

FAMIGLIE DI MODELLI POLINOMIALI

- Nella determinazione di un modello di una grandezza incognita (variabile dipendente y) in funzione di altre grandezze (variabili indipendenti u_i), una soluzione semplice è quella di ipotizzare una dipendenza di y da u_i di tipo polinomiale
- Questo approccio si è rivelato molto utile nella stima del rendimento volumetrico e nella determinazione di un modello della pressione nel collettore
- Come scegliere tra tutti i possibili modelli (problema della scelta della struttura del modello)? Secondo le tecniche e gli indici precedenti o mediante il metodo della stepwise regression

Nei problemi di stima di un modello che spieghi il legame tra q variabili indipendenti u_1, u_2, \dots, u_q , e la variabile dipendente y , è naturale inizialmente cercare un modello di regressione del tipo

$$y = \theta_1 u_1 + \theta_2 u_2 + \dots + \theta_q u_q$$

lineare non solo nei parametri θ_i da stimarsi, ma anche nelle variabili u_i . Tuttavia questa ipotesi è in molti casi eccessivamente limitativa (si pensi al problema di stima di un modello per il rendimento volumetrico).

Può essere conveniente rilassare l'ipotesi di linearità nelle u_i e considerare modelli di tipo polinomiale nelle variabili indipendenti.

Un esempio

Supponiamo $q=2$ e ipotizziamo che la dipendenza di y da u_1 e u_2 possa essere di tipo lineare e/o quadratico rispetto a ogni variabile:

$$y = \theta_0 + \theta_1 u_1 + \theta_2 u_2 + \theta_3 u_1^2 + \theta_4 u_2^2 + \theta_5 u_1 u_2 + \theta_6 u_1^2 u_2 + \theta_7 u_1 u_2^2 + \theta_8 u_1^2 u_2^2$$

□

In questo modo, se q è il numero delle variabili indipendenti e m è la potenza di ordine massimo che può assumere ogni variabile, il numero di possibili regressori è $(m+1)^q$. Naturalmente il modello completo sarà in generale sovradimensionato e il numero di possibili regressori da considerare tenderà a crescere in modo esplosivo all'aumentare di m e q . Inoltre la dipendenza di y da alcuni regressori potrà essere scarsa o nulla. Il problema è quindi quello di ridurre la complessità del modello eliminando i regressori superflui. E' ancora un problema di *scelta della struttura del modello* per il quale viene nel seguito fornito un altro procedimento possibile: la *stepwise regression*.

Si osservi comunque che l'aver considerato potenze intere positive o nulle è comunque arbitrario e limitativo, così come l'essersi limitati a modelli di tipo polinomiale.

STIMA DEL RENDIMENTO VOLUMETRICO η

- Il rendimento volumetrico non è costante, ma dipende da varie grandezze, tra cui primariamente il numero di giri n , la pressione nel collettore p , la temperatura nel collettore T . Anche altre variabili, ad esempio l'angolo di apertura della valvola a farfalla, l'angolo di anticipo accensione, il titolo possono influenzare il valore di η .
- Per semplicità si è supposta una dipendenza di tipo polinomiale da n, p, T :

$$\eta = f(n, p, T)$$

e lineare nei parametri incogniti.

- Tra tutti i modelli identificati, i più interessanti sono riassunti nella seguente tabella (EQM = errore quadratico medio = $\sqrt{SSR/N}$)

Struttura del modello	EQM	Massima deviazione standard
$a_0 + a_1 n + a_2 n^2 + a_3 n^3 + a_4 p_{man}$	0.0110	13.9%
$a_0 + a_1 n + a_2 n^2 + a_3 n^3 + a_4 n^4 + a_5 p_{man}$	0.0083	26.9%
$a_0 + a_1 n + a_2 n^2 + a_3 n^3 + a_4 n^4 + a_5 p_{man} + a_6 p_{man}^2$	0.0082	79.3%
$a_0 + a_1 n + a_2 n^2 + a_3 n^3 + a_4 n^4 + a_5 n^5 + a_6 p_{man}$	0.0081	37.9%

- Non si evidenzia una dipendenza significativa da T (perché le prove sono state effettuate in ambiente condizionato?)
- Tra i modelli selezionati il secondo è quello che garantisce un migliore compromesso tra semplicità, valore di EQM e deviazione standard dei parametri stimati.

- I risultati che si ottengono con il modello 2 sui dati di validazione sono riportati nelle figure seguenti

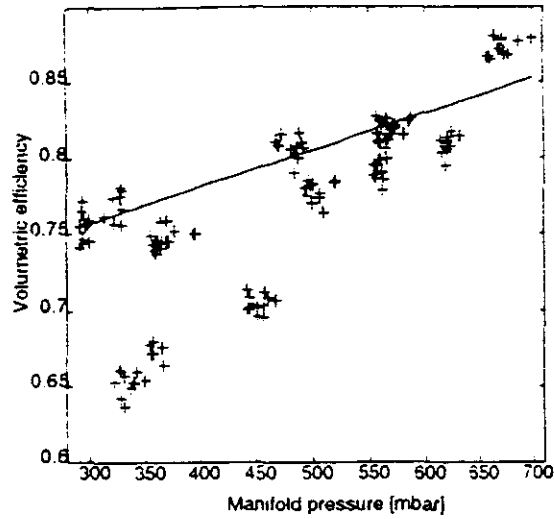


Figure 7: Polynomial model identification of η_v . Plot of $a_0 + \text{mean}(a_1n + a_2n^2 + a_3n^3 + a_4n^4) + a_5p_{man}$. The symbol "+" indicates the validation data.

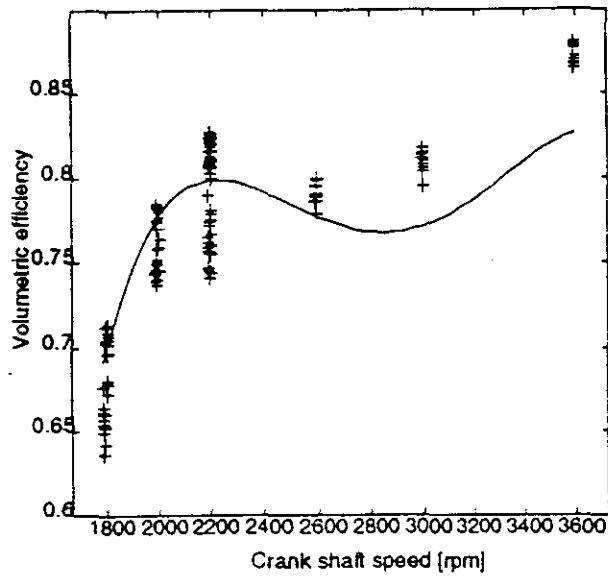


Figure 6: Polynomial model identification of η_v . Plot of $a_0 + a_1n + a_2n^2 + a_3n^3 + a_4n^4 + \text{mean}(a_5p_{man})$. The symbol "+" indicates the validation data.

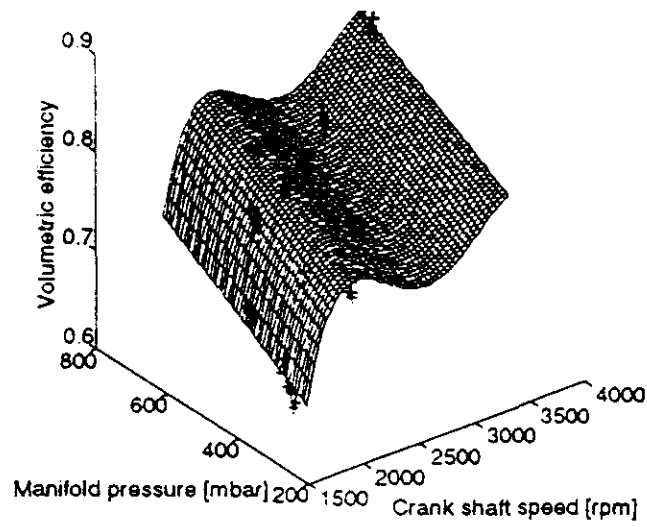


Figure 5: Polynomial model identification of η_v as a function of n , p_{man} (model 2). The symbol "+" indicates the validation data.

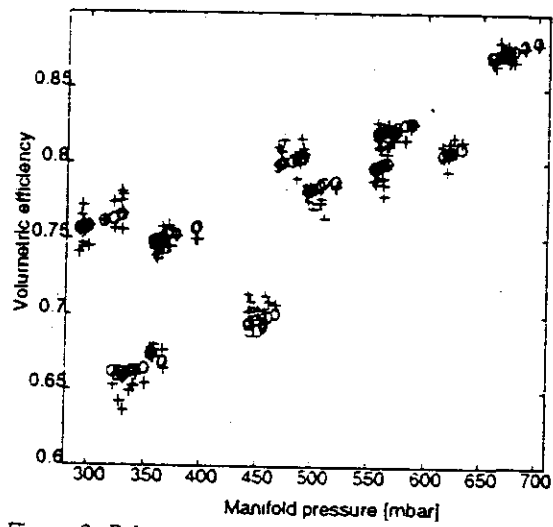


Figure 8: Polynomial model identification of η_v . Validation (+) versus predicted (o) data.

MODELLO DELLA PRESSIONE NEL COLLETTORE

L'equazione della dinamica nel collettore di aspirazione è

$$\dot{p}(t) = -\frac{V}{120\bar{V}}\eta p(t)n(t) + \frac{RT(t)}{\bar{V}}\dot{m}(t)$$

Avendo a disposizione i valori di V (cilindrata), \bar{V} (volume collettore), T e conoscendo la dipendenza di \dot{m} dall'angolo di apertura della valvola a farfalla e la forma identificata di η , il modello è completamente definito

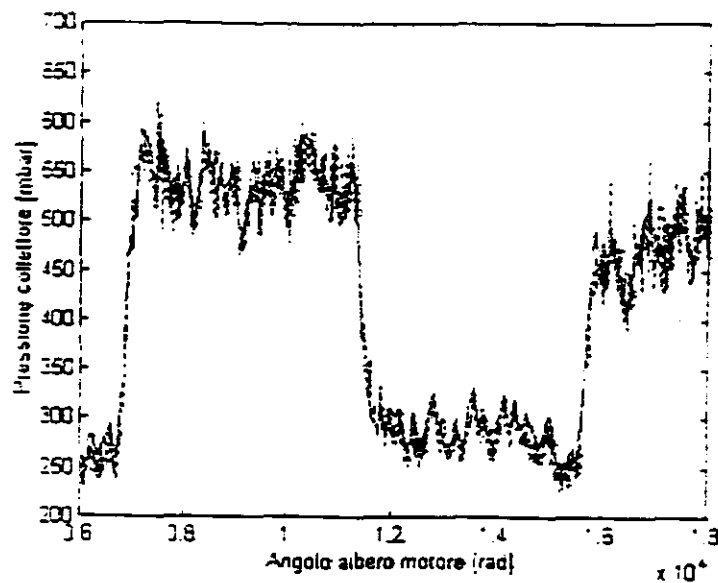


Fig. 4.40 - Pressione nel collettore misurata (linea tratteggiata) e predetta (linea continua) in parte della seconda prova dinamica.

STEPWISE REGRESSION

Supponiamo di dover scegliere la struttura del modello e di dover identificare i relativi parametri avendo a disposizione $r=(m+1)^q$ possibili regressori. Come primo passo dividiamo l'insieme dei dati in dati di stima, utilizzati appunto per la stima dei parametri, e dati di validazione, impiegati per decidere se è opportuno inserire nel modello i singoli regressori, cioè per la valutazione di $J(\mathcal{LS})$. Quindi la procedura si sviluppa secondo i seguenti passi:

1. Si identifichino, con il metodo LS , r modelli distinti considerando soltanto un regressore alla volta ed utilizzando i dati di stima. Per ogni modello identificato si valuti, sui dati di validazione, il valore della cifra di merito $J(\mathcal{LS})$. Si selezioni il regressore più significativo, cioè quello che minimizza tale cifra di merito e si definisca con J_{min} il corrispondente valore calcolato. Tale regressore entrerà nei modelli considerati al passo successivo.

Esempio

Al primo passo il termine $\theta_1 u_1$ è quello che corrisponde al valore minimo di J tra gli r calcolati, quindi il modello temporaneamente selezionato è $y = \theta_1 u_1$ a cui corrisponde un dato valore di J_{min}

2. Si estenda temporaneamente il modello selezionato inserendo, uno alla volta, i regressori rimanenti, si stimino i parametri degli $r-1$ modelli così ottenuti, e si valutino i corrispondenti valori di J .

Esempio

Vengono stimati i parametri dei modelli $y = \theta_0 + \theta_1 u_1$, $y = \theta_1 u_1 + \theta_2 u_2$, ..., $y = \theta_1 u_1 + \theta_2 u_1^2 + \theta_3 u_2^2$ e vengono calcolati i corrispondenti valori di J .

3. Se esiste qualche regressore che migliora il criterio LS rispetto al valore J_{min} assunto dallo stesso indice e relativo al modello selezionato al passo precedente, si includa nel modello il regressore che produce il decremento massimo di J e si aggiorni corrispondentemente il valore di J_{min} . In caso contrario si vada al passo 5.

Esempio

Il modello $y = \theta_1 u_1 + \theta_2 u_2$ produce il massimo decremento di J rispetto al valore di J_{min} precedente.

4. Si eliminino uno alla volta tutti i regressori inseriti nel modello, si ristimino i parametri dei modelli così ottenuti e i corrispondenti valori di J . Si verifichi così se la loro eliminazione riduce il valore di J rispetto al J_{min} precedentemente trovato. Se è così, si elimini dal modello il regressore in esame. Quando questo passo è completato si vada al passo 2.

Questo passo è giustificato dal fatto che può esserci un regressore che se considerato singolarmente è più significativo di altri, mentre perde di significatività quando nel modello sono stati inseriti altri termini. Evidentemente il passo non ha senso se nel modello vi sono soltanto due regressori.

Esempio

Supponiamo che nel modello, dopo aver inserito secondo le modalità precedenti i regressori $\theta_1 u_1$ e $\theta_2 u_2$, sia stato aggiunto il termine $\theta_3 u_1 u_2$. Il modello selezionato è dunque $y = \theta_1 u_1 + \theta_2 u_2 + \theta_3 u_1 u_2$ a cui corrisponde un certo valore J_{min} . Ci si deve chiedere se la cifra di merito associata al modello $y = \theta_2 u_2 + \theta_3 u_1 u_2$ (sino ad ora non considerato) sia minore di J_{min} .

5. Stop.

CONCLUSIONI

- I modelli polinomiali sono molto utilizzati. Il problema connesso con il loro impiego è dovuto al grande numero di potenziali regressori che è necessario considerare a priori
- Una analisi preliminare dei dati (a volte basta solo guardarli!) consente spesso di restringere l'insieme dei regressori potenzialmente candidati. In alternativa si possono usare metodi di stima non parametrica (splines, modelli additivi) per avere un'idea preliminare. Questo punto però non verrà approfondito nel seguito
- La stima dei singoli modelli è di norma semplice, i criteri soggettivi e oggettivi visti in precedenza aiutano nella individuazione di un buon modello candidato
- Un esempio non banale ha mostrato le potenzialità dell'approccio
- Il metodo della stepwise regression è molto efficace e consentirà di ottenere ottimi risultati nella stima di modelli dinamici

ASPETTI COMPUTAZIONALI DELLA STIMA LS

- Le equazioni normali possono essere mal condizionate e un piccolo errore sui dati può provocare un errore molto grande sulla soluzione
- E' necessario ricorrere a metodi efficienti per la loro soluzione
- Tra i metodi più diffusi vi sono: la decomposizione di Cholesky, la decomposizione ortogonale, la decomposizione ai valori singolari
- E' di grande importanza pratica normalizzare i dati

UN ESEMPIO

Supponiamo di avere il modello

$$y(t) = a_1 + a_2 u(t)$$

dove i parametri a_1 e a_2 (da stimare) hanno il valore (vero)

$$a_1 = -3.44952372, a_2 = 0.45457142$$

Supponiamo di usare per la stima i seguenti dati sperimentali (affetti da rumore)

$$Y = \begin{bmatrix} 1 \\ 1.2 \\ 1.25 \\ 1.267 \\ 1.268 \\ 1.274 \end{bmatrix}, \quad U = \begin{bmatrix} 10 \\ 10.1 \\ 10.2 \\ 10.3 \\ 10.4 \\ 10.5 \end{bmatrix}$$

In corrispondenza le equazioni normali sono

$$\begin{bmatrix} 6 & 61.5 \\ 61.5 & 630.55 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 7.259 \\ 74.4843 \end{bmatrix}$$

la cui soluzione porta alla stima corretta dei parametri

se, per qualche errore nella determinazione di u e di y , il sistema a cui si perviene invece è

$$\begin{bmatrix} 6 & 61.5 \\ 61.5 & 630.6 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 7.259 \\ 74.48 \end{bmatrix}$$

la soluzione che si ottiene è

$$a_1 = -2.218222217, a_2 = 0.33444444$$

Un errore relativo di 10^{-4} nei dati ha prodotto un errore relativo dell'ordine dell'unità nei risultati

Il problema è malcondizionato

Nel caso di problemi mal condizionati, anche se i dati sono abbastanza precisi, è necessario utilizzare algoritmi efficienti nella soluzione del sistema di equazioni lineari.

Gran parte degli algoritmi più efficienti sono già disponibili nei packages di calcolo più diffusi (MATLAB, MATRIX-X).

METODI DI SOLUZIONE DEL PROBLEMA LS

$$(\Phi' \Phi) \Theta = \Phi' Y$$

con $\Phi \in \mathbb{R}^{N,q}$, $Y \in \mathbb{R}^{q,l}$

- Decomposizione di Cholesky di $\Phi' \Phi$
- Decomposizione ortogonale di Φ
- Decomposizione ai valori singolari di Φ

DECOMPOSIZIONE DI CHOLESKY

Si determina una matrice $U \in R^{q,q}$ triangolare superiore

$$U = \begin{pmatrix} u_1 & * & * & * & * \\ 0 & u_2 & * & * & * \\ 0 & 0 & \dots & * & * \\ 0 & 0 & 0 & u_{q-1} & * \\ 0 & 0 & 0 & 0 & u_q \end{pmatrix}, \quad u_i > 0, \quad i = 1, \dots, q$$

tale che

$$\Phi' \Phi = U' U$$

Quindi

$$\Phi' \Phi \Theta = \Phi' Y$$

\Downarrow

$$U' U \Theta = \Phi' Y$$

\Downarrow

$$U' q = \Phi' Y$$

$$U \Theta = q$$

sono sistemi triangolari per i quali esistono efficienti metodi di soluzione

E' il metodo computazionalmente meno oneroso, ma non è particolarmente robusto dal punto di vista numerico

DECOMPOSIZIONE ORTOGONALE

Si determinano due matrici $Q \in \mathbb{R}^{N,q}$ e $R \in \mathbb{R}^{q,q}$ tali che

$$\Phi = QR$$

Inoltre Q e R hanno la seguente struttura

$$R = \begin{vmatrix} 1 & * & * & * & * \\ 0 & 1 & * & * & * \\ 0 & 0 & \dots & * & * \\ 0 & 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 & 1 \end{vmatrix}$$

$$Q'Q = D = \begin{vmatrix} d_1 & 0 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & d_{q-1} & 0 \\ 0 & 0 & 0 & 0 & d_q \end{vmatrix}, \quad d_i > 0, \quad i = 1, \dots, q$$

le colonne di Q sono ortogonali (ortonormali se $d_i=1$)

METODO RISOLUTIVO DEL PROBLEMA LS

$$\Phi'Y - \Phi'\Phi\Theta = 0$$

\Downarrow

$$\Phi'(Y - \Phi\Theta) = 0$$

\Downarrow

$$R'Q'(Y - QR\Theta) = R'(Q'Y - Q'QR\Theta) = R'(Q'Y - DR\Theta) =$$

\Downarrow

$$Dq = Q'Y$$

$$R\Theta = q$$

sono un sistema triangolare e uno diagonale per i quali esistono efficienti metodi di soluzione

Determinazione della fattorizzazione QR:

- metodo di Gram-Schmidt
- metodo di Gram-Schmidt modificato
- metodo di Householder

DECOMPOSIZIONE AI VALORI SINGOLARI

Si determinano tre matrici $U \in \mathbb{R}^{N,q}$, $V \in \mathbb{R}^{q,q}$ e $\Sigma \in \mathbb{R}^{q,q}$ tali che

$$\Phi = U\Sigma V'$$

Inoltre U e V sono ortogonali, $\|U'x\| = \|x\|$, e Σ è diagonale:

$$U'U = I_q, V'V = I_q, \Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_r & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & \dots \end{pmatrix}, \sigma_i > 0, i = 1, \dots, r$$

La pseudoinversa Φ^+ di Φ è

$$\Phi^+ = V\Sigma^+U'$$

con

$$\Sigma^+ = \begin{pmatrix} 1/\sigma_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/\sigma_r & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & \dots \end{pmatrix}, \sigma_i > 0, i = 1, \dots, r$$

METODO RISOLUTIVO DEL PROBLEMA LS

$$\|\Phi\Theta - Y\| = \|U'(\Phi\Theta - Y)\| = \|U'(\Phi VV'\Theta - Y)\| =$$

↓

$$= \|U'\Phi VV'\Theta - U'Y\| = \|U'U\Sigma VV'\Theta - U'Y\| =$$

↓

$$\|\Sigma V'\Theta - U'Y\|$$

il minimo si ottiene ponendo

$$\Theta = (\Sigma V')^{-1} U'Y = V'^{-1} \Sigma^+ U'Y = V' \Sigma^+ U'Y = \Phi^+ Y$$

Il metodo garantisce un'elevata precisione numerica al costo di un notevole onere computazionale. E' da impiegarsi in particolare quando $\text{rango}(\Phi) < q$.

NORMALIZZAZIONE DEI DATI

Affinché il sistema di equazioni normali

$$(\Phi' \Phi) \Theta = \Phi' Y$$

sia numericamente ben condizionato, è opportuno che i dati che formano la matrice ϕ e il vettore Y siano normalizzati. Per questo è possibile seguire diverse modalità: sia x una variabile casuale

$$E[x] = \bar{x} \quad , \quad E[(x - \bar{x})^2] = s^2$$

allora, anziché usare x , si può usare

- $z = x / \bar{x}$

- $z = \frac{x - \bar{x}}{\sigma}$

In questo caso risulta

$$E[z] = \frac{E[x] - \bar{x}}{\sigma} = 0 \quad , \quad E[z^2] = \frac{E[(x - \bar{x})^2]}{\sigma^2} = 1$$

le variabili vengono quindi normalizzate in modo da avere tutte valore atteso nullo e varianza unitaria. Certamente, una volta stimati i parametri relativamente a z , è necessario rielaborare il modello per determinare i parametri in x .

CONCLUSIONI

- **Attenzione al condizionamento del problema.** Spesso (soprattutto nel caso in cui si usino modelli di tipo polinomiale non ben scelti) il sistema è mal condizionato e le stime non sono affidabili. Questo perché le colonne della matrice Φ sono "quasi" linearmente dipendenti
- **Normalizzare i dati è un'operazione DA FARSI SEMPRE.** Anche se non vi sono teorie molto sviluppate a riguardo, l'esperienza dimostra che dati non normalizzati impediscono di ottenere buoni risultati. Questo è vero non solo per i minimi quadrati lineari, ma in generale per tutte le tecniche di stima, ad esempio quelle basate su reti neurali

MINIMI QUADRATI NON LINEARI (NLS)

- Se il modello è non lineare nei parametri la soluzione del problema ai minimi quadrati non è in forma chiusa
- E' necessario utilizzare metodi di soluzione di tipo iterativo
- I maggiori problemi sono dovuti alla presenza di minimi locali, alla inizializzazione degli algoritmi, alla velocità di convergenza, ai test di fine

Nel caso di modelli non lineari nei parametri, avendo come di consueto N rilevazioni disponibili della variabile dipendente y e delle variabili indipendenti u_1, u_2, \dots, u_q , si è visto che vale la relazione

$$Y = \Phi(U, \Theta)$$

con il consueto significato dei vettori Y, U, Φ, Θ . Anche in questo caso si può formulare il problema ai minimi quadrati che consiste nella minimizzazione rispetto al vettore dei parametri incogniti Θ del funzionale

$$JNLS(\Theta) = (Y - \Phi(U, \Theta))'(Y - \Phi(U, \Theta))$$

In questo caso tuttavia non vi è più una formula esplicita, ma la soluzione va trovata per via iterativa sfruttando uno dei metodi iterativi della Ricerca Operativa.

Metodi iterativi di ottimizzazione

Tali metodi, posto che Θ^i sia la stima di Θ all'iterazione i -esima, determinano la stima Θ^{i+1} all'iterazione successiva secondo la relazione generale

$$\Theta^{i+1} = \Theta^i + \alpha^i f^i$$

dove f^i è la direzione di ricerca ottenuta in base alle informazioni raccolte sulla forma di $JNLS(\Theta)$ alle iterazioni precedenti, mentre α^i è un coefficiente da scegliersi in modo fisso ($\alpha^i = \alpha^{i+1}$) o variabile con le iterazioni. In generale per α^i piccolo si migliorano le capacità di convergenza degli algoritmi a scapito della velocità della convergenza stessa.

CLASSIFICAZIONE DEI METODI ITERATIVI

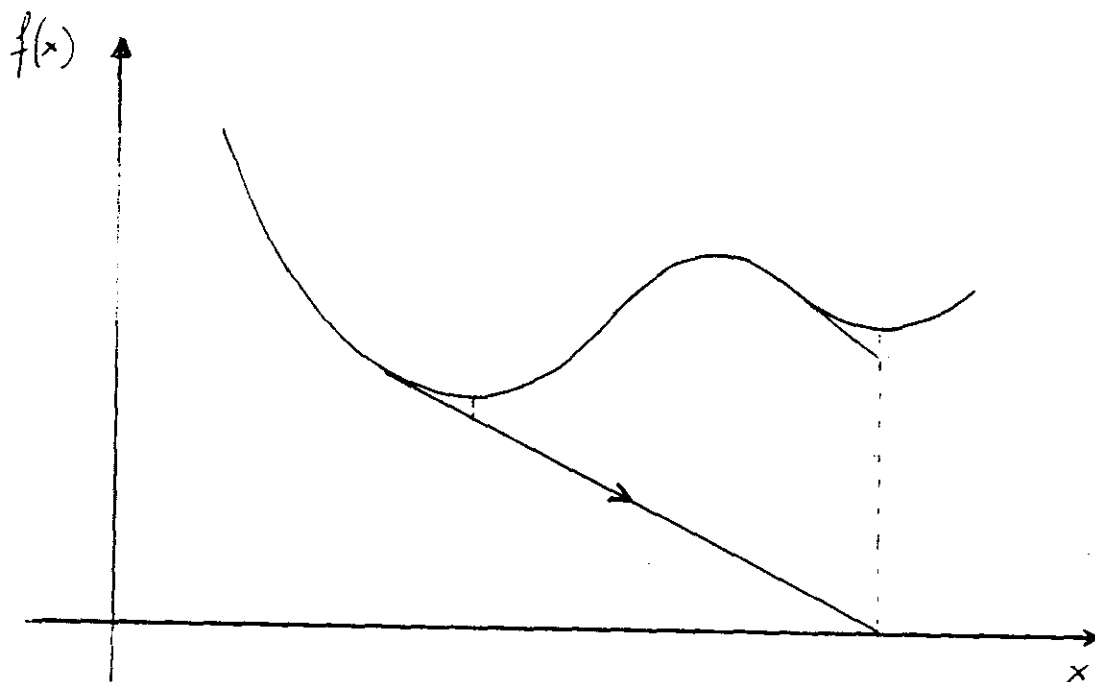
I metodi di minimizzazione normalmente impiegati possono essere classificati come:

- metodi che si basano sulla sola valutazione (analitica o numerica) di $J^{NLS}(\Theta)$ in corrispondenza dei valori Θ^i ;
- metodi che fanno uso della valutazione (analitica o numerica) del vettore gradiente $\left. \frac{dJ^{NLS}(\Theta)}{d\Theta} \right|_{\Theta=\Theta^i}$;
- metodi che prevedono la valutazione (analitica o numerica) del gradiente valutato e della matrice Hessiana $\left. \frac{d^2 J^{NLS}(\Theta)}{d\Theta^2} \right|_{\Theta=\Theta^i}$.

ovviamente maggiori sono le informazioni disponibili (gradiente, matrice Hessiana), migliore sono in genere le prestazioni degli algoritmi. Tralasciando i metodi del primo tipo, utilizzati quando non è disponibile la forma analitica di J e non si vuole procedere alla determinazione numerica del gradiente, si analizzano rapidamente nel seguito le caratteristiche degli altri due approcci.

Nonlinear Least Squares

Va comunque osservato che nessuno di questi metodi offre garanzie di convergenza verso il minimo globale. In base all'inizializzazione considerata (θ^0) si troveranno quindi dei minimi locali. Per questa ragione è opportuno ripetere la procedura di minimizzazione a partire da diverse condizioni iniziali.



Inoltre si ricorda che nell'implementazione di un algoritmo iterativo hanno grande importanza i test di fine, normalmente basati sul numero massimo di iterazioni, sulla norma del gradiente della cifra di merito e sulla norma della differenza tra i parametri stimati in due iterazioni successive.

Il metodo del gradiente

Si osservi innanzitutto che il gradiente della cifra di merito è dato da

$$\frac{dJ^{NLS}(\Theta)'}{d\Theta} = 2 \frac{d\Phi(U, \Theta)'}{d\Theta} (-Y + \Phi(U, \Theta))$$

Poiché il gradiente rappresenta la massima direzione (locale) di accrescimento della funzione, è naturale allora modificare le stime lungo la direzione ad esso opposta, cioè porre

$$\Theta^{i+1} = \Theta^i + \alpha^i \left. \frac{d\Phi(U, \Theta)'}{d\Theta} \right|_{\Theta = \Theta^i} (Y - \Phi(U, \Theta))$$

dove, come detto, α^i è il passo di discesa. La scelta di α^i può essere effettuata risolvendo, ad ogni iterazione, un altro problema di ottimizzazione e precisamente

$$\min_{\alpha^i} J^{NLS}(\Theta^{i+1})$$

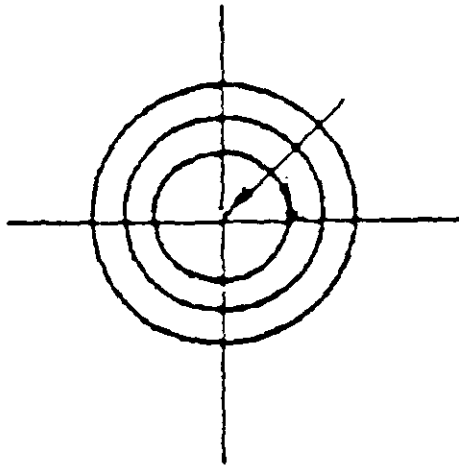
Questo problema può facilmente essere risolto con l'algoritmo di Fibonacci. In questo modo le prestazioni vengono nettamente migliorate, in particolare in prossimità del minimo.

Qualsiasi metodo evolutivo si utilizzi, è comunque opportuno inserire "ogni tanto" un'iterazione fatta con il metodo del gradiente in versione ottima.

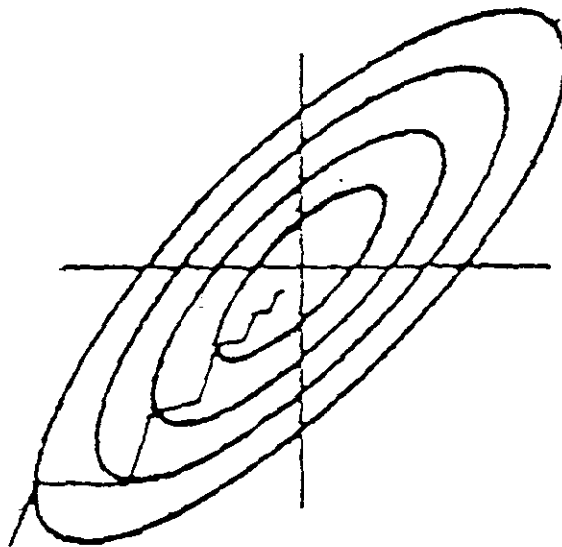
Nonlinear Least Squares

Un problema connesso con l'uso del metodo del gradiente è il grande numero di iterazioni spesso richiesto per ottenere la soluzione

Esempio: linee di livello di una funzione di due variabili



in questo caso va tutto bene



in questo caso sono necessarie moltissime iterazioni

Il metodo di Newton-Raphson

Sviluppando $J^{NLS}(\Theta)$ attorno a Θ^i e arrestandosi al secondo termine dello sviluppo si ottiene

$$J^{NLS}(\Theta) \cong J_a^{NLS}(\Theta) = \\ = J^{NLS}(\Theta^i) + \left. \frac{dJ^{NLS}(\Theta)}{d\Theta} \right|_{\Theta=\Theta^i} (\Theta - \Theta^i) + \frac{1}{2} (\Theta - \Theta^i)' \left. \frac{d^2 J^{NLS}(\Theta)}{d\Theta^2} \right|_{\Theta=\Theta^i} (\Theta - \Theta^i)$$

la cifra di merito è dunque approssimata con la funzione quadratica. Derivando ed uguagliando a zero tale funzione si ottiene

$$\frac{dJ_a^{NLS}(\Theta)}{d\Theta} = \left. \frac{dJ^{NLS}(\Theta)}{d\Theta} \right|_{\Theta=\Theta^i} + (\Theta - \Theta^i)' \left. \frac{d^2 J^{NLS}(\Theta)}{d\Theta^2} \right|_{\Theta=\Theta^i} = 0$$

cioè

$$\Theta = \Theta^i - \left. \frac{d^2 J^{NLS}(\Theta)}{d\Theta^2} \right|_{\Theta=\Theta^i}^{-1} \left. \frac{dJ^{NLS}(\Theta)}{d\Theta} \right|_{\Theta=\Theta^i}$$

e quindi il metodo consiste nel porre

$$\Theta^{i+1} = \Theta^i - \left. \frac{d^2 J^{NLS}(\Theta)}{d\Theta^2} \right|_{\Theta=\Theta^i}^{-1} \left. \frac{dJ^{NLS}(\Theta)}{d\Theta} \right|_{\Theta=\Theta^i}$$

Si osservi che risulta

$$\frac{d^2 J^{NLS}(\Theta)}{d\Theta^2} = 2 \frac{d\Phi(U, \Theta)' d\Phi(U, \Theta)}{d\Theta} + 2(\Phi(U, \Theta)' - Y') \frac{d^2 \Phi(U, \Theta)}{d\Theta^2}$$

Il metodo di Gauss-Newton

I metodi che fanno uso della matrice Hessiana, o metodi del secondo ordine, possono essere computazionalmente onerosi. Si usano quindi spesso metodi che approssimano tale matrice pur facendo ricorso soltanto al calcolo del gradiente. Nell'identificazione, il più diffuso è il metodo di Gauss-Newton che può essere così dedotto.

Sviluppando $\Phi(U, \Theta)$ attorno a Θ^i si ottiene

$$\Phi(U, \Theta) \cong \Phi(U, \Theta^i) + \left. \frac{d\Phi(U, \Theta)}{d\Theta} \right|_{\Theta=\Theta^i} (\Theta - \Theta^i)$$

e quindi

$$\begin{aligned} JNLS(\Theta) &= (Y - \Phi(U, \Theta))' (Y - \Phi(U, \Theta)) \cong \\ &\cong (Y - \Phi(U, \Theta^i) - \left. \frac{d\Phi(U, \Theta)}{d\Theta} \right|_{\Theta=\Theta^i} (\Theta - \Theta^i))' (Y - \Phi(U, \Theta^i) - \left. \frac{d\Phi(U, \Theta)}{d\Theta} \right|_{\Theta=\Theta^i} (\Theta - \Theta^i)) \end{aligned}$$

ancora una volta uguagliando a zero la derivata prima di questa espressione rispetto a Θ si ottiene

$$\Theta = \Theta^i + \left[\left. \frac{d\Phi(U, \Theta)'}{d\Theta} \frac{d\Phi(U, \Theta)}{d\Theta} \right]_{\Theta=\Theta^i}^{-1} \left. \frac{d\Phi(U, \Theta)'}{d\Theta} \right|_{\Theta=\Theta^i} (Y - \Phi(U, \Theta^i))$$

da cui, introducendo ancora il parametro α^i

$$\Theta^{i+1} = \Theta^i + \alpha^i \left[\left. \frac{d\Phi(U, \Theta)'}{d\Theta} \frac{d\Phi(U, \Theta)}{d\Theta} \right]_{\Theta=\Theta^i}^{-1} \left. \frac{d\Phi(U, \Theta)'}{d\Theta} \right|_{\Theta=\Theta^i} (Y - \Phi(U, \Theta^i))$$

Si osservi che il metodo di Gauss Newton è strettamente legato a quello di Newton-Raphson, è sufficiente ipotizzare nel primo tuttavia la matrice Hessiana è approssimata unicamente in base alla conoscenza del gradiente della cifra di merito rispetto ai parametri incogniti.

Nel corso delle iterazioni può anche accadere che

$$\det \left. \frac{d\Phi(U, \Theta)' d\Phi(U, \Theta)}{d\Theta} \right|_{\Theta=\Theta^i} \cong 0$$

E' stata allora proposta la variante di Levenberg-Marquardt che consiste nell'utilizzare la formula

$$\Theta^{i+1} = \Theta^i + \alpha^i \left. \frac{d\Phi(U, \Theta)' d\Phi(U, \Theta)}{d\Theta} \right|_{\Theta=\Theta^i}^{-1} + \beta^i I \left. \frac{d\Phi(U, \Theta)}{d\Theta} \right|_{\Theta=\Theta^i} (Y - \Phi(U, \Theta))$$

Dove β^i è zero se il determinante è (significativamente) diverso da zero, mentre è posto ad un valore piccolo in caso contrario così da consentire la prosecuzione dell'algoritmo di minimizzazione.

STIMA DELLA VARIANZA DEL DISTURBO, VALIDAZIONE E CONFRONTO TRA MODELLI

Supponiamo che il modello sia

$$Y = \Phi(U, \Theta) + V$$

con le consuete ipotesi su V e supponiamo di aver determinato la stima Θ^{NLS} . Linearizzando nell'intorno della soluzione si ha

$$Y = \Phi(U, \Theta^{NLS}) + \Phi^*(U, \Theta^{NLS})(\Theta - \Theta^{NLS}) + V$$

dove $\Phi^*(U, \Theta^{NLS}) = \left. \frac{d\Phi(U, \Theta)}{d\Theta} \right|_{\Theta = \Theta^{NLS}}$. Pur di definire $\bar{Y} = Y - \Phi(U, \Theta^{NLS})$ e $\bar{\Theta} = \Theta - \Theta^{NLS}$, il modello si può scrivere nella forma

$$\bar{Y} = \Phi^*(U, \Theta^{NLS})\bar{\Theta} + V$$

Questo modello linearizzato, valido nell'intorno di Θ^{NLS} , consente di calcolare con le tecniche illustrate nel caso lineare, la varianza del disturbo e della stima. Consente inoltre la validazione del modello mediante il test χ^2 e il confronto tra vari modelli per la selezione del modello ottimo.

IDENTIFICAZIONE DELL'EQUAZIONE DELLA PRESSIONE NEL
COLLETTORE DIRETTAMENTE DAI DATI

Nell'equazione della dinamica nel collettore di aspirazione è

$$\dot{p}(t) = -\frac{V}{120\bar{V}} \eta p(t)n(t) + \frac{RT(t)}{\bar{V}} \dot{m}(t)$$

spesso \bar{V} e T non sono noti. Inoltre può essere costoso fare prove statiche per ricavare la dipendenza di \dot{m} dall'angolo di apertura della valvola α a farfalla o del rendimento volumetrico η da n, p, T . Si ha invece una (o più) registrazione di α, p, n nel tempo

Si può ipotizzare che

$$\eta = a_0 + a_1 n + a_2 n^2 + a_3 n^3 + a_4 n^4 + a_5 p$$

e

$$\dot{m} = a_6 + a_7 \alpha + a_8 \alpha^2$$

Per valori fissati dei parametri a_i si può integrare l'equazione differenziale precedente e determinare l'andamento della pressione stimata p^*

Si può quindi impostare il problema di ottimizzazione

$$\min_{a_i} \sum_{j=1}^N (p(j) - p^*(j))^2$$

chiaramente di tipo non lineare nei parametri incogniti

In questo modo si può identificare un modello mediante tecniche di stima non lineare che fanno uso del gradiente (o di una sua stima) e/o dell'Hessiano

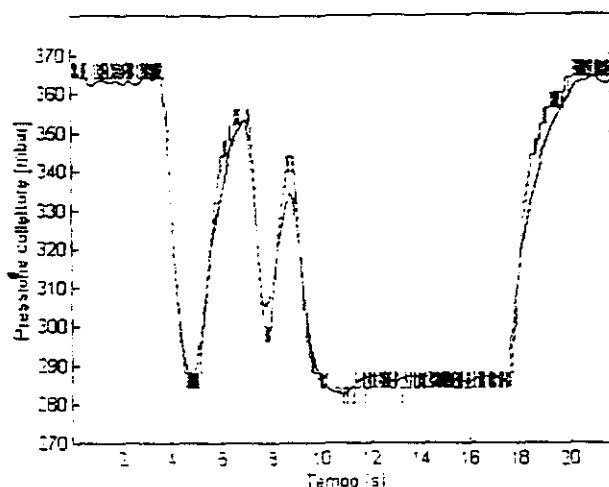


Fig. 4.26 - Quinta prova dinamica: in linea tratteggiata è rappresentata la pressione misurata nel collettore e in linea continua quella simulata dal modello identificato.

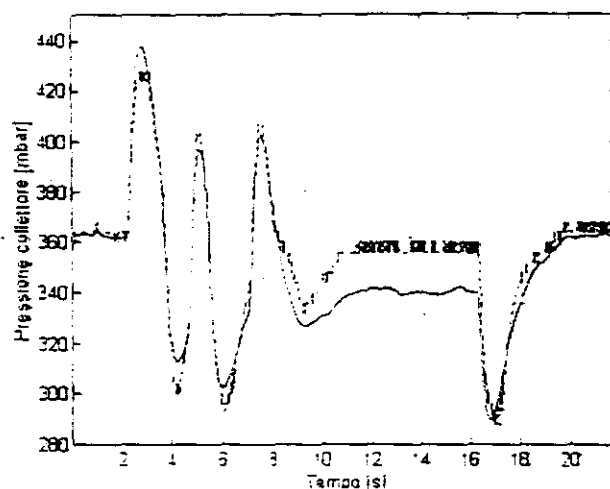


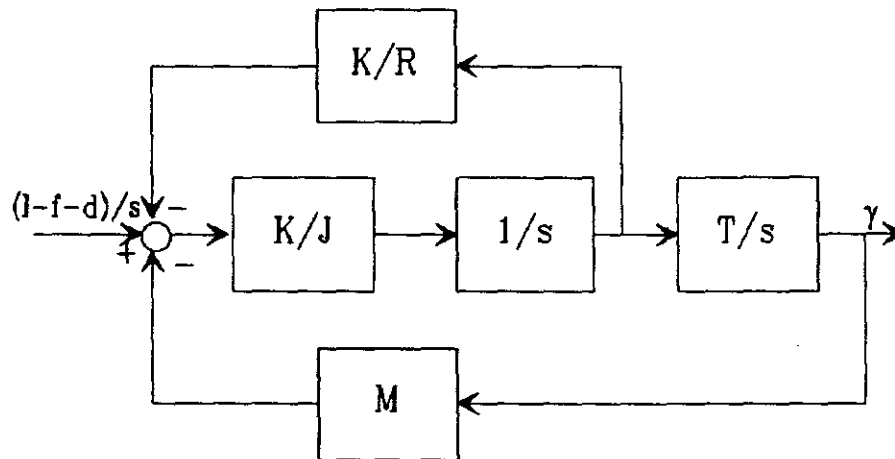
Fig. 4.27 - Terza prova dinamica: in linea tratteggiata è rappresentata la pressione misurata nel collettore ed in linea continua quella simulata dal modello identificato.

Tabella 4.6 - Errore d'uscita e di predizione nelle prove di validazione, generati dal modello fisico della pressione nel collettore di alimentazione

Numero prova	Errore quadratico medio di uscita (in mbar)	Errore quadratico medio di predizione (in mbar)
1	10.62	2.18
3	11.04	2.30
4	10.10	2.02
5	4.21	1.83
6	7.36	1.81

IDENTIFICAZIONE DEI PARAMETRI DEL CORPO FARFALLATO

Si è visto in precedenza che , dato il modello del corpo farfallato



è necessario stimare i valori di K (costante di velocità), J (momento di inerzia), f (forza di attrito cinetico), M (costante elastica delle molle). Anche in questo caso, definendo con γ l'angolo vero e con γ^* quello stimato dal modello, i parametri possono essere determinati minimizzando il funzionale

$$\min \sum_{j=1}^N (\gamma(j) - \gamma^*(j))^2$$

dove N è il numero di misure disponibili

Risultati dell'identificazione

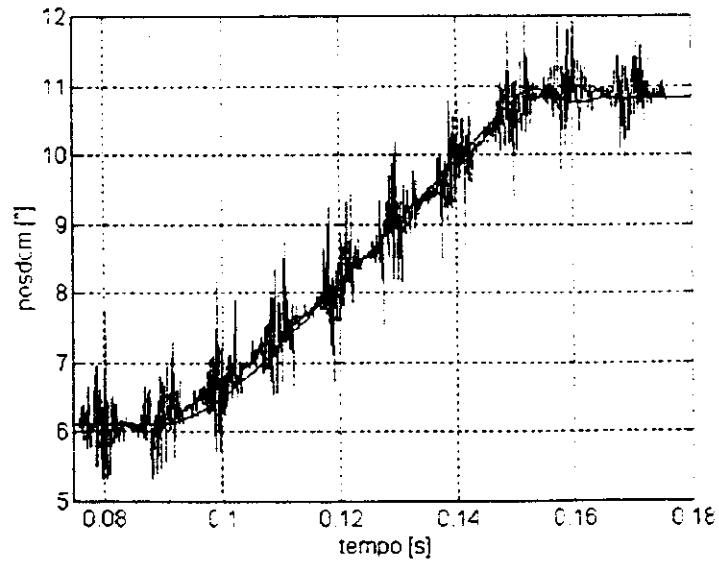


Figura 5.21. Transitorio reale da 0% a 50% e simulato da 0% a 49%.

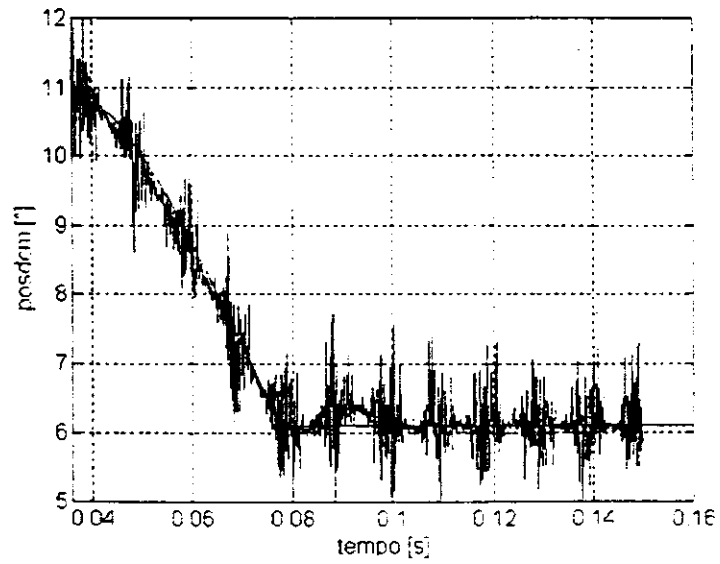


Figura 5.19. Transitorio da 50% a 20%.

CONCLUSIONI

- Spesso i programmi di ottimizzazione richiedono soltanto la valutazione della funzione e provvedono automaticamente alla valutazione numerica del gradiente e alla stima della matrice Hessiana
- E' bene ripetere l'ottimizzazione partendo da diverse condizioni iniziali per ridurre il problema dei minimi locali
- Attenzione: se il numero di parametri non è piccolo (<3-4) il tempo richiesto può essere elevato
- Spesso nei pacchetti (MATLAB, MatrixX) si trovano programmi di ottimizzazione dove il funzionale di costo è generico e altri basati su algoritmi pensati in modo specifico per problemi (non lineari) ai minimi quadrati. E' ovviamente opportuno usare questi ultimi se il problema lo consente