Minimi quadrati e problemi di distanza minima

Consideriamo una matrice rettangolare B, con elementi b_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, m$, con $m < n$ (quindi, più righe che colonne). Vogliamo "risolvere" il sistema lineare

$$B\mathbf{x} = \mathbf{b}\,, \tag{1}$$

dove $\mathbf{b} \in \mathbf{R}^n$ è un vettore dato.

Dal teorema della dimensione dell'algebra lineare sappiamo che la soluzione, se esiste, è unica se e solo se B è di rango massimo (cioè esiste un minore $m \times m$ di B che è non-singolare). Dal Teorema di Rouché–Capelli si sa che la soluzione esiste se e solo se il rango di B coincide con il rango della matrice $(B|\mathbf{b})$, ottenuta aggiungendo a B la colonna data dal termine noto $\mathbf{b}$.

Quello che qui vogliamo fare è però dare un senso alla "soluzione" di (1) per B di rango massimo e per *qualunque* valore di $\mathbf{b} \in \mathbf{R}^n$, cioè anche se il rango di $(B|\mathbf{b})$ è diverso da m (il che vuol dire, in questa situazione, che è uguale a $m + 1$).

L'ipotesi è dunque solamente:

(*) B sia di rango massimo e sia $\mathbf{b} \in \mathbf{R}^n$.

- "Risolvere" significa minimizzare

Bisogna dare un senso opportuno alla parola "soluzione": l'idea è di considerare "soluzione" il vettore $\mathbf{x}^* \in \mathbf{R}^m$ che minimizza la distanza fra $B\mathbf{x}$ e $\mathbf{b}$, cioè

$$\mathbf{x}^* \in \mathbf{R}^m \text{ è tale da minimizzare } \psi(\mathbf{x}) = \|B\mathbf{x} - \mathbf{b}\|^2 = \sum_{i=1}^n \Big(\sum_{j=1}^m b_{ij} x_j - b_i \Big)^2 . \tag{2}$$

Per prima cosa mostriamo che ψ ha effettivamente un minimo: dal momento che è continua (è un polinomio di secondo grado in m variabili, dunque è infinitamente differenziabile), è sufficiente mostrare che $\psi(\mathbf{x}) \to +\infty$ per $\|\mathbf{x}\| \to +\infty$[(1)]. Siccome si ha $\psi(\mathbf{x}) \geq (\|B\mathbf{x}\| - \|\mathbf{b}\|)^2$, basta provare che $\|B\mathbf{x}\| \to +\infty$. Ora

$$\|B\mathbf{x}\|^2 = B\mathbf{x} \cdot B\mathbf{x} = B^T B\mathbf{x} \cdot \mathbf{x}\,,$$

e quindi la matrice quadrata $B^T B$ è semi-definita positiva (cioè $B^T B\mathbf{x} \cdot \mathbf{x} \geq 0$ per ogni $\mathbf{x}$). D'altra parte $B^T B\mathbf{x} \cdot \mathbf{x} = 0$ se e solo se $B\mathbf{x} = \mathbf{0}$, e questo avviene se e solo se $\mathbf{x} = \mathbf{0}$, dal momento che B è di rango massimo. Quindi $B^T B$ è definita positiva (cioè $B^T B\mathbf{x} \cdot \mathbf{x} > 0$ per ogni $\mathbf{x} \neq \mathbf{0}$). Essendo ovviamente simmetrica ($(B^T B)^T = B^T (B^T)^T = B^T B$), si deduce che il suo autovalore minimo $\mu_{\min}$ è strettamente positivo, e che

$$\|B\mathbf{x}\|^2 = B^T B\mathbf{x} \cdot \mathbf{x} \geq \mu_{\min} \|\mathbf{x}\|^2 \to +\infty\,.$$

Chiamiamo dunque $\mathbf{x}^*$ un punto di minimo di ψ: esso è un punto stazionario di ψ e per individuarlo calcoliamo il gradiente di ψ. Si ha, per $k = 1, \ldots, m$:

$$\begin{aligned}
\frac{\partial \psi}{\partial x_k}(\mathbf{x}) &= \frac{\partial}{\partial x_k} \Big[\sum_{i=1}^n \Big(\sum_{j=1}^m b_{ij} x_j - b_i \Big)^2 \Big] \\
&= 2 \sum_{i=1}^n \Big(\sum_{j=1}^m b_{ij} x_j - b_i \Big) \frac{\partial}{\partial x_k} \Big(\sum_{j=1}^m b_{ij} x_j - b_i \Big) \\
&= 2 \sum_{i=1}^n \Big(\sum_{j=1}^m b_{ij} x_j - b_i \Big) \sum_{j=1}^m b_{ij} \frac{\partial x_j}{\partial x_k}\,.
\end{aligned}$$

---

(1) Se $\psi(\mathbf{x}) \to +\infty$, fissata la soglia $M = |\psi(\mathbf{0})|$ esiste $r > 0$ per cui per $\|\mathbf{x}\| > r$ si ha $\psi(\mathbf{x}) \geq M$. D'altro canto, dal Teorema di Weierstrass ψ ha un minimo sull'insieme chiuso e limitato $B_r = \{\mathbf{x} \in \mathbf{R}^m \mid \|\mathbf{x}\| \leq r\}$, cioè si ha $\psi(\mathbf{x}) \geq \nu$ per $\|\mathbf{x}\| \leq r$, avendo definito $\nu = \min_{B_r} \psi(\mathbf{x})$; in particolare $\psi(\mathbf{0}) \geq \nu$. Di conseguenza, per $\|\mathbf{x}\| > r$ si ha $\psi(\mathbf{x}) \geq M \geq \psi(\mathbf{0}) \geq \nu$, e cosí ν risulta il valore di minimo di ψ non solo in B_r ma in tutto $\mathbf{R}^m$.

Siccome

$$\frac{\partial x_j}{\partial x_k} = \begin{cases} 1 & \text{se } j = k \\ 0 & \text{se } j \neq k \end{cases} ,$$

si conclude che

$$\frac{\partial \psi}{\partial x_k}(\mathbf{x}) = 2 \sum_{i=1}^{n} \Big(\sum_{j=1}^{m} b_{ij} x_j - b_i \Big) b_{ik} \,.$$

D'altra parte $b_{ik} = (B)_{ik} = (B^T)_{ki}$, quindi possiamo riscrivere

$$\operatorname{grad} \psi(\mathbf{x}) = 2\, B^T (B\mathbf{x} - \mathbf{b}) \,.$$

Ogni punto stazionario $\mathbf{x}$ di ψ dunque soddisfa

$$B^T B \mathbf{x} = B^T \mathbf{b} \,. \tag{3}$$

Avendo già verificato che, se B è di rango massimo, la matrice $B^T B$ è simmetrica e definita positiva, dunque non-singolare, la soluzione di (3) è unica, quindi c'è un solo punto stazionaro di ψ, e questo punto stazionario è l'unico punto di minimo $\mathbf{x}^*$ di ψ.

In conclusione, per "risolvere" il sistema (1), nel senso dato da (2), bisogna trovare l'unica soluzione $\mathbf{x}^*$ del sistema (3). Questa soluzione viene chiamata "soluzione ai minimi quadrati".

- Un esempio: retta di regressione lineare

Si vuole rispondere alla domanda: qual è la retta $y = \alpha x + \beta$ che "passa" per tre o più punti non allineati? Se i punti sono (x_i, y_i), $i = 1, \ldots, n$, con $n \geq 3$, si vuole risolvere il sistema

$$\begin{cases} \alpha x_1 + \beta = y_1 \\ \alpha x_2 + \beta = y_2 \\ \ldots \\ \alpha x_n + \beta = y_n \,. \end{cases} \tag{4}$$

Se i punti (x_i, y_i) non sono allineati, il sistema (4) non ha soluzione. Però si può provare a "risolverlo" nel senso dei minimi quadrati.

La matrice che esprime il sistema lineare (4) è la matrice $n \times 2$

$$B = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \ldots & \\ x_n & 1 \end{pmatrix} .$$

Se per almeno una coppia di indici distinti i_1 e i_2 si ha $x_{i_1} \neq x_{i_2}$ (cioè se nel piano (x, y) i punti non sono tutti su una retta verticale) la matrice B risulta di rango massimo. Dunque basta risolvere

$$B^T B \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = B^T \begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix} .$$

Si ha

$$B^T B = \begin{pmatrix} \displaystyle\sum_{i=1}^{n} x_i^2 & \displaystyle\sum_{i=1}^{n} x_i \\ \displaystyle\sum_{i=1}^{n} x_i & n \end{pmatrix} \quad , \quad B^T \begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix} = \begin{pmatrix} \displaystyle\sum_{i=1}^{n} x_i y_i \\ \displaystyle\sum_{i=1}^{n} y_i \end{pmatrix} ,$$

per cui

$$
\begin{aligned}
\alpha &= \frac{n\sum_{i=1}^{n} x_i y_i - \Big(\sum_{i=1}^{n} x_i\Big)\Big(\sum_{i=1}^{n} y_i\Big)}{n\sum_{i=1}^{n} x_i^2 - \Big(\sum_{i=1}^{n} x_i\Big)^2} \\
\beta &= \frac{\Big(\sum_{i=1}^{n} x_i^2\Big)\Big(\sum_{i=1}^{n} y_i\Big) - \Big(\sum_{i=1}^{n} x_i\Big)\Big(\sum_{i=1}^{n} x_i y_i\Big)}{n\sum_{i=1}^{n} x_i^2 - \Big(\sum_{i=1}^{n} x_i\Big)^2}\,.
\end{aligned} \tag{5}
$$

- Un altro problema: polinomio di distanza minima

La domanda è: qual è il polinomio di grado N che ha distanza minima da una funzione F assegnata?

Il primo problema è definire il concetto di "distanza fra funzioni". Possiamo procedere in questo modo: se abbiamo due funzioni F e G, definite e (per esempio) continue in un intervallo $[a, b]$, un prodotto scalare fra di loro può essere definito come

$$
\langle F, G\rangle = \int_a^b F(x)\, G(x)\, dx\,.
$$

[Controllate che questa definizione soddisfi tutte le proprietà di un prodotto scalare.]

Avendo un prodotto scalare si ha una norma: $\|F\|_* = \langle F, F\rangle^{1/2}$ (scriviamo $\|F\|_*$ per distinguerla dalla norma euclidea di un vettore $\mathbf{a}$, che è indicata con $\|\mathbf{a}\|$); avendo una norma si ha una distanza: $\mathrm{dist}(F, G) = \|F - G\|_*$.

La domanda iniziale dunque si riformula cosí: data una funzione F, definita e continua in $[a, b]$, determinare il polinomio $P_N(x)$ di grado N che minimizza la distanza (al quadrato) da F, cioè che minimizza

$$
\|P_N - F\|_*^2 = \int_a^b (P_N(x) - F(x))^2 dx\,.
$$

Siccome un polinomio di grado N si scrive come $P_N(x) = \sum_{j=0}^N a_j x^j$, si tratta di trovare i coefficienti a_j, $j = 0, 1, \ldots, N$, che minimizzano la funzione di $N + 1$ variabili

$$
Q(a_0, \ldots, a_N) = \int_a^b \Big(\sum_{j=0}^{N} a_j x^j - F(x)\Big)^2 dx\,.
$$

Si può dimostrare che in $\mathbf{R}^{N+1}$ questa funzione Q ha un valore minimo[(2)]. Come al solito, per trovare un punto di minimo cerchiamo un punto stazionario. Calcoliamo il gradiente di Q (derivando sotto il segno

[(2)] Se proprio lo volete sapere... Scriviamo $\mathbf{a} = (a_0, \ldots, a_N)$. La funzione Q è un polinomio di secondo grado nelle $N + 1$ variabili $a_0, a_1, \ldots, a_N$, dunque è infinitamente differenziabile in $\mathbf{R}^{N+1}$. Come nel caso precedente, per vedere che ha un minimo basta mostrare che $Q(\mathbf{a}) \to +\infty$ per $\|\mathbf{a}\| \to +\infty$. Si ha $Q(\mathbf{a}) \geq (\|P_N\|_* - \|F\|_*)^2$, quindi basta dimostrare che $\|P_N\|_* = \|\sum_{j=0}^N a_i x^j\|_* \to +\infty$ per $\|\mathbf{a}\| \to +\infty$. Definiamo $q(\mathbf{a}) = \|\sum_{j=0}^N a_j x^j\|_*$. Una prima osservazione è che $q(\mathbf{a}) = 0$ se e solo se il polinomio $\sum_{j=0}^N a_j x^j$ è nullo, cioè se e solo se $\mathbf{a} = \mathbf{0}$. Siccome q è una funzione continua, dal Teorema di Weierstrass ha minimo m_1 sull'insieme chiuso e limitato $S_1 = \{\mathbf{v} \in \mathbf{R}^{N+1} \mid \|\mathbf{v}\| = 1\}$, e, per quanto appena detto, questo minimo deve soddisfare $m_1 > 0$. Siccome si ha $q(t\mathbf{w}) = tq(\mathbf{w})$ per ogni $t > 0$ ed ogni $\mathbf{w} \in \mathbf{R}^{N+1}$, preso $\mathbf{a} \in \mathbf{R}^{N+1}$, $\mathbf{a} \neq \mathbf{0}$, si deduce

$$
q(\mathbf{a}) = q\Big(\|\mathbf{a}\|\frac{\mathbf{a}}{\|\mathbf{a}\|}\Big) = \|\mathbf{a}\|\, q\Big(\frac{\mathbf{a}}{\|\mathbf{a}\|}\Big) \geq m_1\|\mathbf{a}\| \to +\infty
$$

(si è considerato $t = \|\mathbf{a}\| > 0$ e $\mathbf{w} = \frac{\mathbf{a}}{\|\mathbf{a}\|} \in S_1$).

di integrale, operazione che è giustificata nel caso in questione):

$$
\begin{aligned}
\frac{\partial Q}{\partial a_k}(a_0,\ldots,a_N) &= \int_a^b \frac{\partial}{\partial a_k}\Big(\sum_{j=0}^N a_j x^j - F(x)\Big)^2 dx \\
&= 2\int_a^b \Big(\sum_{j=0}^N a_j x^j - F(x)\Big)\frac{\partial}{\partial a_k}\Big(\sum_{j=0}^N a_j x^j - F(x)\Big) dx \\
&= 2\int_a^b \Big(\sum_{j=0}^N a_j x^j - F(x)\Big)\Big(\sum_{j=0}^N x^j \frac{\partial a_j}{\partial a_k}\Big) dx \,.
\end{aligned}
$$

Siccome

$$
\frac{\partial a_j}{\partial a_k} = \begin{cases} 1 & \text{se } j = k \\ 0 & \text{se } j \neq k \end{cases} \,,
$$

si conclude che

$$
\frac{\partial Q}{\partial a_k}(a_0,\ldots,a_N) = 2\int_a^b \Big(\sum_{j=0}^N a_j x^j - F(x)\Big) x^k \, dx \,.
$$

Se definiamo per $j,k = 0,\ldots,N$ gli elementi a_{kj} della matrice A e le componenti f_k del vettore $\mathbf{f}$ come

$$
a_{kj} = \int_a^b x^{j+k}\, dx \quad , \quad f_k = \int_a^b F(x)\, x^k \, dx \,,
$$

abbiamo ottenuto che $\operatorname{grad} Q = 2(A\mathbf{a} - \mathbf{f})$, e dunque i punti stazionari di Q sono le soluzioni del sistema lineare

$$
A\mathbf{a} = \mathbf{f} \,. \tag{6}
$$

La matrice A è chiaramente simmetrica, ed è anche definita positiva. Infatti, per $\mathbf{v} \in \mathbf{R}^{N+1}$ si ha

$$
\begin{aligned}
A\mathbf{v}\cdot\mathbf{v} &= \sum_{k,j=0}^N a_{kj} v_j v_k = \sum_{k,j=0}^N v_j v_k \int_a^b x^j x^k \, dx \\
&= \int_a^b \Big(\sum_{j=0}^N v_j x^j\Big)\Big(\sum_{k=0}^N v_k x^k\Big) dx = \int_a^b \Big(\sum_{j=0}^N v_j x^j\Big)^2 dx \,.
\end{aligned}
$$

Quindi $A\mathbf{v}\cdot\mathbf{v} \geq 0$, e $A\mathbf{v}\cdot\mathbf{v} = 0$ se e solo se il polinomio $\sum_{j=0}^N v_j x^j$ è nullo, cioè se e solo se tutti i coefficienti v_j sono nulli (ossia $\mathbf{v} = \mathbf{0}$).

Si può dunque concludere che la matrice A è non-singolare, e il sistema (6) ha soluzione unica, cioè la funzione Q ha un unico punto stazionario, l'unico suo punto di minimo. In conclusione, il polinomio $P_N(x) = \sum_{j=0}^N a_j x^j$ di distanza minima dalla funzione F è univocamente determinato dai coefficienti $a_0, a_1, \ldots, a_N$ soluzione del sistema (6).

[**Approfondimento sul tema:**

Si noti che il problema potrebbe essere semplificato se avessimo rappresentato il polinomio P_N tramite una base polinomiale più "astuta" (lo spazio dei polinomi di grado N definiti in un intervallo $[a,b]$ sono uno spazio vettoriale di dimensione $N+1$, dunque un qualunque sistema di $N+1$ polinomi di grado N fra loro linearmente indipendenti sono una base). La base che abbiamo utilizzato nell'esempio precedente sembra la più semplice, essendo data dai polinomi $1, x, \ldots, x^N$. Ma se ad essa applichiamo il procedimento di ortonormalizzazione di Gram–Schmidt, utilizzando il prodotto scalare $\langle\cdot,\cdot\rangle$, troviamo una base ortonormale di polinomi. Chiamiamoli $L_j(x)$, $j = 0,1,\ldots,N$: in particolare si verifica facilmente che L_j è di grado j. Ora si cercano i coefficienti $\widehat{a}_0, \widehat{a}_1, \ldots, \widehat{a}_N$ del polinomio $P_N(x) = \sum_{j=0}^N \widehat{a}_j L_j(x)$ in modo tale che esso sia di distanza minima dalla funzione F. Ripetendo il procedimento di minimizzazione ora presentato si arriva facilmente a verificare che il vettore $\widehat{\mathbf{a}} \in \mathbf{R}^{N+1}$ cercato è la soluzione del sistema lineare

$$\widehat{A}\,\widehat{\mathbf{a}} = \widehat{\mathbf{f}}\,, \tag{7}$$

ove

$$\widehat{a}_{kj} = \int_a^b L_j(x)L_k(x)\,dx \quad , \quad \widehat{f}_k = \int_a^b F(x)L_k(x)\,dx\,.$$

Ora un'osservazione importante: siccome la base di polinomi L_j è ortonormale rispetto al prodotto scalare $\langle\cdot,\cdot\rangle$, si ha che

$$\int_a^b L_j(x)L_k(x)\,dx = \begin{cases} 1 & \text{se } j = k \\ 0 & \text{se } j \neq k \end{cases}\,,$$

e di conseguenza la matrice $\widehat{A}$ è la matrice identità! La soluzione di $\widehat{A}\,\widehat{\mathbf{a}} = \widehat{\mathbf{f}}$ è dunque $\widehat{\mathbf{a}} = \widehat{\mathbf{f}}$!

In conclusione: utilizzando la base di polinomi ortonormali L_j (si chiamano polinomi di Legendre), il polinomio di distanza minima dalla funzione F è

$$P_N(x) = \sum_{j=0}^{N} \widehat{a}_j L_j(x) \quad , \quad \widehat{a}_j = \int_a^b F(x)L_j(x)\,dx\,,$$

di calcolo immediato (se si hanno a disposizione i polinomi di Legendre L_j...).

Va infine osservato che questi polinomi, oltre a poter essere determinati ogni volta che occorra a partire da $1, x, \ldots, x^N$ tramite il procedimento di Gram–Schmidt, sono stati calcolati esplicitamente per ogni grado N tramite una formula di ricorrenza e sono riportati nei libri che trattano di questi argomenti. Per esempio, nell'intervallo $[-1, 1]$ si ha

$$L_0(x) = \frac{1}{\sqrt{2}}\,,\; L_1(x) = \frac{\sqrt{3}}{\sqrt{2}}\,x\,,\; L_2(x) = \frac{\sqrt{5}}{\sqrt{2}}\Big(\frac{3}{2}\,x^2 - \frac{1}{2}\Big)\,,\; L_3(x) = \frac{\sqrt{7}}{\sqrt{2}}\Big(\frac{5}{2}\,x^3 - \frac{3}{2}\,x\Big)\,,\; \ldots\,.]$$

- Un altro problema: "polinomio" trigonometrico di distanza minima

Un problema simile al precedente è quello di trovare un "polinomio" trigonometrico Q_N di distanza minima da una data funzione (continua) F. (Per "polinomio" trigonometrico si intende una funzione del tipo

$$Q_N(x) = A_0 + \sum_{j=1}^{N} [A_j \cos(jx) + B_j \sin(jx)]\,;$$

per esempio, noi consideriamo qui nel seguito una funzione della forma $Q_N(x) = \sum_{j=1}^{N} B_j \sin(jx)$, e l'intervallo $[0, \pi]$.)

Si tratta dunque di minimizzare la funzione

$$S(B_1, \ldots, B_N) = \int_0^\pi \Big(\sum_{j=1}^{N} B_j \sin(jx) - F(x)\Big)^2 dx\,,$$

che è un polinomio di secondo grado rispetto alle N variabili $\mathbf{B} = (B_1, \ldots, B_N)$.

[In modo molto simile al caso precedente si può vedere che S ha minimo in $\mathbf{R}^N$. L'unico passaggio che richiede un attimo di riflessione è l'affermazione che $\int_0^\pi \Big(\sum_{j=1}^{N} B_j \sin(jx)\Big)^2 dx = 0$ se e solo se $\mathbf{B} = \mathbf{0}$: ma questo sarà chiaro fra un attimo...]

Come al solito cerchiamo i punti stazionari, derivando sotto al segno di integrale:

$$\begin{aligned}
\frac{\partial S}{\partial B_k}(B_1,\dots,B_N) &= \int_0^\pi \frac{\partial}{\partial B_k}\Big(\sum_{j=1}^N B_j \sin(jx) - F(x)\Big)^2 dx \\
&= 2\int_0^\pi \Big(\sum_{j=1}^N B_j \sin(jx) - F(x)\Big)\frac{\partial}{\partial B_k}\Big(\sum_{j=1}^N B_j \sin(jx) - F(x)\Big) dx \\
&= 2\int_0^\pi \Big(\sum_{j=1}^N B_j \sin(jx) - F(x)\Big)\Big(\sum_{j=1}^N \sin(jx)\frac{\partial B_j}{\partial B_k}\Big) dx\,.
\end{aligned}$$

Siccome

$$\frac{\partial B_j}{\partial B_k} = \begin{cases} 1 & \text{se } j = k \\ 0 & \text{se } j \neq k \end{cases}\,,$$

si conclude che

$$\frac{\partial S}{\partial B_k}(B_1,\dots,B_N) = 2\int_0^\pi \Big(\sum_{j=1}^N B_j \sin(jx) - F(x)\Big)\sin(kx)\,dx\,.$$

Ma il sistema $\sin(jx)$, $j = 1,\dots,N$ è ortogonale rispetto al prodotto scalare $\langle\cdot,\cdot\rangle$ considerato nell'intervallo $[0,\pi]$, dato che vale (integrazione per parti...)

$$\int_0^\pi \sin(jx)\,\sin(kx)\,dx = \begin{cases} \frac{\pi}{2} & \text{se } j = k \\ 0 & \text{se } j \neq k \end{cases}\,. \tag{8}$$

Dunque imponendo che $\frac{\partial S}{\partial B_k} = 2\int_0^\pi \Big(\sum_{j=1}^N B_j \sin(jx) - F(x)\Big)\sin(kx)\,dx = 0$ per ogni $k = 1,\dots,N$, si verifica immediatamente che l'unico punto stazionario di S soddisfa

$$B_k = \frac{2}{\pi}\int_0^\pi F(x)\sin(kx)\,dx\,,\ k = 1,\dots,N\,.$$

Si conclude quindi che il "polinomio" trigonometrico di minima distanza da F è dato da

$$Q_N(x) = \sum_{j=1}^N B_j \sin(jx)\quad,\quad B_j = \frac{2}{\pi}\int_0^\pi F(x)\sin(jx)\,dx\,.$$

Si può vedere che mandando N all'infinito la distanza di Q_N da F tende a 0, cioè che F si può sviluppare in serie: la sua serie di Fourier! [Per chi non ha mai sentito parlare di una serie questa affermazione è un po' misteriosa: diciamo che F coincide con la "somma" degli infiniti addendi $B_j \sin(jx)$, $j = 1, 2, \dots$.]

[Ora è anche chiaro che $\int_0^\pi \Big(\sum_{j=1}^N B_j \sin(jx)\Big)^2 dx = 0$ se e solo se $\mathbf{B} = \mathbf{0}$: infatti innanzitutto si ha $\sum_{j=1}^N B_j \sin(jx) = 0$ per ogni $x \in [0,\pi]$, e dunque moltiplicando per $\sin(kx)$ e integrando su $[0,\pi]$ si ottiene

$$\int_0^\pi \sum_{j=1}^N B_j \sin(jx)\,\sin(kx) = 0\,,\ k = 1,\dots,N\,.$$

Quindi usando (8) ne segue subito $\frac{\pi}{2}B_k = 0$ per $k = 1,\dots,N$, cioè $\mathbf{B} = \mathbf{0}$.]