Linear approximation with Legendre polynomials

Approssimazione Costruttiva

Proserpio Lorenzo - N°928121

Giugno 2021

Si richiede di trovare la miglior approssimazione per le funzioni $f_\rho(x) = x^\rho$ con $\rho \in \mathbb{R}$ su $\Omega = (-1, 1)$, che siano anche in $\mathcal{L}^2(\Omega)$, tramite un'approssimazione lineare usando i primi n polinomi di Legendre. La relazione introdurrà prima lo spazio funzionale e dimostrerà che tali polinomi sono effettivamente un sistema ortogonale e completo per tale spazio, infine si studierà il comportamento dell'errore teorico e nella pratica al variare di n e dei ρ ammissibili.

Lo spazio $\mathcal{L}^2(\Omega)$

Tale spazio è definito come:

$$\mathcal{L}^2(\Omega) \coloneqq \left\{ f : (-1, 1) \to \mathbb{R} \;\middle|\; f \text{ misurabile },\; \int_{-1}^{1} |f|^2 < \infty \right\}$$

specifico che con funzione misurabile intendo misurabile secondo la misura di Lebesgue e che spesso si accetta anche come codominio $\mathbb{C}$, ma in questa relazione ci limiteremo a funzioni a valori reali. Talvolta ci si riferisce alla seconda proprietà che devono avere le f come "essere quadrato-sommabile". La norma "naturale" in questo spazio è la seguente:

$$\|f\|_{\mathcal{L}^2(\Omega)} = \left(\int_{-1}^{1} |f|^2 \right)^{1/2}$$

Si noti inoltre che è uno spazio lineare (con la somma definita come somma tra funzioni e la moltiplicazione per uno scalare definita in modo ovvio), questo discende dalla linearità dell'integrale, dalla disuguaglianza triangolare e dall'omogeneità della norma. Inoltre il teorema di Riesz-Fischer ci assicura che sia anche completo. Con tali presupposti abbiamo che è uno spazio di Banach, in realtà ha altre proprietà che vedremo.

ρ ammissibili

Affinchè f_ρ appartenga a $\mathcal{L}^2(\Omega)$ devono essere funzioni misurabili a valori reali e valere che:

$$\|f_\rho\|^2_{\mathcal{L}^2(\Omega)} < \infty$$

innanzitutto devono essere definite su $(-1, 1)$ il che ci porta a scartare tutti i $\rho \in \mathbb{R} \setminus \mathbb{Q}$. Affinchè siano misurabili e a valori reali basta che siano continue quasi ovunque in $(-1, 1)$, queste due condizioni ci portano a scartare tutti i ρ che abbiano, ridotti ai minimi termini, denominatore pari. Infine, affinchè sia rispettata la condizione sulla norma è necessario che $\rho > -1/2$. Dunque la famiglia dei ρ ammissibili è:

$$\mathscr{R} \coloneqq \left\{ \rho \in \mathbb{Q} \;\middle|\; \rho > -1/2 \text{ e } \rho = \frac{r}{2t+1} \text{ con } \; r \in \mathbb{N}, t \in \mathbb{Z} \right\}$$

La base dei polinomi di Legendre

I polinomi di Legendre $(P_k(x))_{k\in\mathbb{N}}$ sono polinomi in x di grado k a coefficienti reali definiti a partire dalla seguente funzione generatrice:

$$\frac{1}{\sqrt{1 - 2xz + z^2}} = \sum_{k=0}^{\infty} P_k(x) z^k$$

vi è anche un'altra formula equivalente per ricavare il k-esimo polinomio di Legendre chiamata formula di Rodrigues:

$$P_k(x) = \frac{1}{2^k k!} \frac{\mathrm{d}^k}{\mathrm{d}x^k} [(x^2 - 1)^k]$$

useremo alternativamente le due formulazioni qualora una risulti più pratica dell'altra per dimostrare una proprietà.

Lemma 1. *In $\mathcal{L}^2(\Omega)$ dotato del prodotto interno:*

$$\langle f, g\rangle = \int_{-1}^{1} fg \quad \textit{vale che} \quad \langle P_k, P_i\rangle = \frac{2}{2k+1}\delta_{ki}$$

Dimostrazione. La dimostrazione si articola in 3 passaggi:

1. $\exists r_p \in \mathbb{P}_p$ polinomio di grado p a coefficienti reali con $p \leq k$ tale che:

$$\frac{\mathrm{d}^p}{\mathrm{d}x^p}[(x^2-1)^k] = r_p(x)(x^2-1)^{k-p}$$

tale osservazione è facilmente verificabile per induzione su p. Infatti per $p = 1$ è banalmente verificata e supposto che valga per p:

$$\frac{\mathrm{d}^{p+1}}{\mathrm{d}x^{p+1}}[(x^2-1)^k] = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}^p}{\mathrm{d}x^p}[(x^2-1)^k]\right) = \frac{\mathrm{d}}{\mathrm{d}x}\left(r_p(x)(x^2-1)^{k-p}\right)$$

e l'RHS è un polinomio della forma $r_{p+1}(x)(x^2-1)^{k-p-1}$ con $r_{p+1}(x) = r_p'(x)(x^2-1) + 2xr_p(x)(k-p)$.

2. Mostriamo che per $k \neq m$ si ha che:

$$\int_{-1}^{1} P_k(x)P_m(x)\mathrm{d}x = 0$$

a tal fine possiamo assumere *WLOG* che $k > m$, con tale assunzione possiamo avere che $\frac{\mathrm{d}^m}{\mathrm{d}x^m}[P_m(x)] = c_m$ con c_m costante, mentre $\frac{\mathrm{d}^{k-m}}{\mathrm{d}x^{k-m}}[P_k(x)] = \frac{1}{k!2^k}r_{k-m}(x)(x^2-1)^m$. A questo punto la nostra tesi si ottiene integrando semplicemente per parti m volte e osservando che ad ogni passo il primo fattore è nullo (per quanto mostrato al punto 1.). Esattamente come segue:

$$k!2^k\int_{-1}^{1} P_kP_m = 0 - k!2^k\int_{-1}^{1}\frac{\mathrm{d}^{k-1}}{\mathrm{d}x^{k-1}}\left[(x^2-1)^k\right]P_m' = \cdots =$$

$$= (-1)^m \cdot c_m \cdot k!2^k\int_{-1}^{1}\frac{\mathrm{d}^{k-m}}{\mathrm{d}x^{k-m}}\left[(x^2-1)^k\right] = (-1)^m \cdot c_m\left[\frac{\mathrm{d}^{k-m-1}}{\mathrm{d}x^{k-m-1}}\left[(x^2-1)^k\right]\right]_{-1}^{1} =$$

$$= (-1)^m \cdot c_m\left[r_{k-m-1}(x)(x^2-1)^{m+1}\right]_{-1}^{1} = 0$$

3. Mostriamo infine che:

$$\int_{-1}^{1} P_k^2(x)\mathrm{d}x = \frac{2}{2k+1}$$

per fare cio' useremo la formula ($\star$) ricorsiva (dimostrata nella sezione *L'implementazione*):

$$kP_k(x) = (2k-1)xP_{k-1}(x) - (k-1)P_{k-2}(x)$$

Innanzitutto notiamo che per $P_0(x) = 1$ e $P_1(x) = x$ l'asserto vale. Poi procediamo per induzione su k:

$$\int_{-1}^{1} P_k^2 = \int_{-1}^{1} P_k \cdot \frac{2k-1}{k}xP_{k-1}$$

per l'ortogonalità tra P_k e P_{k-2} e la formula ($\star$). Ora riapplico ($\star$) su xP_k ed ottengo:

$$\int_{-1}^{1}\frac{2k-1}{k}xP_{k-1}P_k = \frac{2k-1}{2k+1}\int_{-1}^{1} P_{k-1}^2$$

per l'ortogonalità tra P_{k+1} e P_{k-1}. A questo punto uso l'ipotesi induttiva:

$$\int_{-1}^{1} P_k^2 = \frac{2k-1}{2k+1}\cdot\frac{2}{2k-1} = \frac{2}{2k+1}$$

□

Lemma 2. *Nelle ipotesi del lemma soprastante vale che:* $(b_k(x))_{k\in\mathbb{N}} := \left(\sqrt{\frac{2k+1}{2}}P_k(x)\right)_{k\in\mathbb{N}}$ *è una base ortonormale e completa di* $\mathcal{L}^2(\Omega)$

Dimostrazione. L'ortonormalità discende, in modo ovvio, dal Lemma 1. Che sia una base e sia completa deriva dal teorema di Weierstrass. Infatti, come abbiamo detto sopra, $P_k(x)$ è un polinomio a coefficienti reali di grado k e dunque vale lo stesso discorso per i $b_k(x)$. Abbiamo ora che vi è una biezione "naturale" tra $\mathcal{B} := \{b_0, b_1, b_2, ...\}$ e $\mathcal{B}_2 := \{1, x, x^2, ...\}$. Dunque siccome sappiamo da Weierstrass che $\mathcal{B}_2$ è una base completa per lo spazio delle funzioni continue e quadrato-sommabili su un intervallo limitato e chiuso lo è pure $\mathcal{B}$ per $\mathcal{L}^2(\Omega)$, sia perchè l'insieme delle funzioni continue è denso in $\mathcal{L}^2(\overline{\Omega})$ sia perchè in norma $\mathcal{L}^2$ non "importa" cosa succede su insiemi di misura nulla e quindi è valido pure sull'aperto $(-1, 1)$. □

I due lemmi sopra ci assicurano che $\mathcal{L}^2(\Omega)$, con $\langle \cdot, \cdot \rangle$ definito come sopra, sia uno spazio di Hilbert $\mathcal{H}$ *separabile*, di cui $\mathcal{B}$ è un sistema completo. Dunque si ha che:

$$\forall f \in \mathcal{H} \Longrightarrow f = \sum_{k=0}^{\infty} \langle f, b_k \rangle b_k = \sum_{k=0}^{\infty} f_k b_k \ \text{ in } \mathcal{H}$$

dove la convergenza è rispetto alla norma $\|\cdot\|_{\mathcal{L}^2(\Omega)}$, cioè significa che f e il suo sviluppo in serie con i polinomi di Legendre possono differire al più su insiemi di misura nulla.

L'implementazione

Al fine di implementare in MATLAB i polinomi di Legendre si è usata la seguente proprietà:

Lemma 3. *I polinomi di Legendre $(b_k(x))_{k \in \mathbb{N}}$ soddisfano la seguente equazione ricorsiva* $(\star)$*:*

$$(k + 1)b_{k+1}(x) = (2k + 1)xb_k(x) - kb_{k-1}(x)$$

Dimostrazione. Derivo la funzione generatrice dei polinomi di Legendre rispetto a z a destra e sinistra e ottengo:

$$(x - z) \cdot \frac{1}{\sqrt{1 - 2xz + z^2}} = \left(1 - 2xz + z^2\right) \sum_{k=1}^{\infty} kP_k(x)z^{k-1}$$

ora sostituisco il termine a sinistra con la sua definizione e cambio gli indici del termine a destra:

$$(x - z) \sum_{k=0}^{\infty} P_k(x)z^k = \left(1 - 2xz + z^2\right) \sum_{k=0}^{\infty} (k + 1)P_{k+1}(x)z^k =$$

$$\Downarrow$$

$$\sum_{k=0}^{\infty} xP_k(x)z^k - \sum_{k=1}^{\infty} P_{k-1}(x)z^k = \sum_{k=0}^{\infty} (k + 1)P_{k+1}(x)z^k - \sum_{k=1}^{\infty} 2kxP_k(x)z^k + \sum_{k=2}^{\infty} (k - 1)P_{k-1}(x)z^k$$

uguagliando i rispettivi coefficienti e ricordando che $P_0(x) = 1$ e $P_1(x) = x$ si ottiene l'asserto. □

Utilizzando la formula $(\star)$, anzichè la formula di Rodrigues, si minimizza l'errore di troncamento dovuto alla precisione di macchina. Per calcolare i coefficienti $\langle f, b_k \rangle$ abbiamo implementato la quadratura col metodo dei trapezi su una griglia a passo costante h facendo attenzione a non avere nessun nodo in 0 (basta usare il trucco di scegliere i nodi in numero pari), siccome è punto di limite per alcune f_ρ. Inoltre sfruttiamo pure il fatto che i b_k sono alternatamente funzioni pari o dispari, quindi $\langle f_\rho, b_k \rangle = 0$ se, $\left(\text{ricordando che } \rho \text{ è del tipo } \frac{r}{2t+1}\right)$, r e k hanno diversa parità, questo piccolo trucco ci risparmierà dei conti e renderà anche il risultato più preciso. I file sono *Legendre.m* che contiene il *main*, *LegPol.m* che è una funzione che calcola i primi n polinomi di Legendre valutati in un punto x e restituisce dunque un vettore $1 \times k$ e *q_ trapezi* che non è altro che il metodo di quadratura con i trapezi. Discutiamo ora in dettaglio le operazioni che svolge *Legendre.m*. Nell'ordine: definisce la *mesh* e imposta il ρ nella forma $\frac{r}{2t+1}$ e il vettore dei casi da valutare; dichiara la funzione f_ρ, in questo caso abbiamo dovuto utilizzare la funzione *nthroot* nei casi in cui ρ non sia intero altrimenti MATLAB prende le radici complesse piuttosto che quelle reali; ad ogni passo calcola la matrice contente i primi n polinomi di Legendre valutati nei punti di *mesh*; valuta f_ρ nei punti di *mesh*; calcola mediante quadratura i coefficienti $\langle f_\rho, b_k \rangle$; calcola la serie troncata di Legendre valutata nei nodi di *mesh*; plot delle soluzioni (sconsigliato se si valuta un grande numero di casi); calcolo e plot dell'errore in norma $\mathcal{L}^2(\Omega)$; calcolo e print dell'EOC.

Analisi teorica dell'errore

Tenendo tutte le notazioni usate in precedenza definiamo $\forall n \geq 1$:

$$\mathcal{S}_n := \text{span}\{b_0, ..., b_{n-1}\} \quad \text{e} \quad \mathcal{E}_n(f) := \inf_{s \in \mathcal{S}_n} \|f - s\|_{\mathcal{L}^2(\Omega)}$$

Notiamo che tale insieme è lineare (cioè $\alpha\mathcal{S}_n = \mathcal{S}_n$ e $\mathcal{S}_n + \mathcal{S}_n = \mathcal{S}_n$). Inoltre, dalla disuguaglianza di Bessel sappiamo che:

$$(\star\star) \quad \mathcal{E}_n(f) = \left(\sum_{k=n}^{\infty} |\langle f, b_k \rangle|^2 \right)^{1/2}$$

Vedremo che la velocità di convergenza dell'approssimazione lineare con i primi n polinomi di Legendre dipende dall'appartenenza di f_ρ ad un determinato spazio di Sobolev. In particolare ci occupiamo di questi spazi:

$$\mathcal{W}^{q,2}(\Omega) := \{f \in \mathcal{L}^2(\Omega) \mid f^{(1)}, ..., f^{(q)} \text{ esistono e sono in } \mathcal{L}^2(\Omega)\}$$

con $f^{(j)}$ intendo la j-esima derivata debole di f. Mostriamo ora il seguente lemma:

Lemma 4. *Se $f_\rho \in \mathcal{W}^{q,2}(\Omega)$ con $q < \infty$ allora $\forall k > q$ si ha che $|\langle f_\rho, b_k\rangle|^2 \leq C_{k,\rho}^2 \cdot \dfrac{2k+1}{2} \cdot \dfrac{1}{(2^k k!)^2}$ con $C_{k,\rho} > 0$ costante.*

Dimostrazione. Userò la formula di Rodrigues e l'integrazione per parti q volte:

$$\sqrt{\frac{2k+1}{2}} \int_\Omega f_\rho P_k = (-1)^q \cdot \sqrt{\frac{2k+1}{2}} \cdot \frac{1}{2^k k!} \int_\Omega f_\rho^{(q)} \frac{\mathrm{d}^{k-q}}{\mathrm{d}x^{k-q}} \left[(x^2-1)^k\right]$$

ora abbiamo che $k > k - q > 0$ e quindi possiamo usare la stessa formula del punto 1. del **Lemma 1.**; prima però esplicitiamo la derivata q-esima di f_ρ, notando che la f_ρ ha problemi di derivata debole in $\mathcal{L}^2(\Omega)$ solo se l'esponente di x diventa minore di 0 e dunque, esplicitandola, l'integrale diventa in modulo:

$$\sqrt{\frac{2k+1}{2}} \cdot \frac{1}{2^k k!} \cdot \prod_{i=0}^{q-1}(\rho - i) \int_\Omega x^{\rho-q} r_{k-q}(x)(x^2-1)^q \leq C_{k,\rho} \cdot \sqrt{\frac{2k+1}{2}} \cdot \frac{1}{2^k k!}$$

siccome da integrare resta una funzione in $\mathcal{L}^1(\Omega)$ e dunque l'integrale è una certa costante e la produttoria è una costante pure lei. ☐

Dal lemma, con le ipotesi come sopra, emerge che gli f_k^2 con $k > q$ sono strettamente decrescenti e decadono più velocemente di una qualsiasi funzione del tipo k^{-2r}. Possiamo dunque stimare (usando la disuguaglianza triangolare e la concavità della funzione radice quadrata):

$$\left(\sum_{k=0}^{\infty} |\langle f_\rho, b_k\rangle|^2\right)^{1/2} \leq \left[\sum_{k=0}^{q}\left(\|f_\rho^{(k)}\|^4_{\mathcal{L}^2(\Omega)} + \|b_k\|^4_{\mathcal{L}^2(\Omega)}\right)\right]^{1/2} + \left[\sum_{k=q+1}^{\infty} C_{k,\rho}^2 \cdot \frac{2k+1}{2} \cdot \frac{1}{(2^k k!)^2}\right]^{1/2}$$

vediamo che bastano i primi q (o $q+1$ o $q-1$) termini per avere un'approssimazione più che buona. Notiamo inoltre che se $f_\rho \in \mathcal{C}^\infty(\Omega)$ (è un polinomio), vale, facendo una dimostrazione simile al **Lemma 4.**, che tutti i coefficienti f_k con $k > \deg(f_\rho)$ sono 0 (per il semplice fatto che, una volta derivato $(k+1)$-volte tale polinomio diventa costante zero...). I casi differenti da prendere in esame (per $n \to \infty$) sono:

- $-1/2 < \rho < 1/2$ (tolto il caso $\rho = 0$): non appartiene a nessun $\mathcal{W}^{q,2}(\Omega)$ l'errore decadrà più lentamente di $\mathcal{O}(n^{-1})$;
- $\rho \in \mathbb{N}$: sono polinomi, la maggior parte dell'errore decadrà in corrispondenza della serie troncata il cui n-esimo polinomio di Legendre ha grado pari al grado di f_ρ dopodichè, la nostra funzione è perfettamente rappresentata e l'errore è 0;
- $\rho > 1/2$ con $\rho \in \mathbb{Q} \setminus \mathbb{N}$: hanno esattamente $\lfloor \rho + 1/2 \rfloor$ derivate deboli in $\mathcal{L}^2$, dunque stanno in $\mathcal{W}^{q,2}(\Omega)$ con $q = \lfloor \rho + 1/2 \rfloor$, l'errore decadrà come $\mathcal{O}(n^{-\lfloor \rho + 1/2 \rfloor})$.

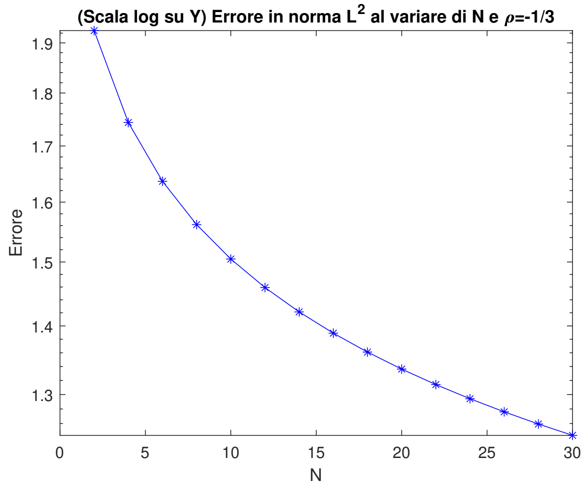
Risultati dell'implementazione

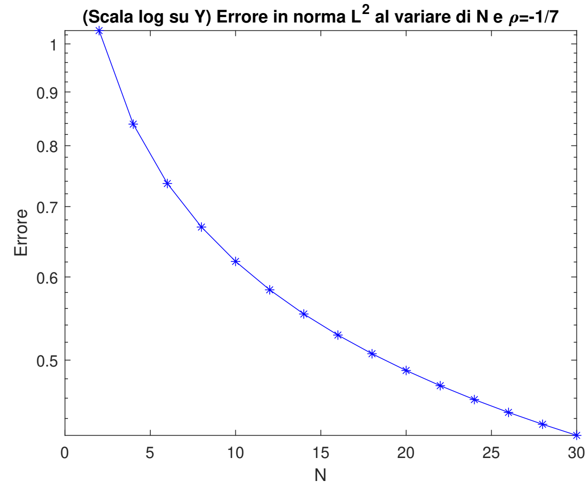
La formula $(\star\star)$, sebbene desiderabile, non puo' essere implementata sul calcolatore (a causa del numero infinito di termini). Affinchè sia computabile, l'errore verrà valutato (con una formulazione equivalente) come la norma $\mathcal{L}^2$ della differenza tra f_ρ e la sua approssimazione lineare a n termini, cioè:

$$\mathcal{E}_n(f) = \left(\int_{-1}^{1} \left| f_\rho - \sum_{k=0}^{n-1} \langle f, b_k\rangle b_k \right|^2\right)^{1/2}$$

nella pratica tale quantità verrà calcolata sempre per quadratura mediante trapezi sulla griglia definita in precedenza. Abbiamo anche fatto attenzione che la *mesh* risulti sufficientemente fitta affinchè l'errore nella quadratura (quando possibile, cioè non nel caso $\rho \in \mathbb{N}$) non "inquini" l'errore dell'approssimazione. A seguire vi sono dei grafici degli errori, in scala logaritmica sulle ordinate, al variare di N per le casistiche viste nella parte teorica. Inoltre per il terzo caso ($\rho > 0$ con $\rho \in \mathbb{Q} \setminus \mathbb{N}$) vi è pure un confronto con la funzione $cost \cdot N^{-\lfloor \rho + 1/2 \rfloor}$ dove la costante è stata stimata con il rapporto tra errore per N_{finale} e la stessa quantità elevata alla $(-\lfloor \rho + 1/2 \rfloor)$. Hanno tutti rispettato le nostre previsioni teoriche.

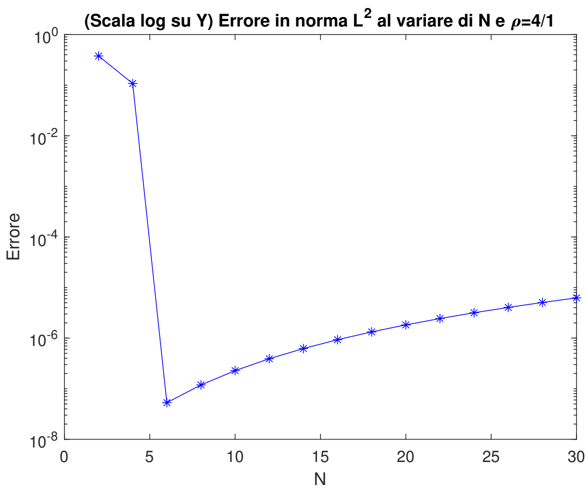
Caso 1: $-1/2 < \rho < 1/2$

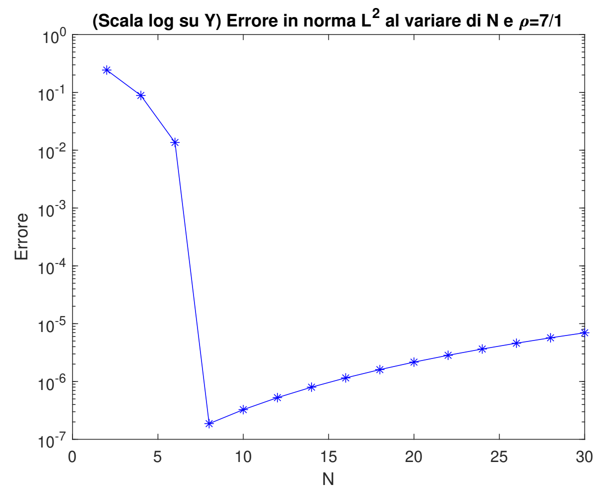




Andando a vedere L'EOC calcolato si vede che è sempre molto minore di 1, dunque la convergenza è effettivamente più lenta di $\mathcal{O}(n^{-1})$.

Caso 2: $\rho \in \mathbb{N}$





Si vede che il risultato migliore ce l'abbiamo nel caso in cui ci fermiamo al termine n-esimo di grado pari al grado di f_ρ dopodichè l'errore tende ad aumentare (la colpa è l'*underflow* dei coefficienti e dei limiti dell'integrazione con i trapezi), in realtà è un problema facilmente risolvibile mettendo come condizione che se $\rho \in \mathbb{N}$ i coefficienti f_k per $k > \deg(f_\rho)$ devono essere 0.

Caso 3: $\rho > 0$ **con** $\rho \in \mathbb{Q} \setminus \mathbb{N}$