

## 2.8 APPROSSIMAZIONE AI MINIMI QUADRATI

Il problema dell'approssimazione ai minimi quadrati è uno dei più classici che si possa incontrare nelle applicazioni ed è motivato dal fatto che molto spesso non ha senso chiedere che la funzione interpolante passi esattamente per le condizioni di interpolazione date, poichè i dati sono affetti da errori. In genere si possono distinguere due tipi di approssimazione ai minimi quadrati: quella *discreta* e quella *continua*. Inizieremo con l'esaminare quella discreta.

Si supponga di avere il solito problema di interpolazione delle coppie  $(x_i, y_i)$  per  $i = 0, \dots, n$ . Si supponga che i valori  $y_i$  siano affetti da certi errori derivanti ad esempio dal fatto che sono stati ottenuti per via sperimentale mediante misurazioni. Allora non ha molto senso richiedere che la funzione interpolante passi esattamente per le ordinate  $y_i$ . Sarà sufficiente cercare una funzione  $q(x)$ , appartenente ad una determinata classe di funzioni  $\mathcal{I}$ , che nei nodi  $x_i$  si discosti il meno possibile dai valori  $y_i$ , cioè tale da risolvere il problema di minimo

$$\min_{y(x) \in \mathcal{I}} \sum_{i=0}^n [y(x_i) - y_i]^2 = \sum_{i=0}^n [q(x_i) - y_i]^2. \quad (2.39)$$

La classe di funzioni  $\mathcal{I}$  in cui cercare la  $q(x)$  può essere quella dei polinomi, quella delle funzioni esponenziali, delle funzioni trigonometriche, etc.

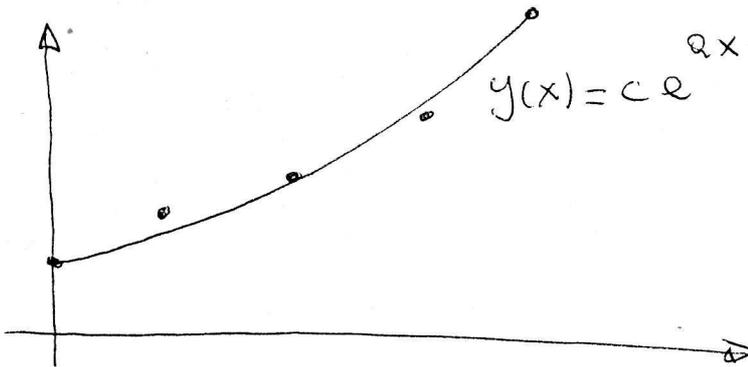


FIG. 2.11 - Minimi quadrati mediante una funzione esponenziale.

**2.8.1 Approssimazione polinomiale ai minimi quadrati nel discreto**

La funzione  $q(x)$  in (2.39) può essere cercata nell'insieme dei polinomi  $\Pi_m$  di grado  $m$  con  $m \leq n$ . In particolare deve risultare  $m < n$ , poiché se fosse  $m = n$  allora avremmo che  $y(x)$  coinciderebbe proprio con il polinomio interpolante di Lagrange di grado  $n$ . Si cerca allora  $q_m(x)$  nella forma

$$q_m(x) = \sum_{j=0}^m a_j x^j$$

ed il problema in (2.39) equivale a trovare i coefficienti  $a_0, a_1, \dots, a_m$ , tali da minimizzare la seguente funzione

$$f(a_0, a_1, \dots, a_m) = \sum_{i=0}^n \left[ \sum_{j=0}^m a_j x_i^j - y_i \right]^2. \tag{2.40}$$

La minimizzazione della  $f(a_0, a_1, \dots, a_m)$  si ottiene imponendo che si annullino le derivate parziali della  $f$  rispetto a ciascuna variabile  $a_k$ , cioè

$$\frac{\partial}{\partial a_k} \sum_{i=0}^n \left[ \sum_{j=0}^m a_j x_i^j - y_i \right]^2 = 0, \quad k = 0, \dots, m,$$

da cui

$$\sum_{i=0}^n \left[ \sum_{j=0}^m a_j x_i^j - y_i \right] x_i^k = 0, \quad k = 0, \dots, m$$

cioè

$$\sum_{j=0}^m c_{jk} a_j = \sum_{i=0}^n x_i^k y_i, \quad \text{per } k = 0, \dots, m, \tag{2.41}$$

dove

$$c_{jk} = \sum_{i=0}^n x_i^{j+k}, \quad \text{per } j, k = 0, \dots, m.$$

Ponendo ora  $\mathbf{z} = (a_0, a_1, \dots, a_m)^T$ ,  $\mathbf{b} = (y_0, y_1, \dots, y_n)^T$  e

$$\mathbf{A} = \begin{pmatrix} 1 & x_0 & x_0^2 & \dots & \dots & x_0^m \\ 1 & x_1 & x_1^2 & \dots & \dots & x_1^m \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & \dots & x_n^m \end{pmatrix}_{(n+1) \times (m+1)}$$

$$\mathbf{A}^T = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_0 & x_1 & \dots & x_n \\ \dots & \dots & \dots & \dots \\ x_0^m & x_1^m & \dots & x_n^m \end{pmatrix}_{(m+1) \times (n+1)}$$

il sistema in (2.41) può essere riscritto come

$$\mathbf{A}^T \mathbf{A} \mathbf{z} = \mathbf{A}^T \mathbf{b} \quad (2.42)$$

che prende il nome di sistema delle *equazioni normali*. Si verifica inoltre la matrice delle derivate parziali seconde della  $f(a_0, a_1, \dots, a_m)$ , cioè

$$H(a_0, a_1, \dots, a_m) = \left( \frac{\partial^2 f}{\partial a_i \partial a_j} \right) = \mathbf{A}^T \mathbf{A}$$

risulta essere definita positiva e quindi la soluzione di (2.42) è il vettore di minimo per la funzione in (2.40).

**TEOREMA 2.6.** *La matrice  $\mathbf{A}^T \mathbf{A}$  in (2.42) è simmetrica, non singolare e definita positiva.*

**DIM.** La matrice  $\mathbf{A}^T \mathbf{A}$  è ovviamente simmetrica e risulta non singolare. Infatti se così non fosse allora il sistema  $\mathbf{A}^T \mathbf{A} \mathbf{z} = 0$  avrebbe anche una soluzione  $\mathbf{d} = (d_0, d_1, \dots, d_m)^T$  diversa da quella banale. In corrispondenza di questa soluzione avremmo che  $\mathbf{d}^T \mathbf{A}^T \mathbf{A} \mathbf{d} = 0$ , da cui  $\mathbf{A} \mathbf{d} = 0$ . Ma dire che  $\mathbf{A} \mathbf{d} = 0$  equivale a dire che

$$\sum_{j=0}^m d_j x_i^j = 0, \quad i = 0, \dots, n$$

da cui segue che il polinomio di grado  $m$  di coefficienti  $d_0, d_1, \dots, d_m$  avrebbe  $n+1$  zeri  $x_0, \dots, x_n$ , il che è assurdo essendo  $n > m$ . Possiamo allora concludere che la matrice  $\mathbf{A}^T \mathbf{A}$  è nonsingolare. Per dimostrare che è anche definita positiva si osservi che  $\mathbf{z}^T \mathbf{A}^T \mathbf{A} \mathbf{z} \geq 0$  per ogni  $\mathbf{z} \neq 0$ . In particolare non può essere  $\mathbf{z}^T \mathbf{A}^T \mathbf{A} \mathbf{z} = 0$  altrimenti, come nella dimostrazione della non singolarità, si ricadrebbe in un assurdo ■

Il grado del polinomio ai minimi quadrati può essere dato a priori, osservando la distribuzione dei valori delle  $y_i$  da approssimare, oppure si può considerare il polinomio ai minimi quadrati di grado via via crescente e fermarsi quando lo scarto

$$\sum_{i=0}^n [q_m(x_i) - y_i]^2 = \delta_m,$$

non varia molto passando da  $m$  a  $m+1$ .

La soluzione del sistema (2.42) potrà ad esempio essere determinata con il metodo di Cholesky (vedi [3,14]). Comunque si deve osservare che per  $m$  elevato la matrice  $\mathbf{A}^T \mathbf{A}$  diventa mal condizionata ed in questo caso è opportuno determinare

$$\begin{aligned} \mathbf{A}_{k \cdot}^T &= \begin{pmatrix} x_0^k & x_1^k & \dots & x_\mu^k \end{pmatrix} \\ \mathbf{A}_{\cdot j} &= \begin{pmatrix} x_0^j \\ x_1^j \\ \vdots \\ x_\mu^j \end{pmatrix} \\ c_{kj} &= (\mathbf{A}_{k \cdot}^T) \cdot (\mathbf{A}_{\cdot j}) = \sum_{\lambda=0}^{\mu} x_\lambda^{k+j} \end{aligned}$$

il polinomio ai minimi quadrati mediante altre tecniche evitando la risoluzione di (2.42).

Per mettere in evidenza il mal condizionamento del sistema normale, consideriamo nodi  $x_0, x_1, \dots, x_n$  in  $[0,1]$ , ed osserviamo che per  $n$  abbastanza grande avremo

$$\sum_{i=0}^n x_i^{j+k} \cong (n+1) \int_0^1 x^{j+k} dx = (n+1) \frac{1}{j+k+1}, \quad j, k = 0, \dots, m,$$

da cui segue che  $\mathbf{A}^T \mathbf{A} \cong (n+1) \mathbf{H}$ , dove  $\mathbf{H}$  è la matrice di Hilbert seguente

$$\mathbf{H} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots & \cdots & \frac{1}{m+1} \\ \frac{1}{2} & \frac{1}{3} & \cdots & \cdots & \cdots & \frac{1}{m+2} \\ \vdots & \cdot & \cdot & \cdots & \cdots & \cdots \\ \cdot & \cdot & \cdot & \cdots & \cdots & \cdots \\ \cdot & \cdot & \cdot & \cdots & \cdots & \cdots \\ \frac{1}{m+1} & \frac{1}{m+2} & \cdots & \cdots & \cdots & \frac{1}{2m+1} \end{pmatrix}.$$

La matrice  $\mathbf{H}$  è mal condizionata anche per valori di  $m$  relativamente piccoli ( $m \geq 9$ ). La causa del mal condizionamento del sistema normale è dovuta al fatto che  $\{1, x, x^2, \dots, x^m\}$  non è una *buona* base per l'insieme dei polinomi di grado  $m$ , nel senso che alcuni elementi di questa base approssimano elementi *linearmente dipendenti*.

Il sistema normale può essere risolto in modo più efficiente osservando che il problema di minimo in (2.40) equivale al problema di determinare il vettore  $\mathbf{w}$  che verifica

$$\min_z \|\mathbf{b} - \mathbf{A}z\|^2 = \|\mathbf{b} - \mathbf{A}\mathbf{w}\|^2$$

dove  $\|\cdot\|$  denota la norma Euclidea su vettori, ed il vettore  $\mathbf{w}$  che verifica il problema di minimo precedente coincide con il vettore soluzione di (2.42).

Il precedente problema di minimo ammette un'unica soluzione quando la matrice  $\mathbf{A}$  ha rango massimo o equivalentemente quando  $\mathbf{A}^T \mathbf{A}$  è non singolare (vedi [3]). La soluzione del problema di minimo può essere determinata usando la fattorizzazione  $\mathbf{QR}$  di Householder della matrice rettangolare  $\mathbf{A}$ , con  $\mathbf{Q}$  matrice unitaria ed  $\mathbf{R}$  matrice con le stesse dimensioni di  $\mathbf{A}$ , in particolare

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_1 \\ 0 \end{pmatrix}$$

con  $\mathbf{R}_1$  matrice triangolare superiore di dimensione  $m+1$ . Allora si avrà

oss. Se  $\mathbf{A}$  è di rango massimo  $(m+1)$  allora

$$\mathbf{A}^T \mathbf{A} z = \mathbf{A}^T \mathbf{b} \Leftrightarrow \min_z \|\mathbf{b} - \mathbf{A}z\|^2$$

$$\|b - Az\|^2 = \|QQ^T b - QRz\|^2 = \|c - Rz\|^2 = \|c_2\|^2 + \|c_1 - R_1 z\|^2$$

dove

$$c = Q^T b = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}; \quad Rz = \begin{pmatrix} R_1 z \\ 0 \end{pmatrix};$$

con  $c_1 \in^{m+1}$ ,  $c_2 \in^{n-m}$ . Quindi la precedente norma sarà minimizzata dal vettore  $w$  soluzione del seguente sistema lineare

$$R_1 z = c_1$$

e la norma  $\|c_2\|$  rappresenterà il residuo relativo al vettore di minimo  $w$ , cioè

$$\|b - Aw\| = \|c_2\|.$$

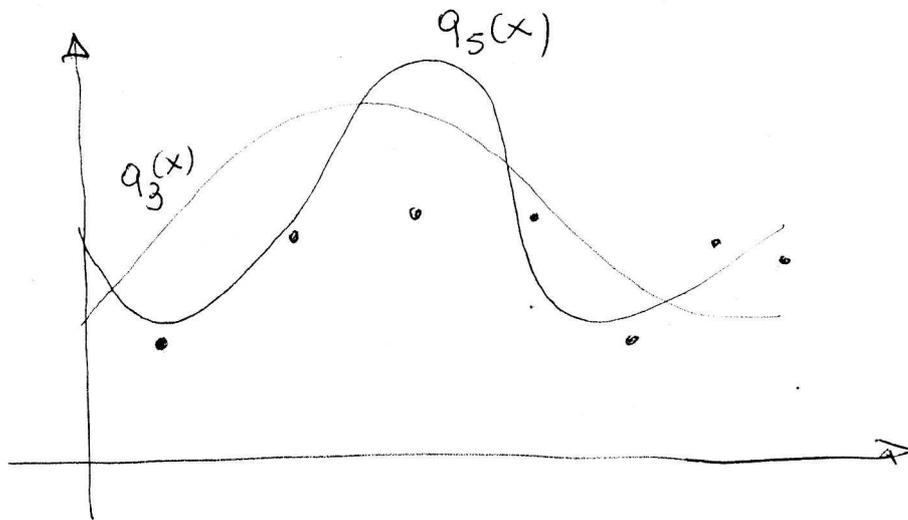


FIG. 2.12 - Minimi quadrati con polinomi di grado 3 e 5.

Oss.  $Q^T Q = I \Rightarrow \|x\|_2 = \|Qx\|_2$

$$\|x\|_2 = \sqrt{x^T x} = \sqrt{x^T Q^T Q x} = \|Qx\|_2$$

### 2.8.2 Parametri statistici

Diamo ora alcune definizioni, tipiche della Statistica, che ci saranno utili nel seguito. Assegnati  $n$  valori  $x_1, x_2, \dots, x_n$ , la funzione

$$f(x) = \frac{1}{n} \sum_{i=1}^n (x - x_i)^2$$

denota una misura della distanza del valore  $x$  dall'insieme di valori assegnato. Il punto di minimo di questa funzione è dato da

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

e si dice *media aritmetica* degli  $n$  numeri dati. Il valore della  $f(x)$  nel punto di minimo  $\bar{x}$  si dice *varianza* e si denota con

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (\bar{x} - x_i)^2$$

mentre si dice *deviazione standard o scarto quadratico medio* il valore seguente

$$\sigma = \left( \frac{1}{n} \sum_{i=1}^n (\bar{x} - x_i)^2 \right)^{1/2}$$

e si può facilmente verificare che *le varianze è anche uguale a:*

$$\sigma^2 = \left( \frac{1}{n} \sum_{i=1}^n x_i^2 \right) - \bar{x}^2.$$

La varianza e la deviazione standard misurano la *dispersione* dei dati rispetto al loro valor medio, se  $\sigma$  è piccolo allora i dati saranno raggruppati attorno al valor medio. Assegnate le coppie  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , si possono quindi definire i valori medi

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i,$$

le varianze

$$\sigma_x^2 = \left( \frac{1}{n} \sum_{i=1}^n x_i^2 \right) - \bar{x}^2, \quad \sigma_y^2 = \left( \frac{1}{n} \sum_{i=1}^n y_i^2 \right) - \bar{y}^2,$$

$x_1, \dots, x_n$  possono essere visti come i valori assunti da una variabile aleatoria  $X$

e la covarianza

$$c_{xy} = \frac{1}{n} \sum_{i=1}^n (\bar{x} - x_i)(\bar{y} - y_i) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y}.$$

### 2.8.3 Retta di regressione lineare e coefficiente di correlazione

Naturalmente a causa della semplicità di calcolo l'approssimazione ai minimi quadrati mediante una retta

$$y(x) = a_0 + a_1 x \quad (2.43)$$

è tra le più utilizzate nelle applicazioni. In questo caso

$$A^T = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_0 & x_1 & \dots & x_n \end{pmatrix}_{2 \times (n+1)} \quad A = \begin{pmatrix} 1 & x_0 \\ 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}_{(n+1) \times 2} \quad A^T \cdot A \quad 2 \times 2$$

il sistema delle equazioni normali risulta essere di dimensione 2 e dato da

$$\begin{pmatrix} 1 & \frac{1}{n+1} \sum_{i=0}^n x_i \\ \frac{1}{n+1} \sum_{i=0}^n x_i & \frac{1}{n+1} \sum_{i=0}^n x_i^2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} \frac{1}{n+1} \sum_{i=0}^n y_i \\ \frac{1}{n+1} \sum_{i=0}^n x_i y_i \end{pmatrix}$$

mentre la soluzione risulta essere

$$a_1 = \frac{\frac{1}{n+1} \sum_{i=0}^n x_i y_i - \bar{x} \bar{y}}{\frac{1}{n+1} \sum_{i=0}^n x_i^2 - \bar{x}^2} = \frac{c_{xy}}{\sigma_x^2}, \quad a_0 = \bar{y} - a_1 \bar{x},$$

dove la covarianza  $c_{xy}$ , la varianza  $\sigma_x$ , ed i valori medi  $\bar{x}$ ,  $\bar{y}$ , sono definiti dalle coppie di dati da interpolare  $(x_i, y_i)$  per  $i = 0, \dots, n$ . I valori dei coefficienti della retta (detta anche retta di *regressione lineare*) risultano determinati purché la varianza  $\sigma_x$  sia diversa da zero, cioè purché le coppie  $(x_i, y_i)$  per  $i = 0, \dots, n$  non siano allineate verticalmente fra loro. Dal valore di  $a_0$  segue pure che la retta di regressione lineare passa per il punto  $(\bar{x}, \bar{y})$  definito dai valori medi.

$$\begin{bmatrix} 1 & \bar{x} \\ \bar{x} & (\sigma_x^2 + \bar{x}^2) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \frac{1}{n+1} \sum x_i y_i \end{bmatrix}$$

$$a_1 = \frac{\bar{y}(\sigma_x^2 + \bar{x}^2) - \bar{x} \cdot \frac{1}{n+1} \sum x_i y_i}{\sigma_x^2 + \bar{x}^2 - \bar{x}^2} = \frac{\bar{y}(\sigma_x^2 + \bar{x}^2) - \bar{x}(\bar{x} \cdot y + c_{xy})}{\sigma_x^2}$$

$$\frac{\bar{y} \sigma_x^2 - \bar{x} c_{xy}}{\sigma_x^2}$$

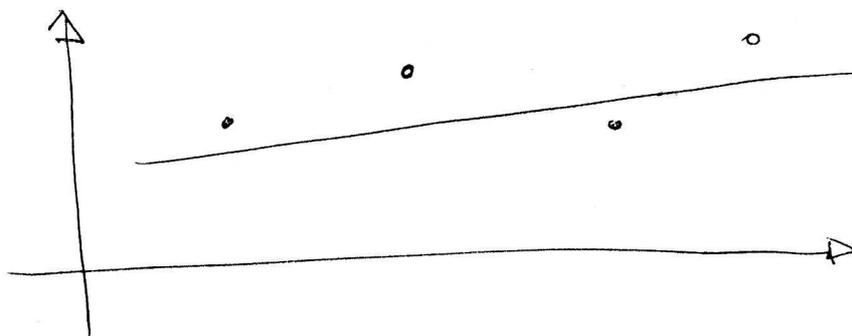


FIG. 2.13 - Retta di regressione lineare.

Naturalmente possiamo porci la domanda se l'insieme di coppie  $(x_i, y_i)$  per  $i = 0, \dots, n$  sia opportunamente rappresentato dalla retta di regressione lineare. Infatti possono esserci altre rette che rappresentano in modo simile l'insieme di coppie date. Come esempio si consideri l'insieme di coppie del piano

$$(1, 1), (1, 3), (1, 5), (3, 1), (3, 3), (3, 5), (5, 1), (5, 3), (5, 5)$$

la cui retta di regressione lineare è la retta  $y = 3$  (vedi Figura 2.14).

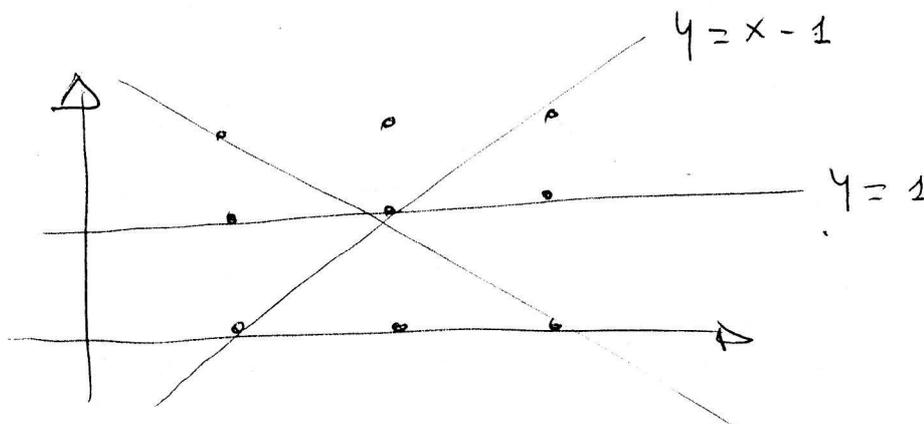


FIG. 2.14 - Esempio di retta di regressione lineare.

Si verifica facilmente che anche la retta  $x = 3$  o la retta  $y = x$  rappresentano allo stesso modo l'insieme di coppie date. È quindi necessario un criterio che riesca a stabilire se è ragionevole rappresentare i dati con un'unica retta. A tal fine si consideri la retta

$$x(y) = a'_0 + a'_1 y \quad (2.44)$$

di regressione lineare per il problema di interpolazione delle coppie

$$(y_0, x_0), (y_1, x_1), \dots, (y_n, x_n)$$

con

$$a'_1 = \frac{\frac{1}{n+1} \sum_{i=0}^n x_i y_i - \bar{x} \bar{y}}{\frac{1}{n+1} \sum_{i=0}^n y_i^2 - \bar{y}^2} = \frac{c_{xy}}{\sigma_y^2}, \quad a'_0 = \bar{x} - a'_1 \bar{y}, \quad \bar{y} = \frac{1}{a'_1} \bar{x} - \frac{a'_0}{a'_1}$$

Dalla (2.44) si ha

$$y = \frac{1}{a'_1} x - \frac{a'_0}{a'_1}$$

da cui segue che anche questa retta passa per  $(\bar{x}, \bar{y})$ . Se  $a_1$  e  $a'_1$  sono non nulli si può definire il *coefficiente di correlazione* come il rapporto

$$r = \sqrt{a_1 a'_1} = \sqrt{\frac{c_{xy}^2}{\sigma_x^2 \sigma_y^2}} \quad \rho = \frac{c_{xy}}{\sigma_x \sigma_y}$$

Si può dimostrare che  $r$  è compreso tra 0 ed 1. Se  $r = 1$  allora  $a_1 = 1/a'_1$ , le due rette di regressione lineare hanno lo stesso coefficiente angolare e passano per uno stesso punto  $(\bar{x}, \bar{y})$ , cioè coincidono. In questo caso esiste un'unica retta di regressione lineare ed i dati possono essere ben rappresentati da questa retta. Se  $r < 1$  allora le due rette non coincidono, e più  $r$  è vicino a 0 più i punti sono dispersi rispetto alle due rette. Nell'esempio precedente delle 9 coppie di valori si ha  $c_{xy} = 0$  da cui  $r = 0$  e si può concludere che i punti non sono ben rappresentati da una retta ed un'approssimazione ai minimi quadrati con un polinomio di grado maggiore di 1 sarà più conveniente.

Ci sono dei casi in cui la funzione ai minimi quadrati di (2.39) pur essendo non lineare può essere determinata con la tecnica di regressione lineare esposta. Ad esempio se la funzione ai minimi quadrati è cercata nelle classi di funzioni dipendenti da due parametri e definite dalle seguenti espressioni

$$y(x) = a + b \log x$$

$$y(x) = ax^b$$

$$y(x) = ae^{bx}$$

$$y(x) = \frac{1}{a + be^{-x}}$$

allora mediante un cambiamento di variabili si può ricondurre il calcolo a quello della determinazione della retta di regressione lineare.

Nel primo caso, posto  $x' = \log x$ , si ottiene

$$y = ax' + b$$

e si calcola la retta di regressione lineare per il problema  $(x'_i, y_i)$ , per  $i = 0, \dots, n$ , in modo da ottenere i parametri  $a$  e  $b$ .

Nel secondo caso, se  $a, x, y > 0$  si pone  $x' = \log x$ ,  $y' = \log y$ , da cui si ottiene

$$y' = \log a + bx'$$

e con il metodo della retta di regressione lineare si calcolano  $b$  e  $\log a$ .

Il terzo caso è simile al precedente e se  $a, y > 0$  si pone  $y' = \log y$ , in modo da ottenere

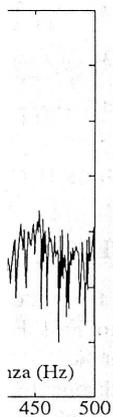
$$y' = \log a + bx.$$

Infine, nell'ultimo caso si pone  $x' = e^{-x}$ ,  $y' = 1/y$ , in modo da ottenere

$$y' = a + bx'.$$

da una distribuzione  
4. Calcoliamo, quindi,

ghezza  $n$  del vettore,



spetto alle frequenze.  
eressano le frequenze  
nmetria della trasfor-  
tarsi a considerare le  
re allora costruito nel

nponenti moltiplicate  
ruzione

semilogy( $f, Pyy$ )

produce il grafico riprodotto in Figura 11.15, dal quale si vedono i due picchi corrispondenti alle frequenze 50 Hz e 120 Hz corrispondenti al segnale  $y$ .

La procedura seguita in precedenza per la stima dello spettro di frequenza è stata indicata a solo scopo illustrativo. In realtà, stime migliori possono essere ottenute mediante altre funzioni (`spectrum`, `cceps`) per le quali rinviamo alla guida di utilizzo del package.

### 11.3 Introduzione al filtro di Kalman

In questo paragrafo vengono introdotte le idee essenziali relative alla *teoria del filtro di Kalman*, che rappresenta uno degli strumenti più interessanti nel *filtraggio* e nella *regolarizzazione (smoothing)* dei segnali, nonché nel campo della *previsione* in condizioni *dinamiche*<sup>8</sup>. Dal punto di vista matematico lo studio del filtro di Kalman può essere inquadrato nell'ambito del *metodo dei minimi quadrati*. Più precisamente, come vedremo, da una parte si estende il metodo al caso in cui la stima ottimale di una quantità fissata viene modificata da successive valutazioni della quantità (*metodo dei minimi quadrati ricorsivo*) e dall'altra al caso in cui la quantità stessa da stimare è la soluzione di un sistema dinamico e quindi varia successivamente (*identificazione ricorsiva*). Pertanto, per una migliore comprensione possono essere opportuni alcuni richiami del metodo dei minimi quadrati (cfr. Appendice A e Capitolo 8).

#### 11.3.1 Metodo dei minimi quadrati con peso

In forma matriciale il metodo dei minimi quadrati può essere visto come un metodo per definire e calcolare la soluzione di un sistema sovradeterminato  $Ax = b$ , con  $A$  matrice di  $m$  righe e  $n$  colonne e  $m \geq n$ . Per il seguito supporremo che la matrice  $A$  abbia rango  $n$ , ossia che le sue colonne siano dei vettori indipendenti. Il vettore  $b \in \mathbb{R}^m$  corrisponde ai valori delle *misurazioni* e  $x \in \mathbb{R}^n$  è il vettore dei *parametri* da stimare. La soluzione del sistema secondo i minimi quadrati è il vettore  $\hat{x}$  che minimizza la lunghezza euclidea  $\|e\|^2$  del vettore errore  $e = b - Ax$ . Geometricamente,  $A\hat{x}$  è la proiezione di  $b$  sullo spazio generato dalle colonne di  $A$ ;

<sup>8</sup>La teoria è stata introdotta da Kalman (1960) e Kalman, Bucy (1961) per superare, in sostanza, le difficoltà inerenti alla *teoria di Wiener-Kolmogorov* ([155], 1949), per la quale è cruciale l'ipotesi che i processi che descrivono i segnali e i rumori siano di tipo *stazionario*. Tra le numerose applicazioni *real world* del filtro di Kalman segnaliamo, in particolare, il suo utilizzo nell'*analisi delle immagini*, nei *processi chimici* (stima, previsione e controllo di inquinamento), in *fluidodinamica* (previsione del flusso), nei *sistemi di comunicazione* (demodulazione e equalizzazione dei segnali), nell'*industria aerospaziale* (stima della posizione e della velocità di un oggetto mobile).

inoltre, il vettore  $\hat{\mathbf{x}}$  è la soluzione del seguente sistema lineare, detto *sistema delle equazioni normali*

$$\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{b} \Rightarrow \hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} \quad (11.43)$$

Supponiamo, ora, che le  $m$  misurazioni abbiano una differente attendibilità. Si può tenere conto di tale informazione introducendo una matrice  $\mathbf{W}$  con la quale pesare opportunamente le varie componenti del vettore residuo. Si considera, cioè,  $\|\mathbf{W}\mathbf{e}\|^2$  come quantità da minimizzare e il sistema delle equazioni normali si modifica nel seguente modo

$$(\mathbf{W}\mathbf{A})^T \mathbf{W}\mathbf{A} \mathbf{x} = (\mathbf{W}\mathbf{A})^T \mathbf{W}\mathbf{b} \Rightarrow \mathbf{A}^T \mathbf{W}^T \mathbf{W} \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{W}^T \mathbf{W} \mathbf{b}$$

Posto  $\mathbf{C} = \mathbf{W}^T \mathbf{W}$ , e supponendo che la matrice  $\mathbf{A}^T \mathbf{C} \mathbf{A}$  sia invertibile, la soluzione è formalmente<sup>9</sup> data da  $\hat{\mathbf{x}} = \mathbf{L}\mathbf{b}$ , con  $\mathbf{L} = (\mathbf{A}^T \mathbf{C} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{C}$ .

**Scelta della matrice  $\mathbf{C} = \mathbf{W}^T \mathbf{W}$**

Il problema della scelta *ottimale* della matrice  $\mathbf{W}$  è risolto sotto opportune condizioni dal *teorema di Gauss-Markov* analizzato nel Capitolo 8. Brevemente, supponendo che gli errori  $\mathbf{e} = \mathbf{b} - \mathbf{A}\mathbf{x}$  provengano da una popolazione di errori, con una distribuzione di probabilità di tipo *normale*, e indicata con  $\mathbf{V}$  la *matrice di covarianza* (che supporremo definita positiva, e quindi invertibile<sup>10</sup>)

$$\mathbf{V} = E(\mathbf{e}\mathbf{e}^T)$$

la scelta

$$\mathbf{C} = \mathbf{W}^T \mathbf{W} = \mathbf{V}^{-1} \quad (11.44)$$

fornisce uno stimatore

$$\hat{\mathbf{x}} = \mathbf{L}\mathbf{b} = (\mathbf{A}^T \mathbf{V}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{V}^{-1} \mathbf{b} \quad (11.45)$$

*lineare, non distorto e ottimale*. Ricordiamo che non distorto significa che il valore atteso dell'errore  $\mathbf{x} - \hat{\mathbf{x}}$ , tra il valore incognito del parametro  $\mathbf{x}$  e il valore  $\hat{\mathbf{x}}$  stimato a partire dalle osservazioni  $\mathbf{b}$ , è nullo. In effetti, si ha

$$E(\mathbf{x} - \hat{\mathbf{x}}) = E(\mathbf{x} - \mathbf{L}\mathbf{b}) = E(\mathbf{x} - \mathbf{L}\mathbf{A}\mathbf{x} - \mathbf{L}\mathbf{e}) = E((\mathbf{I} - \mathbf{L}\mathbf{A})\mathbf{x}) = 0$$

Infine, *ottimale* significa che la scelta  $\mathbf{C} = \mathbf{V}^{-1}$  minimizza la varianza dell'errore

$$\mathbf{P} = E[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T] = (\mathbf{A}^T \mathbf{V}^{-1} \mathbf{A})^{-1}$$

<sup>9</sup>Ricordiamo che per quanto riguarda la *stabilità numerica* possono essere più convenienti opportune decomposizioni della matrice  $\mathbf{W}\mathbf{A}$  (cfr. Capitolo 2).

<sup>10</sup>In effetti, la matrice di covarianza è non invertibile quando almeno una delle varianze è *nulla*, ossia quando alcune delle misure  $b_i$  sono esatte; questo significa che le equazioni corrispondenti potrebbero essere risolte esattamente, anziché con il metodo dei minimi quadrati.

La matrice  $\mathbf{P}$  fa errori aspettati  $\mathbf{A}^T \mathbf{V}^{-1} \mathbf{A} = \mathbf{A}^T$  *matrice di informazione del della attendibilità*

Il caso in cui uno corrisponde informazione è  $\mathbf{A}$

### 11.3.2 Metodo

Supponiamo di sulla base di un misurazioni, con nuta. Utilizzando  $\mathbf{x}_0$  nella seguente

L'errore  $\mathbf{x} - \mathbf{x}_0$  è

Nel caso in cui una stima ottimale a partire da  $\mathbf{x}_0$  è

Nell'ipotesi di covarianza  $\mathbf{V}$  de

Pertanto la mat

$$\mathbf{P}_1^{-1} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix}$$

dalla quale si ha

che fornisce l'in Sottolineiamo che soltanto dalle lo

are, detto sistema delle

$$\mathbf{b} \quad (11.43)$$

te attendibilità. Si può  
W con la quale pesare  
considera, cioè,  $\|\mathbf{W}\mathbf{e}\|^2$   
normali si modifica nel

$$\mathbf{A}^T \mathbf{W}^T \mathbf{W} \mathbf{b}$$

invertibile, la soluzione

opportune condizioni  
vemente, supponendo  
errori, con una distri-  
matrice di covarianza

$$(11.44)$$

$$(11.45)$$

gnifica che il valore  
stro  $\mathbf{x}$  e il valore  $\hat{\mathbf{x}}$

$$\mathbf{A} \mathbf{x} = 0$$

ianza dell'errore

re più convenienti op-

delle varianze è nulla,  
azioni corrispondenti  
rati.

La matrice  $\mathbf{P}$  fornisce gli errori aspettati in  $\hat{\mathbf{x}}$ , mentre la matrice  $\mathbf{V}$  fornisce gli errori aspettati nelle osservazioni  $\mathbf{b}$ . L'inversa della matrice  $\mathbf{P}$ , ossia la matrice  $\mathbf{A}^T \mathbf{V}^{-1} \mathbf{A} = \mathbf{A}^T \mathbf{C} \mathbf{A}$  è un indicatore importante nella teoria del filtraggio ed è chiamata *matrice di informazione* (o matrice di Fisher), in quanto misura il contenuto di informazione dell'esperimento. In maniera schematica, essa aumenta all'aumentare della attendibilità di  $\mathbf{b}$ , ossia al diminuire di  $\mathbf{V}$ .

Il caso in cui gli errori sono indipendenti e distribuiti normalmente con varianza uno corrisponde al caso di *rumori bianchi*; si ha allora  $\mathbf{C} = \mathbf{I}$  e la matrice di informazione è  $\mathbf{A}^T \mathbf{A}$ .

### 11.3.2 Metodo dei minimi quadrati ricorsivo

Supponiamo di aver stimato il vettore  $\mathbf{x}$  mediante il metodo dei minimi quadrati sulla base di un primo insieme di misurazioni. Indicheremo con  $\mathbf{b}_0$  tale insieme di misurazioni, con  $\mathbf{V}_0$  la corrispondente matrice di covarianza e con  $\mathbf{x}_0$  la stima ottenuta. Utilizzando i risultati ottenuti nel paragrafo precedente possiamo esprimere  $\mathbf{x}_0$  nella seguente forma

$$\mathbf{x}_0 = (\mathbf{A}_0^T \mathbf{V}_0^{-1} \mathbf{A}_0)^{-1} \mathbf{A}_0^T \mathbf{V}_0^{-1} \mathbf{b}_0 \quad (11.46)$$

L'errore  $\mathbf{x} - \mathbf{x}_0$  ha media nulla e la sua matrice di covarianza è data da

$$\mathbf{P}_0 = \mathbf{E}((\mathbf{x} - \mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)^T) = (\mathbf{A}_0^T \mathbf{V}_0^{-1} \mathbf{A}_0)^{-1} \quad (11.47)$$

Nel caso in cui siano disponibili altri dati  $\mathbf{b}_1$ , poniamoci il problema di calcolare una stima ottimale della soluzione  $\mathbf{x}_1$  del sistema combinato  $\mathbf{A}_0 \mathbf{x} = \mathbf{b}_0$ ,  $\mathbf{A}_1 \mathbf{x} = \mathbf{b}_1$  a partire da  $\mathbf{x}_0$  e  $\mathbf{b}_1$  senza ripartire con il calcolo da  $\mathbf{b}_0$ .

Nell'ipotesi che gli errori  $\mathbf{e}_1$  siano indipendenti dagli errori  $\mathbf{e}_0$ , la matrice di covarianza  $\mathbf{V}$  degli errori totali  $[\mathbf{e}_0, \mathbf{e}_1]^T$  è di tipo diagonale a blocchi

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_0 & 0 \\ 0 & \mathbf{V}_1 \end{bmatrix}$$

Pertanto la matrice  $\mathbf{A}^T \mathbf{C} \mathbf{A}$  nell'equazione relativa al vettore  $\mathbf{x}_1$  è

$$\mathbf{P}_1^{-1} = \begin{bmatrix} \mathbf{A}_0 \\ \mathbf{A}_1 \end{bmatrix}^T \begin{bmatrix} \mathbf{V}_0 & 0 \\ 0 & \mathbf{V}_1 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{A}_0 \\ \mathbf{A}_1 \end{bmatrix} = \mathbf{A}_0^T \mathbf{V}_0^{-1} \mathbf{A}_0 + \mathbf{A}_1^T \mathbf{V}_1^{-1} \mathbf{A}_1 \quad (11.48)$$

dalla quale si ha la seguente importante relazione

$$\mathbf{P}_1^{-1} = \mathbf{P}_0^{-1} + \mathbf{A}_1^T \mathbf{V}_1^{-1} \mathbf{A}_1 \quad (11.49)$$

che fornisce l'incremento nell'informazione fornito dal secondo insieme di misure. Sottolineiamo che la relazione (11.49) non dipende dai valori attuali di  $\mathbf{b}_0$  o  $\mathbf{b}_1$ , ma soltanto dalle loro proprietà statistiche.

Osserviamo, ora, che dall'equazione normale  $\mathbf{A}^T \mathbf{V}^{-1} \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{V}^{-1} \mathbf{b}$ , si ha

$$\mathbf{x}_1 = \mathbf{P}_1 \begin{bmatrix} \mathbf{A}_0 \\ \mathbf{A}_1 \end{bmatrix}^T \mathbf{V}^{-1} \begin{bmatrix} \mathbf{b}_0 \\ \mathbf{b}_1 \end{bmatrix} = \mathbf{P}_1 (\mathbf{A}_0^T \mathbf{V}_0^{-1} \mathbf{b}_0 + \mathbf{A}_1^T \mathbf{V}_1^{-1} \mathbf{b}_1) \quad (11.50)$$

da cui

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{P}_1 (\mathbf{P}_0^{-1} \mathbf{x}_0 + \mathbf{A}_1^T \mathbf{V}_1^{-1} \mathbf{b}_1) \\ &= \mathbf{P}_1 (\mathbf{P}_1^{-1} \mathbf{x}_0 - \mathbf{A}_1^T \mathbf{V}_1^{-1} \mathbf{A}_1 \mathbf{x}_0 + \mathbf{A}_1^T \mathbf{V}_1^{-1} \mathbf{b}_1) \\ &= \mathbf{x}_0 + \mathbf{K}_1 (\mathbf{b}_1 - \mathbf{A}_1 \mathbf{x}_0) \end{aligned} \quad (11.51)$$

La matrice

$$\mathbf{K}_1 = \mathbf{P}_1 \mathbf{A}_1^T \mathbf{V}_1^{-1}$$

viene chiamata *matrice di guadagno* (gain matrix). In questa maniera si è ottenuta una *formula ricorsiva*; essa permette di calcolare  $\mathbf{x}_1$  a partire da  $\mathbf{x}_0$ , anziché da  $\mathbf{b}_0$ .

Naturalmente, se  $\mathbf{b}_1 = \mathbf{A}_1 \mathbf{x}_0$  la stima migliore è ancora  $\mathbf{x}_1 = \mathbf{x}_0$ ; ossia le nuove misure sono esattamente consistenti con il valore originale  $\mathbf{x}_0$ . In caso contrario, l'*errore di predizione*  $\mathbf{b}_1 - \mathbf{A}_1 \mathbf{x}_0$  è amplificato mediante la matrice di guadagno  $\mathbf{K}_1$  per fornire la correzione a  $\mathbf{x}_0$ . Il procedimento può essere iterato a partire da  $\mathbf{P}_1$  e  $\mathbf{x}_1$  per ottenere  $\mathbf{P}_2$  e  $\mathbf{x}_2$ . In maniera generale, a partire dalla coppia  $(\mathbf{P}_{i-1}, \mathbf{x}_{i-1})$  e dai valori osservati  $\mathbf{b}_i$  si ottiene  $(\mathbf{P}_i, \mathbf{x}_i)$  mediante le seguenti formule

$$\begin{aligned} \mathbf{P}_i^{-1} &= \mathbf{P}_{i-1}^{-1} + \mathbf{A}_i^T \mathbf{V}_i^{-1} \mathbf{A}_i \\ \mathbf{x}_i &= \mathbf{x}_{i-1} + \mathbf{K}_i (\mathbf{b}_i - \mathbf{A}_i \mathbf{x}_{i-1}) \quad \text{ove } \mathbf{K}_i = \mathbf{P}_i \mathbf{A}_i^T \mathbf{V}_i^{-1} \end{aligned}$$

► **Esempio 11.4** Come semplice esemplificazione, consideriamo una serie  $b_1, \dots, b_m$  di valutazioni sperimentali di una determinata grandezza incognita  $x$ . Supponiamo inoltre che le successive valutazioni siano indipendenti e distribuite normalmente con la stessa varianza  $\sigma^2$ . Si verifica facilmente che in questo caso si ha

$$\mathbf{A}^T = [1 \ 1 \ \dots \ 1], \quad \mathbf{P} = (\mathbf{A}^T \mathbf{V}^{-1} \mathbf{A})^{-1} = \frac{\sigma^2}{m}$$

e la soluzione  $\hat{x}$  è la media dei valori  $b_i$ . Supponendo di aggiungere successivamente una nuova stima, le formule iterative precedenti diventano

$$\mathbf{P}_i^{-1} = \mathbf{P}_{i-1}^{-1} + \frac{1}{\sigma^2}, \quad \mathbf{x}_i = \mathbf{x}_{i-1} + \mathbf{K}_i (\mathbf{b}_i - \mathbf{A}_i \mathbf{x}_{i-1})$$

ove  $\mathbf{A}_i = [1]$ . La prima relazione dice che  $\mathbf{P}_i^{-1}$  è uguale a  $i/\sigma^2$ , mentre dalla seconda relazione, poiché  $\mathbf{K}_i = \mathbf{P}_i \mathbf{A}_i^T \mathbf{V}_i^{-1} = [1/i]$ , si ha

$$\mathbf{x}_i = \mathbf{x}_{i-1} + \frac{1}{i} (\mathbf{b}_i - \mathbf{x}_{i-1}) = \frac{1}{i} \mathbf{b}_i + \frac{i-1}{i} \left( \frac{\mathbf{b}_1 + \mathbf{b}_2 + \dots + \mathbf{b}_{i-1}}{i-1} \right) = \frac{\mathbf{b}_1 + \mathbf{b}_2 + \dots + \mathbf{b}_i}{i}$$

Naturalmente, l'idea ha la sua efficacia nel caso più generale in cui  $\mathbf{b}$  e  $\mathbf{x}$  sono vettori e  $\mathbf{V}$  e  $\mathbf{P}$  sono matrici. ■

### 11.3.3 Filtro

Nello schema cc riferiscono alla grandezza che  $\mathbf{b}_i$  si riferiscono nel caso dei seg una grandezza mantenere om con  $\mathbf{x}_i$  (anziché, al generico passo precedente  $\mathbf{x}_i$  d

ove  $\mathbf{F}_i$  è una m acquisisce nuov lineare di  $\mathbf{x}_i$ , ci misurazione, si

Il problema cc separare il segr è rappresentato

Utilizzando il mediante il m sistema linear

ove  $a_i, b_i$  non sono necessariamente numeri finiti. Il convesso  $K$  è allora detto di tipo *locale*. L'estensione dell'algoritmo precedente è ovvia. Si definisce  $x_i^{k+1}$  come la soluzione in  $K_i$  del seguente problema di minimo unidimensionale

$$f(x_1^{k+1}, \dots, x_{i-1}^{k+1}, x_i^{k+1}, x_{i+1}^k, \dots, x_n^k) \leq f(x_1^{k+1}, \dots, x_{i-1}^{k+1}, z, x_{i+1}^k, \dots, x_n^k), \quad \forall z \in K_i$$

per  $i = 1, \dots, n$ . Per tale algoritmo si estende il risultato di *convergenza* del Teorema 5.12.

▼ **Osservazione 5.4** Se il convesso  $K$  non è della forma (5.74), allora il metodo può non convergere per determinati punti di partenza; si consideri come controesempio il calcolo del minimo della funzione  $f(x) = x_1^2 + x_2^2$  sull'insieme  $K$  definito da

$$K = \{x \mid x \in \mathbb{R}^2, x_1 \geq 0, x_2 \geq 0, x_1 + x_2 \geq 1\}$$

che ammette come minimo il punto  $x^* = [1/2, 1/2]^T$ . Partendo da  $x^0 = [0, 1]$ , l'algoritmo si blocca in tale punto. ■

Terminiamo con un'estensione al caso vincolato del metodo SOR che abbiamo visto nel Capitolo 2 per la risoluzione dei sistemi lineari. Supponiamo che  $f$  sia della forma

$$f(x) = \frac{1}{2}(Ax, x) - (b, x)$$

ove  $A$  è una matrice simmetrica definita positiva, e  $K$  di tipo *locale*

$$K = \prod_{i=1}^n K_i$$

L'algoritmo risulta definito dalle formule seguenti, ove  $i = 1, 2, \dots, n$

$$\begin{cases} x_i^{k+1/2} = -\frac{1}{a_{ii}} \left( \sum_{j=1}^{i-1} a_{ij} x_j^{k+1} + \sum_{j=i+1}^n a_{ij} x_j^k - b_i \right) \\ x_i^{k+1} = P_{K_i} \left( (1 - \omega)x_i^k + \omega x_i^{k+1/2} \right) \end{cases}$$

Nelle ipotesi precedenti, si può dimostrare come per il caso non vincolato, che il metodo converge per  $0 < \omega < 2$ .

#### 5.5.4 Minimi quadrati non lineari

Il problema può avere *origine* in diversi contesti applicativi: *identificazione di parametri, curve-fitting*, ecc. Per esempi nell'ambito della identificazione dei parametri si vedano i Capitoli 12 e 13. Il problema può essere formulato, in forma generale, nel seguente modo. È dato un *modello matematico*

$$y = F(t, x) \tag{5.75}$$

alla funzione di

T. Applicando  
uente funzione

azione rispetto  
a punto in cui

necessaria per  
ioni non dif-  
e il seguente

ed è di classe  
problema di

problema di  
odo seguente

(5.74)

cioè una dipendenza funzionale di  $\mathbf{y}$  da  $\mathbf{t}$  e  $\mathbf{x}$ . Tale funzione può essere data esplicitamente, oppure più in generale può essere la soluzione, ad esempio, di un problema differenziale.

Il vettore  $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$  è il *vettore dei parametri*. Il vettore  $\mathbf{t} = [t_1, t_2, \dots, t_k]^T$  rappresenta le *variabili indipendenti*, mentre  $\mathbf{y} = [y_1, y_2, \dots, y_r]^T$  è il vettore delle *variabili dipendenti*, o *osservate* (per il seguito supporremo  $r = 1$ ). Supponendo di conoscere, in corrispondenza ai valori  $t^1, t^2, \dots, t^m$  della variabile indipendente, i valori *sperimentali*  $y^*$  di  $y$ , si definiscono per  $i = 1, 2, \dots, m$  i *residui*

$$f_i(\mathbf{x}) = F(\mathbf{t}^i, \mathbf{x}) - y^*(\mathbf{t}^i) \quad (5.76)$$

e si pone  $\mathbf{f}(\mathbf{x}) = [f_1, f_2, \dots, f_m]^T$ . Il problema è, allora, quello di determinare il vettore  $\mathbf{x}$  in modo che i residui siano *minimi*. Per precisare il problema, è necessario introdurre un particolare *stimatore*, cioè una particolare *distanza* in  $\mathbb{R}^m$ . Quando si sceglie come distanza la norma euclidea<sup>14</sup>  $\|\cdot\|_2$ , il problema diventa il seguente, che viene detto *problema dei minimi quadrati*

$$\min_{\mathbf{x} \in \mathbb{R}^n} S(\mathbf{x}) = \min_{\mathbf{x} \in \mathbb{R}^n} \mathbf{f}(\mathbf{x})^T \mathbf{f}(\mathbf{x}) \quad (5.77)$$

Naturalmente, nelle applicazioni possono esistere ulteriori vincoli per i parametri  $\mathbf{x}$ . In particolare, se i parametri  $\mathbf{x}$  del modello hanno un significato fisico, si ha usualmente  $\mathbf{x} \geq 0$ . Per il seguito, tuttavia considereremo, per semplicità, solo il caso di parametri *non vincolati*.

Supporremo, inoltre, che il modello matematico  $F$  sia regolare, in particolare che sia possibile calcolare le *derivate*

$$J_{ij} = \frac{\partial f_i}{\partial x_j}, \quad i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n$$

La matrice  $\mathbf{J}(\mathbf{x}) = [J_{ij}]$  di ordine  $m \times n$  è la *matrice Jacobiana*.

▼ **Osservazione 5.5** Quando  $m = n$ , per ogni soluzione del sistema

$$\mathbf{f}(\mathbf{x}) = 0$$

si ha  $S(\mathbf{x}) = 0$ . Nel quadro precedente, quindi, rientra come caso particolare il problema della risoluzione di un sistema non lineare. ■

▼ **Osservazione 5.6** Ovviamente, tutti i metodi di ottimizzazione visti nelle sezioni precedenti (*gradiente, gradiente-coniugato, quasi-Newton,...*) si applicano anche al problema dei minimi quadrati. C'è, comunque, l'interesse a vedere se, data la forma particolare di tale problema, si possano costruire metodi ad hoc più efficienti. ■

<sup>14</sup>Più in generale, si potrebbe prendere una norma euclidea *pesata*, cioè del tipo  $\|\cdot\|_A$ , con  $A$  matrice definita positiva, corrispondente, ad esempio, alla *matrice di covarianza* dei dati sperimentali. Tale scelta ha proprietà statistiche interessanti; in effetti (cfr. Capitolo 8), se i dati  $y^*$  sono affetti da errori distribuiti con legge gaussiana, essa fornisce la minima varianza per  $\mathbf{x}$ .

**Gradiente e h**

Con un semplic

ove  $\nabla^2 f_i(\mathbf{x})$  soi

► **Esempio 5.**  
1, ..., 4 mediante  
 $\mathbf{J}(\mathbf{x}) \in \mathbb{R}^{4 \times 2}$

Inoltre,  $\nabla S(\mathbf{x}) \in$

e  $\nabla^2 S(\mathbf{x}) \in \mathbb{R}^{2 \times 2}$

$\nabla^2 S(\mathbf{x}) :$

Una condizion

Se allora per r

In questa forr  
del termine  $\sum$   
termine è, "pi  
"buono", ossi:  
"quasi lineari"  
ragionevole la

$$J^T(\lambda) = \begin{bmatrix} t^1 e^{t^1 x_1} & t^2 e^{t^2 x_1} & t^3 e^{t^3 x_1} & t^4 e^{t^4 x_1} \\ t^1 e^{t^1 x_2} & t^2 e^{t^2 x_2} & t^3 e^{t^3 x_2} & t^4 e^{t^4 x_2} \end{bmatrix}$$

$$J^T J = \begin{bmatrix} \sum_{i=1}^4 (t^i e^{t^i x_1})^2 & \sum_{i=1}^4 (t^i)^2 e^{t^i(x_1+x_2)} \\ \sum_{i=1}^4 (t^i)^2 e^{t^i(x_1+x_2)} & \sum_{i=1}^4 (t^i)^2 e^{t^i(x_1+x_2)} \end{bmatrix}$$

sere data esplicita, di un problema

re  $t = [t_1, t_2, \dots, t_r]^T$  è il vettore (con  $r = 1$ ). Supponiamo che  $t$  sia una variabile indipendente e i residui

$$(5.76)$$

determinare il minimo, è necessario  $\mathbb{R}^m$ . Quando si ha il seguente, che

$$(5.77)$$

er i parametri fisici, si ha semplicità, solo il

in particolare

**Gradiente e Hessiana di  $S(x)$**

Con un semplice calcolo si ottiene

$$\nabla S(x) = 2J(x)^T f(x)$$

$$\nabla^2 S(x) = 2J(x)^T J(x) + 2 \sum_{i=1}^m \nabla^2 f_i(x) f_i(x)$$

ove  $\nabla^2 f_i(x)$  sono le matrici hessiane delle funzioni  $f_i(x)$ .

► **Esempio 5.25** Come illustrazione, supponiamo di voler fittare i dati  $(t^i, y(t^i))$ ,  $i = 1, \dots, 4$  mediante il modello  $F(t, x) = \exp(tx_1) + \exp(tx_2)$ , con  $t \in \mathbb{R}$ . In questo caso si ha  $J(x) \in \mathbb{R}^{4 \times 2}$

$$J(x) = \begin{bmatrix} t^1 e^{t^1 x_1} & t^1 e^{t^1 x_2} \\ t^2 e^{t^2 x_1} & t^2 e^{t^2 x_2} \\ t^3 e^{t^3 x_1} & t^3 e^{t^3 x_2} \\ t^4 e^{t^4 x_1} & t^4 e^{t^4 x_2} \end{bmatrix}$$

$$f_i(x) = F(t^i, x) - y^*(t^i), i=1-4$$

$$= e^{t^i x_1} + e^{t^i x_2} - y^*(t^i)$$

$$\frac{\partial f_i}{\partial x_1} = t^i e^{t^i x_1}, \quad \frac{\partial f_i}{\partial x_2} = t^i e^{t^i x_2}$$

Inoltre,  $\nabla S(x) \in \mathbb{R}^2$

$$\nabla S(x) = 2J(x)^T f(x) = 2 \left[ \sum_{i=1}^4 f_i(x) t^i e^{t^i x_1}, \sum_{i=1}^4 f_i(x) t^i e^{t^i x_2} \right]^T$$

e  $\nabla^2 S(x) \in \mathbb{R}^{2 \times 2}$

$$\nabla^2 S(x) = 2 \begin{bmatrix} \sum_{i=1}^4 (t^i)^2 e^{t^i x_1} (f_i(x) + e^{t^i x_1}) & \sum_{i=1}^4 (t^i)^2 e^{t^i(x_1+x_2)} \\ \sum_{i=1}^4 (t^i)^2 e^{t^i(x_1+x_2)} & \sum_{i=1}^4 (t^i)^2 e^{t^i x_2} (f_i(x) + e^{t^i x_2}) \end{bmatrix}$$

Una condizione necessaria in un punto di minimo per (5.77) è la seguente

$$\nabla S(x) = 0 \Rightarrow J(x)^T f(x) = 0 \tag{5.78}$$

Se allora per risolvere (5.78) utilizziamo il metodo di Newton, si ottiene

$$\begin{cases} x^1 & \text{arbitrario} \\ x^{k+1} = x^k + \lambda_k s^k & \nabla^2 S(x^k) s^k = -\nabla S(x^k) \end{cases}$$

In questa forma, tuttavia, il metodo risulta costoso, in quanto richiede il calcolo del termine  $\sum_{i=1}^m \nabla^2 f_i(x) f_i(x)$ . Osserviamo, d'altra parte, che il contributo di tale termine è, "piccolo" vicino alla soluzione, in due situazioni: quando il modello è "buono", ossia quando i residui  $f_i(x)$  sono piccoli, e quando le funzioni  $f_i$  sono "quasi lineari", ossia i termini  $\nabla^2 f_i(x)$  sono piccoli. In tali casi risulta, pertanto, ragionevole la seguente approssimazione

re il problema

lle sezioni pre-  
ve al problema  
particolare di

$\|\cdot\|_A$ , con A  
dati sperimen-

$$J^T(\underline{x}) = \begin{bmatrix} t^1 e^{t^1 x_1} & t^1 e^{t^1 x_2} \\ t^2 e^{t^2 x_1} & t^2 e^{t^2 x_2} \\ t^3 e^{t^3 x_1} & t^3 e^{t^3 x_2} \\ t^4 e^{t^4 x_1} & t^4 e^{t^4 x_2} \end{bmatrix} \quad J^T J = \begin{bmatrix} \sum_{i=1}^4 (t^i e^{t^i x_1})^2 & \sum_{i=1}^4 (t^i)^2 e^{t^i(x_1+x_2)} \\ \sum_{i=1}^4 (t^i)^2 e^{t^i(x_1+x_2)} & \sum_{i=1}^4 (t^i)^2 e^{t^i(x_1+x_2)} \end{bmatrix}$$

sere data esplici-  
, di un problema

re  $t = [t_1, t_2, \dots, y_r]^T$  è il vettore  $r = 1$ ). Supponiamo che  $t$  sia una variabile indipendente e i residui

$$(5.76)$$

determinare il  
na, è necessario  $\mathbb{R}^m$ . Quando si  
il seguente, che

$$(5.77)$$

er i parametri  
to fisico, si ha  
PLICITÀ, solo il

in particolare

**Gradiente e Hessiana di  $S(x)$**

Con un semplice calcolo si ottiene

$$\nabla S(x) = 2J(x)^T f(x)$$

$$\nabla^2 S(x) = 2J(x)^T J(x) + 2 \sum_{i=1}^m \nabla^2 f_i(x) f_i(x)$$

ove  $\nabla^2 f_i(x)$  sono le matrici hessiane delle funzioni  $f_i(x)$ .

► **Esempio 5.25** Come illustrazione, supponiamo di voler fittare i dati  $(t^i, y(t^i))$ ,  $i = 1, \dots, 4$  mediante il modello  $F(t, x) = \exp(tx_1) + \exp(tx_2)$ , con  $t \in \mathbb{R}$ . In questo caso si ha  $J(x) \in \mathbb{R}^{4 \times 2}$

$$J(x) = \begin{bmatrix} t^1 e^{t^1 x_1} & t^1 e^{t^1 x_2} \\ t^2 e^{t^2 x_1} & t^2 e^{t^2 x_2} \\ t^3 e^{t^3 x_1} & t^3 e^{t^3 x_2} \\ t^4 e^{t^4 x_1} & t^4 e^{t^4 x_2} \end{bmatrix}$$

$$f_i(x) = F(t^i, x) - y^*(t^i), \quad i=1, \dots, 4$$

$$= e^{t^i x_1} + e^{t^i x_2} - y^*(t^i)$$

$$\frac{\partial f_i}{\partial x_1} = t^i e^{t^i x_1}, \quad \frac{\partial f_i}{\partial x_2} = t^i e^{t^i x_2}$$

Inoltre,  $\nabla S(x) \in \mathbb{R}^2$

$$\nabla S(x) = 2J(x)^T f(x) = 2 \left[ \sum_{i=1}^4 f_i(x) t^i e^{t^i x_1}, \sum_{i=1}^4 f_i(x) t^i e^{t^i x_2} \right]^T$$

e  $\nabla^2 S(x) \in \mathbb{R}^{2 \times 2}$

$$\nabla^2 S(x) = 2 \begin{bmatrix} \sum_{i=1}^4 (t^i)^2 e^{t^i x_1} (f_i(x) + e^{t^i x_1}) & \sum_{i=1}^4 (t^i)^2 e^{t^i(x_1+x_2)} \\ \sum_{i=1}^4 (t^i)^2 e^{t^i(x_1+x_2)} & \sum_{i=1}^4 (t^i)^2 e^{t^i x_2} (f_i(x) + e^{t^i x_2}) \end{bmatrix}$$

Una condizione necessaria in un punto di minimo per (5.77) è la seguente

$$\nabla S(x) = 0 \Rightarrow J(x)^T f(x) = 0 \tag{5.78}$$

Se allora per risolvere (5.78) utilizziamo il metodo di Newton, si ottiene

$$\begin{cases} x^1 & \text{arbitrario} \\ x^{k+1} = x^k + \lambda_k s^k & \boxed{\nabla^2 S(x^k) s^k = -\nabla S(x^k)} \end{cases}$$

In questa forma, tuttavia, il metodo risulta costoso, in quanto richiede il calcolo del termine  $\sum_{i=1}^m \nabla^2 f_i(x) f_i(x)$ . Osserviamo, d'altra parte, che il contributo di tale termine è, "piccolo" vicino alla soluzione, in due situazioni: quando il modello è "buono", ossia quando i residui  $f_i(x)$  sono piccoli, e quando le funzioni  $f_i$  sono "quasi lineari", ossia i termini  $\nabla^2 f_i(x)$  sono piccoli. In tali casi risulta, pertanto, ragionevole la seguente approssimazione

$$\nabla^2 S(x) \approx 2J(x)^T J(x)$$

ire il problema

lle sezioni pre-  
ie al problema  
particolare di

$\| \cdot \|_A$ , con  $A$   
dati sperimentali  
i dati  $y^*$  sono  
 $x$ .

$$f_i(\underline{x}) = F(t^i, \underline{x}) - y^*(t^i), \quad i = 1, \dots, m$$

$$\nabla f_i(\underline{x}) = \begin{bmatrix} \frac{\partial f_i}{\partial x_1} \\ \vdots \\ \frac{\partial f_i}{\partial x_m} \end{bmatrix}$$

$$\nabla^2 f_i(\underline{x}) = \begin{bmatrix} \frac{\partial^2 f_i}{\partial x_1^2} & \frac{\partial^2 f_i}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f_i}{\partial x_1 \partial x_m} \\ \frac{\partial^2 f_i}{\partial x_m \partial x_1} & \frac{\partial^2 f_i}{\partial x_m \partial x_2} & \dots & \frac{\partial^2 f_i}{\partial x_m \partial x_m} \end{bmatrix}_{m \times m}$$

Assumendo tale approssimazione si ottiene il cosiddetto *metodo di Gauss-Newton*. Quindi, nel metodo di Gauss-Newton la direzione  $\mathbf{s}^k$  è calcolata come soluzione del seguente sistema lineare delle *equazioni normali*

$$\mathbf{J}(\mathbf{x}^k)^T \mathbf{J}(\mathbf{x}^k) \mathbf{s}^k = -\mathbf{J}(\mathbf{x}^k)^T \mathbf{f}(\mathbf{x}^k)$$

Naturalmente, per risolvere il sistema delle equazioni normali sono opportune le "precauzioni numeriche" indicate nel caso lineare. Ricordiamo, in particolare le tecniche basate sulle decomposizioni **QR**, **SVD**.

Nel caso generale, ossia quando il termine  $\sum_{i=1}^m \nabla^2 f_i(\mathbf{x}) f_i(\mathbf{x})$  non è trascurabile, esistono metodi che introducono opportune simulazioni di tale termine. Ad esempio, nel *metodo ibrido di Powell* si ha l'iterazione

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \lambda_k \mathbf{p}^k$$

ove

$$\mathbf{p}^k = \beta_k \mathbf{s}^k - \gamma_k \nabla S_k$$

Si utilizza, cioè, una direzione che è una combinazione opportuna della direzione di Gauss-Newton e quella del gradiente (cfr. Figura 5.37).

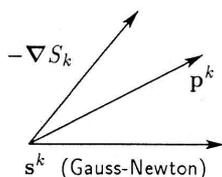


Figura 5.37. Metodo ibrido di Powell.

Nel *metodo di Levenberg-Marquardt*<sup>15</sup>, la direzione di ricerca  $\mathbf{s}^k$  è ottenuta risolvendo il seguente sistema lineare

$$(\mathbf{J}(\mathbf{x}^k)^T \mathbf{J}(\mathbf{x}^k) + \delta_k \mathbf{D}_k) \mathbf{s}^k = -\mathbf{J}(\mathbf{x}^k)^T \mathbf{f}(\mathbf{x}^k)$$

ove  $\mathbf{D}_k$  è una opportuna matrice definita positiva e  $\delta_k$  è un parametro opportuno. Osserviamo che se  $\delta_k$  è sufficientemente grande la matrice  $\mathbf{J}(\mathbf{x}^k)^T \mathbf{J}(\mathbf{x}^k) + \delta_k \mathbf{D}_k$  è definita positiva e  $\mathbf{s}^k$  è una direzione di discesa, mentre per  $\delta_k = 0$  si riottiene il metodo di Gauss-Newton. Si ha ancora, quindi, come direzione  $\mathbf{s}^k$  una *combinazione tra Gauss-Newton e gradiente*. Nei due metodi la parte, numericamente più "delicata", riguarda, ovviamente, la *scelta dei vari parametri*. Nelle implementazioni più

<sup>15</sup>Tale metodo è stato proposto da Levenberg (1944) e da Marquardt (1963).

utilizzate la sc  
sulla base dei 1

◆ **Esercizio 5**  
lungo la retta  $x$   
livello  $f(\mathbf{x}) = f$

◆ **Esercizio 5**  
punti di minimo

◆ **Esercizio 5**

ove  $x > 0, b > 0$

◆ **Esercizio 5**  
funzione

con  $\mathbf{x}^1 = [0, 0]$ .

◆ **Esercizio 5**  
retta  $3x_1 + 2x_2$

◆ **Esercizio 5**

con  $f_1(\mathbf{x}) = x_2$   
Gauss-Newton

Gauss-Newton.  
e soluzione del

o opportune le  
i particolare le

è trascurabile,  
e. Ad esempio,

lla direzione di

o ottenuta risol-

tro opportuno.  
 $J(\mathbf{x}^k) + \delta_k \mathbf{D}_k$  è  
0 si riottiene il  
a combinazione  
nente più "deli-  
ementazioni più

utilizzate la scelta dei parametri è di tipo *adattivo*; essi vengono, cioè, aggiornati sulla base dei risultati di minimizzazione successivamente ottenuti.

◆ **Esercizio 5.37** Mostrare che se  $\bar{\mathbf{x}}$  è un punto di minimo di una funzione  $f(\mathbf{x})$ ,  $\mathbf{x} \in \mathbb{R}^n$ , lungo la retta  $\mathbf{x} = \bar{\mathbf{x}} + \lambda \mathbf{d}$ , con  $\mathbf{d} \in \mathbb{R}^n$ , allora tale retta è tangente in  $\bar{\mathbf{x}}$  alla superficie di livello  $f(\mathbf{x}) = f(\bar{\mathbf{x}})$ .

◆ **Esercizio 5.38** Fornire un esempio di una funzione strettamente convessa che non ha punti di minimo.

◆ **Esercizio 5.39** Applicare il metodo di Newton per minimizzare le seguenti funzioni

$$(a) (n-1)x + bx^{1-n}, \quad (b) (n-2)x^2 + 2bx^{2-n} \quad (n > 2)$$

$$(c) x^{n+1} - (n+1)bx, \quad (d) n \lg n + bx^{-n},$$

$$(e) (n-3)x^{(n+3)/2} + (n+3)x^{-(n-3)/2} \quad (n \neq 3)$$

ove  $x > 0$ ,  $b > 0$ , e  $n > 1$ .

◆ **Esercizio 5.40** Applicare il metodo di Newton per la minimizzazione della seguente funzione

$$f(x_1, x_2) = x_1^4 + 6x_1x_2 + 1.5x_2^2 + 36x_2 + 405$$

con  $\mathbf{x}^1 = [0, 0]$ .

◆ **Esercizio 5.41** Determinare il punto di minimo di  $f(x_1, x_2) = \frac{1}{2}(3x_1^2 + 4x_2^2)$  lungo la retta  $3x_1 + 2x_2 = 4$ . Mostrare che i vettori  $\mathbf{p} = [1, 1]^T$  e  $\mathbf{q} = [4, -3]^T$  sono coniugati.

◆ **Esercizio 5.42** Sia  $S: \mathbb{R}^2 \rightarrow \mathbb{R}$  definita da

$$S(\mathbf{x}) = \mathbf{f}(\mathbf{x})^T \mathbf{f}(\mathbf{x})$$

con  $f_1(x) = x_2 - x_1^2$  e  $f_2(x) = 1 - x_1$ . Analizzare il metodo di Newton e il metodo di Gauss-Newton per la ricerca del punto di minimo  $[1, 1]^T$  di  $S$ .

$$F(t, \underline{x}) \quad \underline{x} = [x_1 \dots x_n]$$

$$(t^i, y^*(t^i)) \quad i=1 \dots m \quad \text{misure}$$

$$f_i(x) = F(t^i, x) - y^*(t^i)$$

$$\min_{\underline{x}} \sum_{i=1}^m [F(t^i, x) - y^*(t^i)]^2 =$$

$$\min_{\underline{x}} \sum_{i=1}^m [f_i(x)]^2 = \min_{\underline{x}} f^T(x) \cdot f(x)$$

dove  $f(x) = [f_1(x), \dots, f_m(x)]^T$

Posto:  $S(x) = f^T(x) \cdot f(x)$  allora:

$$\min_{\underline{x}} S(x)$$

$$\exists \quad J_{ij} = \frac{\partial f_i}{\partial x_j} \quad i=1 \dots m, \quad j=1 \dots n$$

$$1) \quad \nabla S(x) = 2 J^T(x) f(x)$$

$$2) \quad \nabla^2 S(x) = 2 J^T(x) \cdot J(x) + 2 \sum_{i=1}^m \nabla^2 f_i(x) f_i(x)$$

Punti stazionari  $\Rightarrow$

$$\nabla S(x) = 0 \Leftrightarrow J^T(x) f(x) = 0$$

Metodo di Newton:

$$\underline{x}^{k+1} = \underline{x}^k + \lambda_k \underline{d}^k$$

con  $\underline{d}^k = - \nabla^2 S(\underline{x}^k) \cdot \nabla S(\underline{x}^k)$  (metodo esatto)

Si approssima:

$$\nabla^2 S(\underline{x}) \approx 2 \mathbf{J}^T(\underline{x}) \cdot \mathbf{J}(\underline{x})$$

metodo di Gauss-Newton.

$$\left[ \mathbf{J}^T(\underline{x}^k) \mathbf{J}(\underline{x}^k) \right] \underline{d}^k = - \mathbf{J}^T(\underline{x}^k) f(\underline{x}^k), \quad k=0,1,2,\dots$$

Esempio

$$(t^i, y^*(t^i)) \quad i=1,2,3,4, \quad \boxed{m=4}$$

$$F(t, x) = e^{tx_1} + e^{tx_2}, \quad t \in \mathbb{R}; \quad x = [x_1, x_2] \quad \boxed{m=2}$$

Residui:

$$f_i(x_1, x_2) = F(t^i, x) - y^*(t^i) = e^{t^i x_1} + e^{t^i x_2} - y^*(t^i), \quad i=1,2,3,4$$

$$Jf(x) = \left( \frac{\partial f_i}{\partial x_j} \right) = \begin{bmatrix} t^1 e^{t^1 x_1} & t^1 e^{t^1 x_2} \\ t^2 e^{t^2 x_1} & t^2 e^{t^2 x_2} \\ t^3 e^{t^3 x_1} & t^3 e^{t^3 x_2} \\ t^4 e^{t^4 x_1} & t^4 e^{t^4 x_2} \end{bmatrix}_{4 \times 2}; \quad J^T = \begin{bmatrix} t^1 e^{t^1 x_1} & t^1 e^{t^1 x_2} \\ t^2 e^{t^2 x_1} & t^2 e^{t^2 x_2} \\ t^3 e^{t^3 x_1} & t^3 e^{t^3 x_2} \\ t^4 e^{t^4 x_1} & t^4 e^{t^4 x_2} \end{bmatrix}$$

$$\nabla S(x) = 2 J^T(x) f(x) = 2 \begin{bmatrix} \sum_{i=1}^4 t^i e^{t^i x_1} f_i(x) \\ \sum_{i=1}^4 t^i e^{t^i x_2} f_i(x) \end{bmatrix}$$

$$\nabla^2 S(x) = 2 \begin{bmatrix} \sum_i (t^i)^2 (e^{t^i x_1})^2 & \sum_i (t^i)^2 e^{t^i(x_1+x_2)} \\ \sum_i (t^i)^2 e^{t^i(x_1+x_2)} & \sum_i (t^i)^2 (e^{t^i x_2})^2 \end{bmatrix} + 2 \sum_{i=1}^4 \nabla^2 f_i(x) \cdot f_i(x)$$

$$\nabla f_i(x) = \begin{bmatrix} t^i e^{t^i x_1} \\ t^i e^{t^i x_2} \end{bmatrix} \quad \nabla^2 f_i(x) = \begin{bmatrix} (t^i)^2 e^{t^i x_1} & 0 \\ 0 & (t^i)^2 e^{t^i x_2} \end{bmatrix}$$

$$\sum_{i=1}^m \nabla^2 f_i(x) \cdot f_i(x) = \left[ \begin{array}{c|c} \sum_{i=1}^4 (t^i)^2 e^{t^i x_1} \cdot f_i(x) & 0 \\ \hline 0 & \sum_{i=1}^4 (t^i)^2 e^{t^i x_2} \cdot f_i(x) \end{array} \right]$$

metodo di Gauss-Newton:

$$\cancel{J^T} [J^T(x^k) J(x^k)] \underline{d}^k = -J^T(x^k) f(x^k), \quad k=0,1,$$

cioè  $x^k = (x_1^k, x_2^k)$ :

$$\begin{bmatrix} \sum (t^i e^{t^i x_1^k})^2 & \sum (t^i)^2 e^{t^i(x_1^k + x_2^k)} \\ \sum (t^i)^2 e^{t^i(x_1^k + x_2^k)} & \sum (t^i e^{t^i x_2^k})^2 \end{bmatrix} \underline{d}^k = - \begin{bmatrix} \sum_{i=1}^4 t^i e^{t^i x_1^k} f_i(x_1^k, x_2^k) \\ \sum_{i=1}^4 t^i e^{t^i x_2^k} f_i(x_1^k, x_2^k) \end{bmatrix}$$

dove  $f_i(x_1^k, x_2^k) = e^{t^i x_1^k} + e^{t^i x_2^k} - y_i^*(t^i)$ ;  $i=1,2,3,4$

$$\begin{bmatrix} 1 & \bar{x} \\ \bar{x} & \sigma_x^2 + \bar{x}^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \bar{y} \\ C_{xy} + \bar{x} \cdot \bar{y} \end{bmatrix}$$

$$a_1 = \frac{\begin{vmatrix} 1 & \bar{y} \\ \bar{x} & C_{xy} + \bar{x}\bar{y} \end{vmatrix}}{\sigma_x^2} = \frac{C_{xy} + \cancel{\bar{x} \cdot \bar{y}} - \bar{x} \bar{y}}{\sigma_x^2} = \frac{C_{xy}}{\sigma_x^2}$$

$$a_0 = \frac{\begin{vmatrix} \bar{y} & \bar{x} \\ C_{xy} + \bar{x}\bar{y} & \sigma_x^2 + \bar{x}^2 \end{vmatrix}}{\sigma_x^2} = \frac{\sigma_x^2 \cdot \bar{y} + \cancel{\bar{x}^2 \cdot \bar{y}} - \bar{x} C_{xy} - \cancel{\bar{x}^2 \cdot \bar{y}}}{\sigma_x^2}$$

$$= \bar{y} - \bar{x} \frac{C_{xy}}{\sigma_x^2} = \bar{y} - a_1 \bar{x}$$

$$\boxed{\bar{y} = a_1 \bar{x} + a_0}$$

Il punto medio sta sulle rette di regressione lineare

