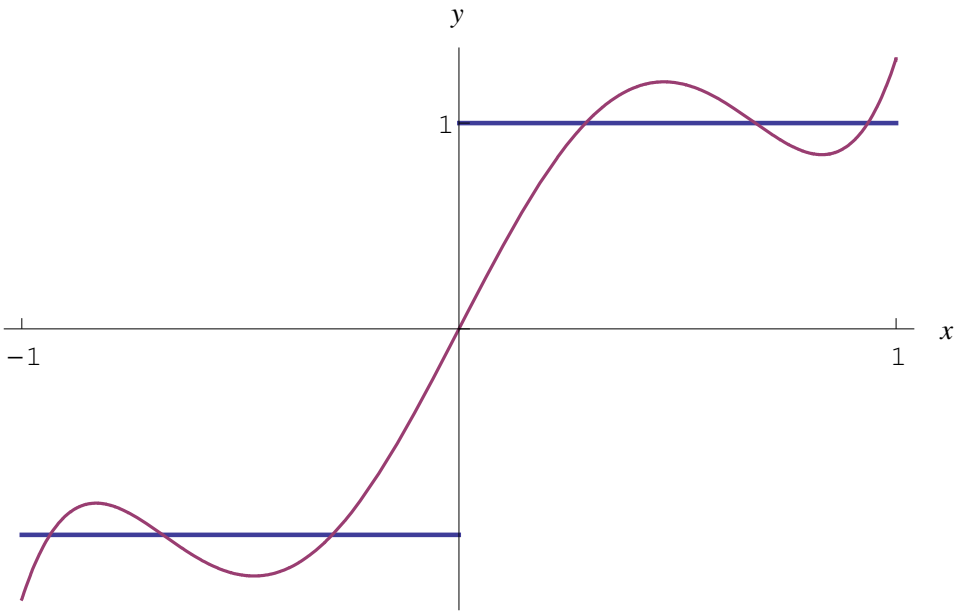
SCIENTIA – http://www.scientiajournal.orf
International Review of Scientific Synthesis – ISSN 2282-2119
Monografia 00x – 2015
MATEMATICA OPEN SOURCE – HTTP://WWW.EXTRABYTE.INFO

Il miglior polinomio approssimante

Marcello Colozzo



Sommario

Nell'approssimazione di una funzione mediante un sistema di polinomi $\{\varphi_k\}$ linearmente indipendenti, è necessario minimizzare l'errore qudratico medio:

$$\left\langle \varepsilon_n^2 \right\rangle = \frac{1}{b-a}\int_a^b \left[f(x) - \sum_{k=1}^{n} a_k \varphi_k(x) \right]^2 dx,$$

che risulta essere una funzione reale delle n variabili reali $a_1, a_2, ..., a_n$. Come è noto, la ricerca del minimo assoluto implica lo studio della forma quadratrica nelle variabili ausiliarie $\lambda_1, \lambda_2, ..., \lambda_n$:

$$\phi(\lambda_1, \lambda_2, ..., \lambda_n) = \sum_{h=1}^{n}\sum_{k=1}^{n} \left. \frac{\partial^2 \left\langle \varepsilon_n^2 \right\rangle}{\partial a_h \partial a_k} \right|_{\tilde{P}} \lambda_h \lambda_k,$$

nei punti critici $\tilde{P}$, cioè tali che $\nabla \left\langle \varepsilon_n^2 \right\rangle \big|_{\tilde{P}} = 0$. Per $n > 2$ lo studio di tale forma quadratica risulta difficoltoso. In questa monografia proponiamo un efficiente algoritmo basato sull'algebra delle matrici.

Sia f un qualunque elemento dello spazio funzionale $C([a,b])$ delle funzioni continue in $[a,b] \subseteq \mathbb{R}$, dotato di prodotto scalare definito da:

$$\langle f, g \rangle = \int_a^b f(x)\, g(x)\, dx, \quad \forall f, g \in C([a,b]) \tag{1}$$

Ci proponiamo di determinare il miglior polinomio approssimante f nel senso dei minimi quadrati. Per essere più specifici, dopo aver assegnato un ordine di approssimazione $n \in \mathbb{N} - \{0,1\}$, "costruiamo" un sistema linearmente indipendente di polinomi $\{\varphi_1, \varphi_2, ..., \varphi_n\}$ quali elementi di $C([a,b])$. Come è noto, il metodo dei minimi quadrati consiste nel minimizzare l'errore quadratico medio. In simboli:

$$f \sim \tau_n = \sum_{k=1}^{n} a_k \varphi_k \Longleftrightarrow \left\langle \varepsilon_n^2 \right\rangle = \frac{1}{b-a} ||f - \tau_n||^2 \text{ assume un minimo assoluto}$$

Sviluppando $||f - \tau_n||^2$ secondo la (1) si perviene:

$$\left\langle \varepsilon_n^2 \right\rangle = \frac{1}{b-a} \left(||f||^2 - 2 \sum_{k=1}^{n} a_k c_k + \sum_{k=1}^{n} \sum_{k'=1}^{n} a_k a_{k'} \langle \varphi_k, \varphi_{k'} \rangle \right), \tag{2}$$

onde $\left\langle \varepsilon_n^2 \right\rangle$ risulta essere una funzione reale delle n variabili reali. In questa equazione i numeri reali c_k sono i coefficienti di Fourier:

$$c_k = \int_a^b \varphi_k(x)\, f(x)\, dx \qquad (k = 1, 2, ..., n) \tag{3}$$

Applichiamo, dunque, il procedimento standard per la ricerca degli estremi assoluti della funzione reale $\left\langle \varepsilon_n^2 \right\rangle$ delle n variabili reali $a_1, a_2, ..., a_n$. Si ricordi che tale procedimento è basato su un teorema che richiede la continuità di $\left\langle \varepsilon_n^2 \right\rangle$ e delle sue derivate parziali seconde nei punti interni del dominio di definizione. Dalla (2) vediamo che $\left\langle \varepsilon_n^2 \right\rangle$ è di classe C^∞ su $\mathbb{R}^n$, per cui possiamo applicare il suddetto teorema. Iniziamo a determinare i punti estremali (o punti critici) della funzione. A tale scopo, riscriviamo la (2) come:

$$\left\langle \varepsilon_n^2 \right\rangle (\mathbf{a}) = \frac{1}{b-a} \left(||f||^2 - 2 \sum_{k=1}^{n} a_k c_k + \sum_{k=1}^{n} \sum_{k'=1}^{n} a_k a_{k'} \langle \varphi_k, \varphi_{k'} \rangle \right), \tag{4}$$

dove $\mathbf{a} = (a_1, a_2, ..., a_n) \in \mathbb{R}^n$. In altri termini, i numeri reali a_k sono le componenti di un vettore di $\mathbb{R}^n$ nella base canonica $\{\mathbf{e}_k\}$ di tale spazio vettoriale. Le coordinate dei punti critici della funzione (4) sono le soluzioni dell'equazione vettoriale:

$$\nabla \left\langle \varepsilon_n^2 \right\rangle = \mathbf{0}, \tag{5}$$

ovvero del sistema di n equazioni scalari nelle n incognite $a_1, a_2, ..., a_n$:

$$\frac{\partial \left\langle \varepsilon_n^2 \right\rangle}{\partial a_h} = 0 \qquad (h = 1, 2, ..., n) \tag{6}$$

Riesce:

$$\frac{\partial \left\langle \varepsilon_n^2 \right\rangle}{\partial a_h} = \frac{1}{b-a} \left[-2 \sum_{k=1}^{n} \delta_{kh} c_k + \sum_{k=1}^{n} \sum_{k'=1}^{n} \langle \varphi_k, \varphi_{k'} \rangle \frac{\partial}{\partial a_k} (a_k a_{k'}) \right]$$

Calcoliamo a parte la doppia sommatoria a secondo membro. Osservando che:

$$\frac{\partial}{\partial a_k}\left(a_k a_{k'}\right) = \delta_{kh} a_{k'} + a_k \delta_{k'h},$$

si ha:

$$\begin{aligned}
&\sum_{k=1}^{n}\sum_{k'=1}^{n} \langle \varphi_k, \varphi_{k'} \rangle \frac{\partial}{\partial a_k}\left(a_k a_{k'}\right) \\
&= \sum_{k=1}^{n}\sum_{k'=1}^{n} \left(\langle \varphi_k, \varphi_{k'} \rangle \delta_{kh} a_{k'} + \langle \varphi_k, \varphi_{k'} \rangle a_k \delta_{k'h} \right)
\end{aligned}$$

Scambiando le sommatorie nel primo termine a secondo membro:

$$\sum_{k=1}^{n}\sum_{k'=1}^{n} \langle \varphi_k, \varphi_{k'} \rangle \frac{\partial}{\partial a_k}\left(a_k a_{k'}\right) = \sum_{k'=1}^{n}\sum_{k=1}^{n} \langle \varphi_k, \varphi_{k'} \rangle \delta_{kh} a_{k'} + \sum_{k=1}^{n}\sum_{k'=1}^{n} \langle \varphi_k, \varphi_{k'} \rangle a_k \delta_{k'h} \tag{7}$$

δ_{kh} cancella nella sommatoria su k tutti e soli i termini con $k \neq h$:

$$\sum_{k'=1}^{n}\sum_{k=1}^{n} \langle \varphi_k, \varphi_{k'} \rangle \delta_{kh} a_{k'} = \sum_{k'=1}^{n} \langle \varphi_h, \varphi_{k'} \rangle a_{k'}$$

Trattandosi di un indice muto, scriviamo:

$$\sum_{k'=1}^{n}\sum_{k=1}^{n} \langle \varphi_k, \varphi_{k'} \rangle \delta_{kh} a_{k'} = \sum_{k=1}^{n} \langle \varphi_h, \varphi_k \rangle a_k$$

Passiamo al secondo termine (a secondo membro della (7)):

$$\sum_{k=1}^{n}\sum_{k'=1}^{n} \langle \varphi_k, \varphi_{k'} \rangle a_k \delta_{k'h} = \sum_{k=1}^{n} \langle \varphi_k, \varphi_h \rangle a_k \underset{\langle \varphi_k, \varphi_h \rangle = \langle \varphi_h, \varphi_k \rangle}{=} \sum_{k=1}^{n} \langle \varphi_h, \varphi_k \rangle a_k$$

Finalmente:

$$\sum_{k=1}^{n}\sum_{k'=1}^{n} \langle \varphi_k, \varphi_{k'} \rangle \frac{\partial}{\partial a_k}\left(a_k a_{k'}\right) = 2\sum_{k=1}^{n} \langle \varphi_h, \varphi_k \rangle a_k,$$

per cui

$$\frac{\partial \left\langle \varepsilon_n^2 \right\rangle}{\partial a_h} = \frac{2}{b-a}\left(-c_h + \sum_{k=1}^{n} \langle \varphi_h, \varphi_k \rangle a_k \right) \tag{8}$$

Pertanto il gradiente di $\langle \varepsilon_n^2 \rangle$ è:

$$\begin{aligned}
\nabla \left\langle \varepsilon_n^2 \right\rangle &= \frac{2}{b-a}\sum_{h=1}^{n}\left(-c_h + \sum_{k=1}^{n} \langle \varphi_h, \varphi_k \rangle a_k \right)\mathbf{e}_h \qquad (9) \\
&= \frac{2}{b-a}\left(-\sum_{h=1}^{n} c_h \mathbf{e}_h + \sum_{h=1}^{n}\sum_{k=1}^{n} \langle \varphi_h, \varphi_k \rangle a_k \mathbf{e}_h \right) \\
&= \frac{2}{b-a}\left(-\mathbf{c} + \sum_{h=1}^{n}\sum_{k=1}^{n} \langle \varphi_h, \varphi_k \rangle a_k \mathbf{e}_h \right),
\end{aligned}$$

essendo

$$\mathbf{c} = \sum_{h=1}^{n} c_h \mathbf{e}_h \tag{10}$$

il *vettore di Fourier* della funzione f secondo la base $\{\varphi_k\}$ di V_n, cioè il vettore le cui componenti nella base canonica di $\mathbb{R}^n$ sono le coordinate di Fourier di f. Esplicitiamo la doppia sommatoria nell'ultimo termine della (9):

$$\sum_{h=1}^{n}\sum_{k=1}^{n}\langle\varphi_h,\varphi_k\rangle\, a_k\mathbf{e}_h = \sum_{h=1}^{n} b_h\mathbf{e}_h,$$

dove:

$$b_h = \sum_{k=1}^{n}\langle\varphi_h,\varphi_k\rangle\, a_k \qquad (h=1,2,...,n)$$

sono le componenti di un vettore di $\mathbb{R}^n$ che denotiamo con $\mathbf{b}$, onde:

$$\nabla\left\langle\varepsilon_n^2\right\rangle = \frac{2}{b-a}\left(\mathbf{b}-\mathbf{c}\right) \tag{11}$$

Per quanto detto:

$$\mathbf{b} = \sum_{h=1}^{n}\left(\sum_{k=1}^{n}\langle\varphi_h,\varphi_k\rangle\, a_k\right)\mathbf{e}_h$$

Da ciò vediamo che $\mathbf{b}$ è il risultato dell'applicazione di un endomorfismo A:

$$\hat{A}\mathbf{a} = \mathbf{b},$$

la cui matrice rappresentativa nella base canonica $\{\mathbf{e}_k\}$ di $\mathbb{R}^n$ è:

$$A = \begin{pmatrix} \langle\varphi_1,\varphi_1\rangle & \langle\varphi_1,\varphi_2\rangle & ... & \langle\varphi_1,\varphi_n\rangle \\ \langle\varphi_2,\varphi_1\rangle & \langle\varphi_2,\varphi_2\rangle & ... & \langle\varphi_2,\varphi_n\rangle \\ ... & ... & ... & ... \\ \langle\varphi_n,\varphi_1\rangle & \langle\varphi_n,\varphi_2\rangle & ... & \langle\varphi_n,\varphi_n\rangle \end{pmatrix}$$

Simbolicamente:

$$\hat{A} \doteq A,$$

dove il simbolo $\doteq$ denota "rappresentato da". In maniera analoga:

$$\mathbf{a} \doteq X = \begin{pmatrix} a_1 \\ a_2 \\ ... \\ a_n \end{pmatrix}$$

Riesce:

$$\hat{A}\mathbf{a} \doteq AX = \begin{pmatrix} \sum_{k=1}^{n}\langle\varphi_1,\varphi_k\rangle\, a_k \\ \sum_{k=1}^{n}\langle\varphi_2,\varphi_k\rangle\, a_k \\ ... \\ \sum_{k=1}^{n}\langle\varphi_n,\varphi_k\rangle\, a_k \end{pmatrix}$$

In tal modo il gradiente di G si esprime attraverso l'azione dell'endomorfismo $\hat{A}$ sul vettore $\mathbf{a}$:

$$\nabla\left\langle\varepsilon_n^2\right\rangle = \frac{2}{b-a}\left(\hat{A}\mathbf{a}-\mathbf{c}\right), \tag{12}$$

per cui l'equazione vettoriale (5) equivale alla seguente equazione operatoriale:

$$\hat{A}\mathbf{a} = \mathbf{c}, \tag{13}$$

che nella base canonica di $\mathbb{R}^n$ si traduce nell'equazione matriciale:

$$AX = C, \tag{14}$$

dove

$$C = \begin{pmatrix} c_1 \\ c_2 \\ ... \\ c_n \end{pmatrix}$$

Quindi la (13) scritta nella base canonica conduce al sistema di equazioni lineari:

$$\begin{cases} \langle \varphi_1, \varphi_1 \rangle a_1 + \langle \varphi_1, \varphi_2 \rangle a_2 + ... + \langle \varphi_1, \varphi_n \rangle a_n = c_1 \\ \langle \varphi_2, \varphi_1 \rangle a_1 + \langle \varphi_2, \varphi_2 \rangle a_2 + ... + \langle \varphi_2, \varphi_n \rangle a_n = c_2 \\ ... \\ \langle \varphi_n, \varphi_1 \rangle a_1 + \langle \varphi_n, \varphi_2 \rangle a_2 + ... + \langle \varphi_n, \varphi_n \rangle a_n = c_n \end{cases} \tag{15}$$

In altri termini, le coordinate dei punti estremali di $\langle \varepsilon_n^2 \rangle$ sono le soluzioni del sistema (15) che è un sistema di n equazioni lineari nelle n incognite $a_1, a_2, ..., a_n$, di coefficienti i prodotti scalari $\langle \varphi_h, \varphi_k \rangle$ e i cui termini noti sono i coefficienti di Fourier della funzione assegnata. Per il teorema **??** la funzione f è univocamente determinata dai suoi coefficienti di Fourier $(c_1, c_2, ..., c_n)$. In particolare:

$$c_1 = c_2 = ... = c_n = 0 \Longleftrightarrow f(x) = 0, \ \ \forall x \in [a, b]$$

In altri termini, il sistema (15) è omogeneo se e solo se f è la funzione identicamente nulla in $[a, b]$. Inoltre, comunque prendiamo $f \in C([a, b])$, la matrice dei coefficienti del sistema si scrive:

$$A = \begin{pmatrix} \langle \varphi_1, \varphi_1 \rangle & \langle \varphi_1, \varphi_2 \rangle & ... & \langle \varphi_1, \varphi_n \rangle \\ \langle \varphi_2, \varphi_1 \rangle & \langle \varphi_2, \varphi_2 \rangle & ... & \langle \varphi_2, \varphi_n \rangle \\ ... & ... & ... & ... \\ \langle \varphi_n, \varphi_1 \rangle & \langle \varphi_n, \varphi_2 \rangle & ... & \langle \varphi_n, \varphi_n \rangle \end{pmatrix},$$

che è la matrice rappresentativa di $\hat{A} \in End(\mathbb{R}^n)$ nella base canonica di $\mathbb{R}^n$. Per chiarezza riassumiamo i risultati raggiunti. Assegnato il sistema linearmente indipendente $\{\varphi_1, \varphi_2, ..., \varphi_n\} \subset C([a, b])$ mediante il quale vogliamo approssimare la nostra funzione $f \in C([a, b])$, è univocamente determinato un endomorifismo $\hat{A} \in End(\mathbb{R}^n)$ la cui matrice rappresentativa nella base canonica è $\hat{A} \doteq A = (\langle \varphi_h, \varphi_k \rangle)$. Posto

$$\left\langle \varepsilon_n^2 \right\rangle (\mathbf{a}) \overset{def}{=} \left\langle \varepsilon_n^2 \right\rangle = \frac{1}{b-a} \left\| f - \sum_{k=1}^{n} a_k \varphi_k \right\|,$$

si ha $\nabla \langle \varepsilon_n^2 \rangle = \frac{2}{b-a} \left(\hat{A}\mathbf{a} - \mathbf{c} \right)$, essendo $\mathbf{c} = \Sigma_k \langle \varphi_k, f \rangle \mathbf{e}_k$. Possono verificarsi i seguenti casi:

1. $\det A \neq 0$, f non è identicamente nulla in $[a, b]$.
2. $\det A = 0$, f non è identicamente nulla in $[a, b]$.
3. f è identicamente nulla in $[a, b]$.

Nel caso 1 il sistema (15) è normale e la sua unica soluzione si ottiene applicando il teorema di Cramer:

$$\exists! \left(\tilde{a}_1, \tilde{a}_2, ..., \tilde{a}_n\right) \in \mathbb{R}^n \mid \tilde{a}_k = \frac{\Delta_k}{\det A},$$

dove Δ_k è il determinante della matrice quadrata ricavata da A sostituendo la colonna *k-esima* con quella dei termini noti. Ne consegue che nel caso 1 esiste un solo punto critico $\tilde{P}\left(\tilde{a}_1, \tilde{a}_2, ..., \tilde{a}_n\right)$. Nel caso 2 il sistema è non normale e la condizione di compatibilità è espressa dal teorema di Rouchè-Capelli. Precisamente, denotando con B la matrice dei coefficienti e dei termini noti:

$$B = \begin{pmatrix} \langle \varphi_1, \varphi_1 \rangle & \langle \varphi_1, \varphi_2 \rangle & ... & \langle \varphi_1, \varphi_n \rangle & c_1 \\ \langle \varphi_2, \varphi_1 \rangle & \langle \varphi_2, \varphi_2 \rangle & ... & \langle \varphi_2, \varphi_n \rangle & c_2 \\ ... & ... & ... & ... & ... \\ \langle \varphi_n, \varphi_1 \rangle & \langle \varphi_n, \varphi_2 \rangle & ... & \langle \varphi_n, \varphi_n \rangle & c_n \end{pmatrix},$$

si ha che il sistema (15) è compatibile se e solo se $rango\,(A) = rango\,(B)$. In tal caso il rango del sistema è $p = rango\,(A) = rango\,(B) < n$, per cui esistono ∞^{n-p} soluzioni. Senza perdita di generalità supponiamo che sia:

$$\begin{vmatrix} \langle \varphi_1, \varphi_1 \rangle & \langle \varphi_1, \varphi_2 \rangle & ... & \langle \varphi_1, \varphi_p \rangle \\ \langle \varphi_2, \varphi_1 \rangle & \langle \varphi_2, \varphi_2 \rangle & ... & \langle \varphi_2, \varphi_p \rangle \\ ... & ... & ... & ... \\ \langle \varphi_p, \varphi_1 \rangle & \langle \varphi_p, \varphi_2 \rangle & ... & \langle \varphi_p, \varphi_p \rangle \end{vmatrix} \neq 0,$$

per cui il sistema (15) è equivalente a

$$\begin{cases} \langle \varphi_1, \varphi_1 \rangle a_1 + \langle \varphi_1, \varphi_2 \rangle a_2 + ... + \langle \varphi_1, \varphi_p \rangle a_p = c_1 - \left(\langle \varphi_1, \varphi_{p+1} \rangle a_{p+1} + ... + \langle \varphi_1, \varphi_n \rangle a_n\right) \\ \langle \varphi_2, \varphi_1 \rangle a_1 + \langle \varphi_2, \varphi_2 \rangle a_2 + ... + \langle \varphi_2, \varphi_p \rangle a_p = c_2 - \left(\langle \varphi_2, \varphi_{p+1} \rangle a_{p+1} + ... + \langle \varphi_2, \varphi_n \rangle a_n\right) \\ ... \\ \langle \varphi_p, \varphi_1 \rangle a_1 + \langle \varphi_p, \varphi_2 \rangle a_2 + ... + \langle \varphi_p, \varphi_p \rangle a_p = c_n - \left(\langle \varphi_2, \varphi_{p+1} \rangle a_{p+1} + ... + \langle \varphi_2, \varphi_n \rangle a_n\right) \end{cases}, \tag{16}$$

che si risolve con il teorema di Cramer, assumendo le $n-p$ incognite $a_{p+1}, ..., a_n$ come parametri. Ne consegue che nel caso 2 esistono infiniti punti critici. Infine, nel caso 3 il sistema è omogeneo:

$$\begin{cases} \langle \varphi_1, \varphi_1 \rangle a_1 + \langle \varphi_1, \varphi_2 \rangle a_2 + ... + \langle \varphi_1, \varphi_n \rangle a_n = 0 \\ \langle \varphi_2, \varphi_1 \rangle a_1 + \langle \varphi_2, \varphi_2 \rangle a_2 + ... + \langle \varphi_2, \varphi_n \rangle a_n = 0 \\ ... \\ \langle \varphi_n, \varphi_1 \rangle a_1 + \langle \varphi_n, \varphi_2 \rangle a_2 + ... + \langle \varphi_n, \varphi_n \rangle a_n = 0 \end{cases} \tag{17}$$

Se $\det A \neq 0$ il sistema (17) ammette la sola soluzione banale $a_1 = a_2 = ... = a_n = 0$. Se $\det A = 0$, detto p il rango di A, i.e. rango del sistema, si ha che (17) ammette ∞^{n-p} autosoluzioni (cioè soluzioni non nulle).

Da tale analisi emerge che la condizione $\det A \neq 0$ è vitale per l'autoconsistenza del metodo di approssimazione che stiamo elaborando. Infatti, se f è identicamente nulla in $[a, b]$, la funzione (2) diviene:

$$\left\langle \varepsilon_n^2 \right\rangle (a_1, a_2, ..., a_n) = \frac{1}{b-a} \sum_{k=1}^{n} \sum_{k'=1}^{n} a_k a_{k'} \left\langle \varphi_k, \varphi_{k'} \right\rangle,$$

mentre la (11) si scrive:

$$\nabla \left\langle \varepsilon_n^2 \right\rangle = \frac{2}{b-a} \hat{A}\mathbf{a},$$

per cui l'equazione operatoriale per i punti critici diviene:

$$\hat{A}\mathbf{a} = \mathbf{0} \Longleftrightarrow \mathbf{a} \in \ker \hat{A} = \left\{ \mathbf{x} \in \mathbb{R}^n \mid \hat{A}\mathbf{x} = \mathbf{0} \right\}$$

Per una nota proprietà:

$$R\left(\hat{A}\right) + N\left(\hat{A}\right) = \dim \mathbb{R}^n = n,$$

dove $R\left(\hat{A}\right)$ è il rango dell'endomorfismo $\hat{A}$, ovvero $R\left(\hat{A}\right) = \dim \hat{A}\left(\mathbb{R}^n\right)$, essendo $\hat{A}\left(\mathbb{R}^n\right) = \left\{\hat{A}\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\right\}$ l'immagine di $\mathbb{R}^n$ mediante $\hat{A}$. Il termine $N\left(\hat{A}\right)$ è, invece, la *nullità* di $\hat{A}$, cioè $N\left(\hat{A}\right) = \dim \ker \hat{A}$. Risulta poi:

$$R\left(\hat{A}\right) = rango\left(A\right), \quad \forall \{\mathbf{e}_k\} \text{ base di } \mathbb{R}^n,$$

onde:

$$\det A \neq 0 \Longrightarrow rango\left(\hat{A}\right) = R\left(\hat{A}\right) = n \Longrightarrow N\left(\hat{A}\right) = 0 \Longrightarrow \ker \hat{A} = \{\mathbf{0}\},$$

cioè $\ker \hat{A}$ è il sottospazio improprio di $\mathbb{R}^n$, ovvero il sottospazio il cui unico elemento è il vettore nullo. Ne concludiamo che se $\det A \neq 0$ e f è la funzione identicamente nulla in $[a, b]$, si ha:

$$\hat{A}\mathbf{a} = \mathbf{0} \Longleftrightarrow \mathbf{a} = \mathbf{0},$$

i.e. l'unico punto critico è $P\left(0, 0, ..., 0\right)$, avendosi :

$$\left\langle \varepsilon_n^2 \right\rangle (0, 0, ..., 0) = 0 \Longleftrightarrow \min_{\mathbb{R}^n} \left\langle \varepsilon_n^2 \right\rangle = 0$$

Tali risultati sono consistenti, poichè se f è identicamente nulla, la migliore approssimazione di f mediante polinomi è il polinomio identicamente nullo, i.e. il polinomio $\tau_n = \Sigma_k a_k \varphi_k$ con coefficienti a_k tutti nulli. Viceversa, se $\det A \neq 0$ esistono ∞^{n-p} polinomi non nulli che approssimano la funzione identicamente nulla. E tale risultato è manifestamente inconsistente.

Riprendiamo l'equazione operatoriale:

$$\hat{A}\mathbf{a} = \mathbf{c} \tag{18}$$

$\hat{A}$ è un endomorfismo simmetrico, giacchè $A \in \mathcal{S}_R\left(n\right)$, essendo quest'ultimo il sottospazio vettoriale di $M_{\mathbb{R}}\left(n, n\right)$[1] delle matrici simmetriche $n \times n$. Comè ben noto dall'Algebra lineare, ogni matrice simmetrica è riducibile a una matrice diagonale. Più precisamente, gli endomorfismi simmetrici ammettono una base ortogonale di autovettori $\{\mathbf{u}_k\}$ con autovalori reali. Abbiamo[2]:

$$\hat{A}\mathbf{u}_k = \lambda_k \mathbf{u}_k \quad (k = 1, 2, ..., n)$$

tali che

$$\lambda_k \in \mathbb{R}, \quad \mathbf{u}_k \cdot \mathbf{u}_h = 0, \quad \forall h \neq k$$

Il sistema ortogonale $\{\mathbf{u}_k\}$ può essere normalizzato, ottenendo una base ortonormale di $\mathbb{R}^n$. In tale base l'endomorfismo $\hat{A}$ è rappresentato da una matrice diagonale:

$$\hat{A} \doteq A_{diag} = \begin{pmatrix} \lambda_1 & 0 & ... & 0 \\ ... & \lambda_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \lambda_n \end{pmatrix}$$

Come è noto, la matrice diagonale A_{diag} è legata alla matrice A da una relazione del tipo (relazione di similitudine):

$$A_{diag} = R^{-1} A R, \tag{19}$$

dove R è la matrice di passaggio dalla base canonica $\{\mathbf{e}_k\}$ alla base $\{\mathbf{u}_k\}$. Trattandosi di basi ortonormali si ha che R è ortogonale, onde la relazione di similudine (19) è in realtà una relazione di congruenza:

$$A_{diag} = R^T A R$$

[1] $M_{\mathbb{R}}\left(n, n\right)$ è lo spazio vettoriale delle matrici $n \times n$ sui reali.

[2] Per semplicità consideriamo pari a 1 la molteplicità algebrica di singolo autovalore λ_k.

Ciò si esprime dicendo che ogni matrice simmetrica è congruente a una matrice diagonale. Per quanto riguarda la determinazione di autovalori e autovettori, ricordiamo che gli autovalori sono gli zeri del polinomio caratteristico, i.e. le radici dell'equazione caratteristica (o equazione secolare):

$$\det\left(A - \lambda \bar{I}_n\right) = 0, \tag{20}$$

dove $\bar{I}_n$ è la matrice identità di ordine n. Senza perdita di generalità, supponiamo di avere n radici semplici $\lambda_1, \lambda_2, ..., \lambda_n$, per cui le componenti nella base canonica dell'autovettore $\mathbf{u}_k \overset{def}{=} \mathbf{u}^{(k)} = \Sigma_{h=1}^{n} u_h^{(k)} \mathbf{e}_h$ si ottengono risolvendo il sistema omogeneo:

$$\begin{cases} \left(\langle \varphi_1, \varphi_1 \rangle - \lambda_1\right) u_1^{(k)} + \langle \varphi_1, \varphi_2 \rangle u_2^{(k)} + ... + \langle \varphi_1, \varphi_n \rangle u_n^{(k)} = 0 \\ \langle \varphi_2, \varphi_1 \rangle u_1^{(k)} + \left(\langle \varphi_2, \varphi_2 \rangle - \lambda_2\right) u_2^{(k)} + ... + \langle \varphi_2, \varphi_n \rangle u_n^{(k)} = 0 \\ ... \\ \langle \varphi_n, \varphi_1 \rangle u_1^{(k)} + \langle \varphi_n, \varphi_2 \rangle u_1^{(k)} + ... + \left(\langle \varphi_n, \varphi_n \rangle - \lambda_n\right) u_n^{(k)} = 0 \end{cases} \tag{21}$$

In forza della (20) tale sistema ammette infinite autosoluzioni definite a meno di una costante di normalizzazione. Al sottospazio vettoriale $V_n = L\left(\{\varphi_k\}\right) = \{\Sigma_k a_k \varphi_k \mid a_k \in \mathbb{R},\ \varphi_k \in C\left([a,b]\right)\}$ corrisponde $\mathbb{R}^n = \{(a_1, a_2, ..., a_n) \mid a_k \in \mathbb{R}\}$. Si tratta di spazi vettoriali isodimensionali, quindi isomorfi. Gli elementi della *n-pla* ordinata $(a_1, a_2, ..., a_n)$ sono le componenti del vettore $\mathbf{a}$ nella base canonica $\{\mathbf{e}_k\}$ di $\mathbb{R}^n$, ma sono anche le componenti di τ_n nella base $\{\varphi_k\}$ di V_n. Quindi, alla base $\{\mathbf{e}_k\}$ di $\mathbb{R}^n$ corrisponde la base $\{\varphi_k\}$ di V_n. Ciò implica che al cambiamento di base $\{\mathbf{e}_k\} \to \left\{\mathbf{u}^{(k)}\right\}$ corrisponde in V_n il cambiamento di base:

$$\{\varphi_k\} \to \{\psi_k\},$$

dove $\{\psi_k\}$ è una base ortogonale di V_n. Quindi:

$$A_{diag} = \begin{pmatrix} \langle \psi_1, \psi_1 \rangle & 0 & ... & 0 \\ ... & \langle \psi_2, \psi_2 \rangle & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \langle \psi_n, \psi_n \rangle \end{pmatrix},$$

cioè $\lambda_k = ||\psi_k||^2$. Se $\{\psi_k\}$ è ortonormale, si ha $\lambda_k = 1$. In tal caso la (18) si scrive:

$$\hat{A}\mathbf{a} = \mathbf{c} \Longleftrightarrow A_{diag} X = C,$$

da cui le soluzioni:

$$a_k = \frac{c_k}{\langle \psi_k, \psi_k \rangle}, \quad (k = 1, 2, ..., n)$$

Se $\{\psi_k\}$ è ortonormale:

$$a_k = c_k, \quad (k = 1, 2, ..., n)$$

onde l'unico punto estremale ha per coordinate le coordinate di Fourier della funzione assegnata f, confermando il risultato ottenuto dalla (**??**).

Ritorniamo al caso generale in cui i polinomi di base di V_n non sono ortogonali. Abbiamo visto che se la matrice A è non singolare, esiste uno ed un solo punto critico $\tilde{P}\left(\tilde{a}_1, \tilde{a}_2, ..., \tilde{a}_n\right)$ della funzione $\langle \varepsilon_n^2 \rangle$. Per stabilire se si tratta di un punto di minimo, dobbiamo determinare le derivate parziali seconde. A tale scopo, riscriviamo la (8) come

$$\frac{\partial \langle \varepsilon_n^2 \rangle}{\partial a_i} = \frac{2}{b-a}\left(-c_i + \sum_{k=1}^{n} \langle \varphi_k, \varphi_i \rangle a_k\right),$$

per cui:

$$\frac{\partial^2 \langle \varepsilon_n^2 \rangle}{\partial a_i \partial a_j} = \frac{2}{b-a}\left(0 + \sum_{k=1}^{n} \langle \varphi_k, \varphi_i \rangle \delta_{kj}\right)$$

Cioè:

$$\frac{\partial^2 \langle \varepsilon_n^2 \rangle}{\partial a_h \partial a_k} = \frac{2}{b-a} \langle \varphi_h, \varphi_k \rangle \tag{22}$$

Senza perdita di generalità consideriamo il caso particolare $n = 2$. Quindi:

$$\langle \varepsilon_n^2 \rangle (\mathbf{a}) = \frac{1}{b-a} \left[||f||^2 - 2\mathbf{a} \cdot \mathbf{c} + a_1^2 \langle \varphi_1, \varphi_1 \rangle + 2a_1 a_2 \langle \varphi_1, \varphi_2 \rangle + a_2^2 \langle \varphi_2, \varphi_2 \rangle \right]$$

Le (8) si scrivono:

$$\begin{aligned} \frac{\partial \langle \varepsilon_n^2 \rangle}{\partial a_1} &= \frac{2}{b-a} \left(-c_1 + a_1 \langle \varphi_1, \varphi_1 \rangle + a_2 \langle \varphi_1, \varphi_2 \rangle \right) \\ \frac{\partial \langle \varepsilon_n^2 \rangle}{\partial a_1} &= \frac{2}{b-a} \left(-c_2 + a_1 \langle \varphi_1, \varphi_2 \rangle + a_2 \langle \varphi_2, \varphi_2 \rangle \right), \end{aligned} \tag{23}$$

quali componenti del gradiente di $\langle \varepsilon_n^2 \rangle$:

$$\nabla \langle \varepsilon_n^2 \rangle = \frac{2}{b-a} \left(\hat{A}\mathbf{a} - \mathbf{c} \right)$$

Le (22):

$$\begin{aligned} \frac{\partial^2 \langle \varepsilon_n^2 \rangle}{\partial a_1^2} &= \frac{2}{b-a} \langle \varphi_1, \varphi_1 \rangle \\ \frac{\partial^2 \langle \varepsilon_n^2 \rangle}{\partial a_2^2} &= \frac{2}{b-a} \langle \varphi_2, \varphi_2 \rangle \\ \frac{\partial^2 \langle \varepsilon_n^2 \rangle}{\partial a_1 \partial a_2} &= \frac{2}{b-a} \langle \varphi_1, \varphi_2 \rangle \end{aligned} \tag{24}$$

Risolviamo:

$$\hat{A}\mathbf{a} = \mathbf{c} \iff \begin{cases} \langle \varphi_1, \varphi_1 \rangle a_1 + \langle \varphi_1, \varphi_2 \rangle a_2 = c_1 \\ \langle \varphi_2, \varphi_1 \rangle a_1 + \langle \varphi_2, \varphi_2 \rangle a_2 = c_2 \end{cases} \tag{25}$$

Riesce:

$$\det A = \langle \varphi_1, \varphi_1 \rangle \langle \varphi_2, \varphi_2 \rangle - \langle \varphi_1, \varphi_2 \rangle^2,$$

da cui vediamo che $\det A \neq 0, \quad \forall \varphi_1, \varphi_2 \neq 0$. Abbiamo quindi un unico punto estremale $\tilde{P}(\tilde{a}_1, \tilde{a}_2)$ con:

$$\begin{aligned} \tilde{a}_1 &= \frac{c_1 \langle \varphi_2, \varphi_2 \rangle - c_2 \langle \varphi_1, \varphi_2 \rangle}{\langle \varphi_1, \varphi_1 \rangle \langle \varphi_2, \varphi_2 \rangle - \langle \varphi_1, \varphi_2 \rangle^2} \\ \tilde{a}_2 &= \frac{c_2 \langle \varphi_1, \varphi_1 \rangle - c_1 \langle \varphi_2, \varphi_1 \rangle}{\langle \varphi_1, \varphi_1 \rangle \langle \varphi_2, \varphi_2 \rangle - \langle \varphi_1, \varphi_2 \rangle^2} \end{aligned} \tag{26}$$

Si noti che

$$(\tilde{a}_1, \tilde{a}_2) = (c_1, c_2) \iff \{\varphi_k\} \text{ è ortonormale,}$$

confermando quanto detto in precedenza. Per stabilire la natura del punto critico $\tilde{P}$, dobbiamo calcolare l'hessiano in tale punto:

$$H(\tilde{a}_1, \tilde{a}_2) = \left. \frac{\partial^2 \langle \varepsilon_n^2 \rangle}{\partial a_1^2} \right|_{(\tilde{a}_1, \tilde{a}_2)} \left. \frac{\partial^2 \langle \varepsilon_n^2 \rangle}{\partial a_2^2} \right|_{(\tilde{a}_1, \tilde{a}_2)} - \left[\left. \frac{\partial^2 \langle \varepsilon_n^2 \rangle}{\partial a_1 \partial a_2} \right|_{(\tilde{a}_1, \tilde{a}_2)} \right]^2$$

Osserviamo che le derivate seconde di $\langle \varepsilon_n^2 \rangle$ sono delle costanti. In ogni caso, sostituendo i loro valori (eq. (24)) nell'equazione precedente si ottiene:

$$H(\tilde{a}_1, \tilde{a}_2) = \frac{4}{b-a} \det A \tag{27}$$

Risulta $H(\tilde{a}_1, \tilde{a}_2) \neq 0$, poichè è $\det A \neq 0$. Ne consegue che $\tilde{P}(\tilde{a}_1, \tilde{a}_2)$ non è punto di sella comunqe prendiamo il sistema $\{\varphi_k\}$. Ricordiamo, poi, che per il noto criterio dell'hessiano affinchè $\tilde{P}(\tilde{a}_1, \tilde{a}_2)$ sia un punto di minimo relativo, deve aversi:

$$H(\tilde{a}_1, \tilde{a}_2) > 0, \quad \left.\frac{\partial^2 \langle \varepsilon_n^2 \rangle}{\partial a_1^2}\right|_{(\tilde{a}_1, \tilde{a}_2)} > 0$$

Riesce $\left.\frac{\partial^2 \langle \varepsilon_n^2 \rangle}{\partial a_1^2}\right|_{(\tilde{a}_1, \tilde{a}_2)} > 0$, in quanto $\left.\frac{\partial^2 \langle \varepsilon_n^2 \rangle}{\partial a_1^2}\right|_{(\tilde{a}_1, \tilde{a}_2)} = \frac{2}{b-a} ||\varphi_k||^2$, mentre

$$H(\tilde{a}_1, \tilde{a}_2) > 0 \Longleftrightarrow \det A > 0$$

Ne concludiamo che $\tilde{P}(\tilde{a}_1, \tilde{a}_2)$ è punto di minimo relativo per $G(\mathbf{a})$ se e solo se $\det A > 0$. Tale punto è manifestamente di minimo assoluto per la funzione $\langle \varepsilon_n^2 \rangle (a_1, a_2)$.

Per $n > 2$ la ricerca del minimo di $\langle \varepsilon_n^2 \rangle$ richiede lo studio della forma quadratica:

$$\phi(\lambda_1, \lambda_2, ..., \lambda_n) = \sum_{h=1}^{n} \sum_{k=1}^{n} \left.\frac{\partial^2 \langle \varepsilon_n^2 \rangle}{\partial a_h \partial a_k}\right|_{\tilde{P}} \lambda_h \lambda_k$$

nelle variabili ausiliarie[3] $\lambda_1, \lambda_2, ..., \lambda_n$. Precisamente:

A. ϕ è definita positiva $\Longrightarrow P_0$ è punto di minimo relativo proprio;

B. ϕ è definita negativa $\Longrightarrow P_0$ è punto di massimo relativo proprio;

C. ϕ è indefinita $\Longrightarrow P_0$ non è punto di estremo relativo.

Ma, per quanto precede, il minimo assoluto di $\langle \varepsilon_n^2 \rangle$ può essere determinato con l'algebra delle matrici, ovvero risolvendo il sistema di equazioni lineari (15). Abbiamo, quindi, il seguente algoritmo:

Sia data la funzione $f \in C([a, b])$. Stabiliamo un ordine di approssimazione individuato da un intero positivo n, dopodichè scegliamo ad arbitrio un sistema linearmente indipendente di polinomi $\{\varphi_1, \varphi_2, ..., \varphi_n\}$. Calcoliamo la norma al quadrato della funzione:

$$||f||^2 = \int_a^b [f(x)]^2 dx,$$

quindi le coordinate di Fourier di f:

$$c_k = \langle \varphi_k, f \rangle = \int_a^b \varphi_k(x) f(x) dx,$$

e i prodotti scalari:

$$\langle \varphi_h, \varphi_k \rangle = \int_a^b \varphi_h(x) \varphi_k(x) dx$$

[3] Solitamente definite da:

$$\lambda_k = \frac{a_k - \tilde{a}_k}{\rho},$$

dove $\rho = dist\left(P, \tilde{P}\right) = \sqrt{\sum_{k=1}^{n} (a_k - \tilde{a}_k)^2}$, con $P \neq \tilde{P}$.

A questo punto non dobbiamo fare altro che risolvere il sistema di equazioni lineari (15). Per quanto visto nel caso particolare $n = 2$, riesce $\det A \neq 0$ per ogni sistema $\{\varphi_k\}$ linearmente indipendente. Quindi l'unica soluzione $(\tilde{a}_1, \tilde{a}_2, ..., \tilde{a}_n)$ è, se $\det A > 0$, punto di minimo di $\langle \varepsilon_n^2 \rangle$:

$$\min_{\mathbb{R}^n} \langle \varepsilon_n^2 \rangle = G(\tilde{a}_1, \tilde{a}_2, ..., \tilde{a}_n)$$

Esempio 1 *Approssimare la funzione*

$$f : [-1, 1] \to \arcsin x^2$$

mediante $\tau_8(x) = \Sigma_{k=1}^{8} a_k \varphi_k$, *dove:*

$$\begin{aligned}
\varphi_1(x) &= 1 + x^3 + x^5 \\
\varphi_2(x) &= 1 - x^2 \\
\varphi_3(x) &= x^4 - x^5 \\
\varphi_4(x) &= 2 + x + x^2 - x^3 + x^4 + 3x^5 \\
\varphi_5(x) &= -3 + 4x - 2x^2 \\
\varphi_6(x) &= -1 + x + x^2 + x^3 + x^4 \\
\varphi_7(x) &= -4 + x^2 - 7x^3 + x^5 \\
\varphi_8(x) &= 5 - 3x^2 + 4x^3 + 10x^5
\end{aligned}$$

Svolgimento.
Calcolando gli integrali, si trova:
Norma al quadrato di f

$$||f||^2 = \int_{-1}^{1} (1 + \arcsin x)^2 \, dx = \frac{\pi^2}{2} - 2$$

Coordinate di Fourier di f

$$\begin{aligned}
c_1 &= \int_{-1}^{1} \left(1 + x^3 + x^5\right)(1 + \arcsin x) \, dx = 2 + \frac{13}{48}\pi \\
c_2 &= \frac{4}{3} \\
c_3 &= \frac{2}{5} - \frac{11}{96}\pi \\
c_4 &= \frac{76}{15} + \frac{7}{16}\pi \\
c_5 &= -\frac{22}{3} + \pi \\
c_6 &= -\frac{14}{15} + \frac{13}{32}\pi \\
c_7 &= -\frac{1}{48}(352 + 47\pi) \\
c_8 &= 8 + \frac{85}{48}\pi
\end{aligned}$$

Prodotti scalari dei vettori di base $\langle \varphi_h, \varphi_k \rangle$ e quindi il sistema di equazioni lineari:

$$\begin{cases} \frac{2018}{693}a_1 + \frac{4}{3}a_2 - \frac{2}{495}a_3 + \frac{3196}{495}a_4 - \frac{482}{105}a_5 + \frac{82}{315}a_6 - \frac{346}{33}a_7 + \frac{9752}{693}a_8 = 2 + \frac{13}{48}\pi \\ \frac{4}{3}a_1 + \frac{16}{15}a_2 + \frac{4}{35}a_3 + \frac{64}{24}a_4 - \frac{68}{15}a_5 - \frac{20}{21}a_6 - \frac{76}{15}a_7 + \frac{88}{15}a_8 = \frac{4}{3} \\ -\frac{2}{495}a_1 + \frac{4}{35}a_2 + \frac{40}{99}a_3 + \frac{346}{495}a_4 - \frac{102}{35}a_5 - \frac{2}{5}a_6 + \frac{206}{3465}a_7 - \frac{1084}{693}a_8 = \frac{2}{5} - \frac{11}{96}\pi \\ \frac{3196}{495}a_1 + \frac{64}{24}a_2 + \frac{346}{495}a_3 + \frac{54158}{3465}a_4 - \frac{516}{35}a_5 + \frac{52}{315}a_6 - \frac{16012}{693}a_7 + \frac{98716}{3465}a_8 = \frac{76}{15} + \frac{7}{16}\pi \\ -\frac{482}{105}a_1 - \frac{68}{15}a_2 - \frac{102}{35}a_3 - \frac{516}{35}a_4 + \frac{574}{15}a_5 + \frac{246}{35}a_6 + \frac{346}{21}a_7 - \frac{1096}{105}a_8 = -\frac{22}{3} + \pi \\ \frac{82}{315}a_1 - \frac{20}{21}a_2 - \frac{2}{5}a_3 + \frac{52}{315}a_4 + \frac{246}{35}a_5 + \frac{886}{315}a_6 - \frac{34}{63}a_7 + \frac{976}{315}a_8 = -\frac{14}{15} + \frac{13}{32}\pi \\ -\frac{346}{33}a_1 - \frac{76}{15}a_2 + \frac{206}{3465}a_3 - \frac{16012}{693}a_4 + \frac{346}{21}a_5 - \frac{34}{63}a_6 + \frac{18878}{495}a_7 - \frac{8368}{165}a_8 = -\frac{1}{48}(352 + 47\pi) \\ \frac{9752}{693}a_1 + \frac{88}{15}a_2 - \frac{1084}{693}a_3 + \frac{98716}{3465}a_4 - \frac{1096}{105}a_5 + \frac{976}{315}a_6 - \frac{8368}{165}a_7 + \frac{256864}{3465}a_8 = 8 + \frac{85}{48}\pi \end{cases} \tag{28}$$

Riesce $\det A = 0$, per cui il sistema se è compatibile è indeterminato. Tentando, tuttavia, di trovare almeno una soluzione (ad esempio, con un sistema di computer algebra), si ha:

$$\tau_8(x) = -\left(4 + \frac{19509\pi}{20480}\right)\varphi_1(x) + \left(6 + \frac{5397\pi}{4096}\right)\varphi_2(x) - \left(4 + \frac{4053\pi}{4096}\right)\varphi_3(x) + \frac{903\pi}{20480}\varphi_3(x) - \left(1 + \frac{21\pi}{128}\right)\varphi_4(x) + \left(4 + \frac{9681\pi}{10240}\right)\varphi_5(x)$$

In fig. 1 riportiamo il grafico di f e del polinomio approssimante.

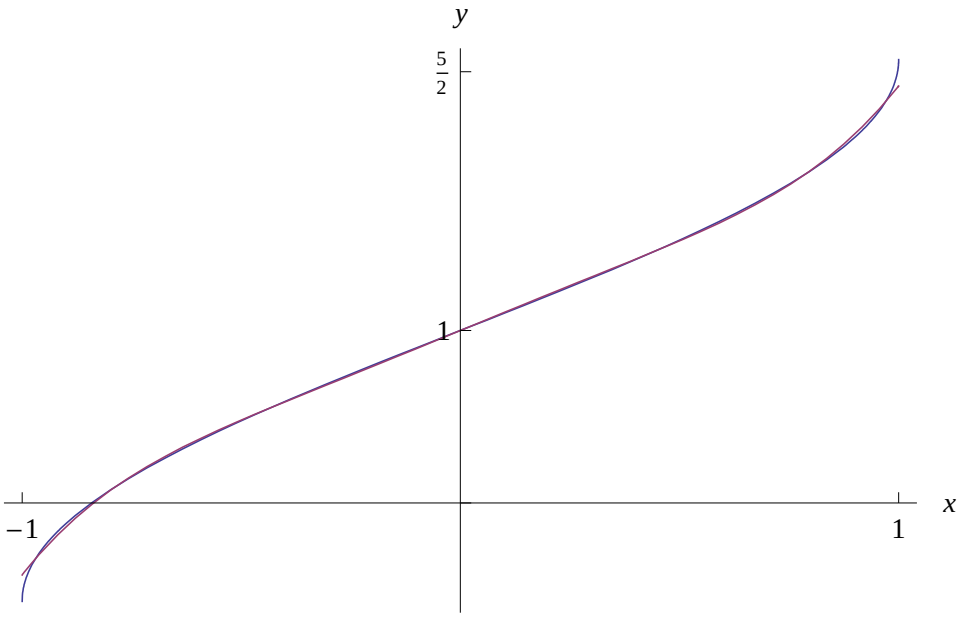


Figura 1: Approssimazione della funzione $f(x) = 1 + \arcsin x$ mediante $\tau_8(x)$.

Esempio 2 *Approssimare la funzione*

$$f(x) = \begin{cases} 1, & per\ 0 < x \leq 1 \\ 0, & per\ x = 0 \\ -1, & per\ -1 < x < 0 \end{cases} \tag{29}$$

mediante $\tau_8(x) = \Sigma_{k=1}^{8} a_k \varphi_k$, *dove* φ_k *sono i polinomi dell'esempio precedente.*

Svolgimento.

Si noti che $f \in C([-1, 1])$, poichè $x = 0$ è un punto di discontinuità di prima specie. Trattandosi di una discontinuità finita, possiamo comunque tentare un'approssimazione. Calcolando le coordinate di Fourier e gli elementi di matrice di A, si perviene all'approssimazione graficata in fig. 2.

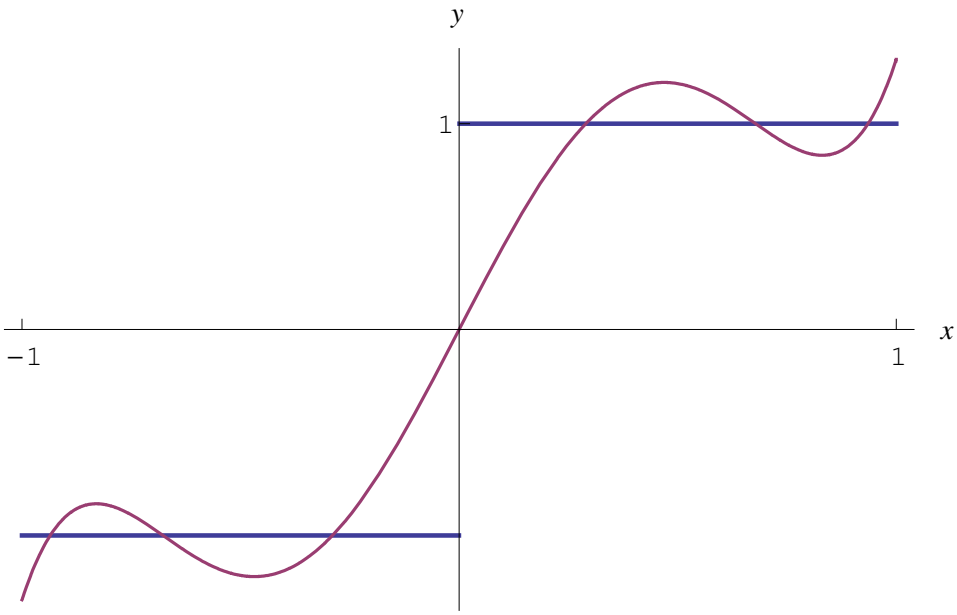


Figura 2: Approssimazione della funzione (29) con $\tau_8(x)$.