Algoritmo di stima "minimi quadrati ricorsivi" per sistemi con ingressi e uscite vettoriali

Lorenzo Magliocchetti Arrigo Marchiori Michele Marino

Ottobre 2006

Sommario

Lo scopo di questo documento è ricavare le formule che descrivono l'algoritmo di stima dei "minimi quadrati ricorsivi", applicato a sistemi con grandezze vettoriali in ingresso e in uscita. In letteratura è possiblie trovare numerose descrizioni di tale algoritmo, ma con grandezze scalari.

L'argomento è introdotto a partire dalla formula dei "minimi quadrati batch".

1 Sistema e stimatore

Consideriamo un generico sistema con ingressi vettoriali $\mathbf{x}$ di dimensione qualunque e uscite $\mathbf{y} \in \mathbb{R}^m$.

L'obiettivo è stimare alcuni parametri $\beta \in \mathbb{R}^n$ di questo sistema. A tale scopo, si utilizza uno stimatore che sia caratterizzato da una dipendenza lineare dal vettore dei parametri da stimare:

$$\tilde{\mathbf{y}} = \mathbf{F}(\mathbf{x}, \beta) = \mathbf{\Phi}(\mathbf{x})\beta$$

dove $\mathbf{\Phi}$ è una funzione non necessariamente lineare dell'ingresso $\mathbf{x}$.

La "bontà" dei valori assegnati ai parametri dello stimatore deve essere giudicata in base ad un criterio; normalmente si definisce una funzione di penalità da minimizzare.

1.1 Criterio dei minimi quadrati *batch*

Si effettuano j "esperimenti", cioè misure dell'ingresso e dell'uscita del sistema. I valori si organizzano in vettori di $\mathbb{R}^{jm}$:

$$\mathbf{x_m} = \begin{bmatrix} \mathbf{x_1} \\ \mathbf{x_2} \\ \vdots \\ \mathbf{x_j} \end{bmatrix} \qquad \mathbf{y_m} = \begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \\ \vdots \\ \mathbf{y_j} \end{bmatrix}$$

Per ogni esperimento si considera l'errore di stima, inteso come la differenza tra l'uscita del sistema e quella dello stimatore:

$$\mathbf{e_i} = \mathbf{y_i} - \tilde{\mathbf{y}}_\mathbf{i} = \mathbf{y_i} - \mathbf{\Phi}(\mathbf{x_i})\beta \tag{1}$$

Incolonnando gli errori di tutti gli esperimenti, si ottiene il vettore $\mathbf{e_m} \in \mathbb{R}^{jm}$:

$$\mathbf{e_m} = \begin{bmatrix} \mathbf{e_1} \\ \mathbf{e_2} \\ \vdots \\ \mathbf{e_j} \end{bmatrix} = \mathbf{y_m} - \begin{bmatrix} \mathbf{\Phi}(\mathbf{x_1}) \\ \mathbf{\Phi}(\mathbf{x_2}) \\ \vdots \\ \mathbf{\Phi}(\mathbf{x_j}) \end{bmatrix} \beta = \mathbf{y_m} - \bar{\mathbf{\Phi}}(\mathbf{x_m})\beta \tag{2}$$

con $\bar{\mathbf{\Phi}}(\mathbf{x_m}) \in \mathbb{R}^{jm \times n}$.

La funzione di penalità che si considera è la seguente:

$$V_j\left(\beta\right) = \frac{1}{2}\mathbf{e_m}^\top \mathbf{R}\mathbf{e_m} \tag{3}$$

dove $\mathbf{R}$ è una matrice simmetrica e definita positiva che ha la funzione di pesare opportunamente gli esperimenti. Con $\mathbf{R} = \mathbf{I}$, ad esempio, tutti gli esperimenti hanno lo stesso peso.

Sostituendo la (2) nella (3), derivando $V_j\left(\beta\right)$ rispetto a β, annullando il risultato e risolvendo rispetto a β, si ottiene:

$$\beta_{opt} = \left(\bar{\mathbf{\Phi}}(\mathbf{x_m})^\top \mathbf{A}\bar{\mathbf{\Phi}}(\mathbf{x_m})\right)^{-1} \bar{\mathbf{\Phi}}(\mathbf{x_m})^\top \mathbf{A}\mathbf{y_m} \tag{4}$$

dove $\mathbf{A} = \left(\mathbf{R} + \mathbf{R}^\top\right)$, anch'essa simmetrica e definita positiva.

Perché questo criterio funzioni, la matrice $\bar{\mathbf{\Phi}}(\mathbf{x_m})^\top \mathbf{A}\bar{\mathbf{\Phi}}(\mathbf{x_m})$ deve essere invertibile; questo accade se l'ingresso $\mathbf{x_m}$ del sistema è sufficientemente eccitante.

1.2 Criterio dei minimi quadrati ricorsivi

Per poter aggiornare "in linea" i parametri dello stimatore, bisogna modificare il criterio dei minimi quadrati, fornendone una formulazione ricorsiva. In questo modo, ad ogni "passo" di computazione k è possibile aggiornare la stima, tenendo conto di ciò che si è calcolato nei passi precedenti.

Si desidera quindi adattare la relazione (4). Omettendo la dipendenza di $\bar{\mathbf{\Phi}}$ da $\mathbf{x_m}$ e riscrivendo l'espressione in funzione del passo k si ottiene:

$$\beta\left(k\right) = \left(\bar{\mathbf{\Phi}}^\top(k)\mathbf{A}\left(k\right)\bar{\mathbf{\Phi}}\left(k\right)\right)^{-1} \bar{\mathbf{\Phi}}^\top(k)\mathbf{A}\left(k\right)\mathbf{y_m}\left(k\right) \tag{5}$$

dove:

$$\bar{\mathbf{\Phi}}\left(k\right) \in \mathbb{R}^{km \times n}, \qquad \mathbf{y_m}\left(k\right) \in \mathbb{R}^{km}, \qquad \mathbf{A}\left(k\right) \in \mathbb{R}^{km \times km}$$

La struttura di $\bar{\mathbf{\Phi}}\left(k\right)$ è la stessa dell'eq. (2), cioè ad ogni passo si aggiunge in basso una nuova sottomatrice. Di conseguenza, la matrice al passo $k+1$ si può scrivere:

$$\bar{\mathbf{\Phi}}\left(k+1\right) = \begin{bmatrix} \bar{\mathbf{\Phi}}\left(k\right) \\ \mathbf{\Phi_{k+1}} \end{bmatrix} \quad \in \mathbb{R}^{(k+1) \times m} \tag{6}$$

dove $\mathbf{\Phi_{k+1}} \in \mathbb{R}^{m \times n}$.

Anche la matrice $\mathbf{A}\left(k\right)$ ed il vettore $\mathbf{y_m}\left(k\right)$ si aggiornano ad ogni passo:

$$\mathbf{y_m}\left(k+1\right) = \begin{bmatrix} \mathbf{y_m}\left(k\right) \\ \mathbf{y_{k+1}} \end{bmatrix} \quad \in \mathbb{R}^{(k+1)m} \tag{7}$$

Nell'aggiornamento di $\mathbf{A}$, però, la parte che si riferisce al passato viene moltiplicata per un coefficiente λ:

$$\mathbf{A}\left(k+1\right)=\begin{bmatrix}\lambda\mathbf{A}\left(k\right) & \\ & \mathbf{A_{k+1}}\end{bmatrix}\quad\in\mathbb{R}^{(k+1)m\times(k+1)m},\qquad 0<\lambda\leq 1 \tag{8}$$

λ è detto *"fattore di dimenticanza"*: si può vedere, ad esempio, che al passo $k+j$ la matrice $\mathbf{A}(k+j)$ contiene la sottomatrice $\mathbf{A}(k)$ moltiplicata per λ^j. Se $\lambda=1$, non si effettua alcuna dimenticanza.

Definiamo la matrice $\mathbf{P}(k)$ attraverso la sua inversa:

$$\mathbf{P}^{-1}(k)=\bar{\mathbf{\Phi}}^\top(k)\mathbf{A}\left(k\right)\bar{\mathbf{\Phi}}\left(k\right)\quad\in\mathbb{R}^{n\times n} \tag{9}$$

Al passo $k+1$, secondo le definizioni appena date (6) e (8), ponendo $\lambda=c^2$, si ottiene:

$$\begin{aligned}\mathbf{P}^{-1}(k+1) &= \bar{\mathbf{\Phi}}^\top(k+1)\mathbf{A}(k+1)\bar{\mathbf{\Phi}}(k+1)=\\ &= \begin{bmatrix}\bar{\mathbf{\Phi}}^\top(k) & \mathbf{\Phi}^\top_{\mathbf{k+1}}\end{bmatrix}\begin{bmatrix}\lambda\mathbf{A}\left(k\right) & \\ & \mathbf{A_{k+1}}\end{bmatrix}\begin{bmatrix}\bar{\mathbf{\Phi}}\left(k\right)\\ \mathbf{\Phi_{k+1}}\end{bmatrix}=\\ &= \begin{bmatrix}\lambda\bar{\mathbf{\Phi}}^\top(k)\mathbf{A}\left(k\right) & \mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{A_{k+1}}\end{bmatrix}\begin{bmatrix}\bar{\mathbf{\Phi}}\left(k\right)\\ \mathbf{\Phi_{k+1}}\end{bmatrix}=\\ &= \lambda\bar{\mathbf{\Phi}}^\top(k)\mathbf{A}\left(k\right)\bar{\mathbf{\Phi}}\left(k\right)+\mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{A_{k+1}}\mathbf{\Phi_{k+1}}=\\ &= \lambda\mathbf{P}^{-1}(k)+\mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{A_{k+1}}\mathbf{\Phi_{k+1}}\end{aligned} \tag{10}$$

Ora si consideri l'espressione di β (5) al passo $k+1$:

$$\beta(k+1)=\left(\bar{\mathbf{\Phi}}^\top(k+1)\mathbf{A}(k+1)\bar{\mathbf{\Phi}}(k+1)\right)^{-1}\bar{\mathbf{\Phi}}^\top(k+1)\mathbf{A}(k+1)\mathbf{y_m}(k+1)$$

riconoscendo nella prima parentesi l'espressione di $\mathbf{P}^{-1}(k+1)$ dalla (9) e sostituendo le espressioni di $\bar{\mathbf{\Phi}}(k+1)$ (6), $\mathbf{A}(k+1)$ (8) e $\mathbf{y_m}(k+1)$ (7) si ottiene:

$$\begin{aligned}\beta(k+1) &= \mathbf{P}(k+1)\bar{\mathbf{\Phi}}^\top(k+1)\mathbf{A}(k+1)\mathbf{y_m}(k+1)=\\ &= \mathbf{P}(k+1)\begin{bmatrix}\bar{\mathbf{\Phi}}^\top(k) & \mathbf{\Phi}^\top_{\mathbf{k+1}}\end{bmatrix}\begin{bmatrix}\lambda\mathbf{A}\left(k\right) & \\ & \mathbf{A_{k+1}}\end{bmatrix}\mathbf{y_m}(k+1)=\\ &= \mathbf{P}(k+1)\begin{bmatrix}\lambda\bar{\mathbf{\Phi}}^\top(k)\mathbf{A}\left(k\right) & \mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{A_{k+1}}\end{bmatrix}\begin{bmatrix}\mathbf{y_m}\left(k\right)\\ \mathbf{y_{k+1}}\end{bmatrix}=\\ &= \mathbf{P}(k+1)\left(\lambda\bar{\mathbf{\Phi}}^\top(k)\mathbf{A}\left(k\right)\mathbf{y_m}\left(k\right)+\mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{A_{k+1}}\mathbf{y_{k+1}}\right)\end{aligned} \tag{11}$$

L'espressione di $\mathbf{P}(k+1)$ si ottiene invertendo la (10):

$$\mathbf{P}(k+1)=\left(\lambda\mathbf{P}^{-1}(k)+\mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{A_{k+1}}\mathbf{\Phi_{k+1}}\right)^{-1}$$

siccome:

$$\left(\mathbf{A}+\mathbf{BCD}\right)^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{DA}^{-1}\mathbf{B}+\mathbf{C}^{-1}\right)^{-1}\mathbf{DA}^{-1} \tag{12}$$

allora:

$$\mathbf{P}(k+1)=\frac{1}{\lambda}\mathbf{P}(k)-\frac{1}{\lambda}\mathbf{P}(k)\mathbf{\Phi}^\top_{\mathbf{k+1}}\left(\frac{1}{\lambda}\mathbf{\Phi_{k+1}}\mathbf{P}(k)\mathbf{\Phi}^\top_{\mathbf{k+1}}+\mathbf{A}^{-1}_{\mathbf{k+1}}\right)^{-1}\frac{1}{\lambda}\mathbf{\Phi_{k+1}}\mathbf{P}(k) \tag{13}$$

Per semplificare l'espressione si introduce il termine $\mathbf{\Gamma}(k) \in \mathbb{R}^{m\times m}$. Esso è invertibile, perché definito positivo.

$$\begin{aligned}
\mathbf{\Gamma}(k) &= \left(\mathbf{\Phi_{k+1}}\mathbf{P}(k)\mathbf{\Phi}^\top_{\mathbf{k+1}} + \lambda\mathbf{A}^{-1}_{\mathbf{k+1}}\right)^{-1} \\
\mathbf{P}(k+1) &= \tfrac{1}{\lambda}\left(\mathbf{P}(k) - \mathbf{P}(k)\mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{\Gamma}(k)\mathbf{\Phi_{k+1}}\mathbf{P}(k)\right)
\end{aligned} \tag{14}$$

Sostituendo questa espressione appena trovata in quella di $\beta(k+1)$ (11) si ottiene:

$$\begin{aligned}
\beta(k+1) &= \mathbf{P}(k)\bar{\mathbf{\Phi}}^\top(k)\mathbf{A}\left(k\right)\mathbf{y_m}\left(k\right) + \frac{1}{\lambda}\mathbf{P}(k)\mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{A_{k+1}}\mathbf{y_{k+1}} + \\
&\quad -\mathbf{P}(k)\mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{\Gamma}(k)\mathbf{\Phi_{k+1}}\mathbf{P}(k)\bar{\mathbf{\Phi}}^\top(k)\mathbf{A}(k)\mathbf{y_m}\left(k\right) + \\
&\quad -\frac{1}{\lambda}\mathbf{P}(k)\mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{\Gamma}(k)\mathbf{\Phi_{k+1}}\mathbf{P}(k)\mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{A_{k+1}}\mathbf{y_{k+1}}
\end{aligned} \tag{15}$$

Abbandonando per un attimo questa espressione, si consideri che per definizione di $\beta(k)$ (5) e di $\mathbf{P}(k)$ (9):

$$\beta(k) = \mathbf{P}(k)\bar{\mathbf{\Phi}}^\top(k)\mathbf{A}\left(k\right)\mathbf{y_m}\left(k\right) \tag{16}$$

e per la seconda delle (14):

$$\begin{aligned}
\mathbf{P}(k+1)\mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{A_{k+1}} &= \frac{1}{\lambda}\left(\mathbf{P}(k)\mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{A_{k+1}} - \mathbf{P}(k)\mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{\Gamma}(k)\mathbf{\Phi_{k+1}}\mathbf{A_{k+1}}\right) = \\
&= \frac{1}{\lambda}\mathbf{P}(k)\mathbf{\Phi}^\top_{\mathbf{k+1}}\left(\mathbf{A_{k+1}} - \mathbf{\Gamma}(k)\mathbf{\Phi_{k+1}}\mathbf{P}(k)\mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{A_{k+1}}\right)
\end{aligned} \tag{17}$$

L'espressione tra parentesi è pari a $\lambda\mathbf{\Gamma}(k)$:

$$\mathbf{A_{k+1}} - \mathbf{\Gamma}(k)\mathbf{\Phi_{k+1}}\mathbf{P}(k)\mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{A_{k+1}} = \lambda\mathbf{\Gamma}(k)$$

infatti, raggruppando per $\mathbf{\Gamma}(k)$:

$$\mathbf{\Gamma}(k)\left(\lambda\mathbf{I} + \mathbf{\Phi_{k+1}}\mathbf{P}(k)\mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{A_{k+1}}\right) = \mathbf{A_{k+1}}$$

moltiplicando a destra entrambi i membri per $\mathbf{A}^{-1}_{\mathbf{k+1}}$ si ottiene:

$$\mathbf{\Gamma}(k)\left(\lambda\mathbf{A}^{-1}_{\mathbf{k+1}} + \mathbf{\Phi_{k+1}}\mathbf{P}(k)\mathbf{\Phi}^\top_{\mathbf{k+1}}\right) = \mathbf{I}$$

questa si rivela un'identità se si considera l'espressione di $\mathbf{\Gamma}(k)$ (14):

$$\mathbf{\Gamma}(k)\mathbf{\Gamma}^{-1}(k) = \mathbf{I}$$

Di conseguenza, la (17) diventa:

$$\mathbf{P}(k+1)\mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{A_{k+1}} = \mathbf{P}(k)\mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{\Gamma}(k) \tag{18}$$

Inserendo questa e la (16) nell'espressione di $\beta\left(k+1\right)$ che si stava calcolando (15) si ottiene:

$$\begin{aligned}
\beta\left(k+1\right) &= \beta\left(k\right) + \frac{1}{\lambda}\mathbf{P}(k)\mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{A_{k+1}}\mathbf{y_{k+1}} + \\
&\quad -\mathbf{P}(k+1)\mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{A_{k+1}}\mathbf{\Phi_{k+1}}\mathbf{P}(k)\bar{\mathbf{\Phi}}^\top(k)\mathbf{A}(k)\mathbf{y_m}\left(k\right) + \\
&\quad -\frac{1}{\lambda}\mathbf{P}(k+1)\mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{A_{k+1}}\mathbf{\Phi_{k+1}}\mathbf{P}(k)\mathbf{\Phi}^\top_{\mathbf{k+1}}\mathbf{A_{k+1}}\mathbf{y_{k+1}}
\end{aligned}$$

$$
\begin{aligned}
&= \beta(k) + \frac{1}{\lambda}\mathbf{P}(k+1)\mathbf{\Phi}_{\mathbf{k+1}}^{\top}\mathbf{A}_{\mathbf{k+1}}\mathbf{\Gamma}^{-1}(k)\mathbf{A}_{\mathbf{k+1}}\mathbf{y}_{\mathbf{k+1}} + \\
&\quad -\mathbf{P}(k-1)\mathbf{\Phi}_{\mathbf{k+1}}^{\top}\mathbf{A}_{\mathbf{k+1}}\mathbf{\Phi}_{\mathbf{k+1}}\mathbf{P}(k)\Big(\bar{\mathbf{\Phi}}^{\top}(k)\mathbf{A}(k)\mathbf{y}_{\mathbf{m}}(k) + \\
&\quad +\frac{1}{\lambda}\mathbf{\Phi}_{\mathbf{k+1}}^{\top}\mathbf{A}_{\mathbf{k+1}}\mathbf{y}_{\mathbf{k+1}}\Big) \\
&= \beta(k) + \mathbf{P}(k+1)\mathbf{\Phi}_{\mathbf{k+1}}^{\top}\mathbf{A}_{\mathbf{k+1}}\Big(\frac{1}{\lambda}\mathbf{\Gamma}^{-1}(k)\mathbf{A}_{\mathbf{k+1}}\mathbf{y}_{\mathbf{k+1}} + \\
&\quad -\mathbf{\Phi}_{\mathbf{k+1}}\mathbf{P}(k)\Big(\bar{\mathbf{\Phi}}^{\top}(k)\mathbf{A}(k)\mathbf{y}_{\mathbf{m}}(k) + \frac{1}{\lambda}\mathbf{\Phi}_{\mathbf{k+1}}^{\top}\mathbf{A}_{\mathbf{k+1}}\mathbf{y}_{\mathbf{k+1}}\Big)\Big) \\
&= \beta(k) + \mathbf{P}(k+1)\mathbf{\Phi}_{\mathbf{k+1}}^{\top}\mathbf{A}_{\mathbf{k+1}}\Big(\frac{1}{\lambda}\Big(\mathbf{\Gamma}^{-1}(k) + \\
&\quad -\mathbf{\Phi}_{\mathbf{k+1}}\mathbf{P}(k)\mathbf{\Phi}_{\mathbf{k+1}}^{\top}\Big)\mathbf{A}_{\mathbf{k+1}}\mathbf{y}_{\mathbf{k+1}} + \\
&\quad -\mathbf{\Phi}_{\mathbf{k+1}}\beta(k)\Big)
\end{aligned}
$$

Sostituendo l'espressione di $\mathbf{\Gamma}^{-1}(k)$ (14) si vede che l'espressione più interna, tra parentesi, diventa:

$$\mathbf{\Phi}_{\mathbf{k+1}}\mathbf{P}(k)\mathbf{\Phi}_{\mathbf{k+1}}^{\top} + \mathbf{A}_{\mathbf{k+1}}^{-1} - \mathbf{\Phi}_{\mathbf{k+1}}\mathbf{P}(k)\mathbf{\Phi}_{\mathbf{k+1}}^{\top} = \lambda\mathbf{A}_{\mathbf{k+1}}^{-1}$$

Di conseguenza, l'espressione di $\beta(k+1)$ risulta:

$$\beta(k+1) = \beta(k) + \mathbf{P}(k+1)\mathbf{\Phi}_{\mathbf{k+1}}^{\top}\mathbf{A}_{\mathbf{k+1}}\left(\mathbf{y}_{\mathbf{k+1}} - \mathbf{\Phi}_{\mathbf{k+1}}\beta(k)\right) \tag{19}$$

Rispetto al caso *batch*, un segnale di ingresso non eccitante non pregiudica la stabilità numerica di questo criterio. Nel caso in cui un parametro non sia stimabile in base agli ingressi e alle uscite misurate, esso semplicemente non viene aggiornato.

Riferimenti bibliografici

[1] G. Ciccarella, P. Marietti, A. Trifiletti: *Strumentazione e misure elettroniche, seconda edizione*, Casa Editrice Ambrosiana