

Corso di Analisi Numerica - AN410

Parte 4: approssimazione polinomiale

Roberto Ferretti



- Le principali strategie di approssimazione
- Il polinomio interpolatore: forme di Lagrange e Newton
- L'errore di interpolazione per funzioni regolari
- Convergenza del polinomio interpolatore: strategie di infittimento
- Il polinomio di Hermite
- Approssimazione per Errore Quadratico Minimo

Le principali strategie di approssimazione

Il problema di **approssimare una funzione** (nel caso più semplice, di una sola variabile) si pone qualora sia necessario fornire una espressione analitica approssimata di una funzione:

- che sia nota solo tramite **punti del suo grafico**;
- che **non possa essere calcolata tramite operazioni elementari** (ad esempio le funzioni trascendenti);
- su cui **non si possano effettuare in modo elementare ed esplicito alcune operazioni** (ad esempio, l'integrazione).

Il caso di gran lunga più frequente è quello in cui la approssimazione di f si scrive in forma di combinazione lineare:

$$f(x) = c_0\phi_0(x) + c_1\phi_1(x) + \cdots + c_n\phi_n(x) + E_n(x). \quad (1)$$

- La scelta tipica della famiglia $\{\phi_k\}$ è una base polinomiale, anche se non necessariamente la base canonica $1, x, x^2, \dots, x^n$ (altre scelte: esponenziali complessi, funzioni razionali)
- La famiglia $\{\phi_k\}$ si suppone densa in uno spazio opportuno X contenente f (si vuole che $\|E_n\|_X \rightarrow 0$). Lavorando con polinomi, tipicamente $X \equiv C^0$ o $X \equiv L^2$) (teoremi di densità di Weierstrass)

Polinomio di Taylor: è il polinomio $T_n \in \mathbb{P}_n$ tale che le sue prime n derivate in un punto x_0 coincidono con quelle di $f(x)$:

$$T_n^{(k)}(x_0) = f^{(k)}(x_0) \quad (k = 0, \dots, n)$$

- Questa approssimazione è la migliore possibile nell'intorno di x_0 , ma peggiora molto velocemente allontanandosene, quindi è poco adatta ad approssimare $f(x)$ in tutto un intervallo
- La convergenza (uniforme) ad $f(x)$ avviene sotto la ipotesi di analiticità, quindi per un insieme di funzioni molto ristretto e di scarsa rilevanza pratica

Polinomio interpolatore: è il polinomio $\Pi_n \in \mathbb{P}_n$ tale che il suo valore coincide in $n + 1$ punti x_0, \dots, x_n con quello di $f(x)$:

$$\Pi_n(x_i) = f(x_i) \quad (i = 0, \dots, n)$$

- Questa approssimazione non richiede la conoscenza delle derivate di f (anzi neanche la loro esistenza), ed utilizza informazioni che provengono da tutto un intervallo
- Come si vedrà, la convergenza ad $f(x)$ dipende in modo non banale dalla regolarità di f , ma anche dalla disposizione dei punti x_0, \dots, x_n

Polinomio di Hermite: combina le due strategie precedenti. E' il polinomio $H_m \in \mathbb{P}_m$ tale che il suo valore, insieme con il valore di un certo numero di derivate, coincide in $n + 1$ punti x_0, \dots, x_n con i corrispondenti valori di $f(x)$:

$$H_m^{(p)}(x_i) = f^{(p)}(x_i) \quad (i = 0, \dots, n; p = 0, \dots, m_i - 1).$$

- Questa strategia richiederebbe la conoscenza di alcune derivate di f , che in qualche caso si possono sostituire con derivate stimate
- In genere il polinomio di Hermite non si utilizza nella sua forma più generale

Errore Quadratico Minimo: si utilizza per descrivere un numero “grande” di punti con un polinomio π di grado “basso” che minimizza lo scarto quadratico rispetto ad m coppie $(x_1, y_1), \dots, (x_m, y_m)$:

$$\sum_{i=1}^m [\pi(x_i) - y_i]^2 = \min_{q \in \mathbb{P}_n} \sum_{i=1}^m [q(x_i) - y_i]^2 \quad (m \gg n)$$

- Questo tipo di approssimazione si utilizza tipicamente per dati affetti da deviazioni aleatorie, in particolare misure sperimentali, e corrisponde in statistica bayesiana al criterio della massima verosimiglianza
- Generalmente in questo caso il polinomio π viene costruito esplicitamente nella base $1, x, \dots, x^n$

Errore di minima norma: si sceglie il polinomio $p \in \mathbb{P}_n$ che minimizza la norma dell'errore:

$$\|p - f\| = \min_{q \in \mathbb{P}_n} \|q - f\|$$

- Pur basandosi su un criterio generale ed elegante, questa strategia di approssimazione porta a caratterizzare il polinomio p ma **non ne fornisce in generale una espressione esplicita**
- Un caso relativamente esplicito è quello delle **serie di Fourier troncate**, che corrispondono alla approssimazione in norma L^2

Il polinomio interpolatore: forme di Lagrange e Newton

Se i punti (detti *nodi*) x_0, \dots, x_n sono distinti, esiste ed è unico il polinomio $\Pi_n \in \mathbb{P}_n$ tale che:

$$\Pi_n(x_i) = f(x_i) \quad (i = 0, \dots, n)$$

- La dimostrazione dell'unicità si basa sul teorema fondamentale dell'algebra
- La dimostrazione dell'esistenza si basa sulla costruzione di un polinomio che soddisfa le condizioni di interpolazione

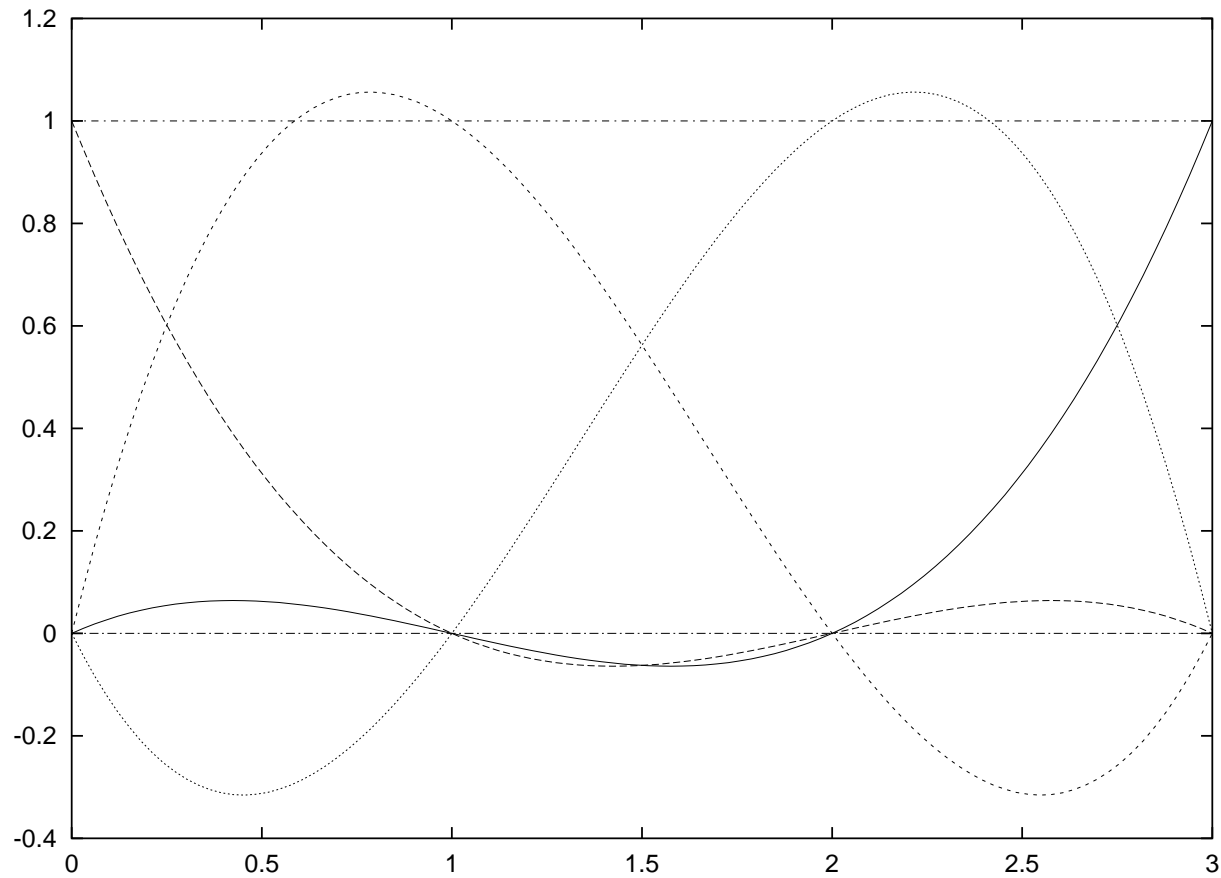
Polinomio di Lagrange: si indica con questo nome il polinomio

$$\Pi_n(x) = \sum_{i=0}^n f(x_i) L_i(x) \quad (2)$$

dove le funzioni $L_i(x)$ (che costituiscono la cosiddetta *base di Lagrange*) sono definite da

$$L_i(x) = \prod_{j \neq i} \frac{x - x_j}{x_i - x_j} \quad (i = 0, \dots, n) \quad (3)$$

- Le funzioni $L_i \in \mathbb{P}_n$ sono linearmente indipendenti e quindi **costituiscono una base in \mathbb{P}_n**
- Si ha che $L_i(x_k) = \delta_{ik}$, e di conseguenza, $\Pi_n(x) = f(x_i)$.



Base di Lagrange $L_i(x)$ relativa ai nodi $x_0 = 0, \dots, x_3 = 3$

- La **complessità computazionale** dominante del polinomio di Lagrange è data dal calcolo delle $n + 1$ funzioni di base $L_i(x)$, che richiede per ogni funzione $n + 1$ somme ed n prodotti al numeratore ed altrettante operazioni al denominatore, per un totale di $O(4n^2)$ operazioni in virgola mobile
- Questa complessità è associata ad ogni calcolo di $\Pi_n(x)$
- Aggiungendo un nodo il calcolo va sostanzialmente rifatto di nuovo (non è possibile utilizzare efficientemente Π_{n-1} per calcolare Π_n)

- In presenza di perturbazioni δ_i di modulo $|\delta_i| \leq \delta$ sui valori $f(x_i)$, il calcolo del polinomio di Lagrange fornisce

$$\Pi_n(x) + \Delta = \sum_{i=0}^n [f(x_i) + \delta_i] L_i(x) = \Pi_n(x) + \sum_{i=0}^n \delta_i L_i(x)$$

e la perturbazione risultante Δ si maggiora come

$$|\Delta| \leq \delta \sum_{i=0}^n |L_i(x)|$$

La funzione (detta *funzione di Lebesgue*) $\sum_{i=0}^n |L_i(x)|$ è quindi un coefficiente di amplificazione delle perturbazioni δ_i

- Al crescere di n , la costante di Lebesgue (ovvero il massimo della funzione di Lebesgue) può divergere molto velocemente

Polinomio di Newton: si indica con questo nome il polinomio

$$\Pi_n(x) = f[x_0] + f[x_0, x_1](x - x_0) + \cdots + f[x_0, \dots, x_n](x - x_0) \cdots (x - x_{n-1}) \quad (4)$$

dove le costanti $f[\dots]$ (dette *differenze divise* della funzione f) sono definite ricorsivamente nel modo seguente:

$$\begin{cases} f[x_0, \dots, x_k] = \frac{f[x_1, \dots, x_k] - f[x_0, \dots, x_{k-1}]}{x_k - x_0}, \\ f[x_0] = f(x_0). \end{cases} \quad (5)$$

- In questo caso la base in \mathbb{P}_n è costituita dalle funzioni 1 , $(x - x_0)$, $(x - x_0)(x - x_1), \dots, (x - x_0) \cdots (x - x_{n-1})$ (si tratta di polinomi di grado crescente e quindi linearmente indipendenti)

Le **differenze divise** della funzione f sono calcolate in forma di tavola a partire dai valori della funzione:

$$f[x_0] = f(x_0)$$

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

$$f[x_1] = f(x_1)$$

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

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

$$f[x_2] = f(x_2)$$

- Un tipico problema di questa costruzione è la **perdita di cifre significative per sottrazione**, che ha luogo sottraendo valori simili (e sempre più simili all'infittirsi dei nodi)

- La **complessità computazionale** dominante del polinomio di Newton è data dal calcolo della tavola delle differenze divise, che richiede un totale di $O(3n^2/2)$ operazioni in virgola mobile

- Il calcolo di $\Pi_n(x)$ ha invece **complessità lineare**, se posto nella forma

$$\begin{aligned} \Pi_n(x) = & f[x_0] + (x - x_0) \left(f[x_0, x_1] + (x - x_1) \left(f[x_0, x_1, x_2] + \dots + \right. \right. \\ & \left. \left. + (x - x_{n-2}) \left(f[x_0, \dots, x_{n-1}] + (x - x_{n-1}) f[x_0, \dots, x_n] \right) \dots \right) \right) \end{aligned}$$

- Aggiungere un nodo richiede di aggiungere una riga alla tavola delle differenze e ricalcolare il polinomio o anche il solo termine di grado più alto (**calcolare Π_n a partire da Π_{n-1} ha quindi complessità lineare**)

- Si presentano per il polinomio di Newton gli stessi problemi di stabilità già visti per il polinomio di Lagrange, che sono insiti nell'operazione di interpolazione
- Questa instabilità può essere spiegata anche con il fatto che la costruzione della tavola delle differenze è una operazione instabile ad esempio a causa della perdita di cifre significative per sottrazione, che produce errori crescenti al crescere dell'ordine delle differenze (e quindi al crescere del grado delle funzioni di base)

L'errore di interpolazione per funzioni regolari

Definito l'intervallo $I = [\min(x, x_0, \dots, x_n), \max(x, x_0, \dots, x_n)]$, se $f \in C^{n+1}(I)$, allora l'errore di interpolazione si può rappresentare come:

$$f(x) - \Pi_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \omega_n(x) \quad (6)$$

dove $\omega_n(x) = (x - x_0)(x - x_1) \cdots (x - x_n)$, e $\xi \in I$.

- Da questa formula di rappresentazione non è ovvio che la differenza $f - \Pi_n$ possa andare a zero al crescere del grado di interpolazione
- Il punto cruciale è di conoscere il comportamento del polinomio ω_n (che può avere grado alto) e delle derivate successive di f

Stime di errore più maneggevoli possono essere ottenute:

- **maggiorando il valore assoluto di $f^{(n+1)}$** (ed eventualmente di ω_n se si vuole ottenere una **stima uniforme su I**):

$$|f(x) - \Pi_n(x)| \leq \frac{\sup |f^{(n+1)}|}{(n+1)!} |\omega_n(x)| \leq \frac{\sup |f^{(n+1)}|}{(n+1)!} \sup_I |\omega_n(x)| \quad (7)$$

- **maggiorando ulteriormente anche il valore assoluto di ω_n con $|I|^{n+1}$** (questo si ottiene dalla maggiorazione $|x - x_i| \leq |I|$):

$$\sup_I |f(x) - \Pi_n(x)| \leq \frac{\sup |f^{(n+1)}|}{(n+1)!} |I|^{n+1} \quad (8)$$

La stima (8):

- E' di applicazione più semplice della (7), non dipendendo dalla posizione dei nodi
- E' estremamente rozza (la maggiorazione $|x - x_i| \leq |I|$ non è in genere ottimale) a meno che $|x - x_i| \approx |I|$, ovvero se si interpola lontano dai nodi, situazione che si evita quanto possibile
- Fornisce comunque un ordine di infinitesimo dell'errore, quando la distanza tra i nodi possa essere trattata come un parametro di discretizzazione (caso delle interpolazioni composite)

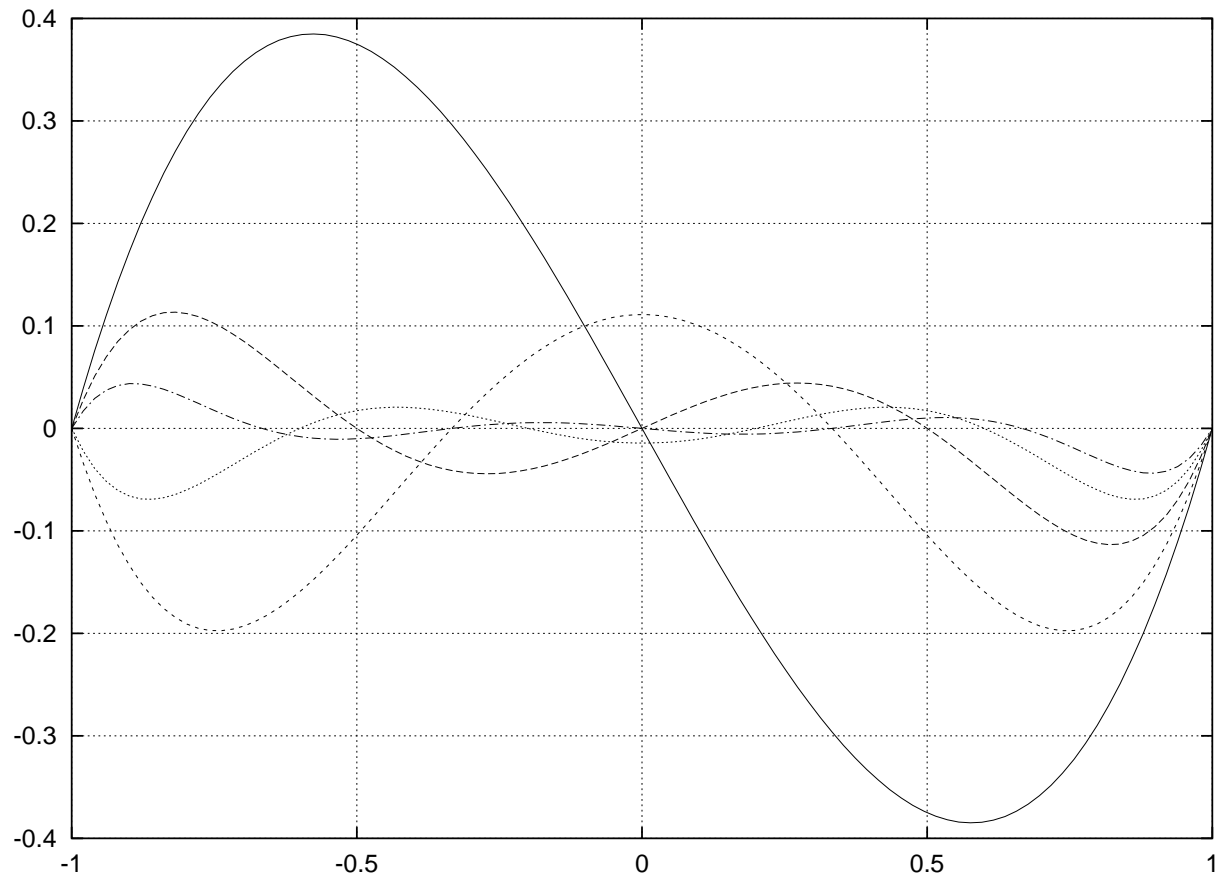
Condizione sufficiente perché la maggiorazione (8) converga a zero (quindi che $\Pi_n \rightarrow f$ uniformemente in I) per $n \rightarrow \infty$ è che

$$\sup_I |f^{(n+1)}| = o\left(\frac{(n+1)!}{|I|^{n+1}}\right)$$

- Questa è una condizione sufficiente di analiticità per f , simile a quella che garantisce la convergenza a zero del resto di Lagrange
- L'insieme di funzioni per cui si ottiene convergenza uniforme del polinomio interpolatore è quindi troppo ristretto se non si fanno ipotesi sulla collocazione dei nodi (cioè se non si usa la stima (7))

Nodi equidistanti: si tratta della strategia più frequente di disposizione dei nodi

- In molte situazioni, l'utilizzo di questa strategia può essere obbligato per il fatto di **interpolare dei dati tabulati ad intervalli costanti**
- Esistono **risultati generali di convergenza** per la interpolazione a nodi equidistanti, che però **richiedono comunque la analiticità**
- Le proprietà dei polinomi ω_n possono essere studiate in un quadro abbastanza generale



Grafici dei polinomi $\omega_n(x)$ ($2 \leq n \leq 6$, nodi **equidistanti** in $[-1, 1]$)

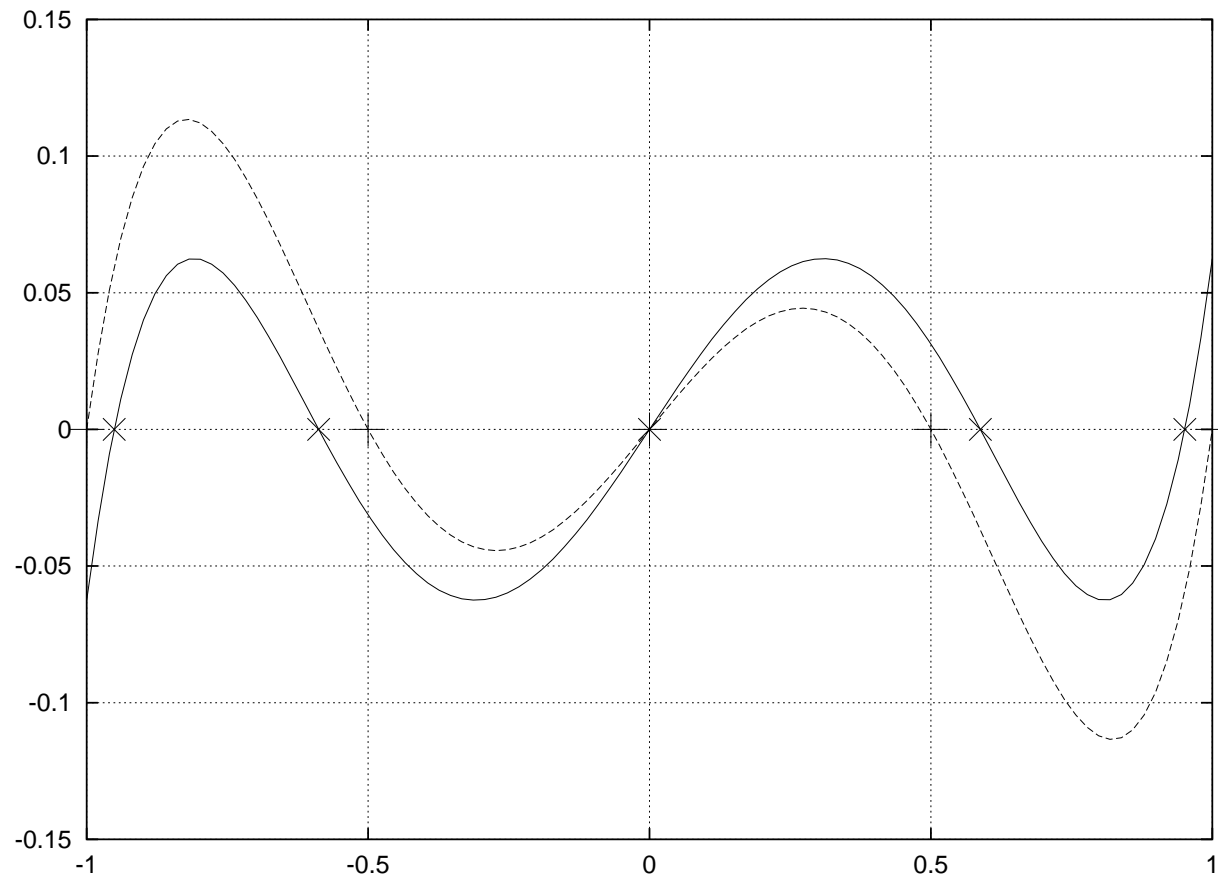
- Al crescere del grado n , $\sup_I |\omega_n(x)| \rightarrow 0$
- La convergenza è però più veloce al centro dell'intervallo, mentre alle estremità ci sono oscillazioni più forti
- Questa considerazione indica che la strategia di interpolazione a nodi equidistanti fornisce risultati buoni al centro dell'intervallo di interpolazione ma che si degradano fortemente ai bordi
- Distribuendo in modo opportuno i nodi (in particolare, infittendoli alle estremità) si può ottenere una accuratezza più uniforme in tutto l'intervallo

Nodi di Chebyshev: si collocano i nodi in corrispondenza degli zeri di un polinomio di Chebyshev di grado $n + 1$, ovvero se $I = [a, b]$:

$$\begin{cases} x_j = a + \frac{b-a}{2} \left(1 - \cos \frac{j\pi}{n}\right) & \text{se } x_0 = a, x_n = b \\ x_j = a + \frac{b-a}{2} \left(1 - \cos \frac{(2j+1)\pi}{2n+2}\right) & \text{se } x_j \in (a, b) \end{cases}$$

- In questo caso, ω_n è esso stesso un polinomio di Chebyshev di grado $n + 1$ a meno di costanti moltiplicative
- Si dimostra che questa scelta dei nodi minimizza la quantità

$$\|\omega_n\|_\infty = \sup_I |\omega_n(x)|$$



Grafici dei polinomi $\omega_4(x)$ (nodi equidistanti e nodi Chebyshev)

- Con questa scelta dei nodi, si ha ancora $\sup_I |\omega_n(x)| \rightarrow 0$, ma tutti gli estremi di ω_n presentano valori di modulo uguale
- La accuratezza della interpolazione aumenta ai bordi e peggiora nella regione centrale dell'intervallo, risultando comunque globalmente migliore
- Si può anche dimostrare che la costante di Lebesgue cresce più lentamente che nel caso dei nodi equidistanti al crescere di n

Convergenza del polinomio interpolatore: strategie di infittimento

Il fatto che, per una distribuzione generica od anche uniforme dei nodi, si ottenga convergenza del polinomio interpolatore per $n \rightarrow \infty$ solo sulle funzioni analitiche è un risultato di scarsa utilità, tenendo conto che i polinomi sono densi in uno spazio molto più grande (quello delle funzioni continue).

Il risultato più generale di convergenza delle approssimazioni per interpolazione in una base generica $\{\phi_0, \dots, \phi_n\}$ stima l'errore a partire dalla costante di Lebesgue e dall'errore della migliore approssimazione

Supponendo $f, \phi_0, \dots, \phi_n \in C^0([a, b])$, e $\phi_i(x_j) = \delta_{ij}$, definendo $X_n = \text{span}\{\phi_0, \dots, \phi_n\}$,

$$\varepsilon_n(f) = \min_{p_n \in X_n} \max_{x \in [a, b]} |f(x) - p_n(x)| \quad (\text{distanza di } f \text{ da } X_n)$$

$$\Lambda_n = \max_{x \in [a, b]} \sum_{i=0}^n |\phi_i(x)| \quad (\text{costante di Lebesgue})$$

si ha:

$$\max_{x \in [a, b]} \left| f(x) - \sum_{i=0}^n f(x_i) \phi_i(x) \right| \leq (1 + \Lambda_n) \varepsilon_n(f).$$

- Mentre la distanza di f da X_n è praticamente un dato del problema (una volta fissata la base dell'interpolazione), questo risultato mostra che il ruolo chiave è quello della costante di Lebesgue.
- Le due situazioni tipiche che permettono di recuperare la convergenza in ipotesi abbastanza generali sono una scelta opportuna dei nodi (ad esempio i nodi di Chebyshev) e le interpolazioni composite in cui i nodi vengono infittiti senza aumentare il grado del polinomio.
- Nella prima situazione la costante di Lebesgue cresce “lentamente”, nella seconda non cresce affatto all'aumentare del numero di nodi.

Nel caso dei **nodi di Chebyshev**, si costruisce appunto un unico polinomio interpolatore di grado n e si ottiene la convergenza di Π_n ad f al crescere di n . Più precisamente, **se $f \in C^k(I)$ ($k \geq 1$), il polinomio interpolatore $\Pi_n(x)$ converge a $f(x)$ per $n \rightarrow \infty$, e vale la stima**

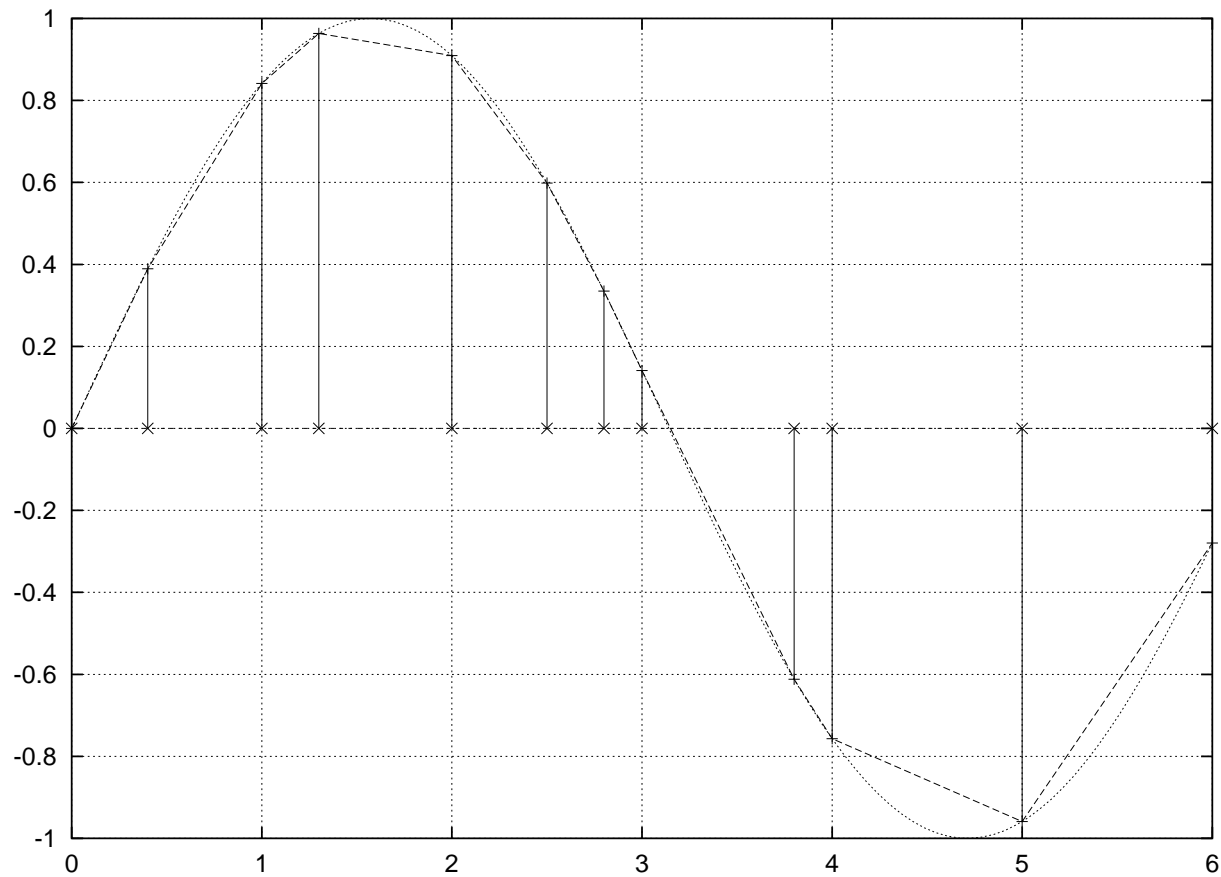
$$|f(x) - \Pi_n(x)| \leq Cn^{-k}$$

per ogni $x \in I$.

- In questo tipo di infittimento, **la crescita “lenta” della costante di Lebesgue** può essere dimostrata anche per **altre famiglie di nodi non uniformi** (ad esempio quelli di Gauss–Legendre)

Nel caso delle interpolazioni composite, si tratta di suddividere l'intervallo di interpolazione in sottointervalli $[a_i, b_i]$ di ampiezza $b_i - a_i = H_i \leq H$ costruendo un diverso polinomio interpolatore di grado n per ogni sottointervallo. La approssimazione risultante $\Pi_{n,H}$ è polinomiale a tratti.

- Il parametro di discretizzazione in questo caso è H : cioè si migliora la approssimazione riducendo la misura ed aumentando il numero dei sottointervalli, ma senza aumentare il grado del polinomio interpolatore in ogni sottointervallo



Interpolata di grado $n = 1$ a tratti della funzione $f(x) = \sin x$

La convergenza delle approssimazioni composite si ottiene applicando la maggiorazione (8) ad ogni sottointervallo. Più precisamente, se $f \in C^{n+1}(I)$ ($n \geq 0$), il polinomio interpolatore composito $\Pi_{n,H}(x)$ converge a $f(x)$ per $H \rightarrow 0$, e vale la stima

$$|f(x) - \Pi_{n,H}(x)| \leq \frac{\sup_I |f^{(n+1)}(x)|}{(n+1)!} H^{n+1}$$

per ogni $x \in I$.

- In questa strategia non è necessario che le ampiezze elementari H_i siano costanti: in pratica la massima efficienza dell'approssimazione si ottiene per $H_i \sim \left[\sup_{I_i} |f^{(n+1)}(x)| \right]^{-1/(n+1)}$

- Nel confronto tra le due strategie, la interpolazione di tipo Chebyshev presenta errori minori a parità di numero di valutazioni della funzione. Richiede però di lavorare a nodi non equidistanti ed in generale irrazionali; inoltre le singolarità della funzione hanno effetto su tutto l'intervallo di interpolazione.
- Le interpolazioni composite sono più adatte ad approssimare funzioni non troppo regolari, anche grazie alla possibilità di variare in modo adattativo le ampiezze H_i . Permettono inoltre di trattare più facilmente geometrie complesse del dominio di interpolazione (problema che si può porre nella approssimazione di funzioni di più variabili)

[indice](#)

Il polinomio di Hermite

Questo tipo di approssimazione è una combinazione delle strategie di interpolazione e sviluppo di Taylor. Dati $n + 1$ nodi x_0, \dots, x_n ed $n + 1$ interi m_0, \dots, m_n tali che $\sum_i m_i = m + 1$, chiamiamo *polinomio di Hermite* il polinomio $H_m \in \mathbb{P}_m$ tale che in ogni nodo x_i il suo valore, insieme con il valore delle sue derivate fino all'ordine $m_i - 1$, uguaglia i corrispondenti valori di $f(x)$:

$$H_m^{(p)}(x_i) = f^{(p)}(x_i) \quad (i = 0, \dots, n; p = 0, \dots, m_i - 1).$$

- In particolare, la formula di Taylor si ottiene ponendo $n = 0$, mentre il polinomio interpolatore si ottiene ponendo $m_i \equiv 1$

Il polinomio di Hermite di grado m relativo a $n + 1$ nodi x_0, \dots, x_n si può scrivere come:

$$H_m(x) = \sum_{i=0}^n \sum_{k=0}^{m_i-1} f^{(k)}(x_i) L_{ik}(x) \quad (9)$$

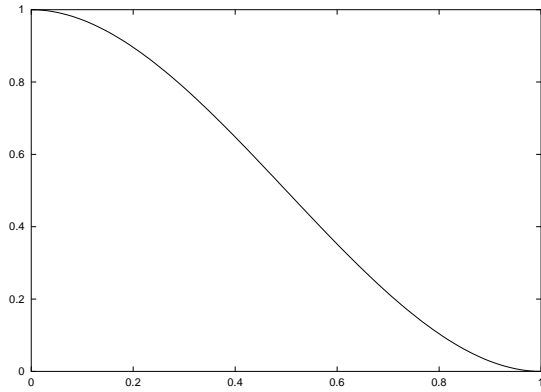
dove i polinomi di base (di grado m) $L_{ik}(x)$ sono caratterizzati dalle condizioni

$$L_{ik}^{(p)}(x_j) = \begin{cases} 1 & \text{se } i = j \text{ e } p = k \\ 0 & \text{altrimenti} \end{cases} \quad (10)$$

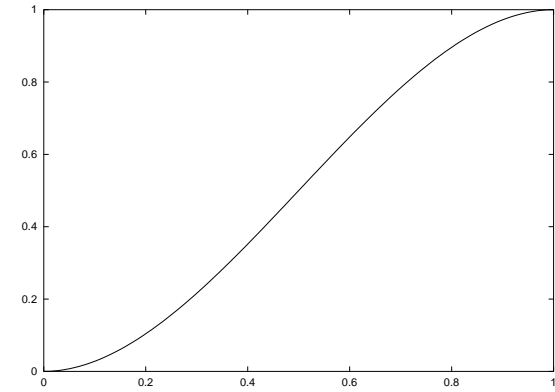
In base alla proprietà (10) (analogamente a quanto fatto per il polinomio di Lagrange) è facile verificare che (9) è effettivamente il polinomio che soddisfa le condizioni di Hermite.

Esempio: poniamo $x_0 = 0$, $x_1 = 1$, $m_0 = m_1 = 2$ (si tratta di una situazione tipica di utilizzo del polinomio di Hermite). Le funzioni L_{ik} sono date da:

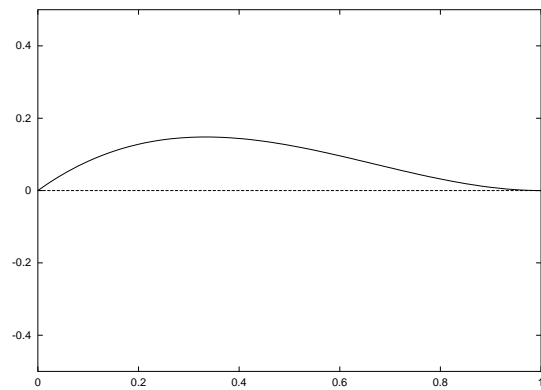
	$i = 0$	$i = 1$
$k = 0$	$L_{00}(x) = 2x^3 - 3x^2 + 1$	$L_{10}(x) = -2x^3 + 3x^2$
$k = 1$	$L_{01}(x) = x^3 - 2x^2 + x$	$L_{11}(x) = x^3 - x^2$



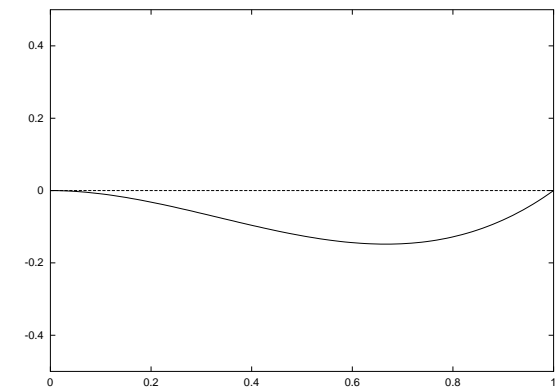
$L_{00}(x)$



$L_{10}(x)$



$L_{01}(x)$



$L_{11}(x)$

Nel caso più generale, i polinomi L_{ik} si calcolano per $i = 0, \dots, n$ in forma ricorrente a partire da $k = m_i - 1$:

$$\begin{cases} L_{i,m_i-1}(x) = l_{i,m_i-1}(x) \\ L_{ik}(x) = l_{ik}(x) - \sum_{p=k+1}^{m_i-1} l_{ik}^{(p)}(x_i) L_{ip}(x) \quad (k = m_i - 2, \dots, 0), \end{cases}$$

dove

$$l_{ik}(x) = \frac{(x - x_i)^k}{k!} \prod_{j \neq i} \left(\frac{x - x_j}{x_i - x_j} \right)^{m_j} \quad (k = 0, \dots, m_i - 1).$$

Questa costruzione è molto complessa e raramente si implementa nella sua generalità.

Definito l'intervallo $I = [\min(x, x_0, \dots, x_n), \max(x, x_0, \dots, x_n)]$, se $f \in C^{m+1}(I)$, allora l'errore di approssimazione si può rappresentare come:

$$f(x) - H_m(x) = \frac{f^{(m+1)}(\xi)}{(m+1)!} \Omega_m(x)$$

dove $\Omega_m(x) = (x - x_0)^{m_0} (x - x_1)^{m_1} \dots (x - x_n)^{m_n}$, $\xi \in I$.

- Questa formula di rappresentazione contiene come casi particolari sia il resto di Lagrange per la formula di Taylor, sia l'errore di interpolazione (6)

[indice](#)

Approssimazione per Errore Quadratico Minimo

In questa metodologia di approssimazione, si cerca di approssimare un numero m “grande” di punti con un modello più semplice (spesso una semplice retta) nella forma

$$\pi(x) = a_0\phi_0(x) + \cdots + a_n\phi_n(x) \quad (n \ll m)$$

Per approssimare i dati “al meglio”, senza richiedere il passaggio per tutti i punti, si minimizza un indice quadratico di scarto:

$$\begin{aligned} r(a) = r(a_0, \dots, a_n) &= \sum_{i=1}^m [\pi(x_i) - y_i]^2 = \\ &= \sum_{i=1}^m [a_0\phi_0(x_i) + \cdots + a_n\phi_n(x_i) - y_i]^2 \end{aligned}$$

- Questa strategia si rende necessaria tipicamente **quando i dati siano affetti da deviazioni aleatorie** (la interpolazione è infatti **molto instabile rispetto alle perturbazioni**)
- In questo caso, l'approssimazione per errore quadratico minimo corrisponde in statistica Bayesiana al criterio di **massima verosimiglianza**
- Se $\phi_i(x) = x^i$ e $n + 1 = m$, la minimizzazione dello scarto quadratico **coincide con la interpolazione** (il polinomio interpolatore viene però calcolato nella base naturale)

Ponendo $a = (a_0 \cdots a_n)^t$, $y = (y_1 \cdots y_m)^t$, e

$$\Phi = \begin{pmatrix} \phi_0(x_1) & \cdots & \phi_n(x_1) \\ \vdots & & \vdots \\ \phi_0(x_m) & \cdots & \phi_n(x_m) \end{pmatrix}$$

Lo scarto r si scrive come una **forma quadratica nella variabile a**

$$r(a) = (\Phi a - y)^t (\Phi a - y) = a^t \Phi^t \Phi a - 2a^t \Phi^t y - y^t y$$

- la matrice Hessiana $2\Phi^t\Phi$ è **semidefinita positiva** (ed in realtà definita positiva, a meno di scelte “infelici” dei nodi x_i)
- Il punto di minimo è soluzione del sistema di **condizioni di stazionarietà** $\Phi^t\Phi a = \Phi^t y$ (detto **sistema delle equazioni normali**)

Il caso più tipico di applicazione di questa tecnica è la **approssimazione di grado $n = 1$** (detta *retta ai minimi quadrati*):

$$y = a_1x + a_0,$$

in cui i parametri a_0 e a_1 devono soddisfare il sistema:

$$\begin{cases} a_0 \sum_i 1 + a_1 \sum_i x_i = \sum_i y_i \\ a_0 \sum_i x_i + a_1 \sum_i x_i^2 = \sum_i x_i y_i \end{cases}$$

(in cui ovviamente $\sum_i 1 = m$, numero dei punti)

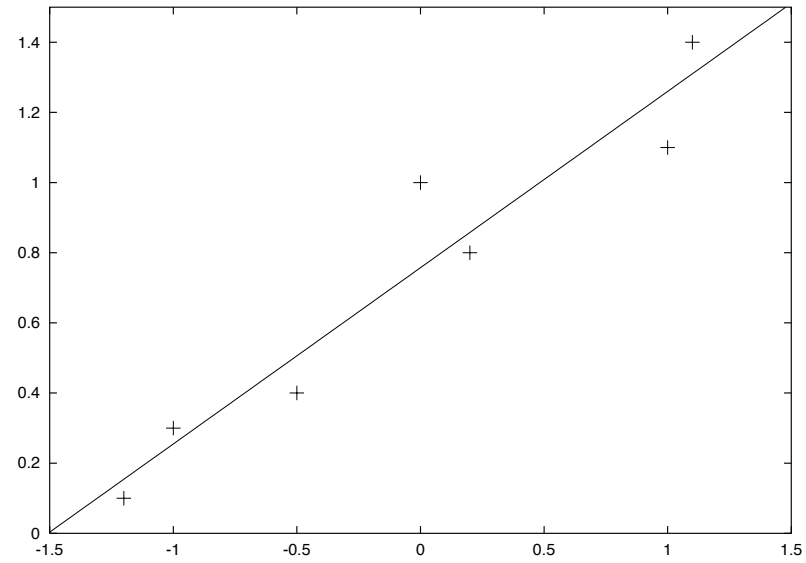
Esempio: data la tabella di punti

x_i	-1.2	-1.0	-0.5	0.0	0.2	1.0	1.1
y_i	0.1	0.3	0.4	1.0	0.8	1.1	1.4

il sistema delle equazioni normali associato è

$$\begin{cases} 7 a_0 - 0.4 a_1 = 5.1 \\ -0.4 a_0 + 4.94 a_1 = 2.18 \end{cases}$$

la cui soluzione, con sei cifre decimali, è $a_0 = 0.757292$, $a_1 = 0.502615$.



Retta di minimo errore quadratico per la tabella dell'esempio

indice