9.3 Il metodo dei minimi quadrati in formalismo matriciale

Nel caso si debba applicare il metodo minimi quadrati con molti parametri risulta vantaggioso utilizzare il formalismo matriciale. Oltre alla compattezza della notazione questo formalismo permette una maggiore comprensione teorica del metodo. Per introdurre questo formalismo riprendiamo in esame il caso dell'applicazione del metodo dei minimi quadrati quando è prevista una dipendenza funzionale lineare tra le grandezze in esame X e Y del tipo caso già trattato nel paragrafo 9.2.

$$Y = \beta_0 + \beta_1 X$$

dove β_0 e β_1 sono i parametri da determinare[8]. Per quanto ci siamo limitati a due parametri per un paragone con i calcoli del paragrafo 9.2, le formule in forma matriciale a cui arriveremo sono valide per un generico numero k di parametri, in altre parole si applicano anche a dipendenze di Y da X del tipo polinomiale: $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \ldots + \beta_k X^k$.

Siano $x_1, x_2, \ldots, x_n$ i valori di X in corrispondenza dei quali sono state misurate le grandezze $y_1, y_2, \ldots, y_n$. Le grandezze x_i hanno incertezze trascurabili e supponiamo che le y_i siano tra loro indipendenti e che abbiano incertezze pari a σ_i. Definiamo inoltre il peso $w_i = 1/\sigma_i^2$ di ogni valore y_i. Per ognuna delle N misurazioni possiamo scrivere:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_1 + e_1 \\ y_2 &= \beta_0 + \beta_1 x_2 + e_2 \\ &\ldots \\ y_N &= \beta_0 + \beta_1 x_N + e_N \end{aligned} \tag{9.16}$$

nelle (9.16) è stato introdotto l'*errore* e_i (detto anche residuo) che tiene conto delle fluttuazioni delle grandezze y_i. Gli errori e_i hanno le seguenti proprietà:

$$\mathbb{E}[e_i] = 0, \qquad \mathbb{V}\text{ar}[e_i] = \mathbb{V}\text{ar}[y_i] = \sigma_i^2.$$

Introduciamo le matrici: $\boldsymbol{Y}$ matrice $(n \times 1)$ delle osservazioni, $\boldsymbol{X}$ matrice $(n \times 2)$ dei valori delle X, $\boldsymbol{\beta}$ matrice (2×1) dei parametri[9], la matrice $\boldsymbol{e}$ $(n \times 1)$ degli errori e la matrice $\boldsymbol{W}$ dei pesi.

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_N \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \ldots & \ldots \\ 1 & x_N \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \quad \boldsymbol{e} = \begin{bmatrix} e_1 \\ e_2 \\ \ldots \\ e_N \end{bmatrix}, \boldsymbol{W} = \begin{bmatrix} w_1 & 0 & \ldots & 0 \\ 0 & w_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & w_n \end{bmatrix}$$

Usando queste matrici le N equazioni (9.16) si scrivono in forma compatta come:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e} \tag{9.17}$$

Il metodo dei minimi quadrati pesato afferma che il valore migliore dei parametri è quello che minimizza la somma dei residui (e_i) al quadrato pesati, che in notazione matriciale si scrive:

$$\sum_{i=1}^{N} w_i {e_i}^2 = [e_1 \; e_2 \; \ldots \; e_N] \begin{bmatrix} w_1 & 0 & \ldots & 0 \\ 0 & w_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & w_N \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \ldots \\ e_N \end{bmatrix} \equiv \boldsymbol{e}^T \boldsymbol{W} \boldsymbol{e}$$

[8]La corrispondenza con la notazione usata nel paragrafo 9.2 è: $a = \beta_0$, $b = \beta_1$. Utilizzare questa definizione dei parametri risulta utile per l'applicazione del formalismo matriciale.

[9]Nel caso di un modello polinomiale di grado k la matrice $\boldsymbol{\beta}$ avrà dimensioni $(k \times 1)$.

Usando la (9.17) otteniamo l'espressione da minimizzare:

$$(\boldsymbol{Y} - \boldsymbol{X\beta})^T \boldsymbol{W} (\boldsymbol{Y} - \boldsymbol{X\beta})$$

Prima di differenziare rispetto a β sviluppiamo questa espressione. Si ha:

$$\begin{aligned}(\boldsymbol{Y} - \boldsymbol{X\beta})^T \boldsymbol{W} (\boldsymbol{Y} - \boldsymbol{X\beta}) &= \boldsymbol{Y}^T\boldsymbol{WY} - \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{WY} - \boldsymbol{Y}^T\boldsymbol{WX\beta} + \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{WX\beta} \\ &= \boldsymbol{Y}^T\boldsymbol{WY} - 2\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{WY} + \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{WX\beta}\end{aligned}$$

dove si è usata la proprietà che dice che la matrice trasposta di un prodotto di matrici è uguale al prodotto delle singole matrici trasposte ma in ordine inverso e che il trasposto di uno scalare è lo scalare stesso Derivando rispetto a β e uguagliando a zero si ha (si ricordi che questa espressione è uno scalare):

$$\frac{\partial}{\partial \beta} \sum w_i e_i^2 = -2\boldsymbol{X}^T\boldsymbol{WY} + 2\boldsymbol{X}^T\boldsymbol{WX\beta} = 0$$

da cui otteniamo la seguente equazione per determinare i parametri β, detta "*equazione normale*":

$$\boldsymbol{X}^T\boldsymbol{WX\beta} = \boldsymbol{X}^T\boldsymbol{WY} \tag{9.18}$$

In formalismo matriciale le stime $\hat{\boldsymbol{\beta}}$ dei parametri $\boldsymbol{\beta}$ sono dati dalla soluzione dell'equazione (9.18):

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{WX})^{-1}\boldsymbol{X}^T\boldsymbol{WY} \tag{9.19}$$

La (9.19) non è nulla di nuovo rispetto a quanto visto nel paragrafo 9.2. Per conferma ed esercizio calcoliamo espressamente le matrici che compaiono nella (9.19) nel caso già esaminato del fit ad una retta:

$$\boldsymbol{X}^T\boldsymbol{WX} = \begin{bmatrix} \sum w_i & \sum w_i x_i \\ \sum w_i x_i & \sum w_i x_i^2 \end{bmatrix}, \ (\boldsymbol{X}^T\boldsymbol{WX})^{-1} = \frac{1}{\Delta}\begin{bmatrix} \sum w_i x_i^2 & -\sum w_i x_i \\ -\sum w_i x_i & \sum w_i \end{bmatrix},$$

$$\boldsymbol{X}^T\boldsymbol{WY} = \begin{bmatrix} \sum w_i y_i \\ \sum w_i x_i y_i \end{bmatrix} \tag{9.20}$$

dove $\Delta = \sum w_i \sum w_i x_i^2 - (\sum w_i x_i)^2$. In conclusione nel caso di due parametri si ottiene la soluzione:

$$\hat{\boldsymbol{\beta}} \equiv \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \frac{1}{\Delta}\begin{bmatrix} \sum w_i x_i^2 \sum w_i y_i - \sum w_i x_i \sum w_i x_i y_i \\ \sum w_i \sum w_i x_i y_i - \sum w_i x_i \sum w_i y_i \end{bmatrix} =$$

che sono esattamente uguali alle soluzioni precedente trovate. Vedi le (9.7).

9.3.1 Matrice di covarianza dei parametri

La forma matriciale del calcolo dei parametri dati dalla (9.19) permette di arrivare rapidamente all'espressione della cosiddetta *matrice di covarianza* dei parametri stimati con il metodo dei minimi quadrati. In generale date N variabili aleatorie $\boldsymbol{x} = \{x_1, x_2, \ldots, x_N\}$, si definisce matrice di covarianza e alle volte indicata con $\boldsymbol{\Sigma}$, la matrice:

$$\begin{bmatrix} \mathbb{C}\text{ov}[x_1, x_1] & \mathbb{C}\text{ov}[x_1, x_2] & \ldots & \mathbb{C}\text{ov}[x_1, x_N] \\ \mathbb{C}\text{ov}[x_2, x_1] & \mathbb{C}\text{ov}[x_2, x_2] & \ldots & \mathbb{C}\text{ov}[x_2, x_N] \\ \ldots & \ldots & \ldots & \ldots \\ \mathbb{C}\text{ov}[x_N, x_1 & \mathbb{C}\text{ov}[x_2, x_2] & \ldots & \mathbb{C}\text{ov}[x_2, x_N] \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \ldots & \rho_{1N}\sigma_1\sigma_N \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \ldots & \rho_{2N}\sigma_2\sigma_N \\ \ldots & \ldots & \ldots & \ldots \\ \rho_{N1}\sigma_1\sigma_2 & \rho_{12}\sigma_1\sigma_2 & \ldots & \sigma_N^2 \end{bmatrix}$$

dove nell'ultima matrice sono state introdotte le varianze e i coefficienti di correlazione delle variabili x_i. La matrice di covarianza è simmetrica e i suoi elementi diagonali sono le varianze delle variabili. Se le N variabili x_i sono indipendenti tra loro la matrice di covarianza è diagonale ($\rho_{ij} = 0,\ i \neq j$).

Si noti che $\hat{\boldsymbol{\beta}}$ può essere interpretata come variabile aleatoria poiché, come indica la (9.19), è una funzione della variabile aleatoria y e quindi possiamo calcolarne la matrice di covarianza. Per definizione la matrice di covarianza di $\hat{\boldsymbol{\beta}}$ è:

$$\mathbb{E}[(\hat{\beta} - \mathbb{E}[\hat{\boldsymbol{\beta}}])(\hat{\boldsymbol{\beta}} - \mathbb{E}[\hat{\boldsymbol{\beta}}])^T] \tag{9.21}$$

Dimostriamo che $\mathbb{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$, dalle (9.19) e (9.17) otteniamo:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}) = \boldsymbol{\beta} + (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{e} \tag{9.22}$$

Prendendo i valori attesi dei due membri:

$$\mathbb{E}[\hat{\boldsymbol{\beta}}] = \mathbb{E}[\boldsymbol{\beta} + (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{e}] = \mathbb{E}[\boldsymbol{\beta}] + (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\,\mathbb{E}[\boldsymbol{e}] = \boldsymbol{\beta}$$

dove si è usato $\mathbb{E}[\boldsymbol{e}] = 0$. Poiché $\mathbb{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$, dalla (9.22) sostituendo $\boldsymbol{\beta}$ con $\mathbb{E}[\hat{\boldsymbol{\beta}}]$, ricaviamo:

$$\hat{\boldsymbol{\beta}} - \mathbb{E}[\hat{\boldsymbol{\beta}}] = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{e}$$

inseriamo infine questo risultato nella (9.21):

$$\begin{aligned}\boldsymbol{\Sigma}\left[\hat{\boldsymbol{\beta}}\right] \equiv \mathbb{E}[(\hat{\boldsymbol{\beta}} - \mathbb{E}[\hat{\boldsymbol{\beta}}])(\hat{\boldsymbol{\beta}} - \mathbb{E}[\hat{\boldsymbol{\beta}}])^T] &= \mathbb{E}\left[[(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{e}][(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{e}]^T\right] \\ &= \mathbb{E}\left[(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{e}\boldsymbol{e}^T\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\right] \\ &= (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\,\mathbb{E}[\boldsymbol{e}\boldsymbol{e}^T]\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\end{aligned} \tag{9.23}$$

La matrice $\boldsymbol{e}\boldsymbol{e}^T$ ha elementi $e_ie_j, (i,j = 1, N)$ e considerando che $\mathbb{E}[e_ie_j] = \sigma_i^2\delta_{ij}$, in quanto per ipotesi le grandezze y_i sono indipendenti, il valore atteso $\mathbb{E}[\boldsymbol{e}\boldsymbol{e}^T]$ è una matrice diagonale $N \times N$ di elementi σ_i^2. Ricordando che $\boldsymbol{W}$ è una matrice diagonale di elementi $w_i = 1/\sigma_i^2$, si ha: $\mathbb{E}[\boldsymbol{e}\boldsymbol{e}^T]\boldsymbol{W} = \boldsymbol{I}$ dove $\boldsymbol{I}$ è la matrice identità. Con queste osservazione la (9.23) diviene:

$$(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}$$

Infine possiamo scrivere a matrice di covarianza $\boldsymbol{\Sigma}\left[\hat{\boldsymbol{\beta}}\right]$ dei parametri β_i:

$$\boldsymbol{\Sigma}\left[\hat{\boldsymbol{\beta}}\right] = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1} \tag{9.24}$$

Come esercizio di calcolo, applichiamo questa relazione al caso del fit lineare. Usando le matrici (9.20) precedentemente trovate otteniamo con facili calcoli:

$$\begin{bmatrix} \mathbb{V}\text{ar}[b_0] & \mathbb{C}\text{ov}[b_0, b_1] \\ \mathbb{C}\text{ov}[b_1, b_0] & \mathbb{V}\text{ar}[b_1] \end{bmatrix} = \frac{1}{\Delta}\begin{bmatrix} \sum w_ix_i^2 & -\sum w_ix_i \\ -\sum w_ix_i & \sum w_i \end{bmatrix}$$

Come era prevedibile, le varianze sui parametri, coincidenti con i termini diagonali della matrice, sono uguali a quelle ottenute con la formula di propagazione delle incertezze date dalle (9.11). Inoltre il formalismo matriciale ha permesso di ottenere facilmente la covarianza dei parametri.

x_i	y_i	u_i
40	4.16	0.03
60	2.30	0.06
70	1.48	0.05
80	0.99	0.04
87	0.70	0.07
97	0.58	0.06
105	0.47	0.07
113	0.40	0.07
120	0.49	0.06
130	0.76	0.02
135	0.97	0.05
145	1.45	0.05
160	2.51	0.06

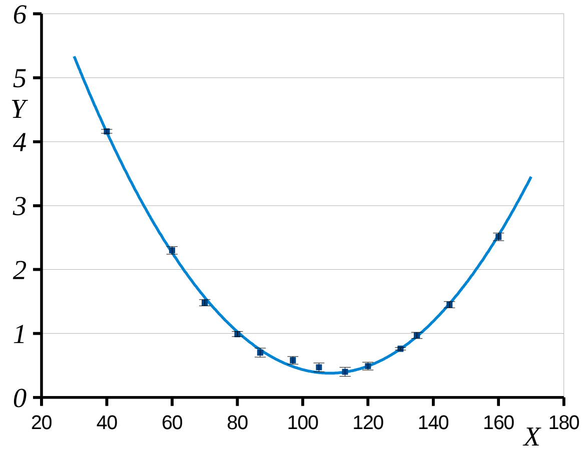

Figura 9.4: Tabelle dei dati, grafico dei punti sperimentali con le relative incertezze e fit parabolico ai dati ottenuto con il metodo dei minimi quadrati come indicato nel testo.

Minimi Quadrati parabolico. Come esempio pratico, applichiamo il metodo descritto nel paragrafo precedente per ottenere un *fit* parabolico a dati che si ipotizza abbiano un andamento di potenza di ordine 2. I dati sui quali si vuole adattare una parabola sono riportati nella tabella inserita nella figura 9.4. La curva che si vuole adattare all'andamento sperimentale è:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

Definendo ora la matrice $\boldsymbol{X}$ e la matrice $\boldsymbol{\beta}$ come:

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \dots & \dots & \dots \\ 1 & x_N & x_N^2 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix},$$

e usando per le le matrici $\boldsymbol{Y}$ e $\boldsymbol{W}$ le definizioni già date, possiamo esplicitare la ??? in questo caso come

$$\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} = \begin{bmatrix} \sum w_i & \sum w_i x_i & \sum w_i x_i^2 \\ \sum w_i x_i & \sum w_i x_i^2 & \sum w_i x_i^3 \\ \sum w_i x_i^2 & \sum w_i x_i^3 & \sum w_i x_i^4 \end{bmatrix}, \quad \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y} = \begin{bmatrix} \sum w_i y_i \\ \sum w_i x_i y_i \\ \sum w_i x_i^2 y_i \end{bmatrix}$$

Il sistema lineare da risolvere per trovare la stima dei parametri con i minimi quadrati è:

$$(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})\boldsymbol{\beta} = \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y} \tag{9.25}$$

Il calcolo algebrico della matrice $(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}$ porta ad espressioni troppo complesse per cui passeremo alla valutazione numerica delle matrici. Con i dati della tabella infigura 9.4 otteniamo:

$$\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} = \begin{bmatrix} 7.1595\times10^3 & 7.4308\times10^5 & 8.6813\times10^7 \\ 7.4308\times10^5 & 8.6813\times10^7 & 1.0821\times10^{10} \\ 8.6813\times10^7 & 1.0821\times10^{10} & 1.3994\times10^{12} \end{bmatrix}, \quad \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y} = \begin{bmatrix} 1.065\times10^4 \\ 8.729\times10^5 \\ 9.244\times10^7 \end{bmatrix} \tag{9.26}$$

$$(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1} = \begin{bmatrix} 7.5476\times10^{-3} & -1.7258\times10^{-4} & 8.6626\times10^{-7} \\ -1.7258\times10^{-4} & 4.2646\times10^{-6} & -2.2271\times10^{-8} \\ 8.6626\times10^{-7} & -2.2271\times10^{-8} & 1.1919\times10^{-10} \end{bmatrix}, \tag{9.27}$$

I parametri che minimizzano la somma dei residui quadrati sono:

$$= (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y} = \begin{bmatrix} 9.8518 \\ -0.174910 \\ 8.0737\times10^{-4} \end{bmatrix} \tag{9.28}$$

Con i valori dei parametri ottenuti in questo modo possiamo scrivere la funzione che meglio si adatta ai dati sperimentali secondo il metodo dei minimi quadrati:

$$y(x) = \beta_0 + \beta_1 x + \beta_2 x^2 \tag{9.29}$$

L'andamento della parabola (9.29) è mostrato, assieme ai punti sperimentali nella figura 9.4. Ricordando che la (9.27) è la matrice di covarianza dei parametri β_i otteniamo i parametri con la loro incertezza standard:

$$\beta_0 = 9.852 \pm 0.087, \quad \beta_1 = -0.1749 \pm 0.0021, \quad \beta_2 = (8.07 \pm 0.11)\cdot 10^{-4}$$

Non si deve tuttavia dimenticare che i valori di questi parametri sono correlati poiché i termini non diagonali della matrice di covarianza sono diversi da zero. E' utile per comprendere l'entità della correlazione calcolare la matrice di correlazione $\boldsymbol{\rho}$ i cui elementi si ottengono da quella di covarianza tramite la relazione $\rho_{ij} = \mathsf{cov}_{ij}/\sigma_i\sigma_j$:

$$\boldsymbol{\rho} = \begin{bmatrix} 1 & -0.962 & 0.913 \\ & 1 & -0.988 \\ & & 1 \end{bmatrix}, \tag{9.30}$$

Si noti la forte correlazione tra i tre parametri che indica come questi valori siano strettamente legati tra loro. Ad esempio se β_0 avesse una fluttuazione positiva β_1 avrebbe con alta probabilità una fluttuazione negativa e β_2 positiva.