0^1 \frac{1}{1+x^2}\,dx = \frac{3}{8}\Big(\frac{1}{2} + 9\frac{23}{130}\Big) + E = \frac{51}{65} + E = 0.784615\ldots + E.$$

Ricaviamo ora la approssimazione fornita dalla formula di quadratura di Gauss-Legendre a tre nodi. In $[-1, 1]$ per una funzione integranda qualsiasi f, la formula era

$$\int_{-1}^{1} f(t)\,dt = \frac{5}{9}(f(-\xi) + f(\xi)) + \frac{8}{9}f(0) + E, \quad \xi = \sqrt{\frac{3}{5}}$$

(si vedano i calcoli nel II esonero). Quindi

$$\begin{aligned}
\int_0^1 f(x)\,dx &= \frac{1}{2}\int_{-1}^{1} f(\frac{1}{2} + t\frac{1}{2})\,dt = \frac{1}{2}\int_{-1}^{1}\tilde{f}(t)\,dt = \frac{1}{2}\Big(\frac{5}{9}(\tilde{f}(-\xi) + \tilde{f}(\xi)) + \frac{8}{9}\tilde{f}(0) + \tilde{E}\Big) \\
&= \frac{1}{2}\Big(\frac{5}{9}(f(\frac{1}{2} - \xi\frac{1}{2}) + f(\frac{1}{2} + \xi\frac{1}{2})) + \frac{8}{9}f(\frac{1}{2}) + \tilde{E}\Big) \\
&= \frac{1}{18}\Big(5(\frac{1}{1 + \frac{1}{4}\frac{(\sqrt{5}-\sqrt{3})^2}{5}} + \frac{1}{1 + \frac{1}{4}\frac{(\sqrt{5}+\sqrt{3})^2}{5}}) + 8\frac{4}{5}\Big) + E \\
&= \frac{1}{9}(\frac{25 \cdot 28}{181} + \frac{16}{5}) + E = 0.78526\ldots + E\ .
\end{aligned}$$

Poiché $\mathcal{I} = \pi/4 = 0.78539816\ldots$, la formula di Gauss-Legendre a tre nodi risulta più precisa della formula di Newton-Cotes a quattro nodi, nell'approssimazione di $\mathcal{I}$.

ii) Siano $x_i = i/n$, $i = 0, 1, \ldots, n$, e $f(x) = 1/(1 + x^2)$. Allora esistono $\eta_i \in (x_i, x_{i+1})$ ed $\eta \in (0, 1)$ per cui

$$\begin{aligned}\int_0^1 f(x)\,dx &= \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f(x)\,dx = \sum_{i=0}^{n-1} \Big[\frac{1/n}{2}(f(x_i) + f(x_{i+1})) - \frac{1}{12n^3} f''(\eta_i)\Big] \\ &= \frac{1}{n}\Big(\frac{f(0)}{2} + \frac{f(1)}{2} + \sum_{i=1}^{n-1} f(x_i)\Big) - \frac{1}{12n^2} f''(\eta) \\ &= \frac{1}{n}\big(\,\frac{1}{2} + \frac{1}{4} + \sum_{i=1}^{n-1} 1/(1 + (i/n)^2)\,\big) - \frac{1}{12n^2} f''(\eta) = t_{n+1} - \frac{1}{12n^2} f''(\eta).\end{aligned}$$

Cerchiamo una limitazione superiore per l'errore che si commette approssimando $\mathcal{I}$ con la formula composta dei trapezi t_{n+1} appena ottenuta. Poiché la funzione $f'''(x) = 24\frac{x(1-x^2)}{(1+x^2)^4}$ è positiva in $(0, 1)$, la funzione $f''(x) = 2\frac{3x^2-1}{(1+x^2)^3}$ è non decrescente in $[0, 1]$. Inoltre, $f''(0) = -2$, $f''(1) = 1/2$. Dunque per ogni $x \in [0, 1]$ vale la maggiorazione $|\frac{1}{12n^2} f''(x)| \leq \frac{1}{12n^2} 2 = \frac{1}{6n^2}$.

Possiamo dunque concludere che l'errore $|\mathcal{I} - t_{n+1}|$ sarà minore di $\varepsilon > 0$ (piccolo quanto si vuole) se $\frac{1}{6n^2} < \varepsilon$, ovvero se $n > \sqrt{1/(6\varepsilon)}$.

ESERCIZIO 3 Sia A non singolare con autovalori reali positivi. Quindi in particolare A è invertibile.

i) Ci chiediamo per quali valori di $\alpha \in \mathbb{R}$ vale la formula $A^{-1} = \alpha \sum_{k=0}^{+\infty} (I - \alpha A)^k$.

Non vale sicuramente per $\alpha = 0$. Supponiamo $\alpha \neq 0$. Allora dimostrare la formula è equivalente a dimostrare l'identità:

$$\frac{1}{\alpha} A^{-1} = (I - (I - \alpha A))^{-1} = \sum_{k=0}^{+\infty} (I - \alpha A)^k. \qquad (*)$$

Ora, comunque presa una matrice M, si ha che $(I - M)(I + M + M^2 + \ldots + M^{k-1}) = I - M^k$. Se M soddisfa la condizione $M^k \to O$, per $k \to +\infty$, ovvero $\rho(M) < 1$, allora esiste $(I - M)^{-1}$ e $I - M^k \to I$, per $k \to +\infty$. Unendo queste due osservazioni si può dire che se $\rho(M) < 1$ allora

$$\|(I - M)^{-1} - (I + M + M^2 + \ldots + M^{k-1})\| \leq \|(I - M)^{-1}\| \|I - (I - M^k)\| \to 0,$$

cioè per $(I - M)^{-1}$ vale la rappresentazione $(I - M)^{-1} = \lim_{k \to +\infty} \sum_{s=0}^{k-1} M^s = \sum_{s=0}^{+\infty} M^s$.

Ne segue che una condizione sufficiente affinché valga (*) è che $\rho(I - \alpha A)$ sia minore di 1, ovvero che α sia tale che $\max_i |1 - \alpha\lambda_i| < 1$, essendo λ_i gli autovalori di A. Visto che i λ_i sono reali positivi e la scelta di α è ristretta ad $\mathbb{R}$, otteniamo dunque le seguenti equivalenti condizioni su α:

$$-1 < 1 - \alpha\lambda_i < 1,\ \forall i,\ \Leftrightarrow\ 0 < \alpha < 2/\lambda_i,\ \forall i,\ \Leftrightarrow\ 0 < \alpha < 2/\max_i \lambda_i = 2/\rho(A).$$

Riassumendo, se $0 < \alpha < 2/\rho(A)$, allora la formula (*) è sicuramente vera.

ii) Si chiede di dare una limitazione superiore (e inferiore) per gli α di cui al punto i) – ovvero per $2/\rho(A)$ – nel caso in cui A è definita come segue: $A_{ii} = i$, $A_{ij} = A_{ji} = 1/(2^{j-1})$, $i < j$ $(1 \leq i, j \leq n)$.

Si nota innanzitutto che A è reale simmetrica e, quindi, ha autovalori reali. Inoltre, l'i-esimo cerchio di Gershgorin K_i di A ha

$$\text{centro } \ i \ \text{ e raggio } \ \frac{i-1}{2^{i-1}} + \sum_{j=i+1}^{n} \frac{1}{2^{j-1}}$$

$(i = 1, 2, \ldots, n)$. Quindi, per il I teorema di Gershgorin, gli autovalori di A, dovendo stare nell'unione dei cerchi K_i, devono necessariamente essere maggiori o uguali a $(1/2)^{n-1}$, cioè sono positivi. Ne segue che possiamo applicare il risultato del punto i): se α è tale che $0 < \alpha < 2/\rho(A)$, allora l'inversa della nostra matrice A ammette la rappresentazione (*).

Sia $\lambda_{\max}$ un autovalore di A tale che $\lambda_{\max} = \rho(A)$. Notiamo che l'$n-1$-esimo e l'n-esimo cerchio hanno, rispettivamente,

$$\text{centro } \ n-1 \ \text{ e raggio } \ \frac{n-2}{2^{n-2}} + \frac{1}{2^{n-1}},$$

$$\text{centro } \ n \ \text{ e raggio } \ \frac{n-1}{2^{n-1}},$$

e, quindi, hanno intersezione vuota, se

$$1 > \left(\frac{n-2}{2^{n-2}} + \frac{1}{2^{n-1}}\right) + \left(\frac{n-1}{2^{n-1}}\right) = \frac{3n-4}{2^{n-1}},$$

ovvero se $n \geq 5$. Ne segue che per $n \geq 5$, per il II e per il III teorema di Gershgorin (applicabile perché A non avendo zeri è ovviamente irriducibile), $\lambda_{\max}$ appartiene alla parte interna del cerchio K_n, dunque

$$n - (n-1)/(2^{n-1}) < \lambda_{\max} = \rho(A) < n + (n-1)/(2^{n-1}).$$

Riassumendo, per la nostra matrice A si ha

$$\frac{2}{n + (n-1)/(2^{n-1})} < \frac{2}{\rho(A)} < \frac{2}{n - (n-1)/(2^{n-1})},$$

e, dunque, prendendo $\alpha \leq \frac{2}{n+(n-1)/(2^{n-1})}$ siamo sicuri che per l'inversa di A vale la rappresentazione (*). Si noti che la limitazione superiore $\frac{2}{n+(n-1)/(2^{n-1})}$ è facilmente calcolabile, mentre la limitazione superiore $\frac{2}{\rho(A)}$ no.

II ESONERO Analisi Numerica, Matematica - 22 Gennaio 2013

1) Data l'equazione non lineare $x = c/(1+x^2)$, $c > 0$, trovare i valori di c per cui

i) L'equazione ha una unica radice in $\mathbb{R}$.

ii) L'equazione ha una unica radice in $(0, c)$.

iii) La successione $x_{k+1} = c/(1 + x_k^2)$, $k = 0, 1, 2, \ldots$, è convergente per ogni $x_0 \in [0, c]$.

Usare il metodo di Newton per risolvere l'equazione in $\mathbb{R}$, scegliendo $f(x)$ e x_0 in modo opportuno. Studiarne l'ordine di convergenza.

2) Siano $x_0, x_1, \ldots, x_{n-1}$ numeri reali distinti, $p_{n-1}(x) = \sum_{i=0}^{n-1} a_i x^i$, e $y_0, y_1, \ldots, y_{n-1}$ numeri reali qualsiasi.

i) Osservare che le condizioni $p_{n-1}(x_k) = y_k$, $k = 0, 1, \ldots, n-1$, sono equivalenti ad un sistema lineare $A\mathbf{a} = \mathbf{y}$, scrivendo esplicitamente A.

ii) Dimostrare che A, nota come matrice di Vandermonde, è non singolare (sugg. usare la teoria sull'interpolazione vista a lezione).

Siano $\omega = e^{\mathbf{i}2\pi/n}$, $x_k = \omega^k$, $k = 0, 1, \ldots, n-1$, $p_{n-1}(x) = \sum_{i=0}^{n-1} a_i x^i$, $y_0, y_1, \ldots, y_{n-1}$ numeri complessi qualsiasi, e $A\mathbf{a} = \mathbf{y}$ il sistema lineare equivalente alle condizioni $p_{n-1}(x_k) = y_k$, $k = 0, 1, \ldots, n-1$. Dimostrare che la soluzione di tale sistema è $\mathbf{a} = \frac{1}{n} QA\mathbf{y}$, essendo Q la matrice di permutazione con $Q_{ij} = 1$ se $i = j = 1$ e se $i = n - j + 2$ $(1 \le i, j \le n)$.

3) Sia $\mathcal{I} = \int_1^2 \frac{1}{x}\, dx = \log_e 2$.

i) Calcolare le approssimazioni di $\mathcal{I}$ fornite dalle formule di quadratura di Newton-Cotes e Gauss-Legendre a tre nodi e confrontarle con $\mathcal{I}$.

ii) Posto $x_i = 1 + i\frac{1}{2n}$, $i = 0, 1, \ldots, 2n$, scrivere precisamente le approssimazioni t_{2n+1} e s_{2n+1} di $\mathcal{I}$ ottenute con le formule composte del trapezio e di Simpson, definite in termini dei $2n+1$ valori $f(x_i)$, $i = 0, 1, \ldots, 2n$, e $\forall\, \varepsilon > 0$ determinare ν_ε^t e ν_ε^s tali che $\forall\, n > \nu_\varepsilon^t$ e $\forall\, n > \nu_\varepsilon^s$ valgono, rispettivamente, le disuguaglianze

$$|t_{2n+1} - \mathcal{I}| < \varepsilon, \quad |s_{2n+1} - \mathcal{I}| < \varepsilon.$$

Nota: $\int_a^b f(x)\, dx = \frac{b-a}{2}[f(a)+f(b)] - \frac{(b-a)^3}{12} f''(\xi) = \frac{b-a}{6}[f(a)+4f(\frac{a+b}{2})+f(b)] - \frac{1}{90}(\frac{b-a}{2})^5 f''''(\eta)$

ESERCIZIO 1 Sia data l'equazione non lineare

$$x = \frac{c}{1+x^2} =: g(x), \quad c > 0.$$

Disegnando i grafici delle due funzioni $y = x$ e $y = c/(1+x^2)$, $c > 0$, è evidente che essi si incontrano in uno ed in uno solo punto la cui ascissa è positiva. Quindi l'equazione data ha una unica radice reale per ogni $c > 0$ e tale radice è positiva. Quindi la risposta alla domanda i) è: $\forall\, c > 0$.

Inoltre, poiché $g(x)|_{x=0} = c > 0 = x|_{x=0}$ e $g(x)|_{x=c} = c/(1+c^2) < c = x|_{x=c}$, tale radice è nell'intervallo $(0, c)$, $\forall\, c > 0$. Quindi, $\forall\, c > 0$, l'equazione ha una unica radice reale α_c e $\alpha_c \in (0, c)$. Quindi la risposta alla domanda ii) è: $\forall\, c > 0$.

Ora vorremmo che $\forall\, x_0 \in [0, c]$, la successione $x_{k+1} = g(x_k)$, $k = 0, 1, \ldots$, converga ad α_c. Una condizione sufficiente affinché ciò avvenga è che $|g'(x)| < 1$ per ogni $x \in [0, c]$. Allora cerchiamo di imporre questa condizione.

Essendo $g'(x) = -2cx/((1 + x^2)^2)$, si ha $g'(0) = 0$, $-1 = -(1 + c^2)^2/((1 + c^2)^2) < g'(c) = -2c^2/((1 + c^2)^2) < 0\ \forall\, c > 0$, $g'(+\infty) = 0^-$, $g'(x) \leq 0\ \forall\, x \in [0, +\infty)$. Dunque la funzione regolare $g'(x)$ deve avere almeno un punto di minimo globale $x_{\min}$ in $(0, +\infty)$. Trovando $x_{\min}$ (vedendo dove si annulla g'') e imponendo $g'(x_{\min}) > -1$, avremo che $g'(x) > -1$ in $[0, +\infty)$ ed in particolare in $[0, c]$.

Si ha

$$g''(x) = 2c\frac{3x^4 + 2x^2 - 1}{(1 + x^2)^4} = 2c\frac{3x^2 - 1}{(1 + x^2)^3} \begin{cases} < 0 & 0 \leq x < 1/\sqrt{3} \\ = 0 & x = 1/\sqrt{3} \\ > 0 & x > 1/\sqrt{3} \end{cases}.$$

Quindi $x_{\min} = 1/\sqrt{3}$, e $g'(1/\sqrt{3}) = -9c/(8\sqrt{3}) > -1$ se e solo se $c < 8\sqrt{3}/9$.

Riassumendo, se $c < 8\sqrt{3}/9$ allora $-1 < g'(x) \leq 0$ per ogni $x \in [0, +\infty)$ e in particolare per ogni $x \in [0, c]$, quindi $|g'(x)| < 1$, $\forall\, x \in [0, c]$, e siccome in $(0, c)$ c'è una unica radice α_c dell'equazione $x = g(x)$, per un risultato visto a lezione la successione $x_{k+1} = g(x_k)$, $k = 0, 1, \ldots$, deve convergere ad α_c per ogni scelta di $x_0 \in [0, c]$.

Si noti che se $c \geq 8\sqrt{3}/9$, allora $g'(1/\sqrt{3}) \leq -1$ e $1/\sqrt{3} \in (0, c)$, cioè c'è almeno un punto in $(0, c)$, $1/\sqrt{3}$, ove $g'(x) \leq -1$; dunque se $c \geq 8\sqrt{3}/9$, senza ulteriori studi non si può dire che comunque preso $x_0 \in [0, c]$ la successione $x_{k+1} = g(x_k)$, $k = 0, 1, \ldots$, converge.

Quindi l'affermazione iii) è vera per $c \in (0, 8\sqrt{3}/9)$.

Domanda. Per c grande α_c può diventare un punto di repulsione per la successione $x_{k+1} = g(x_k)$? L'affermazione iii) è vera per ogni $c > 0$? Si dimostra con ragionamenti geometrici non ovvi che l'affermazione iii) vale anche per $c \in [8\sqrt{3}/9, 49\sqrt{3}/48)$, e che tale risultato può essere ulteriormente migliorato. Allora ci si pone la seguente domanda: indipendendentemente dal valore di $c > 0$, è vero che se $x_0 \in [0, c]$ allora $x_k \in [c/(1 + c^2), c]\ \forall\, k$. Se si', allora ...

Moltiplicando l'equazione data per $1 + x^2$, si osserva che essa è equivalente all'equazione

$$f(x) := x^3 + x - c = 0.$$

Poiché $f'(x) = 3x^2 + 1 > 0\ \forall\, x \in \mathbb{R}$ e $f''(x) = 6x$, la funzione f è una cubica crescente su tutto $\mathbb{R}$ con un punto di flesso in 0. Quindi, $f''(x)f(x) > 0$ per ogni $x \in (\alpha_c, +\infty)$ dove α_c è la radice di f, che sappiamo essere in $(0, c)$. Ne segue che se, ad esempio, inilizializziamo la successione di Newton

$$x_{k+1} = x_k - f(x_k)/f'(x_k) = x_k - (x_k^3 + x_k - c)/(3x_k^2 + 1),\ k = 0, 1, \ldots$$

con $x_0 \geq c$, otteniamo una successione monotona decrescente $\{x_k\}$ convergente ad α_c con ordine di convergenza almeno quadratico ($f'(\alpha_c) \neq 0!$) indipendentemente dal valore di $c > 0$. In realtà, graficamente risulta che la successione $\{x_k\}$ generata dal metodo di Newton converge ad α_c per ogni scelta di $x_0 \in \mathbb{R}$.

ESERCIZIO 2

i) Si ha $p_{n-1}(x_k) = y_k$, $k = 0, 1, \ldots, n-1$, se e solo se $a_0 + a_1 x_k + a_2 x_k^2 + \ldots + a_{n-1} x_k^{n-1} = y_k$, $k = 0, 1, \ldots, n-1$, se e solo se

$$A\mathbf{a} = \begin{bmatrix} 1 & x_0 & x_0^2 & \cdot & x_0^{n-1} \\ 1 & x_1 & x_1^2 & \cdot & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdot & x_2^{n-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{n-1} & x_{n-1}^2 & \cdot & x_{n-1}^{n-1} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \cdot \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \cdot \\ y_{n-1} \end{bmatrix} = \mathbf{y}.$$

ii) Se la matrice A del precedente sistema lineare fosse singolare, allora il sistema o non avrebbe nessuna soluzione o ne avrebbe infinite. Ma ciò è come dire che o non esiste p_{n-1} (polinomio interpolatore di grado al più $n-1$, tale che $p_{n-1}(x_k) = y_k$, $k = 0, 1, \ldots, n-1$) o ne esistono infiniti. La prima eventualità si verifica solo se almeno due degli x_k, ad es. x_i e x_j, coincidono e $y_i \neq y_j$ (essendo $p_{n-1}(x_i) = p_{n-1}(x_j)$, le condizioni $p_{n-1}(x_i) = y_i$ e $p_{n-1}(x_j) = y_j$, $y_i \neq y_j$, sono incompatibili); la seconda eventualità si verifica solo se almeno due degli x_k, ad es. x_i e x_j, coincidono e $y_i = y_j$ (le condizioni $p_{n-1}(x_i) = y_i$ e $p_{n-1}(x_j) = y_j$ sarebbero identiche, ovvero p_{n-1} dovrebbe soddisfare meno di n condizioni di interpolazione e quindi non sarebbe unico). Ma gli x_k per ipotesi sono diversi tra loro, quindi nessuna delle due eventualtà può verificarsi. Ne segue che la matrice di Vandermonde A deve essere non singolare.

Nelle ipotesi $x_k = \omega^k$, $k = 0, 1, \ldots, n-1$, $\omega = e^{\mathbf{i}2\pi/n}$, $p_{n-1}(x) = \sum_{s=0}^{n-1} a_s x^s$, le condizioni $p_{n-1}(x_k) = y_k$, $y_k \in \mathbb{C}$, sono equivalenti al seguente sistema lineare

$$A\mathbf{a} = \begin{bmatrix} 1 & 1 & 1 & \cdot & 1 \\ 1 & \omega & \omega^2 & \cdot & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \cdot & \omega^{2(n-1)} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \omega^{n-1} & \omega^{(n-1)2} & \cdot & \omega^{(n-1)(n-1)} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \cdot \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \cdot \\ y_{n-1} \end{bmatrix} = \mathbf{y}.$$

Per verificare la tesi, ovvero che la soluzione di tale sistema è $\mathbf{a} = \frac{1}{n} QA\mathbf{y}$, con

$$Q = \begin{bmatrix} 1 & & & \\ & & & 1 \\ & & \iddots & \\ & 1 & & \end{bmatrix}$$

di permutazione, è sufficiente far vedere che $\frac{1}{n} QA^2$ è la matrice identica. Ma, per $i, j = 1, \ldots, n$ si ha

$$[A^2]_{ij} = \sum_{k=1}^{n} a_{ik} a_{kj} = \sum_{k=1}^{n} \omega^{(i-1)(k-1)} \omega^{(k-1)(j-1)} = \sum_{k=1}^{n} (\omega^{(i+j-2)})^{k-1},$$

quindi $[A^2]_{ij} = (1-\omega^{(i+j-2)n})/(1-\omega^{(i+j-2)}) = 0$ se $\omega^{(i+j-2)} \neq 1$ e $[A^2]_{ij} = n$ se $\omega^{(i+j-2)} = 1$, ovvero $[A^2]_{ij} = n$ se $i+j-2 = 0$ oppure $i+j-2 = n$, e $[A^2]_{ij} = 0$ altrimenti. Ne segue che $A^2 = nQ$, da cui l'uguaglianza $\frac{1}{n}QA^2 = Q^2 = I$ (ogni matrice di permutazione ha come inversa la sua trasposta).

ESERCIZIO 3

i) Nel caso dell'esercizio in cui $f(x) = 1/x$ e $[a,b] = [1,2]$, la formula di Simpson (Newton-Cotes a tre nodi) produce il valore

$$\frac{b-a}{6}\Big[f(a) + 4f(\frac{a+b}{2}) + f(b)\Big] = \frac{1}{6}(1 + 4\frac{2}{3} + \frac{1}{2}) = \frac{25}{36} = 0.69\overline{4}.$$

Ricaviamo, per una f generica, la formula di Gauss-Legendre a tre nodi nell'intervallo $[-1,1]$ e poi, da questa, la formula di Gauss-Legendre a tre nodi nell'intervallo $[a,b]$. I tre nodi della prima sono gli zeri del polinomio di Legendre di grado 3 relativo all'intervallo $[-1,1]$,

$$\begin{aligned} p_3(x) &= C_3((x^2-1)^3)^{(3)} = C_3(6x(x^2-1)^2)^{(2)} = C_3(6(x^2-1)^2 + 24x^2(x^2-1))^{(1)} \\ &= C_3(24x(x^2-1) + 48x(x^2-1) + 48x^3) = C_3(120x^3 - 72x) = x^3 - \tfrac{3}{5}x \end{aligned}$$

(C_3 si è scelta in modo che p_3 sia monico), cioè $x_1 = -t$, $x_2 = 0$, $x_3 = t$, $t = \sqrt{3/5}$. Quindi la formula di Gauss-Legendre a tre nodi nell'intervallo $[-1,1]$ dovrà essere del tipo $\int_{-1}^{1} f(x)\,dx = A_1(f(-t) + f(t)) + A_2 f(0) + E$, $t = \sqrt{3/5}$ ($A_3 = A_1$ perché sia il peso che i nodi sono simmetrici rispetto al centro dell'intervallo di integrazione). Sappiamo che, per la particolare scelta dei nodi, imponendo $E = 0$ per f polinomi di grado al più $n-1 = 2$ (n è il numero dei nodi, $n = 3$ nel nostro caso), automaticamente si avrà $E = 0$ per tutti i polinomi di grado al più $2n-1 = 5$. La condizione $E = 0$ per $f(x) = 1$ e $f(x) = x^2$ produce le condizioni $2 = 2A_1 + A_2$, $\frac{2}{3} = A_1\frac{6}{5}$ da cui $A_1 = \frac{5}{9}$, $A_2 = \frac{8}{9}$ (per $f(x) = x$ si ha $0 = 0 + E$, cioè $E = 0$ è vera $\forall\, A_1, A_2$). Possiamo dunque concludere che la formula di Gauss-Legendre a tre nodi nell'intervallo $[-1,1]$ è

$$\int_{-1}^{1} f(x)\,dx = \frac{5}{9}\Big(f(-t) + f(t)\Big) + \frac{8}{9}f(0) + E, \quad t = \sqrt{3/5},$$

e, di conseguenza, quella di Gauss-Legendre a tre nodi nell'intervallo $[a,b]$ è $\int_a^b f(x)\,dx = \frac{b-a}{2}\int_{-1}^{1} f(\frac{a+b}{2} + t\frac{b-a}{2})\,dt = \frac{b-a}{2}\int_{-1}^{1}\tilde{f}(t)\,dt = \frac{b-a}{2}(\frac{5}{9}(\tilde{f}(-t) + \tilde{f}(t)) + \frac{8}{9}\tilde{f}(0) + \tilde{E})$, ovvero

$$\int_a^b f(x)\,dx = \frac{b-a}{2}[\frac{5}{9}\Big(f(\frac{a+b}{2} - t\frac{b-a}{2}) + f(\frac{a+b}{2} + t\frac{b-a}{2})\Big) + \frac{8}{9}f(\frac{a+b}{2})] + E, \quad t = \sqrt{3/5}.$$

Quindi, nel caso dell'esercizio in cui $f(x) = 1/x$ e $[a,b] = [1,2]$, la formula di Gauss-Legendre a tre nodi produce il valore

$$\frac{1}{2}[\frac{5}{9}\Big(1/(\frac{3}{2} - t\frac{1}{2}) + 1/(\frac{3}{2} + t\frac{1}{2})\Big) + \frac{8}{9}\frac{2}{3}] = 0.69312169.. .$$

Quest'ultimo ha due cifre precise in più rispetto al valore fornito dalla formula di Simpson (Nota: $\log_e 2 = 0.6931471806..$).

ii) Posto $f(x) = 1/x$, $x_i = 1 + i\frac{1}{2n}$ e $f_i = f(1 + i\frac{1}{2n}) = 1/(1 + i\frac{1}{2n})$, $i = 0, 1, \ldots, 2n-1, 2n$, utilizzando la Nota, si ottengono le identità

$$\begin{aligned}
\int_1^2 f(x)\,dx &= \sum_{i=1}^{2n} \int_{x_{i-1}}^{x_i} f(x)\,dx = \sum_{i=1}^{2n} \Big[\frac{x_i - x_{i-1}}{2}(f_{i-1} + f_i) - \frac{(x_i - x_{i-1})^3}{12} f''(\xi_i)\Big] \\
&= \frac{1}{2n} \sum_{i=1}^{2n} \frac{1}{2}(f_{i-1} + f_i) - \Big(\frac{1}{2n}\Big)^3 \frac{1}{12} \sum_{i=1}^{2n} f''(\xi_i) \\
&= \frac{1}{2n}\Big[\frac{f_0 + f_{2n}}{2} + \sum_{i=1}^{2n-1} f_i\Big] - \Big(\frac{1}{2n}\Big)^2 \frac{1}{12} f''(\xi) \\
&= t_{2n+1} - \Big(\frac{1}{2n}\Big)^2 \frac{1}{12} f''(\xi), \quad t_{2n+1} = \frac{1}{2n}\Big[\frac{f_0 + f_{2n}}{2} + \sum_{i=1}^{2n-1} f_i\Big],
\end{aligned}$$

$$\begin{aligned}
\int_1^2 f(x)\,dx &= \sum_{i=1}^{n} \int_{x_{2i-2}}^{x_{2i}} f(x)\,dx = \sum_{i=1}^{n} \Big[\frac{x_{2i} - x_{2i-2}}{6}(f_{2i-2} + 4f_{2i-1} + f_{2i}) - \frac{1}{90}(\frac{x_{2i} - x_{2i-2}}{2})^5 f''''(\eta_i)\Big] \\
&= \frac{1}{2n}\frac{1}{3} \sum_{i=1}^{n} (f_{2i-2} + 4f_{2i-1} + f_{2i}) - \frac{1}{90}(\frac{1}{2n})^5 \sum_{i=1}^{n} f''''(\eta_i) \\
&= \frac{1}{2n}\frac{1}{3}[f_0 + f_{2n} + 2\sum_{i=1}^{n-1} f_{2i} + 4\sum_{i=1}^{n} f_{2i-1}] - \frac{1}{180}(\frac{1}{2n})^4 f''''(\eta) \\
&= s_{2n+1} - \Big(\frac{1}{2n}\Big)^4 \frac{1}{180} f''''(\eta), \quad s_{2n+1} = \frac{1}{2n}\frac{1}{3}\Big[f_0 + f_{2n} + 2\sum_{i=1}^{n-1} f_{2i} + 4\sum_{i=1}^{n} f_{2i-1}\Big],
\end{aligned}$$

per opportuni $\xi, \eta \in [1, 2]$, dove t_{2n+1} e s_{2n+1} sono note rispettivamente come formule (di quadratura) composte del trapezio e di Simpson.

Usando le espressioni dell'errore per tali formule di quadratura su visualizzate e maggiorazioni in $[1, 2]$ di f'' e f'''', si può contare il numero di valutazioni della funzione integranda sufficienti affinché l'errore commesso da t_{2n+1} e s_{2n+1} nell'approssimare $\int_1^2 f(x)\,dx = \log_e 2$ sia minore di $\varepsilon > 0$, fissato arbitrariamente piccolo.

Infatti, poiché $f''(x) = 2/(x^3)$ e $f''''(x) = 24/(x^5)$, valgono le maggiorazioni $|f''(x)| \leq 2$, $|f''''(x)| \leq 24$, $\forall\, x \in [1, 2]$. Quindi $\forall\, \varepsilon > 0$, si ha

$$|\int_1^2 f(x)\,dx - t_{2n+1}| < \varepsilon, \quad |\int_1^2 f(x)\,dx - s_{2n+1}| < \varepsilon,$$

se rispettivamente $\frac{1}{6}(\frac{1}{2n})^2 < \varepsilon$ e $\frac{24}{180}(\frac{1}{2n})^4 < \varepsilon$, ovvero se rispettivamente $n > \nu_\varepsilon^t := \lceil (1/(\varepsilon 24))^{1/2} \rceil$ e $n > \nu_\varepsilon^s := \lceil (1/(\varepsilon 120))^{1/4} \rceil$. Questo risultato ci dice in pratica che se alla formula di Simpson composta servono 100 valutazioni della funzione integranda f per ottenere una certa precisione di approssimazione, alla formula del trapezio, per ottenere una precisione simile, ne occorrerebbero circa $2 * 100^2$.

APPENDICE

Osservazioni sull'equazione non lineare $x = c/(1+x^2)$, $c > 0$, successive al secondo esonero di Analisi numerica 1, e precedenti allo scritto d'esame (Gennaio 2013))

Supponiamo $c = 8/(3\sqrt{3})$ ($> 8/(3\sqrt{3})$), ovvero $g'(1/\sqrt{3}) = -1$ (< -1). Ovvero, non è più vero che $\forall\, x \in [0, c]$ vale la disuguaglianza $|g'(x)| < 1$.

Per $c = 8/(3\sqrt{3})$ ($> 8/(3\sqrt{3})$) si ha $g(1/\sqrt{3}) = 2/\sqrt{3}$ ($> 2/\sqrt{3}$).

Calcoliamo l'ascissa β_c del punto di intersezione tra la retta per $(1/\sqrt{3}, g(1/\sqrt{3}) = 3c/4)$ con pendenza $g'(1/\sqrt{3}) = -c3\sqrt{3}/8$ e la bisettrice. Ovviamente, essendo $1/\sqrt{3}$ un punto di flesso per g (cioè, dopo $1/\sqrt{3}$ la g' aumenta), si ha $\alpha_c > \beta_c$ e se imponiamo $g'(\beta_c) \geq -1$ avremo senz'altro che $g'(\alpha_c) > -1$.

Retta richiesta: $y = 3c/4 - (x - 1/\sqrt{3})3c\sqrt{3}/8$. Ascissa del punto di intersezione di essa con la bisettrice $y = x$: $\beta_c = 9c/(8 + c3\sqrt{3})$. Valore di g' in β_c: $-18c^2(8 + c3\sqrt{3})^3/(((8 + c3\sqrt{3})^2 + 81c^2)^2)$. È evidente che per c maggiore di un tot si avrà $g'(\beta_c) < -1$, ma questo non esclude che, invece, per tale c, $g'(\alpha_c) > -1$.

Se imponiamo $g'(\beta_c) = -18c^2(8 + c3\sqrt{3})^3/(((8 + c3\sqrt{3})^2 + 81c^2)^2) \geq -1$, avremo che per $c \in (0, tat)$, si ha $g'(\alpha_c) > -1$ e questo probabilmente con $tat > 8/(3\sqrt{3})$. Ad esempio, per $c = 8/(3\sqrt{3})$ si ha $\beta_c = 3/(2\sqrt{3})$ e $g'(\beta_c) = -c16\sqrt{3}/49$. Quindi, $g'(\beta_c) \geq -1$ se $c \leq 49/(16\sqrt{3})$. In altre parole, si può dire che $g'(\alpha_c) > -1$ anche per $c \in [8/(3\sqrt{3}), 49/(16\sqrt{3})]$ (anche se per c in tale intervallo non è più vero che $|g'(x)| < 1\ \forall\, x \in [0, c]$).

Cambiamo metodo.

A) Si dimostra che se $c > 2$ allora:

$g(g(\frac{1}{2}(c \pm \sqrt{c^2-4}))) = \frac{1}{2}(c \pm \sqrt{c^2-4})$ (Nota: $\frac{1}{2}(c \pm \sqrt{c^2-4})$ sono le radici di $\xi^2 - c\xi + 1 = 0$, quindi *il loro prodotto deve fare* 1). Graficamente l'identità $g(g(\xi)) = \xi$ per $\xi \neq g(\xi)$ corrisponde al fatto che g per $c > 2$ ha un punto di simmetria rispetto alla bisettrice, quindi la successione $x_{k+1} = g(x_k) = c/(1 + x_k^2)$ se parte dall'ascissa di tale punto ritorna a tale ascissa dopo una iterazione. Nota: $\frac{1}{2}(c \pm \sqrt{c^2-4}) \in (0, c)$.

$g'(\frac{1}{2}(c \pm \sqrt{c^2-4})) = -(c \pm \sqrt{c^2-4})/(\frac{1}{2}c^2(c \pm \sqrt{c^2-4}) - c) =: -p^{\pm}$, e $p^+p^- = 4/(c^2)$ che è minore di 1 (cioè $p^+ < 1/p^-$) se $c > 2$. Se $c > 2$, inoltre, allora $p^- > p^+$. Da tale risultato si intuisce graficamente che per $c > 2$ si ha $g'(\alpha_c) < -1$, cioè α_c è un punto di repulsione per la successione $x_{k+1} = g(x_k)$ ($g(x) = c/(1 + x^2)$).

Come conseguenza di queste osservazioni si può dire che:

Se $c > 2$ allora esiste $x_0 \in [0, c]$ tale che la successione $x_{k+1} = g(x_k)$, $k = 0, 1, 2, \ldots$, non converge ad α_c: $x_0 = \frac{1}{2}(c \pm \sqrt{c^2-4})$. Vedi il punto i) dell'esercizio 1 del primo appello di Febbraio di Analisi numerica 1.

Un conto: $g'''(x) = 24cx(1 - x^2)/((1 + x^2)^4)$.

B) Si dimostra che *per $c > 2$ il punto fisso α_c è sempre un punto di repulsione ($g'(\alpha_c) < -1$)*. Vedi il punto ii) dell'esercizio 1 del primo appello di Febbraio di Analisi numerica 1.

Per $0 < c < 2$ il punto fisso α_c è sempre un punto di attrazione ($0 > g'(\alpha_c) > -1$).

Dimostrazione (tipo il caso $c > 2$).

Sia $g(x) = c/(1+x^2)$, $c > 0$. Allora $\{x : 0 < x < 1,\, g'(x) \leq -1\} \subset \{x : g(x) > x\}$.

Infatti, sia $x \in \{x : 0 < x < 1,\, g'(x) \leq -1\}$. Allora $2cx/((1+x^2)^2) \geq 1 \Rightarrow g(x) = c/(1+x^2) \geq (1+x^2)/(2x) > (x^2+x^2)/(2x) = x$.

Se $c < 2$ allora $g(1) = c/2 < 1$, quindi $0 < \alpha_c < 1$ (per la non crescenza di g in $[0, +\infty)$). Quindi, se fosse anche $g'(\alpha_c) \leq -1$, per l'inclusione dimostrata si dovrebbe avere $g(\alpha_c) > \alpha_c$, il che è assurdo. Ne segue che $g'(\alpha_c) > -1$. $\square$

Per $c = 2$ si ha l'equazione $x = g(x) = 2/(1+x^2)$ che ha come radice 1 ($\alpha_2 = 1$). Inoltre, $g(1) = 1$, $g'(1) = -1$, $g''(1/\sqrt{3}) = 0$, $g(1/\sqrt{3}) = 3/2$. Graficamente si vede che la radice $\alpha_c = 1$ dovrebbe essere un punto di attrazione per $x_{k+1} = g(x_k)$.

I LEZIONE

Sia $A \in \mathbb{C}^{m\times n}$, $m \geq n$, e $\mathbf{b} \in \mathbb{C}^m$. Supponiamo che le colonne di A siano linearmente indipendenti. In generale $\mathbf{b}$ non è detto che appartenga allo spazio generato dalle colonne di A. Questo è come dire che il sistema lineare $A\mathbf{x} = \mathbf{b}$ in generale non ha soluzione, se non nel senso dei minimi quadrati che andremo a precisare.

Si cerca $\mathbf{x}$ tale che la misura in norma euclidea del vettore residuo $\mathbf{b} - A\mathbf{x}$ sia minima:

$$\|\mathbf{b} - A\mathbf{x}\|_2^2 = \min_{\mathbf{z}} \|\mathbf{b} - A\mathbf{z}\|_2^2 = \min_{z_j} \sum_{i=1}^{m} |b_i - \sum_{j=1}^{n} a_{ij} z_j|^2.$$

Osserviamo che $\|\mathbf{b} - A\mathbf{z}\|_2^2 = (\mathbf{b} - A\mathbf{z})^H(\mathbf{b} - A\mathbf{z}) = \mathbf{z}^H A^H A\mathbf{z} - \mathbf{z}^H A^H \mathbf{b} - \mathbf{b}^H A\mathbf{z} + \|\mathbf{b}\|^2 = \mathbf{z}^H A^H A\mathbf{z} - 2\Re(\mathbf{z}^H A^H \mathbf{b}) + \|\mathbf{b}\|^2 =: F(\mathbf{z})$, dunque $F(\mathbf{y}+\mathbf{z}) = (\mathbf{y}+\mathbf{z})^H A^H A(\mathbf{y}+\mathbf{z}) - 2\Re((\mathbf{y}+\mathbf{z})^H A^H \mathbf{b}) + \|\mathbf{b}\|^2 = F(\mathbf{y}) - 2\Re(\mathbf{z}^H A^H \mathbf{b}) + \mathbf{z}^H A^H A\mathbf{y} + \mathbf{y}^H A^H A\mathbf{z} + \mathbf{z}^H A^H A\mathbf{z} = F(\mathbf{y}) + \mathbf{z}^H A^H A\mathbf{z} - 2\Re(\mathbf{z}^H A^H(\mathbf{b} - A\mathbf{y}))$. Ne segue che, se $\mathbf{x}$ è l'unico vettore di $\mathbb{C}^n$ tale che $A^H A\mathbf{x} = A^H \mathbf{b}$, allora $F(\mathbf{x}+\mathbf{z}) = F(\mathbf{x}) + \mathbf{z}^H A^H A\mathbf{z} > F(\mathbf{x})$ per ogni $\mathbf{z} \in \mathbb{C}^n$, $\mathbf{z} \neq \mathbf{0}$.

In altre parole, il vettore $\mathbf{x} \in \mathbb{C}^n$ tale che

$$\|\mathbf{b} - A\mathbf{x}\|_2 = \min_{\mathbf{z}\in\mathbb{C}^n} \|\mathbf{b} - A\mathbf{z}\|_2$$

esiste, è unico, e coincide con la soluzione del sistema lineare $A^H A\mathbf{x} = A^H \mathbf{b}$.

Esercizio. Dimostrare che la matrice $A^H A$ è definita positiva.

Esercizio. Nel caso $m = n$ dimostrare che il numero di condizionamento di $A^H A$ nella norma spettrale soddisfa l'identità $\mu_2(A^H A) = \mu_2(A)^2$. Se invece $m > n$, cosa si può dire?

Illustriamo un modo per calcolare $\mathbf{x}$. Supponiamo di conoscere una decomposizione QR di A, ovvero di conoscere una matrice unitaria Q $m \times m$ e una matrice $m \times n$ R triangolare superiore con $R_{ii} \neq 0$, $i = 1, \ldots, n$, e $R_{ij} = 0$, $i > j$, per cui $A = QR$. Si osserva che

$$A^H A\mathbf{x} = A^H \mathbf{b} \;\Leftrightarrow\; (QR)^H QR\mathbf{x} = (QR)^H \mathbf{b} \;\Leftrightarrow\; R^H Q^H QR\mathbf{x} = R^H Q^H \mathbf{b} \;\Leftrightarrow\; R^H R\mathbf{x} = R^H Q^H \mathbf{b}.$$

Chiamiamo ora R' la matrice triangolare superiore $n \times n$ tale che $R = \begin{bmatrix} R' \\ O \end{bmatrix}$, e Q' la matrice $m \times n$ (unitariamente ortonormale per colonne) tale che $Q = [Q' \; Q'']$. Allora

$$\begin{aligned} R^H R\mathbf{x} = R^H Q^H \mathbf{b} \;&\Leftrightarrow\; [(R')^H \; O]\begin{bmatrix} R' \\ O \end{bmatrix}\mathbf{x} = [(R')^H \; O]\begin{bmatrix} (Q')^H \\ (Q'')^H \end{bmatrix}\mathbf{b} \\ &\Leftrightarrow\; (R')^H R'\mathbf{x} = (R')^H (Q')^H \mathbf{b} \;\Leftrightarrow\; R'\mathbf{x} = (Q')^H \mathbf{b}\,, \end{aligned}$$

quindi si ha il seguente

Teorema. Sia $A \in \mathbb{C}^{m\times n}$, $\mathbf{b} \in \mathbb{C}^m$, $m \geq n$. Conoscendo la decomposizione $QR = [Q' \; Q'']\begin{bmatrix} R' \\ O \end{bmatrix}$ di A, il vettore $\mathbf{x}$ per cui $\|\mathbf{b} - A\mathbf{x}\|_2 = \min_{\mathbf{z}} \|\mathbf{b} - A\mathbf{z}\|_2$ esiste, è unico e si può calcolare risolvendo il sistema triangolare superiore $R'\mathbf{x} = (Q')^H\mathbf{b}$. Inoltre, per $\|\mathbf{b} - A\mathbf{x}\|_2$ vale l'espressione $\|\mathbf{b} - A\mathbf{x}\|_2 = \|(Q'')^H\mathbf{b}\|_2$.

Dimostrazione. $A = QR = [Q'\ Q'']\begin{bmatrix} R' \\ O \end{bmatrix}$ implica

$$\begin{aligned}
&\min_{\mathbf{z}} \|\mathbf{b} - A\mathbf{z}\|_2^2 = \min_{\mathbf{z}} \|\mathbf{b} - [Q'\ Q'']\begin{bmatrix} R' \\ O \end{bmatrix}\mathbf{z}\|_2^2 \\
&= \min_{\mathbf{z}} \|\begin{bmatrix} (Q')^H \\ (Q'')^H \end{bmatrix}\mathbf{b} - \begin{bmatrix} R' \\ O \end{bmatrix}\mathbf{z}\|_2^2 = \|(Q'')^H\mathbf{b}\|_2^2
\end{aligned}$$

dove l'ultima uguaglianza vale per $\mathbf{z} = \mathbf{x} := (R')^{-1}(Q')^H\mathbf{b}$. $\square$

Vediamo due possibili procedimenti per calcolare Q ed R tali che $A = QR$.

1) *Procedimento di Gram-Schmidt*. Siano $\mathbf{a}_i \in \mathbb{C}^m$, $i = 1, \ldots, n$, $m \geq n$, le colonne linearmente indipendenti di A, cioè $A = [\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_n]$.

Sia $\mathbf{v}_1 = (1/\|\mathbf{a}_1\|_2)\mathbf{a}_1$.

Sia $\tilde{\mathbf{v}}_2 = \mathbf{a}_2 - r_{12}\mathbf{v}_1$, con r_{12} tale che $\mathbf{v}_1^H\tilde{\mathbf{v}}_2 = \mathbf{v}_1^H(\mathbf{a}_2 - r_{12}\mathbf{v}_1) = 0$, cioè $r_{12} = \mathbf{v}_1^H\mathbf{a}_2$. Sia $\mathbf{v}_2 = (1/\|\tilde{\mathbf{v}}_2\|_2)\tilde{\mathbf{v}}_2$ ($\tilde{\mathbf{v}}_2 \neq \mathbf{0}$ perché altrimenti $\mathbf{a}_2$ sarebbe linearmente dipendente da $\mathbf{a}_1$).

...

Sia $\tilde{\mathbf{v}}_j = \mathbf{a}_j - \sum_{i=1}^{j-1} r_{ij}\mathbf{v}_i$, con r_{ij} tali che $\mathbf{v}_s^H\tilde{\mathbf{v}}_j = \mathbf{v}_s^H(\mathbf{a}_j - \sum_{i=1}^{j-1} r_{ij}\mathbf{v}_i) = 0$, $s = 1, \ldots, j-1$, cioè $r_{sj} = \mathbf{v}_s^H\mathbf{a}_j$. Sia $\mathbf{v}_j = (1/\|\tilde{\mathbf{v}}_j\|_2)\tilde{\mathbf{v}}_j$ ($\tilde{\mathbf{v}}_j \neq \mathbf{0}$ perché altrimenti $\mathbf{a}_j$ sarebbe linearmente dipendente da $\mathbf{a}_1, \ldots, \mathbf{a}_{j-1}$).

...

Riassumendo, posto $r_{11} = \|\mathbf{a}_1\|_2$ e $\mathbf{v}_1 = (1/r_{11})\mathbf{a}_1$, per $j = 2, 3, \ldots, n$ si effettuano le seguenti operazioni:

$$\begin{aligned}
&1)\ r_{ij} = \mathbf{v}_i^H\mathbf{a}_j,\ i = 1, \ldots, j-1, \\
&2)\ \tilde{\mathbf{v}}_j = \mathbf{a}_j - \textstyle\sum_{i=1}^{j-1} r_{ij}\mathbf{v}_i, \\
&3)\ r_{jj} = \|\tilde{\mathbf{v}}_j\|_2\ (\neq 0 \text{ se } \mathbf{a}_1, \ldots, \mathbf{a}_j \text{ indipendenti}), \\
&4)\ \mathbf{v}_j = (1/r_{jj})\tilde{\mathbf{v}}_j.
\end{aligned}$$

Vediamo da un punto di vista matriciale cosa si sta facendo. Valgono le identità $r_{jj}\mathbf{v}_j = \mathbf{a}_j - \sum_{i=1}^{j-1} r_{ij}\mathbf{v}_i$, $j = 1, \ldots, n$, ovvero $\mathbf{a}_j = \sum_{i=1}^{j} r_{ij}\mathbf{v}_i$, $j = 1, \ldots, n$, ovvero

$$\begin{bmatrix} \mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdot & r_{1n} \\ 0 & r_{22} & \cdot & r_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & 0 & r_{nn} \end{bmatrix} = Q'R' = [Q'\ Q'']\begin{bmatrix} R' \\ O \end{bmatrix} = QR$$

dove Q'' si può scegliere arbitrariamente in modo che Q sia unitaria.

2) *Procedimento di Givens* (caso A 3×2). Si vuole triangolarizzare la seguente matrice:

$$A = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}.$$

Si potrebbero utilizzare trasformazioni elementari di Gauss o trasformazioni di Givens:

$$\begin{bmatrix} 1 & 0 \\ -c/a & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ 0 & d - bc/a \end{bmatrix}, \quad \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \sqrt{a^2+c^2} & \alpha b - \beta d \\ 0 & \beta b + \alpha d \end{bmatrix},$$

$\alpha = a/\sqrt{a^2+c^2}$, $\beta = -c/\sqrt{a^2+c^2}$. Il metodo di Gauss richiede $a \neq 0$, si procede solo con Givens. Quindi abbiamo ottenuto che

$$\begin{bmatrix} \alpha & -\beta & 0 \\ \beta & \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix} = \begin{bmatrix} \sqrt{a^2+c^2} & \alpha b - \beta d \\ 0 & \beta b + \alpha d \\ e & f \end{bmatrix},$$

$\alpha = a/\sqrt{a^2+c^2}$, $\beta = -c/\sqrt{a^2+c^2}$. Ora si determinano nuovi α, β, per cui

$$\begin{bmatrix} \alpha & 0 & -\beta \\ 0 & 1 & 0 \\ \beta & 0 & \alpha \end{bmatrix} \begin{bmatrix} \sqrt{a^2+c^2} & b' \\ 0 & d' \\ e & f \end{bmatrix} = \begin{bmatrix} \sqrt{a^2+c^2+e^2} & \alpha b' - \beta f \\ 0 & d' \\ 0 & \beta b' + \alpha f \end{bmatrix},$$

$\alpha = \sqrt{a^2+c^2}/\sqrt{a^2+c^2+e^2}$, $\beta = -e/\sqrt{a^2+c^2+e^2}$. Ora determiniamo nuovi α, β, per cui

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \alpha & -\beta \\ 0 & \beta & \alpha \end{bmatrix} \begin{bmatrix} \sqrt{a^2+c^2+e^2} & \alpha b' - \beta f \\ 0 & d' \\ 0 & f' \end{bmatrix} = \begin{bmatrix} \sqrt{a^2+c^2+e^2} & \alpha b' - \beta f \\ 0 & \sqrt{(d')^2+(f')^2} \\ 0 & 0 \end{bmatrix},$$

$\alpha = d'/\sqrt{(d')^2+(f')^2}$, $\beta = -f'/\sqrt{(d')^2+(f')^2}$. Abbiamo introdotto tre matrici unitarie reali Q_1, Q_2, Q_3 tali che $Q_3Q_2Q_1A$ è triangolare superiore.

Più in generale, data $A = [\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_n]$, $\mathbf{a}_i \in \mathbb{R}^m$, $m \geq n$, si intuisce facilmente come costruire matrici unitarie $Q_i = \{\alpha_i, \beta_i\}$ di Givens $m \times m$ tali che $\ldots Q_3Q_2Q_1A = R$ con R triangolare superiore.

Nota: Nel caso più generale in cui $\mathbf{a}_i \in \mathbb{C}^m$, invece delle rotazioni di Givens reali, si possono utilizzare le rotazioni di Givens complesse o le riflessioni di Householder per raggiungere lo stesso scopo.

Nota: Per l'esistenza della decomposizione QR di A, con R triangolare superiore ed R_{ii} arbitrari, non è necessario che i vettori $\mathbf{a}_j$ siano linearmente indipendenti. Se ad esempio $\mathbf{a}_1$ e $\mathbf{a}_2$ fossero indipendenti, ma $\mathbf{a}_3$ fosse dipendente da $\mathbf{a}_1$ e $\mathbf{a}_2$ allora, dopo aver messo zero l'elemento $3,3$ di R, si passerebbe a triangolarizzare la quarta colonna.

Osservazione sugli algoritmi di triangolarizzazione

Nei metodi diretti per la risoluzione di sistemi lineari, di solito si trasforma il sistema dato $A\mathbf{x} = \mathbf{b}$ nel sistema equivalente $E_{n-1}\cdots E_1A\mathbf{x} = E_{n-1}\cdots E_1\mathbf{b}$, ove $U = E_{n-1}\cdots E_1A$ è triangolare superiore. Se le matrici utilizzate E_i in tale trasformazione sono le matrici elementari di Gauss allora $\|E_i\| > 1$ (vedi l'esercizio seguente), quindi $\|U\|$, essendo $\|U\| \leq \|E_{n-1}\|\cdots\|E_1\|\|A\|$, potrebbe anche essere molto più grande di $\|A\|$. Se le E_i sono invece unitarie, allora $\|U\| = \|A\|$ (perché?). Ne segue che tra gli algoritmi di triangolarizzazione di una matrice (o di risoluzione di un sistema lineare triangolarizzando la matrice dei coefficienti), quelli che utilizzano E_i unitarie sono i più stabili (Givens e Householder migliori di Gauss).

Esercizio. Mostrare che la norma spettrale della matrice $E = \begin{bmatrix} 1 & 0 \\ \mathbf{u} & I \end{bmatrix}$ è maggiore di 1 se $\mathbf{u} \neq \mathbf{0}$.
Risoluzione. $\|E\|_2^2 \geq \|E\mathbf{e}_1\|_2^2 = 1 + \|\mathbf{u}\|_2^2 > 1$.

Esercizio. Siano $A \in \mathbb{C}^{n\times n}$, $\mathbf{z} \in \mathbb{C}^n$. Dimostrare che se $Q \in \mathbb{C}^{n\times n}$ è unitaria allora $\|QA\|_{2,F} = \|AQ\|_{2,F} = \|A\|_{2,F}$, $\|Q\mathbf{z}\|_2 = \|\mathbf{z}\|_2$. Se Q è inoltre tale che $Q\mathbf{z} = \alpha\mathbf{e}_1$, allora $|\alpha| = \|\mathbf{z}\|_2$.

Esercizio. Sia Q unitaria ed $\mathcal{L}$ lo spazio delle matrici della forma QDQ^H, con D diagonale generica. Data $A \in \mathbb{C}^{n\times n}$ scrivere D_A diagonale tale che

$$\|QD_AQ^H - A\|_F = \min_D \|QDQ^H - A\|_F .$$

Dimostrare che se A è hermitiana (definita positiva), allora anche QD_AQ^H è hermitiana (definita positiva).

Esercizio. Sia $\varepsilon \in \mathbb{R}$ tale che $0 < \varepsilon < 1$, e

$$A = \begin{bmatrix} 1+\varepsilon & 1-\varepsilon & 0 \\ 1 & -2 & 1 \\ 0 & \frac{1}{2} & 3\mathbf{i} \end{bmatrix}.$$

Sono vere le seguenti affermazioni?
i) $\rho(A)$ non è autovalore di A.
ii) A ha due autovalori reali e uno complesso non reale.
iii) A è non singolare.

Esercizio. Dimostrare che $\rho(A) \leq \|A\|$ per ogni norma matriciale
Risoluzione. Se $A\mathbf{x} = \lambda\mathbf{x}$ con $\mathbf{x} \neq \mathbf{0}$, allora $AX = \lambda X$, con $X = [\,\mathbf{x}\ \mathbf{0}\ \cdots\ \mathbf{0}\,] \neq O$, dunque $\|A\|\|X\| \geq \|AX\| = \|\lambda X\| = |\lambda|\|X\|$, $\|X\| \neq 0$. Da cui la tesi $|\lambda| \leq \|A\|$.

APPENDICE

Esercizio (relazione con la teoria dei minimi quadrati in approssimazione). Sia $\{\varphi(\mathbf{a},\mathbf{z}) : \mathbf{z} \in \mathbb{C}^n\}$ una classe di funzioni di $\mathbf{a} \in \mathbb{C}^l$. Ad esempio, tale classe puo essere formata da funzioni lineari $\{\sum_{i=1}^n z_i\varphi_i(\mathbf{a}) : z_i \in \mathbb{C}\}$. Supponiamo di avere m dati $(\mathbf{a}_i, y_i)$, $\mathbf{a}_i \in \mathbb{C}^l$, $y_i \in \mathbb{C}$, $i = 1,\dots,m$, $\mathbf{a}_i \neq \mathbf{a}_j$ se $i \neq j$, ottenuti osservando, per $i = 1,\dots,m$, il valore y_i in corrispondenza della situazione $\mathbf{a}_i$ ($m \geq n$).

Cercare $\mathbf{x} \in \mathbb{C}^n$ che realizza il seguente minimo

$$\min_{\mathbf{z}\in\mathbb{C}^n} \sum_{i=1}^m |y_i - \varphi(\mathbf{a}_i,\mathbf{z})|^2, \ \min_{z_i\in\mathbb{C}} \sum_{i=1}^m |y_i - \sum_{s=1}^n z_s\varphi_s(\mathbf{a}_i)|^2$$

vuol dire poter utilizzare $\varphi(\mathbf{a},\mathbf{x})$ o $\sum_{s=1}^n x_s\varphi_s(\mathbf{a})$ per indovinare il valore y prodotto dalla situazione $\mathbf{a}\neq\mathbf{a}_i$.

Esempio: Siano $n=l$ e $\varphi_s(\mathbf{a})=a_s$. Allora

$$\min_{z_i\in\mathbb{C}}\sum_{i=1}^m|y_i-\sum_{s=1}^n z_s\varphi_s(\mathbf{a}_i)|^2=\min_{z_i\in\mathbb{C}}\sum_{i=1}^m|y_i-\sum_{s=1}^n z_s a_{i,s}|^2=\min_{\mathbf{z}\in\mathbb{C}^n}\sum_{i=1}^m|y_i-(A\mathbf{z})_i|^2$$

(sistemi sovradeterminati).

Esempio: Siano $l=1$ e $\varphi_s(a)=a^{s-1}$. Allora

$$\min_{z_i\in\mathbb{C}}\sum_{i=1}^m|y_i-\sum_{s=1}^n z_s\varphi_s(a_i)|^2=\min_{z_i\in\mathbb{C}}\sum_{i=1}^m|y_i-\sum_{s=1}^n z_s a_i^{s-1}|^2$$

(approssimazione ai minimi quadrati con polinomi).

II LEZIONE

Minimi quadrati in approssimazione (caso discreto)

Esercizio. Determinare la retta p di migliore approssimazione nel senso dei minimi quadrati della parabola $f(x)=x^2$ nei punti $-1,0,1$. Confrontarla con la retta di Chebycev di f in $[-1,1]$.
Risoluzione. Visto che f è pari e i punti $-1,0,1$ sono simmetrici rispetto 0, è evidente che $p(x)=c$ per qualche $c\in\mathbb{R}$. Dobbiamo allora minimizzare, al variare di c, la funzione $(p(-1)-f(-1))^2+(p(0)-f(0))^2+(p(1)-f(1))^2=(c-1)^2+c^2+(c-1)^2=2(c-1)^2+c^2=3c^2-4c+2$. Dunque si conclude facilmente che la retta cercata è $p(x)=2/3$. La retta di Chebycev richiesta è per definizione il polinomio p di grado minore o uguale a 1 che minimiza $\max_{x\in[-1,1]}|q(x)-x^2|$ al variare di tutti i polinomi q di grado minore o uguale a 1. Visto che $f(x)=x^2$ è simmetrica rispetto al centro dell'intervallo $[-1,1]$ dove è richiesta la retta, è evidente graficamente che $p(x)=(\min_{x\in[-1,1]}f(x)+\max_{x\in[-1,1]}f(x))/2=1/2$. Quindi i due criteri di approssimazione, minmax e minimi quadrati, danno due risultati diversi.

Discussione su tanti modi per approssimare (minimi quadrati, minmax, interpolazione; dati discreti o continui), con diverse classi di funzioni ($\varphi(x,c_0,c_1,\ldots,c_n)$ o $\sum_{i=0}^n c_i\varphi_i(x)$, polinomi, polinomi trigonometrici, funzioni raxionali, Pade', e^A con Pade, ...).

Esercizio. Determinare la parabola p di migliore approssimazione nel senso dei minimi quadrati della funzione $f(x)=|x|$ nei punti $-1,-1/2,0,1/2,1$. Confrontarla con la parabola di migliore approssimazione minmax di f in $[-1,1]$.
Risoluzione. Visto che f è pari e i punti $-1,-1/2,0,1/2,1$ sono simmetrici rispetto 0, è evidente che $p(x)=a+bx^2$ per $a,b\in\mathbb{R}$. Dobbiamo allora minimizzare, al variare di a,b, la funzione $(p(-1)-f(-1))^2+(p(-1/2)-f(-1/2))^2+(p(0)-f(0))^2+(p(1/2)-f(1/2))^2+(p(1)-f(1))^2=(p(0)-f(0))^2+2(p(1/2)-f(1/2))^2+2(p(1)-f(1))^2=a^2+2(a+b/4-$

$1/2)^2 + 2(a+b-1)^2$. Annullando le derivate parziali rispetto ad a e b, si ottengono le condizioni $5a + 3b/4 = 3/2$, $5a + 17b/4 = 9/2$. Dunque si conclude che la parabola cercata è $p(x) = 6/35 + 6x^2/7$.

Teoria dei minimi quadrati per polinomi, base standard, basi opportune e ortogonali

Fissati $m+1$ numeri reali distinti x_i, $i = 0, 1, \ldots, m$, ove sono noti i valori di una funzione f, si vuole minimizzare

$$\sum_{i=0}^{m} (q_n(x_i) - f(x_i))^2 =: \|q_n(\mathbf{x}) - f(\mathbf{x})\|_2^2$$

al variare di q_n tra i polinomi di grado al più n. Posto $q_n(x) = \sum_{j=0}^{n} a_j x^j$, occorre minimizzare

$$\begin{aligned} F(\mathbf{a}) &= \sum_{i=0}^{m} \Big(\sum_{j=0}^{n} a_j x_i^j - f(x_i) \Big)^2 = \sum_{i=0}^{m} \Big(\sum_{j,k=0}^{n} a_j a_k x_i^{j+k} + f(x_i)^2 - 2 f(x_i) \sum_{j=0}^{n} a_j x_i^j \Big) \\ &= \sum_{j,k=0}^{n} a_j a_k \sum_{i=0}^{m} x_i^{j+k} + \sum_{i=0}^{m} f(x_i)^2 - 2 f(x_i) \sum_{j=0}^{n} a_j \sum_{i=0}^{m} f(x_i) x_i^j \\ &= \mathbf{a}^T A \mathbf{a} - 2\mathbf{a}^T \mathbf{b} + c, \;\; A_{ij} = \sum_{i=0}^{m} x_i^{j+k}, \;\; b_j = \sum_{i=0}^{m} f(x_i) x_i^j, \;\; c = \sum_{i=0}^{m} f(x_i)^2, \;\; j = 0, 1, \ldots, n, \end{aligned}$$

Risultato: $F(A^{-1}\mathbf{b}) < F(\mathbf{a})$, $\forall\, \mathbf{a} \in \mathbb{R}^{n+1}$. Infatti, essendo $\nabla F(\mathbf{a}) = 2(A\mathbf{a} - \mathbf{b})$, il vettore $A^{-1}\mathbf{b}$ è l'unico punto estremale per F. Inoltre, essendo $\nabla^2 F(\mathbf{a}) = 2A$ costante e definita positiva (perché?), $A^{-1}\mathbf{b}$ deve essere un punto di minimo assoluto globale.

Dunque il polinomio cercato è $p_n(x) = \sum_{j=0}^{n} a_j x^j$ con $\mathbf{a} = A^{-1}\mathbf{b}$. Notare che $\|p_n(\mathbf{x})\|_2^2 = \sum_{i=0}^{m} p_n(x_i)^2 = \sum_{i=0}^{m} (\sum_{j=0}^{n} a_j x_i^j)^2 = \sum_{k,j=0}^{n} a_k a_j \sum_{i=0}^{m} x_i^{j+k} = \mathbf{a}^T A \mathbf{a} = \mathbf{b}^T A^{-1} \mathbf{b}$.

Esercizio. Dimostrare che $F(\mathbf{a} + \mathbf{d}) = F(\mathbf{a}) + 2\mathbf{d}^T (A\mathbf{a} - \mathbf{b}) + \mathbf{d}^T A \mathbf{d}$, $\forall\, \mathbf{a}, \mathbf{d} \mathbb{R}^{n+1}$. Dunque $F(A^{-1}\mathbf{b}) < F(\mathbf{x})$ per ogni $\mathbf{x} \neq A^{-1}\mathbf{b}$. Inoltre, osservare che $\|p_n(\mathbf{x}) - f(\mathbf{x})\|_2^2 = F(A^{-1}\mathbf{b}) = c - \mathbf{b}^T A^{-1} \mathbf{b} = \|f(\mathbf{x})\|^2 - \|p_n(\mathbf{x})\|_2^2$.

Osservazione importante Se i nodi x_i sono distribuiti uniformemente in $[0, 1]$, allora si può dire che

$$A_{jk} \Delta x = \Delta x \sum_{i=0}^{m} x_i^{j+k} \approx \int_0^1 x^{j+k}\, dx = \frac{1}{j+k+1}, \;\; 0 \leq j, k \leq n\,.$$

In altre parole, la matrice A del sistema lineare che occorre risolvere per calcolare il polinomio di migliore approssimazione nel senso dei minimi quadrati della tabella $(x_i, f(x_i))$, $i = 0, 1, \ldots, m$, è a meno di un multiplo vicina alla matrice di Hilbert, la quale, come è ben noto, ha un numero di condizionamento che cresce molto velocemente con m.

Minimi quadrati con dati continui: cercare l'unico polinomio $p_n(x)$ tale che $\int_a^b (f(x) - p_n(x))^2\, dx \leq \int_a^b (f(x) - q_n(x))^2\, dx$ per ogni polinomio q_n di grado minore o uguale ad n.

Esercizio. Se f è simmetrica (antisimmetrica) in $[a,b]$, anche anche p_n è simmetrica (antisimmetrica) in $[a,b]$.

Interpolazione, formula di Lagrange

Dati $n+1$ numeri reali distinti x_0, x_1, ..., x_n ed $n+1$ numeri reali qualsiasi y_0, y_1, ..., y_n, si vuole determinare un polinomio p tale che $p(x_i) = y_i$, $i = 0, 1, \ldots, n$, ovvero un polinomio interpolante la tabella τ_{n+1} di $n+1$ dati (x_i, y_i), $i = 1, \ldots, n$.

Prima di verificare la effettiva esistenza di un polinomio p interpolante la tabella τ_{n+1}, osserviamo che se p esistesse, allora ogni altro polinomio

$$\tilde{p}(x) = p(x) + q(x) \prod_{i=0}^{n} (x - x_i) \tag{$*$}$$

interpolerebbe τ_{n+1}, e, viceversa, ogni polinomio $\tilde{p}$ interpolante τ_{n+1} dovrebbe ammettere la rappresentazione (*). Da questo ragionamento segue che si può avere l'unicità del polinomio interpolatore richiedendo che il suo grado non superi n (in questo caso, infatti, $q(x)$, nella rappresentazione, dovrà essere per forza identicamente zero).

A questo punto se esibiamo un polinomio di grado al più n interpolante la tabella τ_{n+1}, automaticamente avremo dimostrato l'esistenza e l'unicità di un polinomio con tali caratteristiche.

Cominciamo con l'introdurre i seguenti polinomi di Lagrange associati ai nodi x_i:

$$l_i(x) = \prod_{j=0,\, j \neq i}^{n} \frac{x - x_j}{x_i - x_j}.$$

Ciascun l_i ha grado esattamente n, in x_i vale 1 e in x_j, $j \neq i$, vale zero. I polinomi l_i costituiscono una base per l'insieme dei polinomi q_n di grado al più n, infatti per tali q_n vale ovviamente la seguente rappresentazione: $q_n(x) = \sum_{i=0}^{n} q_n(x_i) l_i(x)$ (perché?).

Esercizio. Rappresentare il polinomio 1, di grado minore o uguale di n, utilizzando la base per l'insieme dei polinomi di grado minore o uguale ad n di Lagrange associata a $n+1$ numeri distinti x_i: $\{l_0(x), l_1(x), \ldots, l_n(x)\}$.

Il seguente polinomio

$$p(x) = \sum_{i=0}^{n} y_i l_i(x), \quad l_i(x) = \prod_{j=0,\, j \neq i}^{n} \frac{x - x_j}{x_i - x_j}$$

ha grado al più n (essendo combinazione lineare di polinomi di grado esattamente n) e $p(x_j) = y_j$, $j = 0, 1, \ldots, n$. Quindi, abbiamo dimostrato che esiste ed è unico un polinomio $p(x)$ interpolante τ_{n+1} e tale polinomio ha grado al più n.

È chiaro che $p(x)$ potrebbe avere effettivamente grado inferiore ad n, ad esempio se la tabella è formata da dati allineati allora il polinomio interpolatore sarà ovviamente una retta.

Esercizio. Dati $x_0 < x_1 < \ldots < x_{n-1} < x_n$ scrivere in termini del polinomio $p(x)$ interpolante τ_{n+1} il polinomio $\tilde{p}$ di grado minimo tale che $\tilde{p}(x_j) = y_j$, $j = 0, 1, \ldots, n$, $\tilde{p}'(x_0) = y_0'$, $\tilde{p}'(x_n) = y_n'$.
Risoluzione. Sappiamo che deve essere $\tilde{p}(x) = p(x) + c(x) \prod_{i=0}^{n}(x - x_i)$. Poiché $\tilde{p}'(x) = p'(x) + c'(x) \prod_{i=0}^{n}(x - x_i) + c(x) \sum_{j=0}^{n} \prod_{i=0,\, i\neq j}^{n}(x - x_i)$, si ha

$$\tilde{p}'(x_k) = p'(x_k) + c(x_k) \prod_{i=0,\, i\neq k}^{n} (x_k - x_i).$$

Dunque la condizione $\tilde{p}'(x_k) = y_k'$ è soddisfatta se $c(x_k) = (y_k' - p'(x_k))/ \prod_{i=0,\, i\neq k}^{n}(x_k - x_i)$. Allora basta porre

$$c(x) = (y_0' - p'(x_0))/ \prod_{i=1}^{n}(x_0 - x_i)\frac{x - x_n}{x_0 - x_n} + (y_n' - p'(x_n))/ \prod_{i=0}^{n-1}(x_n - x_i)\frac{x - x_0}{x_n - x_0}$$

e le condizioni $\tilde{p}'(x_0) = y_0'$, $\tilde{p}'(x_n) = y_n'$ sono soddisfatte.

III LEZIONE

Polinomi densi nelle funzioni continue (Weierstrass)
Sia $f \in C^0([0, 1])$. Comunque fissato $\varepsilon > 0$ esiste $\nu_\varepsilon \in \mathbb{N}$ tale che $\forall\, n > \nu_\varepsilon$

$$\|f - B_n(\cdot; f)\|_{\infty,[0,1]} < \varepsilon,\ B_n(x; f) = \sum_{k=0}^{n} f(\frac{k}{n})[\binom{n}{k} x^k(1 - x)^{n-k}]$$

($\|g\|_{\infty,[a,b]} = \max_{x\in[a,b]} |g(x)|$). Nota: $f(x) = \sum_{k=0}^{n} f(x)\binom{n}{k} x^k(1 - x)^{n-k}$.

Senza dimostrazione.

Polinomio minmax di funzioni continue
Sia $f \in C^0([a, b])$. Esiste ed è unico p_n polinomio di grado minore o uguale ad n tale che

$$\gamma = \|f - p_n\|_{\infty,[a,b]} = \min_{g_n,\, \partial g_n \leq n} \|f - g_n\|_{\infty,[a,b]}.$$

Inoltre, esistono almeno $n + 2$ punti $x_0 < x_1 < \ldots < x_n < x_{n+1}$, $x_i \in [a, b]$ per cui $f(x_i) - p_n(x_i) = (-1)^i\gamma$ (oppure $(-1)^{i+1}\gamma$).

Senza dimostrazione.

Esercizio. Se $f \in C^0([a, b])$ è simmetrica (antisimmetrica) rispetto ad $(a + b)/2$, allora p_n il polinomio di migliore approssimazione minmax di f in $[a, b]$ di grado al più n deve essere simmetrico (antisimmetrico) rispetto ad $(a + b)/2$.

Esercizio (retta di Chebycev). Calcolare la retta di migliore approssimazione minmax di $g(x) = 2/(1 + x^2)$ in $[-1, 1]$. Scrivere l'equazione della retta di migliore approssimazione minmax di $g(x)$ regolare convessa (concava) in $[a, b]$.

Polinomio monico di norma sup minima in $[-1,1]$ *(Chebycev)*

Proposizione. Se p_{n-1} è il polinomio di grado minore o uguale di $n-1$ di migliore approssimazione minmax di x^n in $[-1,1]$, cioè

$$\mu = \|x^n - p_{n-1}(x)\|_{\infty,[-1,1]} = \min_{q_{n-1},\, \partial q_{n-1} \leq n-1} \|x^n - q_{n-1}(x)\|_{\infty,[-1,1]},$$

allora $x^n - p_{n-1}(x) = \frac{1}{2^{n-1}} T_n(x)$, dove $T_n(x)$ è l'n-esimo polinomio di Chebycev.

Prima di dimostrare la Proposizione, vediamo un sua immediata conseguenza:

Corollario.

$$\min \left\{ \max_{x\in[-1,1]} |q_n(x)| : \ q_n \text{ monico di grado } n \right\} = \max_{x\in[-1,1]} |\frac{1}{2^{n-1}} T_n(x)| = \frac{1}{2^{n-1}}.$$

Dimostrazione (della Proposizione). Sappiamo che il polinomio p_{n-1} esiste ed è unico. Sia $y(x) = x^n - p_{n-1}(x)$. Sappiamo che ci sono almeno $n+1$ x_i, con $x_0 < x_1 < \ldots < x_n$, $x_i \in [-1,1]$, tali che $y(x_i) = (-1)^i \mu$ $[(-1)^{i+1}\mu]$, $i = 0, 1, \ldots, n$. Si nota che $y'(x)$ è un polinomio di grado $n-1$ e che ha $n-1$ zeri reali distinti, ovvero $x_1, x_2, \ldots, x_{n-1}$. Se x_0 fosse maggiore di -1, y' avrebbe almeno n zeri reali distinti e quindi dovrebbe essere nullo, il che non può essere. Dunque $x_0 = -1$. Per un ragionamento analogo si ha $x_n = 1$. Si consideri ora il polinomio di grado $2n$ $y(x)^2 - \mu^2$. Esso si annulla negli x_i, $i = 0, 1, \ldots, n$, e, poiché la sua derivata, $2y'(x)y(x)$, si annulla negli x_i, $i = 1, \ldots, n-1$, questi ultimi sono zeri doppi. Abbiamo dunque tutti gli zeri di $y(x)^2 - \mu^2$, ne segue che esiste una costante c per cui

$$y(x)^2 - \mu^2 = c(x^2 - 1) y'(x)^2.$$

La costante c si ottiene imponendo che il coefficiente di x^{2n} del secondo membro sia 1. Dunque $c = 1/n^2$.

$$\frac{n^2}{1-x^2} = \frac{y'(x)^2}{\mu^2 - y(x)^2}, \quad \frac{n}{\sqrt{1-x^2}} = \pm \frac{y'(x)}{\mu\sqrt{1-(y(x)/\mu)^2}}, \quad n \arccos x + c = \pm \arccos \frac{y(x)}{\mu},$$

da cui $y(x) = \mu \cos(n \arccos x + c)$, $\pm\mu = 1 - p_{n-1}(1) = \mu \cos c$. Dunque $c = k\pi$, $k \in \mathbb{Z}$. Riassumendo, se p_{n-1} è il polinomio di migliore approssimazione minmax di x^n in $[-1,1]$, allora

$$x^n - p_{n-1}(x) = y(x) = \pm\mu \cos(n \arccos x)$$

(l'uguaglianza si è ottenuta supponendo $x^2 < 1$).

Prima di concludere ora occorre fare una osservazione. Si considerino i seguenti polinomi $T_0(x) = 1$, $T_1(x) = x$, $T_2(x) = 2xT_1(x) - T_0(x) = 2x^2 - 1$, $T_{i+1}(x) = 2xT_i(x) - T_{i-1}(x)$, $i = 1, 2, \ldots$, ovviamente ben definiti su tutto $\mathbb{R}$. Si noti che il coefficiente di x^n di $T_n(x)$ è 2^{n-1}, cioè $T_n(x) = 2^{n-1}x^n + \ldots$.

Esercizio. Trovare gli zeri e gli estremali di T_n.

Risoluzione. $\cos(n \arccos x) = 0$ sse $n \arccos x = \pi/2 + k\pi$ sse $x = \cos((2k+1)\pi/2n)$, $k = 0, 1, \ldots, n-1$. $\sin(n \arccos x)/\sqrt{1-x^2} = 0$ sse $n \arccos x = k\pi$ sse $x = \cos(k\pi/n)$, $k = 1, \ldots, n-1$.

Si consideri poi la funzione $c_n(x) = \cos(n \arccos x)$, ben definita $\forall\, x \in [-1, 1]$. Si noti che $c_0(x) = 1$, $c_1(x) = x$, $c_2(x) = 2x^2 - 1$, e, poiché $c_{n+m}(x) + c_{|n-m|}(x) = 2c_n(x)c_m(x)$, vale anche l'uguaglianza $c_{n+1}(x) = 2c_n(x)x - c_{n-1}(x)$. Ne segue l'osservazione: $c_n(x) = \cos(n \arccos x) = T_n(x)|_{[-1,1]}$.

Dunque, in $(-1, 1)$ si ha $x^n - p_{n-1}(x) = \pm\mu c_n(x) = \pm\mu T_n(x)|_{(-1,1)}$ ed ora è evidente che in questa identità, per essere vera, si deve scegliere il segno $+$ e si deve porre $\mu = 1/2^{n-1}$.

Calcolare il polinomio interpolatore in un punto (Aitken-Neville)

Indichiamo con $I_{01\ldots n}(x)$ il polinomio interpolatore la tabella (x_i, y_i), $i = 0, 1, \ldots, n$. Allora

$$I_{01\ldots n}(x) = \frac{1}{x_k - x_h} \det \begin{bmatrix} I_{0\ldots h-1\, h+1\ldots n}(x) & x - x_k \\ I_{0\ldots k-1\, k+1\ldots n}(x) & x - x_h \end{bmatrix} = \frac{1}{x_n - x_0} \det \begin{bmatrix} I_{12\ldots n}(x) & x - x_n \\ I_{01\ldots n-1}(x) & x - x_0 \end{bmatrix}.$$

Esempio. Si vuole calcolare in x il polinomio $I_{0123}(x)$ interpolante la tabella $(0, -1)$, $(1/3, 0)$, $(2/3, 0)$, $(1, 2)$:

$$I_0 = -1, \quad I_1 = 0, \quad I_2 = 0, \quad I_3 = 2,$$

$$I_{01}(x) = \frac{1}{x_1 - x_0} \det \begin{bmatrix} I_1(x) & x - x_1 \\ I_0(x) & x - x0 \end{bmatrix} = \frac{1}{h}(x - x_1),$$

$$I_{12}(x) = \frac{1}{x_2 - x_1} \det \begin{bmatrix} I_2(x) & x - x_2 \\ I_1(x) & x - x_1 \end{bmatrix} = 0,$$

$$I_{23}(x) = \frac{1}{x_3 - x_2} \det \begin{bmatrix} I_3(x) & x - x_3 \\ I_2(x) & x - x_2 \end{bmatrix} = \frac{1}{h} 2(x - x_2),$$

$$I_{012}(x) = \frac{1}{x_2 - x_0} \det \begin{bmatrix} I_{12}(x) & x - x_2 \\ I_{01}(x) & x - x_0 \end{bmatrix} = \frac{1}{2h}(-\frac{1}{h}(x - x_1)(x - x_2)),$$

$$I_{123}(x) = \frac{1}{x_3 - x_1} \det \begin{bmatrix} I_{23}(x) & x - x_3 \\ I_{12}(x) & x - x_1 \end{bmatrix} = \frac{1}{2h}(\frac{2}{h}(x - x_2)(x - x_1)),$$

$$\begin{aligned} I_{0123}(x) &= \frac{1}{x_3 - x_0} \det \begin{bmatrix} I_{123}(x) & x - x_3 \\ I_{012}(x) & x - x_0 \end{bmatrix} \\ &= \frac{1}{3h}\Big[\frac{1}{h^2}(x - x_2)(x - x_1)(x - x_0) + \frac{1}{2h^2}(x - x_1)(x - x_2)(x - x_3)\Big] \\ &= 9(x - \frac{2}{3})(x - \frac{1}{3})x + \frac{9}{2}(x - \frac{1}{3})(x - \frac{2}{3})(x - 1) \\ &= 9(x - \frac{1}{3})(x - \frac{2}{3})(\frac{3}{2}x - \frac{1}{2}) = \frac{27}{2}(x - \frac{1}{3})^2(x - \frac{2}{3}). \end{aligned}$$

Si verifica che il polinomio $I_{0123}(x)$ effettivamente per $x = 0, 1/3, 2/3, 1$ assume i valori -1, 0, 0, 2. Si indovina il grafico di $I_{0123}(x)$ osservando che è una cubica con derivata nulla per $x = 1/3$.

Esercizio. Data la funzione $g(x) = c/(1 + x^2)$, $c > 0$, si vuole trovare la parabola di migliore approssimazione minmax di g in $\mathcal{I} = [-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}]$ (si noti che $\pm 1/\sqrt{3}$ sono punti di flesso per g, quindi, effettivamente, g nell'intervallo $\mathcal{I}$ somiglia molto a una parabola e ha quindi senso cercare di approssimarla con una parabola).

Risoluzione. Cominciamo con l'osservare che la parabola che cerchiamo deve essere una funzione pari, come la g (vedi l'Osservazione). Quindi sarà del tipo $\alpha x^2+\beta$. Inoltre, sappiamo che $g(x) - (\alpha x^2 + \beta)$ deve alternare in $\mathcal{I}$ il suo valore massimo almeno quattro volte, e per la parità e per la concavità sia di g che di $\alpha x^2 + \beta$ (ci si aspetta ovviamente che α venga minore di zero) si può indovinare che le volte saranno cinque. Più precisamente, possiamo dire che devono valere le seguenti identità:

$$g(x_i) - (\alpha x_i^2 + \beta) = (-1)^i \mu \ \ [(-1)^{i+1}\mu], \quad \mu = \max_{x\in\mathcal{I}} |g(x) - (\alpha x^2 + \beta)|, \quad i = 0, 1, 2, 3, 4,$$

con $x_0 = -1/\sqrt{3}$, $x_1 = -\xi$, $x_2 = 0$, $x_3 = \xi$, $x_4 = 1/\sqrt{3}$. Dobbiamo dunque determinare $\alpha < 0, \beta > 0, \mu > 0$, e $\xi \in (0, 1/\sqrt{3})$ che soddisfano le identità suddette.

Devono essere verificate le uguaglianze

$$g(0) - (\alpha 0^2 + \beta) = g(1/\sqrt{3}) - (\alpha(1/\sqrt{3})^2 + \beta), \quad (\alpha 0^2 + \beta) - g(0) = g(\xi) - (\alpha\xi^2 + \beta)$$

e, nello stesso tempo, $|(\alpha\xi^2 + \beta) - g(\xi)|$ deve essere maggiore o uguale di ogni altro $|(\alpha x^2 + \beta) - g(x)|$, $x \in [0, 1/\sqrt{3}]$.

La prima uguaglianza diventa $c - \beta = \frac{3}{4}c - \frac{\alpha}{3} - \beta$, da cui $\alpha = -3c/4$. Quindi la parabola che stiamo cercando sarà del tipo $-3cx^2/4 + \beta$. La seconda uguaglianza diventa $\beta - c = c/(1 + \xi^2) + 3c\xi^2/4 - \beta$, da cui la seguente formula per β:

$$\beta = \frac{1}{2}c\Big(1 + 1/(1 + \xi^2) + 3\xi^2/4\Big).$$

Come abbiamo detto, nello stesso tempo, la quantità

$$c/(1+\xi^2)+3c\xi^2/4-\beta = -\frac{1}{2}c+\frac{1}{2}c/(1+\xi^2)+\frac{3}{8}c\xi^2 = \frac{1}{2}c\Big(-1+1/(1+\xi^2)+\frac{3}{4}\xi^2\Big) = \frac{1}{2}c\varphi(\xi)$$

deve essere in modulo massima (in $[0, 1/\sqrt{3}]$). Poiché $\varphi(0) = \varphi(1/\sqrt{3}) = 0$, e $\varphi'(\xi) = \xi(3\xi^4+6\xi^2-1)/(2(1+\xi^2)^2)$ è negativa per gli ξ positivi per cui $0 < \xi^2 < -1+\frac{2}{3}\sqrt{3}$ e positiva per gli ξ positivi per cui $\xi^2 > -1 + \frac{2}{3}\sqrt{3}$, la funzione φ assume valori negativi in $(0, 1/\sqrt{3})$ ed è minima (ovvero, il suo modulo è massimo) per $\xi = \sqrt{-1 + \frac{2}{3}\sqrt{3}}$.

Abbiamo quindi trovato il punto ξ incognito, di alternanza per la parabola minmax di g in $\mathcal{I}$: $\xi = \sqrt{-1 + \frac{2}{3}\sqrt{3}}$. Sostituendo tale valore nell'espressione di cui sopra di β in termini di ξ, otteniamo il coefficiente β della parabola minmax di g in $\mathcal{I}$: $\beta = \frac{1}{2}c(\sqrt{3} + \frac{1}{4})$.

Riassumendo, la parabola di migliore approssimazione minmax di g in $\mathcal{I}$ è

$$-3cx^2/4+\frac{1}{2}c(\sqrt{3}+\frac{1}{4}),$$

i punti di alternanza sono $x_0=-1/\sqrt{3}$, $x_1=-\sqrt{-1+\frac{2}{3}\sqrt{3}}$, $x_2=0$, $x_3=\sqrt{-1+\frac{2}{3}\sqrt{3}}$, $x_4=1/\sqrt{3}$, e poiché $\beta-c=\frac{1}{2}c(\sqrt{3}-\frac{7}{4})<0$, il valore μ alternato è $\frac{1}{2}c(\frac{7}{4}-\sqrt{3})$ e in zero la parabola minmax e sotto la g.

IV LEZIONE

Si disegna la cubica interpolante la tabella $(0,-1)$, $(1/3,0)$, $(2/3,0)$, $(1,2)$, ottenuta nella precedente lezione, notando che se i dati $-1,0,0,2$ della tabella fossero imprecisi tale cubica molto probabilmente non rifletterebbe l'andamento della funzione che si vuole approssimare (conoscendo i suoi valori in $0,1/3,2/3,1$).

Si dimostra la formula di Aitken-Neville, che, data la tabella (x_i,y_i), $i=0,1,\dots,n$, esprime il polinomio $I_{01\dots n}(x)$ interpolante la tabella negli $n+1$ punti x_i, $i=0,1,\dots,n$, in termini del polinomio $I_{01\dots h-1\,h+1\dots n}(x)$ interpolante la tabella sugli n punti $x_0,x_1,\dots,x_{h-1},x_{h+1},\dots,x_n$ e del polinomio $I_{01\dots k-1\,k+1\dots n}(x)$ interpolante la tabella sugli n punti $x_0,x_1,\dots,x_{k-1},x_{k+1},\dots,x_n$.

La dmostrazione consiste nell'osservare che l'espressione di Aitken-Neville è un polinomio di grado al piu n che nei nodi x_j vale y_j come $I_{01\dots n}(x)$ (scrivere i dettagli!).

Minimi quadrati con dati discreti in spazi di funzioni lineari

Dati: $f(x_0)$, $f(x_1)$, ..., $f(x_m)$, con x_i reali distinti.

Classe di funzioni: $a_0\varphi_0(x)+a_1\varphi_1(x)+\ldots+a_n\varphi_n(x)$, con $\varphi_i(x)$ linearmente indipendenti.

Ipotesi: $m\geq n$.

Allora la seguente somma di quadrati

$$\sum_{r=0}^{m}(\sum_{j=0}^{n}a_j\varphi_j(x_r)-f(x_r))^2,\ a_i\in\mathbb{R},$$

è minima se gli a_i risolvono il seguente sistema lineare:

$$(\sum_{r=0}^{m}\varphi_i(x_r)\varphi_j(x_r))_{i,j=0}^{n}(a_j)_{j=0}^{n}=(\sum_{r=0}^{m}f(x_r)\varphi_i(x_r))_{i=0}^{n}$$

(stessi passaggi del caso $\varphi_j(x)=x^j$). Si osserva che, per la lineare indipendenza delle funzioni φ_j, la matrice dei coefficienti di tale sistema è definita positiva (dimostrarlo!).

Inoltre,

$$\sum_{r=0}^{m}\varphi_i(x_r)\varphi_j(x_r)=\begin{cases}1 & i=j\\ 0 & i\neq j\end{cases},\quad 0\leq i,j\leq n,\quad\Rightarrow\quad a_i=\sum_{r=0}^{m}f(x_r)\varphi_i(x_r)$$

(Nota: l'ipotesi che si fa sulle φ_j in quest'ultima affermazione non è altro che una condizione di ortonormalità rispetto al "prodotto scalare" $(g,h)=\sum_{r=0}^m g(x_r)h(x_r)$).

Esempio (classe di funzioni polinomiali).

Si vuole calcolare la retta p di migliore approssimazione nel senso dei minimi quadrati della tabella $f(x_0)$, $f(x_1)$, …, $f(x_m)$, con x_i reali distinti, usando una base ortonomale per polinomi. Si cercano innanzitutto $\varphi_0(x)=\gamma$ e $\varphi_1(x)=\alpha x+\beta$ per cui

$$\sum_{r=0}^m \varphi_0(x_r)^2=1,\quad \sum_{r=0}^m \varphi_0(x_r)\varphi_1(x_r)=0,\quad \sum_{r=0}^m \varphi_1(x_r)^2=1.$$

Si vede subito che tali identità sono soddisfatte se valvono le seguenti condizioni:

$$\gamma=\frac{1}{\sqrt{m+1}},\quad \alpha\sum_{r=0}^m x_r+\beta(m+1)=0,\quad \alpha^2\sum_{r=0}^m x_r^2+2\alpha\beta\sum_{r=0}^m x_r+\beta^2(m+1)=1.$$

Dalla seconda di queste ultime, segue che β deve avere il valore $\beta=-\alpha(\sum_{r=0}^m x_r)/(m+1)$, che sostituito nella terza produce la seguente equazione in α: $(\sum_{r=0}^m x_r^2-(\sum_{r=0}^m x_r)^2/(m+1))\alpha^2=1$. Dunque,

$$\varphi_0(x)=\frac{1}{\sqrt{m+1}},\quad \varphi_1(x)=\alpha x+\beta,\ \alpha=\frac{1}{\sqrt{\sum_{r=0}^m x_r^2-\frac{1}{m+1}(\sum_{r=0}^m x_r)^2}},\ \beta=-\frac{\alpha}{m+1}\sum_{r=0}^m x_r.$$

Ora che abbiamo le funzioni φ_0 e φ_1, ricordando che $a_0=\sum_{r=0}^m f(x_r)\varphi_0(x_r)$ e $a_1=\sum_{r=0}^m f(x_r)\varphi_1(x_r)$, si può scrivere immediatamente la retta p:

$$p(x)=a_0\varphi_0(x)+a_1\varphi_1(x),\quad a_0=\frac{1}{\sqrt{m+1}}\sum_{r=0}^m f(x_r),\quad a_1=\alpha\sum_{r=0}^m f(x_r)x_r+\beta\sum_{r=0}^m f(x_r).$$

Usiamo il risultato ottenuto per trovare la retta di migliore approssimazione nel senso dei minimi quadrati della tabella $(0,-1)$, $(1/3,0)$, $(2/3,0)$, $(1,2)$. In questo caso si ha

$$m=3,\quad \sum_{r=0}^3 x_r=2,\quad \sum_{r=0}^3 x_r^2=\frac{14}{9},\quad \alpha=\frac{1}{\sqrt{14/9-\frac{1}{4}(2)^2}}=\frac{3}{\sqrt5},\quad \beta=-\frac{\alpha}{4}2=-\frac{3}{2\sqrt5},$$

quindi i primi due polinomi ortonormali rispetto al prodotto scalare $(g,h)=\sum_{r=0}^3 g(r/3)h(r/3)$, sono

$$\varphi_0(x)=\frac12,\quad \varphi_1(x)=\frac{3}{\sqrt5}(x-\frac12).$$

Dunque, la retta di migliore approssimazione nel senso dei minimi quadrati della tabella $(r/3,f(r/3))$, $r=0,1,2,3$, è

$$a_0\frac12+a_1\frac{3}{\sqrt5}(x-\frac12),\quad a_0=\frac12\sum_{r=0}^3 f(r/3),\quad a_1=\frac{3}{\sqrt5}\sum_{r=0}^3 f(r/3)r/3-\frac{3}{2\sqrt5}\sum_{r=0}^3 f(r/3).$$

Nel nostro caso

$$\sum_{r=0}^{3} f(r/3) = 1, \quad \sum_{r=0}^{3} f(r/3)r/3 = 2$$

dunque $a_0 = \frac{1}{2}$, $a_1 = \frac{3}{\sqrt{5}}2 - \frac{3}{2\sqrt{5}} = \frac{9}{2\sqrt{5}}$, e la retta è

$$\frac{1}{2}\frac{1}{2} + \frac{9}{2\sqrt{5}}\frac{3}{\sqrt{5}}(x - \frac{1}{2}) = -\frac{11}{10} + \frac{27}{10}x.$$

Si disegna la retta ottenuta e si osserva che se i dati $-1, 0, 0, 2$ della tabella fossero imprecisi, questa retta molto probabilmente sarebbe una approssimazione della funzione f (di cui la tabella è un campionamento) migliore del polinomio interpolante la tabella.

Esercizio. Calcolare la parabola p di migliore approssimazione nel senso dei minimi quadrati della generica tabella $f(x_0)$, $f(x_1)$, ..., $f(x_m)$, con x_i reali distinti.

Esempio (classe di funzioni trigonometriche).

Esempio (classe di funzioni razionali (che non sono lineari!)).

Formula per l'errore nell'interpolazione (supponendo la funzione f che si sta interpolando sufficientemente regolare)

Siano $x_i \in \mathbb{R}$, $i = 0, 1, \ldots, n$, numeri reali distinti e $y_i \in \mathbb{R}$, $i = 0, 1, \ldots, n$, numeri reali qualsiasi. Supponiamo che gli y_i siano i valori assunti in x_i da una funzione f sufficientemente regolare in un intervallo aperto contenente gli x_i. Sia p il polinomio di grado al più n tale che $p(x_i) = y_i = f(x_i)$, $\forall\, i$. Siano $R(t) = f(t) - p(t)$, $\omega(t) = \prod_{i=0}^{n}(t - x_i)$. Scelto x nell'intervallo di cui sopra e non coincidente con nessuno dei nodi, poniamo $F(t) = \omega(t)R(x) - \omega(x)R(t)$.

Sia $\mathcal{I}$ il minimo intervallo reale aperto la cui chiusura contiene x e tutti i punti x_i, cioè $\mathcal{I} = (a, b)$, $a = \min\{x_i, x\}$, $b = \max\{x_i, x\}$. Si osserva che F si annulla negli $n+2$ punti distinti x_i e x. Dunque F' si annulla in almeno $n+1$ punti distinti di $\mathcal{I}$. F'' si annulla in almeno n punti distinti di $\mathcal{I}$. F''' si annulla in almeno $n-1$ punti distinti di $\mathcal{I}$, ..., $F^{(n+1)}$ si annulla in almeno un punto di $\mathcal{I}$. Quindi esiste $\xi \in \mathcal{I}$ tale che $0 = F^{(n+1)}(\xi) = (n+1)!R(x) - \omega(x)f^{(n+1)}(\xi)$, ovvero tale che

$$R(x) = f(x) - p(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\omega(x).$$

Si osserva che l'espressione trovata per $R(x)$ vale anche se x è uno dei nodi.

V LEZIONE

Sia $f : \mathbb{R} \to \mathbb{R}$ una funzione sufficientemente regolare in un aperto contenente l'intervallo $[a, b]$. Siano x_i, $i = 0, 1, \ldots, n$, numeri reali distinti e $p(x)$ il polinomio interpolante la tabella $(x_i, f(x_i))$, $i = 0, 1, \ldots, n$. Sappiamo che esiste ξ nel minimo intervallo contenente gli x_i e $x \in \mathbb{R}$ per cui

$$f(x) - p(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\prod_{j=0}^{n}(x - x_j).$$

Ci si chiede: come vanno scelti gli x_i affinché nell'intervallo $[a,b]$ il polinomio $p(x)$ approssimi $f(x)$ il meglio possibile? Osservando che

$$\forall\, x \in [a,b], \quad |f(x)-p(x)| \le \frac{1}{(n+1)!} \max_{t\in[a,b]} |f^{(n+1)}(t)| \max_{x\in[a,b]} |\prod_{j=0}^{n}(x-x_j)|,$$

viene naturale scegliere, se possibile, gli x_i in modo che sia minima la quantità $\max_{x\in[a,b]} |\prod_{j=0}^{n}(x-x_j)|$. Quindi, dobbiamo risolvere il seguente problema:

Problema: Calcolare il polinomio $\omega_{ott}(x)$ monico di grado $n+1$ tale che

$$\max_{x\in[a,b]} |\omega_{ott}(x)| = \min_{\omega} \max_{x\in[a,b]} |\omega_{ott}(x)|$$

dove il minimo è effettuato su tutti i polinomi monici ω di grado $n+1$.

Tale problema è già stato risolto nel caso $[a,b]=[-1,1]$. Dobbiamo dunque semplicemente "trasformare" quel risultato da $[-1,1]$ ad $[a,b]$. Sia $\omega(x)$ un generico polinomio monico in x di grado $n+1$. Cominciamo con l'osservare che

$$\omega(x) = \omega(\frac{a+b}{2} + t\frac{b-a}{2}) = \tilde{\omega}(t) = \Big(\frac{b-a}{2}\Big)^{n+1} q(t)$$

ove q è un generico polinomio monico in t di grado $n+1$. Dunque

$$\max_{x\in[a,b]} |\omega(x)| = \Big(\frac{b-a}{2}\Big)^{n+1} \max_{t\in[-1,1]} |q(t)|,$$

$$\begin{aligned}
\min_{\omega} \max_{x\in[a,b]} |\omega(x)| &= \Big(\frac{b-a}{2}\Big)^{n+1} \min_{q} \max_{t\in[-1,1]} |q(t)| \\
&= \Big(\frac{b-a}{2}\Big)^{n+1} \max_{t\in[-1,1]} |\frac{1}{2^n} T_{n+1}(t)| \\
&= \Big(\frac{b-a}{2}\Big)^{n+1} \frac{1}{2^n} \\
&= \Big(\frac{b-a}{2}\Big)^{n+1} \frac{1}{2^n} \max_{x\in[a,b]} |T_{n+1}\Big(\frac{x-(a+b)/2}{(b-a)/2}\Big)|.
\end{aligned}$$

Riassumendo, il polinomio ω_{ott} che risolve il Problema di cui sopra è

$$\omega_{ott}(x) = \Big(\frac{b-a}{2}\Big)^{n+1} \frac{1}{2^n} T_{n+1}\Big(\frac{x-(a+b)/2}{(b-a)/2}\Big),$$

e, di conseguenza, gli x_i per cui il polinomio $p(x)$ approssima $f(x)$ il meglio possibile in $[a,b]$ (ovvero, per cui il fattore $\prod_{j=0}^{n}(x-x_j)$ nella formula del resto ha modulo minimo per $x\in[a,b]$), sono i nodi di Chebycev:

$$x_j = \frac{a+b}{2} + \frac{b-a}{2}\cos\frac{(2j+1)\pi}{2(n+1)}, \quad j=0,1,\dots,n.$$

Osserviamo che tali nodi sono tutti all'interno dell'intervallo (a,b) (come era prevedibile) e, all'aumentare di n, tendono ad accumularsi più agli estremi che al centro di tale intervallo.

I nodi x_i dell'interpolazione di una f potrebbero essere scelti secondo altri criteri, non meno significativi. Uno di questi richiede che gli x_i siano scelti in modo che eventuali piccoli errori sulla tabella non implichino grosse variazioni sul polinomio interpolatore. Più precisamente, ricordando l'espressione di Lagrange di $p(x)$,

$$p(x) = \sum_{i=0}^{n} f(x_i) l_i(x), \quad l_i(x) = \prod_{j=0,\, j\neq i}^{n} \frac{(x - x_j)}{(x_i - x_j)},$$

notiamo che la sostituzione dei dati $f(x_i)$ con dati perturbati $f^*(x_i)$, produrrebbe un polinomio perturbato $p^*(x)$ tale che

$$|p(x) - p^*(x)| \leq \sum_{i=0}^{n} |f(x_i) - f^*(x_i)||l_i(x)| \leq \varepsilon \sum_{i=0}^{n} |l_i(x)|$$

(si suppone di conoscere una limitazione superiore per le possibili perturbazioni). A questo punto diventa interessante studiare la funzione fattore di ε nella maggiorazione di cui sopra. Se non vi fosse il modulo essa sarebbe identicamente uguale a 1 (perché?). Ma il modulo c'è.

Esercizio. Mostrare che $\sum_{i=0}^{n} l_i(x) = 1$ e $\sum_{i=0}^{n} |l_i(x)| \geq 1$.

In letteratura, fissato un intervallo $[a,b]$ si studia il seguente numero

$$l_n = \max_{x\in[a,b]} \sum_{i=0}^{n} |l_i(x)|,$$

(costante di Lebesgue) e la possibilità di minimizzarlo al variare degli $x_j = x_j^{(n)}$ in $[a,b]$ e al variare di n. Ad esempio si mostra che se i nodi sono scelti equidistanti in $[a,b]$, allora $l_n \geq c2^n/(n\sqrt{n})$. Inoltre, si mostra che comunque si scelgano i nodi, l_n è sempre maggiore o uguale di $cn \log n$. Infine si prova che se gli $x_j^{(n)}$ sono i nodi di Chebycev allora l_n è abbastanza piccolo, ma non ottimale. Presso l'Università di Potenza si sono studiate successioni di nodi ottimali.

Esercizio. Siano $f \in C^0([a,b])$ e $x_i \in [a,b]$, $i = 0, 1, \ldots, n$, distinti. Sia p il polinomio interpolante f negli x_i. Mostrare che per ogni $x \in [a,b]$,

$$|f(x) - p(x)| \leq (\min_q \max_{x\in[a,b]} |f(x) - q(x)|)(1 + \sum_{i=0}^{n} |l_i(x)|)$$

dove il minimo è effettuato su tutti i polinomi q di grado al più n.

Risoluzione. $|f(x) - p(x)| \leq |f(x) - q(x)| + |q(x) - p(x)| = |f(x) - q(x)| + |\sum_{i=0}^{n}(q(x_i) - f(x_i))l_i(x)| \leq |f(x) - q(x)| + \sum_{i=0}^{n} |q(x_i) - f(x_i)|\,|l_i(x)|$. Completare la dimostrazione. $\square$

Esercizio. Calcolare la parabola che meglio approssima nel senso dei minimi quadrati $|x|$ in $[-1,1]$ e $\cos x$ in $[0,\pi/3]$, usando rispettivamente i punti $-1,-1/2,0,1/2,1$ e $0,\pi/6,\pi/4,\pi/3$. Calcolare i polinomi che interpolano le funzioni $|x|$ e $\cos x$ nei punti indicati. Confrontare i grafici delle funzioni ottenute.

Esercizio. Trovare una espressione esplicita per il polinomio di migliore approssimazione minmax della funzione $f(x)=1/(1+x)$ nell'intervallo $[0,1]$.

Teoria essenziale sui polinomi ortogonali

I polinomi ortogonali, studiati in questa sezione, permettono di introdurre formule di quadratura ottimali per l'approssimazione di integrali (vedi più avanti).

Due funzioni f e g sono ortogonali tra loro rispetto al peso $\omega(x)>0$ e all'intervallo $(a,b)\subset\mathbb{R}$, se $\int_a^b \omega(x)f(x)g(x)\,dx=0$. Si vuole costruire un polinomio $p_j(x)$ monico di grado j ortogonale ad ogni polinomio di grado inferiore. Supponendo di avere già costruito i polinomi $p_j(x)$, $j=0,1,\ldots,n$, con tali proprietà, si vuole costruire $p_{n+1}(x)$. Quindi cerchiamo p_{n+1} monico di grado $n+1$ tale che

$$\int_a^b \omega(x)p_{n+1}(x)q_n(x)\,dx=0$$

per ogni polinomio q_n di grado minore o uguale a n. Sia $q_n(x)=a_0x^n+a_1x^{n-1}+\ldots$. Tenuto conto che $p_n(x)$ è monico, si ha che $q_n(x)-a_0p_n(x)$ è uguale a un polinomio $q_{n-1}(x)$ di grado minore o uguale a $n-1$. Dunque $q_n(x)=a_0p_n(x)+q_{n-1}(x)$. Analogamente, q_{n-1} si può rappresentare come segue: $q_{n-1}(x)=a_0'p_{n-1}(x)+q_{n-2}(x)$. Ne segue che il generico polinomio q_n può essere scritto in termini di p_n e p_{n-1} attraverso una formula del tipo:

$$q_n(x)=a_0p_n(x)+a_0'p_{n-1}(x)+q_{n-2}(x).$$

Allora, la condizione di ortogonalità caratterizzante p_{n+1} diventa: soddisfare l'identità

$$\int_a^b \omega(x)p_{n+1}(x)(a_0p_n(x)+a_0'p_{n-1}(x)+q_{n-2}(x))\,dx=0$$

per ogni $a_0,a_0'\in\mathbb{R}$ e per ogni q_{n-2} polinomio di grado al più $n-2$. Ma tale richiesta è equivalente alle tre seguenti equazioni:

$$\int_a^b \omega(x)p_{n+1}(x)p_n(x)\,dx=0,\quad \int_a^b \omega(x)p_{n+1}(x)p_{n-1}(x)\,dx=0,\quad \int_a^b \omega(x)p_{n+1}(x)q_{n-2}(x)\,dx=0,\ \forall\, q_{n-2}.$$

Cerchiamo come soluzione delle precedenti tre equazioni un polinomio $\tilde{p}_{n+1}$ della forma $\tilde{p}_{n+1}(x)=(x+b_n)p_n(x)+c_np_{n-1}(x)$. Un tale $\tilde{p}_{n+1}$ soddisfa la terza equazione per ogni scelta dei parametri b_n e c_n. Dunque tali parametri possono essere scelti in modo che $\tilde{p}_{n+1}$ soddisfi anche le altre due equazioni:

$$b_n=-\frac{\int_a^b \omega(x)xp_n(x)^2\,dx}{\int_a^b \omega(x)p_n(x)^2\,dx},\quad c_n=-\frac{\int_a^b \omega(x)xp_n(x)p_{n-1}(x)\,dx}{\int_a^b \omega(x)p_{n-1}(x)^2\,dx}.$$

Esercizio. Mostrare la seguente altra espressione per c_n: $c_n = -\frac{\int_a^b \omega(x)p_n(x)^2\,dx}{\int_a^b \omega(x)p_{n-1}(x)^2\,dx}$.
Risoluzione. $p_n = (x+b_{n-1})p_{n-1} + c_{n-1}p_{n-2} \Rightarrow p_n^2 = (x+b_{n-1})p_{n-1}p_n + c_{n-1}p_{n-2}p_n$.

Una volta noti $p_0(x)$ ($p_0(x) = 1$) e $p_1(x)$ (trovare p_1!), possiamo dunque calcolare tutti i successivi polinomi ortogonali. Notiamo che essi sono univocamente determinati. Se infatti fosse vero che

$$\int_a^b \omega(x)p_{n+1}(x)q_n(x)\,dx = 0, \quad \int_a^b \omega(x)\tilde{p}_{n+1}(x)q_n(x)\,dx = 0, \quad \forall\, q_n\ \partial q_n \le n,$$

per due diversi polinomi monici di grado $n+1$, allora si avrebbe

$$\int_a^b \omega(x)(p_{n+1}(x) - \tilde{p}_{n+1}(x))q_n(x)\,dx = 0$$

anche per $q_n = p_{n+1} - \tilde{p}_{n+1}$, il che è assurdo a meno che $p_{n+1} - \tilde{p}_{n+1}$ è identicamente zero.

Concludiamo con una importante osservazione. *Il polinomio ortogonale p_n ha le sue n radici tutte reali e contenute nell'intervallo aperto* (a,b). La dimostrazione di questa osservazione si suddivide in quattro passi. Il risultato del primo passo implicherà i risultati degli altri tre.

a) Si dimostra che ogni fattore di p_n deve cambiare segno in (a,b). Sia infatti $p_n(x) = q_r(x)\tilde{q}_{n-r}(x)$, con $1 \le r \le n-1$. Allora, per la proprietà che caratterizza p_n, si deve avere $\int_a^b \omega(x)q_r(x)\tilde{q}_{n-r}(x)^2\,dx = 0$. Ma $\omega(x)\tilde{q}_{n-r}(x)^2 \ge 0$ in (a,b), dunque esiste $\xi \in (a,b)$ tale che $q_r(\xi) = 0$ e q_r cambia segno in ξ. In particolare, ogni fattore di p_n possiede almeno uno zero reale in (a,b).

Si dimostra che p_n non può avere radici complesse non reali. Sia infatti $\alpha + \mathbf{i}\beta$ una tale radice. Essendo p_n a coefficienti reali, anche $\alpha - \mathbf{i}\beta$ deve essere radice di p_n. Dunque il polinomio non negativo $q_2(x) = (x-(\alpha-\mathbf{i}\beta))(x-(\alpha+\mathbf{i}\beta)) = (x-\alpha)^2 + \beta^2$ deve dividere p_n, il che contraddice a).

Si dimostra che p_n non può avere radici reali non semplici. Sia infatti α una tale radice. Allora il polinomio non negativo $q_2(x) = (x-\alpha)^2$ deve dividere p_n, il che non è possibile per a).

Si dimostra che p_n non può avere radici reali fuori dell'intervallo (a,b). Sia infatti α una tale radice. Supponiamo $\alpha \le a$. Allora il polinomio $q_1(x) = x - \alpha$, positivo in (a,b), deve dividere p_n, il che di nuovo per a) non è possibile. Stessa conclusione se $\alpha \ge a$.

Esercizio. Dimostrare che se la funzione peso ω è simmetrica rispetto al centro dell'intervallo (a,b), allora anche gli zeri di p_n sono simmetrici rispetto al centro dell'intervallo (a,b).

VI LEZIONE

Risultati sulla "convergenza" dei polinomi interpolatori

Sia f ben definita in $[a,b] \subset \mathbb{R}$. Per ogni $n \in \mathbb{N}$, siano $x_j^{(n)}$, $j = 0,1,\dots,n$, numeri reali distinti in $[a,b]$, e $p_n(x)$ il polinomio interpolante f negli $x_j^{(n)}$. Ci chiediamo se e quando è vero che, per $n \to +\infty$,

$$|f(x) - p_n(x)| \to 0, \quad \forall\, x \in [a,b]. \qquad (*)$$

Elenchiamo dei risultati noti a riguardo.

1 Se gli $x_j^{(n)}$ sono scelti equidistanti in $[a,b]$ ($x_j^{(n)} = a + j(b-a)/n$), allora esiste $f \in C^\infty([a,b])$ per cui (*) non è vero. Esempio: $f(x) = 1/(1+x^2)$, $[a,b] = [-5,5]$.

2 Se $f \in C^2([a,b])$, scegliendo opportunamente la successione degli $x_j^{(n)}$, (*) è vera.

3 Comunque scelta la successione degli $x_j^{(n)}$, esiste $f \in C^0([a,b])$ per cui (*) non è vera.

4 Se $f \in C^\infty([a,b])$ con derivate tutte limitate dalla stessa costante M, allora (*) è vera.

Dimostrazione. Si prova solo il risultato 4). È noto che per ogni n e per ogni $x \in [a,b]$, esistono $\xi_n \in (a,b)$ per cui

$$|f(x) - p_n(x)| = \frac{1}{(n+1)!}|f^{(n+1)}(\xi_n)||\prod_{j=0}^{n}(x - x_j^{(n)})| \le \frac{M}{(n+1)!}(b-a)^{n+1}$$

e la quantità a destra, per $n \to +\infty$, tende a zero. $\square$

Esercizio. Il risultato (4) non sembra poter essere vero. Ad esempio, infittendo i punti solo in $[a,(a+b)/2)$ come si può ottenere la convergenza a zero di $f(x) - p_n(x)$ quando x è in $((a+b)/2, b]$?

Per maggiori particolari, vedere il libro di Davis sull'approssimazione funzionale.

Formule di quadratura per l'approssimazione di integrali

Siano $x_0 < x_1 < \ldots < x_n$, $l_i(x)$ i corrispondenti polinomi di Lagrange, e f una funzione ben definita in un aperto contenente $[a,b]$ e gli x_i. Allora, per $x \in [a,b]$, la funzione $f(x)$ può essere rappresentata tramite la seguente formula: $f(x) = \sum_{i=0}^n f(x_i)l_i(x) + R(x)$, dove $R(x)$ è ovviamente sempre ben definito (se f è regolare allora $R(x) = f^{(n+1)}(\xi)\prod_{j=0}^n (x-x_j)/(n+1)!$ con ξ nel minimo intervallo contenente x e gli x_i; si noti però che è sempre valida una formula per $R(x)$ del tipo $R(x) = g(x, x_0, x_1, \ldots, x_n)\prod_{j=0}^n (x - x_j)$, da cui si deduce che $R(x)$ può essere grande se x è lontano dagli x_j, per tale formula vedi l'appendice).

Moltiplicando questa rappresentazione di $f(x)$ per una data funzione peso $\omega(x)$ positiva in (a,b), e integrando da a a b, si ottiene la seguente identità:

$$\int_a^b \omega(x)f(x)\,dx = \sum_{i=0}^n f(x_i)A_i + E, \quad A_i = \int_a^b \omega(x)l_i(x)\,dx, \quad E = \int_a^b \omega(x)R(x)\,dx. \qquad (*)$$

La formula $\sum_{i=0}^n f(x_i)A_i$, $A_i = \int_a^b \omega(x)l_i(x)\,dx$, è detta formula di quadratura "interpolatoria" e si può usare per approssimare l'integrale $\int_a^b \omega(x)f(x)\,dx$.

Elenchiamo fatti importanti sulle formule di quadratura interpolatorie.

- Si ha $E = 0$ se f è un polinomio di grado minore o uguale di n. Quindi, una volta decisi i nodi x_i della formula di quadratura, i suoi coefficienti A_i possono essere calcolati, anziché usando la formula in (*), imponendo l'identità $E = 0$ per $f(x) = 1, x, x^2, \ldots, x^n$ (o per $f(x) = \varphi_0(x), \varphi_1(x), \ldots, \varphi_n(x)$ dove le φ_j formano una base qualunque per l'insieme dei polinomi di grado al più n). Ad esempio, imporre il minimo, cioè $E = 0$ per $f(x) = 1$, produce la seguente condizione sugli A_i:

$$\int_a^b \omega(x)\,dx = \sum_{i=0}^{n} A_i.$$

- Se i nodi x_i sono simmetrici rispetto al centro dell'intervallo (a, b) e la funzione peso ω pure, allora $A_i = A_{n-i}$.

Dimostrazione (nell'ipotesi $(a + b)/2 = 0$). Si dimostra che $A_0 = A_n$, lasciando al lettore la dimostrazione del caso generale.

$$l_0(x) = \prod_{j=1}^{n} \frac{x - x_j}{x_0 - x_j}, \quad l_0(-x) = \prod_{j=1}^{n} \frac{-x - x_j}{x_0 - x_j} = \prod_{j=1}^{n} \frac{x + x_j}{-x_0 + x_j} = \prod_{j=1}^{n} \frac{x - x_{n-j}}{x_n - x_{n-j}} = l_n(x), \quad \Rightarrow$$

$$A_n = \int_a^b \omega(x) l_n(x)\,dx = \int_a^b \omega(x) l_0(-x)\,dx = -\int_b^a \omega(-y) l_0(y)\,dy = \int_a^b \omega(y) l_0(y)\,dy = A_0.$$

- Se $\omega(x) = 1$ e $x_i = a + i(b - a)/n$, $i = 0, 1, \ldots, n$ (nodi equidistanti), allora le corrispondenti formule di quadratura sono note come formule di Newton-Cotes (chiuse). Esse sono esatte per polinomi di grado al più n se n è dispari (come sapevamo), e sono esatte per polinomi di grado al più $n + 1$ se n è pari. Infatti, per $f(x) = (x - (a + b)/2)^{n+1}$ si ha gratis l'identità $E = 0$, essendo ovviamente $\int_a^b f(x)\,dx = 0$ e $\sum_{i=0}^{n} A_i f(x_i) = 0$ per la simmetria dei nodi x_i rispetto ad $(a + b)/2$ e per l'identità $A_i = A_{n-i}$.

- Se si considerano anche i nodi incogniti, allora esiste una loro scelta per cui le formule di quadratura sono esatte ($E = 0$) per ogni polinomio di grado al più $2n + 1$. Tali formule sono dette Gaussiane.

- Le formule di quadratura non potranno mai essere esatte per polinomi di grado maggiore di $2n + 1$ perché ad esempio per $f(x) = \prod_{j=0}^{n}(x - x_j)^2$, l'integrale $\int_a^b \omega(x) f(x)\,dx$ è positivo, mentre $\sum_{i=0}^{n} f(x_i) A_i = 0$.

Formule di quadratura di Newton-Cotes (N.C.)

Troviamo le formule di Newton-Cotes relative all'intervallo $[-1, 1]$ e ai nodi $-1 + i\frac{2}{n}$, $i = 0, 1, \ldots, n$. Ovvero, troviamo $A_i = A_i^{(n)}$ tali che l'identità

$$\int_{-1}^{1} f(x)\,dx = \sum_{i=0}^{n} A_i^{(n)} f(-1 + i\frac{2}{n}) + E, \;\; A_i = \int_{-1}^{1} l_i(x)\,dx, \;\; E = \int_{-1}^{1} R(x)\,dx,$$

è vera con $E = 0$ per ogni f polinomio di grado al più n. Per farlo è utile ricordare che $A_i = A_{n-i}$.

Per $n = 1$ si ha $\int_{-1}^{1} f(x)\,dx = A_0 f(-1) + A_1 f(1) + E = A_0(f(-1) + f(1)) + E$. Imponendo $E = 0$ per $f(x) = 1$ si ottiene la condizione $A_0 = 1$. Quindi abbiamo la formula di N.C. a due nodi in $[-1, 1]$:

$$\int_{-1}^{1} f(x)\,dx = f(-1) + f(1) + E. \tag{$*$}$$

Da questa segue subito la formula di N.C. a due nodi o del "trapezio" in $[a, b]$:

$$\int_{a}^{b} f(x)\,dx = \frac{b-a}{2}(f(a) + f(b)) + E, \quad E = -\frac{h^3}{12} f^{(2)}(\eta), \ \ h = b - a \tag{$trap$}$$

(porre $x = (a + b)/2 + t(b - a)/2$ e usare (*); $\eta \in (a, b)$; si omette la dimostrazione dell'espressione dell'errore).

Per $n = 2$ si ha $\int_{-1}^{1} f(x)\,dx = A_0 f(-1) + A_1 f(0) + A_2 f(1) + E = A_0(f(-1) + f(1)) + A_1 f(0) + E$. Imponendo $E = 0$ per $f(x) = 1$ e $f(x) = x^2$ (per $f(x) = x$ è superfluo perché E è già zero) si ottengono le condizioni $2 = 2A_0 + A_1$, $\frac{2}{3} = 2A_0$, cioè $A_0 = \frac{1}{3}$, $A_1 = \frac{4}{3}$. Quindi abbiamo la formula di N.C. a tre nodi in $[-1, 1]$:

$$\int_{-1}^{1} f(x)\,dx = \frac{1}{3}(f(-1) + f(1)) + \frac{4}{3} f(0) + E.$$

Da questa segue la formula di N.C. a tre nodi o di "Simpson" in $[a, b]$:

$$\int_{a}^{b} f(x)\,dx = \frac{b-a}{2}\Big[\frac{1}{3}(f(a) + f(b)) + \frac{4}{3} f\Big(\frac{a+b}{2}\Big)\Big] + E, \quad E = -\frac{h^5}{90} f^{(4)}(\eta), \ \ h = \frac{b-a}{2} \tag{$simps$}$$

($\eta \in (a, b)$; si omette la dimostrazione dell'espressione dell'errore).

Esercizio. Calcolare con le formule di quadratura del trapezio e di Simpson (o Newton-Cotes rispettivamente a due e a tre nodi) due approssimazioni dell'integrale $\int_1^2 \frac{1}{x}\,dx = \log_e 2$. Risoluzione: Con la formula del trapezio $h\frac{1}{2}(f(a) + f(b))$, $h = b - a$, si approssima l'integrale $\int_a^b f(x)\,dx$. Quindi, nel nostro caso $\frac{1}{2}(1 + \frac{1}{2}) = \frac{3}{4}$ approssima $\int_1^2 \frac{1}{x}\,dx = \log_e 2$. Con la formula di Simpson $h[\frac{1}{3}(f(a) + f(b)) + \frac{4}{3} f(\frac{a+b}{2})]$, $h = \frac{b-a}{2}$, si approssima l'integrale $\int_a^b f(x)\,dx$. Quindi, nel nostro caso $\frac{1}{2}(\frac{1}{2} + \frac{8}{9}) = \frac{25}{36}$ approssima $\int_1^2 \frac{1}{x}\,dx = \log_e 2$.

Esercizio. Abbiamo trovato le formule di Newton-Cotes per $n = 1, 2$. Trovarle per $n = 3, 4$.

Formule di quadratura composte

Esempi di formule di quadratura non interpolatorie si ottengono "componendo" quelle interpolatorie. Si introduce ora la formula composta dei trapezi. Sia f definita in $[a, b]$ e si vuole ottenere una approssimazione dell'integrale $\int_a^b f(x)\,dx$. Poniamo $x_i = a + i(b-a)/n$, $f_i = f(x_i)$, $i = 0, 1, \ldots, n$, e $h = (b-a)/n$. Allora esistono $\eta_i \in (x_{i-1}, x_i)$ ed $\eta \in (a,b)$ per cui

$$\begin{aligned}
\int_a^b f(x)\,dx &= \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} f(x)\,dx \\
&= \sum_{i=1}^{n} \Big[\frac{h}{2}(f_{i-1}+f_i) - \frac{h^3}{12} f''(\eta_i)\Big] \\
&= h\Big(\frac{f_0+f_n}{2} + \sum_{i=1}^{n-1} f_i\Big) - \frac{h^3}{12} n f''(\eta) \\
&= h\Big(\frac{f_0+f_n}{2} + \sum_{i=1}^{n-1} f_i\Big) - \frac{h^2}{12}(b-a) f''(\eta).
\end{aligned}$$

La formula di quadratura cercata è $I_h := h(\frac{f_0+f_n}{2} + \sum_{i=1}^{n-1} f_i)$. Osserviamo che, se $f \in C^2([a,b])$, allora $I_h \to \int_a^b f(x)\,dx$ se $n \to +\infty$ e l'ordine di convergenza è $O(h^2) = O(1/n^2)$. Tuttavia, dalla teoria dell'integrazione di Riemann sappiamo che basta la continuità di f in $[a,b]$ per avere che $I_h \to \int_a^b f(x)\,dx$.

Il resto $-\frac{h^2}{12}(b-a)f''(\eta)$, se f è sufficientemente regolare, ammette l'ulteriore rappresentazione

$$\sum_{j=1}^{k} h^{2j} \frac{B_{2j}(0)}{(2j)!}[f^{(2j-1)}(b) - f^{(2j-1)}(a)] + u_{k+1}$$

($u_{k+1} = O(h^{2k+2})$), ove i $B_{2j}(0)$ sono i numeri di Bernoulli: $B_{2j}(0) = (-1)^{j+1}(2(2j)!/((2\pi)^{2j})) \sum_{r=1}^{+\infty} 1/r^{2j}$. Tale formula giustifica l'efficienza del procedimento di estrapolazione di Romberg con la formula dei trapezi per la approssimazione di integrali (argomento più avanzato).

Si può comporre allo stesso modo la formula di Simpson ed ottenere una formula di quadratura non interpolatoria in generale molto più precisa della formula composta dei trapezi. Vedi la prossima lezione.

Spezzare l'intervallo $[a,b]$ in tanti sottointervalli, approssimare f in tali sottointervalli con polinomi di grado basso, unire i risultati e far crescere il numero di sottointervalli è un modo per ottenere approssimazioni sempre migliori di una funzione f in $[a,b]$ alternativo a quello di aumentare il grado del polinomio approssimante f in $[a,b]$ (definito via interpolazione, minmax o minimi quadrati).

Ad esempio le formule composte dei trapezi e di Simpson si ottengono in pratica calcolando le aree sotto approssimazioni regolari a tratti (rette a tratti o parabole a tratti) della funzione f che si vuole integrare. Tali approssimazioni sono, in $[a,b]$, solo continue. Accenniamo al fatto che si possono ottenere approssimazioni di f polinomiali a tratti in $[a,b]$, più regolari in $[a,b]$. Consideriamo l'esempio delle spline cubiche. Dati $x_0 = a < x_1 < \ldots < x_{n-1} < x_n = b$ e f definita in $[a,b]$, è possibile imporre le seguenti condizioni:

1) $g(x_i) = f(x_i)$, $i = 0, 1, \dots, n$,
2) $g|_{[x_i, x_{i+1}]}$ è un polinomio di grado minore o uguale a tre ($\forall\, i$),
3) $g \in C^2([a, b])$.

Infatti, abbiamo a disposizione $4n$ parametri (ovvero, i parametri che definiscono gli n polinomi di grado al più 3 distribuiti negli n intervalli $[x_i, x_{i+1}]$). La condizione 1) occupa $2n$ di questi parametri (il polinomio dell'intervallo $[x_i, x_{i+1}]$ nell'estremo destro e sinistro di tale intervallo deve assumere il valore della f). Affinché g sia C^1 dobbiamo imporre altre $n-1$ condizioni (il polinomio a destra di x_i e quello a sinistra di x_i devono avere derivate coincidenti in x_i). Idem affinché g sia C^2. Rimangono dunque due parametri liberi che possono essere scelti in vari modi (ad es. tali che $g''(a) = g''(b) = 0$, spline naturali, o tali che $g'(a) = f'(a)$, $g'(b) = f'(b)$).

APPENDICE alla VI lezione

Interpolazione e differenze divise

Dati x_i, $i = 0, 1, \dots, n$, numeri reali distinti, $x \neq x_i$, ed f definita su un aperto contenente x e gli x_i, diamo le seguenti definizioni:

$$f[x, x_0] = \frac{f(x_0) - f(x)}{x_0 - x},$$

da cui $f(x) = f(x_0) + (x - x_0) f[x, x_0]$,

$$f[x, x_0, x_1] = \frac{f[x_0, x_1] - f[x, x_0]}{x_1 - x},$$

da cui $f(x) = f(x_0) + (x - x_0) f[x_0, x_1] + (x - x_0)(x - x_1) f[x, x_0, x_1]$,

$$f[x, x_0, x_1, x_2] = \frac{f[x_0, x_1, x_2] - f[x, x_0, x_1]}{x_2 - x},$$

da cui $f(x) = f(x_0) + (x - x_0) f[x_0, x_1] + (x - x_0)(x - x_1) f[x_0, x_1, x_2] + (x - x_0)(x - x_1)(x - x_2) f[x, x_0, x_1, x_2]$,

$\dots$, e, infine,

$$f[x, x_0, x_1, x_2, \dots, x_n] = \frac{f[x_0, x_1, x_2, \dots, x_n] - f[x, x_0, x_1, \dots, x_{n-1}]}{x_n - x},$$

da cui la seguente rappresentazione per $f(x)$:

$$f(x) = p(x) + R(x),$$

dove

$$\begin{aligned} p(x) &= f(x_0) + (x - x_0) f[x_0, x_1] + (x - x_0)(x - x_1) f[x_0, x_1, x_2] \\ &\quad + (x - x_0)(x - x_1)(x - x_2) f[x_0, x_1, x_2, x_3] + \dots \\ &\quad + (x - x_0)(x - x_1)(x - x_2) \cdots (x - x_{n-1}) f[x_0, x_1, x_2, x_3, \dots, x_n] \end{aligned}$$

$$\begin{aligned} &= \sum_{i=0}^{n} f[x_0,\ldots,x_i]\prod_{j=0}^{i-1}(x-x_j),\\ R(x) &= f[x,x_0,x_1,x_2,\ldots,x_n]\prod_{j=0}^{n}(x-x_j). \end{aligned}$$

Si osserva che $p(x_0) = f(x_0)$. Se ripetessimo le definizioni di cui sopra, anziché usando l'ordine $x_0, x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_n$, usando l'ordine $x_j, x_1, \ldots, x_{j-1}, x_0, x_{j+1}, \ldots, x_n$, si otterrebbe la rappresentazione

$$f(x) = \tilde{p}(x) + \tilde{R}(x),$$

dove

$$\begin{aligned} \tilde{p}(x) &= f(x_j) + (x-x_j)f[x_j,x_1] + (x-x_j)(x-x_1)f[x_j,x_1,x_2]\\ &\quad +(x-x_j)(x-x_1)(x-x_2)f[x_j,x_1,x_2,x_3] + \ldots\\ &\quad +(x-x_j)(x-x_1)(x-x_2)\cdots(x-x_0)\cdots(x-x_{n-1})f[x_j,x_1,x_2,\ldots,x_0,\ldots,x_n],\\ \tilde{R}(x) &= f[x,x_j,x_1,x_2,\ldots,x_0,\ldots,x_n]\prod_{j=0}^{n}(x-x_j). \end{aligned}$$

Si osserva che $\tilde{p}(x_j) = f(x_j)$. Ma $f[x, x_j, x_1, x_2, \ldots, x_0, \ldots, x_n] = f[x, x_0, x_1, \ldots, x_j, \ldots, x_n]$ (vedi più avanti), dunque deve essere $\tilde{p}(x) = p(x)$. Ne segue che $p(x)$ è il polinomio interpolante f nei punti x_i.

Esercizio. I polinomi 1, $x-x_0$, $(x-x_0)(x-x_1)$, ..., $(x-x_0)(x-x_1)\cdots(x-x_{n-1})$ costituiscono una base per l'insieme dei polinomi di grado al più n. Quanto visto sopra prova che i coefficienti rispetto a questa base del polinomio interpolante f nei nodi x_i sono le *differenze divise* $f[x_0, \ldots, x_i]$. Mostrare che effettivamente la soluzione del sistema di $n+1$ equazioni lineari $\sum_{i=0}^{n} c_i \prod_{j=0}^{i-1}(x_k - x_j) = f(x_k)$, $k = 0, 1, \ldots, n$, è data da $c_i = f[x_0, \ldots, x_i]$, $i = 0, 1, \ldots, n$.

Si dimostra ora la seguente rappresentazione per la differenza divisa relativa a $k+1$ numeri reali distinti

$$f[x_0,x_1,\ldots,x_k] = \sum_{i=0}^{k}\frac{f(x_i)}{\omega'(x_i)}, \quad \omega(x) = \prod_{j=0}^{k}(x-x_j), \tag{$*$}$$

dalla quale segue che il suo valore non dipende dall'ordine dei numeri, cioè, se $i_0, i_1, \ldots, i_k$ è una permutazione di $0, 1, \ldots, k$, allora

$$f[x_0,x_1,\ldots,x_k] = f[x_{i_0},x_{i_1},\ldots,x_{i_k}].$$

L'identità (*) si dimostra per induzione. Supponiamo (*) vera per $k = 0, 1, \ldots, n-1$, e dimostriamola per $k = n$. Siano $\omega_l(x) = \prod_{j=0}^{n-1}(x - x_j)$ e $\omega_r(x) = \prod_{j=1}^{n}(x - x_j)$. Allora

$$\begin{aligned} f[x_0,x_1,\ldots,x_n] &= \frac{f[x_1,\ldots,x_{n-1},x_n] - f[x_0,x_1,\ldots,x_{n-1}]}{x_n - x_0}\\ &= \frac{1}{x_n-x_0}\Big(\sum_{i=1}^{n}\frac{f(x_i)}{\omega_r'(x_i)} - \sum_{i=0}^{n-1}\frac{f(x_i)}{\omega_l'(x_i)}\Big). \end{aligned}$$

Osserviamo che $\omega(x) = (x - x_0)\omega_r(x) = \omega_l(x)(x - x_n)$, dunque $\omega'(x) = \omega_r(x) + (x - x_0)\omega_r'(x) = \omega_l(x) + \omega_l'(x)(x - x_n)$. Ne segue che $\omega'(x_i) = (x_i - x_0)\omega_r'(x_i)$, $i = 1, \ldots, n$, e $\omega'(x_i) = \omega_l'(x_i)(x_i - x_n)$, $i = 0, \ldots, n-1$. Allora

$$
\begin{aligned}
f[x_0, x_1, \ldots, x_n] &= \frac{1}{x_n - x_0}\Big(\sum_{i=1}^{n} \frac{f(x_i)(x_i - x_0)}{\omega'(x_i)} - \sum_{i=0}^{n-1} \frac{f(x_i)(x_i - x_n)}{\omega'(x_i)}\Big) \\
&= \frac{1}{x_n - x_0}\sum_{i=0}^{n}\Big(\frac{f(x_i)(x_i - x_0)}{\omega'(x_i)} - \frac{f(x_i)(x_i - x_n)}{\omega'(x_i)}\Big) \\
&= \frac{1}{x_n - x_0}\sum_{i=0}^{n}\frac{f(x_i)}{\omega'(x_i)}[(x_i - x_0) - (x_i - x_n)] \\
&= \sum_{i=0}^{n}\frac{f(x_i)}{\omega'(x_i)}.
\end{aligned}
$$

Si lascia al lettore la dimostrazione della base dell'induzione.

VII LEZIONE

Supponiamo di aver trovato $A_i^{(n)}$ tali che l'identità

$$\int_{-1}^{1} f(x)\,dx = \sum_{i=0}^{n} A_i^{(n)} f(-1 + i\frac{2}{n}) + E$$

è vera con $E = 0$ per ogni f polinomio di grado al più n. Allora

$$
\begin{aligned}
\int_a^b f(x)\,dx &= \tfrac{b-a}{2}\int_{-1}^{1} f(\tfrac{a+b}{2} + t\tfrac{b-a}{2})\,dt = \tfrac{b-a}{2}\int_{-1}^{1}\tilde{f}(t)\,dt \\
&= \tfrac{b-a}{2}\big(\textstyle\sum_{i=0}^{n} A_i^{(n)}\tilde{f}(-1 + i\tfrac{2}{n}) + \tilde{E}\big) \\
&= \tfrac{b-a}{2}\textstyle\sum_{i=0}^{n} A_i^{(n)} f(a + i\tfrac{b-a}{n}) + \tfrac{b-a}{2}\tilde{E}
\end{aligned}
$$

Quindi

$$\int_a^b f(x)\,dx = h\frac{n}{2}\sum_{i=0}^{n} A_i^{(n)} f(a + ih) + c_n \times \begin{cases} h^{n+2} f^{(n+1)}(\eta_n) & n \text{ dispari} \\ h^{n+3} f^{(n+2)}(\eta_n) & n \text{ pari} \end{cases}, \quad h = \frac{b-a}{n} \tag{newtoncotes}$$

(si omette la dim dell'espressione dell'errore).

Abbiamo visto i casi particolari $n = 1, 2, 3, 4$. Si può vedere che per n circa 10 e oltre alcuni dei coefficienti $A_i^{(n)}$ vengono negativi. Questo può essere un inconveniente. Infatti, in generale, invece della formula $Q = \sum_i A_i f_i$, per approssimare $\int_a^b \omega(x) f(x)\,dx$, si usa effettivamente $Q^* = \sum_i A_i f_i^*$, dove le effettive f_i^* sono solo approssimazioni delle vere f_i, dunque

$$|Q - Q^*| = \Big|\sum_i A_i f_i - \sum_i A_i f_i^*\Big| \leq \sum_i |A_i||f_i - f_i^*| \leq \varepsilon \sum_i |A_i|.$$

Se gli A_i cambiano segno, la quantità $\sum |A_i|$ può crescere al crescere di n, e, di conseguenza, la formula di quadratura effettiva Q^* potrebbe avere un valore molto diverso da quello della formula di quadratura vera Q, quindi Q^* potrebbe non essere una approssimazione di $\int_a^b \omega(x) f(x)\, dx$ buona quanto Q.

Siccome, in generale, $\sum_i A_i = \int_a^b \omega(x)\, dx$, sarebbe sufficiente che gli $A_i^{(n)}$ siano positivi per ogni n per avere l'errore $|\sum A_i f_i - \sum A_i f_i^*|$ limitato, $\forall\, n$, dalla stessa quantità $\varepsilon \int_a^b \omega(x)\, dx$.

Le formule di Newton-Cotes non hanno gli $A_i^{(n)}$ positivi. Quelle composte del trapezio e di Simpson invece sì. Ricordiamo la prima, già calcolata in precedenza,

$$\int_a^b f(x)\, dx = h[\frac{f(a)}{2} + \sum_{i=1}^{n-1} f(a+ih) + \frac{f(b)}{2}] - \ldots, \ h = \frac{b-a}{n}$$

(Nota: usa $n+1$ nodi (valori di f) ed è esatta solo per polinomi di grado ≤ 1!). Per quanto riguarda la formula di Simpson composta, posto $h = (b-a)/(2n)$, si ha

$$\begin{aligned}
\int_a^b f(x)\, dx &= \sum_{i=1}^n \int_{x_{2i-2}}^{x_{2i}} f(x)\, dx = \sum_{i=1}^n [\tfrac{h}{3}(f_{2i-2} + 4f_{2i-1} + f_{2i}) - \tfrac{h^5}{90} f^{(4)}(\eta_i)] \\
&= \tfrac{h}{3}(f_0 + 4f_1 + 2f_2 + 4f_3 + \ldots + 4f_{2n-1} + f_{2n}) - \tfrac{h^5}{90} n f^{(4)}(\eta) \\
&= \tfrac{h}{3}(f_0 + 4f_1 + 2f_2 + 4f_3 + \ldots + 4f_{2n-1} + f_{2n}) - \tfrac{h^4}{180}(b-a) f^{(4)}(\eta)
\end{aligned}$$

(Nota: usa $2n+1$ nodi (valori di f) ed è esatta solo per polinomi di grado ≤ 2!).

C'è una classe di formule di quadratura, alternative a quelle di Newton-Cotes, i cui coefficienti sono sempre positivi, ovvero per qualsiasi numero di nodi. Sono le formule di quadratura di Gauss, che, usando n nodi, sono esatte per tutti i polinomi di grado $\leq 2n-1$.

Dati $[a,b] \subset \mathbb{R}$ e ω ben definita e positiva in (a,b), tale che $\int_a^b \omega(x)\, dx < +\infty$, si vogliono cercare x_i e A_i tali che

$$\int_a^b \omega(x) P_{2n-1}(x)\, dx = \sum_{i=1}^n A_i P_{2n-1}(x_i), \ \forall\, P_{2n-1} \text{ polinomio}, \ \partial P_{2n-1} \leq 2n-1\,. \tag{condizGauss}$$

Notiamo che $A_i = \int_a^b \omega(x) l_i(x)$ (si scelga $f(x) = l_i(x)$ in (condizGauss)). Notiamo anche che l'uguaglianza non può essere soddisfatta per $f(x) = \Pi_i (x - x_i)^2$, quindi non possono ottenersi formule di quadratura (del tipo $\sum_{i=1}^n A_i f(x_i)$) esatte per polinomi di grado $\leq 2n$.

Sia $p_n(x) = \Pi_j (x - x_j)$. Esistono q_{n-1} ed r_{n-1} polinomi di grado al più $n-1$ tali che $P_{2n-1}(x) = p_n(x) q_{n-1}(x) + r_{n-1}(x)$, quindi la condizione (condizGauss) diventa

$$\int_a^b \omega(x) P_{2n-1}(x)\, dx = \sum_{i=1}^n A_i r_{2n-1}(x_i)$$

ovvero, per ogni r_{n-1} e q_{n-1} di grado al più $n-1$, devono verificarsi le seguenti due uguaglianze:

$$\int_a^b \omega(x) p_n(x) q_{n-1}(x)\, dx = 0, \quad \int_a^b \omega(x) r_{n-1}(x)\, dx = \sum_{i=1}^n A_i r_{2n-1}(x_i).$$

Sappiamo che la prima condizione ammette una unica soluzione p_n e che tale soluzione ha n zeri reali e distinti in (a, b). Quindi la prima condizione definisce univocamente i nodi x_i della formula di quadratura *ottimale* che stiamo costruendo: essi devono essere gli zeri del polinomio n-esimo ortogonale rispetto all'intervallo $[a, b]$ e al peso $\omega(x)$. Una volta stabiliti i nodi, per determinare i coefficienti A_i si procede soddisfacendo la seconda condizione, cioè imponendo che la formula sia esatta per tutti i polinomi di grado minore o uguale a $n-1$. Automaticamente, a causa della particolare scelta degli x_i, la formula sarà esatta anche per polinomi di grado maggiore, fino al grado $2n-1$.

Gli A_i dovranno venire necessariamente tutti positivi. Per provare questa affermazione, osserviamo innanzitutto che se P_{2n-1} è un polinomio di grado al più $2n-1$ tale che $P_{2n-1}(x_j) = 0$, $\forall\, j$, allora necessariamente $\int_a^b \omega(x) P_{2n-1}(x)\, dx = 0$ (perché?). Poi notiamo che $l_i(x) - l_i(x)^2$, dove gli l_i sono i soliti polinomi di Lagrange ($l_i(x) = \Pi_{j\neq i}(x - x_j)/\Pi_{j\neq i}(x_i - x_j)$),è effettivamente un polinomio di grado al più $2n-1$ tale che $l_i(x_j) - l_i(x_j)^2 = 0$, $\forall\, i, j$. Quindi, $\int_a^b \omega(x)(l_i(x) - l_i(x)^2)\, dx = 0$, e, di conseguenza, $A_i = \int_a^b \omega(x) l_i(x)^2\, dx > 0$.

Un metodo per costruire l'ennesimo polinomio ortogonale.

Siano Q ed R due funzioni sufficientemente regolari. Osserviamo che

$$\begin{aligned}
\int_a^b Q^{(n)}(x)R(x)\, dx &= Q^{(n-1)}(x)R(x)|_a^b - \int_a^b Q^{(n-1)}(x)R^{(1)}(x)\, dx \\
&= Q^{(n-1)}(x)R(x)|_a^b - Q^{(n-2)}(x)R^{(1)}(x)|_a^b + \int_a^b Q^{(n-2)}(x)R^{(2)}(x)\, dx \\
&= Q^{(n-1)}(x)R(x)|_a^b - Q^{(n-2)}(x)R^{(1)}(x)|_a^b \\
&\quad + \ldots + (-1)^{n-1} Q(x)R^{(n-1)}(x)|_a^b + (-1)^n \int_a^b Q(x)R^{(n)}(x)\, dx\,.
\end{aligned}$$

Dunque, se Q fosse tale che $Q^{(n-1)}(a) = Q^{(n-1)}(b) = 0$, $Q^{(n-2)}(a) = Q^{(n-2)}(b) = 0$, ..., $Q(a) = Q(b) = 0$, allora si avrebbe

$$\int_a^b Q^{(n)}(x)R(x)\, dx = 0,\ \forall\, R(x) \text{ polinomio, } \partial R \leq n-1$$

e se p_n polinomio di grado n monico fosse tale che $\omega(x)p_n(x) = Q^{(n)}(x)$, allora p_n dovrebbe necessariamente essere l'n-esimo polinomio ortogonale rispetto all'intervallo $[a, b]$ e alla funzione peso $\omega(x)$.

Esempio 1. I polinomi di Legendre, ortogonali rispetto a $[-1, 1]$ e $\omega(x) = 1$. Sia $Q(x)$ tale che $Q^{(n-1)}(-1) = Q^{(n-1)}(1) = 0$, $Q^{(n-2)}(-1) = Q^{(n-2)}(1) = 0$, ..., $Q(-1) = Q(1) = 0$, e $Q^{(n)}(x)$ è un polinomio di grado esattamente n. È evidente che Q deve essere un polinomio di grado esattamente $2n$ e che deve avere -1 e 1 come radici di molteplicità n. Quindi, necessariamente, $Q(x) = C_n(x^2-1)^n$ e $p_n(x) = C_n((x^2-1)^n)^{(n)}$. La costante C_n va scelta in modo che p_n risulti monico, visto che p_n così è stato definito. Ad esempio,

$$\begin{aligned}
p_2(x) &= C_2((x^2-1)^2)^{(2)} = C_2(4x(x^2-1))^{(1)} \\
&= C_2(4(x^2-1) + 8x^2) = C_2(12x^2 - 4) = x^2 - \tfrac{1}{3}
\end{aligned}$$

$$
\begin{aligned}
p_3(x) &= C_3((x^2-1)^3)^{(3)} = C_3(6x(x^2-1)^2)^{(2)} \\
&= C_3(6(x^2-1)^2 + 6x2(x^2-1)2x)^{(1)} = C_3(6(x^2-1)^2 + 24x^2(x^2-1))^{(1)} \\
&= C_3(24x(x^2-1) + 48x(x^2-1) + 48x^3) = C_3(120x^3 - 72x) \\
&= x^3 - \tfrac{3}{5}x, \quad x_1 = -\sqrt{\tfrac{3}{5}},\ x_2 = 0, x_3 = \sqrt{\tfrac{3}{5}}
\end{aligned}
$$

Esempio 2. I polinomi di Laguerre, ortogonali rispetto a $(0,+\infty)$ e $\omega(x) = e^{-x}$. Sia $Q(x)$ tale che $Q^{(n-1)}(0) = Q^{(n-1)}(+\infty) = 0$, $Q^{(n-2)}(0) = Q^{(n-2)}(+\infty) = 0$, ..., $Q(0) = Q(+\infty) = 0$, e $Q^{(n)} = e^{-x}p_n(x)$. Se prendiamo $Q(x) = C_n x^n e^{-x}$, si osserva facilmente che tutte le condizioni richieste sono soddisfatte, quindi $p_n(x) = C_n e^x (x^n e^{-x})^{(n)}$, con C_n tale che p_n è monico.

Esempio 3. I polinomi di Hermite, ortogonali rispetto a $(-\infty,+\infty)$ e $\omega(x) = e^{-x^2}$. Sia $Q(x)$ tale che $Q^{(n-1)}(-\infty) = Q^{(n-1)}(+\infty) = 0$, $Q^{(n-2)}(-\infty) = Q^{(n-2)}(+\infty) = 0$, ..., $Q(-\infty) = Q(+\infty) = 0$, $Q^{(n)} = e^{-x^2}p_n(x)$. La scelta $Q(x) = C_n e^{-x^2}$ verifica tutte le condizioni, quindi $p_n(x) = C_n e^{x^2}(e^{-x^2})^{(n)}$, con C_n tale che $p_n(x) = x^n + \ldots$.

VIII LEZIONE

Dati $[a,b] \subset \mathbb{R}^n$ e $\omega(x)$ ben definita e positiva in (a,b) e tale che $\int_a^b \omega(x)\,dx < +\infty$, esiste ed è unica una successione di polinomi $\{p_n(x)\}_{n=0}^{+\infty}$ con p_n di grado n, monico e tale che $\int_a^b \omega(x)p_n(x)q_{n-1}(x)\,dx = 0\ \forall\, q_{n-1}$ polinomio di grado $\leq n-1$. Il generico polinomio p_n di tale successione di *polinomi ortogonali* (rispetto $[a,b]$, ω) ha n zeri reali, distinti, contenuti in (a,b) e, ovviamente, non ha altri zeri in $\mathbb{C}$. Siano essi $x_i = x_i^{(n)}$, $x_1 < x_2 < \ldots < x_n$.

Esercizio. È vero che se ω è simmetrica rispetto al centro dell'intervallo $[a,b]$ allora anche gli x_i lo sono? Cioè, $x_i = \frac{a+b}{2} - \xi_i \Rightarrow x_{n-i+1} = \frac{a+b}{2} + \xi_i$?

Siano ora $A_i = A_i^{(n)}$ gli n numeri reali univocamente definiti imponendo per ogni f polinomio, $\partial f \leq n-1$, la seguente identità

$$
\int_a^b \omega(x)f(x)\,dx = \sum_{i=1}^n A_i f(x_i) \qquad (condizXesattezzaGauss)
$$

(nota: $A_i = \int_a^b \omega(x)l_i(x)\,dx$, $l_i(x) = \Pi_{j\neq i}(x-x_j)/\Pi_{j\neq i}(x_i - x_j)$; nota: gli A_i si ottengono imponendo l'uguaglianza per gli n elementi di una base generica dell'insieme dei polinomi di grado $\leq n-1$).

Esercizio. Osservare che per $[a,b] = [-1,1]$, $\omega(x) = 1/\sqrt{1-x^2}$ si ha $A_i = A = \pi/n\ \forall\, i$.

Sappiamo che allora l'identità (condizXesattezzaGauss) vale in realtà per ogni f polinomio, $\partial f \leq 2n-1$. Più precisamente si può dimostrare che se $f \in C^{2n}([a,b])$ allora

$$
\exists\, \eta \in (a,b) \mid \int_a^b \omega(x)f(x)\,dx = \sum_{i=1}^n A_i f(x_i) + \frac{f^{(2n)(\eta)}}{(2n)!}\int_a^b \omega(x)\Pi_j(x-x_j)^2\,dx
$$

I polinomi di Hermite, Laguerre e Legendre sono, rispettivamente,

$$p_n^{He}(x) = C_n^{He} e^{x^2}(e^{-x^2})^{(n)}, \quad p_n^{La}(x) = C_n^{La} e^{x}(x^n e^{-x})^{(n)}, \quad p_n^{Le}(x) = C_n^{Le}((x^2-1)^n)^{(n)}.$$

Calcoliamo $p_3^{He}(x)$. Si ha

$$\begin{aligned} p_3(x) &= C_3 e^{x^2}(e^{-x^2})^{(3)} = C_3 e^{x^2}(-2xe^{-x^2})^{(2)} \\ &= C_3 e^{x^2}(-2e^{-x^2} + 4x^2 e^{-x^2})^{(1)} = C_3 e^{x^2}(12xe^{-x^2} - 8x^3 e^{-x^2}) \\ &= C_3(12x - 8x^3) = x^3 - \tfrac{3}{2}x = x(x^2 - \tfrac{3}{2}). \end{aligned}$$

Quindi $x_1 = -\sqrt{\frac{3}{2}}$, $x_2 = 0$, $x_3 = \sqrt{\frac{3}{2}}$. Calcoliamo la corrispondente formula di quadratura di Gauss, esatta per polinomi di grado minore o uguale a $2n-1=5$. Essa è del tipo

$$\int_{-\infty}^{+\infty} e^{-x^2} f(x)\,dx = A_1 f(-\sqrt{\frac{3}{2}}) + A_2 f(0) + A_3 f(\sqrt{\frac{3}{2}}) + E$$

dove, sappiamo, $A_1 = A_3$ (ω simm e nodi simm $\Rightarrow$ A_i simmetrici). Imponendo $E=0$ per f polinomi di grado minore o uguale di $n-1=2$, cioè per $f(x) = 1, x, x^2$ (nota: per $f(x) = x$ è superfluo farlo, già sappiamo che in quel caso $E=0$), si ottengono le condizioni: $t = 2A_1 + A_2$, $u = 3A_1$, dove $t = \int_{-\infty}^{+\infty} e^{-x^2}\,dx$, $u = \int_{-\infty}^{+\infty} x^2 e^{-x^2}\,dx$, e, quindi, la seguente formula di quadratura di Gauss-Hermite a tre nodi:

$$\int_{-\infty}^{+\infty} e^{-x^2} f(x)\,dx = \frac{u}{3}(f(-\sqrt{\frac{3}{2}} + f(\sqrt{\frac{3}{2}}) + (t - \frac{2}{3}u) f(0) + E$$

Ricordiamo che la formula rappresenta semplicemente l'integrale $\int_{-\infty}^{+\infty} e^{-x^2} P(x)\,dx$ dove $P(x)$ è il polinomio interpolatore della f nei punti $-\sqrt{\frac{3}{2}}, 0, \sqrt{\frac{3}{2}}$.

Esercizio. Calcolare t e, per parti, u. Osservare che $u = \frac{1}{2}t$.

Calcoliamo $p_4^{He}(x)$. Si ha

$$\begin{aligned} p_4(x) &= C_4 e^{x^2}(e^{-x^2})^{(4)} = C_4 e^{x^2}(12xe^{-x^2} - 8x^3 e^{-x^2})^{(1)} \\ &= C_4 e^{x^2}(12e^{-x^2} - 48x^2 e^{-x^2} + 16x^4 e^{-x^2}) \\ &= C_4(12 - 48x^2 + 16x^4) = x^4 - 3x^2 + \tfrac{3}{4}. \end{aligned}$$

Quindi $x_1 = -\sqrt{\frac{3+\sqrt{6}}{2}}$, $x_2 = -\sqrt{\frac{3-\sqrt{6}}{2}}$, $x_3 = \sqrt{\frac{3-\sqrt{6}}{2}}$, $x_4 = \sqrt{\frac{3+\sqrt{6}}{2}}$. Calcoliamo la corrispondente formula di quadratura di Gauss, esatta per polinomi di grado minore o uguale a $2n-1=7$. Essa è del tipo

$$\int_{-\infty}^{+\infty} e^{-x^2} f(x)\,dx = A_1(f(-\sqrt{\frac{3+\sqrt{6}}{2}}) + f(\sqrt{\frac{3+\sqrt{6}}{2}})) + A_2(f(-\sqrt{\frac{3-\sqrt{6}}{2}}) + f(\sqrt{\frac{3-\sqrt{6}}{2}})) + E.$$

Imponendo $E = 0$ per f polinomi di grado minore o uguale di $n-1 = 3$, cioè per $f(x) = 1, x^2$ (nota: per $f(x) = x, x^3$ è superfluo farlo, già sappiamo che in quei casi $E = 0$), si ottengono le condizioni: $t = 2A_1 + 2A_2$, $u = (3+\sqrt{6})A_1 + (3-\sqrt{6})A_2$, dove t, u sono come prima. Completare l'esercizio.

Calcoliamo $p_3^{La}(x)$. Si ha

$$\begin{aligned} p_3(x) &= C_3 e^x (x^3 e^{-x})^{(3)} = C_3 e^x (3x^2 e^{-x} - x^3 e^{-x})^{(2)} \\ &= C_3 e^x (6x e^{-x} - 6x^2 e^{-x} + x^3 e^{-x})^{(1)} \\ &= C_3 e^x (6e^{-x} - 18x e^{-x} + 9x^2 e^{-x} - x^3 e^{-x}) \\ &= C_3 (6 - 18x + 9x^2 - x^3) = x^3 - 9x^2 + 18x - 6. \end{aligned}$$

Non è semplice calcolare una espressione esplicita di x_1, x_2, x_3. Però si possono calcolare con Newton delle approssimazioni buone quanto si vuole di tali radici di p_3 , dapprima localizzandole in intervalli e poi partendo dall'estremo destro o sinistro di tali intervalli in modo che la successione generata da Newton sia monotona convergente. Si può prevedere che x_1, x_2, x_3 non siano troppo lontano da 0 essendo la funzione peso maggiore vicino a zero.

Poiché $p_3(0) = -6$, $p_3(1) = 4$, $p_3(2) = 2$, $p_3(3) = -6$, $p_3(6) = -6$, $p_3(7) > 0$, $p_3'(3 \pm \sqrt{3}) = 0$, $p_3''(3) = 0$, le radici cercate sono in $(0, 1)$, $(2, 3)$, $(6, 7)$, e, prendendo ξ_0 rispettivamente uguale a 0, 3, 7, la corrispondente successione $\xi_{i+1} = \xi_i - p_3(\xi_i)/p_3'(\xi_i)$, $i = 0, 1, 2, \ldots$, convergerà in modo monotono (da sinistra, da destra, da destra) a x_1, x_2, x_3.

Calcoliamo la corrispondente formula di quadratura di Gauss, esatta per polinomi di grado minore o uguale a $2n - 1 = 5$. Essa è del tipo

$$\int_{-\infty}^{+\infty} e^{-x} f(x)\, dx = A_1 f(x_1) + A_2 f(x_2) + A_3 f(x_3) + E$$

dove, stavolta, i coefficienti A_i potrebbero essere diversi tra loro. Imponendo $E = 0$ per f polinomi di grado minore o uguale di $n - 1 = 2$, cioè per $f(x) = 1, x, x^2$ (nota: stavolta è essenziale farlo anche per $f(x) = x$), si ottengono le condizioni: $t = A_1 + A_2 + A_3$, $u = A_1 x_1 + A_2 x_2 + A_3 x_3$, $v = A_1 x_1^2 + A_2 x_2^2 + A_3 x_3^2$, dove $t = \int_0^{+\infty} e^{-x}\, dx =$, $u = \int_0^{+\infty} x e^{-x}\, dx =$ e $v = \int_0^{+\infty} x^2 e^{-x}\, dx =$.

Ricordiamo che la formula ottenuta rapresenta semplicemente l'integrale $\int_0^{+\infty} e^{-x} P(x)\, dx$ dove $P(x)$ è il polinomio interpolatore della f nei punti x_1, x_2, x_3.

Esercizio. Calcolare le approssimazioni fornite da Newton-Cotes a tre nodi e Gauss-Legendre a tre nodi per l'integrale $\int_1^2 \frac{1}{x}\, dx$, e confrontarle con il valore esatto $\log_e 2$.

Risoluzione numerica di problemi differenziali di Cauchy, Medodo di Eulero, Metodo di Eulero implicito

... da scrivere ...

IX LEZIONE

Risoluzione numerica di problemi differenziali di Cauchy, metodi di Taylor (cenni), di Runge-Kutta (cenni). Introduzione ai metodi multistep (via interpolazione)

In generale, i metodi utilizzati per la risoluzione di problemi differenziali di Cauchy si dividono in metodi a un passo (one-step) e a più passi (multi-step). I metodi di Eulero e di Eulero implicito, visti nella precedente lezione, sono esempi di metodi a un passo. Si possono ricavare metodi a un passo precisi quanto si vuole troncando più avanti lo sviluppo di Taylor della soluzione esatta $y(t)$ del problema di Cauchy, $y(x+h) = y(x) + hy'(x) + \frac{h^2}{2}y''(x) + \ldots + \frac{h^p}{p!}y^{(p)}(x) + \ldots$. Ad esempio, per migliorare Eulero, ovvero per avere una approssimazione $\eta(x+h)$ di $y(x+h)$ tale che $|\eta(x+h) - y(x+h)| = O(|h|^3)$ (nell'ipotesi $\eta(x) = y(x)$), è sufficiente porre

$$\eta(x+h) = y(x)+hy'(x)+\frac{h^2}{2}y''(x) = y(x)+hf(x,y(x))+\frac{h^2}{2}[f_x(x,y(x))+f_y(x,y(x))f(x,y(x))].$$

Si ottiene così il metodo di Taylor di ordine 2 (2 perché $|\eta(x+h) - y(x+h)| = O(|h|^3) \Rightarrow |\frac{\eta(x+h)-y(x)}{h} - \frac{y(x+h)-y(x)}{h} = O(|h|^2)$):

$$\begin{array}{l} \eta(x_0) = y(x_0) = y_0, \text{ per } i = 0, 1, 2, \ldots : \\ \eta(x_i + h) = \eta(x_i) + hf(x_i, \eta(x_i)) + \frac{h^2}{2}[f_x(x_i, \eta(x_i)) + f_y(x_i, \eta(x_i))f(x_i, \eta(x_i))], \\ x_{i+1} = x_i + h. \end{array}$$

Ogni passo di tale metodo richiede la valutazione di tre funzioni, f, f_x, f_y, in $(x_i, \eta(x_i))$. Nel metodo di Taylor di ordine 3, le funzioni da valutare in $(x_i, \eta(x_i))$ sono 6: f, f_x, f_y, f_{xx}, f_{xy}, f_{yy}. Più in generale si vede facilmente che per utilizzare un metodo di Taylor di ordine p occorre valutare $1+2+\ldots+p$ funzioni diverse in $(x_i, \eta(x_i))$. Notiamo che tali calcoli possono essere molto dispendiosi poiché le derivate parziali della f di solito sono più complicate della f. Inoltre, se la f del problema di Cauchy non ammette derivate parziali, diventa impossibile usare i metodi di Taylor per ottenere maggiore precisione nell'integrazione del problema di Cauchy: l'unico metodo di tale classe che si potrebbe usare sarebbe Eulero. Inoltre, i metodi di Taylor di ordine maggiore di 1 non sono facilmente applicabili quando occorre risolvere problemi di Cauchy associati a sistemi di equazioni differenziali. Più precisamente, dato il problema

$$\mathbf{y}'(t) = \mathbf{f}(t, \mathbf{y}(t)),\ t \in (a, b),\ \mathbf{y}(x_0) = \mathbf{y}_0,\ x_0 \in (a, b),$$

si può applicare immediatamente Eulero (o Eulero implicito) per integrarlo,

$$\underline{\eta}(x_i + h) = \underline{\eta}(x_i) + \mathbf{f}(x_i, \underline{\eta}(x_i)),\ x_{i+1} = x_i + h,$$

ma come applicare, ad esempio, il metodo di Taylor di ordine 2? Nel caso dei sistemi, anche solo la definizione di tale metodo sarebbe tutt'altro che banale e, in ogni caso, coinvolgerebbe matrici di derivate parziali di funzioni.

Quindi, anche se la f ($\mathbf{f}$) è sufficientemente regolare, i metodi di Taylor appaiono molto dispendiosi, e, nel caso di sistemi, pressoché inutilizzabili. Per questo motivo essi non si usano molto.

Al loro posto si usa l'altra classe importante di metodi a un passo: i metodi Runge-Kutta. In essi si riesce ad ottenere una maggiore precisione andando a valutare la f, oltre che in $(x_i, \eta(x_i))$, anche in altri punti opportuni, calcolabili con la sola conoscenza di $\eta(x_i)$ ed h. Ci sono metodi Runge-Kutta di ordine p che ad ogni passo richiedono p valutazioni della f (per $p \leq 4$), quindi sono molto meno dispendiosi dei corrispondenti metodi dello stesso ordine di Taylor. Inoltre, essendo definiti solo in termini della f, sono immediatamente applicabili anche nel caso di sistemi di equazioni differenziali. Il loro studio è tuttavia omesso, per ragioni di tempo.

Nei metodi a un passo, $\eta(x_i + h)$ può essere calcolato conoscendo solo h ed $\eta(x_i)$. Nei metodi multi-step occorre conoscere anche altri valori $\eta(x_k)$, $k < i$. Introduciamo ora, usando la teoria dell'interpolazione, una particolare classe di metodi multi-step lineari, omettendo lo studio della consistenza-convergenza di tali metodi.

Sia $y(t)$ la soluzione del seguente problema di Cauchy:

$$y'(t) = f(t, y(t)),\ t \in (a,b),\ y(x_0) = y_0,\ x_0 \in (a,b).$$

Siano $h > 0$ e $x_j = x_0 + jh$, $j = 0, -1, 1, -2, 2, \ldots$. Allora, per $i \geq 0$, $s \geq -1$, posto $l_k(t) = \prod_{j=-s,\, j\neq k}^{1} (t - x_{n+j})/(x_{n+k} - x_{n+j})$, $k = -s, \ldots, 1$, e $B_k = \int_{-i}^{1} l_k(x_n + rh)\, dr$, si ha che

$$\begin{aligned}
y(x_{n+1}) - y(x_{n-i}) &= \int_{x_{n-i}}^{x_{n+1}} y'(t)\, dt = \int_{x_{n-i}}^{x_{n+1}} f(t, y(t))\, dt \\
&= \int_{x_{n-i}}^{x_{n+1}} \Big(\sum_{k=-s}^{1} f(x_{n+k}, y(x_{n+k})) l_k(t) + R(t) \Big)\, dt \\
&= \sum_{k=-s}^{1} f(x_{n+k}, y(x_{n+k})) \int_{x_{n-i}}^{x_{n+1}} l_k(t)\, dt + \int_{x_{n-i}}^{x_{n+1}} R(t)\, dt \\
&= h \sum_{k=-s}^{1} B_k f(x_{n+k}, y(x_{n+k})) + \tilde{R}.
\end{aligned}$$

In altre parole, posto $\varphi(\sigma) = y(x_{n-i}) + h \sum_{k=-s}^{0} B_k f(x_{n+k}, y(x_{n+k})) + hB_1 f(x_{n+1}, \sigma) + \tilde{R}$, il valore $y(x_{n+1})$ è punto fisso di φ. Notiamo inoltre che, se h è sufficientemente piccolo, si ha $|\varphi'(y(x_{n+1}))| = |hB_1 f_y(x_{n+1}, y(x_{n+1}))| < 1$ e quindi la curva $\varphi(\sigma)$ in $y(x_{n+1})$ taglia la bisettrice σ. Ne segue che la curva $\psi(\sigma) = y(x_{n-i}) + h \sum_{k=-s}^{0} B_k f(x_{n+k}, y(x_{n+k})) + hB_1 f(x_{n+1}, \sigma)$ – minima traslazione della φ – taglia la bisettrice σ in un punto vicino a $y(x_{n+1})$, cioè è ben definito un numero $\eta(x_{n+1})$ punto fisso di ψ,

$$\eta(x_{n+1}) = y(x_{n-i}) + h \sum_{k=-s}^{0} B_k f(x_{n+k}, y(x_{n+k})) + hB_1 f(x_{n+1}, \eta(x_{n+1})) = \psi(\eta(x_{n+1})).$$

Potendosi assumere $|\psi'(\eta(x_{n+1}))| = |hB_1 f_y(x_{n+1}, \eta(x_{n+1}))| < 1$, tale punto fisso si può calcolare come limite della successione $\eta(x_{n+1}))^{j+1} = \psi(\eta(x_{n+1}))^j$, $j = 0, 1, \ldots$.

Ovviamente, se $B_1 = 0$, la funzione ψ (φ) non dipende da σ, dunque il valore di $\eta(x_{n+1})$ è ricavabile esplicitamente.

Il calcolo dei coefficienti B_k delle formule multi-step

Sappiamo che i B_k devono soddisfare la seguente identità:

$$y(x_{n+1}) = y(x_{n-i}) + h \sum_{k=-s}^{1} B_k y'(x_{n+k}) + \tilde{R}, \tag{$*$}$$

dove $\tilde{R} = 0$ $(R(t) = 0)$ se $y'(t)$ è un polinomio di grado al più $s+1$, ovvero se $y(t)$, la soluzione del problema di Cauchy, è un polinomio di grado al più $s+2$.

Allora possiamo trovare i B_k imponendo direttamente che l'uguaglianza (*) sia verificata con $\tilde{R} = 0$ quando al posto di $y(t)$ mettiamo polinomi $\varphi_i(t)$ di grado i, $i = 0, 1, 2, \ldots$:

$\tilde{R} = 0$ se $y(x) = 1$: $1 = 1$;

$\tilde{R} = 0$ se $y(x) = x - x_n$: $1 = -i + \sum_{k=-s}^{1} B_k$;

$\tilde{R} = 0$ se $y(x) = (x - x_n)^2$: $1 = i^2 + 2\sum_{k=-s}^{1} B_k k$;

$\tilde{R} = 0$ se $y(x) = (x - x_n)^3$: $1 = -i^3 + 3\sum_{k=-s}^{1} B_k k^2$;

...

Esempi

$s = -1$:

$1 = -i + B_1$,

$\eta(x_{n+1}) = y(x_{n-i}) + h(1 + i)f(x_{n+1}, \eta(x_{n+1}))$ ($i = 0$: Eulero implicito).

$s = 0$:

ordine 1:

$1 = -i + B_0 + B_1$, esempio: $B_1 = 0$, $i = 0$, $B_0 = 1$,

$\eta(x_{n+1}) = y(x_n) + hf(x_n, y(x_n))$ (Eulero).

ordine 2:

$1 = -i + B_0 + B_1$, $1 = i^2 + 2B_1$, $B_0 = (1 + i)^2/2$, $B_1 = (1 - i^2)/2$,

$\eta(x_{n+1}) = y(x_{n-i}) + h\frac{(1+i)^2}{2} f(x_n, y(x_n)) + h\frac{1-i^2}{2} f(x_{n+1}, \eta(x_{n+1}))$

($i = 0$: trapezio; $i = 1$: mid-point).

$s = 1$:

ordine 3:

$1 = -i + B_{-1} + B_0 + B_1$, $1 = i^2 + 2(-B_{-1} + B_1)$, $1 = -i^3 + 3(B_{-1} + B_1)$, $B_0 = 1 + i - \frac{1+i^3}{3}$, $B_1 = \frac{1}{2}(\frac{1+i^3}{3} + \frac{1-i^2}{2})$, $B_{-1} = \frac{1}{2}(\frac{1+i^3}{3} - \frac{1-i^2}{2})$;

$i = 1$: $B_0 = 4/3$, $B_1 = 1/3$, $B_{-1} = 1/3$,

$\eta(x_{n+1}) = y(x_{n-1}) + h[\frac{1}{3} f(x_{n-1}, y(x_{n-1})) + \frac{4}{3} f(x_n, y(x_n)) + \frac{1}{3} f(x_{n+1}, \eta(x_{n+1}))]$ (Milne-Simpson).

Esercizio. Si consideri il problema

$$y'(t) = -2ty(t)^2,\ y(-1) = 1/2,\ y(t) = 1/(1+t^2).$$

Valutare $y(0)$ con le approssimazioni $\eta(0; 1/n)$, $n = 1, 2, 3$, fornite dal metodo di Eulero. Osservare che tali approssimazioni di $y(0)$ migliorano al crescere di n.

Esercizio. Si consideri il problema

$$y'(t) = y(t)^2,\ y(1) = -1,\ y(t) = -1/t.$$

Valutare $y(1+h)$ con i metodi di Eulero, Eulero implicito e trapezio ponendo $h = 1/4$.

Esercizio. Si consideri il problema

$$y'(t) = -200ty(t)^2,\ y(-1) = 1/101,\ y(t) = 1/(1+100t^2).$$

Valutare $y(0)$ con le approssimazioni $\eta(0; 1/n)$, $n \in \mathbb{N}$, fornite dal metodo di Eulero (usare una calcolatrice). Si osserva che inizialmente più n cresce, più le $\eta(0; 1/n)$ approssimano bene $y(0)$, ma poi, per n maggiore di un certo ν, le approssimazioni cominciano a peggiorare. Perché?

Metodi multi-step lineari generici

In un generico metodo multi-step lineare, supponendo noti $y(x_k)$, $k \leq n$, si definisce come approssimazione di $y(x_{n+1})$ il numero $\eta(x_{n+1})$ definito dalla seguente identità

$$\eta(x_{n+1}) + \sum_{j=-i}^{0} C_j y(x_{n+j}) = h \sum_{k=-s}^{0} B_k f(x_{n+k}, y(x_{n+k})) + hB_1 f(x_{n+1}, \eta(x_{n+1})).$$

I coefficienti B_k, C_j, B_1 sono scelti in modo che $\eta(x_{n+1}) = y(x_{n+1})$ quando $y(t)$ è un polinomio generico di grado al più tot. Allora tot-1 si dice ordine del metodo.

Vi sono almeno tre sottoclassi significative di tali metodi:

1) $i = 0$, $C_0 = -1$: Adams. Quelli espliciti sono noti come Adams-Bashforth (es. Eulero), quelli impliciti come Adams-Moulton (es. trapezio).

2) $i = 1$, $C_0 = 0$, $C_{-1} = -1$: se espliciti sono noti come Nyström (mid-point), se impliciti come Milne-Simpson generalizzati (es. Milne-Simpson).

3) $B_k = 0$, $k \leq 0$, $B_1 \neq 0$: backward difference formulas, sono indicati per risolvere problemi stiff .

Nota: un possibile esercizio potrebbe essere di mostrare che un metodo one-step o multi-step ha un certo ordine.

Sul metodo di Eulero

Concludiamo studiando più approfonditamente il metodo di Eulero: convergenza di $\eta(x; h_n)$ per $n \to +\infty$, andamento di $\eta(x_i; h)$ per h fissato, andamento dei numeri $\eta^*(x_i; h)$ e $\eta^*(x; h_n)$ con i quali si ha effettivamente a che fare quando si implementa Eulero su un calcolatore.

Dalle equazioni

$$\eta(x_i+h;h) = \eta(x_i;h) + hf(x_i,\eta(x_i;h)),$$
$$y(x_i+h) = y(x_i) + hf(x_i,y(x_i)) + \tfrac{h^2}{2}y''(\xi_i),\ \xi \in (x_i, x_i+h),$$

segue che

$$\begin{aligned} y(x_i+h)-\eta(x_i+h;h) &= y(x_i)-\eta(x_i;h)+h[f(x_i,y(x_i))-f(x_i,\eta(x_i;h))]+\frac{h^2}{2}y''(\xi_i), \\ |y(x_i+h)-\eta(x_i+h;h)| &\leq (1+|h|L)|y(x_i)-\eta(x_i;h)|+\frac{|h|^2}{2}M \end{aligned}$$

(si suppone che f, f_x, f_y siano limitate su $[a,b]\times\mathbb{R}$). Dunque, grazie al seguente risultato (la cui dimostrazione è lasciata al lettore)

Lemma. Se $\xi_{i+1} \leq (1+\delta)\xi_i + B$, $i=0,1,2,\ldots$, allora $\xi_i \leq e^{i\delta}\xi_0 + B(e^{i\delta}-1)/\delta$

possiamo concludere che

$$|y(x_i)-\eta(x_i;h)| \leq \frac{|h|M}{2L}(e^{i|h|L}-1), \quad i=0,1,2,\ldots$$

$(y(x_0)=\eta(x_0;h))$.

Convergenza: posto $h=h_n := (x-x_0)/n$, fatti n passi, si ottiene un valore $\eta(x;h_n)$ tale che $|y(x)-\eta(x;h_n)| \leq |h_n|M(e^{|x-x_0|L}-1)/(2L)$, dunque per $n\to+\infty$ la successione $|y(x)-\eta(x;h_n)|$ deve tendere a zero.

Andamento: per h fissato positivo, anche piccolissimo, per i grande $\eta(x_i;h)$ può non aver nulla a che fare con $y(x_i)$.

Esercizio. Studiare la propagazione dell'errore dovuto al fatto che si usa il calcolatore, dunque invece di

$$\eta(x_i+h;h) = \eta(x_i;h) + hf(x_i,\eta(x_i;h))$$

si calcola

$$\eta^*(x_i+h;h) = \eta^*(x_i;h) + f(x_i,\eta^*(x_i;h)) + \varepsilon_i,\ |\varepsilon_i| \leq \varepsilon,$$

cioè trovare una maggiorazione per $|\eta(x;h_n)-\eta^*(x;h_n)|$ (usando sempre il *Lemma* di cui sopra). Dedurre da tale maggiorazione che quando n diventa grande, la quantità $|\eta(x;h_n)-\eta^*(x;h_n)|$ potrebbe esplodere.

X LEZIONE

Esercitazioni

• Data l'equazione $x=g(x)$, si discute graficamente la successione $x_{k+1}=g(x_k)$, $k=0,1,\ldots$, al variare della scelta del valore iniziale x_0. Si disegnano le funzioni x e $g(x)$ e si sceglie x_0 sull'asse delle x. Si valuta $x_1 := g(x_0)$ e lo si disegna sull'asse delle x osservando che esso non è altro che l'ascissa del punto di intersezione tra la funzione bisettrice $y=x$

e la retta $y = g(x_0)$. Si valuta poi $x_2 := g(x_1)$ e lo si disegna sull'asse x con lo stesso procedimento. E così via. Da questo studio grafico si possono dedurre le seguenti condizioni sufficienti per la convergenza della successione $x_{k+1} := g(x_k)$, $k = 0, 1, \ldots$,

i) esiste un α tale che $\alpha = g(\alpha)$,

ii) $g(x)$ in $(\alpha-\delta, \alpha+\delta)$ ha il grafico racchiuso nell'insieme aperto delimitato dalla bisettrice $y = x$ e dalla retta $y - g(\alpha) = -(x - \alpha)$ (se $g \in C^1$, quest'ultima condizione è verificata se e solo se $|g'(\alpha)| < 1$),

iii) $x_0 \in (\alpha - \delta, \alpha + \delta)$,

condizioni trovate già nella teoria.

Esempi

L'equazione $x = 1+\log x$ ha $\alpha = 1$ come punto fisso. Si osserva graficamente che se $x_0 > 1$ allora la successione $x_{k+1} = 1 + \log x_k$ converge a 1 in modo monotono, ma lentamente. Se invece $0 < x_0 < 1$, la successione si allontana dal punto fisso e a una certa iterazione non è ben definita. (Notare che $g'(1) = 1$).

L'equazione $x = 2+\log x$ ha due punti fissi, uno maggiore e uno minore di 1, siano essi 1^+ e 1^-. Si osserva graficamente che se x_0 è maggiore di 1^-, allora la successione $x_{k+1} = 2+\log x_k$ converge a 1^+. Se invece $0 < x_0 < 1^-$, la successione si allontana dai punti fissi e a una certa iterazione non è ben definita. (Notare che $|g'(1^+)| < 1$ e $|g'(1^-)| > 1$). Quindi tramite il metodo $x_{k+1} = 2 + \log x_k$ si può ottenere una approssimazione buona quanto si vuole solo di 1^+. Come calcolare l'altra radice 1^-? Basta notare che l'equazione $x = 2+\log x$ è equivalente all'equazione $x = e^{x-2}$ (se ad esempio il log è inbase e) e che la g di quest'ultima è tale che $|g'(1^-)| < 1$ e $|g'(1^+)| > 1$, quindi 1^- è un punto di attrazione per la successione $x_{k+1} = e^{x_k-2}$.

Esercizi

Studiare le equazioni $x = \cos x$, $x = 1/(1 + x^2)$.

Porre $g(x) = 1/(1+x/(1+x)) = (1+x)/(2x+1)$ e studiare la convergenza della successione $x_{k+1} = g(x_k)$ al variare di x_0. ($1/\sqrt{2}$ punto di attrazione; $-1/\sqrt{2}$ punto di repulsione, ...).

Porre $g(x) = 1/(1+x/(1+x/(1+x))) = (1+2x)/(1+3x+x^2)$. Dimostrare che l'equazione $x = g(x)$ ha tre radici reali e distinte. Studiare la convergenza della successione $x_{k+1} = g(x_k)$ al variare di x_0. ($g'(x) = -(2x^2+2x+1)/((1+3x+x^2)^2) < 0$, il denominatore di g si annulla in $(-3 \pm \sqrt{5})/2$, ...).

- Dato $p_n(x) = a_n x^n + \ldots$ polinomio di grado esattamente n ($a_n \neq 0$), minimizzare $\max_{x\in[-1,1]} |p_n(x) - q(x)|$ al variare di q tra tutti i polinomi di grado al più $n - 1$. Si osserva che

$$\begin{aligned}
\max_{x\in[-1,1]} |p_n(x) - q(x)| &= \max_{x\in[-1,1]} |a_n(\tfrac{1}{a_n}p_n(x) - \tfrac{1}{a_n}q(x)| \\
&= |a_n| \max_{x\in[-1,1]} |\tfrac{1}{a_n}p_n(x) - \tfrac{1}{a_n}q(x)|, \\
\min_q \max_{x\in[-1,1]} |p_n(x) - q(x)| &= |a_n| \min_q \max_{x\in[-1,1]} |\tfrac{1}{a_n}p_n(x) - \tfrac{1}{a_n}q(x)| = |a_n| \tfrac{1}{2^{n-1}}
\end{aligned}$$

dove l'ultima uguaglianza è ottenuta scegliendo q tale che $\frac{1}{a_n}p_n(x) - \frac{1}{a_n}q(x) = \frac{1}{2^{n-1}}T_n(x)$, ovvero

$$q(x) = p_n(x) - \frac{a_n}{2^{n-1}} T_n(x).$$

Esempio

Sia $p_3(x) = 1 + x + x^2/2 + x^3/(3!)$. Per ogni $x \in \mathbb{R}$ sappiamo che esiste ξ nel minimo intervallo contenente 0 e x tale che $|e^x - p_3(x)| = e^\xi x^4/(4!)$. Ne segue che $|e^x - p_3(x)| \leq e/(4!)$ se $x \in [-1, 1]$. Sia $q(x) = p_3(x) - \frac{1}{3!4} T_3(x)$ il polinomio di grado al più due che meglio approssima $p_3(x)$ in $[-1, 1]$. Allora, per ogni $x \in [-1, 1]$ si ha

$$|e^x - q(x)| \leq |e^x - p_3(x)| + |p_3(x) - q(x)| \leq e/(4!) + \frac{1}{3!4}.$$

Quindi, economizzando una moltiplicazione ($q(x)$ ha grado 2, minore del grado di p_3) riusciamo ad approssimare e^x ancora abbastanza bene.

• L'approssimare polinomiale di una funzione f fornita dalla tecnica dei minimi quadrati–caso discreto è il polinomio $p_n(x)$ di grado al più n che rende minima la somma di quadrati $\sum_{i=0}^m \omega(x_i)(f(x_i) - q_n(x_i))^2$, essendo $m \geq n$ e gli x_i $m+1$ numeri reali distinti fissati, al variare di q_n tra tutti i polinomi di grado al più n (si suppone f ben definita negli x_i e $\omega(x_i) > 0$).

Alternativamente si può applicare la tecnica dei minimi quadrati–caso continuo, per approssimare f. È infatti univocamente ben definito e si può pensare di calcolare il polinomio $p_n(x)$ di grado al più n che rende l'integrale $\int_a^b \omega(x)(f(x) - q_n(x))^2\,dx$ minimo al variare di q_n tra tutti i polinomi di grado al più n (si suppone f ben definita in (a, b) e $\omega > 0$ in (a,b)). Si osserva che

$$\begin{aligned}
& \int_a^b \omega(x)(f(x) - q_n(x))^2\,dx = \int_a^b \omega(x)f(x)^2\,dx + \int_a^b \omega(x)q_n(x)^2\,dx - 2\int_a^b \omega(x)f(x)q_n(x)\,dx \\
= & \int_a^b \omega(x)f(x)^2\,dx + \int_a^b \omega(x)(\textstyle\sum_{i=0}^n \alpha_i r_i(x))^2\,dx - 2\int_a^b \omega(x)f(x)\sum_{i=0}^n \alpha_i r_i(x)\,dx \\
= & \int_a^b \omega(x)f(x)^2\,dx + \textstyle\sum_{i,j=0}^n \alpha_i\alpha_j \int_a^b \omega(x)r_i(x)r_j(x)\,dx - 2\sum_{i=0}^n \alpha_i \int_a^b \omega(x)f(x)r_i(x)\,dx \\
= & \int_a^b \omega(x)f(x)^2\,dx + \alpha^T A\alpha - 2\alpha^T\mathbf{b},\ \ A_{ij} = \int_a^b \omega(x)r_i(x)r_j(x)\,dx,\ \ b_i = \int_a^b \omega(x)f(x)r_i(x)\,dx.
\end{aligned}$$

Quindi il polinomio cercato è $p_n(x) = \sum_{i=0}^n \alpha_i r_i(x)$ con $\alpha = [\alpha_0\,\alpha_1\ \cdot\ \alpha_n]^T = A^{-1}\mathbf{b}$ (notare che A è definita positiva). Osserviamo che la matrice A varia al variare della base $\{r_i\}$ che si sceglie per l'insieme dei polinomi q_n di grado al più n. Se ad esempio $r_i(x) = x^i$ e $[a, b] = [0, 1]$, allora A è la matrice di Hilbert (dimostrarlo!), ben nota perché il suo numero di condizionamento tende a crescere molto rapidamente (al crescere di n). Per non dover risolvere alcun sistema lineare, conviene scegliere gli r_i come quei polinomi di grado esattamente i tali che $\int_a^b \omega(x)r_i(x)r_j(x)\,dx = 0$ per $i \neq j$. In tale modo A risulta diagonale e per gli α_i si ha la semplice formula $\alpha_i = \int_a^b \omega(x)f(x)r_i(x)\,dx / \int_a^b \omega(x)r_i(x)^2\,dx$. Vedi la teoria dei polinomi ortogonali.

Esercizio. Calcolare l'integrale $\int_{-1}^1 \frac{T_n(x)T_m(x)}{\sqrt{1-x^2}}\,dx$ per ogni $n, m \geq 0$, essendo T_n e T_m rispettivamente i polinomi di Chebycev di grado n ed m.

• Determinare i coefficienti delle formule di quadratura di Gauss-Chebycev, cioè i numeri reali A_i per cui l'uguaglianza

$$\int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} f(x)\, dx = \sum_{i=1}^{n} A_i f(\cos \frac{(2i-1)\pi}{2n})$$

è vera ogni volta che f è un polinomio di grado al più $n-1$. (Automaticamente, per la particolare scelta dei nodi, l'uguaglianza sarà vera per ogni f polinomio di grado al più $2n-1$).

Si osserva che $A_i = \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} l_i(x)\, dx$ dove l_i è il polinomio di grado al più $n-1$ tale che $l_i(\cos \frac{(2j-1)\pi}{2n}) = \delta_{ij}$, $1 \leq i, j \leq n$. Si può dimostrare che gli $A_i^{(n)}$, $i = 1, \ldots, n$, sono tutti uguali tra loro. Dunque $A_i^{(n)} = A^{(n)}$ deve essere tale che $nA^{(n)} = \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}}\, dx$, da cui $A^{(n)} = \pi/n$.

Alcuni metodi per risolvere problemi differenziali ordinari di Cauchy
$y'(t) = f(t, y(t))$, $t \in (a, b)$, $y(x_0) = y_0$, $x_0 \in (a, b)$
(Lambert, p.23). Siano $h > 0$, e $x_i = x_0 + ih$, $i = 0, \pm 1, \pm 2, \ldots$. Nel seguito si elencano diversi metodi per generare approssimazioni $\eta(x_i)$ dei valori esatti $y(x_i)$. Sia $\eta(x_0) = y(x_0)$.

Esempio di Runge-Kutta a tre passi:

$$\begin{aligned}
k_1 &= f(x_i, \eta(x_i)), \\
k_2 &= f(x_i + \frac{1}{3}h, \eta(x_i) + \frac{1}{3}hk_1), \\
k_3 &= f(x_i + \frac{2}{3}h, \eta(x_i) + \frac{2}{3}hk_2), \\
\eta(x_{i+1}) - \eta(x_i) &= \frac{h}{4}[k_1 + 3k_3].
\end{aligned}$$

Esempio di Runge-Kutta a due passi implicito:

$$\begin{aligned}
k_1 &= f(x_i, \eta(x_i)), \\
k_2 &= f(x_i + h, \eta(x_i) + \frac{1}{2}hk_1 + \frac{1}{2}hk_2), \\
\eta(x_{i+1}) - \eta(x_i) &= \frac{h}{2}[k_1 + k_2].
\end{aligned}$$

Esercizio. Il classico metodo Runge-Kutta del quarto ordine

$$\begin{aligned}
&k_1 = f(x_i, \eta(x_i)), \\
&k_2 = f(x_i + h/2, \eta(x_i) + hk_1/2), \\
&k_3 = f(x_i + h/2, \eta(x_i) + hk_2/2), \\
&k_4 = f(x_i + h, \eta(x_i) + hk_3), \\
&\eta(x_i + h) = \eta(x_i) + h[k_1 + 2k_2 + 2k_3 + k_4]/6,
\end{aligned}$$

si può introdurre utilizzando la formula di quadratura di Simpson. É vero?

L'implementazione dei metodi "multi-step" che seguono richiede la conoscenza di "starting values". Infatti nei metodi multi-step l'approssimazione $\eta(x_{i+1})$ è definita (esplicitamente o implicitamente) in termini delle precedenti approssimazioni $\eta(x_i), \ldots, \eta(x_{i-p+1})$, quindi, se $p \geq 2$, per applicare il metodo la prima volta, ovvero per $i = p - 1$, occorre conoscere le approssimazioni $\eta(x_0), \ldots, \eta(x_{p-1})$ e di queste il problema fornisce solo $\eta(x_0) = y(x_0) = y_0$. Ci sono diverse strategie per calcolare le altre. Una consiste nell'applicare $p - 1$ volte un opportuno metodo ad un passo (ad esempio uno dei tre sopra visualizzati). Per altre si veda il Cap. 4 del Lambert.

Esempio di Two-step:

$$\eta(x_{i+1}) - \eta(x_i) = \frac{h}{3}[3f(x_i, \eta(x_i)) - 2f(x_{i-1}, \eta(x_{i-1}))].$$

Esempio di Two-step implicito:

$$\eta(x_{i+1}) + \eta(x_i) - 2\eta(x_{i-1}) = \frac{h}{4}[f(x_{i+1}, \eta(x_{i+1})) + 8f(x_i, \eta(x_i)) + 3f(x_{i-1}, \eta(x_{i-1}))],$$

Esempio di Three-step:

$$\eta(x_{i+1}) + \frac{1}{4}\eta(x_i) - \frac{1}{2}\eta(x_{i-1}) - \frac{3}{4}\eta(x_{i-2}) = \frac{h}{8}[19f(x_i, \eta(x_i)) + 5f(x_{i-2}, \eta(x_{i-2}))].$$

Esempio di Predittore-Correttore (un two-step esplicito combinato con un two-step implicito):

$$\begin{aligned} \eta^*(x_{i+1}) - 3\eta(x_i) + 2\eta(x_{i-1}) &= \frac{h}{2}[f(x_i, \eta(x_i)) - 3f(x_{i-1}, \eta(x_{i-1}))], \\ \eta(x_{i+1}) - \eta(x_{i-1}) &= h[f(x_{i+1}, \eta^*(x_{i+1})) + f(x_{i-1}, \eta(x_{i-1}))]. \end{aligned}$$

Sopra ci sono diversi esempi di metodi impliciti, sia Runge-Kutta che multi-step. Per ogni i il calcolo di $\eta(x_{i+1})$ in tali metodi richiede la risoluzione di una equazione non lineare (o di un sistema di equazioni non lineari, se il metodo è applicato a un sistema di equazioni differenziali), operazione costosa effettuata di solito con il metodo delle approssimazioni successive ("simple iterations") a meno che il problema differenziale è stiff, caso in cui occorre utilizzare tecniche più raffinate.