

Traccia della prima parte del corso di IAN, a.a. 2012-2013

Dario A. Bini

29 aprile 2014

1 Introduzione

Questo documento contiene una traccia più o meno dettagliata di parte degli argomenti trattati nel corso di Istituzioni di Analisi Numerica (IAN). La traccia degli altri argomenti trattati si trova in un secondo documento sulla risoluzione numerica di equazioni differenziali alle derivate parziali e dagli appunti forniti dalla prof.ssa Ornella Menchi durante lo svolgimento del corso di IAN.

Scopo dell'Analisi Numerica è il progetto e l'analisi di metodi di risoluzione di vari problemi matematici mediante l'uso di sole operazioni aritmetiche fra numeri *floating point*. Nel corso dell'insegnamento di Analisi Numerica ci siamo occupati principalmente di problemi dell'algebra lineare e del discreto quali sistemi lineari, trasformate discrete, interpolazione. Nel corso di Calcolo Scientifico l'interesse si è rivolto verso problemi spettrali (autovalori, SVD) e metodi per sistemi lineari di grandi dimensioni provenienti dalle applicazioni. Nel corso di IAN ci occupiamo principalmente di problemi del "continuo". problemi cioè in cui l'oggetto da "calcolare" è una funzione da \mathbb{R}^n in \mathbb{R} e non un vettore. La motivazione e lo scopo saranno comunque gli stessi: progettare e analizzare metodi di risoluzione che impieghino un numero finito di operazioni aritmetiche. Naturalmente la finitezza del numero di operazioni comporterà l'impossibilità di calcolare esattamente la soluzione per cui sarà necessario approssimarla.

I problemi che andiamo a trattare sono:

- approssimazione di funzioni: data $f(x) : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}$ determinare una $g(x)$ in una classe di funzioni facilmente calcolabili che meglio approssimi $g(x)$;
- integrazione approssimata: data $f(x) : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}$ approssimare $\int_a^b f(x)dx$;
- approssimare la soluzione di certe equazioni differenziali alle derivate parziali che si incontrano nelle applicazioni

I metodi numerici diventano necessità in tutte le situazioni in cui la soluzione (funzione) di cui cerchiamo informazioni non si lascia esprimere in modo esplicito in termini di funzioni elementari. Questa situazione si presenta nella quasi totalità dei casi di interesse.

Il corso è suddiviso in

- parte sui polinomi ortogonali: contiene la definizione e l'analisi delle proprietà di questa classe di polinomi con la tassonomia delle principali tipologie di polinomi; saranno evidenziate analogie tra polinomi ortogonali e matrici tridiagonali con proprietà dell'algebra lineare numerica;
- parte sull'applicazione dei polinomi ortogonali alla costruzione di efficienti formule di integrazione approssimata;
- parte che riguarda i metodi per l'approssimazione di funzioni in generale; particolare attenzione è rivolta alla migliore approssimazione in norma 2 e in norma infinito; saranno utilizzate le proprietà dei polinomi ortogonali; saranno sottolineate le relazioni tra approssimazione e interpolazione con particolare attenzione ai problemi del condizionamento; verrà considerato un risultato generale noto come teorema di Korovkin; verranno introdotte e analizzate le funzioni curvilinee
- parte sulla risoluzione numerica di equazioni differenziali alle derivate parziali: sarà fatta la classificazione in equazioni ellittiche, paraboliche e iperboliche con esempi dalle applicazioni; verrà sviluppata la parte relativa al metodo delle differenze finite dove viene data una impostazione il più possibile "matriciale"; è fatto un cenno ai metodi di Rayleigh-Ritz-Galerkin e di collocazione.

2 Polinomi ortogonali

Nel seguito si denota con \mathcal{P}_n lo spazio vettoriale di dimensione $n + 1$ costituito dai polinomi a coefficienti reali di grado al più n e con \mathcal{P} lo spazio vettoriale di tutti i polinomi a coefficienti reali. Inoltre a, b denotano due costanti della retta reale estesa $\mathbb{R} \cup \{+\infty, -\infty\}$ con $a < b$. In questo modo l'intervallo $[a, b]$ può essere anche del tipo $[-\infty, \beta]$, $[\alpha, +\infty]$, con $\alpha, \beta \in \mathbb{R}$, e $[-\infty, +\infty]$.

Sia $\omega(x) : [a, b] \rightarrow \mathbb{R}$ una funzione a valori nella retta reale estesa tale che $\omega(x) > 0$ per $a < x < b$ ed esista finito $\int_a^b f(x)\omega(x)dx$ per ogni polinomio $f(x) \in \mathcal{P}(x)$. Si osserva che l'applicazione che a una coppia di polinomi $(f(x), g(x))$, $f(x), g(x) \in \mathcal{P}$ associa il numero reale

$$\langle f, g \rangle := \int_a^b f(x)g(x)\omega(x)dx \quad (1)$$

è un prodotto scalare sul \mathcal{P} . L'applicazione che a $f(x) \in \mathcal{P}$ associa $(\int_a^b f(x)\omega(x)dx)^{1/2}$ è una norma su \mathcal{P} che viene denotata con $\|f(x)\|$.

Osservazione 1 Il prodotto scalare appena introdotto verifica la proprietà $\langle xf(x), g(x) \rangle = \langle f(x), xg(x) \rangle$. Non tutti i prodotti scalari su \mathcal{P} verificano questa proprietà. Ad esempio, se consideriamo \mathcal{P}_n , e associamo a $p(x) \in \mathcal{P}_n$ la $(n+1)$ -upla dei suoi coefficienti, possiamo rappresentare lo spazio lineare \mathcal{P}_n con \mathbb{R}^{n+1} . Per cui ogni prodotto scalare su \mathbb{R}^{n+1} induce un prodotto scalare su \mathcal{P}_n . In particolare il prodotto scalare euclideo su \mathbb{R}^{n+1} induce il prodotto scalare tra polinomi $\langle p(x), q(x) \rangle = \sum_{i=0}^n p_i q_i$ dove $p(x) = \sum_{i=0}^n p_i x^i$, $q(x) = \sum_{i=0}^n q_i x^i$, che non soddisfa la proprietà $\langle xf(x), g(x) \rangle = \langle f(x), xg(x) \rangle$. \square

Definizione 2 Dato un prodotto scalare $\langle \cdot, \cdot \rangle$, su \mathcal{P} , un insieme di polinomi $\{p_i(x) \in \mathcal{P}_i, i = 0, 1, \dots\}$ tale che $\deg(p_i) = i$ e $\langle p_i, p_j \rangle = 0$ se $i \neq j$ è detto insieme di polinomi ortogonali *relativamente al prodotto scalare* $\langle \cdot, \cdot \rangle$. \square

I polinomi $p_i(x)/h_i^{1/2}$, $h_i = \langle p_i, p_i \rangle$ sono detti polinomi ortonormali. Ad esempio, i polinomi x^i , $i = 0, 1, \dots$, sono ortonormali rispetto al prodotto scalare indotto dal prodotto euclideo.

Esempi classici di funzioni peso che definiscono il prodotto scalare di tipo integrale sono $\omega(x) = 1$ su $[-1, 1]$, $\omega(x) = \sqrt{1-x^2}$ su $[-1, 1]$, $\omega(x) = 1/\sqrt{1-x^2}$ su $[-1, 1]$, $\omega(x) = e^{-x}$ su $[0, +\infty]$, e^{-x^2} su $[-\infty, +\infty]$.

Osservazione 3 Per un generico prodotto scalare su \mathcal{P} , esistono sempre polinomi ortogonali che possono essere costruiti mediante il procedimento di ortogonalizzazione di Gram-Schmidt a partire dai monomi $1, x, x^2, x^3, \dots$, nel seguente modo:

$$p_0(x) = 1, \quad p_k(x) = x^k + \sum_{i=0}^{k-1} \alpha_{k,i} p_i(x), \quad k = 1, 2, \dots \quad (2)$$

$$\alpha_{k,i} = -\langle x^k, p_i(x) \rangle / \langle p_i(x), p_i(x) \rangle, \quad i = 0, \dots, k-1$$

Col prodotto scalare integrale (1) per calcolare i coefficienti dei polinomi ortogonali di grado al più n occorre calcolare, o approssimare numericamente, $n(n+1)/2$ integrali. Inoltre, per calcolare i coefficienti del generico polinomio p_k di grado k occorre eseguire circa k^2 operazioni aritmetiche per cui il costo totale per calcolare i coefficienti di tutti i polinomi p_k , $k = 0, \dots, n$ è di $O(n^3)$ operazioni. Questo però non è il modo migliore di procedere. Come vedremo tra poco ci sono espressioni più semplici per esprimere i coefficienti di $p_k(x)$.

Inoltre per calcolare il valore che $p_k(x)$ assume in un punto ξ usando la (2) occorrono circa $2k$ operazioni aritmetiche avendo però prima calcolato i valori di $p_i(\xi)$, $i = 0, 1, \dots, k-1$ e i coefficienti $\alpha_{k,i}$. Quindi occorrono $O(k^2)$ operazioni. Anche questo calcolo può essere semplificato mediante una diversa rappresentazione dei polinomi ortogonali. \square

La seguente banale osservazione ha una serie di conseguenze meno immediate e di particolare rilevanza.

Osservazione 4 I polinomi ortogonali $p_0(x), p_1(x), \dots, p_n(x)$ sono linearmente indipendenti e quindi costituiscono una base dello spazio vettoriale \mathcal{P}_n dei polinomi di grado minore o uguale ad n . \square

Dalla precedente osservazione segue il

Teorema 5 Se p_0, p_1, \dots , sono polinomi ortogonali rispetto al prodotto scalare (1) allora per ogni polinomio q di grado n vale $\langle p_i, q \rangle = 0$ per ogni $i > n$.

Dim. Per l'osservazione 4 si ha $q(x) = \sum_{j=0}^n \gamma_j p_j(x)$, $\gamma_n \neq 0$. Se $i > n$ vale allora $\langle p_i, q \rangle = \sum_{j=0}^n \gamma_j \langle p_i, p_j \rangle = 0$. \square

La proprietà espressa nel teorema precedente ci permette di dimostrare il seguente risultato interessante.

Teorema 6 Gli zeri dei polinomi ortogonali rispetto al prodotto scalare (1) sono reali e semplici e stanno in (a, b) .

Dim. Siano x_1, \dots, x_j gli zeri reali e distinti di $p_n(x)$ in (a, b) . Per assurdo supponiamo $j < n$. Supponiamo inoltre che k di questi zeri abbiano molteplicità dispari e $j - k$ pari. Numeriamo gli x_i in modo che i primi k abbiano molteplicità dispari. Definiamo

$$q(x) = \begin{cases} 1 & \text{se } k = 0 \\ \prod_{s=1}^k (x - x_s) & \text{se } k > 0 \end{cases}$$

Allora gli zeri in (a, b) di $p_n(x)q(x)$ hanno molteplicità pari e quindi questo prodotto non cambia segno in (a, b) . Conseguentemente

$$\langle p_n, q \rangle = \int_a^b p_n(x)q(x)\omega(x)dx \neq 0$$

che è assurdo per il teorema 5 poiché il grado di $q(x)$ è $k \leq j < n$ \square

Vale il seguente risultato di minima norma

Teorema 7 Tra tutti i polinomi $p(x)$ di grado n che hanno coefficiente del termine di grado massimo uguale a quello di $p_n(x)$, il polinomio $p_n(x)$ è quello di minima norma.

Dim. Ogni polinomio q di grado n che ha lo stesso coefficiente di grado massimo di $p_n(x)$ si lascia scrivere come $q = p_n(x) + q_{n-1}(x)$, dove $q_{n-1}(x)$ è un polinomio di grado al più $n - 1$. Vale quindi

$$\|q\|^2 = \langle q, q \rangle = \langle p_n, p_n \rangle + \langle q_{n-1}, q_{n-1} \rangle$$

essendo $\langle q_{n-1}, p_n \rangle = 0$ per l'ortogonalità. Risulta allora $\|q\|^2 \geq \|p_n\|^2$ dove l'uguaglianza è raggiunta se e solo se $q_{n-1} = 0$ cioè $q = p$. \square

Un'altra conseguenza del teorema 5 è data dal seguente risultato che ha rilevanza sia teorica che computazionale.

Teorema 8 (Ricorrenza a tre termini) *Siano $p_i(x)$, $i = 0, 1, \dots$, un insieme di polinomi ortogonali su $[a, b]$ rispetto a un prodotto scalare (1). Si denoti $p_0(x) = a_0$, $p_1(x) = a_1x + b_1$. Esistono $A_i, B_i, C_i \in \mathbb{R}$ tali che*

$$p_{i+1}(x) = (xA_{i+1} + B_{i+1})p_i(x) - C_i p_{i-1}(x), \quad i \geq 1, \quad A_{i+1}, C_i \neq 0.$$

Inoltre

$$A_{i+1} = \frac{\langle p_{i+1}, p_{i+1} \rangle}{\langle xp_i, p_{i+1} \rangle}, \quad B_{i+1} = -A_{i+1} \frac{\langle xp_i, p_i \rangle}{\langle p_i, p_i \rangle},$$

$$C_i = A_{i+1} \frac{\langle xp_i, p_{i-1} \rangle}{\langle p_{i-1}, p_{i-1} \rangle} = \frac{A_{i+1}}{A_i} \frac{\langle p_i, p_i \rangle}{\langle p_{i-1}, p_{i-1} \rangle}.$$

Denotando con a_i e b_i rispettivamente i coefficienti di x^i e x^{i-1} in $p_i(x)$, vale $A_{i+1} = \frac{a_{i+1}}{a_i}$, $B_{i+1} = \frac{a_{i+1}}{a_i} \left(\frac{b_{i+1}}{a_{i+1}} - \frac{b_i}{a_i} \right)$, $C_i = \frac{a_{i+1}a_{i-1}}{a_i^2} \frac{h_i}{h_{i-1}}$, dove $h_i = \langle p_i, p_i \rangle$.

Dim. Il polinomio $xp_i(x)$ ha grado $i+1$ per cui xp_i, p_i, \dots, p_0 sono linearmente indipendenti. Esistono allora $\alpha_0, \alpha_1, \dots, \alpha_{i+1} \in \mathbb{R}$ tali che

$$p_{i+1} = \alpha_0 p_0 + \dots + \alpha_i p_i + \alpha_{i+1} xp_i.$$

Per $j \neq i+1$ vale

$$0 = \langle p_j, p_{i+1} \rangle = \alpha_j \langle p_j, p_j \rangle + \alpha_{i+1} \langle p_j, xp_i \rangle = \alpha_j \langle p_j, p_j \rangle + \alpha_{i+1} \langle xp_j, p_i \rangle$$

Poiché se $j \leq i-2$ il polinomio xp_j ha grado minore di i , per il teorema 5 è $\langle xp_j, p_i \rangle = 0$. Quindi $\alpha_j = 0$ per $j = 0, \dots, i-2$ per cui

$$p_{i+1} = \alpha_{i+1} xp_i + \alpha_i p_i + \alpha_{i-1} p_{i-1}. \quad (3)$$

Vale quindi il risultato del teorema con $A_{i+1} = \alpha_{i+1}$, $B_{i+1} = \alpha_i$, $C_i = -\alpha_{i-1}$. Le costanti A_i, B_i, C_i possono essere facilmente espresse in termini di prodotti scalari. Infatti moltiplicando scalarmente per p_{i+1}, p_i e p_{i-1} entrambi i membri della (3) si ottiene

$$\langle p_{i+1}, p_{i+1} \rangle = A_{i+1} \langle xp_i, p_{i+1} \rangle,$$

$$A_{i+1} \langle xp_i, p_i \rangle + B_{i+1} \langle p_i, p_i \rangle = 0$$

$$A_{i+1} \langle p_{i-1}, xp_i \rangle - C_i \langle p_{i-1}, p_{i-1} \rangle = 0$$

da cui $\langle xp_i(x), p_{i+1}(x) \rangle \neq 0$ e si ricavano le espressioni per A_{i+1}, B_{i+1}, C_i . Le altre relazioni seguono da un confronto diretto dei polinomi. Inoltre, poiché $\langle p_{i+1}, p_{i+1} \rangle \neq 0$ ne segue che $A_{i+1} \neq 0$, inoltre dall'espressione $C_i = \frac{A_{i+1}}{A_i} \frac{\langle p_i, p_i \rangle}{\langle p_{i-1}, p_{i-1} \rangle}$ segue $C_i \neq 0$. \square

Osservazione 9 Poiché i polinomi del teorema 8 sono definiti a meno di una costante moltiplicativa, possiamo scegliere i polinomi monici in modo che $a_i = 1$ per ogni $i \geq 1$ per cui $A_i = 1$ per $i \geq 1$. In questo modo risulta

$$A_{i+1} = 1, \quad B_{i+1} = -\frac{\langle xp_i, p_i \rangle}{\langle p_i, p_i \rangle}, \quad C_i = \frac{\langle p_i, p_i \rangle}{\langle p_{i-1}, p_{i-1} \rangle} > 0.$$

□

Osservazione 10 Si noti che la relazione a tre termini data nel teorema 8 ci dice che $p_{i-1}(x)$ è il resto della divisione di $p_{i+1}(x)$ per $p_i(x)$. Per cui i polinomi ortogonali $p_{n+1}(x), p_n(x), \dots, p_0(x)$ possono essere visti come i polinomi generati dall'algoritmo Euclideo applicato a $p_{n+1}(x)$ e $p_n(x)$. □

Osservazione 11 Fissato i , i coefficienti $A_{i+1} = 1, B_{i+1}, C_i$ richiedono il calcolo dei prodotti scalari $\langle p_i, p_i \rangle, \langle p_{i-1}, p_{i-1} \rangle, \langle xp_i, p_i \rangle$. Per cui il loro calcolo per $i = 1, \dots, k$ comporta solamente $2k$ prodotti scalari.

Per calcolare il valore di $p_k(\xi)$ usando la relazione a tre termini, supponendo di avere a disposizione i coefficienti $A_{i+1} = 1, B_{i+1}$ e C_i e di aver calcolato i valori $p_i(\xi)$ per $i = 0, \dots, k-1$, bastano 4 operazioni aritmetiche. Quindi per calcolare tutti i valori $p_i(\xi)$ per $i = 1, \dots, k$ bastano $4k$ operazioni aritmetiche.

In modo analogo si vede che i coefficienti di tutti i polinomi ortogonali di grado al più k sono calcolabili in $O(k^2)$ operazioni.

Un confronto col metodo di ortogonalizzazione di Gram-Schmidt mostra che l'uso della relazione ricorrente a tre termini abbassa sostanzialmente i costi computazionali delle operazioni con polinomi ortogonali. □

Teorema 12 (Formula di Christoffel-Darboux) *Vale*

$$(x-y) \sum_{i=0}^n \frac{1}{h_i} p_i(x) p_i(y) = \gamma_n [p_{n+1}(x) p_n(y) - p_{n+1}(y) p_n(x)] \quad (4)$$

dove $\gamma_n = \frac{a_n}{a_{n+1} h_n} = \frac{1}{h_n A_{n+1}}, h_i = \langle p_i, p_i \rangle$.

Dim. Procediamo per induzione. Se $n = 0$ allora devo dimostrare che $(x-y) \frac{1}{h_0} p_0(x) p_0(y) = \frac{a_0}{a_1 h_0} [p_1(x) p_0(y) - p_1(y) p_0(x)]$. Poiché $p_0(x) = a_0$, e $p_1(x) = a_1 x + b_1$, allora il secondo membro della relazione precedente diventa

$$\frac{a_0}{a_1 h_0} [(a_1 x + b_1) a_0 - (a_1 y + b_1) a_0] = \frac{a_0}{a_1 h_0} a_1 (x-y) a_0,$$

che coincide con $\frac{a_0^2}{h_0} (x-y)$.

Per l'implicazione $n-1 \rightarrow n$, consideriamo il secondo membro di (4) dove sostituiamo al posto di $p_{n+1}(x)$ e di $p_{n+1}(y)$ l'espressione ottenuta con la relazione

a tre termini. Si ottiene quindi

$$\begin{aligned} \gamma_n [((xA_{n+1} + B_{n+1})p_n(x) - C_n p_{n-1}(x))p_n(y) \\ - ((yA_{n+1} + B_{n+1})p_n(y) - C_n p_{n-1}(y))p_n(x)]. \end{aligned}$$

Semplificando si arriva a

$$(x - y)\gamma_n A_{n+1} p_n(x)p_n(y) + \gamma_n C_n (p_n(x)p_{n-1}(y) - p_n(y)p_{n-1}(x))$$

Per l'ipotesi induttiva il secondo addendo coincide con

$$\frac{C_n \gamma_n}{\gamma_{n-1}} (x - y) \sum_{i=0}^{n-1} \frac{1}{h_i} p_i(x)p_i(y).$$

La tesi segue dal fatto che $\frac{\gamma_n C_n}{\gamma_{n-1}} = 1$ e $\gamma_n A_{n+1} = 1/h_n$, per la definizione di γ_n e per le relazioni date nel teorema 8. \square

Osservazione 13 Una dimostrazione alternativa della formula di Christoffel-Darboux è la seguente. Vale

$$\begin{aligned} p_{i+1}(x) &= (A_{i+1}x + B_{i+1})p_i(x) - C_i p_{i-1}(x) && \text{multiplico per } p_i(y) \\ p_{i+1}(x)p_i(y) &= (A_{i+1}x + B_{i+1})p_i(x)p_i(y) - C_i p_{i-1}(x)p_i(y) && \text{scambio variabili} \\ p_{i+1}(y)p_i(x) &= (A_{i+1}y + B_{i+1})p_i(y)p_i(x) - C_i p_{i-1}(y)p_i(x) && \text{sottraggo} \\ p_{i+1}(x)p_i(y) - p_{i+1}(y)p_i(x) & & & \\ &= p_i(x)p_i(y)[x - y]A_{i+1} - C_i [p_{i-1}(x)p_i(y) - p_{i-1}(y)p_i(x)] \end{aligned}$$

Quindi

$$\begin{aligned} [x - y]p_i(x)p_i(y)A_{i+1} &= p_{i+1}(x)p_i(y) - p_{i+1}(y)p_i(x) \\ &\quad + C_i [p_{i-1}(x)p_i(y) - p_{i-1}(y)p_i(x)] \\ \frac{1}{h_i} [x - y]p_i(x)p_i(y) &= \frac{1}{h_i A_{i+1}} [p_{i+1}(x)p_i(y) - p_{i+1}(y)p_i(x)] \\ &\quad + \frac{C_i}{h_i A_{i+1}} [p_{i-1}(x)p_i(y) - p_{i-1}(y)p_i(x)] \end{aligned}$$

dove $C_i = (h_i/h_{i-1})(A_{i+1}/A_i)$. Si ottiene allora

$$\begin{aligned} \frac{1}{h_i} [x - y]p_i(x)p_i(y) &= \frac{1}{h_i A_{i+1}} [p_{i+1}(x)p_i(y) - p_{i+1}(y)p_i(x)] \\ &\quad + \frac{1}{h_{i-1} A_i} [p_{i-1}(x)p_i(y) - p_{i-1}(y)p_i(x)] \end{aligned}$$

Sommando, i termini si elidono a due a due e si ottiene la tesi. \square

Osservazione 14 Applicando la formula di Christoffel Darboux per $x = x_i$ e $y = x_j$, dove x_1, \dots, x_{n+1} sono gli zeri di $p_{n+1}(x)$, si ottiene

$$\sum_{k=0}^n \frac{1}{h_k} p_k(x_i) p_k(x_j) = 0, \quad i \neq j.$$

La matrice $V = (v_{i,j})_{i,j=1,n+1}$, $v_{i,j} = p_{i-1}(x_j)/\sqrt{h_{i-1}}$ è tale che $V^T V$ è una matrice diagonale. Cioè V ha le colonne v_j ortogonali. Possiamo scalare le colonne in modo da farle diventare ortonormali, basta per questo dividerle per la loro norma euclidea

$$\|v_j\| = \sqrt{\sigma_j}, \quad \sigma_j = \sum_{i=0}^n \frac{p_i(x_j)^2}{h_i}.$$

Quindi la matrice $W = V \text{diag}(\sigma_1^{-1}, \dots, \sigma_{n+1}^{-1})$ verifica la proprietà $W^T W = I$ per cui è ortogonale e quindi $W W^T = I$. Quest'ultima relazione si scrive come $\sum_{k=1}^{n+1} w_{i,k} w_{j,k} = \delta_{i,j}$, da cui

$$\sum_{k=1}^{n+1} \frac{p_i(x_k)}{\sqrt{h_i}} \frac{p_j(x_k)}{\sqrt{h_j}} \sigma_k = \delta_{i,j}, \quad \sigma_k = 1 / \sum_{s=0}^n \frac{p_s(x_k)^2}{h_s}. \quad (5)$$

Questa espressione fornisce una sorta di *ortogonalità "discreta"* dei polinomi $p_i(x)$ relativa ai punti x_k e ai pesi σ_k . In particolare, se associamo ai polinomi $p_j(x)$ i vettori $\mathbf{u}^{(j)} = (u_i^{(j)})$ per $j = 0, \dots, n$, ottenuti campionando i polinomi negli zeri di $p_{n+1}(x)$, cioè tali che $u_i^{(j)} = p_j(x_i)$, abbiamo che $\mathbf{u}^{(0)}, \dots, \mathbf{u}^{(n)}$ sono ortogonali rispetto al prodotto scalare (discreto pesato)

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{n+1} u_i v_i \sigma_i.$$

□

Osservazione 15 Se su uno spazio vettoriale \mathcal{V}_n di dimensione n sono assegnati due prodotti scalari $\langle \cdot, \cdot \rangle'$ e $\langle \cdot, \cdot \rangle''$ che coincidono su una base v_1, \dots, v_n di \mathcal{V}_n , cioè $\langle v_i, v_j \rangle' = \langle v_i, v_j \rangle''$, allora tali prodotti scalari coincidono dappertutto, cioè $\langle u, v \rangle' = \langle u, v \rangle''$ per ogni $u, v \in \mathcal{V}_n$. Questo può essere facilmente dimostrato rappresentando u e v nella base v_1, \dots, v_n e applicando la bilinearità del prodotto scalare.

Se consideriamo allora come \mathcal{V}_n lo spazio dei polinomi di grado al più $n-1$ abbiamo che i due prodotti scalari $\langle u, v \rangle' = \int_a^b u(x)v(x)\omega(x)dx$ e $\langle \cdot, \cdot \rangle'' = \sum_{k=1}^n u(x_k)v(x_k)\sigma_k$, dove x_k per $k = 1, \dots, n$ denotano gli zeri del polinomio ortogonale di grado n , coincidono sulla base dei polinomi ortogonali. Quindi coincidono dappertutto. In particolare, se $u(x)$ e $v(x)$ sono polinomi di grado al più $n-1$ si ha

$$\int_a^b u(x)v(x)\omega(x)dx = \sum_{k=1}^n u(x_k)v(x_k)\sigma_k.$$

Inoltre, poiché un polinomio $q(x)$ di grado al più $m = 2n - 2$ lo possiamo fattorizzare come prodotto di due polinomi $q_1(x)$ e $q_2(x)$ di grado al più $n - 1$, cioè $q(x) = q_1(x)q_2(x)$ allora l'integrale $\int_a^b q(x)\omega(x)dx$ possiamo vederlo come prodotto scalare $\langle q_1, q_2 \rangle$ e calcolarlo con la formula di integrazione

$$\int_a^b q(x)\omega(x)dx = \sum_{k=1}^n q(x_k)\sigma_k.$$

Per la precisione occorre dire che il ragionamento precedente vale se la fattorizzazione del polinomio $q(x)$ è sul campo reale. Però non tutti i polinomi di grado $2n - 2$ ammettono una fattorizzazione sul campo reale in due fattori di grado $n - 1$. Si pensi ad esempio a $x^2 + 1 = (x + i)(x - i)$ con i unità immaginaria. Il risultato però può essere esteso facilmente al caso di fattori complessi che esistono sempre per il teorema fondamentale dell'algebra. Infatti un polinomio $p(x)$ a coefficienti complessi si può scrivere come $p(x) = p_R(x) + ip_I(x)$ dove $p_R(x)$, $p_I(x)$ sono polinomi a coefficienti reali. Per cui la parte reale e la parte immaginaria del prodotto di due polinomi a coefficienti complessi possono essere scritti come somma di prodotti di polinomi a coefficienti reali. Per ciascuno dei prodotti può essere applicata la formula di integrazione discreta. \square

Osservazione 16 Dati due polinomi $p(x)$ e $q(y)$ il *bezoutiano* $B(x, y)$ di p e q è il polinomio in due variabili $(p(x)q(y) - p(y)q(x))/(x - y)$. Il bezoutiano ha diverse particolarità interessanti che sono utili in computer algebra. La formula di Cristoffel-Darboux aggiunge una ulteriore proprietà: se p e q sono polinomi ortogonali di grado n e $n + 1$ il loro bezoutiano è somma di prodotti di polinomi ortogonali. Si associa al bezoutiano $B(x, y)$ la matrice $B = (b_{i,j})$ tale che $B(x, y) = \sum_i \sum_j b_{i,j} x^i y^j$ chiamata *matrice bezoutiana*. È immediato verificare che B è simmetrica e che $B(x, y) = \mathbf{x}^T B \mathbf{y}$ dove $\mathbf{x} = (x^i)_{i=0, n-1}$, $\mathbf{y} = (y^i)_{i=0, n-1}$. Poiché $p_k(x)p_k(y) = \mathbf{x}^T \mathbf{p}^{(k)} \mathbf{p}^{(k)T} \mathbf{y}$, dove $\mathbf{p}^{(k)}$ è il vettore dei coefficienti del polinomio $p_k(x)$, la formula di Christoffel Darboux permette di scrivere la matrice bezoutiana di p_n e p_{n+1} come somma delle n diadi simmetriche $\frac{1}{h_k} \mathbf{p}^{(k)} \mathbf{p}^{(k)T}$. La stessa somma di diadi simmetriche permette di scrivere la matrice bezoutiana come $B = L^T L$, dove L è una matrice triangolare inferiore che sulla riga i -esima ha i coefficienti del polinomio $p_i(x)/\sqrt{h_i}$.

Una proprietà interessante della matrice bezoutiana è che la sua fattorizzazione UL a blocchi fornisce quozienti e resti generati dall'algoritmo euclideo applicato a $p(x)$ e $q(x)$. Inoltre la matrice bezoutiana è non singolare se e solo se i polinomi $p(x)$ e $q(x)$ sono primi tra loro. \square

2.1 Polinomi ortogonali e matrici tridiagonali

Sia $T_n(x)$ la matrice definita da

$$T_n(x) = \begin{bmatrix} a_1x + b_1 & -a_0 & & & & \\ -C_1 & A_2x + B_2 & -1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & \ddots & -1 \\ & & & & -C_{n-1} & A_nx + B_n \end{bmatrix}$$

Applicando la regola di Laplace per calcolare $\det T_n(x)$ lungo l'ultima riga si scopre che

$$\det T_n(x) = (A_nx + B_n) \det T_{n-1}(x) - C_{n-1} \det T_{n-2}(x)$$

inoltre $\det T_1(x) = a_1x + b_1 = p_1(x)$ e $\det T_2(x) = (A_2x + B_2)p_1(x) - C_1p_0(x)$. Per cui vale $\det T_n(x) = p_n(x)$.

È interessante osservare che

$$T_n(x) \begin{bmatrix} p_0(x) \\ p_1(x) \\ \vdots \\ p_{n-1}(x) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ p_n(x) \end{bmatrix}$$

per cui il vettore di componenti $(p_i(x))_{i=0:n-1}$ sta nel nucleo di $T_n(x)$ se x è uno zero di $p_n(x)$.

Osservazione 17 Se i polinomi ortogonali vengono normalizzati in modo che $A_i = 1$ allora gli zeri di $p_n(x)$ possono essere visti come gli autovalori della matrice $T_n := -\text{trid}_n(-C_i, B_i, -1)$, dove per semplicità di notazione si è posto $B_1 = b_1$, e gli autovettori corrispondenti hanno per componenti i valori dei polinomi ortogonali calcolati negli zeri di $p_n(x)$. Vale infatti $T_n(x) = xI - T_n$. Si osservi ancora che la matrice ottenuta cambiando segno agli elementi sopra e sotto diagonali ha gli stessi autovalori della matrice originale e autovettori con componenti a segno alterno. \square

Osservazione 18 Assumendo di aver normalizzato i polinomi in modo che $a_0 = A_i = 1$, $i = 1, \dots, n+1$, poiché i C_i sono tutti positivi, è possibile determinare una matrice diagonale $D = \text{diag}_n(d_1, \dots, d_n)$ tale che $D^{-1}T_nD$ è simmetrica. Tale matrice è data da $d_i = \sqrt{h_{i-1}} = \langle p_{i-1}, p_{i-1} \rangle^{1/2}$, $i = 1, \dots, n-1$ e vale

$$D^{-1}T_n(x)D = \begin{bmatrix} b_1 + x & -\sqrt{C_1} & & & & \\ -\sqrt{C_1} & B_2 + x & -\sqrt{C_2} & & & \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & \ddots & -\sqrt{C_{n-1}} \\ & & & & -\sqrt{C_{n-1}} & B_n + x \end{bmatrix}$$

Gli autovettori $\mathbf{v}^{(j)}$ tali che $D^{-1}T_{n+1}(x_j)D\mathbf{v}^{(j)} = 0$ sono $\mathbf{v}^{(j)} = D^{-1}\mathbf{u}^{(j)}$ con $\mathbf{u}^{(j)} = (p_0(x_j), \dots, p_{n-1}(x_j))^T$, dove i polinomi ortogonali $p_i(x)$ sono monici. Dall'ortogonalità dei vettori $\mathbf{v}^{(j)}$ segue che

$$\sum_{k=0}^n \frac{1}{h_k} p_k(x_i) p_k(x_j) = 0, \quad i \neq j.$$

Si osservi che $p_k(x)/\sqrt{h_k}$ è il polinomio ortogonale normalizzato rispetto alla sua norma indotta dal prodotto scalare. Questa proprietà era stata ottenuta anche dalla formula di Christoffel Darboux. \square

Le proprietà degli zeri dei polinomi ortogonali sono quindi strettamente legate alle proprietà degli autovalori delle matrici tridiagonali simmetriche. Nel prossimo paragrafo riportiamo un teorema molto utile, di interesse generale, che ci permette di scoprire proprietà interessanti degli zeri dei polinomi ortogonali

2.2 Il teorema del minimax di Courant-Fischer

Data $A \in \mathbb{R}^{n \times n}$ simmetrica e $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} \neq 0$ si definisce il quoziente di Rayleigh di A in \mathbf{x} l'espressione

$$\frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$

Osservazione 19 Il gradiente del quoziente di Rayleigh è

$$\frac{2}{\mathbf{x}^T \mathbf{x}} (A - \lambda I) \mathbf{x}, \quad \lambda = \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}}.$$

\square

Quindi i punti stazionari del quoziente di Rayleigh sono gli autovettori di A e i valori presi dal quoziente in corrispondenza dei punti stazionari sono i corrispondenti autovalori. Questo porta al seguente primo risultato

Teorema 20 *Siano $\lambda_1 \geq \dots \geq \lambda_n$ gli autovalori della matrice simmetrica $A \in \mathbb{R}^{n \times n}$. Vale*

$$\max_{\mathbf{x} \neq 0} \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \lambda_1, \quad \min_{\mathbf{x} \neq 0} \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \lambda_n.$$

Si può dimostrare molto di più:

Teorema 21 (Teorema di Courant-Fischer del minimax) *Vale*

$$\min_{\substack{U \subset \mathbb{R}^n \\ \dim U = k}} \max_{\substack{\mathbf{x} \in U \\ \mathbf{x} \neq 0}} \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \lambda_{n-k+1}, \quad k = 1, 2, \dots, n$$

$$\max_{\substack{U \subset \mathbb{R}^n \\ \dim U = k}} \min_{\substack{\mathbf{x} \in U \\ \mathbf{x} \neq 0}} \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \lambda_k, \quad k = 1, 2, \dots, n$$

Dim. Si dimostra la versione maxmin. Siano \mathbf{x}_i gli autovettori di A ortonormali e $S = \text{span}(\mathbf{x}_k, \dots, \mathbf{x}_n)$. Allora per ogni sottospazio V_k di dimensione k vale $S \cap V_k \neq \{0\}$ poiché $\dim(S) = n - k + 1$ e $\dim(V_k) = k$. Allora esiste $\mathbf{x} \in S \cap V_k$, $\mathbf{x} = \sum_{i=k}^n \alpha_i \mathbf{x}_i \neq 0$. Vale

$$\frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \frac{\sum_{i=k}^n |\alpha_i|^2 \lambda_i}{\sum_{i=k}^n |\alpha_i|^2} \leq \lambda_k \frac{\sum_{i=k}^n |\alpha_i|^2}{\sum_{i=k}^n |\alpha_i|^2} = \lambda_k.$$

Quindi il minimo su ogni V_k del quoziente di Rayleigh è minore o uguale a λ_k e quindi anche il massimo al variare di V_k dei minimi è minore o uguale a λ_k . Basta quindi fare vedere che esiste un \widehat{V}_k speciale in cui il minimo del quoziente di Rayleigh vale λ_k . Per questo basta scegliere $\widehat{V}_k = \text{span}(\mathbf{x}_1, \dots, \mathbf{x}_k)$. Infatti per $\mathbf{x} \in \widehat{V}_k$ risulta $\mathbf{x} = \sum_{i=1}^k \alpha_i \mathbf{x}_i$ per cui

$$\mathbf{x}^T A \mathbf{x} / \mathbf{x}^T \mathbf{x} = \sum_{i=1}^k \lambda_i \alpha_i^2 \mathbf{x}_i^T \mathbf{x}_i / \sum_{i=1}^k \alpha_i^2 \mathbf{x}_i^T \mathbf{x}_i \geq \lambda_k.$$

La versione MinMax del teorema si ottiene applicando la versione MaxMin alla matrice $-A$. \square

Alcuni corollari interessanti del teorema del minimax.

Corollario 22 *A matrice reale simmetrica $n \times n$ di autovalori $\alpha_1 \geq \dots \geq \alpha_n$. $U \in \mathbb{R}^{n \times (n-1)}$ tale che $U^T U = I_{n-1}$. Allora gli autovalori $\beta_1 \geq \dots \geq \beta_{n-1}$ di $B = U^T A U$ sono tali che*

$$\alpha_1 \geq \beta_1 \geq \alpha_2 \geq \beta_2 \geq \dots \geq \beta_{n-1} \geq \alpha_n.$$

e si dice che gli autovalori di B separano gli autovalori di A .

Dim. Vale

$$\beta_k = \max_{W_k} \min_{\mathbf{y} \in W_k \setminus 0} \frac{\mathbf{y}^T B \mathbf{y}}{\mathbf{y}^T \mathbf{y}} = \min_{\mathbf{y} \in \widehat{W}_k \setminus 0} \frac{\mathbf{y}^T B \mathbf{y}}{\mathbf{y}^T \mathbf{y}} = \min_{\mathbf{y} \in \widehat{W}_k \setminus 0} \frac{\mathbf{y}^T U^T A U \mathbf{y}}{\mathbf{y}^T \mathbf{y}}$$

dove \widehat{W}_k è il sottospazio dove viene preso il massimo. Sia

$$\widehat{V}_k = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = U \mathbf{y}, \mathbf{y} \in \widehat{W}_k\}.$$

Vale $\dim(\widehat{V}_k) = k$ e

$$\frac{\mathbf{y}^T U^T A U \mathbf{y}}{\mathbf{y}^T \mathbf{y}} = \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$

per cui

$$\beta_k = \min_{\mathbf{y} \in \widehat{W}_k \setminus \{0\}} \frac{\mathbf{y}^T U^T A U \mathbf{y}}{\mathbf{y}^T \mathbf{y}} = \min_{\mathbf{x} \in \widehat{V}_k \setminus \{0\}} \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \leq \max_{V_k} \min_{\mathbf{x} \in V_k \setminus \{0\}} \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \alpha_k.$$

La disuguaglianza $\beta_{k-1} \geq \alpha_k$ si ottiene applicando il risultato a $-A$ e $-B$. \square

Corollario 23 *Se A_n è matrice tridiagonale simmetrica, gli autovalori di una qualsiasi sottomatrice principale di A_n di dimensione $(n-1) \times (n-1)$ separano gli autovalori di A .*

Come conseguenza del teorema del minimax segue l'importante proprietà degli zeri dei polinomi ortogonali

Teorema 24 *Gli zeri del polinomio ortogonale $p_n(x)$ separano strettamente gli zeri del polinomio ortogonale $p_{n+1}(x)$.*

Dim. La separazione con disuguaglianza debole segue dal corollario 23 applicato alla matrice tridiagonale simmetrica ottenuta simmetrizzando la matrice $\text{trid}_n(-C_i, B_i, -1)$ alla luce delle osservazioni 17, 18. Per dimostrare la disuguaglianza stretta si supponga per assurdo che λ sia zero di p_n e di p_{n-1} . Dalla relazione a tre termini segue che λ è zero di p_{n-1} . Procedendo per induzione si conclude che λ è zero di p_0 che è assurdo. \square

Si riportano a titolo di curiosità due altri corollari di cui non si dà dimostrazione.

Corollario 25 *Se A, B, C sono matrici reali simmetriche tali che $A = B + C$ allora per gli autovalori $\alpha_i, \beta_i, \gamma_i$ ordinati in modo non crescente vale*

$$\beta_i + \gamma_n \leq \alpha_i \leq \beta_i + \gamma_1.$$

Corollario 26 *Se A, B, C sono matrici reali simmetriche tali che $A = B + C$ e $C = \mathbf{u}\mathbf{u}^T$ per un vettore non nullo \mathbf{u} , allora per gli autovalori α_i, β_i di A e B ordinati in modo non crescente vale*

$$\beta_i \leq \alpha_i \leq \beta_{i+1}$$

3 Rappresentazione di polinomi ortogonali

Oltre alla rappresentazione data mediante la relazione ricorrente a tre termini esistono altri modi per rappresentare i polinomi ortogonali. In questo paragrafo esaminiamo due rappresentazioni diverse, quella basata sulla matrice dei momenti e quella data dalla formula di Rodrigues.

3.1 Matrice dei momenti

Le quantità $\mu_k = \int_a^b x^k \omega(x) dx$, $k = 0, 1, \dots, n$, vengono dette *momenti*. Vale il seguente risultato

Teorema 27 *I seguenti polinomi sono ortogonali rispetto al prodotto scalare*
 $\langle f, g \rangle = \int_a^b f(x)g(x)\omega(x)dx$

$$p_n(x) = \det \begin{bmatrix} \mu_0 & \mu_1 & \mu_2 & \dots & \mu_n \\ \mu_1 & \mu_2 & \mu_3 & \dots & \mu_{n+1} \\ \mu_2 & \mu_3 & \mu_4 & \dots & \mu_{n+2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \mu_{n-1} & \mu_n & \mu_{n+1} & \dots & \mu_{2n-1} \\ 1 & x & x^2 & \dots & x^n \end{bmatrix} \quad (6)$$

Dim. Basta verificare che $\langle x^k, p_n(x) \rangle = 0$ per $k = 0, 1, \dots, n-1$. Per la linearità dell'integrale e la multilinearità del determinante si ha che $\langle x^k, p_n(x) \rangle$ è uguale al determinante della matrice che si ottiene sostituendo l'ultima riga della matrice in (6) con $(\int_a^b x^k, \int_a^b x^{k+1}, \dots, \int_a^b x^{k+n})$, cioè $(\mu_k, \mu_{k+1}, \dots, \mu_{k+n})$. Questo determinante è nullo poiché la matrice ha due righe uguali. \square

Osservazione 28 Col prodotto scalare su $[0, 1]$ definito da $\langle f, g \rangle = \int_0^1 f(x)g(x)dx$, la sottomatrice principale di testa $n \times n$ della matrice dei momenti è la matrice di Hilbert di elementi $(1/(i+j-1))$ per $i, j = 1, \dots, n$. Le sottomatrici principali di testa di dimensione minore o uguale a n di una matrice dei momenti sono *matrici di Hankel*. Cioè i loro elementi dipendono dalla somma degli indici e quindi sono costanti lungo le anti-diagonali. Si può dimostrare che l'inversa di una matrice di Hankel non singolare è una matrice di Bezout. \square

Osservazione 29 Si ricorda che se A è una matrice reale simmetrica definita positiva allora l'applicazione $(u, v) \rightarrow u^T A v$ da $\mathbb{R}^n \times \mathbb{R}^n$ in \mathbb{R} è un prodotto scalare. Se denotiamo H_n la matrice di elementi μ^{i+j-1} , $i, j = 1, \dots, n$, si ha $\mu_{i+j-1} = \langle x^{i-1}, x^{j-1} \rangle$ per $i, j = 1, \dots, n$. Per cui $u^T H_n v = \sum_i \sum_j u_i \mu_{i+j-1} v_j = \langle \sum_i u_i x^{i-1}, \sum_j v_j x^{j-1} \rangle$ non è altro che il prodotto scalare tra il polinomio di coefficienti u_1, \dots, u_n e il polinomio di coefficienti v_1, \dots, v_n . Cioè H_n descrive il prodotto scalare di tipo integrale tra polinomi in termini dei loro coefficienti.

Si osserva ancora che se $H_n = LL^T$ è la fattorizzazione di Cholesky di H_n , dove L è triangolare inferiore $n \times n$, allora dalla condizione $(L^{-1})^T H_n L^{-1} = I$ segue che le righe di $(L^{-1})^T$ sono i coefficienti dei polinomi ortogonali normalizzati in modo da avere norma 1. Avevamo inoltre puntualizzato nell'Osservazione 16, come conseguenza della formula di Christoffel Darboux, che la matrice di Bezout B_n di $p_n(x)$ e $p_{n-1}(x)$ si fattorizza nel prodotto $B = \widehat{L}^T \widehat{L}$ dove \widehat{L} è triangolare inferiore e le colonne di L hanno per elementi i coefficienti del generico polinomio ortogonale normalizzato per cui vale $\widehat{L} = L^{-1}$. Allora, confrontando la relazione $= \widehat{L}^T \widehat{L}$ e $H^{-1} = (L^{-1})^T L^{-1}$ si deduce che $H_n = B_n^{-1}$. \square

3.2 Formula di Rodrigues

Teorema 30 Sia $s(x) \in C^n[a, b]$, $s^{(k)}(a) = s^{(k)}(b) = 0$, $k = 0, 1, \dots, n-1$. Allora la funzione $t(x) = s^{(n)}(x)/\omega(x)$ è ortogonale a ogni polinomio di grado al più $n-1$ col prodotto scalare $\langle f, g \rangle = \int_a^b f(x)g(x)\omega(x)dx$.

Dim. Sia $q(x)$ polinomio di grado $k \leq n-1$. Allora $\langle q, s^{(n)}/\omega(n) \rangle = \int_a^b \omega(x)q(x)\frac{s^{(n)}}{\omega(x)}dx = \int_a^b q(x)s^{(n)}(x)dx$. Integrando per parti si ha

$$\begin{aligned} \left\langle q, \frac{s^{(n)}(x)}{\omega(x)} \right\rangle &= [q(x)s^{(n-1)}(x)]_a^b - \int_a^b q'(x)s^{(n-1)}(x)dx \\ &= - \int_a^b q'(x)s^{(n-1)}(x)dx = \dots = (-1)^n \int_a^b q^{(n)}(x)s(x)dx = 0 \end{aligned}$$

poiché $q^{(n)} = 0$ essendo il grado di q minore di n . □

Osservazione 31 Nelle ipotesi del teorema, se $s^{(n)}(x)/\omega(x)$ è un polinomio di grado n allora esso coincide con $p_n(x)$. □

Ciò permette di esprimere i polinomi ortogonali mediante la *Formula di Rodrigues*

$$\begin{aligned} p_n(x) &= \frac{\beta_n}{\omega(x)} \frac{d^n}{dx^n} s_n(x), \quad n = 0, 1, \dots, \\ \beta_n &\in \mathbb{R}, \quad s_n(x) \in C^n[a, b], \quad s_n^{(k)}(a) = s_n^{(k)}(b) = 0, \quad k = 0, 1, \dots, n-1, \quad (7) \\ &\frac{1}{\omega(x)} \frac{d^n}{dx^n} s_n(x), \quad \text{polinomio di grado } n. \end{aligned}$$

I valori dei β_n sono scelti in modo da normalizzare i polinomi in modo più conveniente a seconda dei casi.

4 Polinomi ortogonali specifici

Nella tabella 1 si riportano le definizioni dei polinomi ortogonali classici di Legendre, Chebyshev di prima e di seconda specie, Laguerre ed Hermite.

I polinomi di Legendre e di Chebyshev sono casi particolari dei *polinomi ultrasferici*, noti anche come polinomi di *Gegenbauer* cioè i polinomi ortogonali su $[-1, 1]$ rispetto al peso $\omega = (1-x^2)^\alpha$, $\alpha > -1$. Per questi polinomi è possibile dare una espressione dei coefficienti in termini della funzione *Gamma* di Eulero

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt, \quad x > 0.$$

$[a, b]$	$\omega(x)$	$s_n(x)$	Nome
$[-1, 1]$	1	$(1 - x^2)^n$	Legendre
$[-1, 1]$	$(1 - x^2)^{-1/2}$	$(1 - x^2)^{n-1/2}$	Chebyshev di prima specie
$[-1, 1]$	$(1 - x^2)^{1/2}$	$(1 - x^2)^{n+1/2}$	Chebyshev di seconda specie
$[0, +\infty]$	e^{-x}	$e^{-x}x^n$	Laguerre
$[-\infty, +\infty]$	e^{-x^2}	e^{-x^2}	Hermite

Tabella 1: Tassonomia dei principali polinomi ortogonali

Tale funzione è estendibile a $x \in \mathbb{C}$, $x \neq 0, -1, -2, \dots$, mediante l'espressione

$$\Gamma(x) = \frac{1}{z} \prod_{n=1}^{\infty} \frac{(1 + \frac{1}{n})^x}{1 + \frac{x}{n}}.$$

Non solo, ma è possibile dare una forma esplicita alla funzione $s(x)$ nella formula di Rodrigues, infatti vale $s(x) = (x^2 - 1)^{n+\alpha}$.

Osservazione 32 La Funzione Gamma di Eulero gode di molte proprietà interessanti tra cui

$$\begin{aligned} \Gamma(x+1) &= x\Gamma(x), \quad x \in \mathbb{R}, \quad x \neq 0, -1, -2, \dots \\ \Gamma(1) &= 1 \\ \Gamma(n+1) &= n!, \quad \text{se } n \text{ è intero} \end{aligned}$$

Vale inoltre

$$\frac{d^n x^\alpha}{dx^n} = \alpha(\alpha-1)\dots(\alpha-n+1)x^{\alpha-n} = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha-m+1)}x^{\alpha-n}$$

Per ulteriori proprietà si veda [2]. □

Teorema 33 Se $\omega(x) = (1 - x^2)^\alpha$, $\alpha > -1$, $s_n(x) = (1 - x^2)^{\alpha+n}$, allora $\frac{s_n^{(n)}(x)}{\omega(x)}$ è polinomio di grado n e vale $s_n^{(k)}(1) = s_n^{(k)}(-1) = 0$, $k = 0, 1, \dots, n-1$. Vale inoltre

$$\frac{s_n^{(n)}(x)}{\omega(x)} = \sum_{j=0}^n (-1)^j \binom{n}{j} \frac{\Gamma(\alpha+n+1)^2}{\Gamma(\alpha+j+1)\Gamma(\alpha+n-j+1)} (1-x)^{n-j} (1+x)^j.$$

Dim. Se $n = 0$ è $\frac{s_0^{(0)}}{\omega(x)} = \frac{(1-x^2)^\alpha}{(1-x^2)^\alpha} = 1$. Se $n > 0$ pongo $z = \alpha + n$ e ho

$$\begin{aligned} \frac{d^k}{dx^k} (1-x^2)^z &= \frac{d^k}{dx^k} [(1-x)^z (1+x)^z] \\ &= \sum_{j=0}^k \binom{k}{j} \frac{d^j}{dx^j} (1-x)^z \frac{d^{k-j}}{dx^{k-j}} (1+x)^z \end{aligned}$$

D'altro canto vale

$$\frac{d^m}{dx^m} x^\theta = \frac{\Gamma(\theta + 1)}{\Gamma(\theta - m + 1)} x^{\theta - m}$$

da cui

$$\begin{aligned} \frac{d^k}{dx^k} (1 - x^2)^z &= \sum_{j=0}^k \binom{k}{j} (-1)^j \frac{\Gamma(z + 1)}{\Gamma(z - j + 1)} (1 - x)^{z-j} \frac{\Gamma(z + 1)}{\Gamma(z - k + j + 1)} (1 + x)^{z-k+j} \\ &= (1 - x^2)^{z-k} \sum_{j=0}^k \binom{k}{j} (-1)^j \frac{\Gamma(z + 1)^2 (1 - x)^{k-j} (1 + x)^j}{\Gamma(z - j + 1) \Gamma(z - k + j + 1)} \end{aligned}$$

che diviso per $\omega(x) = (1 - x^2)^\alpha$ è un polinomio in x . Inoltre, per $k = 0, 1, \dots, n-1$ è $z - k \geq \alpha + 1 > 0$ per cui $s_n^{(k)}(-1) = s_n^{(k)}(1) = 0$. Infine per $k = n$ vale $s_n^{(n)}(x) = (1 - x^2)^{z-n} \phi(x)$, dove $\phi(x)$ è un polinomio, per cui $\frac{s_n^{(n)}(x)}{(1 - x^2)^\alpha}$ è un polinomio \square

Per i polinomi ultrasferici $p_{n,\alpha}(x)$ vale allora la formula

$$p_{n,\alpha}(x) = \beta_{n,\alpha} \frac{1}{(1 - x^2)^\alpha} \frac{d^n}{dx^n} (1 - x^2)^{\alpha+n}$$

Dal teorema precedente si possono ricavare le espressioni di a_n, h_n in funzione di Γ .

Si osservi che vale ancora $p_{n,\alpha}(-x) = (-1)^n p_{n,\alpha}(x)$ per cui $p_{n,\alpha}$ ha coefficienti nulli per le potenze che hanno esponente con la stessa parità di $n + 1$. Conseguentemente il coefficienti B_n nella relazione ricorrente a tre termini sono nulli.

4.1 Polinomi di Legendre

I polinomi di Legendre sono polinomi ultrasferici con $\alpha = 0$, e la relazione ricorrente a tre termini è data da

$$\begin{aligned} p_{n+1}(x) &= \frac{2n+1}{n+1} x p_n(x) - \frac{n}{n+1} p_{n-1}(x) \\ p_0 &= 1, \quad p_1 = x \end{aligned}$$

In particolare

$$p_0(x) = 1, \quad p_1(x) = x, \quad p_2(x) = \frac{1}{2}(3x^2 - 1), \quad p_3(x) = \frac{1}{2}(5x^3 - 3x)$$

Infatti, scrivendo `sort(zeri)'` (l'apice serve per fare scrivere gli autovalori come vettore riga) Octave fornisce i valori

```
ans =
-0.93247  -0.66121  -0.23862   0.23862   0.66121   0.93247
```

4.2 Polinomi di Chebyshev di prima specie

I polinomi di Chebyshev di prima specie, che rientrano nella classe dei polinomi ultrasferici con $\alpha = -1/2$, soddisfano la relazione ricorrente a tre termini

$$\begin{aligned} T_{n+1}(x) &= 2xT_n(x) - T_{n-1}(x) \\ T_0(x) &= 1, \quad T_1(x) = x \end{aligned}$$

Ponendo $x = \cos \theta$, $0 \leq \theta \leq \pi$, dalla relazione a tre termini segue che $T_n(\cos \theta) = \cos n\theta$. La dimostrazione può essere fatta per induzione. Infatti per $n = 0, 1$ la proprietà è banalmente verificata. In generale, per l'ipotesi induttiva vale

$$T_{n+1}(\cos \theta) = 2 \cos \theta \cos n\theta - \cos(n-1)\theta.$$

Da cui

$$\begin{aligned} T_{n+1}(\cos \theta) &= 2 \cos \theta \cos n\theta - \cos n\theta \cos \theta - \sin n\theta \sin \theta \\ &= \cos \theta \cos n\theta - \sin n\theta \sin \theta = \cos(n+1)\theta. \end{aligned}$$

Questo fatto permette di dare una espressione esplicita agli zeri $x_k^{(n)}$ di $T_n(x)$. Infatti vale $x_k^{(n)} = \cos \frac{(2k-1)\pi}{2n}$, $k = 1, \dots, n$. Inoltre su $[-1, 1]$ il polinomio $T_n(x)$ è sempre compreso tra -1 e 1 .

Si osservi anche che $T_n(\cos \theta) = \pm 1$ se e solo se $\cos(n\theta) = \pm 1$, se e solo se $n\theta = k\pi$, se e solo se $\theta = k\pi/n$, $k = 0, 1, \dots, n$. Cioè nell'intervallo $[-1, 1]$ il polinomio $T_n(x)$ assume $n+1$ volte i valori massimi e minimi.

Teorema 34 *Tra tutti i polinomi monici di grado $i \geq 1$ il polinomio monico $\frac{1}{2^{n-1}}T_n(x)$ è quello che ha minima norma infinito su $[-1, 1]$*

Dim. Si è già osservato che $T_n(x)$ assume il valore massimo 1 e il valore minimo -1 per $n+1$ volte. Quindi $\|T_n(x)\|_\infty = 1$ e $\|T_n(x)/2^{n-1}\|_\infty = 1/2^{n-1}$. Supponiamo per assurdo che esista un polinomio p_n monico di grado n tale che $\|p_n(x)\|_\infty < 1/2^{n-1}$. Il polinomio $p_n(x)$ ha massimo minore di $1/2^{n-1}$ e minimo maggiore di $-1/2^{n-1}$. Quindi il polinomio differenza $q(x) = p_n(x) - T_n(x)/2^{n-1}$ di grado al più $n-1$ è negativo dove T_n ha massimo e positivo dove T_n ha minimo. Ciò accade in $n+1$ punti di $[-1, 1]$. Quindi $q(x)$ ha n zeri. Il che è assurdo. \square

La matrice tridiagonale associata ai polinomi di Chebyshev di prima specie è

$$\begin{bmatrix} x & -1 & & & \\ -1 & 2x & -1 & & \\ & -1 & 2x & -1 & \\ & & -1 & 2x & -1 \\ & & & \ddots & \ddots & \ddots \end{bmatrix}$$

Un polinomio $q(x)$ di grado n può essere rappresentato nella base di Chebyshev come $q(x) = \sum_{k=0}^n a_k T_k(x)$. Data l'ortogonalità dei polinomi $T_k(x)$, i valori di a_k possono essere calcolati come

$$a_k = \langle q(x), T_k(x) \rangle / \langle T_k(x), T_k(x) \rangle.$$

Poiché vale la proprietà di ortogonalità discreta (5) si ha

$$\sum_{k=0}^{n-1} T_i(x_k) T_j(x_k) \sigma_k = \delta_{i,j} \sqrt{h_i} \sqrt{h_j}, \quad \sigma_k = 1 / \sum_{s=0}^{n-1} \frac{T_s(x_k)^2}{h_s}$$

dove x_s sono gli zeri di $T_n(x)$, i coefficienti a_k possono essere calcolati come

$$a_k = \sum_{s=0}^{n-1} q(x_s) T_k(x_s) \sigma_s / \sum_{s=0}^{n-1} T_k(x_s) T_k(x_s) \sigma_s.$$

La somma infinita

$$f(x) = \sum_{k=0}^{\infty} a_k T_k(x)$$

se convergente, definisce una funzione $f(x)$ ed è chiamata *serie di Chebyshev* di $f(x)$. Data una funzione $f(x)$ le condizioni a cui deve soddisfare $f(x)$ affinché esista convergente la serie di Chebyshev sono analoghe a quelle valide per la convergenza della serie di Fourier.

Osservazione 35 Per i polinomi di Chebyshev di prima specie i valori di h_k sono dati da

$$\langle T_k(x), T_k(x) \rangle = \int_{-1}^1 T_k(x)^2 \frac{1}{\sqrt{1-x^2}} dx.$$

Con la sostituzione di variabile $x = \cos \theta$ si ottiene quindi

$$h_k = \int_0^{\pi} \cos^2 k\theta d\theta = \begin{cases} \frac{\pi}{2} & \text{se } k \neq 0 \\ \pi & \text{se } k = 0. \end{cases}$$

Inoltre, poiché gli zeri di $T_n(x)$ sono $x_k = \cos \frac{(2k-1)\pi}{2n}$ e vale $T_s(x) = \cos(s\theta)$ con $x = \cos \theta$, si ha $T_s(x_k) = \cos(s \frac{(2k-1)\pi}{2n})$ per cui, $1/\sigma_k = \sum_{s=0}^{n-1} T_s(x_k)^2 / h_s = \frac{2}{\pi} (1/2 + \sum_{s=1}^{n-1} T_s(x_k)^2) = \frac{n}{\pi}$.

Quindi la proprietà di ortogonalità discreta dei polinomi di Chebyshev diventa

$$\sum_{k=1}^n T_i(x_k) T_j(x_k) = \begin{cases} \frac{n}{2} \delta_{i,j} & \text{se } i \neq 0 \\ n \delta_{i,j} & \text{se } i = 0. \end{cases}$$

La matrice $C = (c_{i,j})$ di elementi $c_{i,j} = T_j(x_{i-1})$ definisce la trasformata discreta dei coseni di secondo tipo http://en.wikipedia.org/wiki/Discrete_cosine_transform e gode della proprietà che $C^T C = n \text{diag}(1, 1/2, \dots, 1/2)$. \square

Listing 1: Function cheby1

```
function t=cheby1(m,n)
% function t=cheby1(m,n)
% output: t e' la matrice mxn la cui colonna j-esima contiene i valori
% del polinomio di Chebyshev di prima specie grado j-1 campionato nei
% punti x_i=(-1+ i*2/(m-1)) per i=0,1,...,m-1
t = ones(m,n);
x = [-1:2/(m-1):1]';
t(:,2) = x;
for j=3:n
    t(:,j) = 2*x.*t(:,j-1) - t(:,j-2);
end
```

La function Octave riportata nel listing 1 crea una matrice \mathbf{t} di dimensioni $m \times n$ che ha nella colonna j -esima i valori di $T_{j-1}(x)$ calcolati in m punti equispaziati tra -1 e 1 . Mentre la function riportata nel listing 2 crea una matrice \mathbf{t} di dimensioni $n \times n$ che ha nella colonna j -esima i valori di $T_{j-1}(x)$ calcolati negli n zeri del polinomio di Chebyshev $T_n(x)$ di prima specie di grado n .

In questo modo, con i seguenti comandi

```
t = cheby1(100,6);
x = t(:,2);
plot(x,t(:,2),x,t(:,3),x,t(:,4),x,t(:,5),x,t(:,6));
```

si tracciano i grafici dei primi 6 polinomi di Chebyshev di prima specie riportati nella figura 1. Mentre scrivendo

```
t=chebyshev1(6);
t'*t
```

si ottiene la matrice numericamente diagonale

```
ans =

    6.00000    0.00000    0.00000    0.00000   -0.00000    0.00000
    0.00000    3.00000    0.00000   -0.00000    0.00000   -0.00000
    0.00000    0.00000    3.00000    0.00000   -0.00000    0.00000
    0.00000   -0.00000    0.00000    3.00000    0.00000   -0.00000
   -0.00000    0.00000   -0.00000    0.00000    3.00000    0.00000
    0.00000   -0.00000    0.00000   -0.00000    0.00000    3.00000
```

che conferma l'ortogonalità delle colonne di \mathbf{t} per $n = 6$.

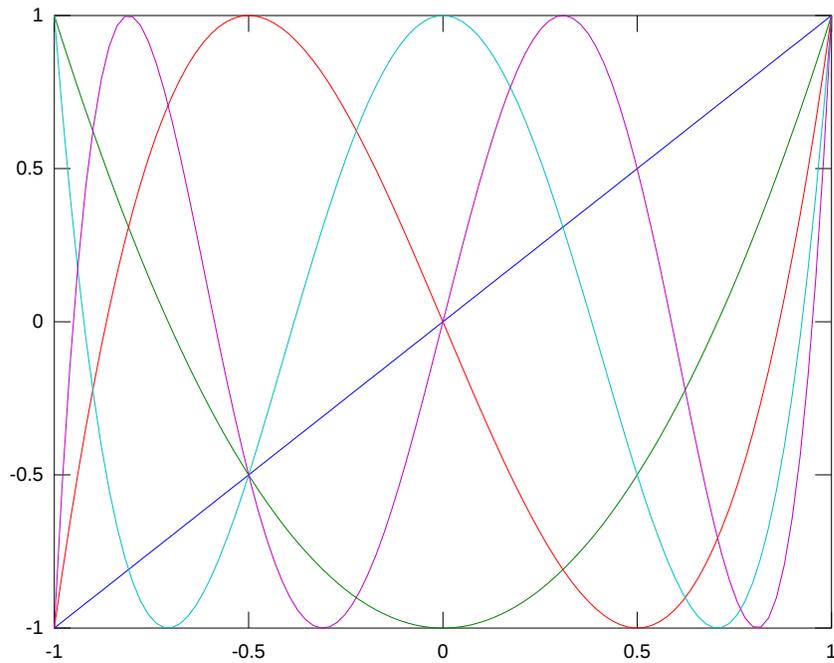


Figura 1: Polinomi di Chebyshev di prima specie

Listing 2: Function chebyshev1

```
function t=chebyshev1(n)
% function t=chebyshev1(n)
% output: t e' la matrice nxn la cui colonna j-esima contiene i valori
% del polinomio di Chebyshev di prima specie grado j-1 campionato negli
% zeri del polinomio di Chebyshev di grado n: x_k=cos((2k-1)pi/(2n))
% per k=1,...,n
t = ones(n);
x = cos([pi/(2*n): pi/n : (2*n-1)*pi/(2*n)]');
t(:,2) = x;
for j=3:n
    t(:,j) = 2*x.*t(:,j-1) - t(:,j-2);
end
```

4.3 Polinomi di Chebyshev di seconda specie

Anche i polinomi di Chebyshev di seconda specie rientrano nella classe dei polinomi ultrasferici. Sono definiti da $\alpha = 1/2$ e soddisfano la relazione ricorrente

$$\begin{aligned} U_{n+1}(x) &= 2xU_n(x) - U_{n-1} \\ U_0(x) &= 1, \quad U_1(x) = 2x \end{aligned}$$

Dalla relazione a tre termini segue che ponendo $x = \cos \theta$ risulta

$$U_n(\cos \theta) = \frac{\sin((n+1)\theta)}{\sin \theta}$$

Quindi gli zeri di $U_n(x)$ sono $x_k^{(n)} = \cos \frac{k\pi}{n+1}$, $k = 1, \dots, n$.

Si osserva che

$$\frac{dT_n}{dx} = \frac{dT_n(\cos \theta)}{d\theta} \cdot \frac{d\theta}{dx} = \frac{d \cos(n\theta)}{d\theta} / \frac{d \cos \theta}{d\theta} = n \frac{\sin(n\theta)}{\sin \theta}$$

che implica

$$T'_n(x) = nU_{n-1}(x).$$

La matrice tridiagonale associata ai polinomi di Chebyshev di seconda specie è

$$\begin{bmatrix} 2x & -1 & & & & \\ -1 & 2x & -1 & & & \\ & -1 & 2x & -1 & & \\ & & -1 & 2x & -1 & \\ & & & \ddots & \ddots & \ddots \end{bmatrix}$$

La function Octave riportata nel listing 3 crea una matrice u di dimensioni $m \times n$ che ha nella colonna j -esima i valori di $U_{j-1}(x)$ calcolati in m punti equispaziati tra -1 e 1 . Mentre la function riportata nel listing 4 crea una matrice u di dimensioni $n \times n$ che ha nella colonna j -esima i valori di $U_{j-1}(x)$ calcolati negli n zeri del polinomio di Chebyshev $U_n(x)$ di seconda specie di grado n .

In questo modo, con i seguenti comandi

```
u = cheby1(100,6);
x = [-1:2/99:1];
plot(x,u(:,2),x,u(:,3),x,u(:,4),x,u(:,5),x,u(:,6));
```

si tracciano i grafici dei primi 6 polinomi di Chebyshev di seconda specie riportati nella figura 2. Mentre scrivendo

```
u = chebyshev2(6);
u*u'
```

si ottiene la matrice numericamente diagonale

Listing 3: Function cheby2

```
function t=cheby2(m,n)
% function t=cheby1(m,n)
% output: t e' la matrice mxn la cui colonna j-esima contiene i valori
% del polinomio di Chebyshev di seconda specie grado j-1 campionato nei
% punti x_i=(-1+ i*2/(m-1)) per i=0,1,...,m-1
t = ones(m,n);
x = [-1:2/(m-1):1]';
t(:,2) = 2*x;
for j=3:n
    t(:,j) = 2*x.*t(:,j-1) - t(:,j-2);
end
```

Listing 4: Function chebyshev2

```
function t=chebyshev2(n)
% function t=chebyshev2(n)
% output: t e' la matrice nxn la cui colonna j-esima contiene i valori
% del polinomio di Chebyshev di seconda specie grado j-1 campionato negli
% zeri del polinomio di Chebyshev di grado n: x_k=cos(k*pi/(n+1))
% per k=1,...,n
t = ones(n);
x = cos([pi/(n+1): pi/(n+1) : n*pi/(n+1)]');
t(:,2) = 2*x;
for j=3:n
    t(:,j) = 2*x.*t(:,j-1) - t(:,j-2);
end
```

1.8592e+01	-1.7048e-15	1.3849e-15	-6.6331e-16	5.0209e-16	-4.0007e-16
-1.7048e-15	5.7259e+00	-4.9922e-16	9.9204e-17	-1.6046e-16	-5.1337e-16
1.3849e-15	-4.9922e-16	3.6823e+00	2.9891e-16	-1.9125e-16	8.2611e-16
-6.6331e-16	9.9204e-17	2.9891e-16	3.6823e+00	-3.9411e-17	-5.7718e-16
5.0209e-16	-1.6046e-16	-1.9125e-16	-3.9411e-17	5.7259e+00	2.0095e-15
-4.0007e-16	-5.1337e-16	8.2611e-16	-5.7718e-16	2.0095e-15	1.8592e+01

4.4 Polinomi di Laguerre

Peso $\omega(x) = e^{-x}$, intervallo $[0, +\infty]$. Coefficienti della ricorrenza a tre termini:
 $A_{n+1} = -\frac{1}{n+1}$, $B_{n+1} = \frac{2n+1}{n+1}$, $C_n = \frac{n}{n+1}$.

$$L_{n+1}(x) = \frac{1}{n+1} ((2n+1-x)L_n(x) - nL_{n-1}(x))$$

$$L_0(x) = 1, \quad L_1(x) = 1 - x$$

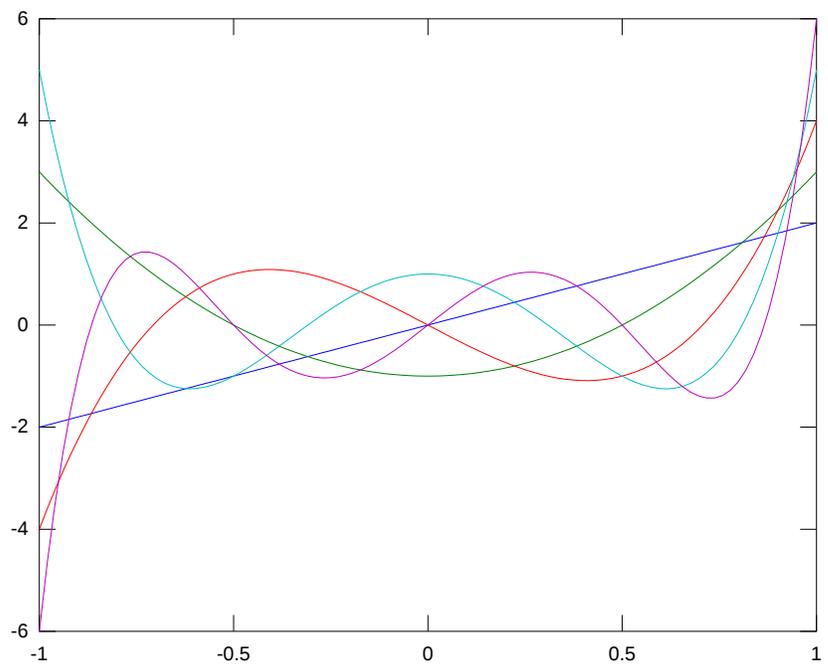


Figura 2: Polynomi di Chebyshev di seconda specie

Matrice tridiagonale associata

$$\text{diag}\left(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots\right) \begin{bmatrix} 1-x & -1 & & & \\ -1 & 3-x & -1 & & \\ & -2 & 5-x & -1 & \\ & & -3 & 7-x & -1 \\ & & & \ddots & \ddots & \ddots \end{bmatrix}$$

4.5 Polinomi di Hermite

Peso $\omega(x) = e^{-x^2}$, intervallo $[-\infty, +\infty]$. Ricorrenza a tre termini:

$$\begin{aligned} H_{n+1}(x) &= 2xH_n(x) - 2nH_{n-1}(x) \\ H_0(x) &= 1, \quad H_1(x) = 2x \end{aligned}$$

Matrice tridiagonale associata

$$\begin{bmatrix} 2x & -1 & & & \\ -2 & 2x & -1 & & \\ & -4 & 2x & -1 & \\ & & -6 & 2x & -1 \\ & & & \ddots & \ddots & \ddots \end{bmatrix}$$

5 Il problema dell'approssimazione di funzioni

Lo scopo è quello di approssimare funzioni continue $f : K \subset \mathbb{R}^n \rightarrow \mathbb{R}$, dove K è un compatto, con polinomi o con funzioni facilmente calcolabili, ad esempio, polinomi o funzioni razionali. La funzione $f(x)$ può essere assegnata esplicitamente in termini di funzioni elementari oppure essere assegnata formalmente come soluzione di una equazione differenziale.

Per semplicità lavoriamo su \mathbb{R} e scegliamo $K = [a, b]$.

La scelta di usare polinomi è saggia dal punto di vista computazionale, ma lo è anche dal punto di vista dell'approssimazione. Vale infatti

Teorema 36 (Weierstrass) *Per ogni $f(x) : K \rightarrow \mathbb{R}$ continua e per ogni $\epsilon > 0$ esiste un polinomio $p_\epsilon(x)$ tale che $\max_{x \in K} |f(x) - p_\epsilon(x)| \leq \epsilon$.*

Denotiamo \mathcal{P}_n l'insieme dei polinomi di grado al più n . Denotiamo $d(f, \mathcal{P}_n) = \inf_{p \in \mathcal{P}_n} \max_{x \in K} |f(x) - p(x)|$.

5.1 Il problema dell'approssimazione lineare

Viene fissato un “modello” di approssimazione scegliendo un insieme di funzioni continue $\{\varphi_0, \varphi_1, \dots\}$ e si cerca di approssimare nel modo migliore possibile una funzione assegnata $f(x) \in C[a, b]$ mediante $g(x) = \sum_{i=0}^n \alpha_i \varphi_i(x)$.

Le funzioni $\varphi_i(x)$ devono essere computazionalmente facili e in grado di approssimare bene la $f(x)$. Ad esempio, i polinomi non sono adatti ad approssimare funzioni che hanno asintoti verticali (in questo caso l'insieme in cui sono definite le funzioni non è un compatto). In questo caso le funzioni razionali si prestano meglio. Similmente per funzioni periodiche sono più adatte le funzioni trigonometriche.

Sia \mathcal{V} spazio vettoriale sul corpo \mathbb{K} ($\mathbb{K} = \mathbb{C}$, $\mathbb{K} = \mathbb{R}$).

Definizione 37 $\|\cdot\| : \mathcal{V} \rightarrow \mathbb{R}$ è una norma se per ogni $x, y \in \mathcal{V}$, $\alpha \in \mathbb{R}$ è

$$\begin{aligned} \|x\| &\geq 0 \\ \|x\| = 0 &\Leftrightarrow x = 0 \\ \|\alpha x\| &= |\alpha| \cdot \|x\| \\ \|x + y\| &\leq \|x\| + \|y\| \quad \text{diseguaglianza triangolare} \end{aligned}$$

Teorema 38 Ogni norma su \mathcal{V} è funzione uniformemente continua nella topologia indotta dalla norma.

Dim. Dalla disequaglianza triangolare segue che $|\|x\| - \|y\|| \leq \|x - y\|$. Quindi $\forall \epsilon > 0 \exists \delta > 0$ ($\delta = \epsilon$): $\|x - y\| \leq \delta \Rightarrow |\|x\| - \|y\|| \leq \epsilon$. \square

Esempi di norme su $C[a, b]$:

$$\begin{aligned} \|f(x)\|_2 &= \left(\int_a^b f(x)^2 dx \right)^{1/2} \\ \|f(x)\|_\infty &= \max_{x \in [a, b]} |f(x)| \end{aligned}$$

Dato un prodotto scalare $\langle \cdot, \cdot \rangle$ su $C[a, b]$, ad esempio $\langle f, g \rangle = \int_a^b f(x)g(x)\omega(x)dx$, si definisce norma indotta dal prodotto scalare

$$\|f(x)\| = \langle f, f \rangle^{1/2}.$$

Problema dell'approssimazione lineare Data $f(x) \in C[a, b]$ e l'insieme di funzioni linearmente indipendenti $\{\varphi_0, \varphi_1, \dots\}$, data una norma $\|\cdot\|$ e un intero n , calcolare α_i^* , $i = 0, 1, \dots, n$ tali che la funzione $g_n(x) = \sum_{i=0}^n \alpha_i^* \varphi_i(x)$ è tale che

$$\|f(x) - g_n(x)\| = \min_{\alpha_0, \dots, \alpha_n} \|f(x) - \sum_{i=0}^n \alpha_i \varphi_i(x)\|.$$

La funzione $g_n(x)$ è detta *funzione di migliore approssimazione*
 $r_n(x) = f(x) - g_n(x)$ è detto *resto dell'approssimazione*
 $\delta_n = \|r_n(x)\|$ è detto *errore assoluto in norma*

Osservazione 39 Se $\|\cdot\| = \|\cdot\|_\infty$ allora la funzione di migliore approssimazione $g_n(x)$ di $f(x)$ è tale che $|r_n(x)| = |f(x) - g_n(x)| \leq \delta_n \forall x \in [a, b]$. In questo caso si parla di *approssimazione uniforme*. Se $\|\cdot\| = \|\cdot\|_2$ non è vero che $|r_n(x)| \leq \delta_n$ per ogni $x \in [a, b]$. Si consideri ad esempio $r_n(x) = x^n : [0, 1] \rightarrow \mathbb{R}$. Vale $\|r_n(x)\|_\infty = 1$ mentre $\|r_n(x)\|_2 = 1/\sqrt{2n+1}$ che può essere arbitrariamente piccola. \square

Teorema 40 Sia $\mathcal{F} = \{f(x) : [a, b] \rightarrow \mathbb{R}\}$ uno spazio di funzioni dotato di norma $\|\cdot\|$, e sia $\mathcal{G}_n \subset \mathcal{F}$ un sottospazio di \mathcal{F} generato da un insieme di funzioni linearmente indipendenti $\{\varphi_0, \varphi_1, \dots, \varphi_n\}$. Allora

- il problema dell'approssimazione lineare su \mathcal{G}_n ha soluzione;
- l'insieme delle soluzioni è convesso;
- la successione $\{\delta_n\}$ è non crescente;
- esiste $\lim_n \delta_n \geq 0$.

Dim. Dimostriamo che il problema ha soluzione. L'idea della dimostrazione consiste nel verificare che l'inf del resto si può restringere all'insieme $\{\alpha = (\alpha_i) \in \mathbb{R}^{n+1} : \|\alpha\|_\infty \leq \gamma\}$ per una opportuna costante γ . Poiché questo insieme è un compatto di \mathbb{R}^{n+1} e la norma è continua esiste il minimo e quindi il problema dell'approssimazione lineare ha soluzione. Tecnicamente si introducono le funzioni $c(\alpha) = \|\sum_{i=0}^n \alpha_i \varphi_i\|$ e $d(\alpha) = \|f - \sum_{i=0}^n \alpha_i \varphi_i\|$ che sono continue. Per cui $c(\alpha)$ ha minimo su $S = \{\alpha \in \mathbb{R}^{n+1} : \|\alpha\|_\infty = 1\}$. Sia γ il minimo. Vale $\gamma \neq 0$ poiché φ_i sono linearmente indipendenti. Vale

$$c(\alpha) = \left\| \sum \frac{\alpha_i}{\|\alpha\|_\infty} \varphi_i \right\| \cdot \|\alpha\|_\infty \geq \gamma \|\alpha\|_\infty$$

inoltre

$$d(\alpha) = \|f - \sum \alpha_i \varphi_i\| \geq \left\| \sum \alpha_i \varphi_i \right\| - \|f\| \geq \gamma \|\alpha\|_\infty - \|f\|.$$

Posto $t = \inf d(\alpha)$ si scelga μ tale che $\mu \geq \frac{t+\|f\|}{\gamma}$. Risulta che se $\|\alpha\|_\infty > \mu$ allora $d(\alpha) \geq \gamma \|\alpha\|_\infty - \|f\| > \gamma \mu - \|f\| \geq t$, per cui

$$t = \inf_{\alpha} d(\alpha) = \inf_{\|\alpha\|_\infty \leq \mu} d(\alpha).$$

Ma poiché $\{\alpha \in \mathbb{R}^{n+1} : \|\alpha\|_\infty \leq \mu\}$ è compatto, l'inf è un minimo. Esiste quindi $g_n(x) \in \mathcal{G}_n$ tale che $\|f(x) - g_n(x)\| = t$ è minimo.

Per dimostrare la convessità: siano g_n e \hat{g}_n due funzioni di \mathcal{G}_n tali che $\|f - g_n\| = \|f - \hat{g}_n\|$ è il minimo. Per $\lambda \in [0, 1]$ vale

$$\begin{aligned} \|f - (\lambda g_n + (1-\lambda)\hat{g}_n)\| &= \|\lambda f - (1-\lambda)f - (\lambda g_n + (1-\lambda)\hat{g}_n)\| \\ &= \|\lambda(f - g_n) + (1-\lambda)(f - \hat{g}_n)\| \\ &\leq \lambda \|f - g_n\| + (1-\lambda) \|f - \hat{g}_n\| = \|f - g_n\| \end{aligned}$$

Quindi anche $\lambda g_n + (1 - \lambda)\hat{g}_n$ è funzione di migliore approssimazione. Per la non crescita dei δ_n basta osservare che δ_{n+1} è il minimo ottenuto su un insieme più ampio. \square

Osservazione 41 Il problema lineare di approssimazione ha una oppure infinite soluzioni. \square

Osservazione 42 Il teorema precedente non garantisce che $\delta_n \rightarrow 0$ e che quindi g_n è una “buona approssimazione”. \square

Definizione 43 \mathcal{F} è strettamente convesso rispetto alla norma $\|\cdot\|$ se per ogni $f, g \in \mathcal{F}$, $f \neq g$, vale

$$\|f\| \leq m, \|g\| \leq m \Rightarrow \|f + g\| < 2m.$$

Osservazione 44 \mathbb{R}^2 con la norma infinito non è strettamente convesso infatti $(1, 0) + (0, 1) = (1, 1)$ e i tre vettori hanno tutti norma infinito 1. Mentre \mathbb{R}^2 con la norma 2 è strettamente convesso. \square

Teorema 45 Se \mathcal{F} è strettamente convesso allora il problema dell'approssimazione lineare ha una sola soluzione

Dim. Se esistono $g_n \neq \hat{g}_n$ funzioni di migliore approssimazione allora

$$\|f - g_n\| = \|f - \hat{g}_n\| = \delta_n$$

quindi per la stretta convessità vale $\|f - g_n + f - \hat{g}_n\| < 2\delta_n$ e quindi la funzione $h(x) = \frac{1}{2}(g_n(x) + \hat{g}_n(x))$ è tale che $\|f - h\| < \delta_n$ che è assurdo. \square

Il teorema precedente non garantisce ancora che $\lim \delta_n = 0$.

5.2 Richiami sugli spazi di Banach e di Hilbert.

Definizione 46 Sia \mathcal{V} spazio vettoriale normato. Una successione $f_n \in \mathcal{V}$ è di Cauchy se $\forall \epsilon > 0 \exists n_0: \forall m, n \geq n_0 \|f_m - f_n\| \leq \epsilon$.

Esempi:

- $\mathcal{V} = C[0, 1]$, $\|f\| = \|f\|_\infty = \max_{x \in [0, 1]} |f(x)|$: $f_n(x) = x^n$ non è di Cauchy. Infatti $x^n - x^{2n}$ per $x = (1/2)^{1/n}$ vale $1/4$ qualsiasi sia n .
- $\mathcal{V} = C[0, 1]$ con la norma $\|f\|_2 = (\int_0^1 f(x)^2 dx)^{1/2}$: $f_n(x) = x^n$ è di Cauchy. Infatti $\int_0^1 (x^m - x^n)^2 dx = \dots$

Definizione 47 Uno spazio vettoriale normato \mathcal{B} è di Banach se ogni successione di Cauchy $\{f_n\}$ converge in norma ad un elemento $f \in \mathcal{B}$. Cioè se esiste $f \in \mathcal{B}$ tale che $\lim_n \|f_n - f\| = 0$.

Esempi di spazi di Banach

- $\mathbb{R}^n, \mathbb{C}^n$;
- $C[a, b]$ con $\|\cdot\|_\infty$;
- funzioni continue da un compatto $K \subset \mathbb{R}$ in \mathbb{R} con $\|\cdot\|_\infty$;
- funzioni analitiche da un aperto $\Omega \subset \mathbb{C}$ in \mathbb{C} con $\|\cdot\|_\infty$;
- $\ell^p = \{(x_i)_{i=1,2,\dots} \in \mathbb{R}^\infty : \sum_i |x_i|^p < +\infty\}$ con $\|x\|_p = (\sum_i |x_i|^p)^{1/p}$, $p \geq 1$;
- $\ell^\infty = \{(x_i)_{i=1,2,\dots} \in \mathbb{R}^\infty : \sup_i |x_i| < +\infty\}$ con $\|x\|_\infty = \sup_i |x_i|$.

Osservazione 48 Nello spazio vettoriale $\{f : [a, b] \rightarrow \mathbb{R} : \int_a^b |f|^p dx < +\infty \text{ (int. di Lebesgue)}\}$ l'applicazione $f \rightarrow \|f\|_p = (\int_a^b |f|^p dx)^{1/p}$ non è norma poiché esistono $f(x) \neq 0$: $\|f\|_p = 0$. Ma diventa una norma nello spazio quoziente con la relazione di equivalenza $f \equiv g$ se $\|f - g\|_p = 0$. Questo spazio è denotato con L^p ed è di Banach. \square

Un prodotto scalare $\langle \cdot, \cdot \rangle$ induce una norma $\|f\| = \langle f, f \rangle^{1/2}$.

Definizione 49 Uno spazio vettoriale \mathcal{H} con prodotto scalare $\langle \cdot, \cdot \rangle$ è di Hilbert se ogni successione di Cauchy converge ad un elemento di \mathcal{H} nella norma $\|\cdot\|$ indotta dal prodotto scalare.

Osservazione 50 Non tutte le norme sono indotte da un prodotto scalare. Non tutti gli spazi di Banach sono di Hilbert. \square

Osservazione 51 Lo spazio $C[a, b]$ col prodotto scalare $\langle f, g \rangle = (\int_a^b f(x)g(x)dx)^{1/2}$ non è di Hilbert. Infatti con $[a, b] = [0, 2]$, la successione

$$f_n(x) = \begin{cases} x^n & \text{se } x \in [0, 1] \\ 1 & \text{se } x \in [1, 2] \end{cases}$$

è di Cauchy, ma la successione non ha limite in $C[0, 2]$. \square

Teorema 52 Una norma sullo spazio \mathcal{V} è indotta da un prodotto scalare se e solo se vale la legge del parallelogramma

$$\|u + v\|^2 + \|u - v\|^2 = 2(\|u\|^2 + \|v\|^2)$$

e vale

$$\begin{aligned}\langle u, v \rangle &= \frac{1}{4}(\|u + v\|^2 - \|u - v\|^2) \quad \text{su } \mathbb{R} \\ \langle u, v \rangle &= \frac{1}{4}(\|u + v\|^2 - \|u - v\|^2 + i(\|u + iv\|^2 - \|u - iv\|^2)) \quad \text{su } \mathbb{C}.\end{aligned}$$

dove i è l'unità immaginaria tale che $i^2 = -1$.

Corollario 53 Uno spazio vettoriale \mathcal{V} con norma indotta da un prodotto scalare è strettamente convesso.

Dim. Dalla legge del parallelogramma si ha $\|u + v\|^2 = 2(\|u\|^2 + \|v\|^2) - \|u - v\|^2$. Per cui, se $u - v \neq 0$ segue che $\|u + v\|^2 < 2(\|u\|^2 + \|v\|^2)$. Quindi, se $\|u\| \leq m$ e $\|v\| \leq m$ e $u \neq v$, allora $\|u + v\|^2 < 2(m^2 + m^2) = 4m^2$. Ne segue $\|u + v\| < 2m$ e quindi la stretta convessità di \mathcal{V} . \square

Osservazione 54 Dalla legge del parallelogramma e dalla disuguaglianza triangolare segue la disuguaglianza di Cauchy-Schwarz, $|\langle x, y \rangle|^2 \leq \langle x, x \rangle \langle y, y \rangle$. Infatti, dalla disuguaglianza triangolare scritta nella forma $(\|u\| - \|v\|)^2 \leq \|u - v\|^2$ segue

$$\|u\|^2 + \|v\|^2 - 2\|u\| \cdot \|v\| \leq \|u - v\|^2$$

da cui, per la legge del parallelogramma, si ha

$$\frac{1}{2}(\|u + v\|^2 + \|u - v\|^2) - \|u - v\|^2 \leq 2\|u\| \cdot \|v\|$$

cioè

$$\frac{1}{2}(\|u + v\|^2 - \|u - v\|^2) \leq 2\|u\| \cdot \|v\|,$$

quindi $\langle u, v \rangle \leq \|u\| \cdot \|v\|$. \square

È utile la seguente definizione

Definizione 55 Sia \mathcal{V} spazio di Hilbert con prodotto scalare $\langle \cdot, \cdot \rangle$ e siano $\varphi_i(x) \in \mathcal{V}$, per $i = 0, 1, \dots$, tali che $\langle \varphi_i, \varphi_j \rangle = 0$ per $i \neq j$. L'insieme ortogonale $\{\varphi_i \in \mathcal{V}, i = 0, 1, 2, \dots\}$, è completo se non esistono $y \in \mathcal{V}$ tali che $\langle \varphi_i, y \rangle = 0$ per ogni i . Cioè non esiste un altro insieme ortogonale di cui l'insieme originale sia sottoinsieme proprio.

Ricordiamo il seguente risultato riguardante insiemi ortogonali completi in uno spazio di Hilbert.

Lemma 56 Sia $\{\varphi_i(x) \in \mathcal{H}, i = 0, 1, 2, \dots\}$ un insieme ortogonale completo nello spazio di Hilbert \mathcal{H} . Allora l'insieme $\mathcal{S} = \{f = \sum_{i=0}^n \alpha_i \varphi_i, n \in \mathbb{N}, \alpha_i \in K\}$ è denso in \mathcal{H} . Cioè per ogni $f \in \mathcal{H}$ e per ogni $\epsilon > 0$ esiste $g_n = \sum_{i=0}^n \alpha_i \varphi_i \in \mathcal{S}$ tale che $\|f - g_n\| \leq \epsilon$.

5.3 Funzione di migliore approssimazione

Teorema 57 Sia \mathcal{V} spazio vettoriale dotato di prodotto scalare e sia $f \in \mathcal{V}$. Siano inoltre $\varphi_0, \dots, \varphi_n \in \mathcal{V}$ linearmente indipendenti. Allora $\exists!$ $g = \sum_{i=0}^n \alpha_i \varphi_i$ tale che $\|g - f\|$ è minima. Inoltre $\|g - f\|$ è minima se e solo se $\langle g - f, v \rangle = 0$ per ogni $v \in \mathcal{W} := \text{span}(\varphi_0, \dots, \varphi_n)$.

Dim. L'esistenza e unicità di g segue dal fatto che \mathcal{V} è strettamente convesso alla luce del teorema 45 e del corollario 53.

Sia $h = f - g$. Dimostriamo ora che $\|h\|$ è minima se e solo se il vettore h è ortogonale a $\mathcal{W} = \text{span}(\varphi_0, \dots, \varphi_n)$, cioè $\langle h, v \rangle = 0$ per ogni $v \in \mathcal{W}$. Se $\|h\|$ è minimo, allora $\|h\|^2 \leq \|h + \gamma v\|^2$ per ogni γ e $v \in \mathcal{W}$, da cui

$$\|h\|^2 \leq \|h + \gamma v\|^2 = \langle h + \gamma v, h + \gamma v \rangle = \|h\|^2 + 2\gamma \langle h, v \rangle + \gamma^2 \|v\|^2$$

cioè $2\gamma \langle h, v \rangle + \gamma^2 \|v\|^2 \geq 0$ per ogni γ . Scelgo $\gamma = -\langle h, v \rangle / \langle v, v \rangle$ ed ho $-\langle h, v \rangle^2 / \langle v, v \rangle \geq 0$, cioè $\langle h, v \rangle = 0$.

Viceversa, se $\langle h, v \rangle = 0$ per ogni $v \in \mathcal{W}$ allora $\|h + \gamma v\|^2 = \langle h + \gamma v, h + \gamma v \rangle = \langle h, h \rangle + \gamma^2 \langle v, v \rangle \geq \|h\|^2$. \square

Il seguente risultato dà una espressione esplicita della soluzione di miglior approssimazione

Teorema 58 Nell'ipotesi del teorema precedente la soluzione di miglior approssimazione g è : $g = \sum_{i=0}^n \alpha_i \varphi_i$ dove

$$A \begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} b_0 \\ \vdots \\ b_n \end{bmatrix}$$

con $A = (a_{i,j})$, $a_{i,j} = \langle \varphi_i, \varphi_j \rangle$, $b_i = \langle f, \varphi_i \rangle$.

Dim. $f - g$ è ortogonale a $\mathcal{W} = \text{span}(\varphi_0, \dots, \varphi_n)$, cioè $\langle f - g, \varphi_j \rangle = 0$ per $j = 0, 1, \dots, n$ e questa è la j -esima equazione. \square

Corollario 59 Se $\varphi_0, \dots, \varphi_n$ sono ortogonali allora $\alpha_i = \langle f, \varphi_i \rangle / \langle \varphi_i, \varphi_i \rangle$. Quindi

$$g_n = \sum_{i=0}^n \frac{\langle f, \varphi_i \rangle}{\langle \varphi_i, \varphi_i \rangle} \varphi_i$$

I coefficienti $\frac{\langle f, \varphi_i \rangle}{\langle \varphi_i, \varphi_i \rangle}$ sono chiamati i *coefficienti di Fourier* di $f(x)$.

Teorema 60 Sia $\{\varphi_i \in \mathcal{H}, i = 0, 1, 2, \dots\}$ un insieme ortogonale nello spazio di Hilbert \mathcal{H} . Sia

$$g_n = \sum_{i=0}^n \frac{\langle f, \varphi_i \rangle}{\langle \varphi_i, \varphi_i \rangle} \varphi_i$$

la funzione di miglior approssimazione di f , $r_n = f - g_n$ il resto e $\delta_n = \|r_n\|$ l'errore di approssimazione. Allora vale

$$\delta_n^2 = \|f\|^2 - \sum_{i=0}^n \frac{\langle f, \varphi_i \rangle^2}{\langle \varphi_i, \varphi_i \rangle} \quad (8)$$

per cui la successione δ_n è non crescente. inoltre la serie $\sum_{i=0}^{\infty} \frac{\langle f, \varphi_i \rangle^2}{\langle \varphi_i, \varphi_i \rangle}$ è convergente e

$$\|f\|^2 \geq \sum_{i=0}^{\infty} \frac{\langle f, \varphi_i \rangle^2}{\langle \varphi_i, \varphi_i \rangle} \quad \text{diseguaglianza di Bessel} \quad (9)$$

quindi la serie in (9) è convergente. Se inoltre l'insieme $\{\varphi_i \in \mathcal{H}, i = 0, 1, 2, \dots\}$ è anche completo vale

$$\|f\|^2 = \sum_{i=0}^n \frac{\langle f, \varphi_i \rangle^2}{\langle \varphi_i, \varphi_i \rangle} + \delta_n^2 = \sum_{i=0}^{+\infty} \frac{\langle f, \varphi_i \rangle^2}{\langle \varphi_i, \varphi_i \rangle} \quad \text{uguaglianza di Parseval} \quad (10)$$

per cui la successione δ_n converge a zero in modo monotono.

Dim. Se g_n è di migliore approssimazione per f rispetto a $\varphi_0, \varphi_1, \dots, \varphi_n$, allora $\delta_n^2 = \|f - g_n\|^2 = \langle f - g_n, f - g_n \rangle = \langle f - g_n, f \rangle$ essendo g_n ortogonale a $h = f - g_n$. Risulta allora che

$$\begin{aligned} \delta_n^2 &= \|f\|^2 - \langle f, g_n \rangle \\ &= \|f\|^2 - \left\langle f, \sum_{i=0}^n \frac{\langle f, \varphi_i \rangle}{\langle \varphi_i, \varphi_i \rangle} \varphi_i \right\rangle \\ &= \|f\|^2 - \sum_{i=0}^n \frac{\langle f, \varphi_i \rangle^2}{\langle \varphi_i, \varphi_i \rangle} \geq 0 \end{aligned}$$

cioè la (8). In particolare, per ogni n vale

$$\|f\|^2 \geq \sum_{i=0}^n \frac{\langle f, \varphi_i \rangle^2}{\langle \varphi_i, \varphi_i \rangle}.$$

Quindi la serie $\sum_{i=0}^{+\infty} \frac{\langle f, \varphi_i \rangle^2}{\langle \varphi_i, \varphi_i \rangle}$ converge e vale

$$\|f\|^2 \geq \sum_{i=0}^{+\infty} \frac{\langle f, \varphi_i \rangle^2}{\langle \varphi_i, \varphi_i \rangle}.$$

Dimostriamo ora che se $\{\varphi_0, \varphi_1, \dots\}$ è completo allora $\|f\|^2 \leq \sum_{i=0}^{+\infty} \frac{\langle f, \varphi_i \rangle^2}{\langle \varphi_i, \varphi_i \rangle}$. Sia $\mathcal{S} = \{\sum_{i=0}^n \alpha_i \varphi_i : n \in \mathbb{N}, \alpha_i \in \mathbb{R}\}$. Per il lemma 56 vale $\bar{\mathcal{S}} = \mathcal{H}$. Quindi per ogni $\epsilon > 0$ esiste $\psi \in \mathcal{S}$ tale che $\|f - \psi\| \leq \epsilon$, $\psi = \sum_{i=0}^n \gamma_i \varphi_i$. Per il teorema 57

esiste ed è unica la soluzione di migliore approssimazione $g = \sum_{i=0}^n \alpha_i^* \varphi_i$ tale che $\|f - g\| \leq \|f - \psi\| \leq \epsilon$. Quindi

$$\epsilon \geq \|f - g\| \geq |\|f\| - \|g\|| \geq \|f\| - \|g\|$$

da cui

$$(\|f\| - \epsilon)^2 \leq \|g\|^2 = \langle g, g \rangle = \sum_{i=0}^n \frac{\langle f, \varphi_i \rangle^2}{\langle \varphi_i, \varphi_i \rangle} \leq \sum_{i=0}^{+\infty} \frac{\langle f, \varphi_i \rangle^2}{\langle \varphi_i, \varphi_i \rangle}$$

per ogni $\epsilon > 0$. Passando al limite per $\epsilon \rightarrow 0$ si ha

$$\|f\|^2 \leq \sum_{i=0}^{+\infty} \frac{\langle f, \varphi_i \rangle^2}{\langle \varphi_i, \varphi_i \rangle}$$

che completa la dimostrazione. \square

Esempio: il prodotto scalare $\langle f, g \rangle = \int_a^b f(x)g(x)\omega(x)dx$ definisce lo spazio di Hilbert delle funzioni da $[a, b]$ in \mathbb{R} per cui è finito l'integrale di Lebesgue $\int_a^b f(x)^2\omega(x)dx$, dove $\omega(x) : (a, b) \rightarrow \mathbb{R}$ è una funzione a valori positivi.

Esempio: $\varphi_i(x) = x^i$, $i = 0, 1, 2, \dots$, non è un sistema ortonormale in $L^2[0, 1]$. Vale $\langle \varphi_i, \varphi_j \rangle = \int_0^1 x^{i+j} dx = 1/(i+j+1)$. La matrice $A = (a_{i,j})_{i,j=0:n}$, $a_{i,j} = 1/(i+j+1)$ è detta matrice di Hilbert. Il suo numero di condizionamento cresce esponenzialmente con n .

Osservazione 61 La scelta $\varphi_i(x) = x^i$ oltre a non fornire un sistema ortonormale di funzioni, è particolarmente infelice poiché la matrice del sistema che fornisce la funzione di migliore approssimazione è fortemente mal condizionata per cui dal punto di vista numerico il problema della sua risoluzione diventa intrattabile già per valori moderati di n . Un'altro inconveniente è che nel passare da n a $n+1$ i calcoli svolti per risolvere il sistema $n \times n$ non possono essere utilizzati per facilitare la risoluzione del sistema $(n+1) \times (n+1)$. \square

Osservazione 62 Si osservi che se $\{\varphi_i\}$ è sistema ortogonale completo, dall'espressione

$$\delta_n^2 = \|f\|_2^2 - \sum_{i=0}^n \frac{\langle \varphi_i, f \rangle^2}{\langle \varphi_i, \varphi_i \rangle}$$

segue che il valore di δ_n è calcolabile esplicitamente. Per cui, data una tolleranza ϵ , il calcolo dei coefficienti di Fourier può essere condotto per $n = 1, 2, 3, \dots$, fintanto che $\delta_n < \epsilon$. Inoltre il passare da n a $n+1$ non comporta il dover risolvere un nuovo sistema lineare come nel caso di un sistema non ortogonale. \square

Osservazione 63 Purtroppo, con un sistema ortogonale completo dato da un prodotto scalare, la relazione $\|\delta_n\| \leq \epsilon$ non implica che $|f(x) - g_n(x)| \leq \epsilon$ per ogni x . Per avere verificata quest'ultima condizione bisogna adottare la norma infinito la quale non è indotta da nessun prodotto scalare.

In particolare $\|h\|_2 \leq \epsilon$ non implica $\|h\|_\infty \leq \epsilon$. Per questo basta scegliere $L^2([0, 1])$ con $h(x) = x^n$ per cui $\|h\|_2^2 = 1/(2n + 1)$, $\|h\|_\infty = 1$. \square

In generale la convergenza in L^2 non implica la convergenza in L^∞ . Però sotto particolari condizioni sì. Vale infatti il seguente

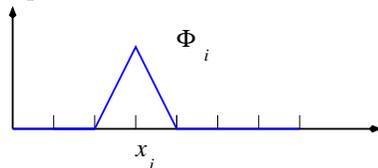
Teorema 64 Per $n \geq 1$ sia $g_n(x)$ la funzione di migliore approssimazione di $f(x) \in C[-1, 1]$ ottenuta scegliendo come φ_i i polinomi di Chebyshev di prima specie e col prodotto scalare ad essi associato. Se $f(x)$ è Lipschitziana allora $g_n(x)$ converge uniformemente a $f(x)$, i.e., $\lim_n \|f(x) - g_n(x)\|_\infty = 0$. Se $f \in C^k[-1, 1]$, $k \geq 1$, allora $\exists \gamma \in \mathbb{R}$, $\gamma > 0$: $\|f - g_n\|_\infty \leq \gamma \frac{\log n}{n^k}$ e vale

$$f(x) = \frac{\alpha_0}{2} + \sum_{i=1}^{+\infty} \alpha_i T_i(x).$$

La serie del teorema precedente è detta Serie di Chebyshev.

Osservazione 65 Dal punto di vista computazionale talvolta è conveniente scegliere delle funzioni $\varphi_i(x)$ che pur non essendo ortogonali verifichino la seguente proprietà: siano a supporto compatto e siano tali che $\text{supp}(\varphi_i) \cap \text{supp}(\varphi_j) = \emptyset$ se $|i - j| \geq k$ per un intero $k > 0$ ragionevolmente piccolo. In questo modo si verifica facilmente che la matrice A del sistema lineare che fornisce i coefficienti della migliore approssimazione è una matrice a banda con $2k + 1$ diagonali. Sistemi lineari con matrici a banda si risolvono in modo poco costoso con le metodologie dell'algebra lineare numerica.

Una possibile scelta a questo riguardo è quella di suddividere l'intervallo $[a, b]$ in intervalli equispaziati $x_i = a + ih$, $h = 1/(b - a)$, $i = 0, 1, \dots, n$, fissare una funzione $\varphi(x)$ che abbia supporto $[-h, h]$ e sia nulla al di fuori di esso e porre $\varphi_i(x) = \varphi(x - x_i)$, $i = 0, \dots, n$. Funzioni usualmente utilizzate nel contesto delle equazioni differenziali sono le *hat functions*



ottenute con $\varphi(x) = \max(0, h - |x|)$. Per l'insieme delle funzioni hat col prodotto scalare $\langle f, g \rangle = \int_a^b f(x)g(x)dx$ la matrice A è tale che $A = (h^3/6)\text{trid}(1, 4, 1)$, cioè è tridiagonale con elementi diagonali uguali a $4(h^3/6)$ e elementi sopra e sottodiagonali uguali a $h^3/6$. La matrice è fortemente dominante diagonale e ha numero di condizionamento indipendente dalla dimensione. Questo dà alla matrice proprietà computazionali rilevanti. \square

6 Alcune considerazioni sull'interpolazione

Finora abbiamo studiato metodi e proprietà di approssimazione di funzioni mediante polinomi. Su spazi di Hilbert con norme derivanti da un prodotto scalare abbiamo caratterizzato la migliore soluzione con resto di minima norma e descritto come calcolarla.

Abbiamo introdotto polinomi ortogonali che ben si prestano all'uso. Per spazi di Banach, ad esempio $C[a, b]$ con la norma infinito la situazione è un po' più complicata. Ci sono comunque metodi per calcolare il polinomio di migliore approssimazione in norma infinito.

Abbiamo poi visto come usando polinomi di Chebyshev, e quindi il prodotto scalare associato, la serie di Fourier, ribattezzata serie di Chebyshev, converge uniformemente a $f(x)$ se f è lipschitziana o di classe $C^k[a, b]$. Per cui la serie di Chebyshev dà un buon approssimante in norma infinito anche se non ottimo.

Ci chiediamo ora cosa si può dire delle approssimazioni ottenute con polinomi di interpolazione.

Definizione 66 Dato un intervallo $[a, b]$ definiamo tavola dei nodi su $[a, b]$ l'insieme $\{x_i^{(n)} \in [a, b] : i = 0, \dots, n, n \in \mathbb{N}, x_i^{(n)} \neq x_j^{(n)}, \text{ se } i \neq j\}$.

Dati un intervallo $[a, b]$ e una tavola di nodi $x_i^{(n)} \in [a, b], i = 0, \dots, n$, e una $f \in C[a, b]$, definiamo $p_n(x) = \sum_{i=0}^n f(x_i^{(n)})L_{i,n}(x)$, dove $L_{i,n}(x)$ sono i polinomi di Lagrange definiti da

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

Poiché $L_{i,n}(x_j^{(n)}) = \delta_{i,j}$, ne segue che $p_n(x_i^{(n)}) = f(x_i^{(n)})$, cioè il polinomio $p_n(x)$ è il polinomio di interpolazione di $f(x)$ relativo ai nodi $x_i^{(n)}, i = 0, 1, \dots, n$.

Definisco l'operatore $\mathcal{A}_n : C[a, b] \rightarrow \mathcal{P}_n$ tale che $\mathcal{A}_n(f) = p_n$. Voglio studiare il condizionamento di \mathcal{A}_n

6.1 Costanti di Lebesgue e condizionamento dell'interpolazione

Denotiamo $f_i = f(x_i^{(n)})$ (tralasciando per semplicità la dipendenza da n). In questo modo il polinomio di interpolazione si scrive come $p_n(x) = \sum_{i=0}^n f_i L_{i,n}(x)$. Se perturbo i valori f_i in \tilde{f}_i ho il polinomio $\tilde{p}_n(x) = \sum_{i=0}^n \tilde{f}_i L_{i,n}(x)$. Per cui

$$|p_n(x) - \tilde{p}_n(x)| \leq \sum_{i=0}^n |f_i - \tilde{f}_i| \cdot |L_{i,n}(x)| \leq \max_i |f_i - \tilde{f}_i| \sum_{i=0}^n |L_{i,n}(x)|.$$

Pongo $\Lambda_n = \max_{x \in [a, b]} \sum_{i=0}^n |L_{i,n}(x)|$ ed ho

$$|p_n(x) - \tilde{p}_n(x)| \leq \Lambda_n \max_i |f_i - \tilde{f}_i|, \quad \forall x \in [a, b],$$

da cui

$$\|p_n - \tilde{p}_n\|_\infty \leq \Lambda_n \|f - \tilde{f}\|_\infty. \quad (11)$$

Cioè la quantità Λ_n esprime la massima amplificazione che può avere $\|p_n - \tilde{p}_n\|_\infty$ sotto perturbazioni dei valori di f . Quindi esprime una maggiorazione del condizionamento numerico dell'operatore \mathcal{A}_n che a $f(x)$ associa il polinomio di interpolazione in relazione alla tavola di nodi scelta.

La costante Λ_n è chiamata *costante di Lebesgue* e dipende unicamente dalla tavola di nodi di interpolazione. Ogni tavola di nodi di interpolazione ha quindi la sua corrispondente costante di Lebesgue.

Da (11) applicata con $g = f - \tilde{f}$, segue $\Lambda_n \geq \frac{\max_x |\mathcal{A}_n(g(x))|}{\max_x |g(x)|}$. Vale la seguente proprietà:

$$\Lambda_n = \max_{g \in C[a,b], \|g\|_\infty=1} \|\mathcal{A}_n(g)\|_\infty =: \|\mathcal{A}_n\|_\infty$$

per dimostrarlo basta osservare che

$$\begin{aligned} \sup_{g \in C[a,b]} \frac{\|\mathcal{A}_n(g)\|_\infty}{\|g\|_\infty} &= \sup_{g \in C[a,b]} \frac{\|\sum g(x_i^{(n)}) L_{i,n}(x)\|_\infty}{\|g\|_\infty} \\ &\leq \frac{\max_i |g(x_i^{(n)})| \max_{x \in [a,b]} \sum_i |L_{i,n}(x)|}{\|g\|_\infty} \\ &\leq \max_{x \in [a,b]} \sum_i |L_{i,n}(x)|. \end{aligned}$$

Inoltre, se si sceglie g in modo che $\|g\|_\infty = 1$ e $g(x_i^{(n)}) = \pm 1$ a seconda del segno di $L_{i,n}(\xi)$, dove ξ è il punto in cui $\sum_i |L_{i,n}(x)|$ prende il valore massimo, si ottiene l'uguaglianza.

Teorema 67 *Sia $f \in C[a, b]$ e $p_n(x)$ il polinomio di interpolazione relativo alla tavola di nodi $x_i^{(n)}$. Sia inoltre $q_n(x)$ il polinomio di migliore approssimazione uniforme. Allora vale*

$$\|f - p_n\|_\infty \leq (1 + \Lambda_n) \|f - q_n\|_\infty.$$

Dim. Siano p_n il polinomio di interpolazione e q_n il polinomio di miglior approssimazione uniforme. Poiché $f - p_n = f - q_n + q_n - p_n$ si ha

$$\|f - p_n\|_\infty \leq \|f - q_n\|_\infty + \|q_n - p_n\|_\infty = \|f - q_n\|_\infty + \|\mathcal{A}_n(q_n - f)\|_\infty \leq (1 + \Lambda_n) \|q_n - f\|_\infty$$

che completa la dimostrazione. \square

Osservazione 68 Se Λ_n è “piccola” allora l'errore dell'interpolazione è poco più grande di quello della migliore approssimazione uniforme. È quindi importante usare delle tavole di nodi che comportino costanti di Lebesgue Λ_n il più piccole possibile. \square

Si riportano i valori di Λ_n per alcune tavole di nodi su $[-1, 1]$

- Punti equispaziati: $\Lambda_n \approx 2^{n+1}/(en \log n)$
- Punti di Chebyshev $x_i^{(n)} = \cos(\frac{(2i+1)\pi}{2(n+1)})$, $i = 0, \dots, n$:

$$\frac{2}{\pi} \log(n+1) + 0.96 \leq \Lambda_n \leq \frac{2}{\pi} \log(n+1) + 1.$$

I nodi $x_i^{(n)}$ sono gli zeri dei polinomi di Chebyshev di prima specie.

Si osservi la enorme differenza nei due casi che sottolinea l'importanza dei nodi di Chebyshev. Per il teorema 67 l'interpolazione sui nodi di Chebyshev, anche se non fornisce la funzione di migliore approssimazione uniforme, dà un'approssimazione che può essere ragionevole data la crescita logaritmica della costante di Lebesgue associata Λ_n .

È noto che $\Lambda_n \geq \frac{2}{\pi} \log(n+1) + 0.818$ per ogni scelta dei nodi ma non è nota la tavola dei nodi che dà il valore minimo di Λ_n . Maggiori informazioni sulle costanti di Lebesgue si trovano in [3].

Un risultato negativo relativo all'interpolazione è il seguente teorema di Faber:

Teorema 69 (Faber) *Non esiste nessuna tavola di nodi su $[a, b]$ tale che per ogni $f \in C[a, b]$ il polinomio di interpolazione relativo a questa tavola di nodi converga uniformemente a f .*

Per dimostrare questo risultato bisogna richiamare il teorema di Banach-Steinhaus

Teorema 70 (Banach-Steinhaus) *X spazio di Banach, Y spazio normato, $\mathcal{F} = \{\mathcal{A} : X \rightarrow Y, \mathcal{A} \text{ lineare e continuo}\}$, $\|\mathcal{A}\| = \sup \frac{\|\mathcal{A}(x)\|}{\|x\|}$. Se per ogni $x \in X$ è $\sup_{\mathcal{A} \in \mathcal{F}} \|\mathcal{A}(x)\| < +\infty$ allora $\sup_{\mathcal{A} \in \mathcal{F}} \|\mathcal{A}\| < +\infty$.*

La dimostrazione del teorema di Faber procede allora nel seguente modo: se per assurdo esiste una tavola dei nodi tale che $\mathcal{A}_n(f)$ converge uniformemente a f per ogni $f \in C[a, b]$, allora $\forall f \in C[a, b]$ è $\lim_n \|\mathcal{A}_n(f) - f\|_\infty = 0$, cioè $\|\mathcal{A}_n(f) - f\|_\infty$ è limitata per ogni n . Si applica allora il teorema di Banach-Steinhaus a $\{\mathcal{A}_n - 1\}$ e si ha che $\|\mathcal{A}_n - 1\|$ è limitata per ogni n ma $\|\mathcal{A}_n - 1\| \geq \|\mathcal{A}_n\| - 1 \geq \Lambda_n - 1 \geq \frac{2}{\pi} \log n - 1$, che è assurdo. \square

7 Osservazioni su “continuo e discreto”

Osservazione 71 Nel caso discreto il problema di approssimare un vettore $\mathbf{f} \in \mathbb{R}^n$ con un vettore $\mathbf{g} \in \mathcal{S} \subset \mathbb{R}^n$ dove \mathcal{S} è il sottospazio generato dalle colonne della matrice A di dimensioni $n \times m$, $m < n$ è

$$\min_{\mathbf{x} \in \mathcal{S}} \|\mathbf{A}\mathbf{x} - \mathbf{f}\|$$

dove $\|\mathbf{x}\| = \langle \mathbf{x}, \mathbf{x} \rangle^{1/2}$. Se A ha rango massimo, cioè le sue colonne formano una base di \mathcal{S} un approccio risolutivo è quello di usare le *equazioni normali*

$$\langle A\mathbf{e}_i, A\mathbf{c} - \mathbf{f} \rangle = 0$$

che nel prodotto euclideo diventano

$$A^T(A\mathbf{x} - \mathbf{f}) = 0.$$

Esse impongono l'ortogonalità del residuo rispetto alle colonne di A , cioè l'ortogonalità del residuo rispetto allo spazio \mathcal{S} esattamente come accadeva nel caso continuo.

Un altro approccio numericamente più stabile nel discreto consiste nel calcolare la fattorizzazione QR della matrice A : $A = QR$, dove Q ha colonne ortogonali e R è triangolare superiore, e poi risolvere

$$\min \|A\mathbf{x} - \mathbf{f}\| = \min \|QR\mathbf{x} - \mathbf{f}\| = \min \|Q(R\mathbf{x} - Q^T\mathbf{f})\| = \min \|R\mathbf{x} - Q^T\mathbf{f}\|$$

che fornisce $y_i = (Q^T\mathbf{f})_i$, $R\mathbf{x} = \mathbf{y}$.

Nel continuo, questo corrisponde a costruirsi prima una base ortogonale di \mathcal{S} e a calcolarsi i coefficienti di Fourier. \square

Osservazione 72 In Octave il grafico della funzione $f(x)$ su $[a, b]$ può essere tracciato calcolando i vettori $\mathbf{f} = (f_i)$, $f_i = f(x_i)$ per $x_i = a + hi$, $i = 0, \dots, n$ con $h = (b - a)/n$ e tracciando la spezzata che unisce i punti (x_i, f_i) con (x_{i+1}, f_{i+1}) .

Il prodotto scalare tra $f(x)$ e $g(x)$ può essere stimato approssimando l'integrale $\int_0^1 f(x)g(x)dx$ con $h \sum_i f_i g_i$, cioè $h \langle \mathbf{f}, \mathbf{g} \rangle = h \mathbf{f}^T \mathbf{g}$.

La norma di $f(x)$ viene quindi approssimata da $(\frac{b-a}{n} \sum_i f_i^2)^{1/2}$. Quindi la discretizzazione del problema continuo conduce ad un problema discreto come descritto nell'osservazione precedente \square

Osservazione 73 Scegliendo la base dei monomi $\varphi_i(x) = x^i$ e campionando ciascun polinomio $\varphi_i(x)$ nei nodi x_i scelti si ottiene la matrice di Vandermonde. Nel discreto la matrice di Hilbert $\langle x^i, x^j \rangle$ è approssimata dal prodotto tra la matrice di Vandermonde e la sua trasposta. \square

Come potrebbe essere definita nel continuo l'analogo di una matrice di Householder?

Osservazione 74 Se $A\mathbf{x} = \mathbf{b}$ è sistema lineare con A matrice definita positiva, allora il funzionale $\varphi(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T A\mathbf{x} - \mathbf{b}^T \mathbf{x}$ ha gradiente $\nabla\varphi(\mathbf{x}) = A\mathbf{x} - \mathbf{b}$ che si annulla se e solo se \mathbf{x} è soluzione del sistema lineare $A\mathbf{x} = \mathbf{b}$. Inoltre l'hessiano di $\varphi(\mathbf{x})$ coincide con la matrice A che è definita positiva. Quindi il funzionale φ è convesso e il suo unico punto di minimo è la soluzione del sistema lineare $A\mathbf{x} = \mathbf{b}$.

Questo fatto permette di interpretare la soluzione di un sistema lineare definito positivo come punto di minimo di un funzionale e quindi permette di costruire dei metodi di risoluzione iterativi così detti del gradiente, quali i metodi della discesa più ripida e del gradiente coniugato. Tali metodi generano una successione di punti x_k in cui il funzionale assume valori decrescenti e tale che $\lim_k \varphi(x_k) = 0$. Proprietà analoghe valgono per operatori differenziali lineari “definiti positivi”. Vedremo ciò nella seconda parte.

La minimizzazione del funzionale $\varphi(x)$ può essere vista come un problema di minima norma. Infatti, introduciamo il prodotto scalare $\langle x, y \rangle = x^T A y$ (è prodotto scalare data la definita positività di A) e la norma conseguente $\|x\|_A := \langle x, x \rangle^{1/2}$. Questa norma è chiamata *norma in energia*. Denotata con u la soluzione del sistema $Ax = b$ risulta allora

$$\begin{aligned} \|x - u\|_A^2 &= \langle x - u, x - u \rangle = x^T A x - 2x^T A u + u^T A u \\ &= x^T A x - 2x^T b + u^T b = 2\varphi(x) + u^T b \end{aligned}$$

Cioè minimizzare il funzionale $\varphi(x)$ su un sottospazio \mathcal{V} di \mathbb{R}^n equivale a trovare il vettore di \mathcal{V} che è più vicino alla soluzione del sistema $Ax = b$ in norma in energia. \square

8 Il teorema di Korovkin

In questa sezione¹ si presenta il Teorema di Korovkin la cui estrema potenza risiede nella semplicità delle ipotesi, nella loro facile verificabilità e nella notevole forza della tesi. In particolare, data una successione di operatori da $C(K)$ in sé, K compatto di \mathbb{R}^n , basta verificare che tali operatori siano lineari e positivi e che siano convergenti per un numero finito di semplici funzioni continue (alcuni polinomi di grado al più due) per concludere che la data successione di operatori è in grado di approssimare ogni funzione continua. Non ultimo pregio del Teorema di Korovkin è la estrema essenzialità della sua dimostrazione che si basa su conoscenze elementari di analisi matematica.

Per meglio definire il quadro in cui ci muoviamo, si riporta la definizione di operatore lineare e positivo (LPO).

Un operatore L è detto *operatore lineare positivo* (LPO) se:

1. È lineare, cioè $L[\lambda f] = \lambda L[f]$ per ogni $\lambda \in \mathbb{R}$ (o \mathbb{C}), e $L[f + g] = L[f] + L[g]$
2. È *positivo*, cioè $L[f](x) \geq 0$ per ogni funzione f tale che $f(x) \geq 0$ per ogni x . Cioè L manda funzioni non negative su tutto il dominio in funzioni non negative su tutto il dominio.

Notate che le due ipotesi implicano che se $f(x) \leq g(x)$ per ogni x nel dominio, allora

$$L[f](x) \leq L[g](x) \quad \text{per ogni } x \text{ nel dominio} \quad (12)$$

¹scritta da Federico Poloni

e che

$$|L[f]| \leq L[|f|]. \quad (13)$$

Sia L_n una successione di operatori. Diciamo che L_n *approssima bene* una funzione f (in una certa norma $\|\cdot\|$) se $\|L_n[f] - f\| \rightarrow 0$ per $n \rightarrow \infty$.

Teorema 75 (Teorema di Korovkin, versione 1-dimensionale) *Sia $K \subseteq \mathbb{R}$ un compatto, e sia $C(K)$ lo spazio delle funzioni continue da K a \mathbb{R} con la norma del sup (convergenza uniforme). Sia $(L_n)_{n=1}^{\infty}$ una successione di LPO su $C(K)$. Se L_n approssima bene le tre funzioni:*

- $x \mapsto 1$ (la funzione costante uguale a 1),
- $x \mapsto x$ (l'identità),
- $x \mapsto x^2$ (la funzione "elevare al quadrato"),

allora L_n approssima bene tutte le funzioni di $C(K)$.

Dim. Siano $f \in C(K)$, e $\varepsilon > 0$ fissati; dobbiamo mostrare che esiste \tilde{n} tale che per ogni $n \geq \tilde{n}$ valga

$$|L_n[f](y) - f(y)| \leq \varepsilon \quad \forall y \in K.$$

Notiamo anche che al posto di ε al membro di destra possiamo mettere anche un'espressione del tipo $\gamma\varepsilon$, a patto che $\gamma > 0$ non dipenda né da n né da a .

Chiamiamo inoltre $M := \|f\|_{\infty}$, $N := \max_{x \in K} |x|$, e δ è tale che $|x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$ (che esiste perché una funzione continua su un compatto è uniformemente continua).

Innanzitutto sfruttiamo l'ipotesi: definiamo per ogni n, y

$$\begin{aligned} \epsilon_0^{(n)}(y) &:= L_n[x \mapsto 1](y) - 1, \\ \epsilon_1^{(n)}(y) &:= L_n[x \mapsto x](y) - y, \\ \epsilon_2^{(n)}(y) &:= L_n[x \mapsto x^2](y) - y^2. \end{aligned}$$

Per l'ipotesi, possiamo prendere \tilde{n} sufficientemente grande, in modo che per $n \geq \tilde{n}$ si abbia $\|\epsilon_i^{(n)}\|_{\infty} < \min(\varepsilon, \varepsilon\delta^2)$ per $i = 0, 1, 2$.

Ora, fissiamo $y \in K$ e per ogni $n \geq \tilde{n}$ scriviamo la stima

$$\begin{aligned} |L_n[f](y) - f(y)| &\leq |L_n[f](y) - f(y)L_n[1](y)| + |f(y)L_n[1](y) - f(y)| \\ &= |L_n[f(x) - f(y)](y)| + \left| \epsilon_0^{(n)}(y)f(y) \right| \\ &\leq L_n[|f(x) - f(y)|](y) + M\varepsilon \end{aligned}$$

(abbiamo usato la (13) nell'ultimo passaggio). Lavoriamo separatamente ora sul primo dei due addendi; si ha

$$\begin{aligned} |f(x) - f(y)| &\leq \begin{cases} \varepsilon & \text{per } |x - y| \leq \delta \\ 2M & \text{per } |x - y| > \delta \end{cases} \\ &\leq \begin{cases} \varepsilon & \text{per } |x - y| \leq \delta \\ \frac{2M}{\delta^2}(x - y)^2 & \text{per } |x - a| > \delta \end{cases} \\ &\leq \varepsilon + \frac{2M}{\delta^2}(x - y)^2 \end{aligned}$$

quindi per la (12) e la (13), abbiamo

$$\begin{aligned} L_n[|f(x) - f(y)|](y) &\leq L_n[\varepsilon + \frac{2M}{\delta^2}(x - y)^2] \\ &= \varepsilon L_n[1](y) + \frac{2M}{\delta^2} (y^2 L_n[1](y) + L_n[x^2](y) - 2y L_n[x](y)) \\ &= \varepsilon(1 + \epsilon_0^{(n)}(y)) + \frac{2M}{\delta^2} (y^2(1 + \epsilon_0^{(n)}(y)) + (y^2 + \epsilon_1^{(n)}(y)) - 2y(y + \epsilon_2^{(n)}(y))) \\ &= \varepsilon(1 + \epsilon_0^{(n)}(y)) + \frac{2M}{\delta^2} (y^2 \epsilon_0^{(n)}(y) + \epsilon_1^{(n)}(y) - 2y \epsilon_2^{(n)}(y)) \end{aligned}$$

Ora, poiché abbiamo scelto \tilde{n} in modo che per $n \geq \tilde{n}$ le quantità $\epsilon_i^{(n)}(y)$ siano in modulo minori di $\delta^2 \varepsilon$ si ha

$$L_n[|f(x) - f(y)|](y) \leq \gamma \varepsilon$$

con $\gamma > 0$ che dipende solo dalle scelte di K e f . Quindi il teorema è dimostrato. \square

Notate che nelle stime compaiono quantità che dipendono da f : quindi la successione di LPO approssima bene tutte le funzioni, ma su alcune funzioni l'approssimazione converge più lentamente che su altre.

Cosa si può dire sulla velocità di convergenza? Se $\|\epsilon_i^{(n)}\|_\infty = O(h(n))$ per $i = 0, 1, 2$, allora $\|f - L_n[f]\|_\infty = O(h(n))$ per ogni f (però occhio che la costante nascosta nella notazione $O(\cdot)$ dipende da f). Quindi se per esempio una successione di LPO converge linearmente, quadraticamente, ... sulle tre funzioni di test, allora converge con lo stesso ordine per tutte le funzioni. Osservate che nella dimostrazione abbiamo richiesto che per $n > \tilde{n}$ valga $|\epsilon_i^{(n)}| \leq \varepsilon/\delta^2$ dove δ è tale che $|x - y| \leq \delta \Rightarrow |f(x) - f(y)| \leq \varepsilon$. La "piccolezza" di δ rispetto a ε ha un ruolo negativo sulla convergenza. Più piccolo è δ e maggiore è il valore di \tilde{n} oltre il quale vale la maggiorazione dell'errore $|L_n[f(x)](y) - f(y)| \leq \gamma \varepsilon$.

Osserviamo ancora che avremmo potuto maggiorare $|f(x) - f(y)|$ con $\varepsilon + \frac{2M}{\delta}|x - y|$, rimuovendo il δ^2 al denominatore. Però in questo caso non avremmo potuto esprimere $|x - y|$ come combinazione lineare dei polinomi $1, x, x^2$. Questo fatto è essenziale per arrivare alla maggiorazione richiesta.

8.1 Generalizzazioni

Teorema 76 (Teorema di Korovkin, versione m -dimensionale) Sia $K \subseteq \mathbb{R}^m$ un compatto, e sia $C(K)$ lo spazio delle funzioni continue da K a \mathbb{R} con la norma del sup (convergenza uniforme). Sia $(L_n)_{n=1}^\infty$ una successione di LPO su $C(K)$. Se L_n approssima bene le tre funzioni:

- $x \mapsto 1$ (la funzione costante uguale a 1),
- $x \mapsto x_i, i = 1, 2, \dots, m$ (la proiezione sulla i -esima componente)
- $x \mapsto x_i^2, i = 1, 2, \dots, m$ (la funzione “elevare al quadrato”),

allora L_n approssima bene tutte le funzioni di $C(K)$.

Vale una versione del teorema di Korovkin per funzioni di variabile reale a valori complessi. In questo caso la monotonia dell'operatore va intesa ristretta a funzioni che prendono valori reali, cioè se f è tale che $f(x) \in \mathbb{R}$ e $f(x) \geq 0$ allora $L[f]$ è una funzione a valori reali tale che $L[f](y) \geq 0$. Per linearità la monotonia si applica anche alla parte reale e alla parte immaginaria della funzione. Ad esempio, se $f(x) = f_1(x) + if_2(x)$, dove $i^2 = -1$, con $f_1(x)$ e $f_2(x)$ funzioni reali non negative, posto $g(y) = L[f(x)](y) = g_1(y) + ig_2(y)$, con $g_1(y), g_2(y)$ funzioni reali, allora la monotonia di L implica che anche $g_1(y)$ e $g_2(y)$ sono non negative.

Teorema 77 (Teorema di Korovkin, versione complessa) Sia $\mathcal{B} = \{f : [0, 2\pi] \rightarrow \mathbb{C}, f(0) = f(2\pi)\}$ (funzioni complesse continue e 2π -periodiche), con la norma del sup (convergenza uniforme). Sia $(L_n)_{n=1}^\infty$ una successione di LPO su \mathcal{B} . Se L_n approssima bene le tre funzioni:

- $x \mapsto 1$,
- $x \mapsto e^{ix}$,
- $x \mapsto e^{-ix}$,

allora L_n approssima bene tutte le funzioni di \mathcal{B} .

Rispetto alla versione 1D, cambia solo l'insieme delle funzioni su cui richiediamo la convergenza tra le ipotesi (il cosiddetto *test di Korovkin*).

8.2 Polinomi di Bernstein

I *polinomi di Bernstein* sono la successione di LPO su $K = [0, 1]$ definita da

$$B_n[f](x) := \sum_{k=0}^n \binom{n}{k} x^k (1-x)^{n-k} f\left(\frac{k}{n}\right)$$

Si verifichi per esercizio che i polinomi di Bernstein sono LPO.

8.3 Un'applicazione: dimostrazione del teorema di approssimazione di Weierstrass

Ci servirà nel seguito questo lemmetto con i binomiali.

Lemma 78 (formule “in-and-out” per i binomiali) *Valgono le seguenti identità.*

1. $\binom{n}{k} = \frac{n}{k} \binom{n-1}{k-1}$
2. $\binom{n}{k} = \frac{n(n-1)}{k(k-1)} \binom{n-2}{k-2}$

Dim. Applica le definizioni. . . □

Teorema 79 (di approssimazione di Weierstrass, caso 1D) *Sia $K \subseteq \mathbb{R}$ un compatto, e $f \in C(K)$. Allora, per ogni $\varepsilon > 0$ esiste un polinomio $p = p_{f,\varepsilon}$ tale che*

$$\|f - p\|_\infty \leq \varepsilon.$$

Dim. Ci basta provare il teorema per $K = [0, 1]$, poi con qualche semplice trasformazione del dominio possiamo estenderlo a tutti gli altri compatti (come?).

Mostriamo che

$$\|f - B_n[f]\|_\infty \rightarrow 0 \quad \text{quando } n \rightarrow \infty. \quad (14)$$

Quindi basta prendere $p = B_n[f]$ per un n sufficientemente grande. La (14) è la tesi del teorema di Korovkin per i B_n , quindi ci basta dimostrare che sono soddisfatte le ipotesi, cioè che $\|f - B_n[f]\|_\infty \rightarrow 0$ per le tre funzioni di test $x \mapsto 1$, $x \mapsto x$, $x \mapsto x^2$

- Per $x \mapsto 1$:

$$B_n[1](y) = \sum_{k=0}^n \binom{n}{k} y^k (1-y)^{n-k} = (y + (1-y))^n = 1,$$

quindi l'errore è costantemente uguale a 0.

- Per $x \mapsto x$:

$$\begin{aligned} B_n[x](y) &= \sum_{k=0}^n \binom{n}{k} y^k (1-y)^{n-k} \frac{k}{n} \\ &= \sum_{k=1}^n \binom{n}{k} y^k (1-y)^{n-k} \frac{k}{n} \\ &= \sum_{k=1}^n \binom{n-1}{k-1} y^k (1-y)^{n-k} \\ &= \sum_{h=0}^{n-1} \binom{n-1}{h} y^{h+1} (1-y)^{n-1-h} = y(1 + (1-y))^{n-1} = y \end{aligned}$$

(abbiamo usato la in-and-out formula e cambiato indice $h := k - 1$) quindi di nuovo l'errore è sempre nullo.

- Per $x \mapsto x^2$: usiamo l'identità

$$\frac{k^2}{n} = \frac{k}{n^2} + \frac{k^2}{n^2} - \frac{k}{n^2} = \frac{1}{n} \frac{k}{n} + \frac{n-1}{n} \frac{k(k-1)}{n(n-1)},$$

le in-and-out formulas, e il risultato del punto precedente:

$$\begin{aligned} B_n[x^2](y) &= \sum_{k=0}^n \binom{n}{k} y^k (1-y)^{n-k} \frac{k^2}{n^2} \\ &= \frac{1}{n} \left(\sum_{k=0}^n \binom{n}{k} y^k (1-y)^{n-k} \frac{k}{n} \right) \\ &\quad + \frac{n-1}{n} \left(\sum_{k=0}^n \binom{n}{k} y^k (1-y)^{n-k} \frac{k(k-1)}{n(n-1)} \right) \\ &= \frac{1}{n} y + \frac{n-1}{n} \left(\sum_{h=0}^{n-2} \binom{n-2}{h} y^{h+2} (1-y)^{n-2-h} \right) \\ &= \frac{1}{n} y + \frac{n-1}{n} y^2 = y^2 - \frac{1}{n} y^2 + \frac{1}{n} y \end{aligned}$$

(stavolta abbiamo fatto il cambio di variabile $h := k - 2$). Quindi $B_n[x \mapsto x^2]$ converge uniformemente a $x \mapsto x^2$ (come ci serviva per provare che valgono le ipotesi del teorema di Korovkin), anche se stavolta c'è un errore di $O(\frac{1}{n})$. Del resto non potevamo aspettarci che l'errore fosse zero anche stavolta: difatti, se nella dimostrazione del teorema di Korovkin si ha $\epsilon_i^n = 0$ per ogni $n \dots$

Quindi i polinomi di Bernstein soddisfano il teorema di Korovkin, e in particolare ci forniscono l'approssimazione che cercavamo sopra. \square

8.4 Caso multidimensionale

Il teorema di Weierstrass funziona pari pari in dimensione maggiore:

Teorema 80 (di approssimazione di Weierstrass) *Sia $K \subseteq \mathbb{R}^m$ un compatto, e $f \in C(K)$. Allora, per ogni $\varepsilon > 0$ esiste un polinomio $p = p_{f,\varepsilon}$ tale che*

$$\|f - p\|_\infty \leq \varepsilon.$$

Dim. Se $K = [0, 1]^m$, funziona lo stesso trucco di sopra. Definiamo i *polinomi di Bernstein m -dimensionali* come

$$B_n^{(m)}[f](y_1, \dots, y_n) := \sum_{\substack{k_1 = 1, \dots, n \\ k_2 = 1, \dots, n \\ \vdots \\ k_m = 1, \dots, n}} \binom{n}{k_1} y_1^{k_1} (1-y_1)^{n-k_1} \binom{n}{k_2} y_2^{k_2} (1-y_2)^{n-k_2} \dots \binom{n}{k_m} y_m^{k_m} (1-y_m)^{n-k_m} f\left(\frac{k_1}{n}, \frac{k_2}{n}, \dots, \frac{k_m}{n}\right)$$

Con un po' di lavoro si riesce a provare che sono un LPO (facile) e che soddisfano le ipotesi del teorema di Korovkin (non è difficile come sembra: si riutilizzano diverse volte i calcoli fatti per il caso 1D . . .), quindi concludiamo come sopra.

Stavolta però non è banale estendere il risultato da $K = [0, 1]^m$ a un qualunque altro compatto. È facile estenderlo con un cambio di variabile a qualunque “cubo” $[-t, t]^m$, ma per passare a un compatto qualsiasi ora serve usare anche questo risultato (con $H = [-t, t]^m$ un cubo “sufficientemente grande”):

Teorema 81 (caso \mathbb{R}^m del teorema di estensione di Tietze) *Sia $H \subseteq \mathbb{R}^m$ compatto, $K \subseteq H$ compatto in H , $f : K \rightarrow \mathbb{R}$ continua. Allora esiste $\tilde{f} : H \rightarrow \mathbb{R}$ che estende f .*

Il risultato è un caso particolare di un teorema più generale di topologia; non conosco una dimostrazione rapida e indolore di questo teorema nel caso \mathbb{R}^m che serve a noi². \square

Riferimenti bibliografici

- [1] D. Bini, M. Capovani, O. Menchi, *Metodi Numerici per l'Algebra Lineare*. Zanichelli, Bologna, 1987.
- [2] R. Bevilacqua, D. Bini, M. Capovani e O. Menchi, *Metodi numerici*, Zanichelli, Bologna 1992.
- [3] Simon J. Smith, Lebesgue constants in polynomial interpolation, *Annales Mathematicae et Informaticae*, v. 33 (2006) pp. 109-123.
www.ektf.hu/tanszek/matematika/ami
- [4] J. Stoer, R. Burlisch, *Introduction to Numerical Analysis*, Third Edition, Springer, 2002.
- [5] W. Rudin, *Real and Complex Analysis*, Second Edition, Tata McGraw-Hill, 1974.

²Quando l'ho chiesto a un paio di analisti mi hanno assicurato che “ma sì, si dovrebbe fare con un po' di lavoro, considerando il modulo di continuità di f su K , che è uniformemente continua, estendendola localmente e usando la compattezza. . .”