

1.1 Variabili aleatorie e nozioni preliminari

Le definizioni di questa sezione si trovano in qualsiasi testo di processi stocastici lineari ed in particolare nel Capitolo 2 in [1].

Definizione 1.1. Una matrice $\Sigma \in \mathbb{R}^{n \times n}$ si dice semidefinita positiva se $\Sigma = \Sigma^T$ e $x^T \Sigma x \geq 0, \forall x \in \mathbb{R}^n$, e viene indicata con $\Sigma \geq 0$. Una matrice semidefinita positiva si dice positiva se $\Sigma > 0$.

Definizione 1.2. La traccia $\text{tr} : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ di una matrice $A \in \mathbb{R}^{n \times n}$ e definita come $\text{tr}(A) = \sum_{i=1}^n A_{ii}$ dove A_{ij} e' elemento $i - j$ della matrice A . La traccia gode delle seguenti proprieta':

1. $\text{tr}(ABC) = \text{tr}(BCA) = \text{tr}(CAB)$, dove le matrici A, B, C hanno dimensione opportuna
2. Sia $x \in \mathbb{R}^n$ e $Q \geq 0$, allora $\|x\|_Q^2 = x^T Q x = \text{tr}(x^T Q x) = \text{tr}(Q x x^T)$
3. Se $Q \geq 0$ e $P_1 \geq P_2$, allora $\text{tr}(Q P_1) \geq \text{tr}(Q P_2)$
4. $P_1 \geq P_2 \implies \text{tr}(P_1) \geq \text{tr}(P_2)$, ma non viceversa.

Definizione 1.3. Una variabile aleatoria si dice gaussiana ed e' indicata con $x \sim \mathcal{N}(\mu, \Sigma)$, dove $\mu \in \mathbb{R}^n$ e $\Sigma \in \mathbb{R}^{n \times n}, \Sigma > 0$, se la sua densita' di probabilita' e' data da $p(x) = \frac{1}{\sqrt{2^n \pi^n \det \Sigma}} e^{-(x-\mu)^T \Sigma^{-1} (x-\mu)}$. E' facile verificare che $\mathbb{E}[x] = \mu$ e $\text{Var}(x) = \mathbb{E}[(x - \mathbb{E}[x])(x - \mathbb{E}[x])^T] = \Sigma$. La condizione $\Sigma > 0$ puo essere rilassata a $\Sigma \geq 0$ tramite l'introduzione delle distribuzioni di Dirac.

Definizione 1.4. Sia $A \in \mathbb{R}^{n \times m}$. Si definisce pseudoinversa della matrice A , l'unica matrice $A^\dagger \in \mathbb{R}^{m \times n}$ che soddisfa le seguenti proprieta':

$$A^\dagger A A^\dagger = A^\dagger, \quad A A^\dagger A = A, \quad (A A^\dagger)^T = A A^\dagger, \quad (A^\dagger A)^T = A^\dagger A$$

Dalla definizione segue che

1. La pseudoinversa della matrice di soli zeri $0_{nm} \in 0^{n \times m}$, e' data da $0_{nm}^\dagger = 0_{mn}$.
2. Se la matrice $A \in \mathbb{R}^{n \times n}$ e' invertibile allora $A^\dagger = A^{-1}$

3. Data la matrice $A = \begin{bmatrix} A_{11} & 0_{nm} \\ 0_{lr} & 0_{lm} \end{bmatrix}$ la sua pseudoinversa e' data da $A^\dagger = \begin{bmatrix} A_{11}^\dagger & 0_{rl} \\ 0_{mn} & 0_{ml} \end{bmatrix}$.

Elenchiamo ora una serie di semplici proprieta' che risulteranno molto utili per derivare in maniera chiara e semplice il filtro di Kalman:

Proposizione 1.1. Siano x, y due variabili aleatorie (non necessariamente gaussiane) con densita' di probabilita' $p(x, y)$. Sia $g(x, y)$ una generica funzione misurabile. Allora

$$\mathbb{E}_{xy}[g(x, y)] = \mathbb{E}_y[\mathbb{E}_{x|y}[g(x, y)|y]]$$

Dimostrazione: Con un piccolo abuso di notazione usiamo \mathbb{E}_x per indicare chiaramente rispetto a quale variabile aleatoria e' calcolata l'aspettazione. In genere tale pedice non viene indicato ed e' sottointeso. Infatti

$$\begin{aligned} \mathbb{E}[g(x, y)] &= \mathbb{E}_{xy}[g(x, y)] = \int_x \int_y g(x, y)p(x, y)dx dy = \int_x \int_y g(x, y)p(x|y)p(y)dx dy \\ &= \int_y \left(\int_x g(x, y)p(x|y)dx \right) p(y)dy = \int_y \mathbb{E}_{x|y}[g(x, y)|y]p(y)dy = \mathbb{E}_y[\mathbb{E}_{x|y}[g(x, y)|y]] \end{aligned}$$

□

Proposizione 1.2. Siano x, y due variabili congiuntamente gaussiane, cioe':

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

1. La combinazione lineare di variabili gaussiane e' ancora una variabile aleatoria gaussiana. Infatti se $z = Ax + By$ dove A, B sono matrici di dimensioni opportune, allora $z \sim \mathcal{N}(\mu_z, \Sigma_z)$, dove

$$\mu_z = A\mu_x + B\mu_y, \quad \Sigma_z = [A \ B] \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} \begin{bmatrix} A^T \\ B^T \end{bmatrix}$$

2. Il condizionamento di variabili aleatorie gaussiane e' ancora una variabile aleatoria gaussiana. Infatti $p(x|y) = \mathcal{N}(\mu_{x|y}, \Sigma_{x|y})$. Se $\Sigma_{yy} > 0$

$$\mu_{x|y} = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y), \quad \Sigma_{x|y} = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}$$

Nel caso sia semplicemente semidefinita positiva $\Sigma_{yy} \geq 0$, la media e la varianza sono date da:

$$\mu_{x|y} = \mu_x + \Sigma_{xy}\Sigma_{yy}^\dagger(y - \mu_y), \quad \Sigma_{x|y} = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^\dagger\Sigma_{yx}$$

dove Σ_{yy}^\dagger e' la pseudoinversa di Σ_{yy} .

Dimostrazione: Vedere Teorema 2.3 e 2.6 in [1].

□

Passiamo ora ad enunciare uno dei piu' importanti teoremi che riguardano la stima di variabili aleatorie.

Teorema 1.1. *Siano $x \in \mathbb{R}^n, y \in \mathbb{R}^m$ due variabili aleatorie (non necessariamente gaussiane), e la funzione $g : \mathbb{R}^m \rightarrow \mathbb{R}^n$ una qualsiasi funzione misurabile. Definiamo $\tilde{x}_g = g(y)$ un generico stimatore di x , e sia $e_g = x - \tilde{x}_g$ il corrispondente errore di stima, che a loro volta sono variabili aleatorie. Lo stimatore $\hat{x} = \hat{g}(y) = \mathbb{E}[x|y]$ e' definito **stimatore a minima varianza** o **stimatore ottimo in media quadratica**, poiche' possiede la seguente proprieta':*

$$\mathbb{E}[(x - \hat{x})(x - \hat{x})^T] \leq \mathbb{E}[(x - \tilde{x}_g)(x - \tilde{x}_g)^T], \forall g()$$

La precedente proprieta' implica che

$$\mathbb{E}[||x - \hat{x}||_Q^2] \leq \mathbb{E}[||x - \tilde{x}_g||_Q^2], \forall g(), \forall Q \geq 0.$$

Se indichiamo con $e = x - \hat{x}$ l'errore dello stimatore ottimo, si ha che

$$\mathbb{E}[eg(y)^T] = 0, \forall g()$$

cioe' lo stimatore e' incorrelato con il le "misure" y . Infine lo stimatore ottimo e' corretto (unbiased) in quanto

$$\mathbb{E}[\hat{x}] = \mathbb{E}[x]$$

Dimostrazione: Questo teorema e' simile al Teorema 2.2 in [1]. Definiamo per semplicita' $\Delta g(y) = \hat{g}(y) - g(y)$.

$$\begin{aligned} \mathbb{E}[(x - \tilde{x}_g)(x - \tilde{x}_g)^T] &= \mathbb{E}_{xy}[(x - \tilde{x}_g)(x - \tilde{x}_g)^T] = \mathbb{E}_{xy}[(x - g(y))(x - g(y))^T] = \\ &= \mathbb{E}_{xy}[(x - g(y) + \hat{g}(y) - \hat{g}(y))(x - g(y) + \hat{g}(y) - \hat{g}(y))^T] \\ &= \mathbb{E}_y[\mathbb{E}_{x|y}[(x - g(y) + \hat{g}(y) - \hat{g}(y))(x - g(y) + \hat{g}(y) - \hat{g}(y))^T | y]] \\ &= \mathbb{E}_y[\mathbb{E}_{x|y}[(x - \hat{g}(y))(x - \hat{g}(y))^T + (x - \hat{g}(y))\Delta g(y)^T + \Delta g(y)(x - \hat{g}(y))^T + \Delta g(y)\Delta g(y)^T | y]] \\ &= \mathbb{E}_y[\mathbb{E}_{x|y}[(x - \hat{g}(y))(x - \hat{g}(y))^T | y] + (\mathbb{E}_{x|y}[x|y] - \hat{g}(y))\Delta g(y)^T + \Delta g(y)(\mathbb{E}_{x|y}[x|y] - \hat{g}(y))^T + \\ &\quad + \Delta g(y)\Delta g(y)^T] \\ &= \mathbb{E}[(x - \hat{x})(x - \hat{x})^T] + \mathbb{E}[\Delta g(y)\Delta g(y)^T] = \mathbb{E}[(x - \hat{x})(x - \hat{x})^T] + M \geq \mathbb{E}[(x - \hat{x})(x - \hat{x})^T] \end{aligned}$$

dove abbiamo usato il fatto che $\mathbb{E}_{x|y}[x|y] = \hat{g}(y)$ e $M \geq 0, \forall g()$. Questo dimostra la prima disuguaglianza del teorema. La seconda disuguaglianza segue utilizzando le proprieta' della traccia indicata nella Definizione 1.2 come segue

$$\begin{aligned} \mathbb{E}[(x - \hat{x})(x - \hat{x})^T] \leq \mathbb{E}[(x - \tilde{x}_g)(x - \tilde{x}_g)^T] &\implies \mathbb{E}[Q(x - \hat{x})(x - \hat{x})^T] \leq \mathbb{E}[Q(x - \tilde{x}_g)(x - \tilde{x}_g)^T] \\ \implies (\text{tr}(\mathbb{E}[Q(x - \hat{x})(x - \hat{x})^T]) &\leq (\text{tr}(\mathbb{E}[Q(x - \tilde{x}_g)(x - \tilde{x}_g)^T]) \implies \mathbb{E}[||x - \hat{x}||_Q^2] \leq \mathbb{E}[||x - \tilde{x}_g||_Q^2] \end{aligned}$$

Passiamo ora a dimostrare l'incorrelazione tra lo stimatore e una qualsiasi funzione dei dati:

$$\begin{aligned} \mathbb{E}[eg(y)^T] &= \mathbb{E}_{xy}[(x - \hat{x})g(y)^T] = \mathbb{E}_y[\mathbb{E}_{x|y}[(x - \hat{x})g(y)^T | y]] = \\ &= \mathbb{E}_y[(\mathbb{E}_{x|y}[x|y] - \hat{x})g(y)^T] = \mathbb{E}_y[(\hat{x} - \hat{x})g(y)^T] = 0 \end{aligned}$$

Infine dimostriamo che lo stimatore ottimo e' corretto (unbiased):

$$\mathbb{E}[\hat{x}] = \mathbb{E}_y[\hat{x}] = \mathbb{E}_y[\mathbb{E}_{x|y}[x|y]] = \mathbb{E}_{xy}[x] = \mathbb{E}_x[x] = \mathbb{E}[x]$$

□

Questo teorema dimostra che lo stimatore ottimo in media quadratica coincide con l'aspettazione condizionata della variabile incognita di interesse x rispetto ai dati misurati y . In genere lo stimatore ottimo e' una funzione non-lineare dei dati e quindi di poca utilita' pratica. Tuttavia, nel caso di processi gaussiani, lo stimatore ottimo si puo' calcolare in modo esplicito e risulta essere una funzione lineare¹ dei dati come indicato nella Proposizione 1.2-2 in quanto $\hat{x} = \mu_{x|y}$. Il filtro di Kalman non e' altro che un modo efficiente per calcolare l'aspettazione condizionata di una variabile aleatoria gaussiana condizionata rispetto ad una serie di misure.

1.2 Processi lineari stocastici e stimatori ottimi

Consideriamo ora un sistema lineare stocastico a tempo discreto del tipo:

$$x_{k+1} = Ax_k + w_k \tag{1.1}$$

$$y_k = Cx_k + v_k \tag{1.2}$$

dove $x_k \in \mathbb{R}^n, y_k \in \mathbb{R}^m, A \in \mathbb{R}^{n \times n}, C \in \mathbb{R}^{m \times n}$, e $w_k \sim \mathcal{N}(0, Q), v_k \sim \mathcal{N}(0, R)$ e $x_0 \sim \mathcal{N}(\bar{x}_0, P_0)$ sono variabili aleatorie gaussiane incorrelate² ed identicamente distribuite (i.i.d), dove $Q \geq 0, R \geq 0, P_0 \geq 0$. Per semplificare la notazione abbiamo indicato $x_k = x(k)$ dove $k = 0, 1, 2, \dots$ sono gli istanti temporali. Definiamo anche le seguenti quantita':

$$\hat{x}_{k|h} = \mathbb{E}[x_k | y_0, \dots, y_h] \tag{1.3}$$

$$P_{k|h} = \mathbb{E}[(x_k - \hat{x}_{k|h})(x_k - \hat{x}_{k|h})^T | y_0, \dots, y_h] \tag{1.4}$$

Lo stimatore $\hat{x}_{k|h}$ viene definito come fitro di kalman. Se $k = h + N$ allora $\hat{x}_{k+N|k}$ viene indicato come predittore a N passi, mentre se $k = h - N$ allora $\hat{x}_{k+N|k}$ viene indicato come interpolatore (smoother) a N passi. La matrice semidefinita positiva $P_{k|h} \geq 0$ corrisponde alla varianza dell'errore di stima.

Prima di derivare le usuali equazioni del filtro di Kalman e' interessante notare come il realta' i risultati descritti nella sezione precedente siano sufficienti per ottenere lo stimatore ottimo dato le misure fino all'istante k . A questo scopo definiamo le nuove variabili aleatorie $\mathbf{x}_k = (x_k, x_{k-1}, \dots, x_1) \in \mathbb{R}^{kn}, \mathbf{y}_k = (y_k, y_{k-1}, \dots, y_1, y_0) \in \mathbb{R}^{(k+1)m}, \mathbf{v}_k =$

¹Da un punto di vista formale la funzione e' *affine* anche se e' uso comune usare impropriamente il termine *lineare*.

²E' possibile generalizzare i risultati che seguono per variabili aleatorie correlate, ma per semplicita' non verranno considerate.

$(v_k, v_{k-1}, \dots, v_1, v_0) \in \mathbb{R}^{(k+1)m}$ e $\mathbf{w}_k = (w_{k-1}, w_{k-2}, \dots, w_0) \in \mathbb{R}^{kn}$ per poter riscrivere la dinamica del sistema stocastico in maniera compatta come segue:

$$\mathbf{x}_k = \begin{pmatrix} I & A & A^2 & \dots & A^{k-1} \\ 0 & I & A & \dots & A^{k-2} \\ 0 & 0 & I & \dots & A^{k-3} \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \dots & I \end{pmatrix} \mathbf{w}_k + \begin{pmatrix} A^k \\ A^{k-1} \\ \vdots \\ A^2 \\ A \end{pmatrix} x_0 = F\mathbf{w}_k + Gx_0 \quad (1.5)$$

$$\mathbf{y}_k = \begin{pmatrix} C & CA & CA^2 & \dots & CA^k \\ 0 & C & CA & \dots & CA^{k-1} \\ 0 & 0 & C & \dots & CA^{k-2} \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \dots & C \end{pmatrix} \mathbf{w}_k + \begin{pmatrix} CA^k \\ CA^{k-1} \\ \vdots \\ CA^2 \\ CA \end{pmatrix} x_0 + \mathbf{v}_k = H\mathbf{w}_k + Lx_0 + \mathbf{v}_k \quad (1.6)$$

Si vede chiaramente che $\mathbf{x}_k, \mathbf{y}_k$ sono variabili aleatorie gaussiane essendo combinazioni lineari di variabili aleatorie gaussiane. Da un punto di vista teorico e' quindi facile ottenere lo stimatore ottimo \hat{x}_k calcolando medie e varianze congiunte delle variabili $\mathbf{x}_k, \mathbf{y}_k$ come indicato nella Proposizione 1.2-1 e poi calcolando la stima condizionata $\mathbb{E}[\mathbf{x}_k | \mathbf{y}_k]$ come indicato nella Proposizione 1.2-2. Lo stimatore ottimo \hat{x}_k e' dato dai primi n elementi del vettore $\mathbb{E}[\mathbf{x}_k | \mathbf{y}_k]$. Si noti come implicitamente si sia calcolato anche l'interpolazione (smoothing) degli stati precedenti $\hat{x}_{h|k}$ per $h = 1, \dots, k-1$. Sebbene molto intuitivo e matematicamente compatto, il calcolo dello stimatore ottimo necessitano l'inversione di matrici estremamente elevate, anzi che crescono al crescere di k . Tuttavia come si puo' notare che le precedenti equazioni hanno una struttura particolare. Quella piu' evidente e' che alcune matrici sono triangolari e quindi possiedono una struttura che puo' essere utilizzata per un calcolo numerico piu' efficiente della semplice inversione. Inoltre nel caso tempo-invariante le matrici sono Toeplitz e questo permette di ottenere molti risultati che riguardano la prestazione a regime, cioe' per $k \rightarrow \infty$. Le equazioni del filtro di Kalman sono infatti un algoritmo ricorsivo numericamente efficiente che appunto sfrutta la particolare struttura del problema.

Bibliography

- [1] Giorgio Picci. *Fitraggio Statistico (Wiener, Levinson, Kalman) e Applicazioni*. Libreria Progetto, 2006.