PSC: Progettazione di sistemi di controllo **a.a. 2010-2011**

Lezione 4 — Ottobre 21, 2010

Docente: Luca Schenato *Stesori: Minucelli, Sterbizzi, Thomaseth*

4.1 Alcuni esempi

Si illustrano in seguito alcuni esempi di applicazione del filtro di Kalman, di cui si riportano le equazioni in forma di informazione:

$$\left.\begin{array}{rcl} \hat{x}_{k+1|k} & = & A\hat{x}_{k|k} \\ P_{k+1|k} & = & AP_{k|k}A^T + Q \end{array}\right\} \text{predizione}$$

$$\left.\begin{array}{rcl} \hat{x}_{k+1|k+1} & = & P_{k+1|k+1}(P_{k+1|k}^{-1}\hat{x}_{k+1|k} + C^T R^{-1} y_{k+1}) \\ P_{k+1|k+1} & = & (P_{k+1|k}^{-1} + C^T R^{-1} C)^{-1} \end{array}\right\} \text{aggiornamento} \tag{4.1}$$

$$\left.\begin{array}{rcl} \hat{x}_{0|-1} & = & \bar{x}_0 \\ P_{0|-1} & = & P_0 \end{array}\right\} \text{inizializzazione}$$

Tutti gli esempi si riferiscono al seguente modello dinamico lineare tempo invariante:

$$\left\{\begin{array}{rcl} x_{k+1} & = & Ax_k + w_k \\ y_k & = & Cx_k + v_k \end{array}\right. \tag{4.2}$$

dove, come di consueto:

$$\left\{\begin{array}{rcl} v_k & \sim & \mathcal{N}(0,R), \quad \mathbb{E}[v_k v_h] = R\delta(k-h) \\ w_k & \sim & \mathcal{N}(0,Q), \quad \mathbb{E}[w_k w_h] = Q\delta(k-h) \\ x_0 & \sim & \mathcal{N}(\bar{x}_0, P_0) \end{array}\right. \tag{4.3}$$

4.1.1 Esempio 1: stima di un parametro scalare

Si vuole stimare un parametro scalare, $\theta \in \mathbb{R}$, mediante una serie di misure rumorose $y_i = \theta + v_i, i = 1 \dots N$. I parametri del modello 4.2 risultano essere:

$$A = C = 1, \quad Q = 0$$

$$R = \sigma^2, \quad P_0 = \infty$$

dal momento che:

$$\left\{\begin{array}{rcl} x_{k+1} & = & x_k = x_0 = \theta \\ y_k & = & \theta + v_k \end{array}\right.$$

Applicando il filtro di Kalman si ricava:

$$\hat{x}_{k|k} = \frac{1}{k}\sum_{i=1}^{k} y_i = \frac{1}{k}y_k + \frac{k-1}{k}\hat{x}_{k-1|k-1}$$

$$P_{k|k} = \frac{\sigma^2}{k}.$$

È evidente che in questo caso lo stimatore di Kalman, dopo aver ricevuto le N misure, coincide con la media campionaria:

$$\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N} y_i.$$

Dato che, inoltre, si ha

$$\hat{x}_{k|k} \sim \mathcal{N}(\theta, \frac{\sigma^2}{k}),$$

l'intervallo di confidenza al 99% risulta essere $\left[\hat{x}_{k|k} - \frac{3\sigma}{\sqrt{k}}, \hat{x}_{k|k} + \frac{3\sigma}{\sqrt{k}}\right]$.

4.1.2 Esempio 2

Nel caso in cui le N misure siano tutte immediatamente disponibili, il modello precedente assume la forma:

$$A = 1, \quad Q = 0$$

$$R = \sigma^2 I_N, \quad P_0 = \infty$$

$$y^0 = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = C\theta + v = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \theta + v$$

Applicando le equazioni del filtro di Kalman si ricava:

$$P_{0|0} = (C^T R^{-1} C)^{-1} = \left([1 \dots 1] \frac{1}{\sigma^2} I \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right)^{-1} = \frac{\sigma^2}{N}$$

$$\hat{\theta} = \hat{x}_{0|0} = \frac{\sigma^2}{N} \left([1 \dots 1] \frac{1}{\sigma^2} I \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \right) = \frac{1}{N}\sum_{i=1}^{N} y_i.$$

Si nota come tali formule coincidano con quelle ricavate nell'esempio precedente, ottenute qui però in forma più generale.

4.1.3 Esempio 3: stima con misure a varianze diverse

Si suppone ora di avere a disposizione due misure rumorose dello stesso parametro:

$$y_1 = \theta + v_1 \qquad v_1 \sim \mathcal{N}(0, \sigma_1^2)$$

$$y_2 = \theta + v_2 \qquad v_2 \sim \mathcal{N}(0, \sigma_2^2)$$

ove $v_1 \perp v_2$. Lo stimatore del parametro assume la forma $\hat{\theta} = a\,y_1 + b\,y_2$, combinazione lineare delle due misure.
Per la correttezza dello stimatore risulta:

$$\mathbb{E}[\hat{\theta}] = \theta = a\,\theta + b\,\theta \quad \Rightarrow a + b = 1 \quad \Rightarrow \hat{\theta} = a\,y_1 + (1-a)\,y_2.$$

La varianza dello stimatore è data dalla formula:

$$\begin{aligned} Var\hat{\theta} &= \mathbb{E}[(\hat{\theta} - \theta)^2] \\ &= \mathbb{E}[(ay_1 + (1-a)y_2 - \theta)^2] \\ &= \mathbb{E}[(a\theta + av_1 + (1-a)\theta + (1-a)v_2 - \theta)^2] \\ &= \mathbb{E}[(av_1 + (1-a)v_2)^2] \\ &= \mathbb{E}[a^2 v_1^2] + \mathbb{E}[(1-a)^2 v_2^2] \\ &= a^2\sigma_1^2 + (1-a)^2\sigma_2^2 \\ &= a^2(\sigma_1^2 + \sigma_2^2) - 2a\sigma_2^2 + \sigma_2^2 \end{aligned}$$

di cui l'andamento in funzione del parametro a è mostrato in Figura 4.1.

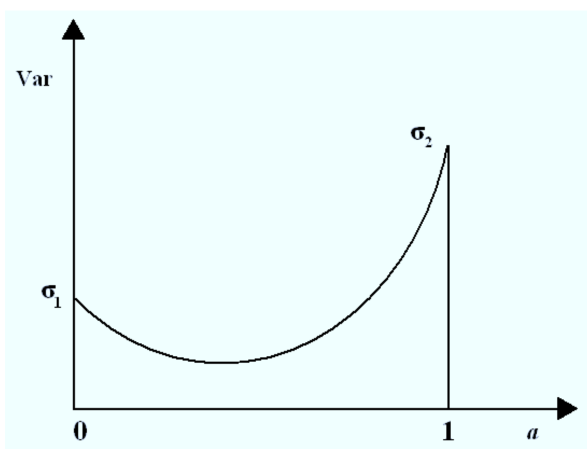

Figura 4.1. Andamento della varianza dello stimatore in funzione del parametro a.

Annullando la derivata rispetto ad a dell'espressione così ottenuta, si ottiene il valore del parametro che minimizza la varianza dello stimatore:

$$\frac{d}{da}(a^2(\sigma_1^2 + \sigma_2^2) - 2a\sigma_2^2 + \sigma_2^2) = 2a(\sigma_1^2 + \sigma_2^2) - 2\sigma_2^2 = 0$$

da cui segue:

$$0 \leq a^* = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \leq 1.$$

Lo stimatore risulta quindi una combinazione lineare convessa delle due misure.
Si nota che se:

- $\sigma_2^2 \gg \sigma_1^2 \Longrightarrow a^* \simeq 1 \Longrightarrow$ viene pesata solo la prima misura
- $\sigma_2^2 = \sigma_1^2 \Longrightarrow a^* = \frac{1}{2} \Longrightarrow$ caso di misure i.i.d.

Lo stimatore corretto a minima varianza è dunque:

$$\hat{\theta} = \frac{\frac{1}{\sigma_1^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} y_1 + \frac{\frac{1}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} y_2.$$

Generalizzando al caso in cui le misure a disposizione siano:

$$y_i = \theta + v_i \qquad v_i \sim \mathcal{N}(0, \sigma_i^2), \quad i = 1 \dots N,$$

con $v_i \perp v_j, i \neq j$, si ha

$$\hat{\theta} = \sum_{i=1}^{N} a_i y_i,$$

sotto le condizioni

$$\sum_{i=1}^{N} a_i = 1, \qquad a_i = \frac{\frac{1}{\sigma_i^2}}{\sum_{l=1}^{N} \frac{1}{\sigma_l^2}}.$$

Si vuole vedere ora come tale risultato si ottenga in modo quasi banale dall'algoritmo di Kalman. Il modello assume la forma:

$$y^0 = C\theta + v$$

dove

$$y^0 = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \quad , \quad C = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \quad , \quad R = \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_N^2 \end{bmatrix}$$

$$A = 1 \quad , \quad Q = 0 \quad , \quad P_0 = \infty\,.$$

Si ottiene:

$$P_{0|0} = \left(\begin{bmatrix} 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1^2} & & \\ & \ddots & \\ & & \frac{1}{\sigma_N^2} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right)^{-1} = \frac{1}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}$$

e dunque:

$$\hat{x}_{0|0} = \frac{\begin{bmatrix} 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1^2} & & \\ & \ddots & \\ & & \frac{1}{\sigma_N^2} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}}{\sum_{i=1}^N \frac{1}{\sigma_i^2}} = \frac{\sum_{i=1}^N \frac{1}{\sigma_i^2} y_i}{\sum_{i=1}^N \frac{1}{\sigma_i^2}} = \sum_{i=1}^N a_i y_i.$$

4.1.4 Esempio 4: stima ai minimi quadrati

Si supponga di avere una serie di dati rumorosi $\{(x_i, y_i)\}_{i=1}^N$, e di volerli interpolare mediante un modello lineare del tipo $y = ax + b$, ove i coefficienti a e b sono incogniti, come mostrato in Figura 4.2.

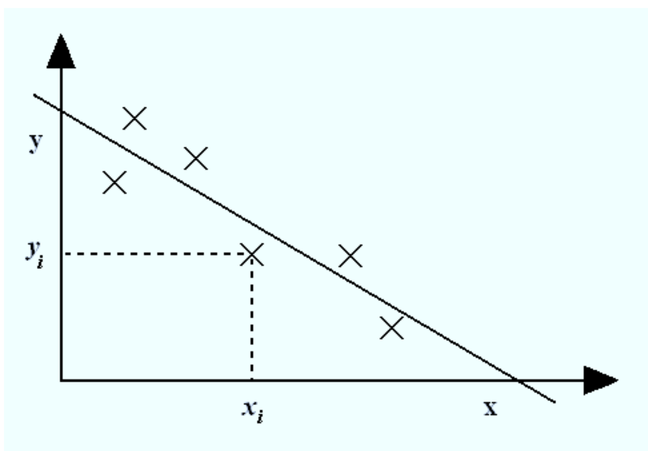

Figura 4.2. Interpolazione lineare di una serie di dati $\{(x_i, y_i)\}_{i=1}^N$.

Si definisce l'errore relativo alla misura i-esima come: $e_i = y_i - (ax_i + b)$.
In forma matriciale, si ha:

$$e = \begin{bmatrix} e_1 \\ \vdots \\ e_N \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} - \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = y - A\theta$$

Un possibile approccio per determinare i coefficienti a e b è quello di minimizzare la somma dei quadrati degli errori delle misure, cioè di trovare:

$$\min_{\theta=(a,b)} \sum_{i=1}^N e_i^2 = \min_{\theta=(a,b)} \|e\|^2 = \min_{\theta=(a,b)} \|y - A\theta\|^2 = \min_{\theta=(a,b)} (y - a\theta)^T (y - A\theta) =$$

$$= \min_{\theta=(a,b)} \|y\|^2 + \theta^T A^T A\theta - 2\theta^T A^T y = \min_{\theta=(a,b)} J(\theta)$$

Si supponga per il momento che la matrice A^TA sia definita positiva, e dunque invertibile. Per determinare il minimo, si annulla la derivata di $J(\theta)$. Tenendo conto delle relazioni:

$$\frac{\partial}{\partial\theta}\theta^T A^T A\theta = 2\theta^T A^T,$$

$$\frac{\partial}{\partial\theta}\theta^T B = B^T,$$

si ottiene facilmente:

$$\frac{\partial}{\partial\theta}J(\theta) = 0 + 2\theta^T AA^T - 2y^T A = 0,$$

da cui:

$$\theta^* = (A^TA)^{-1}A^Ty.$$

Si è dunque determinato il valore ottimo di $\theta = (a, b)$, supponendo che $(A^TA)^{-1}$ esista.
Se la matrice A^TA non fosse invertibile, significa che esiste un vettore $z \neq 0$ tale che $A^TAz = 0$, il che vale se e solo se $Az = 0$, per $z \neq 0$. Ciò significa che la matrice A è singolare, ovvero il suo nucleo non coincide con lo spazio nullo, e si trovano più valori del parametro θ che minimizzano la forma quadratica $\|y - A\theta\|^2$.
Fra tutti i possibili $\theta \in \Theta = \{\theta : \arg\min_\theta \|y - A\theta\|^2\}$, scelgo quello a norma più piccola, cioè $\arg\min_{\theta\in\Theta} \|\theta\|^2$. Si ricava che tale parametro è dato dalla formula:

$$\theta^* = (A^TA)^\dagger A^Ty,$$

che possiede quindi una validità generale, in quanto nel caso A^TA fosse invertibile si ottiene esattamente la formula ricavata prima.
Procedendo, in modo alternativo, con le formule del filtro di Kalman, si possono ottenere anche in questo caso i medesimi risultati. Le ordinate dei dati a disposizione possono essere espresse come:

$$y_k = C_k\theta + v_k \quad k = 1, \ldots, N$$

dove:

$$A = I \quad , \quad Q = 0 \quad , \quad C_k = \begin{bmatrix} x_k & 1 \end{bmatrix}$$
$$v_k \sim \mathcal{N}(0, \sigma^2) \quad , \quad P_0 = \infty \quad , \quad R = \sigma^2 I$$

Il valore del parametro σ^2 è arbitrario in quanto risulterà ininfluente.
Si ottiene dunque il modello:

$$y^0 = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = C\theta + v = \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} + v = A\theta + v$$

da cui

$$P_{0|0} = \left(A^T\frac{1}{\sigma^2}IA\right)^{-1} = \left(A^TA\right)^{-1}\sigma^2$$

e

$$\hat{x}_{0|0} = \hat{\theta} = \left(A^TA\right)^{-1}\sigma^2 A^T\frac{1}{\sigma^2}Iy^0 = \left(A^TA\right)^{-1}A^Ty^0$$

ottenendo lo stesso risultato ricavato precedentemente.