

8.1 Approccio Bayesiana per stima e filtraggio

Si consideri il seguente modello dinamico lineare e tempo invariante:

$$\begin{cases} x_{k+1} = Ax_k + w_k \\ y_k = Cx_k + v_k \end{cases} \quad (8.1)$$

dove, come di consueto:

$$\begin{cases} v_k \sim \mathcal{N}(0, R), & \mathbb{E}[v_k v_h] = R\delta(k-h) \\ w_k \sim \mathcal{N}(0, Q), & \mathbb{E}[w_k w_h] = Q\delta(k-h) \\ x_0 \sim \mathcal{N}(\bar{x}_0, P_0) \end{cases}$$

Si supponga di voler stimare x_k in base alle osservazioni $Y^h = \{y_0, \dots, y_h\}$ fino all'istante h . Si distinguono tre casi:

1. $k > h \implies$ predittore
2. $k = h \implies$ filtro
3. $k < h \implies$ interpolatore

I primi due casi sono stati trattati in precedenza. Rimane quindi da vedere l'interpolatore.

Osservazione: Finora nel contesto della predizione si è sempre parlato di predizione a un passo. Non è però difficile estendere questo concetto alla predizione a 2 (o più) passi. Ricordando le formule per il predittore a 1 passo:

$$\begin{aligned} \hat{x}_{k+1|k} &= A\hat{x}_{k|k} \\ P_{k+1|k} &= AP_{k|k}A^T + Q \end{aligned}$$

e tenendo conto del fatto che per la predizione a 2 passi non si hanno a disposizione ulteriori misure, le espressioni risultano:

$$\begin{aligned} \hat{x}_{k+2|k} &= A\hat{x}_{k+1|k} = A^2\hat{x}_{k|k} \\ P_{k+2|k} &= AP_{k+1|k}A^T + Q = A^2P_{k|k}(A^T)^2 + AQA^T + Q \end{aligned}$$

Si vede quindi che dal secondo passo (in poi) la predizione è basata sul modello (8.1) (si lavora in catena aperta). La generalizzazione alla predizione a n passi è immediata.

□

8.1.1 Approcci all'interpolazione

Gli approcci al calcolo dell'interpolatore di Kalman sono sostanzialmente tre. Si ricordi che si vuole stimare x_k in base alle osservazioni y_0, y_1, \dots, y_T , dove $T > k$.

1. Approccio vettoriale
Costruendo i vettori

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_T \end{bmatrix}, \quad y = \begin{bmatrix} y_0 \\ \vdots \\ y_T \end{bmatrix}, \quad w = \begin{bmatrix} w_0 \\ \vdots \\ w_{T-1} \end{bmatrix}, \quad v = \begin{bmatrix} v_0 \\ \vdots \\ v_T \end{bmatrix}$$

il modello (8.1) può essere riscritto, con ovvio significato dei parametri, in forma compatta:

$$\begin{cases} x = Hx_0 + Lw \\ y = Gx + v = G(Hx_0 + Lw) + v = \bar{G}x_0 + Mw + v \end{cases}$$

Questo approccio, seppur molto intuitivo e matematicamente corretto, non è molto efficiente dal punto di vista computazionale.

2. Approccio di programmazione dinamica

Come evidenziato in Figura 8.1, seguendo questo metodo si fanno due passate con il filtro di Kalman: una in avanti, ottenendo $\hat{x}_{T|T}$, e successivamente una seconda passata all'indietro, usando come misure le stime $\hat{x}_{k|k}$ ricavate alla prima passata, ottenendo $\hat{x}_{k|T}$.

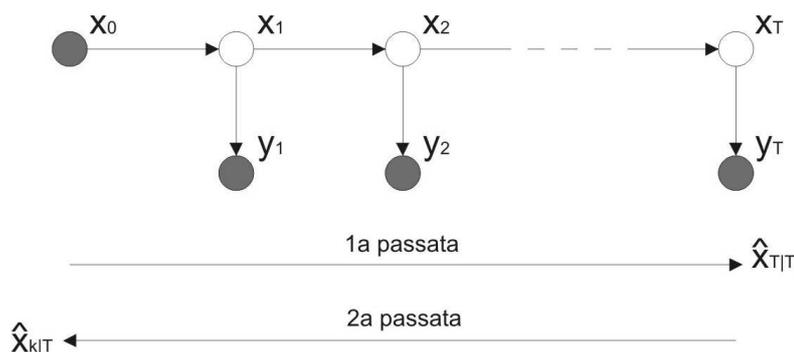


Figura 8.1. Modello markoviano (rappresentazione grafica della dinamica)

3. Approccio forward-backward o parallelo

Invece di fare due passate successive come nell'approccio di programmazione dinamica, si può pensare di far partire il filtro di Kalman all'istante iniziale 0 e contemporaneamente all'istante finale T (facendolo evolvere all'indietro). I due filtri si devono poi incontrare proprio all'istante k , come evidenziato in Figura 8.2, e successivamente i

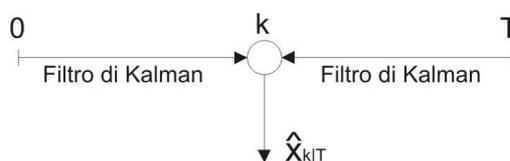


Figura 8.2. Interpolazione forward-backward

risultati ottenuti vanno combinati in modo opportuno. Le espressioni che si ottengono usando questa tecnica sono diverse da quelle ricavate secondo l'approccio 2, ma il risultato è lo stesso.

In questa trattazione verrà sviluppato solo l'approccio forward-backward.

8.1.2 Interpolazione forward-backward

Prima di passare al calcolo dell'interpolatore, si richiamano alcune nozioni preliminari, che torneranno utili più avanti.

Siano x, y due variabili aleatorie gaussiane, con densità di probabilità congiunta data da:

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} \right) \quad (8.2)$$

Valgono allora le seguenti proprietà:

A. Proprietà di condizionamento

Con riferimento alla (8.2), la variabile aleatoria $x|y$ (x condizionata y) ha distribuzione

$$x|y \sim \mathcal{N}(\mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y), \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})$$

B. Proprietà di marginalizzazione

Sempre con riferimento alla (8.2), la variabile aleatoria x ha distribuzione

$$x \sim \mathcal{N}(\mu_x, \Sigma_{xx})$$

C. Si consideri una variabile aleatoria $x \sim \mathcal{N}_x(\mu, \Sigma)$. Allora la sua densità di probabilità è data da:

$$p(x) = c \cdot \exp \left\{ -\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu) \right\}$$

dove la costante di normalizzazione c viene determinata in modo tale da rendere l'integrale di $p(x)$ unitario. Spesso si scrive (omettendo così c) la relazione di proporzionalità:

$$p(x) \propto \exp \left\{ -\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu) \right\}$$

D. L'inversa di una matrice a blocchi è data da:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - CD^{-1}B)^{-1} & * \\ * & * \end{bmatrix}$$

Ora si può procedere alla derivazione delle equazioni dell'interpolatore di Kalman. Per tale scopo si seguirà l'approccio bayesiano.

Si consideri il modello (8.1). Passando alle densità di probabilità, per la prima equazione risulta:

$$x_{k+1} = Ax_k + w_k \implies p(x_{k+1}|x_k) = \mathcal{N}_{x_{k+1}}(Ax_k, Q)$$

dove per il calcolo della varianza è stata sfruttata la relazione:

$$x_{k+1} - \mathbb{E}[x_{k+1}] = x_{k+1} - Ax_k = w_k$$

Analogamente, per la seconda parte del modello si ottiene:

$$y_k = Cx_k + v_k \implies p(y_k|x_k) = \mathcal{N}_{y_k}(Cx_k, R)$$

Così, nel caso particolare di variabili aleatorie gaussiane, le espressioni (8.1) si possono riscrivere nella forma:

$$\begin{cases} p(x_{k+1}|x_k) = \mathcal{N}_{x_{k+1}}(Ax_k, Q) \\ p(y_k|x_k) = \mathcal{N}_{y_k}(Cx_k, R) \end{cases} \quad (8.3)$$

dove la prima equazione rappresenta il modello della dinamica e la seconda equazione il modello della misura.

Se ne deduce che:

$$\begin{aligned} p(x_{k+1}|x_k) &\propto \exp\left\{-\frac{1}{2}(x_{k+1} - Ax_k)^T Q^{-1}(x_{k+1} - Ax_k)\right\} \\ p(y_k|x_k) &\propto \exp\left\{-\frac{1}{2}(y_k - Cx_k)^T R^{-1}(y_k - Cx_k)\right\} \end{aligned}$$

dove è stata applicata la proprietà C e le matrici di varianza R e Q si sono supposte invertibili.

Per ottenere le formule dell'interpolatore è necessario calcolarsi la densità di probabilità $p(x_k|Y^T)$. Prima di fare ciò è opportuno ricavarsi le espressioni del filtro di Kalman (in avanti), ovvero $p(x_k|Y^k)$, usando l'approccio bayesiano.

N.B.: Per la gaussianità del rumore e la linearità del modello (con riferimento a (8.1)) vale:

$$p(x_k|Y^h) = \mathcal{N}_{x_k}(\hat{x}_{k|h}, P_{k|h}) \quad (8.4)$$

□

Il calcolo del filtro di Kalman viene fatto in maniera iterativa; la domanda che ci si pone è come dedurre $p(x_k|Y^k)$ da $p(x_{k-1}|Y^{k-1})$, sfruttando opportunamente le equazioni (8.3), ovvero le espressioni per la dinamica e per la misura.

Nel seguito verranno utilizzate ripetutamente le seguenti proprietà:

- Formula di Bayes

$$p(x, y) = p(x|y)p(y) \quad (8.5)$$

- Marginalizzazione

$$p(x) = \int_y p(x, y) dy = \int_y p(x|y) p(y) dy \quad (8.6)$$

(la seconda uguaglianza segue dalla formula di Bayes)

Si noti che le espressioni che verranno dedotte sono valide anche nel caso più generale di densità di probabilità non gaussiane; l'unico vincolo che si impone è la markovianità. Nel caso gaussiano si vedrà che le relazioni che si ottengono coincidono proprio con quelle del filtro di Kalman (in forma di informazione).

Si scrive quindi:

$$\begin{aligned} p(x_k|Y^k) &= p(x_k|y_k, Y^{k-1}) \\ &= \frac{p(y_k, x_k|Y^{k-1})}{p(y_k|Y^{k-1})} \\ p(x_k|Y^k) &= \frac{p(y_k|x_k, Y^{k-1}) p(x_k|Y^{k-1})}{p(y_k|Y^{k-1})} \end{aligned}$$

dove la regola di Bayes è stata applicata due volte, prima a $p(x_k|y_k, Y^{k-1})$ (usandola al rovescio), poi a $p(y_k, x_k|Y^{k-1})$. Si noti che va sempre tenuto conto anche del condizionamento rispetto a Y^{k-1} .

Ora, essendo $p(y_k|Y^{k-1})$ una costante quando sono note le misure, il tutto può essere riscritto come una relazione di proporzionalità; inoltre, data la markovianità del processo, si ha $p(y_k|x_k, Y^{k-1}) = p(y_k|x_k)$, e perciò risulta:

$$\begin{aligned} p(x_k|Y^k) &\propto p(y_k|x_k) p(x_k|Y^{k-1}) \\ &\propto p(y_k|x_k) \int_{x_{k-1}} p(x_k, x_{k-1}|Y^{k-1}) dx_{k-1} \\ &\propto p(y_k|x_k) \int_{x_{k-1}} p(x_k|x_{k-1}, Y^{k-1}) p(x_{k-1}|Y^{k-1}) dx_{k-1} \\ p(x_k|Y^k) &\propto p(y_k|x_k) \int_{x_{k-1}} p(x_k|x_{k-1}) p(x_{k-1}|Y^{k-1}) dx_{k-1} \end{aligned} \quad (8.7)$$

dove sono state applicate, nell'ordine, la proprietà di marginalizzazione e la regola di Bayes, e infine è stata sfruttata ancora la markovianità del processo, scrivendo $p(x_k|x_{k-1}, Y^{k-1}) =$

$p(x_k|x_{k-1})$. Si noti che fino a questo punto non si è ancora sfruttata la gaussianità, ma solo la markovianità del processo; la (8.7) è quindi di validità generale.

Ora si aggiunge l'ipotesi di gaussianità: ricordando le identità (8.4) e (8.3), l'equazione (8.7) si presenta nella forma:

$$\mathcal{N}_{x_k}(\hat{x}_{k|k}, P_{k|k}) \propto \mathcal{N}_{y_k}(Cx_k, R) \int_{x_{k-1}} \mathcal{N}_{x_k}(Ax_{k-1}, Q) \mathcal{N}_{x_{k-1}}(\hat{x}_{k-1|k-1}, P_{k-1|k-1}) dx_{k-1} \quad (8.8)$$

Dalla seconda equazione del modello (8.3) e dal primo passaggio svolto nella deduzione della (8.7) si osserva che l'integrale delle due densità di probabilità gaussiane nella (8.8) corrisponde alla formula di predizione ad un passo.

Sostituendo le densità di probabilità (gaussiane) nella (8.8), risulta:

$$\begin{aligned} \exp \left\{ -\frac{1}{2}(x_k - \hat{x}_{k|k})^T P_{k|k}^{-1}(x_k - \hat{x}_{k|k}) \right\} &\propto \exp \left\{ -\frac{1}{2}(y_k - Cx_k)^T R^{-1}(y_k - Cx_k) \right\} \cdot \\ &\cdot \int_{x_{k-1}} \left(\exp \left\{ -\frac{1}{2}(x_k - Ax_{k-1})^T Q^{-1}(x_k - Ax_{k-1}) \right\} \cdot \right. \\ &\left. \exp \left\{ -\frac{1}{2}(x_{k-1} - \hat{x}_{k-1|k-1})^T P_{k-1|k-1}^{-1}(x_{k-1} - \hat{x}_{k-1|k-1}) \right\} \right) dx_{k-1} \end{aligned}$$

Il membro a destra deve perciò essere posto nella forma $\exp \left\{ -\frac{1}{2}(x_k - *)^T * (x_k - *) \right\}$, potendosi ricavare così per confronto con il termine a sinistra le espressioni per $\hat{x}_{k|k}$ e $P_{k|k}$.

L'integrale, usando la regola di Bayes, con $(x, y) \longrightarrow (x_k, x_{k-1})$, e ricordando la (8.7), risulta:

$$\int_{x_{k-1}} p(x_k|x_{k-1}) p(x_{k-1}|Y^{k-1}) dx_{k-1} = \int_{x_{k-1}} p(x_k, x_{k-1}|Y^{k-1}) dx_{k-1}$$

Per la gaussianità, il termine a destra del segno di uguaglianza può quindi essere riscritto come:

$$\int_{x_{k-1}} \exp \left\{ -\frac{1}{2} \underbrace{\begin{bmatrix} x_k - \mu_1 \\ x_{k-1} - \mu_2 \end{bmatrix}^T \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} x_k - \mu_1 \\ x_{k-1} - \mu_2 \end{bmatrix}}_{V(x_k, x_{k-1})} \right\} dx_{k-1}$$

Si vede facilmente che questo integrale, per la proprietà di marginalizzazione, è pari a:

$$\exp \left\{ -\frac{1}{2}(x_k - \mu_1)^T \Sigma_{11}^{-1}(x_k - \mu_1) \right\}$$

e perciò è necessario trovare μ_1 e Σ_{11} . A tale scopo si sviluppa l'esponenziale \otimes , ponendo, per semplificare la notazione:

$$\Gamma = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1}$$

Vale allora:

$$V(x_k, x_{k-1}) = \begin{bmatrix} x_k - \mu_1 \\ x_{k-1} - \mu_2 \end{bmatrix}^T \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix} \begin{bmatrix} x_k - \mu_1 \\ x_{k-1} - \mu_2 \end{bmatrix}$$

che scritto per esteso diventa:

$$\begin{aligned} V(x_k, x_{k-1}) &= (x_k - \mu_1)^T \Gamma_{11} (x_k - \mu_1) + (x_k - \mu_1)^T \Gamma_{12} (x_{k-1} - \mu_2) + \dots \\ &\quad \dots + (x_{k-1} - \mu_2)^T \Gamma_{21} (x_k - \mu_1) + (x_{k-1} - \mu_2)^T \Gamma_{22} (x_{k-1} - \mu_2) \\ &= x_k^T \Gamma_{11} x_k - 2\mu_1^T \Gamma_{11} x_k - 2(x_{k-1} - \mu_2)^T \Sigma_{21} x_k + \dots \\ &= x_k^T \Gamma_{11} x_k - 2(\mu_1^T \Gamma_{11} - \mu_2^T \Sigma_{21}) x_k - 2x_{k-1}^T \Sigma_{21} x_k + \dots \end{aligned}$$

dove i puntini nella seconda e nella terza riga indicano termini indipendenti da x_k e x_{k+1} . Rimangono quindi da calcolare le varie Σ_{ij} per arrivare alla conclusione. Per questo si vedano gli appunti della Lezione 9.