

Capitolo 8

Filtraggio non lineare

per gli studenti del corso di
Stima e identificazione

Luigi Chisci, 10 Maggio 2020

8.1 Problema di stima dello stato

Si consideri un generico sistema dinamico a tempo-discreto:

$$\begin{cases} x_{k+1} = f_k(x_k, w_k) \\ y_k = h_k(x_k, v_k) \end{cases} \quad (8.1.1)$$

con

x_k : stato del sistema;

y_k : uscita osservata, che costituisce una misura (generalmente indiretta) dello stato del sistema;

$w_k \sim p_{w_k}(\cdot)$: disturbo di processo (o di sistema), che solitamente modella i fattori ambientali che interferiscono con il sistema;

$v_k \sim p_{v_k}(\cdot)$: disturbo (o rumore) di misura, che tiene conto delle imprecisioni del sensore.

Si assume, inoltre, che i disturbi w_k e v_k siano bianchi, fra loro incorrelati ed incorrelati con lo stato iniziale x_0 .

L'obiettivo è quello di determinare, ad ogni istante, una stima \hat{x}_k dello stato x_k del sistema. Si noti che è sufficiente che una tra le funzioni $f_k(\cdot, \cdot)$ ed $h_k(\cdot, \cdot)$ sia non lineare perché il sistema sia considerato non lineare.

Sia z una generica variabile, di seguito si indicherà con z_k il valore assunto da z al tempo k , e con $z^k = \{z_0, z_1, \dots, z_k\}$ la sequenza dei valori assunti da z dall'istante iniziale all'istante k . Nel seguito si indicherà con $\hat{x}_{t|k}$ la stima dello stato x_t basata sulle osservazioni y^k

Il procedimento di stima consta di due fasi:

- 1) **Correzione** (detta anche *filtraggio* o *aggiornamento alla misura*): partendo dalla stima $\hat{x}_{k|k-1}$ e data la nuova misura, $y_k = h_k(x_k, v_k)$, si determina la stima aggiornata $\hat{x}_{k|k}$ detta anche *stima corretta* o *filtrata*.
- 2) **Predizione** (o *aggiornamento temporale*): partendo da $\hat{x}_{k|k}$ ed utilizzando il modello di stato $x_{k+1} = f_k(x_k, w_k)$ si determina la *stima predetta* o *predizione* $\hat{x}_{k+1|k}$.

In ambito meteorologico ed oceanografico la correzione è chiamata *analisi* mentre la predizione è chiamata *forecast* o *simulazione*. In tali ambiti, l'intero processo di stima viene chiamato *assimilazione dati*.

8.2 Stima ricorsiva Bayesiana

Secondo l'approccio Bayesiano [1], il problema di stima viene interpretato come determinazione della densità di probabilità (*PDF: Probability Density Function*) della quantità da stimare condizionatamente alle osservazioni. Con riferimento alla stima dello stato del sistema (8.1.1) si definiscono le seguenti PDF:

- $p_{k|k}(x) = f(x_k | y^k)$ *PDF filtrata o corretta*
- $p_{k+1|k}(x) = f(x_{k+1} | y^k)$ *PDF predetta.*

Nota la PDF, per determinare la stima si possono poi utilizzare diversi criteri, quali ad esempio il criterio del minimo errore quadratico medio (*MMSE: Minimum Mean Square Error*) oppure il criterio della massima probabilità a posteriori (*MAP: Max A Posteriori Probability*), così definiti:

- criterio MMSE: $\hat{x}_{k|k} = E[x_k | y^k] = \int x p_{k|k}(x) dx$
- criterio MAP: $\hat{x}_{k|k} = \arg \max_x p_{k|k}(x)$

e la corrispondente varianza sarà:

$$P_{k|k} = \int (x - \hat{x}_{k|k})(x - \hat{x}_{k|k})^\top p_{k|k}(x) dx.$$

Quindi, partendo dalla PDF iniziale $p_{1|0}(x)$, la stima Bayesiana consta delle seguenti due fasi:

- **Correzione:** $p_{k|k-1}(x) \xrightarrow[\text{eq. di misura}]{y_k = h_k(x_k, v_k)} p_{k|k}(x)$
- **Predizione:** $p_{k|k}(x) \xrightarrow[\text{modello di stato}]{x_{k+1} = f_k(x_k, w_k)} p_{k+1|k}(x).$

Nel prossimo paragrafo si determinano le equazioni ricorsive per le suddette PDF.

8.2.1 Correzione

Data $p_{k|k-1}(x) = f(x_k | y^{k-1})$ si vuole determinare $p_{k|k}(x) = f(x_k | y^k)$ utilizzando la misura $y_k = h_k(x_k, v_k)$. A tale proposito si utilizza la proprietà:

$$f(a | b, c) = \frac{f(b | a, c) f(a | c)}{f(b | c)}$$

che deriva da $f(a | b, c) f(b | c) = f(a, b | c) = f(b | a, c) f(a | c)$.

Ponendo $a = x_k$, $b = y_k$ e $c = y^{k-1}$, si ha

$$\underbrace{f(x_k | y_k, y^{k-1})}_{p_{k|k}(x)} = \frac{f(y_k | x_k, y^{k-1}) \overbrace{f(x_k | y^{k-1})}^{p_{k|k-1}(x)}}{f(y_k | y^{k-1})}$$

Si noti che:

- $f(x_k | y_k, y^{k-1}) = p_{k|k}(x)$
- $f(x_k | y^{k-1}) = p_{k|k-1}(x)$
- $f(y_k | x_k, y^{k-1}) = f(y_k | x_k)$, detta *verosimiglianza (likelihood)*.

Infatti, il condizionamento in $f(y_k | x_k, y^{k-1})$ rispetto a y^{k-1} è irrilevante dato che la stessa è condizionata anche rispetto a x_k . Quindi, si introduce la verosimiglianza indicata con

$$\ell_k(x, y) = f(y_k = y | x_k = x).$$

Si ha

$$p_{k|k}(x) \propto \ell_k(x, y_k) \cdot p_{k|k-1}(x).$$

Quanto più x è verosimile con l'osservazione y_k , tanto maggiore sarà $p_{k|k}(x)$. Il prodotto $\ell_k(\cdot) \cdot p_{k|k-1}(\cdot)$ sposta la PDF risultante, $p_{k|k}(\cdot)$, verso gli stati più verosimili, vedi figura 8.1. Risulta

$$p_{k|k}(x) = \frac{\ell_k(x, y_k) \cdot p_{k|k-1}(x)}{c}$$

dove c è una costante tale che

$$\int \frac{\ell_k(\xi, y_k) \cdot p_{k|k-1}(\xi)}{c} d\xi = 1 \quad \implies \quad c = \int \ell_k(\xi, y_k) \cdot p_{k|k-1}(\xi) d\xi.$$

Più $\ell_k(\cdot, y_k)$ è stretta, cioè migliore è la qualità del sensore (R piccola), più $\ell_k(\cdot, y_k)$ è informativa.

L'equazione ricorsiva risultante per la correzione, detta *equazione di Bayes*, è quindi:

$$p_{k|k}(x) = \frac{\ell_k(x, y_k) \cdot p_{k|k-1}(x)}{\int \ell_k(\xi, y_k) \cdot p_{k|k-1}(\xi) d\xi}. \quad (8.2.1)$$

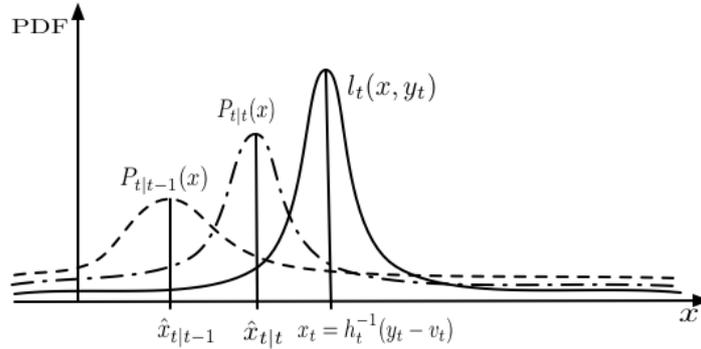


Figura 8.1: Effetto della verosimiglianza nella fase di correzione.

8.2.2 Predizione

Data $p_{k|k}(x) = f(x_k | y^k)$, utilizzando l'equazione di transizione dello stato $x_{k+1} = f_k(x_k, w_k)$, si vuole determinare $p_{k+1|k}(x) = f(x_{k+1} | y^k)$. A tale proposito, si determina la PDF congiunta di x_k ed x_{k+1} e quindi la si marginalizza rispetto a x_k :

$$p_{k+1|k}(x) = f(x_{k+1} = x | y^k) = \int f(x_{k+1}, x_k | y^k) dx_k.$$

Sfruttando la proprietà: $f(a | c) = \int f(a, b | c) db = \int f(a | b, c) f(b | c) db$, con $a = x_{k+1}$, $b = x_k$ e $c = y^k$, si ha

$$\underbrace{f(x_{k+1} | y^k)}_{p_{k+1|k}(x_{k+1})} = \int f(x_{k+1} | x_k, y^k) \underbrace{f(x_k | y^k)}_{p_{k|k}(x_k)} dx_k$$

Considerando $f(x_{k+1} | x_k, y^k)$, il condizionamento rispetto a x_k rende irrilevante il condizionamento rispetto a y^k , cioè

$$f(x_{k+1} | x_k, y^k) = f(x_{k+1} | x_k)$$

da cui, posto $x = x_{k+1}$ e $\xi = x_k$, si ha

$$p_{k+1|k}(x) = \int \underbrace{f(x_{k+1} = x | x_k = \xi)}_{\text{PDF di transizione}} \underbrace{f(x_k = \xi | y^k)}_{p_{k|k}(\xi)} d\xi.$$

La PDF condizionata $f(x_{k+1} | x_k)$ descrive, in modo probabilistico, la transizione dello stato dall'istante k all'istante $k + 1$ e viene per questo detta *PDF di transizione* ed indicata con la notazione

$$\varphi_{k+1|k}(x | \xi) \triangleq f(x_{k+1} = x | x_k = \xi). \quad (8.2.2)$$

Si noti che la PDF di transizione, $\varphi_{k+1|k}(\cdot|\cdot)$, dipende unicamente da:

- modello di evoluzione dello stato $x_{k+1} = f_k(x_k, w_k)$;
- PDF del rumore di processo $p_{w_k}(\cdot)$.

L'equazione ricorsiva risultante per la predizione, detta *equazione di Chapman-Kolmogorov*, è quindi:

$$p_{k+1|k}(x) = \int \varphi_{k+1|k}(x|\xi) p_{k|k}(\xi) d\xi. \quad (8.2.3)$$

8.2.3 Algoritmo di stima Bayesiana

Nell'effettuare la stima i passi di predizione e correzione possono essere scambiati a seconda delle esigenze. L'Algoritmo 1, di seguito riportato, riassume un ciclo temporale del filtro Bayesiano nel caso in cui la stima ricorsiva viene effettuata prima con il passo di correzione e poi con quello di predizione.

Algorithm 1: Algoritmo di stima Bayesiana.

Dati: PDF a priori $p_{1|0}(\cdot)$

for $k = 1, 2, \dots$ **do**

Correzione:

se è disponibile la misura y_k :

$$p_{k|k}(x) = \frac{\ell_k(x, y_k) \cdot p_{k|k-1}(x)}{\int \ell_k(\xi, y_k) \cdot p_{k|k-1}(\xi) d\xi}$$

altrimenti:

$$p_{k|k}(x) = p_{k|k-1}(x)$$

Predizione:

$$p_{k+1|k}(x) = \int \varphi_{k+1|k}(x|\xi) p_{k|k}(\xi) d\xi.$$

end for

8.3 PDF di transizione e verosimiglianza

La determinazione della verosimiglianza $\ell_k(x, y)$ e della PDF di transizione $\varphi_{k+1|k}(x|\xi)$ è generalmente complessa. Se si assume, tuttavia, che le funzioni $h_k(x, v)$ e $f_k(x, w)$ siano invertibili rispetto al secondo argomento (v e, rispettivamente, w), le PDF $\ell_k(\cdot, \cdot)$ e $\varphi_{k+1|k}(\cdot|\cdot)$ possono essere determinate

analiticamente come di seguito specificato. Indicata con $h_k^{-1}(\cdot, \cdot)$ la funzione inversa tale che $v = h_k^{-1}(x, y)$, la funzione di verosimiglianza è data da:

$$\ell_k(x, y) = p_{v_k}(h_k^{-1}(x, y)) \left\{ \left| \det \frac{\partial}{\partial v} [h_k(x, v)] \right| \right\}^{-1} \Bigg|_{v=h_k^{-1}(x, y)}. \quad (8.3.1)$$

Analogamente, se $x = f_k(\xi, w)$ è invertibile rispetto a w , ovvero esiste la funzione inversa $f_k^{-1}(\cdot, \cdot)$ tale che $w = f_k^{-1}(\xi, x)$, allora la PDF di transizione dello stato è data da:

$$\varphi_{k+1|k}(x|\xi) = p_{w_k}(f_k^{-1}(\xi, x)) \left\{ \left| \det \frac{\partial}{\partial w} [f_k(x, w)] \right| \right\}^{-1} \Bigg|_{w=f_k^{-1}(\xi, x)}. \quad (8.3.2)$$

Caso di disturbi additivi

Se inoltre si assume, come è realistico supporre in molti contesti pratici, che i disturbi siano *additivi*, cioè che il sistema dinamico a tempo-discreto sia della forma:

$$\begin{cases} x_{k+1} = f_k(x_k) + w_k \\ y_k = h_k(x_k) + v_k \end{cases} \quad (8.3.3)$$

si possono ricavare facilmente i disturbi conoscendo le altre variabili, cioè:

$$\begin{aligned} w_k &= f_k^{-1}(x_k, x_{k+1}) = x_{k+1} - f_k(x_k) \\ v_k &= h_k^{-1}(x_k, y_k) = y_k - h_k(x_k). \end{aligned} \quad (8.3.4)$$

Si possono così ricavare la PDF di transizione dello stato

$$\varphi_{k+1|k}(x|\xi) = p_{w_k}(x - f_k(\xi)) \quad (8.3.5)$$

e la verosimiglianza

$$\ell_k(x, y) = p_{v_k}(y - h_k(x)). \quad (8.3.6)$$

Caso di disturbi additivi Gaussiani

Se si fa l'ulteriore ipotesi che i disturbi w_k e v_k siano Gaussiani, cioè

$$\begin{aligned} p_{w_k}(w) &= \mathcal{N}(w; 0, Q_k) = \frac{1}{\sqrt{\det(2\pi Q_k)}} \exp\left(-\frac{1}{2}w^\top Q_k^{-1}w\right) \\ p_{v_k}(v) &= \mathcal{N}(v; 0, R_k) = \frac{1}{\sqrt{\det(2\pi R_k)}} \exp\left(-\frac{1}{2}v^\top R_k^{-1}v\right) \end{aligned} \quad (8.3.7)$$

si possono ricavare le seguenti espressioni analitiche per la PDF di transizione dello stato e per la verosimiglianza

$$\begin{aligned}\varphi_{k+1|k}(x|\xi) &= \frac{1}{\sqrt{\det(2\pi Q_k)}} \exp\left(-\frac{1}{2}(x - f_k(\xi))^\top Q_k^{-1}(x - f_k(\xi))\right) \\ \ell_k(x, y) &= \frac{1}{\sqrt{\det(2\pi R_k)}} \exp\left(-\frac{1}{2}(y - h_k(x))^\top R_k^{-1}(y - h_k(x))\right)\end{aligned}\tag{8.3.8}$$

In generale, la soluzione della ricorsione Bayesiana è *infinito-dimensionale*, cioè non è esprimibile con un numero finito di parametri, per cui è impossibile da determinare analiticamente. Vi sono però alcune eccezioni, di seguito elencate.

1. Caso lineare Gaussiano, caratterizzato da funzioni $f_k(x, w)$ e $h_k(x, v)$ lineari nonché da PDF $p_{1|0}(\cdot)$, $p_{w_k}(\cdot)$ e $p_{v_k}(\cdot)$ Gaussiane.
2. Caso lineare con rumore di misura uniforme in assenza di rumore di processo [2].
3. Caso in cui lo stato x_k è discreto e finito, i.e. $x_k \in \{x^1, x^2, \dots, x^\ell\}$ con $\ell < \infty$.
4. Alcune non linearità speciali, cioè funzioni $f_k(\cdot, \cdot)$ e $h_k(\cdot, \cdot)$ particolari, e.g. filtro di Benes [3] e filtro di Daum [4] che lo generalizza.

8.4 Caso lineare Gaussiano: filtro di Kalman

Si consideri un sistema dinamico a tempo-discreto *lineare*, eventualmente tempo-variante

$$\begin{cases} x_{k+1} = A_k x_k + b_k + D_k w_k \\ y_k = C_k x_k + v_k \end{cases}\tag{8.4.1}$$

che deriva dal sistema (8.1.1), nel caso in cui sia $f_k(\cdot, \cdot)$ che $h_k(\cdot, \cdot)$ siano lineari. Se

- $p_{1|0}(x) \sim \mathcal{N}(x; \hat{x}_{1|0}, P_{1|0})$
- w_k, v_k sono congiuntamente Gaussiani ed incorrelati, i.e. $\begin{bmatrix} w_k \\ v_k \end{bmatrix} \sim \mathcal{N}\left(\cdot; 0, \begin{bmatrix} Q_k & 0 \\ 0 & R_k \end{bmatrix}\right)$

allora, per ogni $k \geq 1$, $p_{k|k}(x)$ e $p_{k|k-1}(x)$ risultano Gaussiane, vale a dire

- $p_{k|k-1}(x) \sim \mathcal{N}(x; \hat{x}_{k|k-1}, P_{k|k-1})$
- $p_{k|k}(x) \sim \mathcal{N}(x; \hat{x}_{k|k}, P_{k|k})$

con medie e covarianze $\hat{x}_{k|k-1}$, $\hat{x}_{k|k}$, $P_{k|k-1}$, $P_{k|k}$, che caratterizzano completamente tali PDF fornite dal filtro di Kalman (KF, *Kalman Filter*). KF è soluzione ottima nel senso MMSE della stima Bayesiana nel caso di un sistema dinamico *lineare Gaussiano*. Le equazioni ricorsive del filtro di Kalman sono riportate in dettaglio nell'Algoritmo 2.

Algorithm 2: Filtro di Kalman.

Dati: $\hat{x}_{1|0}$ e $P_{1|0}$; se non si hanno informazioni si può porre $\hat{x}_{1|0} = 0$ e $P_{1|0} = \alpha I_n$ con $\alpha \gg 1$ e I_n matrice identità di ordine n

for $k = 1, 2, \dots$ **do**

Correzione:

si acquisisce la misura y_k

$e_k = y_k - C_k \hat{x}_{k|k-1}$; %errore di stima dell'uscita o *innovazione*

$S_k = R_k + C_k P_{k|k-1} C_k^\top$; %covarianza del processo delle innovazioni

$L_k = P_{k|k-1} C_k^\top S_k^{-1}$; %guadagno di Kalman

$\hat{x}_{k|k} = \hat{x}_{k|k-1} + L_k e_k$

$P_{k|k} = P_{k|k-1} - L_k S_k L_k^\top$

$u_k = g(\hat{x}_{k|k}, \dots)$; %in un sistema di controllo qui viene calcolato u_k

Predizione:

$\hat{x}_{k+1|k} = A_k \hat{x}_{k|k} + b_k$

$P_{k+1|k} = A_k P_{k|k} A_k^\top + D_k Q_k D_k^\top$

end for

Di seguito si elencano alcune delle proprietà significative del filtro di Kalman.

- Il filtro di Kalman è **non-polarizzato**, cioè $E[\hat{x}_{k|k}] = E[x_k]$.
- **Stabilità:** sotto opportune ipotesi di osservabilità, la covarianza dell'errore di stima resta limitata anche se il sistema è instabile, cioè le successioni di matrici $P_{k|k}$ e $P_{k+1|k}$ sono limitate.
- **Ottimalità:** nel caso lineare Gaussiano, il filtro di Kalman è ottimo sia nel senso MMSE che MAP. Nel caso semplicemente lineare (non

Gaussiano) il filtro di Kalman è, comunque, lo stimatore ottimo MMSE tra tutti i filtri lineari, anche se possono esistere filtri non lineari migliori.

- Il filtro di Kalman è **ricorsivo** nel senso che, all'istante k , è sufficiente conoscere $\hat{x}_{k|k-1}$ e $P_{k|k-1}$, e non i valori precedentemente assunti da queste variabili, per poter proseguire nell'elaborazione.
- Se $\dim(x) = n$, il filtro di Kalman richiede un impiego di memoria $O(n^2)$ ed una complessità computazionale $O(n^3)$ per ogni ciclo temporale.

Il vero parametro di sintonizzazione del filtro di Kalman è la covarianza del rumore di processo Q_k . Anche se talvolta non si hanno informazioni su Q_k , assegnandogli un valore opportuno è possibile quantificare l'incertezza sul modello.

La matrice Q ha anche un'interpretazione frequenziale. Anzitutto si osserva che le prestazioni del filtro non dipendono singolarmente dal valore assoluto delle componenti di Q bensì dal loro valore relativo rispetto alle componenti della matrice di covarianza del rumore di misura R . In altri termini, riscalandolo della stessa quantità positiva le matrici Q_k ed R_k non produce alcun effetto sulla successione delle stime $\hat{x}_{k|k}$ e $\hat{x}_{k+1|k}$ ma solo una analogha riscalfatura delle matrici di covarianza $P_{k|k}$ e $P_{k+1|k}$. Tenendo conto che Q_k ed R_k hanno, in generale, dimensioni diverse si può comunque affermare che:

- se $Q \gg R$ (nel senso che le componenti diagonali di Q sono molto maggiori delle componenti diagonali di R) si ha un *filtro a banda elevata*, ovvero che risponde rapidamente alle misure
- se $Q \ll R$ (nel senso che le componenti diagonali di Q sono molto minori delle componenti diagonali di R), si ha un *filtro a banda stretta*, ovvero che risponde lentamente alle misure.

8.5 Stima dello stato di un sistema non lineare

Si consideri un sistema dinamico non lineare a tempo-discreto, eventualmente tempo-variante, con disturbi *additivi*

$$\begin{cases} x_{k+1} = f_k(x_k) + w_k \\ y_k = h_k(x_k) + v_k \end{cases} \quad (8.5.1)$$

Nel caso di sistemi non lineari, una soluzione analitica del problema di filtraggio Bayesiano non è disponibile ma esistono diversi metodi approssimati (subottimi) di stima, tra i quali i *metodi di linearizzazione*, i *metodi di*

campionamento della PDF ed i *metodi a modelli multipli/somma di Gaussiane*. Storicamente la prima estensione del filtro di Kalman al caso non lineare è il *filtro di Kalman esteso* (EKF = *Extended Kalman Filter*), ideato e sviluppato presso la NASA per le missioni Apollo [5]. EKF linearizza, ad ogni istante temporale, il sistema in un intorno della stima disponibile a quell'istante ed applica il filtro di Kalman standard al sistema linearizzato. Si tratta di un metodo semplice che, tuttavia, non sempre fornisce prestazioni soddisfacenti in quanto si basa sulla linearizzazione nell'intorno di una stima che potrebbe essere lontana dallo stato effettivo. I metodi di campionamento della PDF approssimano la PDF continua con una discreta. L'approssimazione può essere effettuata in modo deterministico come nel caso del *filtro di Kalman unscented* (UKF = *Unscented Kalman Filter*) oppure in modo stocastico come nel *filtro a particelle* (PF = *Particle Filter*), detto anche *Filtro Sequenziale Monte Carlo* (SMCF = *Sequential Monte Carlo Filter*). Infine i metodi a *modelli multipli* o a *somma/mistura di Gaussiane* approssimano un sistema non lineare con più modelli lineari, a ciascuno dei quali corrisponde una componente Gaussiana. Ne risulta una PDF che è una combinazione lineare convessa (mistura) di Gaussiane. Di seguito, si esamineranno in maggior dettaglio i metodi di linearizzazione e quelli di campionamento della PDF.

8.5.1 Filtro di Kalman esteso

In particolare, il metodo EKF prevede lo sviluppo in serie di Taylor al primo ordine delle funzioni $f_k(\cdot)$ e $h_k(\cdot)$. La funzione di transizione ad un passo dello stato $f_k(\cdot)$ viene usata per fare predizione quando è nota la stima $\hat{x}_{k|k}$, per cui viene sviluppata in un intorno di $\hat{x}_{k|k}$ come segue

$$x_{k+1} = f_k(\hat{x}_{k|k}) + \frac{\partial f_k}{\partial x}(\hat{x}_{k|k}) \cdot \underbrace{(x_k - \hat{x}_{k|k})}_{\hat{x}_{k|k}} + \underbrace{\Delta f_k(\tilde{x}_{k|k})}_{\text{resto}} + w_k.$$

Viceversa, la funzione di misura $h_k(\cdot)$ viene usata per fare correzione quando è nota la stima $\hat{x}_{k|k-1}$; pertanto viene sviluppata in un intorno di $\hat{x}_{k|k-1}$, cioè

$$y_k = h_k(\hat{x}_{k|k-1}) + \frac{\partial h_k}{\partial x}(\hat{x}_{k|k-1}) \cdot \tilde{x}_{k|k-1} + \underbrace{\Delta h_k(\tilde{x}_{k|k-1})}_{\text{resto}} + v_k.$$

Si devono quindi calcolare i Jacobiani:

$$A_k = \frac{\partial f_k}{\partial x}(\hat{x}_{k|k})$$

prima della predizione e

$$C_k = \frac{\partial h_k}{\partial x}(\hat{x}_{k|k-1})$$

prima della correzione. Si noti che le matrici A_k e C_k risultano tempo-varianti anche se il sistema non lineare originario è tempo-invariante poiché gli Jacobiani devono essere ricalcolati ad ogni istante k . Di seguito si esaminano in dettaglio le fasi di correzione e predizione di EKF.

Correzione: Nella correzione della stima

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + L_k (y_k - \hat{y}_{k|k-1})$$

si approssima $\hat{y}_{k|k-1}$ ponendo $\hat{y}_{k|k-1} \cong h_k(\hat{x}_{k|k-1})$, da cui

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + L_k (y_k - h_k(\hat{x}_{k|k-1})).$$

Si noti che l'innovazione utilizzata per l'aggiornamento della stima è

$$e_k = y_k - h_k(\hat{x}_{k|k-1}) = C_k \tilde{x}_{k|k-1} + \Delta h_k(\tilde{x}_{k|k-1}) + v_k \cong C_k \tilde{x}_{k|k-1} + v_k$$

da cui si deduce che il guadagno di correzione L_k e la matrice di covarianza a posteriori $P_{k|k}$ possono essere ottenute, seppur in modo approssimato, come nel filtro di Kalman standard utilizzando la matrice C_k calcolata come Jacobiano della funzione di misura, cioè

$$\begin{aligned} S_k &= R_k + C_k P_{k|k-1} C_k^\top \\ L_k &= P_{k|k-1} C_k^\top S_k^{-1} \\ P_{k|k} &= P_{k|k-1} - L_k S_k L_k^\top. \end{aligned}$$

Predizione: Nella predizione della stima, si approssima $\hat{x}_{k+1|k}$ ponendo

$$\hat{x}_{k+1|k} \cong f_k(\hat{x}_{k|k}).$$

Per l'aggiornamento temporale della matrice di covarianza $P_{k+1|k} = E[\tilde{x}_{k+1|k} \tilde{x}_{k+1|k}^\top]$, si fa uso della linearizzazione

$$\tilde{x}_{k+1|k} = x_{k+1} - \hat{x}_{k+1|k} = A_k \tilde{x}_{k|k} + \Delta f_k(\tilde{x}_{k|k}) + w_k \cong A_k \tilde{x}_{k|k} + w_k$$

da cui si deduce la stessa equazione di aggiornamento del filtro di Kalman standard

$$P_{k+1|k} = A_k P_{k|k} A_k^\top + Q_k$$

in cui la matrice A_k è calcolata come Jacobiano della funzione di transizione dello stato.

EKF è riassunto nell'Algoritmo 3.

Algorithm 3: Algoritmo EKF nella forma di covarianza.

Dati: $\hat{x}_{1|0}$ e $P_{1|0}$; se non si hanno informazioni si può porre $\hat{x}_{1|0} = 0$ e $P_{1|0} = \alpha I_n$ con $\alpha \gg 1$ e I_n matrice identità di ordine n

Calcolo preliminare dei Jacobiani:

$$A_k = \frac{\partial f_k}{\partial x}(\cdot); \quad C_k = \frac{\partial h_k}{\partial x}(\cdot)$$

for $k = 1, 2, \dots$ **do**

Correzione:

si acquisisce la misura y_k

$$C_k = \frac{\partial h_k}{\partial x}(\hat{x}_{k|k-1})$$

$e_k = y_k - h_k(\hat{x}_{k|k-1});$ %innovazione

$S_k = R_k + C_k P_k C_k^\top;$ %covarianza del processo delle innovazioni

$L_k = P_k C_k^\top S_k^{-1};$ %guadagno di Kalman

$\hat{x}_{k|k} = \hat{x}_{k|k-1} + L_k e_k$

$P_{k|k} = P_{k|k-1} - L_k S_k L_k^\top$

$u_k = g(\hat{x}_{k|k}, \dots);$ %in un sistema di controllo qui viene calcolato u_k

Predizione:

$$A_k = \frac{\partial f_k}{\partial x}(\hat{x}_{k|k})$$

$\hat{x}_{k+1|k} = f_k(\hat{x}_{k|k})$

$P_{k+1|k} = A_k P_{k|k} A_k^\top + Q_k$

end for

Esistono risultati di stabilità locale relativi a EKF [6], anche se questo algoritmo fornisce solitamente prestazioni inferiori ai metodi di campionamento della PDF e ai filtri a modelli multipli/somma di Gaussiane.

8.5.2 Filtro di Kalman unscented

L'Unscented Kalman Filter è un metodo di campionamento deterministico che utilizza la cosiddetta *trasformazione unscented* (UT= Unscented Transform), una tecnica efficace per propagare i momenti di una variabile aleatoria (v.a.) attraverso una trasformazione non lineare. Data una v.a. $X \sim (\bar{x}, P_x)$, di cui sono note media e varianza, se ne vuole approssimare la distribuzione con un numero finito di campioni, i *punti-sigma* x_i , la cui media e varianza campionaria coincidano esattamente con \bar{x} e, rispettivamente,

P_x . Si dimostra che a tale proposito sono sufficienti $n + 1$ campioni, dove n è la dimensione di X , anche se, per motivi di simmetria, se ne utilizzano solitamente $2n + 1$ indicati con x_i per $i = -n, \dots, 0, \dots, n$, con relativi pesi w_i . L'uguaglianza di media e covarianza campionarie dei punti-sigma con media e covarianza della v.a. X impone le relazioni:

$$\bar{x} = \sum_{i=-n}^n w_i x_i \quad (8.5.2)$$

$$P_x = \sum_{i=-n}^n w_i (x_i - \bar{x})(x_i - \bar{x})^\top \quad (8.5.3)$$

Considerando una generica trasformazione non lineare $Y = g(X)$ della v.a. X , ovvero

$$X \sim (\bar{x}, P_x) \xrightarrow{g(\cdot)} Y = g(X) \sim (\bar{y}, P_y)$$

si possono determinare i campioni trasformati

$$y_i = g(x_i) \quad i = -n, \dots, 0, \dots, n$$

da cui si possono ricavare, in modo approssimato, i primi due momenti, media e varianza, della distribuzione di Y tramite:

$$\bar{y} \cong \sum_{i=-n}^n w_i y_i \quad (8.5.4)$$

$$P_y \cong \sum_{i=-n}^n w_i (y_i - \bar{y})(y_i - \bar{y})^\top \quad (8.5.5)$$

Si può dimostrare che se la distribuzione discreta $\{x_i\}_{i=0}^{2n}$ fornisce esattamente i primi due momenti, media e varianza, di X , allora la corrispondente distribuzione discreta $\{y_i\}_{i=0}^{2n}$ approssima correttamente almeno al secondo ordine media e varianza di Y . Si noti che, viceversa, EKF approssima tali momenti solo al primo ordine.

Generazione dei campioni

Data la v.a. $X \sim (\bar{x}, P_x)$, in primo luogo si fattorizza la sua matrice di covarianza come

$$P_x = \Gamma \Gamma^\top = \sum_{i=1}^n \gamma_i \gamma_i^\top \quad (8.5.6)$$

dove $\Gamma = [\gamma_1, \gamma_2, \dots, \gamma_n]$. Si può ottenere questa fattorizzazione in svariati modi, quali ad esempio:

- la fattorizzazione di Cholesky in cui Γ coincide con una matrice triangolare, inferiore o superiore;
- la decomposizione a valori singolari (autovalori-autovettori) $P_x = T\Lambda T^\top$, con T matrice ortogonale degli autovettori e Λ matrice diagonale degli autovalori, che fornisce $\Gamma = T\Lambda^{1/2}$;
- la radice quadrata simmetrica $\Gamma = \sqrt{P_x}$.

Confrontando (8.5.5) con (8.5.6) si ha:

$$\sum_{i=-n}^n w_i (x_i - \bar{x})(x_i - \bar{x})^\top = \sum_{i=1}^n \gamma_i \gamma_i^\top. \quad (8.5.7)$$

Una possibile scelta dei punti-sigma per imporre (8.5.7) è quindi

$$x_i - \bar{x} = \begin{cases} \alpha_i \gamma_i, & i > 0 \\ 0, & i = 0 \\ -\alpha_i \gamma_i, & i < 0 \end{cases} \implies x_i = \begin{cases} \bar{x} + \alpha_i \gamma_i, & i > 0 \\ \bar{x}, & i = 0 \\ \bar{x} - \alpha_i \gamma_i, & i < 0. \end{cases} \quad (8.5.8)$$

Sostituendo (8.5.8) in (8.5.7) e ponendo $w_i = w_{-i}$, si ha

$$\sum_{i=1}^n 2w_i \alpha_i^2 \gamma_i \gamma_i^\top = \sum_{i=1}^n \gamma_i \gamma_i^\top$$

da cui si deduce la relazione $2w_i \alpha_i^2 = 1$ e quindi

$$\alpha_i = \frac{1}{\sqrt{2w_i}} \quad i = 1, 2, \dots, n. \quad (8.5.9)$$

Sostituendo (8.5.8) in (8.5.4), l'uguaglianza della media comporta la relazione

$$\sum_{i=-n}^n w_i = 1. \quad (8.5.10)$$

È prassi comune scegliere tutti i pesi (eccetto eventualmente w_0) uguali, cioè

$$w_{\pm i} = w \quad i = 1, 2, \dots, n$$

per cui la condizione (8.5.10) diventa $w_0 + 2nw = 1$. Pertanto si può scegliere il peso centrale w_0 e, conseguentemente, ricavare i pesi non centrali mediante $w = (1 - w_0)/2n$. Alternativamente si possono scegliere i pesi non centrali w e determinare conseguentemente $w_0 = 1 - 2nw$. In letteratura [7], sono proposte scelte del peso centrale w_0 che permettono di influenzare i momenti

di ordine superiore della distribuzione dei punti-sigma senza alterarne la media e la covarianza. Una buona scelta, in situazioni in cui non si ha nessuna conoscenza a-priori sul tipo di distribuzione della variabile da stimare, è:

$$w_0 = 0, w = \frac{1}{2n} \implies \alpha_i = \alpha = \sqrt{n} \quad i = 1, \dots, n$$

È anche buona norma, per garantire consistenza cioè che la covarianza calcolata non sottostimi quella effettiva, utilizzare pesi w_i^c diversi per il calcolo della covarianza; una scelta consigliata in tal senso è

$$w_0^c = 2, w_i^c = \frac{1}{2n} \quad i = \pm 1, \dots, \pm n$$

Riassumendo i precedenti sviluppi, la trasformazione unscented (UT) viene riportata nell'Algoritmo 4. Si noti che per il suo utilizzo nella stima è spesso richiesto anche il calcolo della cross-covarianza P_{xy} fra le v.a. X e $Y = g(X)$.

Algorithm 4: Algoritmo UT $[\bar{y}, P_y, P_{xy}] = UT(\bar{x}, P_x, g(\cdot))$

Dati: \bar{x} , P_x e funzione $g(\cdot)$

1) Fattorizzazione $P_x = \Gamma \Gamma^\top$ con $\Gamma = [\gamma_1, \dots, \gamma_n]$

2) Calcolo dei pesi:

$$w_0 = 0$$

$$w_0^c = 2$$

$$w_i = w_i^c = \frac{1}{2n} \quad i = \pm 1, \dots, \pm n$$

3) Calcolo dei punti-sigma:

$$x_0 = \bar{x}, \alpha = \sqrt{n}$$

for $i = 1, \dots, n$ **do**

$$\begin{array}{|l} x_i = \bar{x} + \alpha \gamma_i \\ x_{-i} = \bar{x} - \alpha \gamma_i \end{array}$$

end for

4) Calcolo di \bar{y}, P_y, P_{xy} :

for $i = -n, \dots, 0, \dots, n$ **do**

$$| \quad y_i = g(x_i)$$

end for

$$\bar{y} = \sum_{i=-n}^n w_i y_i$$

$$P_y = \sum_{i=-n}^n w_i^c (\bar{y} - y_i)(\bar{y} - y_i)^\top$$

$$P_{xy} = \sum_{i=-n}^n w_i^c (\bar{x} - x_i)(\bar{y} - y_i)^\top.$$

Algorithm 5: Algoritmo UKF.

Dati: $\hat{x}_{1|0}, P_{1|0}$

for $k = 1, 2, \dots$ **do**

Correzione: $(\hat{x}_{k|k-1}, P_{k|k-1}) \xrightarrow{y_k=z_k+v_k, z_k=h_k(x_k)} (\hat{x}_{k|k}, P_{k|k})$

$\begin{bmatrix} \hat{y}_{k|k-1}, P_{k|k-1}^z, P_{k|k-1}^{xy} \end{bmatrix} = \text{UT} (\hat{x}_{k|k-1}, P_{k|k-1}, h_k(\cdot))$

$S_k = P_{k|k-1}^z + R_k$

$L_k = P_{k|k-1}^{xy} S_k^{-1}$

Si acquisisce la misura y_k

$\hat{x}_{k|k} = \hat{x}_{k|k-1} + L_k (y_k - \hat{y}_{k|k-1})$

$P_{k|k} = P_{k|k-1} - L_k S_k L_k^\top$

Predizione: $(\hat{x}_{k|k}, P_{k|k}) \xrightarrow{x_{k+1}=s_{k+1}+w_k, s_{k+1}=f_k(x_k)} (\hat{x}_{k+1|k}, P_{k+1|k})$

$\begin{bmatrix} \hat{x}_{k+1|k}, P_{k+1|k}^s, \cdot \end{bmatrix} = \text{UT} (\hat{x}_{k|k}, P_{k|k}, f_k(\cdot))$

$P_{k+1|k} = P_{k+1|k}^s + Q_k$.

end for

Nell'UKF (vedi algoritmo 5), invece di approssimare al 1°ordine le funzioni $f_k(\cdot)$ e $h_k(\cdot)$ come nell'EKF (mediante linearizzazione in un intorno della stima), si approssimano i momenti della PDF $p_{k|k}(\cdot)$, la stima $\hat{x}_{k|k}$ e la covarianza $P_{k|k}$, tramite i punti-sigma.

8.5.3 Filtro a particelle

L'aumento della potenza di calcolo degli elaboratori ha reso sempre più praticabili i metodi di campionamento che approssimano la distribuzione continua con una distribuzione discreta costituita da un numero molto elevato, N , di campioni, i.e.:

$$p(x) \cong \sum_{i=1}^N w_i \delta(x - x_i) \quad (8.5.11)$$

dove: $\delta(x - x_i)$ è la delta di Dirac centrata sul campione x_i ; $w_i \geq 0$ è il *peso di importanza* relativo al campione x_i con $\sum_{i=1}^N w_i = 1$; la coppia (w_i, x_i) prende il nome di *particella*. Da (8.5.11) si approssimano i momenti della

distribuzione, cioè la media

$$\bar{x} \cong \int x \sum_{i=0}^N w_i \delta(x - x_i) dx = \sum_{i=0}^N w_i \underbrace{\int x \delta(x - x_i) dx}_{x_i} = \sum_{i=0}^N w_i x_i,$$

e la covarianza

$$\begin{aligned} P_x &\cong \int \sum_{i=0}^N w_i \delta(x - x_i) (x - \bar{x}) (x - \bar{x})^\top dx \\ &= \sum_{i=0}^N w_i \underbrace{\int (x - \bar{x}) (x - \bar{x})^\top \delta(x - x_i) dx}_{(x_i - \bar{x}) (x_i - \bar{x})^\top} \\ &= \sum_{i=0}^N w_i (x_i - \bar{x}) (x_i - \bar{x})^\top. \end{aligned}$$

Con riferimento al problema di filtraggio, si assume che, all'istante t , la PDF $p_{k|k-1}(\cdot)$ è parametrizzata dall'insieme delle particelle $\{(w_{k|k-1}^i, x_{k|k-1}^i)\}_{i=1}^N$ nel seguente modo

$$p_{k|k-1}(x) = \sum_{i=1}^N w_{k|k-1}^i \delta(x - x_{k|k-1}^i).$$

Applicando la correzione Bayesiana, si ha

$$\begin{aligned} p_{k|k}(x) &= \frac{\ell_k(x, y_k) p_{k|k-1}(x)}{\int \ell_k(\xi, y_k) p_{k|k-1}(\xi) d\xi} \\ &= \frac{\ell_k(x, y_k) \sum_{i=1}^N w_{k|k-1}^i \delta(x - x_{k|k-1}^i)}{\int \ell_k(\xi, y_k) \sum_{i=1}^N w_{k|k-1}^i \delta(\xi - x_{k|k-1}^i) d\xi} \\ &= \frac{\sum_{i=1}^N w_{k|k-1}^i \ell_k(x_{k|k-1}^i, y_k) \delta(x - x_{k|k-1}^i)}{\sum_{j=1}^N w_{k|k-1}^j \ell_k(x_{k|k-1}^j, y_k)} \\ &= \sum_{i=1}^N w_{k|k}^i \delta(x - x_{k|k}^i) \end{aligned}$$

con

$$\begin{aligned} x_{k|k}^i &= x_{k|k-1}^i \\ w_{k|k}^i &= \frac{\ell_k \left(x_{k|k-1}^i, y_k \right) w_{k|k-1}^i}{\sum_{j=1}^N \ell_k \left(x_{k|k-1}^j, y_k \right) w_{k|k-1}^j} \end{aligned} \quad (8.5.12)$$

per $i = 1, 2, \dots, N$.

In altri termini, la correzione preserva la natura discreta della PDF originaria $p_{k|k-1}(\cdot)$ con gli stessi campioni e pesi di importanza aggiornati con la nuova misura y_k in accordo alla seconda relazione in (8.5.12). Si considera adesso la predizione Bayesiana:

$$\begin{aligned} p_{k+1|k}(x) &= \int \varphi_{k+1|k}(x|\xi) p_{k|k}(\xi) d\xi \\ &= \int \varphi_{k+1|k}(x|\xi) \sum_{i=1}^N w_{k|k}^i \delta(x - x_{k|k}^i) d\xi \\ &= \sum_{i=1}^N w_{k|k}^i \int \varphi_{k+1|k}(x|\xi) \delta(\xi - x_{k|k}^i) d\xi \\ &= \sum_{i=1}^N w_{k|k}^i \varphi_{k+1|k}(x|x_{k|k}^i). \end{aligned} \quad (8.5.13)$$

Si noti che la PDF $p_{k+1|k}(\cdot)$ in (8.5.13) non è altro che una mistura delle PDF di transizione continue $\varphi_{k+1|k}(\cdot|x_{k|k}^i)$ condizionate ai vari campioni $x_{k|k}^i$. In particolare, nel caso di disturbo di processo additivo Gaussiano, $p_{k+1|k}(\cdot)$ si riduce ad una mistura di Gaussiane. In ogni caso, la PDF $p_{k+1|k}(\cdot)$ risultante dalla predizione non è più discreta (mistura di PDF impulsive) come $p_{k|k}(\cdot)$. Un algoritmo di *filtraggio a particelle* (PF = *Particle Filter* o *Sequenziale Monte Carlo* (SMCF= *Sequential Monte Carlo Filter*) deve, dunque, provvedere a campionare la PDF continua (8.5.13). Il campionamento di (8.5.13) può essere effettuato in svariati modi, i.e. utilizzando diverse densità di importanza, ciascuno dei quali caratterizza una diversa variante di filtro a particelle. Il metodo più comunemente usato è quello di generare $x_{k+1|k}^i$ campionando direttamente la mistura di PDF, ovvero nel seguente modo

per $i = 1, \dots, N$:

$$\begin{aligned} \omega_k^i &\sim p_{w_k}(\cdot) \\ x_{k+1|k}^i &= f_k \left(x_{k|k}^j, \omega_k^i \right) \quad \text{con probabilità } w_{k|k}^j. \end{aligned}$$

Questo porta all'approssimazione

$$p_{k+1|k}(x) \cong \sum_{i=1}^N w_{k+1|k} \delta(x - x_{k+1|k}^i)$$

con $w_{k+1|k} = 1/N$ per $i = 1, 2, \dots, N$. L'algoritmo risultante da questa scelta è denominato *filtro a particelle con ricampionamento sequenziale* (SIR-PF = *Sequential Importance Resampling - Particle Filter*) ed è riportato nell'algoritmo 6.

Un metodo generico di approssimazione consiste nell'utilizzo di una PDF di campionamento d'importanza generica $x_{k+1|k}^i \sim q(x, x_{k|k}^i, y_k)$, eventualmente tenendo conto delle misure.

Algorithm 6: Algoritmo SIR-PF.

Dati: Per $i = 1, 2, \dots, N$ si generano i campioni $x_{1|0}^i$ con la PDF $p_{1|0}(\cdot)$ e si pone $w_{1|0}^i = 1/N$

for $k = 1, 2, \dots$ **do**

Correzione:

si acquisisce y_k

for $i = 1, 2, \dots, N$ **do**

$$w_{k|k}^i = \frac{\ell_k(x_{k|k-1}^i, y_k) w_{k|k-1}^i}{\sum_{j=1}^N \ell_k(x_{k|k-1}^j, y_k) w_{k|k-1}^j}$$
$$x_{k|k}^i = x_{k|k-1}^i;$$

end for

$$\hat{x}_{k|k} = \sum_{i=1}^N w_{k|k}^i x_{k|k}^i$$

$$P_{k|k} = \sum_{i=1}^N w_{k|k}^i (x_{k|k}^i - \hat{x}_{k|k}) (x_{k|k}^i - \hat{x}_{k|k})^\top$$

Predizione:

for $i = 1, 2, \dots, N$ **do**

si estrae $j \in \{1, 2, \dots, N\}$ con probabilità $w_{k|k}^j$

si pone $x_{k+1|k}^i \sim \varphi_{k+1|k}(\cdot | x_{k|k}^j)$, che equivale a generare $x_{k+1|k}^i$

con un passo di simulazione a partire da $x_{k|k}^j$, cioè

$$x_{k+1|k}^i = f_k(x_{k|k}^j, \omega_k^i) \text{ con } \omega_k^i \sim p_{\omega_k}(\cdot)$$

$$w_{k+1|k}^i = 1/N$$

end for

$$\hat{x}_{k+1|k} = \sum_{i=1}^N w_{k+1|k}^i x_{k+1|k}^i$$

$$P_{k+1|k} = \sum_{i=1}^N w_{k+1|k}^i (x_{k+1|k}^i - \hat{x}_{k+1|k}) (x_{k+1|k}^i - \hat{x}_{k+1|k})^\top$$

end for

Bibliografia

- [1] A. Jazwinski, *Stochastic Processes and Filtering Theory*. Academic Press, 1970.
- [2] L. Servi and Y. Ho, “Recursive estimation in the presence of uniformly distributed measurement noise,” *IEEE Transactions on Automatic Control*, vol. 26, no. 2, pp. 563–565, 1981.
- [3] V. E. Benes, “Exact finite-dimensional filters for certain diffusions with non linear drift,” *Stochastics*, vol. 5, no. 1, pp. 65–92, 1981.
- [4] F. E. Daum, “Exact finite dimensional nonlinear filters,” *IEEE Transactions on Automatic Control*, vol. 31, no. 7, pp. 616–622, 1986.
- [5] S. F. Schmidt, “Application of state-space methods to navigation problems,” *Advanced Control Systems*, vol. 3, pp. 293–340, 1966.
- [6] K. Reif, S. Gunther, E. Yaz, and R. Unbehauen, “Stochastic stability of the discrete-time extended kalman filter,” *IEEE Transactions on Automatic Control*, vol. 44, no. 4, pp. 714–728, 1999.
- [7] S. Julier and J. Uhlmann, “Unscented filtering and nonlinear estimation,” *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.