Approssimazione minimax

1 Il problema dell'approssimazione lineare

Data una $f(x)$ appartenente allo spazio vettoriale $\mathcal{F}$ delle funzioni reali di variabile reale, si sceglie in $\mathcal{F}$ un *modello*, cioè un insieme di funzioni $\phi_i(x)$, $i = 0, 1, \ldots$, linearmente indipendenti e si considera l'insieme delle combinazioni lineari

$$g_n(x) = \sum_{i=0}^{n} \alpha_i \phi_i(x), \quad n \text{ intero.}$$

Fissato l'intero n, per ottenere una funzione che approssimi la $f(x)$ dovremo determinare i coefficienti α_i in modo opportuno. Formalmente il problema dell'approssimazione lineare viene posto così: fissata in $\mathcal{F}$ una norma, si vogliono determinare $n+1$ coefficienti $\alpha_0^*, \ldots, \alpha_n^*$ tali che

$$\|f - \sum_{i=0}^{n} \alpha_i^* \phi_i\| = \min_{\boldsymbol{\alpha}} \|f - \sum_{i=0}^{n} \alpha_i \phi_i\|,$$

dove $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_n)$. La funzione

$$g_n^*(x) = \sum_{i=0}^{n} \alpha_i^* \phi_i(x)$$

è la *funzione di (migliore) approssimazione* rispetto alla norma fissata e la quantità

$$\delta_n^* = \|f - g_n^*\|$$

ne è l'*errore assoluto in norma.*

Dalla teoria si sa che

- Il problema dell'approssimazione lineare ha soluzione,
- le soluzioni del problema formano un insieme convesso,
- la successione $\{\delta_n^*\}_{n \in \mathbf{N}}$ è monotona non crescente, e quindi esiste il $\lim_{n\to\infty} \delta_n^* \geq 0$.

Ne segue che il problema o ha una sola soluzione o ne ha infinite. Si tratta quindi di stabilire quali ipotesi assicurano l'unicità della soluzione. Inoltre, perché

la funzione $g_n^*(x)$ possa essere considerata una buona approssimazione della $f(x)$, dovrebbe accadere che

$$\lim_{n\to\infty} \delta_n^* = 0, \tag{1}$$

cioè che al crescere di n la successione delle approssimazioni *converga* alla funzione $f(x)$, mentre il teorema assicura solo l'esistenza del limite, ma non che tale limite sia nullo.

I due problemi, dell'unicità e della convergenza, non possono essere affrontati riferendosi a una generica norma: è necessario specificare la particolare norma rispetto alla quale si vuole minimizzare. Entrambi i problemi trovano facile soluzione se la norma che si considera è la norma 2. L'approssimazione in norma 2 viene anche detta *approssimazione ai minimi quadrati* e per essa la (1) vale. Quindi è possibile, fissato ϵ, determinare n in modo che il polinomio approssimante ai minimi quadrati $g_n^*(x)$ soddisfi $\delta_n^* < \epsilon$. Questo però non garantisce che la condizione $|f(x) - g_n^*(x)| < \epsilon$ sia soddisfatta per ogni x dell'intervallo $[a, b]$ di definizione della $f(x)$. Cioè con la norma 2 si ottiene un'approssimazione in media sull'intervallo, che può essere buona in certi punti e meno in altri. Se invece è richiesto di determinare un'approssimazione che soddisfi la condizione in ogni punto x di $[a, b]$, come generalmente avviene quando si calcolano funzioni matematiche con un calcolatore, si deve usare la *norma* ∞, definita da

$$\|f\|_\infty = \max_{x\in[a,b]} |f(x)|.$$

L'approssimazione in norma ∞ viene detta anche *approssimazione minimax*. A differenza dell'approssimazione ai minimi quadrati non esistono metodi espliciti per calcolarla.

2 Approssimazione minimax polinomiale

La norma ∞ non è indotta da nessun prodotto scalare, pertanto la teoria dell'approssimazione su spazi di Hilbert sviluppata per la norma 2 non può essere utilizzata nel caso della norma ∞. In questo paragrafo si esamina il problema dell'approssimazione lineare in norma ∞ quando l'insieme delle funzioni approssimanti è l'insieme dei polinomi.

Sia $\mathcal{P}_n$ la classe dei polinomi di grado minore o uguale ad n e sia $f(x)$ una funzione continua su un intervallo limitato $[a, b]$. Un polinomio $p_n^* \in \mathcal{P}_n$ di *approssimazione minimax* è tale che

$$\|f - p_n^*\|_\infty = \min_{p_n\in\mathcal{P}_n} \|f - p_n\|_\infty.$$

Quindi, posto $\delta^* = \|f - p_n^*\|_\infty$, è

$$\delta^* \le \|f - p_n\|_\infty \quad \text{per ogni } \; p_n \in \mathcal{P}_n.$$

Una caratteristica dei polinomi $p_n^*(x)$ è quella di equioscillare attorno alla funzione $f(x)$. Questa proprietà definisce il polinomio di approssimazione minimax e consente di stabilirne l'unicità.

I punti

$$x_0, x_1, \ldots, x_k, \quad \text{con} \quad a \le x_0 < x_1 < \ldots < x_k \le b,$$

sono detti di *equioscillazione* per p_n se

$$f(x_i) - p_n(x_i) = (-1)^i d, \quad \text{per } i = 0, \ldots, k, \quad \text{dove} \quad |d| = \|f - p_n\|_\infty.$$

Teorema 1 *(di equioscillazione di Chebyshev). Sia $f \in C[a,b]$. Un polinomio $p_n \in \mathcal{P}_n$ è di approssimazione minimax della $f(x)$ se e solo se ha almeno $n+2$ punti di equioscillazione in $[a,b]$. Inoltre il polinomio di approssimazione minimax è unico.*

Dim.

• Sia $p_n \in \mathcal{P}_n$ un polinomio con almeno $n+2$ punti di equioscillazione (per semplicità si suppone che $d > 0$, ma la dimostrazione è perfettamente analoga nel caso opposto). Per dimostrare che $d = \delta^*$ si suppone per assurdo che esista un polinomio $q \in \mathcal{P}_n$ tale che

$$\|f - q\|_\infty < d. \tag{2}$$

Il polinomio $s(x) = p_n(x) - q(x)$ è tale che per $i = 0, \ldots, n+1$

$$s(x_i) = p_n(x_i) - q(x_i) = (f(x_i) - q(x_i)\,) - (f(x_i) - p_n(x_i)\,) = (f(x_i) - q(x_i)\,) - (-1)^i d.$$

Dalla (2), per gli indici i pari, si ha

$$s(x_i) = (f(x_i) - q(x_i)\,) - d < 0,$$

mentre per gli indici i dispari, si ha

$$s(x_i) = (f(x_i) - q(x_i)\,) + d > 0.$$

Perciò il polinomio $s(x)$ assume $n+2$ volte valori di segno opposto, e quindi ha almeno $n+1$ zeri. Ma $s(x)$ ha grado al più n, per cui $s(x)$ deve essere identicamente nullo. Quindi $q(x) \equiv p_n(x)$, e questo è assurdo per la (2). Ne segue che

$$d \le \|f - q\|_\infty \quad \text{per ogni } q \in \mathcal{P}_n,$$

e quindi $d = \delta^*$ e $p_n^*(x) \equiv p_n(x)$.

• Per dimostrare la necessità della condizione, si fa vedere che se per assurdo $r_n^*(x) = f(x) - p_n^*(x)$ assumesse il valore $\pm d$ con segno alternato in un numero k di punti $x_0, x_1, \ldots, x_{k-1}$, con $k \le n+1$, allora esisterebbe un polinomio $q(x)$ tale che $p_n^* + q \in \mathcal{P}_n$ e

$$\|f - (p_n^* + q)\,\|_\infty < \delta^*.$$

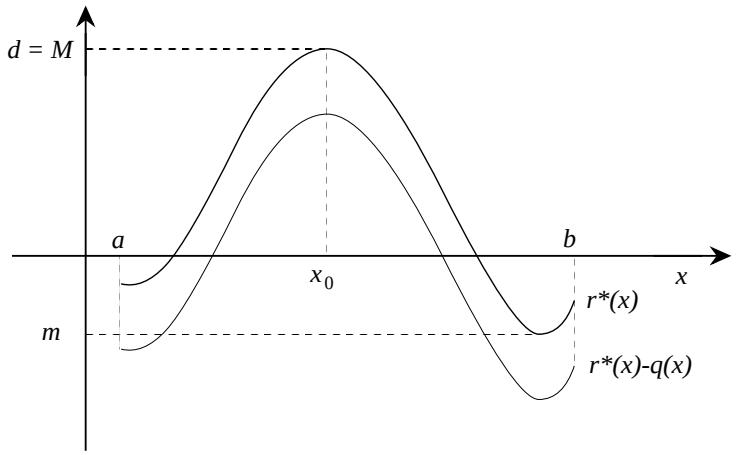


Figure 1: - Caso $k = 1$.

Se fosse $k = 1$, si considerano

$$M = \max_{x\in[a,b]} r^*(x) = d \quad \text{e} \quad m = \min_{x\in[a,b]} r^*(x), \quad \text{con} \quad |m| < M,$$

e si definisce $q(x) = (M + m)/2$ (si veda la figura 1). Quindi $p_n^* + q \in \mathcal{P}_n$. Poiché $m \le r^*(x) \le M = d$, si ha

$$-d < -\frac{M-m}{2} = m - \frac{M+m}{2} \le r^*(x) - q(x) \le M - \frac{M+m}{2} = \frac{M-m}{2} < d,$$

e quindi

$$\|f - (p_n^* + q)\|_\infty = \|r^* - q\|_\infty < d.$$

Se fosse $2 \le k \le n + 1$, per la continuità della funzione $r^*(x)$ esisterebbero $k + 1$ punti $\xi_0, \ldots, \xi_k$, tali che

$$\xi_0 = a \le x_0 < \xi_1 < x_1 < \ldots < \xi_{k-1} < x_{k-1} \le b = \xi_k,$$

in cui $r^*(\xi_i) = 0$ per $i = 1, \ldots, k - 1$. Posto

$$M_i = \max_{x\in[\xi_i,\xi_{i+1}]} r^*(x) \quad \text{e} \quad m_i = \min_{x\in[\xi_i,\xi_{i+1}]} r^*(x), \quad i = 0, 1, \ldots, k-1,$$

e

$$\mu = \min_i \frac{|M_i + m_i|}{2},$$

si consideri il polinomio di grado $k - 1$

$$s(x) = \prod_{i=1}^{k-1} (\xi_i - x),$$

e siano

$$\sigma = \max_{x\in[a,b]} |s(x)| \quad \text{e} \quad q(x) = \frac{\mu\, s(x)}{\sigma}\,.$$

Il polinomio $q(x)$ (si veda la figura 2) è tale che $|q(x)| \le \mu$ e $p_n^* + q \in \mathcal{P}_n$ per $k \le n+1$.

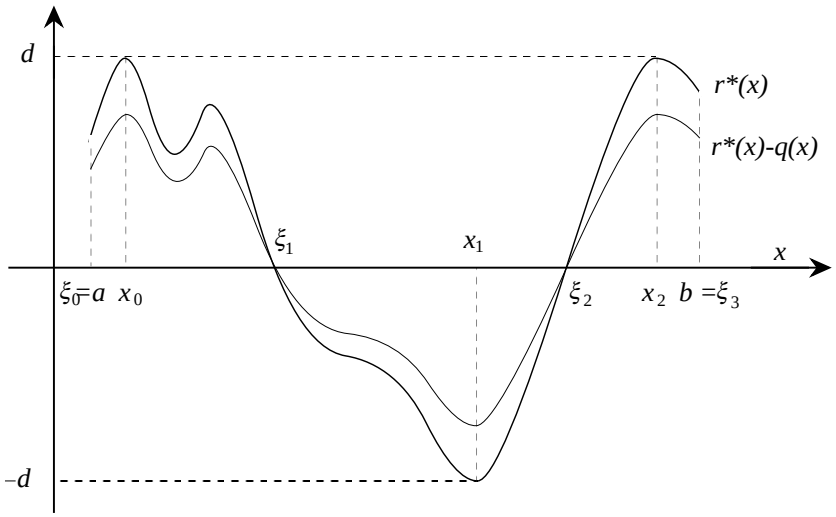


Figure 2: - Caso $k \ge 2$.

Negli intervalli $[\xi_i, \xi_{i+1}]$ con i pari si ha

$$0 \le q(x) \le \mu, \quad m_i \le r^*(x) \le M_i = d, \quad \text{con} \quad |m_i| < M_i$$

e quindi

$$-d < r^*(x) - q(x) < d,$$

in quanto $q(x) = 0$ solo nei punti in cui anche $r^*(x) = 0$, per cui

$$|r^*(x) - q(x)| = |f(x) - (p_n^*(x) + q(x))| < d. \tag{3}$$

Per gli intervalli $[\xi_i, \xi_{i+1}]$ con i dispari la (3) può essere dimostrata in modo del tutto analogo. Allora

$$\|f - (p_n^* + q)\|_\infty < d,$$

da cui l'assurdo.

• Per dimostrare l'unicità del polinomio di approssimazione minimax, si suppone per assurdo che ne esistano due $p_n(x)$ e $\widetilde{p}_n(x)$. Poiché l'insieme delle soluzioni del problema dell'approssimazione lineare è convesso, anche il polinomio $q(x) = (p_n(x) + \widetilde{p}_n(x))/2$ è di approssimazione minimax. Indicato con y uno degli $n+2$

punti di $[a,b]$ di equioscillazione per $q(x)$, è

$$\begin{aligned}\delta^* &= |f(y) - \tfrac{1}{2}\,(p_n(y) + \widetilde{p}_n(y))| \\ &\leq \frac{1}{2}\left[|f(y) - p_n(y)| + |f(y) - \widetilde{p}_n(y)|\right] \leq \frac{1}{2}\,(\delta^* + \delta^*) = \delta^*,\end{aligned}$$

per cui

$$f(y) - p_n(y) = f(y) - \widetilde{p}_n(y),$$

in quanto $|f(y) - p_n(y)|$ e $|f(y) - \widetilde{p}_n(y)|$ non possono diventare più grandi di δ^*. Perciò $p_n(x)$ e $\widetilde{p}_n(x)$ sono polinomi di grado al più n che coincidono in almeno $n+2$ punti e quindi sono identicamente uguali. □

Il teorema di equioscillazione è fondamentale per la costruzione dell'approssimazione minimax. Poiché il resto $r^*(x)$ assume $n+2$ massimi o minimi di segno opposto, esso si annulla, per il teorema di Rolle, in almeno $n+1$ punti $\xi_0, \ldots, \xi_n$, quindi il polinomio $p_n^*(x)$ e la funzione $f(x)$ assumono lo stesso valore negli $n+1$ punti ξ_i, $i = 0, 1, \ldots, n$. Perciò il polinomio $p_n^*(x)$ è il polinomio di interpolazione di grado al più n della $f(x)$ nei punti ξ_i, $i = 0, 1, \ldots, n$, che però non sono noti a priori. Ad esso si possono applicare i risultati ottenuti nella teoria dell'interpolazione: in particolare se $f \in C^{n+1}[a,b]$, allora

$$r^*(x) = (x - \xi_0) \ldots (x - \xi_n)\, \frac{f^{(n+1)}(\eta)}{(n+1)!}\,, \quad \eta \in (a,b). \tag{4}$$

Ne segue che se $f^{(n+1)}(x) \neq 0$ per $x \in (a,b)$, il resto $r^*(x)$ ha esattamente $n+2$ punti di equioscillazione, compresi gli estremi a e b.

Come esempio consideriamo il caso dell'*approssimazione minimax lineare*. Sia $f \in C^2[a,b]$ e sia $f''(x) \neq 0$ in (a,b). Il polinomio $p_1^*(x) = a_1 x + a_0$ di approssimazione minimax è tale che $r^*(x)$ assume massimo e minimo in 3 punti distinti x_0^*, x_1^*, x_2^* di $[a,b]$. Poiché $f''(x) \neq 0$ per $x \in (a,b)$, risulta

$$a = x_0^* < x_1^* < x_2^* = b,$$

e x_1^* è tale che

$$(r^*)'(x_1^*) = 0. \tag{5}$$

Poiché $r^*(x) = f(x) - a_1 x - a_0$, dalla (5) si ottiene l'equazione

$$f'(x_1^*) = a_1.$$

Imponendo poi le condizioni di equioscillazione nei tre punti a, x_1^*, b si ottengono le altre 3 equazioni

$$\begin{cases} f(a) - a_1 a - a_0 = d \\ f(x_1^*) - a_1 x_1^* - a_0 = -d \\ f(b) - a_1 b - a_0 = d, \end{cases}$$

che insieme alla (5) danno per x_1^*, a_1, a_0 e d le espressioni

$$
\begin{aligned}
f'(x_1^*) &= \frac{f(b)-f(a)}{b-a}\,, \\
a_1 &= \frac{f(b)-f(a)}{b-a}\,, \\
a_0 &= \frac{f(a)+f(x_1^*)}{2} - \frac{f(b)-f(a)}{2(b-a)}\,(a+x_1^*), \\
d &= \frac{f(a)-f(x_1^*)}{2} - \frac{f(b)-f(a)}{2(b-a)}\,(a-x_1^*).
\end{aligned}
$$

Una volta determinato dalla prima relazione il punto x_1^* in cui la $f(x)$ ha la derivata uguale al valore del rapporto incrementale dal punto a al punto b, è facile ricavare i coefficienti a_1 e a_0 del polinomio cercato. Poiché $f''(x) \neq 0$ in (a,b), il punto x_1^* esiste ed è unico nell'intervallo (a,b); la sua determinazione può non essere facile e può richiedere un metodo approssimato di risoluzione di equazioni.

Ad esempio, se $f(x) = \mathrm{e}^x$, $a = 0$, $b = 1$, il punto x_1^* è la soluzione dell'equazione

$$\mathrm{e}^x = \mathrm{e} - 1,$$

per cui

$$x_1^* = 0.54133, \quad a_1^* = 1.7183, \quad a_0^* = 0.89407, \quad d = 0.10593.$$

Il polinomio minimax di grado 1 per la funzione $f(x) = \mathrm{e}^x$, $x \in [0,1]$, è allora

$$p_1^*(x) = 1.7183\,x + 0.89407.$$

Nella figura 3 è riportato il grafico del resto $r^*(x)$.

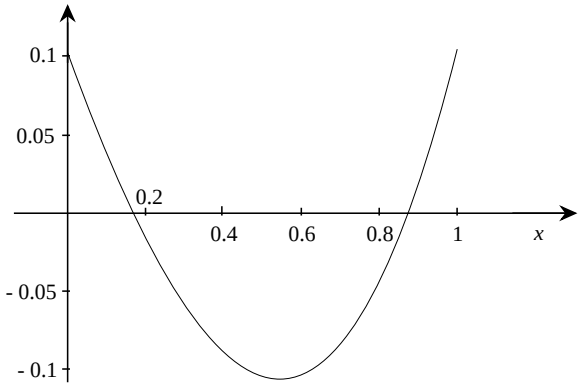


Figure 3: - Resto dell'approssimazione lineare minimax della funzione $f(x) = \mathrm{e}^x$.

Si dice che il resto $r^*(x)$ del polinomio di approssimazione minimax è una funzione *standard* quando ha esattamente $n+2$ punti, compresi gli estremi a e b dell'intervallo, di massimo o minimo locale.

Per quanto visto, se $f \in C^{n+1}[a,b]$ e $f^{(n+1)}(x) \neq 0$ in $[a,b]$, allora il resto risulta standard, come nel caso della funzione vista sopra. Se invece $f^{(n+1)}(x) = 0$ in almeno un punto $x \in [a,b]$, allora è possibile che il resto non sia standard, e questo può creare delle complicazioni nella determinazione dell'approssimazione minimax.

Consideriamo ad esempio la funzione $f(x) = x^3$, per la quale $f''(x) = 0$ in un punto dell'intervallo $[-1,1]$. Non è quindi possibile dire se il resto $r^*(x)$ dell'approssimazione minimax lineare su $[-1,1]$ è una funzione standard oppure no. Supponendo che il resto sia standard, cioè che i punti di massimo o minimo locale di $r^*(x)$ siano tali che

$$a = x_0^* < x_1^* < x_2^* = b,$$

procedendo come sopra, si ottiene il sistema non lineare

$$\begin{cases} 3(x_1^*)^2 = a_1 \\ -1 + a_1 - a_0 = d \\ (x_1^*)^3 - a_1 x_1^* - a_0 = -d \\ 1 - a_1 - a_0 = d \end{cases}$$

da cui si ricavano le due soluzioni

$$x_1^* = \frac{1}{\sqrt{3}}\,, \quad a_1 = 1, \quad a_0 = -\frac{1}{3\sqrt{3}}\,, \quad d = \frac{1}{3\sqrt{3}}\,,$$

e

$$x_1^* = -\frac{1}{\sqrt{3}}\,, \quad a_1 = 1, \quad a_0 = \frac{1}{3\sqrt{3}}\,, \quad d = -\frac{1}{3\sqrt{3}}\,.$$

Poiché il polinomio di approssimazione minimax è unico, le due soluzioni trovate non possono riferirsi ad esso. Quindi il resto non è una funzione standard. Si suppone allora che dei tre punti di massimo o minimo locale due siano interni all'intervallo, cioè ad esempio

$$a < x_0^* < x_1^* < x_2^* = b.$$

Al posto della (5) si ottengono allora le due equazioni

$$3(x_0^*)^2 = a_1 \quad \text{e} \quad 3(x_1^*)^2 = a_1,$$

da cui segue che $x_0^* = -x_1^*$. Imponendo le condizioni di equioscillazione si hanno le tre equazioni

$$\begin{cases} (x_0^*)^3 - a_1 x_0^* - a_0 = d \\ (x_1^*)^3 - a_1 x_1^* - a_0 = -d \\ 1 - a_1 - a_0 = d. \end{cases}$$

Dalla prima e seconda equazione segue che $a_0 = 0$. Per sostituzione risulta che x_1^* è la soluzione dell'equazione

$$2x^3 + 3x^2 - 1 = 0$$

che appartiene all'intervallo $[0,1]$, cioè

$$x_1^* = \frac{1}{2}\,,$$

per cui $a_1^* = \frac{3}{4}$, $a_0^* = 0$, $x_0^* = -\frac{1}{2}$, $d = \frac{1}{4}$, e quindi

$$p_1^*(x) = \frac{3}{4}\,x.$$

È inoltre facile verificare che $r^*(-1) = -d$, cioè i punti di massimo o minimo locale in questo esempio sono 4, anziché 3, due punti interni e due estremi dell'intervallo, come risulta dalla figura 4.

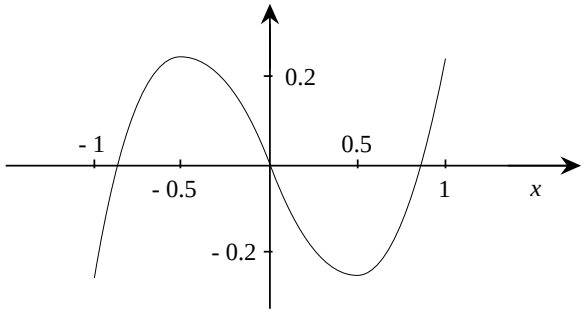


Figure 4: - Resto dell'approssimazione lineare minimax della funzione $f(x) = x^3$.

Volendo invece determinare il polinomio di approssimazione minimax di grado al più 2 per la stessa funzione $f(x) = x^3$ nello stesso intervallo $[-1,1]$, non vi sono difficoltà, perché $f'''(x) \neq 0$ in tutto l'intervallo, e quindi il resto è una funzione standard. D'altra parte il polinomio di primo grado sopra determinato è tale che $p_1^* \in \mathcal{P}_2$, ha 4 punti di oscillazione compresi gli estremi. Per l'unicità del polinomio di migliore approssimazione è $p_2^*(x) = p_1^*(x)$.

Perciò, a parte il caso in cui $f^{(n+1)}(x) \neq 0$ in tutto $[a,b]$, non è possibile dire a priori se il resto sarà una funzione standard. Il seguente teorema assicura la convergenza alla $f(x)$ del polinomio di approssimazione minimax al crescere del grado n.

Teorema 2 *Sia $f \in C[a,b]$. La successione $\{\delta_n^*\}_{n\in\mathbb{N}}$ è monotona non crescente e*

$$\lim_{n\to\infty} \delta_n^* = 0.$$

Dim. Sappiamo già che la successione δ_n^* è monotona non crescente. Per il teorema di Weierstrass si ha che, comunque si fissi una costante $\epsilon > 0$, esistono un intero m e un polinomio $q_m(x)$ di grado al più m, tale che

$$|f(x) - q_m(x)| \leq \epsilon, \qquad \text{per ogni } \; x \in [a,b].$$

Si ha quindi

$$\delta_m^* = \|f - p_m^*\|_\infty \leq \|f - q_m\|_\infty \leq \epsilon.$$

Per la monotonia della successione δ_n^*, ne segue che per ogni $\epsilon > 0$, esiste un intero m tale che per ogni $n \geq m$ è $\delta_n^* \leq \epsilon$. □

La proprietà di convergenza illustrata nel teorema 2 non dà però alcuna informazione sulla velocità con cui la successione dei δ_n^* tende a zero.

Nel caso della funzione $f(x) = \arcsin x$ e $[a, b] = [-1, 1]$, si può verificare che δ_n^* converge più lentamente di $1/n^2$, per cui per ottenere un'approssimazione dell'ordine di 10^{-6} si deve determinare un polinomio di grado maggiore di 10^3. In questo caso quindi, per la lentezza della convergenza della successione dei δ_n^*, non conviene approssimare la funzione in questo modo. Se la $f(x)$ ha una maggiore regolarità, la successione dei δ_n^* converge assai più rapidamente, come risulta dal seguente teorema.

Teorema 3 *(di Jackson). Se $f \in C^k[a, b]$ e se $n > k$, esiste una costante γ, indipendente da n, tale che*

$$\delta_n^* \leq \frac{\gamma \|f^{(k)}\|_\infty}{n^k}.$$

Se $f \in C^\infty[a, b]$ la successione dei δ_n^* converge più rapidamente della successione $\dfrac{1}{n^k}$ per ogni $k \geq 1$.

È possibile dare una limitazione inferiore e superiore di δ_n^* nel caso in cui la $f^{(n+1)}(x)$ non cambi segno in $[a, b]$, sfruttando la seguente proprietà di minimo dei polinomi di Chebyshev.

Lemma 4 *Sia $T_n(x)$, $x \in [-1, 1]$, l'n-esimo polinomio ortogonale di Chebyshev di 1ª specie. Si considerano i polinomi $t_0(x) = 1$ e $t_n(x) = T_n(x)/2^{n-1}$ per $n \geq 1$.*

- *$t_n(x)$ è monico di grado n e $\|t_n\|_\infty = 1/2^{n-1}$ per $n \geq 1$.*
- *Fra tutti i polinomi monici di grado n, $t_n(x)$ è quello che ha la minima norma ∞ sull'intervallo $[-1, 1]$.*

Teorema 5 *Sia $f \in C^{n+1}[a, b]$, con $f^{(n+1)}(x) \neq 0$ in $[a, b]$, e siano m e M, con $0 < m < M$, due costanti tali che*

$$m \leq |f^{(n+1)}(x)| \leq M, \quad \text{per } x \in [a, b],$$

allora

$$\frac{m(b-a)^{n+1}}{2^{2n+1}(n+1)!} \leq \delta_n^* \leq \frac{M(b-a)^{n+1}}{2^{2n+1}(n+1)!}.$$

Dim. Si considera dapprima il caso in cui $[a, b] = [-1, 1]$. Dalla (4) si ha che

$$r_n^*(x) = s(x)\, \frac{f^{(n+1)}(\eta)}{(n+1)!}, \quad s(x) = (x - \xi_0) \ldots (x - \xi_n), \quad \eta \in (-1, 1).$$

Il polinomio $s(x)$ è monico di grado $n+1$, quindi per il lemma 4 è $\|s(x)\|_\infty \geq 1/2^n$. Quindi

$$\frac{m}{2^n(n+1)!} \leq \delta_n^*.$$

Per la limitazione superiore si nota che la funzione

$$s(x) = \frac{r_n^*(x)\,(n+1)!}{f^{(n+1)}(\eta)}$$

ha le stesse oscillazioni di segno di $r_n^*(x)$, cioè ha $n+2$ punti distinti di oscillazione di segno. Inoltre

$$|s(x)| = \Big|\frac{r_n^*(x)\,(n+1)!}{f^{(n+1)}(\eta)}\Big| \geq \Big|\frac{r_n^*(x)\,(n+1)!}{M}\Big|.$$

Supponendo per assurdo che

$$\delta^* > \frac{M}{2^n\,(n+1)!},$$

ne seguirebbe che

$$\|s(x)\| > \frac{1}{2^n},$$

cioè il polinomio $s(x)$ avrebbe in $[-1,1]$ $\;n+2$ punti distinti di oscillazione, in cui assumerebbe valori di segno alterno e modulo maggiore di $1/2^n$. Quindi il polinomio $s(x) - t_{n+1}(x)$ di grado n manterrebbe $n+2$ punti distinti di oscillazione di segno, e quindi dovrebbe annullarsi in $n+1$ punti distinti, il che è assurdo. La tesi risulta così dimostrata nel caso che $a=-1$ e $b=1$.

Se l'intervallo non fosse $[-1,1]$, si trasforma $[a,b]$ in $[-1,1]$, ponendo

$$x = \frac{b-a}{2}\,y + \frac{a+b}{2}.$$

Per la funzione $g(y) = f(x(y))$ si ha

$$m\left(\frac{b-a}{2}\right)^{n+1} \leq g^{(n+1)}(y) \leq M\left(\frac{b-a}{2}\right)^{n+1},$$

quindi il ragionamento fatto sopra si può ripetere per la $g(y)$ tenendo conto del fattore $((b-a)/2)^{n+1}$. □

Il teorema 5, consente di determinare con sufficiente precisione il grado del polinomio che fornisce l'approssimazione minimax con accuratezza prefissata.

Si applica ad esempio il teorema 5 per determinare il grado del polinomio di approssimazione minimax della funzione $f(x) = \mathrm{e}^x$ per $x \in [0,1]$, tale che $\delta_n^* \leq 0.5\ 10^{-6}$. Poiché $m=1$, $M = \mathrm{e}$, risulta

$$\frac{m(b-a)^{n+1}}{2^{2n+1}(n+1)!} \approx 0.678\ 10^{-6} \quad \text{per} \quad n=5,$$
$$\frac{M(b-a)^{n+1}}{2^{2n+1}(n+1)!} = \frac{\mathrm{e}}{2^{2n+1}(n+1)!} \leq 0.5\ 10^{-6} \quad \text{per} \quad n=6.$$

Occorre quindi fissare $n = 6$ e risulta

$$0.242\ 10^{-7} \leq \delta_6^* \leq 0.658\ 10^{-7}.$$

Dal teorema 5 risulta che per un fissato n, al diminuire dell'ampiezza dell'intervallo, segue una drastica riduzione dell'errore. Quindi per ottenere polinomi di approssimazione di grado basso occorre ridurre l'intervallo. Ad esempio, per ottenere il polinomio di approssimazione minimax della funzione $f(x) = \sin x$ tale che $\delta_n^* \leq 0.5\ 10^{-6}$, deve essere $n = 7$ se $[a, b] = [0, \pi/2]$ e $n = 5$ se $[a, b] = [0, \pi/4]$.

3 Algoritmo di Remez

La determinazione effettiva di un polinomio di approssimazione minimax potrebbe in via teorica essere ottenuta con un procedimento analogo a quello visto per il caso lineare. Si dovrebbero però risolvere sistemi di complicate equazioni non lineari, che richiederebbero, salvo casi particolari, l'utilizzazione di metodi iterativi assai onerosi. D'altra parte, a differenza di quanto accade per i polinomi di approssimazione ai minimi quadrati, non esiste alcuna procedura diretta che fornisca i coefficienti del polinomio di approssimazione minimax e non vi è alcuna relazione fra i coefficienti dei polinomi minimax $p_n^*(x)$ e $p_{n+1}^*(x)$. Sono stati perciò studiati specifici metodi iterativi: uno di tali metodi, che va sotto il nome di *(secondo) algoritmo di Remez*, costruisce, a partire da un vettore iniziale $\boldsymbol{x}^{(0)}$ di $n+2$ componenti distinte, una successione di vettori $\{\boldsymbol{x}^{(k)}\}_k$, $k \geq 1$, che converge al vettore dei punti di equioscillazione (la convergenza verrà studiata nel prossimo paragrafo). In teoria il vettore $\boldsymbol{x}^{(0)}$ può essere arbitrario, in pratica conviene sceglierlo opportunamente.

- Sia $\boldsymbol{x}^{(k)} = (x_0^{(k)}, x_1^{(k)}, \ldots, x_{n+1}^{(k)})$ un vettore di $n+2$ componenti, tale che

$$a \leq x_0^{(k)} < x_1^{(k)} < \ldots < x_{n+1}^{(k)} \leq b$$

(se il resto è standard, conviene porre subito $x_0^{(k)} = a$ e $x_{n+1}^{(k)} = b$). Sfruttando le tecniche di costruzione dei polinomi di interpolazione, si calcolano i polinomi $q^{(k)}$ e $s^{(k)} \in \mathcal{P}_{n+1}$ tali che

$$q^{(k)}(x_i^{(k)}) = f(x_i^{(k)}), \quad s^{(k)}(x_i^{(k)}) = (-1)^i, \quad \text{per } i = 0, \ldots, n+1.$$

Se il polinomio $q^{(k)}(x)$ fosse di grado minore di $n+1$, sarebbe necessario scegliere punti $x_i^{(0)}$ diversi. Indicato con $d^{(k)}$ il rapporto fra i coefficienti di grado massimo di $q^{(k)}(x)$ e di $s^{(k)}(x)$, si definisce

$$p_n^{(k)}(x) = q^{(k)}(x) - d^{(k)} s^{(k)}(x).$$

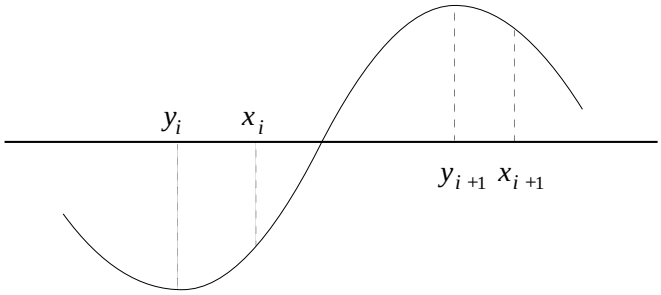
Quindi $p_n^{(k)} \in \mathcal{P}_n$ ed è tale che

$$f(x_i^{(k)}) - p_n^{(k)}(x_i^{(k)}) = (-1)^i d^{(k)},$$

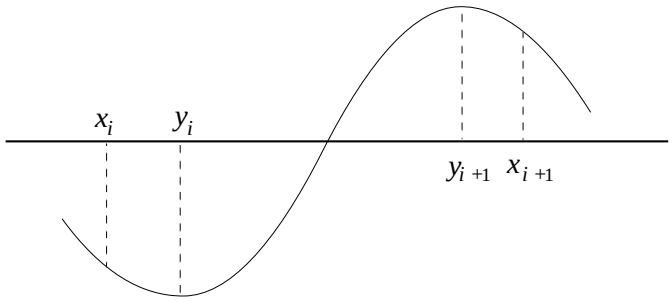
perciò il polinomio $p_n^{(k)}(x)$ oscilla almeno $n+2$ volte in $[a,b]$. Se i punti $x_i^{(k)}$, $i = 0,\ldots,n+1$ fossero tutti punti di massimo o di minimo di $r^{(k)}(x) = f(x) - p_n^{(k)}(x)$, allora $p_n^{(k)}(x)$ sarebbe il polinomio di migliore approssimazione. Naturalmente questo in generale non sarà vero, per cui è possibile determinare $n+2$ punti di massimo o minimo locale

$$a \leq y_0 < y_1 < \ldots < y_{n+1} \leq b$$

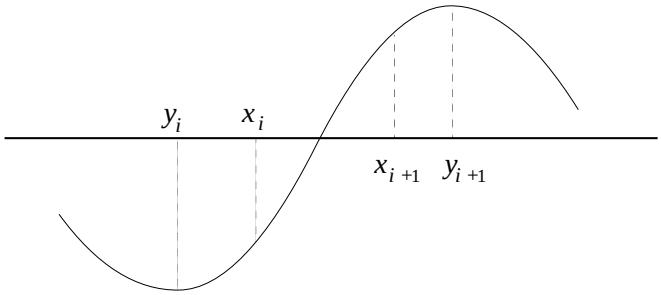
di $r^{(k)}(x)$ in $[a,b]$, tali che $r^{(k)}(y_i)$ abbia lo stesso segno di $r^{(k)}(x_i^{(k)})$. Questa condizione è fondamentale perché garantisce che il resto nei punti y_i , $i = 0,1,\ldots,n+1$, abbia ancora segno alternato. Le seguenti figure illustrano il procedimento di scelta dei punti y_i (per semplicità si è omesso l'indice k).
1^o caso: vi è un solo punto di massimo o minimo di $r(x)$ compreso fra x_i e x_{i+1}.



2^o caso: vi sono due punti di massimo o minimo di $r(x)$ compresi fra x_i e x_{i+1}.



3^o caso: non vi sono punti di massimo o minimo di $r(x)$ compresi fra x_i e x_{i+1}.



Se fra x_i e x_{i+1} vi sono più di due punti di massimo o minimo, si scelgono i punti

corrispondenti a massimi più alti o a minimi più bassi. I punti estremi $x_0 = a$ e $x_{n+1} = b$ possono essere sostituiti da $y_0 \neq a$ e $y_{n+1} \neq b$ solo nel caso che il resto non sia standard. Si ottiene così il vettore

$$\boldsymbol{x}^{(k+1)} = (y_0, y_1, \ldots, y_{n+1}).$$

La determinazione dei punti y_i di massimo o di minimo di $r^{(k)}(x)$ costituisce proprio il principale problema del metodo di Remez. Una possibile via da seguire è quella di approssimare gli zeri della derivata di $r^{(k)}(x)$ se questa non è una funzione troppo complicata, applicando un metodo iterativo a convergenza garantita, come il metodo di bisezione o quello di falsa posizione. Questo modo di procedere è facilitato dal fatto che dopo le prime iterazioni i punti y_i non potranno trovarsi molto distanti dai punti $x_i^{(k)}$.

- Come risulta dal prossimo teorema 8 di convergenza, il vettore iniziale $\boldsymbol{x}^{(0)}$ può essere arbitrario, purché risulti $d^{(0)} \neq 0$. Si possono scegliere $n+2$ punti equidistanti nell'intervallo, con $x_0^{(0)} = a$ e $x_{n+1}^{(0)} = b$, ma come si vedrà successivamente, una scelta migliore dei punti $x_i^{(0)}$ è data da

$$x_i^{(0)} = \frac{b-a}{2} \cos\frac{(n+1-i)\pi}{n+1} + \frac{a+b}{2}, \quad i = 0, \ldots, n+1.$$

Questi punti $x_i^{(0)}$ sono, in ordine crescente, i punti di massimo o minimo del polinomio di Chebyshev T_{n+1}, definito sull'intervallo $[a, b]$, estremi compresi.

- Poiché i punti y_i sono di massimo o di minimo locale di $r^{(k)}(x)$, ne segue che

$$|r^{(k)}(y_i)| \geq |d^{(k)}|, \quad i = 0, \ldots, n+1.$$

Posto

$$m^{(k)} = \min_{i=0,\ldots,n+1} |r^{(k)}(y_i)| \quad \text{e} \quad M^{(k)} = \max_{i=0,\ldots,n+1} |r^{(k)}(y_i)|,$$

si ottiene come criterio di arresto per l'iterazione che il rapporto $M^{(k)}/m^{(k)}$ sia sufficientemente vicino a 1, cioè

$$\left|\frac{M^{(k)}}{m^{(k)}} - 1\right| < \epsilon,$$

dove ϵ è una quantità piccola prefissata. Si possono utilizzare altri criteri di arresto, sfruttando ad esempio la differenza fra i vettori $\boldsymbol{x}^{(k)}$ e $\boldsymbol{x}^{(k+1)}$ oppure fra i coefficienti di due successivi polinomi $p_n^{(k)}(x)$ calcolati. È opportuno comunque fissare un numero massimo di iterazioni.

- Per la costruzione del polinomio $p_n^{(k)}(x)$, anzichè passare attraverso il procedimento di interpolazione, si può risolvere un sistema lineare, utilizzando una base

qualsiasi dello spazio dei polinomi. In particolare si potrebbe scegliere la base dei monomi x^j, $j = 0, \ldots, n$, scrivendo

$$p_n^{(k)}(x) = \sum_{j=0}^{n} a_j^{(k)} x^j,$$

o la base dei polinomi di Chebyshev di 1^a specie scrivendo

$$p_n^{(k)}(x) = \sum_{j=0}^{n}{}' \beta_j^{(k)} T_j(x),$$

dove l'apice vicino alla sommatoria indica che il primo termine deve essere dimezzato. Nel primo caso i coefficienti $a_j^{(k)}$ per $j = 0, \ldots, n$ e $d^{(k)}$ vengono determinati risolvendo il sistema

$$\sum_{j=0}^{n} a_j^{(k)} (x_i^{(k)})^j + (-1)^i d^{(k)} = f(x_i^{(k)}), \quad i = 0, \ldots, n+1, \tag{6}$$

nel secondo caso i coefficienti $\beta_j^{(k)}$ per $j = 0, \ldots, n$ e $d^{(k)}$ vengono determinati risolvendo il sistema

$$\sum_{j=0}^{n}{}' \beta_j^{(k)} T_j(x_i^{(k)}) + (-1)^i d^{(k)} = f(x_i^{(k)}), \quad i = 0, \ldots, n+1. \tag{7}$$

La matrice del sistema (6) è, a meno di una colonna, una matrice di Vandermonde (ved. lemma 6) che per certe scelte degli $x_i^{(k)}$ può essere malcondizionata, mentre la matrice del sistema (7) è solitamente meglio condizionata. In ogni caso la soluzione è unica.

Si vuole ad esempio determinare con il metodo di Remez il polinomio di grado al più 3 di approssimazione minimax per la funzione $f(x) = \mathrm{e}^x$ nell'intervallo $[0, 1]$. Dal teorema 5 si ha che

$$0.325\ 10^{-3} \leq \delta_3^* \leq 0.885\ 10^{-3}.$$

Si considerano inizialmente i punti

$$\begin{aligned}
x_0^{(0)} &= \frac{1}{2}(1 + \cos\pi) = 0, \quad x_1^{(0)} = \frac{1}{2}(1 + \cos\frac{3}{4}\pi) = 0.14645, \\
x_2^{(0)} &= \frac{1}{2}(1 + \cos\frac{1}{2}\pi) = 0.5, \quad x_3^{(0)} = \frac{1}{2}(1 + \cos\frac{1}{4}\pi) = 0.85355, \\
x_4^{(0)} &= \frac{1}{2}(1 + \cos 0) = 1.
\end{aligned}$$

Poiché $f^{(4)}(x) \neq 0$ in $[0, 1]$, il resto risulta una funzione standard: perciò si assumerà in ogni iterazione $x_0^{(k)} = 0$ e $x_4^{(k)} = 1$. Al primo passo si ottiene

$$a_3^{(0)} = 0.27998,\ a_2^{(0)} = 0.42172,\ a_1^{(0)} = 1.0166,\ a_0^{(0)} = 0.99946$$

e $d^{(0)} = 0.54344\ 10^{-3}$, da cui risulta

$$r^{(0)}(x) = \mathrm{e}^x - 0.27998x^3 - 0.42172x^2 - 1.0166x - 0.99946.$$

I punti di massimo o minimo di $r^{(0)}(x)$ sono

$$y_1 = 0.15258,\ y_2 = 0.51245,\ y_3 = 0.85987,$$

e si ottiene

$$\left|\frac{M^{(0)}}{m^{(0)}} - 1\right| \approx 0.497\ 10^{-2}.$$

Assumendo $x_1^{(1)} = y_1$, $x_2^{(1)} = y_2$, $x_3^{(1)} = y_3$, e ripetendo il calcolo si ottiene

$$a_3^{(1)} = 0.27998,\ a_2^{(1)} = 0.42170,\ a_1^{(1)} = 1.0166,\ a_0^{(1)} = 0.99946$$

e $d^{(1)} = 0.54479\ 10^{-3}$. I punti di massimo o minimo di $r^{(1)}(x)$ sono

$$y_1 = 0.15270,\ y_2 = 0.51247,\ y_3 = 0.85977,$$

e si ottiene

$$\left|\frac{M^{(1)}}{m^{(1)}} - 1\right| \approx 0.833\ 10^{-6}.$$

Le successive iterazioni non modificano ulteriormente le prime 5 cifre dei coefficienti ottenuti. Il polinomio $p_3^*(x)$ di approssimazione minimax di e^x nell'intervallo $[0, 1]$ è perciò

$$p_3^*(x) = 0.27998x^3 + 0.42170x^2 + 1.0166x + 0.99946,$$

e risulta $\delta_3^* = 0.545\ 10^{-3}$. Nella figura 5 è riportato il grafico del resto $f(x) - p_3^*(x)$.

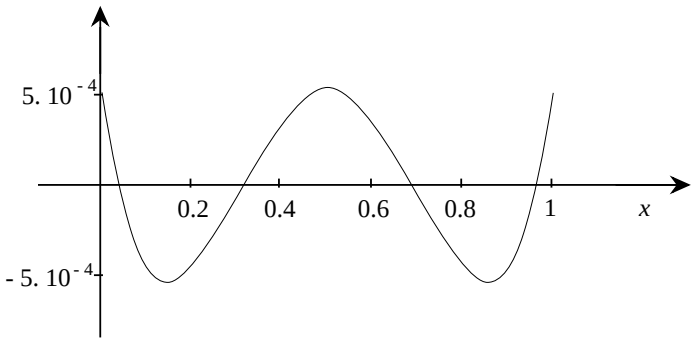


Figure 5: - Resto dell'approssimazione minimax di terzo grado della funzione $f(x) = \mathrm{e}^x$.

4 Convergenza del metodo di Remez

La dimostrazione qui riportata della convergenza del metodo di Remez sfrutta la rappresentazione dei polinomi in termini della base dei monomi. È opportuno premettere alcuni lemmi di carattere generale.

Lemma 6 *Dati $k+1$ punti $x_0, x_1, \ldots, x_k$, la matrice $V^{(k)}$ i cui elementi sono*

$$v_{i,j} = x_i^j, \quad i,\ j = 0, \ldots, k,$$

è detta matrice di Vandermonde e vale

$$\det V^{(k)} = \prod_{\substack{i,j=0,k \\ j>i}} (x_j - x_i). \tag{8}$$

Quindi $\det V^{(k)} \neq 0$ *se e solo se i numeri x_i sono a due a due distinti.*

Dim: si sfrutta la relazione ricorrente

$$\det V^{(k)} = (x_k - x_0)(x_k - x_1) \cdots (x_k - x_{k-1}) \det V^{(k-1)}, \quad \det V^{(0)} = 1,$$

che si ottiene sottraendo dalla j-esima colonna di $V^{(k)}$ la $(j-1)$-esima moltiplicata per x_k, per $j = 2, \ldots, k+1$. □

Lemma 7 *Sia $f(x) \in C[a,b]$. Dati $n+2$ punti distinti in $[a,b]$ con*

$$a \leq x_0 < x_1 < \ldots < x_{n+1} \leq b,$$

- *il sistema*

$$\sum_{i=0}^{n+1} x_i^j\, c_i = 0, \quad per \quad j = 0, \ldots, n, \tag{9}$$

$$\sum_{i=0}^{n+1} |c_i| = 1, \quad con \quad c_0 > 0, \tag{10}$$

ha una e una sola soluzione, le cui componenti c_i sono non nulle e hanno segni alterni, cioè

$$c_i = (-1)^i\, |c_i|, \quad per \quad i = 0, \ldots, n+1; \tag{11}$$

- *il sistema*

$$\sum_{j=0}^{n} x_i^j\, a_j + (-1)^i d = f(x_i), \quad per \quad i = 0, \ldots, n+1, \tag{12}$$

ha una e una sola soluzione, la cui ultima componente d è tale che

$$d = \sum_{i=0}^{n+1} c_i f(x_i) = \sum_{i=0}^{n+1} c_i [f(x_i) - p(x_i)], \tag{13}$$

qualunque sia il polinomio $p \in \mathcal{P}_n$.

Dim: • La matrice del sistema (9) di dimensioni $(n+1) \times (n+2)$ ha rango massimo in quanto i suoi minori di ordine $(n+1)$ sono trasposti di matrici di Vandermonde costruite con punti distinti. Quindi il sistema (9) ha infinite soluzioni non nulle, che dipendono da un parametro moltiplicativo. Assumendo c_0 come tale parametro e imponendo che $c_0 > 0$, dalla condizione (10) segue che la $(n+2)$-upla c_i, $i = 0, \ldots, n+1$, è unica. Si scrive il sistema (9) nella forma

$$\sum_{i=1}^{n+1} x_i^j \, c_i = -x_0^j \, c_0, \quad j = 0, \ldots, n,$$

e si utilizza il metodo di Cramer, usando la (8) per i determinanti coinvolti. Si ottiene

$$c_i = -c_0 \prod_{\substack{j=1 \\ j \neq i}}^{n} \frac{x_0 - x_j}{x_i - x_j}. \tag{14}$$

Poiché i nodi sono tutti distinti, è $c_i \neq 0$ per ogni i. Inoltre c_i ha $2n - i - 1$ fattori negativi, quindi c_i e c_{i+1} hanno segni opposti.

Dalla (9) segue poi che per ogni $p \in \mathcal{P}_n$ è

$$\sum_{i=0}^{n+1} c_i \, p(x_i) = 0. \tag{15}$$

• Per il sistema (12) si nota che l'ultima colonna della matrice del sistema è formata da elementi che sono alternativamente uguali a $+1$ e a -1. Pertanto sviluppando il determinante rispetto a questa colonna, esso risulta uguale in modulo alla somma di determinanti di matrici di Vandermonde, che per la (8) sono positivi in quanto $x_j > x_i$ per $j > i$. Ne segue che il determinante del sistema è non nullo e quindi il sistema ha una e una sola soluzione. Posto

$$s(x) = \sum_{j=0}^{n} a_j x^j,$$

per la (12) è

$$f(x_i) - s(x_i) = (-1)^i \, d,$$

e da (11) e da (10) segue che

$$\sum_{i=0}^{n+1} c_i [f(x_i) - s(x_i)] = \sum_{i=0}^{n+1} (-1)^i \, c_i d = \sum_{i=0}^{n+1} |c_i| d = d.$$

Poiché $s(x)$ è un polinomio di grado n, per la (15) esso può essere sostituito con qualunque $p \in \mathcal{P}_n$. Si ottiene così la (13). □

Il seguente teorema assicura, sotto l'ipotesi di continuità della $f(x)$, la convergenza del metodo di Remez per ogni scelta del punto iniziale $\boldsymbol{x}^{(0)}$.

Teorema 8 *Sia $f(x) \in C[a,b]$ e sia $\{\boldsymbol{x}^{(k)}\}$ la successione di vettori generata applicando l'algoritmo di Remez ad un vettore $\boldsymbol{x}^{(0)}$ arbitrario. Se $d^{(0)} \neq 0$, allora la successione $\{p_n^{(k)}(x)\}$ converge uniformemente su $[a,b]$ a $p_n^*(x)$ per $k \to \infty$.*

Dim: • Si dimostra dapprima che se per ogni k è

$$|d^{(k)}| \geq \delta > 0, \tag{16}$$

i punti $x_i^{(k)}$ restano sempre distinti, cioè tali che

$$|x_{i+1}^{(k)} - x_i^{(k)}| \geq \xi, \quad \text{per ogni } i, \tag{17}$$

dove $\xi > 0$ non dipende da k. Infatti, se per assurdo ciò non fosse, esisterebbe un indice j, $0 \leq j \leq n$, per cui

$$\min \lim_{k\to\infty} (x_{j+1}^{(k)} - x_j^{(k)}) = 0.$$

Considerando il polinomio $q_n \in \mathcal{P}_n$, tale che

$$q_n(x_i^{(k)}) = f(x_i^{(k)}), \quad \text{per} \quad i = 0, \ldots, n+1, \; i \neq j,$$

e applicando il lemma 7 al vettore $\boldsymbol{x}^{(k)} = [x_0^{(k)}, \ldots, x_{n+1}^{(k)}]$ degli $n+2$ punti ottenuti al k-esimo passo del metodo di Remez, per la (13) si avrebbe

$$\begin{aligned} |d^{(k)}| &= \Big| \sum_{i=0}^{n+1} c_i [f(x_i^{(k)}) - q_n(x_i^{(k)})] \Big| = |c_j| \; |f(x_j^{(k)}) - q_n(x_j^{(k)})| \\ &\leq |c_j| \Big[|f(x_j^{(k)}) - f(x_{j+1}^{(k)})| + |q_n(x_{j+1}^{(k)}) - q_n(x_j^{(k)})| \Big]. \end{aligned}$$

Per continuità, per ogni ϵ esisterebbe un k per cui

$$|f(x_j^{(k)}) - f(x_{j+1}^{(k)})| \leq \epsilon \quad \text{e} \quad |q_n(x_{j+1}^{(k)}) - q_n(x_j^{(k)})| \leq \epsilon,$$

e quindi, poiché $|c_j| < 1$, risulterebbe

$$|d^{(k)}| \leq 2\epsilon |c_j| < 2\epsilon$$

in contrasto con la (16). Ne segue che la (17) vale, e per la (14) esiste $\gamma > 0$ tale che $|c_i| \geq \gamma$ per ogni i e k.

• Si dimostra poi che $|d^{(k+1)}| \geq |d^{(k)}|$, per ogni k. Alla $(k+1)$-esima iterazione, poiché $x_i^{(k+1)} = y_i$, $i = 0, \ldots, n+1$, per la (13) è

$$d^{(k+1)} = \sum_{i=0}^{n+1} c_i \, [f(y_i) - p_n^{(k)}(y_i)] = \sum_{i=0}^{n+1} c_i \, r^{(k)}(y_i) = \sum_{i=0}^{n+1} c_i \, |r^{(k)}(y_i)| \operatorname{sgn} r^{(k)}(y_i).$$

Sia i c_i per la (11) che gli $r^{(k)}(y_i)$ hanno segno alternato, quindi i fattori $c_i \operatorname{sgn} r^{(k)}(y_i)$ hanno tutti lo stesso segno

$$|d^{(k+1)}| = \sum_{i=0}^{n+1} |c_i| \, |r^{(k)}(y_i)|. \tag{18}$$

Poiché

$$|r^{(k)}(y_i)| \geq |r^{(k)}(x_i^{(k)})| = |d^{(k)}|, \quad \text{per} \quad i = 0, \ldots, n+1,$$

risulta $|d^{(k+1)}| \geq |d^{(k)}|$ e quindi $|d^{(k)}| \geq |d^{(0)}| > 0$ per ogni k. Inoltre è $|d^{(k)}| \leq \delta^*$, per ogni k, cioè la successione $\{|d^{(k)}|\}$ è monotona non decrescente, limitata da δ^*, e quindi convergente. Posto

$$|d^{(k+1)}| - |d^{(k)}| = \epsilon^{(k)} \geq 0, \tag{19}$$

risulta

$$\lim_{k \to \infty} \epsilon^{(k)} = 0.$$

Dalla (18) si ha

$$|d^{(k+1)}| - |d^{(k)}| = \sum_{i=0}^{n+1} |c_i| \, |r^{(k)}(y_i)| - |d^{(k)}|,$$

e poiché

$$|d^{(k)}| = \sum_{i=0}^{n+1} |c_i| \, |d^{(k)}|,$$

si ha

$$|d^{(k+1)}| - |d^{(k)}| = \sum_{i=0}^{n+1} |c_i| \left[|r^{(k)}(y_i)| - |d^{(k)}| \right].$$

Poiché i punti y_i sono tutti punti di massimo o minimo, esiste un indice j (dipendente da k) tale che

$$|r^{(k)}(y_j)| = \|r^{(k)}\|_\infty,$$

per cui

$$|d^{(k+1)}| - |d^{(k)}| \geq |c_j| \left[\, \|r^{(k)}\|_\infty - |d^{(k)}| \, \right],$$

e dalla (19) segue

$$\|r^{(k)}\|_\infty \leq |d^{(k)}| + \frac{\epsilon^{(k)}}{|c_j|} \leq \delta^* + \frac{\epsilon^{(k)}}{|c_j|}.$$

D'altra parte

$$\delta^* \leq \|r^{(k)}\|_\infty,$$

da cui

$$\delta^* \leq \|f - p_n^{(k)}\|_\infty \leq \delta^* + \frac{\epsilon^{(k)}}{|c_j|}.$$

Poiché i c_i sono limitati inferiormente in modulo, per ogni ϵ esiste un k_0 tale che per $k \geq k_0$ è

$$\delta^* \leq \|f - p_n^{(k)}\|_\infty \leq \delta^* + \epsilon. \tag{20}$$

- Finalmente si dimostra che

$$\lim_{k\to\infty} \|p_n^{(k)} - p_n^*\|_\infty = 0.$$

Infatti, se ciò non fosse, esisterebbe una sottosuccessione di polinomi $\{p_n^{(k_m)}(x)\}$ di $\{p_n^{(k)}(x)\}$ tale che

$$\|p_n^{(k_m)} - p_n^*\|_\infty \geq M, \quad M > 0. \tag{21}$$

Per il teorema di Bolzano-Weierstrass, la successione $\{p_n^{(k_m)}(x)\}$ ammette una sottosuccessione uniformemente convergente $\{p_n^{(k_q)}(x)\}$. Indicato con $q(x)$ il limite di tale sottosuccessione, si ha dalla (20) che

$$\|f - q\|_\infty = \delta^*.$$

Ma dalla (21) segue che $q(x) \neq p_n^*(x)$, ciò che contraddice la proprietà di unicità del polinomio di migliore approssimazione. □

Il prossimo teorema mostra che sotto ipotesi assai generali la convergenza del metodo di Remez è molto rapida, addirittura quadratica.

Teorema 9 *Sia $f \in C^2[a,b]$. Se $r^*(x)$ è standard e tale che $(r^*)''(x_i^*) \neq 0$ nei punti x_i^*, $i = 0, \ldots, n+1$ di equioscillazione, posto $\delta^{(k)} = \|r^{(k)}\|_\infty$, esiste una costante $\gamma \neq 0$, tale che*

$$\frac{\delta^{(k+1)} - \delta^*}{(\,\delta^{(k)} - \delta^*)^2} \leq \gamma, \quad \textit{per} \quad k = 1, 2, \ldots$$

Dim: Poiché $r^*(x)$ è standard, si assume $x_0^* = a$ e $x_{n+1}^* = b$. I punti x_i^*, $i = 1, \ldots, n$, sono stazionari per

$$r^*(x) = f(x) - \sum_{j=0}^{n} a_j^*\, x^j,$$

quindi, posto $\boldsymbol{a} = (a_1, \ldots, a_n)$, si ha

$$f'(x_i^*) - q(\boldsymbol{a}^*, x_i^*) = 0, \quad \text{dove} \quad q(\boldsymbol{a}, x_i) = \sum_{j=1}^{n} a_j\; j\; x_i^{j-1}.$$

Poiché $(r^*)''(x_i^*) \neq 0$, dal teorema del Dini applicato a $(r^*)'(x)$ segue che esiste un intorno U del punto $\boldsymbol{a}^*$ ed esistono n funzioni $x_i = \phi_i(\boldsymbol{a})$ definite in U, tali che $x_i^* = \phi_i(\boldsymbol{a}^*)$ e per cui valgono le relazioni

$$f'(x_i) - q(\boldsymbol{a}, x_i) = 0, \tag{22}$$

per ogni $\boldsymbol{a} \in U$. Si pone $\boldsymbol{\alpha} = (a_0, \boldsymbol{a}, d, x_i)$, in cui le $x_i = \phi_i(\boldsymbol{a})$ sono le funzioni definite in U implicitamente dalle (22), e si considera il sistema non lineare

$$F_i(\boldsymbol{\alpha}) = f(x_i) - \sum_{j=0}^{n} a_j x_i^j - (-1)^i d = 0, \text{ per } i = 0, \ldots, n+1. \tag{23}$$

Per $i = 1, \ldots, n$ si applica a questo sistema il metodo di Newton-Raphson. Alla k-esima iterazione si ha

$$\sum_{m=0}^{n} \frac{\partial F_i}{\partial a_m}\bigg|_{\boldsymbol{\alpha}^{(k-1)}} \left(a_m^{(k)} - a_m^{(k-1)}\right) + \frac{\partial F_i}{\partial d}\bigg|_{\boldsymbol{\alpha}^{(k-1)}} \left(d^{(k)} - d^{(k-1)}\right) = -F_i(\boldsymbol{\alpha}^{(k-1)}), \tag{24}$$

dove

$$\boldsymbol{\alpha}^{(k-1)} = \left(a_0^{(k-1)}, \boldsymbol{a}^{(k-1)}, d^{(k-1)}, x_i^{(k)}\right) \quad \text{e} \quad x_i^{(k)} = \phi_i(\boldsymbol{a}^{(k-1)})$$

e

$$\frac{\partial F_i}{\partial a_m}\bigg|_{\boldsymbol{a}^{(k-1)}} = -\left(x_i^{(k)}\right)^m + \left(f'(x_i^{(k)}) - q(\boldsymbol{a}^{(k-1)}, x_i^{(k)})\right) \frac{\partial \phi_i}{\partial a_m}\bigg|_{\boldsymbol{a}^{(k-1)}}.$$

Dalla (22) risulta che

$$f'(x_i^{(k)}) - q(\boldsymbol{a}^{(k-1)}, x_i^{(k)}) = 0, \tag{25}$$

quindi

$$\frac{\partial F_i}{\partial a_m}\bigg|_{\boldsymbol{\alpha}^{(k-1)}} = -(x_i^{(k)})^m.$$

È facile verificare che tale relazione vale anche per $i = 0$ e $i = n+1$ e per $m = 0$. Quindi la (24) diventa

$$\begin{aligned} &-\sum_{m=0}^{n} (x_i^{(k)})^m \left(a_m^{(k)} - a_m^{(k-1)}\right) - (-1)^i \left(d^{(k)} - d^{(k-1)}\right) \\ &\quad = -\left[f(x_i^{(k)}) - \sum_{j=0}^{n} a_j^{(k-1)} (x_i^{(k)})^j - (-1)^i d^{(k-1)}\right]. \end{aligned}$$

Semplificando si ottiene la relazione

$$\sum_{j=0}^{n} a_j^{(k)} (x_i^{(k)})^j + (-1)^i d^{(k)} = f(x_i^{(k)}), \ i = 0, \ldots, n+1,$$

che coincide con la (6). Per il teorema 8 il metodo di Remez è convergente, quindi esiste un indice k tale che $\boldsymbol{\alpha}^{(k-1)} \in U$, per il quale le (25) sono esplicitabili. La successione $\{a_0^{(k)}, \ldots, a_n^{(k)}\}$ generata con il metodo di Remez coincide con quella generata dal metodo di Newton-Raphson applicato al sistema (23). Pertanto l'ordine di convergenza di tale successione ai coefficienti $a_0^*, \ldots, a_n^*$ di $p_n^*(x)$ è due. □

5 Approssimazione quasi minimax

Il calcolo del polinomio $p_n^*(x)$ di approssimazione minimax richiede, come si è visto, un notevole costo computazionale. Per questo motivo sono stati studiati altri metodi detti di approssimazione *quasi minimax*, che hanno un costo computazionale assai inferiore e consentono di determinare polinomi $p_n(x)$ che sono delle stime ragionevoli di $p_n^*(x)$. Qui si considerano quattro semplici approssimazioni quasi minimax:

- economizzazione,
- serie di Chebyshev troncata,
- interpolazione nei nodi di Chebyshev,
- arresto al primo passo del metodo di Remez.

Queste approssimazionui sfruttano le proprietà dei polinomi di Chebyshev, per cui si assume che l'intervallo $[a, b]$ sia $[-1, 1]$ (se ciò non fosse, si deve prima trasfomare l'intervallo). Saranno richieste alcune relazioni trigonometriche elencate nel seguente lemma.

Lemma 10 *Per $n \geq 1$ e $0 \leq \theta, \varphi \leq \pi$ si considerano la funzione*

$$\sigma(\theta, \varphi) = {\sum_{j=0}^{n}}' \cos j\theta \cos j\varphi$$

e i nodi $\varphi_i = \dfrac{(2i+1)\pi}{2(n+1)}$, per $i = 0, \ldots, n$, e $\psi_i = \dfrac{i\,\pi}{n+1}$, per $i = 0, \ldots, n+1$. Allora

(a) $$\sigma(\varphi_h, \varphi_i) = \delta_{i,h}\,(n+1)/2,$$

(b) $$\sigma(\psi_h, \psi_i) = \begin{cases} \dfrac{n+1}{2}\,\delta_{i,h} + \dfrac{(-1)^{i+h+1}}{2} & \text{per} \quad i = 1, \ldots, n \\ (n+1)\,\delta_{i,h} + \dfrac{(-1)^{i+h+1}}{2} & \text{per} \quad i = 0 \text{ e } i = n+1, \end{cases}$$

(c) $$\max_{0 \leq \theta \leq \pi} \int_0^\pi |\,\sigma(\theta, \varphi)\,|\,d\varphi = \int_0^\pi |\,\sigma(0, \varphi)\,|\,d\varphi\,,$$

(d) $$\max_{0 \leq \theta \leq \pi} \sum_{i=0}^{n} |\,\sigma(\theta, \varphi_i)\,| = \sum_{i=0}^{n} |\,\sigma(0, \varphi_i)\,|\,,$$

(e) $$\sigma(0, \varphi) = {\sum_{j=0}^{n}}' \cos j\varphi = \frac{\sin\left((n + 1/2)\,\varphi\right)}{2\sin(\varphi/2)}\,,$$

(f) $$\sum_{j=1}^{n} (-1)^j \sin\frac{2ij\,\pi}{2n+1} = -\frac{1}{2}\tan\frac{j\,\pi}{2n+1}\,.$$

• Il metodo di *economizzazione* viene di solito applicato ad approssimazioni polinomiali ottenute troncando serie di Taylor. Siano $p_n(x)$ un polinomio monico di grado n e $p_{n-1}(x) \in \mathcal{P}_{n-1}$ il polinomio di approssimazione minimax di $p_n(x)$. Poiché il resto $r(x) = p_n(x) - p_{n-1}(x)$ è un polinomio monico di grado n, per il lemma 4 deve essere $r(x) = t_n(x)$, per cui risulta

$$p_{n-1}(x) = p_n(x) - t_n(x).$$

Sia ora $p_n(x)$ un polinomio di approssimazione di grado n della $f(x)$ sull'intervallo $[-1, 1]$, tale che

$$\|f - p_n\|_\infty \leq \eta,$$

e sia $\epsilon > \eta$ l'approssimazione richiesta. Il metodo di economizzazione consiste nel sostituire al posto di $p_n(x)$ il polinomio $p_{n-1}(x)$ di minimax di $p_n(x)$ che, per quanto visto sopra, è dato da

$$p_{n-1}(x) = p_n(x) - a_n t_n(x),$$

dove a_n è il primo coefficiente di $p_n(x)$. Poiché $\|t_n\|_\infty = 1/2^{n-1}$, è

$$\|p_n - p_{n-1}\|_\infty \leq \frac{|a_n|}{2^{n-1}} \quad \text{e} \quad \|f - p_{n-1}\|_\infty \leq \eta + \frac{|a_n|}{2^{n-1}}\,.$$

Perciò se

$$\eta + \frac{|a_n|}{2^{n-1}} \leq \epsilon,$$

il polinomio $p_{n-1}(x)$ soddisfa l'approssimazione richiesta.

Questo procedimento si applica soprattutto a serie che convergono lentamente, e può essere riapplicato abbassando ogni volta il grado di 1. Se la funzione $f(x)$ è pari oppure dispari, il procedimento determina l'abbassamento del grado di 2.

Applicando ad esempio il metodo di economizzazione al polinomio

$$p_{2n+1}(x) = \sum_{j=0}^{n} (-1)^j \, \frac{x^{2j+1}}{(2j+1)!}\,,$$

che approssima la funzione $f(x) = \sin x$, si ottiene il polinomio

$$q_{2n-1}(x) = \sum_{j=0}^{n} (-1)^j \, \frac{x^{2j+1}}{(2j+1)!} \; - \; \frac{(-1)^n}{(2n+1)!}\, t_{2n+1}(x).$$

Per $n = 1$ si ha

$$p_3(x) = x - \frac{x^3}{3!} = x - 0.16667x^3,$$

da cui si ottiene l'approssimazione lineare di $\sin x$

$$q_1(x) = p_3(x) + \frac{1}{3!}\, t_3(x) = \frac{7}{8}\, x.$$

Per $x \in [-1,1]$ risulta

$$\|f - p_3\|_\infty \approx 0.814\ 10^{-2} \quad \text{e} \quad \|f - q_1\|_\infty \approx 0.419\ 10^{-1}.$$

Per $n = 2$ si ha

$$p_5(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!},$$
$$q_3(x) = p_5(x) - \frac{1}{5!}\, t_5(x) = \frac{383}{384}\, x - \frac{5}{32}\, x^3 = 0.99748\, x - 0.15625\, x^3.$$

Per $x \in [-1,1]$ risulta

$$\|f - p_5\|_\infty \approx 0.196\ 10^{-3} \quad \text{e} \quad \|f - q_3\|_\infty \approx 0.568\ 10^{-3},$$

e quindi nell'intervallo $[-1,1]$ $q_3(x)$ risulta un'approssimazione migliore in norma ∞ di $p_3(x)$.

Per $n = 3$ si ha

$$p_7(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!},$$

da cui applicando successivamente due volte il procedimento si ottiene

$$q_5(x) = p_7(x) + \frac{1}{7!}\, t_7(x) = \frac{46079}{46080}\, x - \frac{959}{5760}\, x^3 + \frac{23}{2880}\, x^5,$$
$$t_3(x) = q_5(x) - \frac{23}{2880}\, t_5(x) = \frac{11491}{11520}\, x - \frac{601}{3840}\, x^3 = 0.99748\, x - 0.15651\, x^3.$$

Per $x \in [-1,1]$ risulta

$$\|f - p_7\|_\infty \approx 0.273\ 10^{-5}, \quad \|f - q_5\|_\infty \approx 0.424\ 10^{-5}, \quad \|f - t_3\|_\infty \approx 0.502\ 10^{-3},$$

e quindi nell'intervallo $[-1,1]$ risulta che $q_5(x)$ è un'approssimazione migliore in norma ∞ di $p_5(x)$ e che $t_3(x)$ è migliore di $q_3(x)$. Nella figura 6 sono riportati nell'intervallo $[0,1]$ i grafici dei resti dei tre polinomi di terzo grado ottenuti: $f(x) - p_3(x)$ (con i pallini), $f - q_3(x)$ (con i quadratini neri) e $f(x) - t_3(x)$ (con linea continua). Come si può notare, i grafici di questi ultimi due resti sono praticamente coincidenti.

• La tecnica dei minimi quadrati consente di determinare esplicitamente i coefficienti dei polinomi che minimizzano l'errore in norma 2. Tali polinomi ovviamente non coincidono con quelli che minimizzano l'errore in norma ∞, ma possono rappresentare delle valide approssimazioni. In particolare conviene usare i polinomi ottenuti troncando la serie di Chebyshev, sia per la maggiore semplicità di calcolo, sia perché i polinomi di Chebyshev convergono rapidamente, oltre che in norma 2, anche in norma ∞ quando la $f(x)$ è derivabile fino ad un ordine elevato.

Posto $x = \cos\theta$, $0 \le \theta \le \pi$, il polinomio $p_n^C(x)$ ottenuto troncando all'$(n+1)$-esimo termine l'espansione della $f(x)$ in polinomi di Chebyshev di 1^a specie è

$$p_n^C(x) = p_n^C(\cos\theta) = {\sum_{j=0}^{n}}' \alpha_j^C \cos j\theta, \tag{26}$$

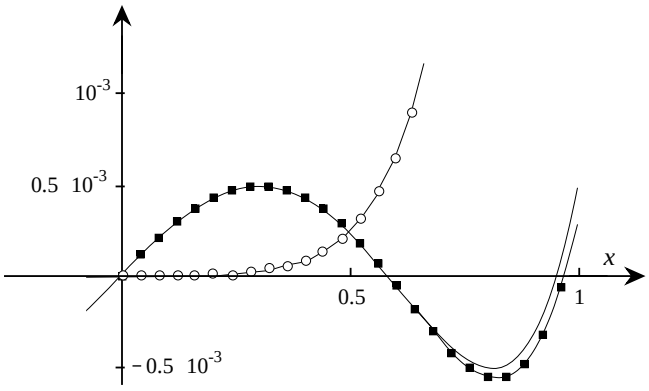


Figure 6: - Resti dei polinomi di terzo grado ottenuti dalla serie di Taylor della funzione $f(x) = \sin x$.

dove

$$\alpha_j^C = \frac{2}{\pi} \int_0^\pi f(\cos\varphi) \cos j\varphi \; d\varphi. \tag{27}$$

Se la convergenza è rapida, al crescere del grado i coefficienti dell'espansione tendono rapidamente a zero, per cui il coefficiente del primo termine trascurato può ragionevolmente stimare l'errore commesso.

• Se gli integrali (27) non possono essere ottenuti per via analitica, occorre approssimarli con una formula di quadratura. Usando la formula dei punti di mezzo con $n+1$ nodi, si ottiene il polinomio

$$p_n^+(x) = p_n^+(\cos\theta) = {\sum_{j=0}^{n}}' \alpha_j^+ \cos j\theta, \tag{28}$$

dove

$$\alpha_j^+ = \frac{2}{n+1} \sum_{i=0}^{n} f(\cos\varphi_i) \cos j\varphi_i, \quad \varphi_i = \frac{(2i+1)\pi}{2(n+1)}, \quad i = 0, \ldots, n. \tag{29}$$

Sostituendo le (29) nella (28) risulta

$$p_n^+(\cos\theta) = \sum_{i=0}^{n} L_i(\cos\theta) f(\cos\varphi_i), \quad \text{dove} \quad L_i^+(\cos\theta) = \frac{2}{n+1}\, \sigma(\theta, \varphi_i).$$

Per la (a) del Lemma 10 è $L_i^+(\cos\varphi_h) = \delta_{i,h}$ per $i, h = 0, \ldots, n$, per cui

$$p_n^+(\cos\varphi_h) = f(\cos\varphi_h),$$

cioè $p_n^+(x)$ è il polinomio di interpolazione della $f(x)$ nei nodi $x_i = \cos\varphi_i$, $i = 0, \ldots, n$. Il polinomio $p_n^+(x)$ rappresenta una valida alternativa al polinomio $p_n^*(x)$ ed è molto usato per la sua semplicità.

• Se per approssimare gli integrali (27) si usa la formula di quadratura dei trapezi con $n+2$ nodi, si ottiene

$$p_n^R(x) = p_n^R(\cos\theta) = \sum_{j=0}^{n}{}' \alpha_j^R \cos j\theta, \tag{30}$$

dove

$$\alpha_j^R = \frac{2}{n+1} \sum_{i=0}^{n+1}{}'' f(\cos\psi_i)\cos j\psi_i, \quad \text{dove} \quad \psi_i = \frac{i\pi}{n+1}, \tag{31}$$

e il doppio apice vicino alla sommatoria indica che il primo e l'ultimo termine devono essere dimezzati. Sostituendo le (31) nella (30) risulta

$$p_n^R(\cos\theta) = \sum_{i=0}^{n+1}{}'' L_i^R(\cos\theta) f(\cos\psi_i), \quad \text{dove} \quad L_i^R(\cos\theta) = \frac{2}{n+1}\,\sigma(\theta,\psi_i).$$

Per la (b) del Lemma 10 è

$$p_n^R(\cos\psi_h) = \sum_{i=0}^{n+1}{}'' L_i^R(\cos\psi_h) f(\cos\psi_i) = f(\cos\psi_h) + \frac{(-1)^{h+1}}{n+1} \sum_{i=0}^{n+1}{}''(-1)^i f(\cos\psi_i).$$

Estendendo la (31) all'indice $j = n+1$ si ha che

$$\alpha_{n+1}^R = \frac{2}{n+1} \sum_{i=0}^{n+1}{}''(-1)^i f(\cos\psi_i)$$

per cui

$$p_n^R(\cos\psi_h) + (-1)^h\,\frac{\alpha_{n+1}^R}{2} = f(\cos\psi_h). \tag{32}$$

Ponendo $d^{(0)} = \alpha_{n+1}^R/2$ e confrontando la (32) con la (7) si vede che il polinomio $p_n^R(x)$ è quello ottenuto con l'applicazione di un passo dell'algoritmo di Remez scegliendo come punti iniziali gli $x_i^{(0)} = \cos\psi_i$ ed esprimendo il polinomio come combinazione di polinomi di Chebyshev (si noti però che qui i nodi $x_i^{(0)}$ sono ordinati in modo decrescente, invece che crescente come nell'ordinamento consueto).

Anche tenendo conto della rapidità di convergenza dell'algoritmo di Remez, il costo computazionale del calcolo dei polinomi $p_n^C(x)$, $p_n^+(x)$ e $p_n^R(x)$ è molto minore di quello del polinomio $p_n^*(x)$. Inoltre, per valori non troppo grossi di n i resti dei polinomi ottenuti con uno di questi metodi sono dello stesso ordine del resto del polinomio di approssimazione minimax, come risulta dal teorema 11, la cui dimostrazione sfrutta le relazioni del lemma 10.

Teorema 11 *Valgono le seguenti maggiorazioni:*

$$\|f - p_n^C\|_\infty \le \delta^* u_n, \quad \|f - p_n^+\|_\infty \le \delta^* v_n, \quad \|f - p_n^R\|_\infty \le \delta^* w_n,$$

dove asintoticamente per $n \to \infty$ *è*

$$u_n \sim \frac{4}{\pi^2}\ \log n, \quad v_n \sim \frac{2}{\pi}\ \log n, \quad w_n \sim \frac{2}{\pi}\ \log n.$$

Dim: • Posto $r^C(x) = f(x) - p_n^C(x)$, la funzione $r^*(x) - r^C(x) = p_n^C(x) - p_n^*(x)$ è un polinomio di grado minore o uguale ad n e quindi può essere espressa nella forma

$$r^*(x) - r^C(x) = r^*(\cos\theta) - r^C(\cos\theta) = {\sum_{j=0}^{n}}' \gamma_j \cos j\theta,$$

dove

$$\gamma_j = \frac{2}{\pi}\ \int_0^\pi \Big(r^*(\cos\varphi) - r^C(\cos\varphi)\Big) \cos j\varphi\ d\varphi.$$

Poiché

$$\int_0^\pi r^C(\cos\varphi) \cos j\varphi\ d\varphi = 0,$$

risulta per $x = \cos\theta$, $0 \le \theta \le \pi$,

$$r^C(x) = r^*(x)\ - \frac{2}{\pi}\ \int_0^\pi r^*(\cos\varphi) {\sum_{j=0}^{n}}' \cos j\theta \cos j\varphi\ d\varphi,$$

da cui, per la (c) e la (e) del Lemma 10 è

$$\|r^C\|_\infty \le \delta^*\ \Big(1 + \frac{2}{\pi}\ \max_{0\le\theta\le\pi} \int_0^\pi |\ \sigma(\theta,\varphi)\ |\ d\varphi\Big) = \delta^*\ u_n$$

dove

$$u_n = 1 + \frac{2}{\pi}\ \int_0^\pi |\ \sigma(0,\varphi)\ |\ d\varphi = 1 + \frac{2}{\pi}\ \int_0^\pi \Big|\ \frac{\sin\left((n+1/2)\,\varphi\right)}{2\sin(\varphi/2)}\ \Big|\ d\varphi.$$

Si considerano i punti $\xi_i = \dfrac{2i\pi}{2n+1}$ per $i = 0, \ldots, n$ e $\xi_{n+1} = \pi$. È

$$S = \int_0^\pi |\ \sigma(0,\varphi)\ |\ d\varphi = \sum_{i=0}^{n} (-1)^i S_i,$$

dove

$$S_i = \int_{\xi_i}^{\xi_{i+1}}\ \frac{\sin\left((n+1/2)\,\varphi\right)}{2\sin(\varphi/2)}\ d\varphi = \int_{\xi_i}^{\xi_{i+1}}\ {\sum_{j=0}^{n}}'\ \cos j\varphi\, d\varphi = \Big[\frac{\varphi}{2} + \sum_{j=1}^{n} \frac{\sin j\varphi}{j}\ \Big]_{\xi_i}^{\xi_{i+1}}$$

Poiché

$$\sum_{i=0}^{n} (-1)^i \Big[\frac{\varphi}{2}\Big]_{\xi_i}^{\xi_{i+1}} = \frac{\pi}{2(2n+1)}$$

e

$$\sum_{i=0}^{n}(-1)^i\Big[\sum_{j=1}^{n}\frac{\sin j\varphi}{j}\Big]_{\xi_i}^{\xi_{i+1}} = -2\sum_{j=1}^{n}\frac{1}{j}\ \sum_{i=0}^{n}(-1)^i\sin\frac{2ij\,\pi}{2n+1}$$

per la (f) del Lemma 10 risulta

$$S = \frac{\pi}{2(2n+1)}\ +\ 2\sum_{j=1}^{n}\frac{1}{j}\ \tan\ \frac{j\,\pi}{2n+1}\,,$$

e quindi

$$\frac{2}{\pi}\,S = \frac{1}{2n+1} + \frac{2}{\pi}\ \sum_{j=1}^{n}\frac{1}{j}\ \tan\frac{j\,\pi}{2n+1} = \frac{1}{2n+1} + \frac{2}{2n+1}\ \sum_{j=1}^{n}\frac{2n+1}{j\pi}\ \tan\frac{j\,\pi}{2n+1}.$$

Poiché per $\varphi \to \pi/2$ il comportamento asintotico di $(1/\varphi)\,\tan\varphi$ è uguale a quello di $(2/\pi)\,\tan\varphi$, il comportamento asintotico di u_n è

$$u_n \sim \frac{4}{(2n+1)\pi}\ \sum_{j=1}^{n}\tan\frac{j\,\pi}{2n+1}.$$

Posto $h = \pi/(2n+1)$, per la formula di quadratura dei rettangoli si ha

$$h\ \sum_{j=1}^{n}\tan\frac{j\pi}{2n+1} \sim \int_{\pi/(2n+1)}^{\pi/2-\pi/(2n+1)}\tan\varphi\ d\varphi = -\log\,\cos\varphi\,\Big|_{\pi/(2n+1)}^{\pi/2-\pi/(2n+1)}$$

$$\sim -\log\,\sin\frac{\pi}{2n+1} \sim \log\,n.$$

Quindi

$$u_n \sim \frac{4}{\pi^2}\,\log n.$$

• Posto $r^+(x) = f(x) - p_n^+(x)$, la funzione $r^*(x) - r^+(x) = p_n^+(x) - p_n^*(x)$ è un polinomio di grado minore o uguale ad n tale che $r^*(x_i) - r^+(x_i) = r^*(x_i)$, dove $x_i = \cos\varphi_i$. Quindi

$$r^*(x) - r^+(x) = \sum_{i=0}^{n} r^*(x_i)L_i^+(x),\ \text{dove}\ L_i^+(x) = L_i^+(\cos\theta) = \frac{2}{n+1}\ \sigma(\theta,\varphi_i),$$

da cui

$$r^+(x) = r^*(x) - \frac{2}{n+1}\ \sum_{i=0}^{n} r^*(x_i)\,\sigma(\theta,\varphi_i).$$

Per la (d) e la (e) del Lemma 10 è

$$\|r^+\|_\infty \le \delta^*\ \Big(1 + \frac{2}{n+1}\ \max_{0\le\theta\le\pi}\sum_{i=0}^{n}\,|\,\sigma(\theta,\varphi_i)\,|\,\Big) = \delta^*\,v_n$$

dove

$$v_n = 1 + \frac{2}{n+1} \sum_{i=0}^{n} \Big| \frac{\sin\,(n+\frac{1}{2})\,\varphi_i}{2\sin\frac{\varphi_i}{2}} \Big| .$$

Poiché

$$\Big| \sin\,(n+\frac{1}{2})\,\varphi_i \Big| = \Big| \sin\Big[\frac{\pi}{2} - \frac{(2i+1)\pi}{4(n+1)}\Big] \Big| = \cos\frac{(2i+1)\pi}{4(n+1)},$$

asintoticamente per $n \to \infty$ è

$$\sum_{i=0}^{n} \Big| \frac{\sin\,(n+\frac{1}{2})\,\varphi_i}{2\sin\frac{\varphi_i}{2}} \Big| \sim \sum_{i=0}^{n} \frac{1}{\varphi_i} = \sum_{i=0}^{n} \frac{2(n+1)}{(2i+1)\pi} \sim \frac{n+1}{\pi}\,\log n,$$

e quindi

$$v_n \sim \frac{2}{\pi}\,\log n.$$

• Dalla (32) segue che anche il polinomio $p^R_{n+1}(x) = p^R_{n+1}(\cos\theta)$ è di interpolazione di $f(x)$, per cui si possono ripetere le stesse considerazioni del caso precedente, sostituendo al grado n il grado $n+1$, e ai nodi φ_i i nodi ψ_i, $i = 0, \ldots, n+1$, definiti in (31). Posto $r^R(x) = f(x) - p^R_{n+1}(x)$, risulta

$$\|r^R\|_\infty \le \delta^* \Big(1 + \frac{2}{n+1} \max_{0\le\theta\le\pi} {\sum_{i=0}^{n+1}}'' \,|\,\sigma(\theta,\psi_i)\,|\Big).$$

Ma a differenza del caso precedente, non si può dare un valore di θ indipendente da n per cui il max viene raggiunto. Si congettura che se n è pari il max venga raggiunto per $\theta = \pi/2$ e che $w_n = v_n$, mentre se n è dispari il max venga raggiunto per θ vicino a $\pi/2$ e che $w_n < v_n$, per $n \ge 3$. Per $n = 1$ è $v_1 = \sqrt{2}$ e $w_1 = 5/2$. Quindi a w_n si può dare la stessa dipendenza asintotica di v_n. □

Nella seguente tabella sono riportati i valori di u_n, v_n e w_n per alcuni valori di n.

n	u_n	v_n	w_n
1	2.436	2.414	2.5
2	2.642	2.667	2.667
5	2.961	3.104	3.094
10	3.223	3.489	3.489
50	3.860	4.466	4.466
100	4.139	4.901	4.901

Per calcolare approssimazioni quasi minimax di grado 3 della funzione e^x nell'intervallo $[0,1]$ si fa il cambiamento di variabile $x = (y+1)/2$ ottenendo la funzione

$$f(y) = e^{(y+1)/2}, \quad y \in [-1,1].$$

- La serie di Taylor di $f(y)$ è

$$e^{(y+1)/2} = \sum_{j=0}^{\infty} \frac{(y+1)^j}{2^j\ j!}\,.$$

Troncando al quinto termine ed applicando il metodo di economizzazione si ottiene il polinomio

$$q(y) = \sum_{j=0}^{4} \frac{(y+1)^j}{2^j\ j!}\ - \frac{1}{3072}\ T_4(y) = 0.03125\,y^3 + 0.20573\,y^2 + 0.82292\,y + 1.6481,$$

da cui

$$p_3(x) = q(2x-1) = 0.25\,x^3 + 0.44792\,x^2 + 1.0104\,x + 0.99967,$$

e risulta

$$\max_{x\in[0,1]} |e^x - p_3(x)| \approx 0.103\ 10^{-1}.$$

Troncando al sesto termine ed applicando il metodo di economizzazione per due volte si ottiene il polinomio

$$\begin{aligned}\overline{q}(y) &= \sum_{j=0}^{5} \frac{(y+1)^j}{2^j\ j!}\ - \frac{1}{61440}\ T_5(y)\ - \frac{1}{2048}\ T_4(y)\\ &= 0.03418\,y^3 + 0.20964\,y^2 + 0.824137\,y + 1.6482,\end{aligned}$$

da cui

$$\overline{p}_3(x) = \overline{q}(2x-1) = 0.27344\,x^3 + 0.42839\,x^2 + 1.0148\,x + 0.99953,$$

e risulta

$$\max_{x\in[0,1]} |e^x - \overline{p}_3(x)| \approx 0.212\ 10^{-2}.$$

- Si calcolano i coefficienti α_j, $j = 0,\ldots,3$, dello sviluppo di $f(y)$ in serie di polinomi di Chebyshev di 1^a specie. Dalla (27), posto $y = \cos\theta$, si ha

$$\alpha_j^C = \frac{2}{\pi}\ \int_0^{\pi} e^{(\cos\varphi+1)/2} \cos j\varphi\ d\varphi, \quad j = 0,\ldots,3,$$

e si ottengono i valori

$$\alpha_0^C = 3.5068,\ \alpha_1^C = 0.85039,\ \alpha_2^C = 0.10521,\ \alpha_3^C = 0.0087221.$$

Facendo nella (26) la sostituzione $\theta = \arccos(2x-1)$, risulta

$$p_3^C(x) = 0.27911\,x^3 + 0.42301\,x^2 + 1.0161\,x + 0.99948,$$
$$\|f - p_3^C(x)\|_\infty \approx 0.572\ 10^{-3}.$$

- Posto $\varphi_i = \dfrac{(2i+1)\pi}{8}$, $i = 0, \ldots, 3$, dalla (29) si ha

$$\alpha_0^+ = 3.5068,\ \alpha_1^+ = 0.85039,\ \alpha_2^+ = 0.10521,\ \alpha_3^+ = 0.0086942,$$

e facendo nella (28) la sostituzione $\theta = \arccos(2x-1)$, risulta

$$p_3^+(x) = 0.27822\,x^3 + 0.42434\,x^2 + 1.0156\,x + 0.99951,$$
$$\|f - p_3^+\|_\infty \approx 0.603\ 10^{-3}.$$

- Fissati i punti $\psi_i = \dfrac{i\pi}{4}$, $i = 0, \ldots, 4$, dalla (31) si ha

$$\alpha_0^R = 3.5068, \quad \alpha_1^R = 0.85039, \quad \alpha_2^R = 0.10521, \quad \alpha_3^R = 0.0087492,$$

e facendo nella (30) la sostituzione $\theta = \arccos(2x-1)$, risulta

$$p_3^R(x) = 0.27998\,x^3 + 0.42172\,x^2 + 1.0166\,x + 0.99946,$$
$$\|f - p_3^R\|_\infty \approx 0.547\ 10^{-3}.$$

I massimi moduli dei resti dei polinomi di grado 3 ottenuti con i metodi quasi minimax per l'approssimazione di e^x sull'intervallo $[0,1]$ sono riportati nella tabella, a confronto con il valore corrispondente del polinomio minimax.

metodo	resto in norma ∞
minimax	$0.545\ 10^{-3}$
serie di Taylor + 1 procedimento di economizzazione	$0.103\ 10^{-1}$
serie di Taylor + 2 procedimenti di economizzazione	$0.212\ 10^{-2}$
serie di Chebyshev troncata	$0.572\ 10^{-3}$
interpolazione nei nodi di Chebyshev	$0.603\ 10^{-3}$
un passo del metodo di Remez	$0.547\ 10^{-3}$

Come conseguenza del teorema 11 si ha che se $f \in C^1[a,b]$, allora la successione dei polinomi di interpolazione di grado n, costruiti assumendo come nodi gli zeri dei polinomi di Chebyshev di 1^a specie, converge alla funzione $f(x)$ su tutto l'intervallo

$[-1, 1]$. Infatti combinando il teorema 3 di Jackson e il teorema 11, asintoticamente risulta

$$\|f - p_n^+\|_\infty \leq \gamma \|f'\|_\infty \frac{v_n}{n}, \quad v_n \sim \frac{2}{\pi} \log n.$$

Se invece $f \notin C^1[a, b]$ è solo continua, la convergenza della successione dei polinomi di interpolazione non è garantita (mentre risulta garantita la convergenza in media, cioè in norma 2).

Si considera ad esempio il polinomio di interpolazione della funzione di Runge

$$f(x) = \frac{1}{1 + x^2},$$

definita sull'intervallo $[-5, 5]$. Come si è visto, assumendo come nodi dei punti equidistanti, al crescere di n il polinomio di interpolazione approssima sempre peggio la $f(x)$. Assumendo invece come nodi i punti di Chebyshev, che nell'intervallo $[-5, 5]$ sono

$$x_i = 5 \cos \frac{(2i + 1)\pi}{2(n + 1)}, \quad i = 0, 1, \ldots, n,$$

si ottiene una successione di polinomi che converge alla $f(x)$. Nella figura 7 sono riportati, al variare di n, i valori della norma $r_n = \|f - p_n\|_\infty$ nell'intervallo $[-5, 5]$ per i polinomi di interpolazione $p_n(x)$ di grado n, assumendo come nodi punti equidistanti (linea con i pallini) e punti di Chebyshev (linea con i quadratini neri). Mentre nel caso dei nodi equidistanti risulta chiaramente la divergenza della successione r_n, nel caso dei nodi di Chebyshev i resti decrescono per $n \leq 20$. Per valori di n più elevati gli errori di arrotondamento distruggono il carattere di convergenza della successione.

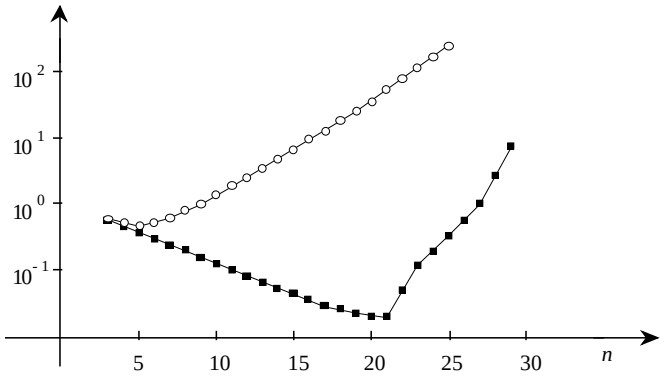


Figure 7: - Errori dei polinomi di interpolazione per la funzione di Runge.

6 Approssimazione minimax rispetto all'errore relativo

La bontà dell'approssimazione $\widetilde{x}$ di un numero reale $x \neq 0$, rappresentato in virgola mobile (*floating point*) viene misurata più frequentemente per mezzo dell'errore

relativo

$$\epsilon = \frac{\widetilde{x} - x}{x}$$

che dell'errore assoluto $\widetilde{x} - x$. Così facendo si prende in considerazione il numero delle cifre significative di x, indipendentemente dalla grandezza di x. Le stesse considerazioni si possono applicare all'approssimazione di funzioni. L'approssimazione minimax consente di determinare, con poche modifiche, anche approssimazioni rispetto all'errore relativo.

Il teorema 1 di equioscillazione di Chebyshev vale anche quando come funzione di errore si considera il resto relativo

$$r(x) = \frac{f(x) - p_n(x)}{f(x)}\,, \tag{33}$$

dove $f(x) \neq 0$, per $x \in [a, b]$. La dimostrazione è analoga a quella svolta per l'errore assoluto.

Ovviamente il polinomio di approssimazione minimax rispetto all'errore relativo non è lo stesso di quello rispetto all'errore assoluto, anzi i due polinomi possono essere abbastanza diversi.

Se $f(x) \neq 0$ in $[a, b]$ il polinomio

$$p_1^*(x) = a_1 x + a_0$$

di approssimazione minimax lineare rispetto all'errore relativo è tale che

$$r^*(x) = \frac{f(x) - p_1^*(x)}{f(x)} \tag{34}$$

assume massimo e minimo locale in tre punti distinti di $[a, b]$

$$a = x_0^* < x_1^* < x_2^* = b.$$

Dovendo essere

$$(r^*)'(x_1^*) = 0,$$

ne segue che x_1^* è soluzione dell'equazione

$$(p_1^*)'(x) f(x) - p_1^*(x) f'(x) = 0,$$

cioè

$$a_1[f(x) - x f'(x)] - a_0 f'(x) = 0.$$

Ne segue che x_1^* è tale che

$$\frac{a_0}{a_1} = \frac{f(x_1^*)}{f'(x_1^*)} - x_1^*. \tag{35}$$

Imponendo le condizioni di equioscillazione nei tre punti a, x_1^*, b, si ottengono altre 3 equazioni

$$\begin{cases} f(a) - a_1 a - a_0 = d\, f(a) \\ f(x_1^*) - a_1 x_1^* - a_0 = -d\, f(x_1^*) \\ f(b) - a_1 b - a_0 = d\, f(b). \end{cases} \tag{36}$$

Ad esempio, nel caso della funzione $f(x) = \sqrt{x}$ per $x \in [1/16, 1]$, il sistema formato dalle (35) e (36) è

$$\begin{cases} x_1^* = a_0/a_1 \\ 4 - a_1 - 16a_0 = 4d \\ \sqrt{x_1^*} - a_1 x_1^* - a_0 = -d\sqrt{x_1^*} \\ 1 - a_1 - a_0 = d, \end{cases}$$

la cui soluzione è $x_1^* = 1/4$, $a_0 = 2/9$, $a_1 = 8/9$, $d = -1/9$. Il polinomio minimax di grado 1 rispetto all'errore relativo della funzione $f(x) = \sqrt{x}$ nell'intervallo $[1/16, 1]$ è allora dato da

$$p_1^*(x) = \frac{1}{9}\,(8x + 2).$$

Nella figura 8 è riportato il corrispondente andamento del resto relativo (34).

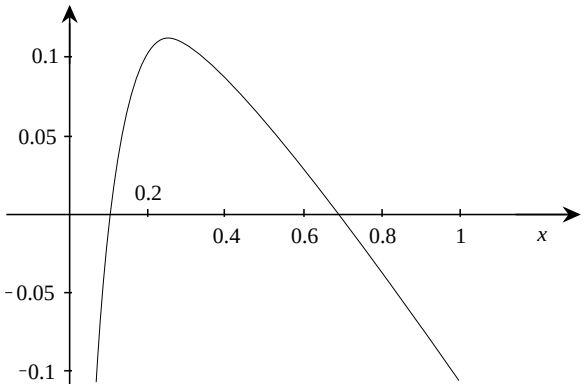


Figure 8: - Resto relativo dell'approssimazione minimax lineare rispetto all'errore relativo della funzione $f(x) = \sqrt{x}$.

Anche l'algoritmo di Remez non richiede modifiche sostanziali. Basta sostituire la (6) con la

$$\sum_{j=0}^{n} a_j^{(k)} (x_i^{(k)})^j + (-1)^i f(x_i^{(k)}) d^{(k)} = f(x_i^{(k)}), \quad i = 0, \ldots, n+1, \tag{37}$$

per ottenere i coefficienti del polinomio $p_n^{(k)} \in \mathcal{P}_n$ tale che

$$\frac{f(x_i) - p_n^{(k)}(x_i)}{f(x_i)} = (-1)^i d^{(k)}.$$

Inoltre i punti $y_i,\ i = 1, \ldots, n$, devono essere i punti di massimo o minimo locale del resto relativo (33). La dimostrazione della convergenza del metodo di Remez così modificato è sostanzialmente uguale a quella del teorema 8.

Ad esempio, determiniamo con il metodo di Remez il polinomio di approssimazione minimax rispetto all'errore relativo di grado al più 3 della funzione $f(x) = \mathrm{e}^x$ nell'intervallo $[0, 1]$. Assumendo gli stessi 5 punti iniziali

$$a = x_0^{(0)} < x_1^{(0)} < x_2^{(0)} < x_3^{(0)} < x_4^{(0)} = b,$$

si ottiene dal sistema (37) la soluzione

$$a_3^{(0)} = 0.27154,\ a_2^{(0)} = 0.43411,\ a_1^{(0)} = 1.0121,\ a_0^{(0)} = 0.99969,\ d^{(0)} = 0.30994\ 10^{-3}.$$

I punti di massimo o minimo del resto relativo della prima iterazione

$$r^{(0)}(x) = \frac{\mathrm{e}^x - 0.27154x^3 - 0.43411x^2 - 1.0121x - 0.99969}{\mathrm{e}^x}$$

interni all'intervallo $[0, 1]$ sono

$$y_1 = 0.12845,\ y_2 = 0.46652,\ y_3 = 0.83786.$$

Assumendo $x_1^{(1)} = y_1,\ x_2^{(1)} = y_2,\ x_3^{(1)} = y_3$ e ripetendo il calcolo, dopo altre due iterazioni si ottiene

$$a_3^{(3)} = 0.27136,\ a_2^{(3)} = 0.43421,\ a_1^{(3)} = 1.0122,\ a_0^{(3)} = 0.99968,\ d^{(3)} = 0.32110\ 10^{-3}.$$

Le successive iterazioni non modificano tali coefficienti, per cui si assume che il polinomio di terzo grado di approssimazione minimax rispetto all'errore relativo di $f(x) = \mathrm{e}^x$ nell'intervallo $[0, 1]$ è dato da

$$p_3^*(x) = 0.27136x^3 + 0.43421x^2 + 1.0122x + 0.99968.$$

La condizione che la $f(x)$ non debba annullarsi nell'intervallo $[a, b]$ è piuttosto pesante, perché esclude la possibilità di approssimazione relativa di funzioni per altri aspetti del tutto regolari, come ad esempio la funzione $\log x$ in un qualunque intervallo che contenga il punto 1. Questa condizione può essere indebolita: infatti si può approssimare con il minimax rispetto all'errore relativo anche una $f(x)$ che nell'intervallo $[a, b]$ abbia un solo zero α di molteplicità 1, cioè tale che il

$$\lim_{x \to \alpha} \frac{f(x)}{x - \alpha}$$

sia finito e diverso da zero. In tal caso si considera la funzione

$$g(x) = \begin{cases} \dfrac{f(x)}{x - \alpha}, & \text{se } \ x \neq \alpha, \\ \lim\limits_{t \to \alpha} \dfrac{f(t)}{t - \alpha}, & \text{se } \ x = \alpha, \end{cases}$$

e si determina il polinomio p^*_{n-1} di migliore approssimazione per la g. Poiché per $x \neq \alpha$ è

$$\frac{|g(x) - p^*_{n-1}(x)|}{|g(x)|} = \Big|\frac{f(x) - (x-\alpha)p^*_{n-1}(x)}{f(x)}\Big|,$$

ne segue che $p^*_n(x) = (x-\alpha)p^*_{n-1}(x)$ è il polinomio di migliore approssimazione minimax della $f(x)$ rispetto all'errore relativo in $[a,b] - \{\alpha\}$.

Ad esempio, nell'intervallo $[1,2]$ la funzione $\log x$ ha uno zero di molteplicità 1. Per determinare il polinomio di secondo grado di approssimazione minimax rispetto all'errore relativo, si considera la funzione

$$g(x) = \begin{cases} \dfrac{\log x}{x-1}, & \text{se } x \neq 1, \\ 1, & \text{se } x = 1. \end{cases}$$

Con il metodo di Remez si ottiene il polinomio lineare

$$p^*_1(x) = -0.30024\,x + 1.2787$$

di approssimazione minimax rispetto all'errore relativo della $g(x)$ e quindi il polinomio

$$p^*_2(x) = (x-1)p^*_1(x) = -0.30024\,x^2 + 1.5789\,x - 1.2787$$

è il polinomio cercato e risulta $\delta^* \approx 0.380\ 10^{-1}$.

7 Approssimazione minimax con vincoli

In alcuni casi le approssimazioni cercate devono soddisfare alcuni vincoli. I casi più frequenti sono:

a) la funzione approssimante deve assumere in uno o più punti gli stessi valori della $f(x)$, oppure avere in uno o più punti gli stessi valori e le stesse derivate della $f(x)$ fino ad un ordine prefissato;

b) la funzione approssimante appartiene ad una sottoclasse della classe dei polinomi, come ad esempio nel caso in cui alcuni coefficienti debbano assumere valori prefissati.

Approssimazioni di questo genere possono essere utilissime, quando si richiede ad esempio che la funzione approssimante abbia lo stesso andamento della $f(x)$, oppure conservi le stesse simmetrie.

In generale, ben poco si può dire, dal punto di vista teorico, per quanto riguarda l'esistenza e l'unicità delle soluzioni di tali problemi. Per alcuni casi particolari, come ad esempio per il caso in cui la funzione approssimante è un polinomio con alcuni coefficienti prefissati, esiste una generalizzazione del teorema 1 di Chebyshev, in cui il numero dei punti di equioscillazione del resto risulta diminuito, per tener conto del minor numero di gradi di libertà.

Ad esempio, l'approssimazione minimax di $f(x) = \sqrt{x}$ rispetto all'errore relativo con un polinomio della forma

$$p_1^*(x) = x + a_0$$

nell'intervallo $[1/16, 1]$, è tale che il resto relativo

$$r^*(x) = 1 - \frac{x + a_0}{\sqrt{x}}$$

ha due punti di equioscillazione. Poiché $(r^*)'(x)$ può annullarsi in un punto solo, non è possibile che i due punti di equioscillazione siano entrambi interni. Supponendo che essi siano gli estremi dell'intervallo, si impongono le condizioni $r^*(1/16) = d$, $r^*(1) = -d$ e si risolve il sistema lineare di due equazioni nelle incognite a_0 e d che così si ottiene. La soluzione è $a_0 = d = 3/20$. Però il polinomio $p(x) = x + 3/20$ non è l'approssimazione minimax cercata, perché il resto $r^*(x)$ risulta avere un punto di massimo interno all'intervallo $[1/16, 1]$, esattamente il punto $x = a_0 = 3/20$, in cui ha il valore $r^*(3/20) = 0.225403$, superiore a d. Di conseguenza uno dei due punti di equioscillazione deve essere interno all'intervallo e l'altro deve coincidere con uno dei due estremi.

Ponendo $a < x_0^* < x_1^* = b$, si ricava

$$x_0^* = a_0 = d = 3 - 2\sqrt{2} = 0.171573 = \delta^*;$$

ponendo invece $a = x_0^* < x_1^* < b$, si ricava

$$x_1^* = a_0 = \frac{9 - 4\sqrt{2}}{16} = 0.20895 \quad \text{e} \quad d = -\frac{3 - 2\sqrt{2}}{2} = -0.085786,$$

ma $|r^*(1)| = 0.20895$.

Confrontando i valori ottenuti risulta che il polinomio che fornisce l'approssimazione minimax cercata è

$$p_1^*(x) = x + 3 - 2\sqrt{2} = x + 0.17157.$$

Nel caso in cui il polinomio cercato debba avere alcuni coefficienti prefissati, il metodo di Remez può essere applicato senza grosse modifiche, sostituendo nella (6) i valori prefissati e diminuendo di conseguenza il numero delle equazioni. Ad esempio, se il polinomio $p_n^*(x)$ deve essere pari, i coefficienti a_j per j dispari devono essere nulli e la (6) assume la forma

$$\sum_{\substack{j=0 \\ j \text{ pari}}}^{n} a_j^{(k)} (x_i^{(k)})^j + (-1)^i d^{(k)} = f(x_i^{(k)}), \quad i = 0, \ldots, \frac{n}{2} + 1.$$

Ad esempio, per determinare il polinomio di approssimazione minimax di $\cos x$ della forma

$$p_2^*(x) = 1 + a_0 x^2 + a_1 x^4$$

nell'intervallo $[0, \frac{\pi}{2}]$, con il metodo di Remez, al posto del sistema (6) si risolve il sistema

$$1 + a_0^{(k)} (x_i^{(k)})^2 + a_1^{(k)} (x_i^{(k)})^4 + (-1)^i d^{(k)} = f(x_i^{(k)}), \quad i = 0, 1, 2.$$

Non si può scegliere $x_0^{(0)} = 0$ perché risulterebbe $d^{(0)} = 0$. Scegliendo invece ad esempio

$$x_0^{(0)} = \frac{\pi}{6}, \quad x_1^{(0)} = \frac{\pi}{3}, \quad x_2^{(0)} = \frac{\pi}{2},$$

dopo 3 iterazioni si ottiene il polinomio

$$p_2^*(x) = 1 - 0.49661\, x^2 + 0.037131\, x^4,$$

per il quale risulta $\delta^* \approx 0.737\ 10^{-3}$.

8 Approssimazione di alcune funzioni elementari

Come anticipato, il minimax è particolarmente indicato per costruire approssimazioni di funzioni da includere nel software. In questo caso è importante tenere presenti aspetti quali l'efficienza computazionale e la stabilità delle approssimazioni che si ottengono. L'uso di proprietà matematiche note delle funzioni è comunque fondamentale, perché consente di ottenere approssimazioni migliori.

La ragione per cui il calcolo della f in un punto x viene ricondotto al calcolo in un altro punto y è che l'efficienza dell'approssimazione dipende in gran parte dall'ampiezza dell'intervallo su cui si opera. È noto ad esempio che l'approssimazione con la formula di Taylor della funzione $\sin x$ è tanto migliore quanto più x è vicino allo zero. Sfruttando le proprietà di periodicità, di simmetria e antisimmetria, il calcolo di $\sin x$ può essere ricondotto a quello di $\sin y$, con $0 \le y \le \pi/2$, in quanto

$$\sin x = \sin(x - 2\pi), \qquad \sin x = -\sin(x - \pi) \quad \text{e} \quad \sin(\pi/2 + x) = \sin(\pi/2 - x).$$

Naturalmente l'applicazione di queste relazioni richiede che il valore di π sia noto con un numero di cifre esatte sufficienti per l'approssimazione che si vuole ottenere.

Un altro esempio classico è quello della funzione $f(x) = \mathrm{e}^x$, per cui valgono le relazioni

$$\mathrm{e}^{-x} = 1/\mathrm{e}^x \quad \text{e} \quad \mathrm{e}^{x+1} = \mathrm{e}\ \mathrm{e}^x,$$

che consentono di ridurre il calcolo a quello dell'esponenziale in un punto compreso fra 0 ed e. Anche in questo caso è richiesta la conoscenza di un numero sufficiente di cifre esatte di e.

Ulteriori riduzioni dell'intervallo si possono ottenere applicando delle relazioni che sfruttano proprietà dipendenti dalla base β dell'aritmetica di macchina (si suppone che $\beta = 2$). Si ottengono così relazioni del tipo

$$f(x) = 2^k f(y) \quad \text{oppure} \quad f(x) = 2^k g(y), \quad \text{con } k \text{ intero},$$

cioè il calcolo della f nel punto x è ricondotto al calcolo della stessa funzione o di un'altra g in un altro punto y. La moltiplicazione per 2^k non richiede in pratica una moltiplicazione effettiva, ma solo l'aggiunta dell'intero k all'esponente della rappresentazione di $f(y)$ o di $g(y)$.

Si esamina ora l'approssimazione in precisione semplice (circa 7 cifre decimali esatte) di alcune funzioni elementari, indicando per ciascuna di esse la tecnica di riduzione dell'intervallo, supponendo in ogni caso che le costanti richieste siano memorizzate con sufficiente precisione e quindi non concorrano al costo.

- Per la **radice quadrata**, sia

$$x = m\, 2^p, \quad \text{con} \quad 1/2 \le m < 1,$$

la rappresentazione in base 2 di x, $x > 0$. Se $p = 2n$, n intero, è

$$\sqrt{x} = \sqrt{m}\, 2^n,$$

e se $p = 2n + 1$, è

$$\sqrt{x} = \sqrt{m/2}\, 2^{n+1}.$$

Perciò il calcolo di $\sqrt{x}$ viene ricondotto al calcolo di $\sqrt{z}$, dove $z = m$ se p è pari e $z = m/2$ se p è dispari.

La riduzione dell'intervallo è richiesta dalla necessità di avere una buona approssimazione iniziale per il metodo iterativo che viene utilizzato, applicato alla funzione

$$h(y) = y^2 - z,$$

il cui zero positivo è proprio $\sqrt{z}$. Poiché la molteplicità della radice è 1, si usa il metodo iterativo delle tangenti

$$y_n = y_{n-1} - \frac{h(y_{n-1})}{h'(y_{n-1})} = \frac{1}{2}\left(y_{n-1} + \frac{z}{y_{n-1}}\right),$$

che ha ordine di convergenza 2. Per l'approssimazione iniziale y_0 si usa una funzione della forma

$$y_0 = a_0 + a_1\, z,$$

in cui i coeficienti a_0 e a_1 vengono determinati con un minimax dell'errore relativo sull'intervallo $[1/2, 1]$. I valori che si ottengono sono

$$a_0 = 0.41731, \quad a_1 = 0.59016,$$

e y_0 risulta avere un errore relativo minore di 0.0075 su tutto l'intervallo. A partire da y_0 vengono fatte due iterazioni del metodo delle tangenti e si ottiene un errore relativo inferiore a 4×10^{-10} su tutto l'intervallo (tre iterazioni sarebbero sufficienti per ottenere la precisione doppia).

- Per le funzioni **seno** e **coseno**, si tiene conto della periodicità e della simmetria e del fatto che

$$\sin\left(\frac{\pi}{4} \pm x\right) = \cos\left(\frac{\pi}{4} \mp x\right). \tag{38}$$

Il calcolo di $\sin x$ e $\cos x$ per ogni x reale può quindi essere ricondotto al calcolo di $\sin x$ e $\cos x$ per $x \in [0, \pi/4]$. Per approssimare $\sin x$ si cerca un polinomio $p_s(x)$ tale che

(1) $p_s(x)$ sia una funzione dispari, e quindi $p_s(0) = 0$,

(2) $\lim_{x\to 0} \dfrac{p_s(x)}{x} = 1$;

per approssimare $\cos x$ si cerca un polinomio $p_c(x)$ tale che

(3) $p_c(x)$ sia una funzione pari,

(4) $p_c(0) = 1$.

Si definsce poi la funzione $g(x)$ che approssima $\sin x$ nel modo seguente

$$g(x) = \begin{cases} p_s(x), & \text{per } \ x \in [0, \dfrac{\pi}{4}], \\ p_c(\dfrac{\pi}{2} - x), & \text{per } \ x \in (\dfrac{\pi}{4}, \dfrac{\pi}{2}] \end{cases} \tag{39}$$

(e analogamente per $\cos x$).

Nell'intervallo $x \in [0, \pi/4]$ le serie di Taylor di $\sin x$ e $\cos x$ convergono abbastanza rapidamente, quindi possono essere prese come base per applicare un procedimento di economizzazione. Per la funzione $\sin x$ si considerano i primi 4 termini della serie di Taylor di

$$\frac{1}{x^2}\left(\frac{\sin x}{x} - 1\right) = -\frac{1}{3!} + \frac{x^2}{5!} - \frac{x^4}{7!} + \frac{x^6}{9!} + O(x^8).$$

Con la trasformazione $x = \dfrac{\pi}{4}\, t$ e il procedimento di economizzazione si determina un polinomio $q(t)$ di quarto grado, da cui si ottiene

$$p_s(x) = x + x^3 q(\frac{4}{\pi}\, x) = x - 0.16667\, x^3 + 0.0083327\, x^5 - 0.00019586\, x^7,$$

e risulta

$$\max_{x\in[0,\pi/4]} |\sin x - p_s(x)| \approx 0.240 \times 10^{-8}, \quad \max_{x\in[0,\pi/4]} \left|\frac{\sin x - p_s(x)}{\sin x}\right| \approx 0.364 \times 10^{-8}.$$

Per la funzione $\cos x$ si considerano i primi 4 termini della serie di Taylor di

$$\frac{\cos x - 1}{x^2} = -\frac{1}{2!} + \frac{x^2}{4!} - \frac{x^4}{6!} + \frac{x^6}{8!} + O(x^8).$$

Procedendo come sopra si ottiene il polinomio

$$p_c(x) = 1 - 0.49999\, x^2 + 0.041661\, x^4 - 0.0013659\, x^6,$$

e risulta

$$\max_{x\in[0,\pi/4]} |\cos x - p_c(x)| \approx 0.924 \times 10^{-7}, \quad \max_{x\in[0,\pi/4]} \left|\frac{\cos x - p_c(x)}{\cos x}\right| \approx 0.129 \times 10^{-6}.$$

In alternativa si costruisce il polinomio di approssimazione minimax di grado 7 per la funzione $\sin x$ e di grado 6 per la funzione $\cos x$, in modo che valgano le proprietà (1) $\div$ (4). Per la funzione $\sin x$ si applica l'algoritmo di Remez a un polinomio della forma

$$p_s(x) = x + \sum_{i=1}^{3} a_i x^{2i+1},$$

scegliendo come nodi iniziali 3 punti in $(0, \pi/4)$ e il punto $\pi/4$. Si ottiene

$$p_s(x) = x - 0.16666\, x^3 + 0.0083320\, x^5 - 0.00019496\, x^7,$$

e risulta

$$\max_{x\in[0,\pi/4]} |\sin x - p_s(x)| \approx 0.212 \times 10^{-8}, \quad \max_{x\in[0,\pi/4]} \left|\frac{\sin x - p_s(x)}{\sin x}\right| \approx 0.637 \times 10^{-8}.$$

Per la funzione $\cos x$ si applica l'algoritmo di Remez a un polinomio della forma

$$p_c(x) = 1 + \sum_{i=1}^{3} a_i x^{2i},$$

scegliendo i nodi iniziali come sopra. Si ottiene

$$p_c(x) = 1 - 0.49999\, x^2 + 0.041656\, x^4 - 0.0013598\, x^6,$$

e risulta

$$\max_{x\in[0,\pi/4]} |\cos x - p_c(x)| \approx 0.346 \times 10^{-7}, \quad \max_{x\in[0,\pi/4]} \left|\frac{\cos x - p_c(x)}{\cos x}\right| \approx 0.489 \times 10^{-7}.$$

La funzione $g(x)$ che approssima $\sin x$ viene poi definita come in (39).

Si nota che la $g(x)$ costruita con il minimax ha un errore relativo più equilibrato. Si esamina quindi se le proprietà di continuità e di crescenza della funzione $\sin x$ sull'intervallo $[0, \pi/2]$ sono verificate anche dalle $g(x)$ trovate. In entrambi i casi $p_s(x)$ è crescente e $p_c(x)$ è decrescente per $x \in [0, \pi/4]$, ma la funzione $g(x)$ non è né continua né crescente in $\pi/4$, infatti

$$p_s(\frac{\pi}{4}) - p_c(\frac{\pi}{4}) \approx \begin{cases} 0.106 \times 10^{-6} & \text{per l'economizzazione,} \\ 0.325 \times 10^{-7} & \text{per il minimax.} \end{cases}$$

Comunque il polinomio ottenuto con il minimax fornisce uno scarto minore ed è quindi preferibile.

- Per la funzione **esponenziale**, posto

$$\mathrm{e}^x = 2^{z/2}, \quad \text{con } z = \frac{2x}{\log 2} = r + s, \quad r = \lceil z \rceil,\ -1 < s \leq 0,$$

risulta

$$\frac{z}{2} = u - v + \frac{s}{2}\,,$$

dove $u = \lceil r/2 \rceil$ è intero, $v = 0$ oppure $1/2$, e $s/2 \in (-1/2, 0]$, e si ha

$$\mathrm{e}^x = 2^u 2^{-v} 2^y, \quad y \in (-1/2, 0].$$

La moltiplicazione per 2^u non viene effettivamente eseguita, perché comporta la sola modifica dell'esponente. È richiesta la memorizzazione delle costanti $\log 2$ e $2^{-1/2}$. Quindi il calcolo di e^x per ogni x reale viene ricondotto al calcolo di

$$2^x \quad \text{per} \quad x \in [-1/2, 0].$$

Data la rapidità di convergenza della serie di Taylor di 2^x, si potrebbe sfruttarla per ottenere l'approssimazione cercata sull'intervallo $[-1/2, 0]$. Il massimo errore relativo dell'approssimazione ottenuta con i primi n termini della serie si verifica per gli x vicini a $-1/2$. Per ottenere 7 cifre decimali esatte occorrono 8 termini. Con la formula che si ottiene applicando l'algoritmo di Remez si ottiene la stessa precisione con costi minori.

Si applica l'algoritmo di Remez a un polinomio di grado 5 della forma

$$p_5(x) = 1 + \sum_{i=1}^{5} a_i x^i,$$

che in 0 vale 1, scegliendo come nodi iniziali il punto -0.5 e 5 punti in $(-0.5, 0)$. Si ottiene

$$p_5(x) = 1 + 0.69315\,x + 0.24022\,x^2 + 0.055477\,x^3 + 0.00950817\,x^4 + 0.0011262\,x^5,$$

e risulta

$$\max_{x\in[-0.5,0]} |2^x - p_5(x)| \approx 0.272 \times 10^{-8}, \quad \max_{x\in[-0.5,0]} \left|\frac{2^x - p_5(x)}{2^x}\right| \approx 0.385 \times 10^{-8}.$$

- Per la funzione **logaritmo**, sia $x = 2^p m$, la rappresentazione in base 2 di x. La mantissa m è tale che $1/2 \le m < 1$. Si pone

$$z = \frac{\sqrt{2}\,m - 1}{\sqrt{2}\,m + 1}\,.$$

Risulta

$$-\frac{\sqrt{2}-1}{\sqrt{2}+1} \le z < \frac{\sqrt{2}-1}{\sqrt{2}+1}\,, \quad \text{e} \quad m = \frac{1}{\sqrt{2}}\,\frac{1+z}{1-z}\,.$$

Quindi il calcolo di $\log x$ per $x > 0$ viene ricondotto al calcolo di

$$f(x) = \log\frac{1+x}{1-x} \quad \text{per} \quad |x| \le \delta, \quad \delta = \frac{\sqrt{2}-1}{\sqrt{2}+1} = 0.17157.$$

La riduzione dell'intervallo risulta di facile applicazione se si può operare con istruzioni di macchina sulla prima cifra di m rappresentata in base 2. È comunque richiesta la memorizzazione delle costanti $\sqrt{2}$ e $\log 2$.

Per approssimare $f(x)$ cerca un polinomio $p_7(x)$ di grado 7 tale che

(1) $p_7(x)$ sia una funzione dispari, e quindi $p_7(0) = 0$,

(2) $\lim_{x\to 0} \dfrac{p_7(x)}{x} = 2.$

Con il procedimento di economizzazione si considerano i primi 4 termini della serie di Taylor di

$$\frac{1}{x^2}\left(\frac{f(x)}{2x} - 1\right) = \frac{1}{3} + \frac{x^2}{5} + \frac{x^4}{7} + \frac{x^6}{9} + O(x^8).$$

Posto $x = t\,\delta$, si determina un polinomio $q(t)$ di quarto grado, da cui si ottiene

$$p_7(x) = 2x + 2x^3 q\Big(\frac{x}{\delta}\Big) = 2x + 0.66666\,x^3 + 0.39989\,x^5 + 0.29553\,x^7,$$

e risulta

$$\max_{|x|\le\delta} |f(x) - p_7(x)| \approx 0.189 \times 10^{-8},\ \max_{|x|\le\delta} \Big|\frac{f(x) - p_7(x)}{f(x)}\Big| \approx 0.525 \times 10^{-8}.$$

In alternativa si determina il minimax polinomiale di grado 7 rispetto all'errore relativo, considerando un polinomio della forma

$$p_7(x) = 2x + \sum_{i=1}^{3} a_i x^{2i+1}.$$

Si scelgono come nodi iniziali 3 punti in $(0, \delta)$ e il punto δ. Si ottiene

$$p_7(x) = 2x + 0.66666\,x^3 + 0.39977\,x^5 + 0.29870\,x^7,$$

e risulta

$$\max_{|x|\le\delta} |f(x) - p_7(x)| \approx 0.283 \times 10^{-9}, \quad \max_{|x|\le\delta} \Big|\frac{f(x) - p_7(x)}{f(x)}\Big| \approx 0.863 \times 10^{-9}.$$

• Fra le funzioni elementari di solito si includono anche la **tangente** e l'**arcotangente**, ma per entrambe è opportuno ricorrere ad approssimazioni razionali, anziché polinomiali, che sarebbero poco efficaci perché di grado troppo elevato.