

Corso di Laurea in Ingegneria e Scienze Informatiche

Studio e applicazione del filtro di Kalman e sue varianti per il tracciamento di corpi celesti

Tesi di laurea in:
METODI NUMERICI

Relatore
Prof.ssa Damiana Lazzaro

Candidato
Marco Buda

Sommario

TODO

Indice

| | |
|--|------------|
| Sommario | iii |
| 1 Introduzione | 1 |
| 2 Fondamenti teorici | 3 |
| 2.1 I sistemi dinamici | 3 |
| 2.2 Il filtro di Kalman | 5 |
| 2.3 Costruzione del filtro lineare | 6 |
| 2.4 Varianti non lineari | 12 |
| 2.4.1 Extended Kalman Filter | 13 |
| 2.4.2 Unscented Kalman Filter | 14 |
| 2.4.3 Metodo Monte Carlo | 16 |
| 2.5 Sviluppi recenti | 18 |
| 2.6 Limitazioni | 18 |
| 3 Modello astronomico | 19 |
| 3.1 Coordinate celesti | 19 |
| 3.2 Fotometria e il sistema H-G | 21 |
| 3.3 Orbite ellittiche | 21 |
| 4 Contributo | 25 |
| 4.1 Approccio diretto | 25 |
| 4.2 Sfide sul tracciamento di angoli | 27 |
| 4.3 Scelta dei vettori di stato e di misurazione | 29 |
| 4.4 Equazioni del sistema | 30 |
| 4.5 Matrici di covarianza | 31 |
| 4.6 Simulazione | 32 |
| 4.7 Applicazione a dati reali | 34 |
| 5 Conclusioni | 37 |

INDICE

| | |
|---------------------|-----------|
| | 39 |
| Bibliografia | 39 |

Elenco delle figure

| | | |
|-----|--|----|
| 2.1 | Diagramma di classificazione delle varianti del filtro di Kalman. . . | 17 |
| 4.1 | Confronto dei metodi per la stima di angoli. | 28 |
| 4.2 | Visualizzazione delle stime sui dati simulati (primi 20 passi), confrontate con l'approccio diretto. | 33 |
| 4.3 | Visualizzazione delle stime sui dati reali dell'asteroide Bennu (prime 130 osservazioni), confrontate con l'approccio diretto. | 35 |

Capitolo 1

Introduzione

TODO

Struttura della tesi TODO

Capitolo 2

Fondamenti teorici

In questo capitolo si definiscono i concetti di base su cui si svilupperanno gli argomenti di questa tesi, insieme ai modelli matematici che li descrivono. Si presenta, inoltre, una costruzione algebrica del filtro di Kalman e si offre una panoramica delle sue varianti.

2.1 I sistemi dinamici

Un sistema dinamico è un qualunque sistema che evolve nel tempo secondo una determinata legge.

A seconda dell'obiettivo, si individuano le proprietà del sistema che interessa esaminare (es.: posizione e velocità di un oggetto), dette variabili di stato, e le si raccoglie in un vettore, generalmente $x \in \mathbb{R}^n$, andando a definire una cosiddetta rappresentazione in spazio di stato.

Nella forma generale, si considera il tempo come continuo ($t \in \mathbb{R}$), per cui si parla di modelli continui, espressi con equazioni differenziali:

$$\frac{d}{dt}(x(t)) = f(t, x(t))$$

Tuttavia, si può decidere di considerare lo stato soltanto in determinati istanti t_k . Ne risultano modelli discreti, su cui si concentrerà questa tesi, i quali possono essere espressi con relazioni di ricorrenza:

$$x_k = f(k, x_{k-1})$$

In genere si sceglie di esplicitare la presenza di ingressi $u \in \mathbb{R}^l$ nel sistema (*control input*, es.: forza di gravità, propulsione...), assunti deterministici, e di misurazioni $z \in \mathbb{R}^m$ (*measurement*), effettuate sulle variabili osservate.

Dunque il modello viene espresso con le equazioni:

$$\begin{aligned} x_k &= f(k, x_{k-1}, u_{k-1}) \\ z_k &= h(k, x_k) \end{aligned}$$

La prima è detta “equazione di stato” e la seconda “trasformazione di uscita”.

Modellizzare la realtà comporta spesso che non si conosca la vera legge di evoluzione f_{reale} , ma solo una sua approssimazione, quindi soggetta a disturbi o imprecisioni (*process noise*, $w \in \mathbb{R}^n$), da cui l'importanza di effettuare misurazioni periodiche per monitorare la reale evoluzione dello stato.

Poiché le misurazioni stesse sono soggette a disturbi (*measurement noise*, $v \in \mathbb{R}^m$), il modello completo diventa:

$$x_k = f(k, x_{k-1}, u_{k-1}) + w_{k-1} \quad (2.1)$$

$$z_k = h(k, x_k) + v_k \quad (2.2)$$

Un caso particolare riguarda i modelli lineari, cioè esprimibili nella seguente forma matriciale:

$$x_k = A_k x_{k-1} + B_k u_{k-1} + w_{k-1} \quad (2.3)$$

$$z_k = H_k x_k + v_k \quad (2.4)$$

In questo caso, $A_k \in \mathbb{R}^{n \times n}$ è detta “matrice dinamica” e determina l'evoluzione dello stato dal passo $k - 1$ al passo k in assenza di ingressi e disturbi, $B_k \in \mathbb{R}^{l \times n}$

è detta “matrice di ingresso” e determina il contributo degli ingressi che agiscono dal passo $k - 1$ al passo k e $H_k \in \mathbb{R}^{m \times n}$ è detta “matrice di uscita” e determina le variabili osservate in base alle variabili da stimare.

In generale, i modelli sono tempo-varianti (o non stazionari), da cui k al pedice di A_k , B_k e H_k , ma si evidenziano i modelli LTI, ossia lineari tempo-invarianti (o lineari stazionari) in cui le matrici A , B e H sono costanti nel tempo.

Dalla combinazione delle imprecisioni nel modello e nelle misurazioni, nasce la necessità di algoritmi che raccolgano ed interpretino i dati osservati per determinare una stima (*state estimate*, \hat{x}) dello stato reale (*ground truth*, x). Questi algoritmi sono detti, appunto, stimatori.

2.2 Il filtro di Kalman

Il filtro di Kalman [1] [2] [3] è uno stimatore lineare ricorsivo che minimizza l’errore quadratico medio.

L’aspetto ricorsivo implica, innanzitutto, che si tratta di un filtro a tempo discreto e si rifà al fatto che ogni stima \hat{x}_k è determinata in base alla stima precedente \hat{x}_{k-1} e alla misurazione attuale z_k , senza richiedere l’utilizzo esplicito di $\hat{x}_0, \hat{x}_1, \dots, \hat{x}_{k-2}$ o z_1, \dots, z_{k-1} . Si tratta di uno dei principali vantaggi del filtro, in quanto riduce sia la complessità temporale che quella spaziale, senza compromettere l’ottimalità. Come anticipato, l’ottimalità è definita dal fatto che, ad ogni passo k , l’algoritmo produce la stima \hat{x}_k che minimizza la quantità $\mathbb{E}[\|e_k\|^2]$, con $e_k = \hat{x}_k - x_k$.

Il filtro è applicabile a un qualsiasi modello lineare a tempo discreto, anche tempo-variante, ossia descritto dalle equazioni 2.3 e 2.4.

Sono posti, però, vincoli di non correlazione sulle variabili aleatorie:

$$\begin{aligned}
\text{Cov}(w_k, v_j) &= 0_{n \times m}, & \forall k, \forall j \\
\text{Cov}(w_k, u_j) &= 0_{n \times l}, & \forall k, \forall 0 \leq j \leq k-1 \\
\text{Cov}(v_k, u_j) &= 0_{m \times l}, & \forall k, \forall 0 \leq j \leq k-2 \\
\text{Cov}(w_k, w_j) &= 0_{n \times n}, & \forall k, \forall j \neq k \\
\text{Cov}(v_k, v_j) &= 0_{m \times m}, & \forall k, \forall j \neq k \\
\text{Cov}(w_k, x_0) &= 0_{n \times n}, & \forall k \\
\text{Cov}(v_k, x_0) &= 0_{m \times n}, & \forall k
\end{aligned}$$

Infine, nelle formulazioni standard del filtro, si richiede che w e v siano privi di *bias*, ossia che $\mathbb{E}[w_k] = \underline{0}$ e $\mathbb{E}[v_k] = \underline{0}$, ad ogni passo k .

2.3 Costruzione del filtro lineare

In letteratura sono riportate diverse derivazioni e dimostrazioni di ottimalità del filtro di Kalman [4, pp. 107-113] [5]. Si presenta qui una costruzione algebrica il più elementare possibile.

Il primo passo è descrivere l'espressione che calcoli la stima \hat{x}_k . Al passo $k \geq 1$ sono disponibili le informazioni riguardo \hat{x}_{k-1} , u_{k-1} e z_k , per cui l'espressione lineare avrà la forma generica:

$$\hat{x}_k = \mathcal{A}_k \hat{x}_{k-1} + \mathcal{B}_k u_{k-1} + \mathcal{K}_k z_k \quad (2.5)$$

Per quanto riguarda il passo $k = 0$, la scelta della stima iniziale \hat{x}_0 è libera, purché sia deterministica. Se si hanno informazioni sulla distribuzione di x_0 , è consigliabile scegliere $\hat{x}_0 = \mathbb{E}[x_0]$.

Una condizione implicita sul filtro richiede che anche l'errore sulle stime gene-

rate e_k sia privo di *bias*, ossia, ad ogni passo k :

$$\mathbb{E}[e_k] = \mathbb{E}[\hat{x}_k - x_k] = \underline{0}$$

Da questa condizione si otterranno informazioni sulle matrici incognite \mathcal{A} , \mathcal{B} e \mathcal{K} . In effetti, considerando per $k \geq 1$ e sostituendo \hat{x}_k con la sua definizione (eq. 2.5), si ottiene:

$$\mathbb{E}[\mathcal{A}_k \hat{x}_{k-1} + \mathcal{B}_k u_{k-1} + \mathcal{K}_k z_k - x_k] = \underline{0}$$

Successivamente, sostituendo z_k con la sua definizione (eq. 2.4):

$$\mathbb{E}[\mathcal{A}_k \hat{x}_{k-1} + \mathcal{B}_k u_{k-1} + \mathcal{K}_k (H_k x_k + v_k) - x_k] = \underline{0}$$

Sostituendo x_k con la relazione di ricorrenza (eq. 2.3) e manipolando i termini:

$$\begin{aligned} & \mathbb{E}[\mathcal{A}_k \hat{x}_{k-1} + \mathcal{B}_k u_{k-1} + \mathcal{K}_k (H_k (A_k x_{k-1} + B_k u_{k-1} + w_{k-1}) + v_k) + \\ & \quad - (A_k x_{k-1} + B_k u_{k-1} + w_{k-1})] = \underline{0} \\ \Rightarrow & \mathbb{E}[\mathcal{A}_k \hat{x}_{k-1} + \mathcal{B}_k u_{k-1} + \mathcal{K}_k H_k A_k x_{k-1} + \mathcal{K}_k H_k B_k u_{k-1} + \mathcal{K}_k H_k w_{k-1} + \\ & \quad + \mathcal{K}_k v_k - A_k x_{k-1} - B_k u_{k-1} - w_{k-1}] = \underline{0} \\ \Rightarrow & \mathbb{E}[\mathcal{A}_k \hat{x}_{k-1} + \mathcal{B}_k u_{k-1} + \mathcal{K}_k H_k A_k x_{k-1} + \mathcal{K}_k H_k B_k u_{k-1} + \mathcal{K}_k H_k w_{k-1} + \\ & \quad + \mathcal{K}_k v_k - A_k x_{k-1} - B_k u_{k-1} - w_{k-1} - \mathcal{A}_k x_{k-1} + \mathcal{A}_k x_{k-1}] = \underline{0} \\ \Rightarrow & \mathbb{E}[\mathcal{A}_k (\hat{x}_{k-1} - x_{k-1}) + (\mathcal{K}_k H_k A_k - A_k + \mathcal{A}_k) x_{k-1} + \\ & \quad + (\mathcal{K}_k H_k B_k - B_k + \mathcal{B}_k) u_{k-1} + (\mathcal{K}_k H_k - I) w_{k-1} + \mathcal{K}_k v_k] = \underline{0} \\ \Rightarrow & \mathcal{A}_k \mathbb{E}[\hat{x}_{k-1} - x_{k-1}] + (\mathcal{K}_k H_k A_k - A_k + \mathcal{A}_k) \mathbb{E}[x_{k-1}] + \\ & \quad + (\mathcal{K}_k H_k B_k - B_k + \mathcal{B}_k) \mathbb{E}[u_{k-1}] + (\mathcal{K}_k H_k - I) \mathbb{E}[w_{k-1}] + \mathcal{K}_k \mathbb{E}[v_k] = \underline{0} \end{aligned}$$

Sfruttando l'ipotesi che le quantità $\hat{x}_{k-1} - x_{k-1}$, w_{k-1} e v_k siano prive di *bias*:

$$(\mathcal{K}_k H_k A_k - A_k + \mathcal{A}_k) \mathbb{E}[x_{k-1}] + (\mathcal{K}_k H_k B_k - B_k + \mathcal{B}_k) \mathbb{E}[u_{k-1}] = \underline{0}$$

Non potendo fare assunzioni su x e u , ne segue che:

$$\mathcal{K}_k H_k A_k - A_k + \mathcal{A}_k = 0_{n \times n} \quad , \quad \mathcal{K}_k H_k B_k - B_k + \mathcal{B}_k = 0_{l \times n}$$

$$\Rightarrow \mathcal{A}_k = (I - \mathcal{K}_k H_k) A_k, \quad \mathcal{B}_k = (I - \mathcal{K}_k H_k) B_k$$

Risulta pratico definire la seguente quantità, per $k \geq 1$, come una stima *a priori*, ossia che non tenga conto della misurazione z_k :

$$\hat{x}_k^- = A_k \hat{x}_{k-1} + B_k u_{k-1}$$

Così facendo, si vanno a distinguere due fasi all'interno di ogni passo: una fase di predizione, o *time update*, in cui si calcola \hat{x}_k^- in base alle conoscenze sul modello, e una fase di correzione, o *measurement update*, in cui si combina \hat{x}_k^- con la misurazione z_k per ottenere una stima ottimale \hat{x}_k .

L'espressione per ricavare \hat{x}_k , ossia la stima *a posteriori*, diventa ora:

$$\begin{aligned} \hat{x}_k &= (I - \mathcal{K}_k H_k) A_k \hat{x}_{k-1} + (I - \mathcal{K}_k H_k) B_k u_{k-1} + \mathcal{K}_k z_k = \\ &= (I - \mathcal{K}_k H_k) (A_k \hat{x}_{k-1} + B_k u_{k-1}) + \mathcal{K}_k z_k = (I - \mathcal{K}_k H_k) \hat{x}_k^- + \mathcal{K}_k z_k = \\ &= \hat{x}_k^- + \mathcal{K}_k (z_k - H_k \hat{x}_k^-) \end{aligned}$$

Proseguendo, si definisca $e_k^- = \hat{x}_k^- - x_k$ come l'errore sulla stima *a priori* e si consideri:

$$\begin{aligned} e_k^- &= \hat{x}_k^- - x_k = (A_k \hat{x}_{k-1} + B_k u_{k-1}) - (A_k x_{k-1} + B_k u_{k-1} + w_{k-1}) = \\ &= A_k (\hat{x}_{k-1} - x_{k-1}) - w_{k-1} = A_k e_{k-1} - w_{k-1} \end{aligned}$$

Si osservi che anche e_k^- è privo di *bias*:

$$\mathbb{E}[e_k^-] = \mathbb{E}[A_k e_{k-1} - w_{k-1}] = A_k \mathbb{E}[e_{k-1}] - E[w_{k-1}] = \underline{0}$$

Si definiscano, per $k \geq 1$, le auto-covarianze degli errori $P_k^- = \text{Cov}(e_k^-, e_k^-)$ e $P_k = \text{Cov}(e_k, e_k)$.

Si ha:

$$\begin{aligned}
 P_k^- &= \mathbb{E} [(e_k^- - \mathbb{E}[e_k^-])(e_k^- - \mathbb{E}[e_k^-])^T] = \mathbb{E} [e_k^- (e_k^-)^T] = \\
 &= \mathbb{E} [(A_k e_{k-1} - w_{k-1})(A_k e_{k-1} - w_{k-1})^T] = \\
 &= \mathbb{E} [(A_k e_{k-1} - w_{k-1}) (e_{k-1}^T A_k^T - w_{k-1}^T)] = \\
 &= A_k \mathbb{E} [e_{k-1} (e_{k-1})^T] A_k^T - A_k \mathbb{E} [e_{k-1} (w_{k-1})^T] + \\
 &\quad - \mathbb{E} [w_{k-1} (e_{k-1})^T] A_k^T + \mathbb{E} [w_{k-1} (w_{k-1})^T] = \\
 &= A_k \mathbb{E} [(e_{k-1} - \underline{0})(e_{k-1} - \underline{0})^T] A_k^T - A_k \mathbb{E} [(e_{k-1} - \underline{0})(w_{k-1} - \underline{0})^T] + \\
 &\quad - \mathbb{E} [(w_{k-1} - \underline{0})(e_{k-1} - \underline{0})^T] A_k^T + \mathbb{E} [(w_{k-1} - \underline{0})(w_{k-1} - \underline{0})^T] = \\
 &= A_k \mathbb{E} [(e_{k-1} - \mathbb{E}[e_{k-1}])(e_{k-1} - \mathbb{E}[e_{k-1}])^T] A_k^T + \\
 &\quad - A_k \mathbb{E} [(e_{k-1} - \mathbb{E}[e_{k-1}])(w_{k-1} - \mathbb{E}[w_{k-1}])^T] + \\
 &\quad - \mathbb{E} [(w_{k-1} - \mathbb{E}[w_{k-1}])(e_{k-1} - \mathbb{E}[e_{k-1}])^T] A_k^T + \\
 &\quad + \mathbb{E} [(w_{k-1} - \mathbb{E}[w_{k-1}])(w_{k-1} - \mathbb{E}[w_{k-1}])^T] = \\
 &= A_k \text{Cov}(e_{k-1}, e_{k-1}) A_k^T - A_k \text{Cov}(e_{k-1}, w_{k-1}) + \\
 &\quad - \text{Cov}(w_{k-1}, e_{k-1}) A_k^T + \text{Cov}(w_{k-1}, w_{k-1})
 \end{aligned}$$

Si osservi che i termini centrali si annullano se e_{k-1} e w_{k-1} sono non correlati. In effetti, analizzando ricorsivamente l'errore e_{k-1} , si trova che le uniche variabili aleatorie da cui esso dipende sono u_0, \dots, u_{k-2} , w_0, \dots, w_{k-2} , v_1, \dots, v_{k-1} e x_0 , ossia variabili con cui w_{k-1} è non correlato per ipotesi.

Ricordando la definizione di P_{k-1} e definendo Q_{k-1} come la auto-covarianza di w_{k-1} , si trova, dunque, l'espressione:

$$P_k^- = A_k P_{k-1} A_k^T + Q_{k-1}$$

La scelta iniziale di P_0 è pressoché libera. Se si hanno informazioni sulla distribuzione di x_0 , è consigliabile utilizzare una stima della sua auto-covarianza. In ogni caso, è necessario avere $P_0 \neq 0_{n \times n}$ e semidefinita positiva.

Per trovare un'espressione per P_k , si consideri inizialmente:

$$\begin{aligned}\hat{x}_k &= (I - \mathcal{K}_k H_k) \hat{x}_k^- + \mathcal{K}_k z_k \\ \Rightarrow e_k &= \hat{x}_k - x_k = (I - \mathcal{K}_k H_k) \hat{x}_k^- + \mathcal{K}_k z_k - x_k\end{aligned}$$

Sostituendo z_k con la sua definizione (eq. 2.4):

$$\begin{aligned}e_k &= (I - \mathcal{K}_k H_k) \hat{x}_k^- + \mathcal{K}_k (H_k x_k + v_k) - x_k = \\ &= (I - \mathcal{K}_k H_k) \hat{x}_k^- + \mathcal{K}_k H_k x_k + \mathcal{K}_k v_k - x_k = \\ &= (I - \mathcal{K}_k H_k) \hat{x}_k^- - (I - \mathcal{K}_k H_k) x_k + \mathcal{K}_k v_k = \\ &= (I - \mathcal{K}_k H_k) (\hat{x}_k^- - x_k) + \mathcal{K}_k v_k = (I - \mathcal{K}_k H_k) e_k^- + \mathcal{K}_k v_k\end{aligned}$$

Dunque, si ha:

$$\begin{aligned}P_k &= \mathbb{E} [(e_k - \mathbb{E}[e_k])(e_k - \mathbb{E}[e_k])^T] = \mathbb{E} [e_k (e_k)^T] = \\ &= \mathbb{E} \left[((I - \mathcal{K}_k H_k) e_k^- + \mathcal{K}_k v_k) ((I - \mathcal{K}_k H_k) e_k^- + \mathcal{K}_k v_k)^T \right] = \\ &= \mathbb{E} \left[((I - \mathcal{K}_k H_k) e_k^- + \mathcal{K}_k v_k) ((e_k^-)^T (I - \mathcal{K}_k H_k)^T + v_k^T \mathcal{K}_k^T) \right] = \\ &= (I - \mathcal{K}_k H_k) \mathbb{E} [e_k^- (e_k^-)^T] (I - \mathcal{K}_k H_k)^T + (I - \mathcal{K}_k H_k) \mathbb{E} [e_k^- (v_k)^T] \mathcal{K}_k^T + \\ &\quad + \mathcal{K}_k \mathbb{E} [v_k (e_k^-)^T] (I - \mathcal{K}_k H_k)^T + \mathcal{K}_k \mathbb{E} [v_k (v_k)^T] \mathcal{K}_k^T\end{aligned}$$

Anche in questo caso, l'espressione si riduce alle covarianze. I termini centrali si annullano, poiché le uniche variabili aleatorie ad influenzare e_k^- sono u_0, \dots, u_{k-2} , w_0, \dots, w_{k-1} , v_1, \dots, v_{k-1} e x_0 , ossia variabili con cui v_k è non correlato per ipotesi. Ricordando la definizione di P_k^- e definendo R_k come la auto-covarianza di v_k , si trova l'espressione provvisoria:

$$P_k = (I - \mathcal{K}_k H_k) P_k^- (I - \mathcal{K}_k H_k)^T + \mathcal{K}_k R_k \mathcal{K}_k^T$$

Si osservi che minimizzare l'errore quadratico medio $\mathbb{E}[\|e_k\|^2]$ equivale a minimizzare la traccia di P_k . In effetti:

$$\mathbb{E}[\|e_k\|^2] = \mathbb{E}[e_k^T e_k] = \mathbb{E}[\text{tr}(e_k^T e_k)] = \mathbb{E}[\text{tr}(e_k (e_k)^T)] = \text{tr}(\mathbb{E}[e_k (e_k)^T]) = \text{tr}(P_k)$$

Dunque, si calcoli:

$$\begin{aligned}
 \text{tr}(P_k) &= \text{tr}\left((I - \mathcal{K}_k H_k) P_k^- (I - \mathcal{K}_k H_k)^T + \mathcal{K}_k R_k \mathcal{K}_k^T\right) = \\
 &= \text{tr}\left((I - \mathcal{K}_k H_k) P_k^- (I - H_k^T \mathcal{K}_k^T) + \mathcal{K}_k R_k \mathcal{K}_k^T\right) = \\
 &= \text{tr}\left(P_k^- - P_k^- H_k^T \mathcal{K}_k^T - \mathcal{K}_k H_k P_k^- + \mathcal{K}_k H_k P_k^- H_k^T \mathcal{K}_k^T + \mathcal{K}_k R_k \mathcal{K}_k^T\right)
 \end{aligned}$$

Essendo P_k^- simmetrica, vale $P_k^- H_k^T \mathcal{K}_k^T = (P_k^-)^T H_k^T \mathcal{K}_k^T = (\mathcal{K}_k H_k P_k^-)^T$, per cui:

$$\text{tr}(P_k) = \text{tr}(P_k^-) - 2 \text{tr}(\mathcal{K}_k H_k P_k^-) + \text{tr}(\mathcal{K}_k H_k P_k^- H_k^T \mathcal{K}_k^T) + \text{tr}(\mathcal{K}_k R_k \mathcal{K}_k^T)$$

Per ricercare \mathcal{K}_k che minimizzi $\text{tr}(P_k)$ si ponga:

$$\begin{aligned}
 \frac{\partial \text{tr}(P_k)}{\partial \mathcal{K}_k} &= 0_{n \times n} \\
 \Rightarrow \frac{\partial \text{tr}(P_k^-)}{\partial \mathcal{K}_k} - 2 \frac{\partial \text{tr}(\mathcal{K}_k H_k P_k^-)}{\partial \mathcal{K}_k} + \frac{\partial \text{tr}(\mathcal{K}_k H_k P_k^- H_k^T \mathcal{K}_k^T)}{\partial \mathcal{K}_k} + \\
 &\quad + \frac{\partial \text{tr}(\mathcal{K}_k R_k \mathcal{K}_k^T)}{\partial \mathcal{K}_k} = 0_{n \times n}
 \end{aligned}$$

Si osservi che $\text{tr}(P_k^-)$ è costante in K_k , per cui il primo termine si annulla. Per calcolare i termini restanti, si utilizzino le seguenti identità, con la seconda valida se N è simmetrica:

$$\frac{\partial \text{tr}(MN)}{\partial M} = N^T, \quad \frac{\partial \text{tr}(MNM^T)}{\partial M} = 2MN$$

Si ottiene:

$$\begin{aligned}
 -2(H_k P_k^-)^T + 2\mathcal{K}_k (H_k P_k^- H_k^T) + 2\mathcal{K}_k R_k &= 0_{n \times n} \\
 \Rightarrow -2P_k^- H_k^T + 2\mathcal{K}_k H_k P_k^- H_k^T + 2\mathcal{K}_k R_k &= 0_{n \times n} \\
 \Rightarrow \mathcal{K}_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1}
 \end{aligned}$$

Esaminando la matrice Hessiana di $\text{tr}(P_k)$, la quale risulta essere semidefinita positiva, si può verificare che il valore trovato rappresenta un minimo globale.

La quantità \mathcal{K}_k è detta matrice dei guadagni di Kalman (*Kalman gain*) e può essere pensata come un indice di affidabilità delle misurazioni rispetto al modello teorico. In effetti, per $R_k \rightarrow 0_{m \times m}$ si ha $\mathcal{K}_k \rightarrow H^{-1}$, per cui $\hat{x}_k \rightarrow z_k$, mentre per $P_k^- \rightarrow 0_{n \times n}$ si ha $\mathcal{K}_k \rightarrow 0_{n \times m}$, per cui $\hat{x}_k \rightarrow \hat{x}_k^-$.

Sostituendo il valore trovato, l'espressione per P_k diventa:

$$\begin{aligned} P_k &= P_k^- - P_k^- H_k^T \mathcal{K}_k^T - \mathcal{K}_k H_k P_k^- + \mathcal{K}_k H_k P_k^- H_k^T \mathcal{K}_k^T + \mathcal{K}_k R_k \mathcal{K}_k^T = \\ &= (I - \mathcal{K}_k H_k) P_k^- - P_k^- H_k^T \mathcal{K}_k^T + \mathcal{K}_k (H_k P_k^- H_k^T + R_k) \mathcal{K}_k^T = \\ &= (I - \mathcal{K}_k H_k) P_k^- - P_k^- H_k^T \mathcal{K}_k^T + \\ &\quad + P_k^- H_k^T \underbrace{(H_k P_k^- H_k^T + R_k)^{-1} (H_k P_k^- H_k^T + R_k)}_I \mathcal{K}_k^T = \\ &= (I - \mathcal{K}_k H_k) P_k^- - P_k^- H_k^T \mathcal{K}_k^T + P_k^- H_k^T \mathcal{K}_k^T = \\ &= (I - \mathcal{K}_k H_k) P_k^- \end{aligned}$$

Riassumendo, l'algoritmo si basa sulle seguenti formule:

$$\hat{x}_0 = \mathbb{E}[x_0]$$

$$P_0 = \text{Cov}(x_0, x_0)$$

$$\hat{x}_k^- = A_k \hat{x}_{k-1} + B u_{k-1}$$

$$P_k^- = A_k P_{k-1} A_k^T + Q_{k-1}$$

$$\mathcal{K}_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1}$$

$$\hat{x}_k = \hat{x}_k^- + \mathcal{K}_k (z_k - H_k \hat{x}_k^-)$$

$$P_k = (I - \mathcal{K}_k H_k) P_k^-$$

2.4 Varianti non lineari

Sulla base della formulazione originale di Kálmán — oggi conosciuta come “Standard” Kalman Filter o KF — sono state sviluppate numerose varianti del filtro, con l'obiettivo principale di estenderne il campo di applicazione, specie a mo-

delli non lineari. Si presenta qui una descrizione delle varianti più conosciute, successivamente schematizzate in fig. 2.1.

2.4.1 Extended Kalman Filter

EKF [2] nasce dalle prime applicazioni delle intuizioni di Kálmán e Bucy al programma Apollo [6]. Si basa sull'utilizzo naturale del polinomio di Taylor di grado 1 per trovare approssimazioni lineari alle equazioni del sistema, in generale non lineari.

Partendo dalla formula generale del polinomio in una variabile reale:

$$f(x) = f(a) + f'(a)(x - a) + o(x - a)$$

e applicandola alla legge di evoluzione f di un sistema dinamico, si ottiene:

$$f(k, x_{k-1}, u_{k-1}) \approx f(k, \hat{x}_{k-1}, u_{k-1}) + J_x f(k, \hat{x}_{k-1}, u_{k-1})(x_{k-1} - \hat{x}_{k-1})$$

Qui, $f'(a)$ è sostituita da $J_x f(k, \hat{x}_{k-1}, u_{k-1})$, ossia dalla matrice jacobiana delle derivate parziali di f rispetto ad x calcolate in $k, \hat{x}_{k-1}, u_{k-1}$.

Da qui in poi verrà indicata con A_k , poiché verrà utilizzata nelle formule del filtro in modo identico alla matrice dinamica dei sistemi dinamici lineari.

$$(A_k)_{ij} = \left. \frac{\partial f_i}{\partial x_j} \right|_{k, \hat{x}_{k-1}, u_{k-1}}$$

Lo stesso ragionamento vale per la funzione di misurazione h :

$$h(k, x_k) \approx h(k, \hat{x}_k) + J_x h(k, \hat{x}_k)(x_k - \hat{x}_k)$$

$$H_k = J_x h(k, \hat{x}_k) \Rightarrow (H_k)_{ij} = \left. \frac{\partial h_i}{\partial x_j} \right|_{k, \hat{x}_k}$$

Dunque, il sistema è descritto dalle seguenti approssimazioni lineari:

$$x_k \approx f(k, \hat{x}_{k-1}, u_{k-1}) + A_k(x_{k-1} - \hat{x}_{k-1}) + w_{k-1}$$

$$z_k \approx h(k, \hat{x}_k) + H_k(x_k - \hat{x}_k) + v_k$$

Una volta determinate le matrici jacobiane A_k e H_k , EKF fa uso delle stesse formule di KF per il calcolo di P_k^- , P_k e K_k . Dove possibile, ossia per il calcolo delle stime, utilizza, invece, le funzioni f e h direttamente:

$$\begin{aligned}\hat{x}_k^- &= f(\hat{x}_{k-1}, u_{k-1}) \\ \hat{x}_k &= \hat{x}_k^- + K_k(z_k - h(\hat{x}_k^-))\end{aligned}$$

Il filtro risultante è poco più complesso di KF, ma con un evidente vantaggio in termini di flessibilità. Tuttavia, il processo di linearizzazione è efficace solo per piccoli errori e può essere soggetto a rapida divergenza.

Un logico sviluppo di queste idee prevede di includere anche i termini di secondo grado nelle espansioni in serie di Taylor. Si parla in questo caso di Second-Order Extended Kalman Filter, abbreviato EKF2 o SO-EKF.

2.4.2 Unscented Kalman Filter

UKF [7] è il più noto dei cosiddetti filtri Sigma-Point. Queste varianti selezionano deterministicamente un piccolo insieme di punti ($2n + 1$ per UKF) attorno a \hat{x}_{k-1} e studiano come questi reagiscono alla trasformazione non lineare, fornendo informazioni più accurate sulla distribuzione delle stime ottenute.

UKF utilizza tre parametri che influenzano la scelta dei Sigma-Point: $\alpha \ll 1$ ne determina la dispersione, mentre $\beta = 2$ e $\kappa = 0$ raramente richiedono di essere modificati.

Una volta determinato il parametro intermedio $\lambda = \alpha^2(n + \kappa) - n$, si calcola la matrice $(\lambda + n)P_{k-1}$ e si utilizza un metodo di decomposizione, come l'algoritmo di Cholesky, per trovare $U = \sqrt{(\lambda + n)P_{k-1}}$.

Ora, indicando con $U_{[i]}$ la i -esima riga di U , i punti selezionati sono dati da:

$$\mathbf{x}_{k-1}^{(i)} = \begin{cases} \hat{x}_{k-1} & \text{se } i = 0 \\ \hat{x}_{k-1} + U_{[i]} & \text{se } i = 1, \dots, n \\ \hat{x}_{k-1} - U_{[i-n]} & \text{se } i = n + 1, \dots, 2n \end{cases}$$

A ciascuno viene anche associato un peso fisso sul calcolo del valore medio e uno sul calcolo della covarianza:

$$W^{(\text{mean}, i)} = \begin{cases} \frac{\lambda}{\lambda+n} & \text{se } i = 0 \\ \frac{1}{2(\lambda+n)} & \text{se } i = 1, \dots, 2n \end{cases}$$

$$W^{(\text{cov}, i)} = \begin{cases} \frac{\lambda}{\lambda+n} + (1 - \alpha^2 + \beta) & \text{se } i = 0 \\ \frac{1}{2(\lambda+n)} & \text{se } i = 1, \dots, 2n \end{cases}$$

Dunque, viene applicata la trasformazione non lineare, per determinare la stima *a priori* e la sua covarianza come somme pesate:

$$\mathbf{x}_k^{(i)} = f(k, \mathbf{x}_{k-1}^{(i)}, u_{k-1}) \quad \forall i = 0, \dots, 2n$$

$$\hat{x}_k^- = \sum_{i=0}^{2n} W^{(\text{mean}, i)} \mathbf{x}_k^{(i)}$$

$$P_k^- = \sum_{i=0}^{2n} W^{(\text{cov}, i)} \left(\mathbf{x}_k^{(i)} - \hat{x}_k^- \right) \left(\mathbf{x}_k^{(i)} - \hat{x}_k^- \right)^T$$

Similmente, nella fase di correzione viene applicata la trasformazione h per trovare una previsione della misurazione \hat{z}_k^- (si mantiene la notazione all'apice per indicare che il calcolo viene effettuato prima della misurazione, seppure non sia necessario migliorare successivamente questa stima). Ne viene poi calcolata sia la auto-covarianza che la covarianza incrociata con la stima *a priori*:

$$\mathbf{z}_k^{(i)} = h(k, \mathbf{x}_{k-1}^{(i)}) \quad \forall i = 0, \dots, 2n$$

$$\hat{z}_k^- = \sum_{i=0}^{2n} W^{(\text{mean}, i)} \mathbf{z}_k^{(i)}$$

$$\text{Cov}(\hat{z}_k^-, \hat{z}_k^-) = \sum_{i=0}^{2n} W^{(\text{cov}, i)} \left(\mathbf{z}_k^{(i)} - \hat{z}_k^- \right) \left(\mathbf{z}_k^{(i)} - \hat{z}_k^- \right)^T$$

$$\text{Cov}(\hat{x}_k^-, \hat{z}_k^-) = \sum_{i=0}^{2n} W^{(\text{cov}, i)} \left(\mathbf{x}_k^{(i)} - \hat{x}_k^- \right) \left(\mathbf{z}_k^{(i)} - \hat{z}_k^- \right)^T$$

A questo punto, la matrice dei guadagni di Kalman viene calcolata come:

$$\mathcal{K}_k = \text{Cov}(\hat{x}_k^-, \hat{z}_k^-) \text{Cov}(\hat{z}_k^-, \hat{z}_k^-)^{-1}$$

Infine, la stima *a posteriori* e la relativa auto-covarianza sono date da:

$$\begin{aligned}\hat{x}_k &= \hat{x}_k^- + \mathcal{K}_k(z_k - \hat{z}_k^-) \\ P_k &= P_k^- - \mathcal{K}_k \text{Cov}(\hat{z}_k^-, \hat{z}_k^-) \mathcal{K}_k^T\end{aligned}$$

UKF permette di determinare stime accurate quanto un polinomio di Taylor di grado 3, ma con la stessa complessità computazionale di EKF base.

Sono disponibili diverse implementazioni di questo algoritmo. In questa tesi si farà uso della classe `filterpy.kalman.UnscentedKalmanFilter`, con i relativi Sigma-Point `filterpy.kalman.sigma_points.MerweScaledSigmaPoints`.

Altre varianti di tipo Sigma-Point includono i filtri Quadrature KF (QKF), fra cui il più conosciuto Gauss-Hermite Kalman Filter (GHKF), basati sulle formule di integrazione numerica (o quadratura numerica) di Gauss, e i Cubature KF (CKF), che utilizzano la “regola di cubatura”, ossia l’applicazione delle stesse regole di quadratura ad uno spazio in coordinate sferiche.

2.4.3 Metodo Monte Carlo

Un’ulteriore strategia per lo studio della distribuzione di una variabile aleatoria sottoposta ad una trasformazione non lineare è il classico metodo Monte Carlo. Diversamente dai metodi Sigma-Point, si considerano campioni cospicui di punti, consentendo di raggiungere precisione arbitraria a costo di un maggiore onere computazionale, il quale, però, può essere facilmente distribuito con tecniche di parallelizzazione. Da queste idee nascono Ensemble KF (EnKF), che richiede distribuzioni iniziali Gaussiane, e i Particle Filters (PF), estremamente flessibili, ma meno efficienti.

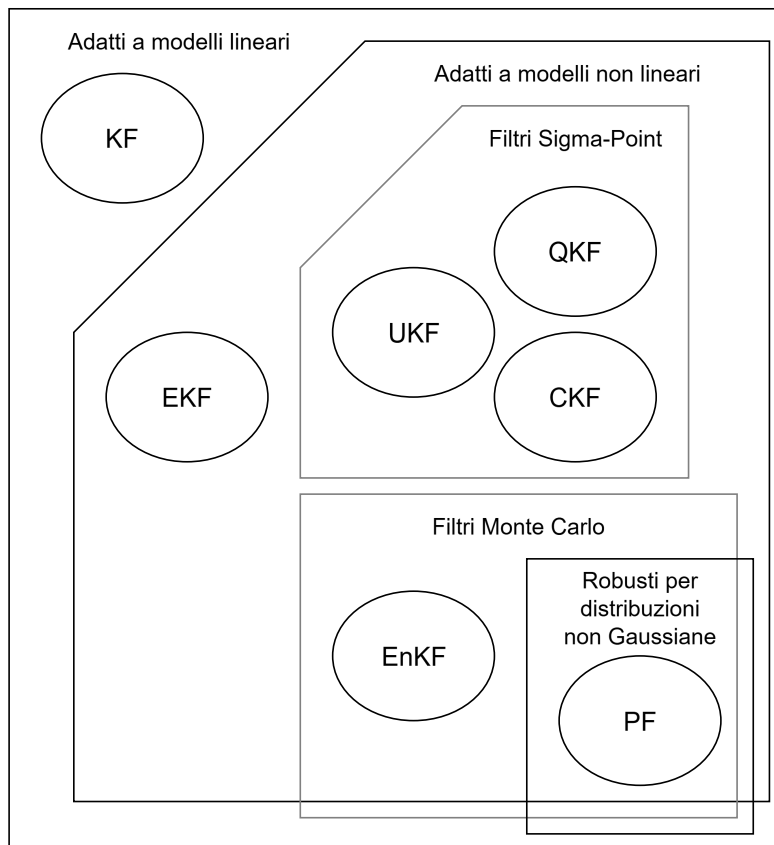


Figura 2.1: Diagramma di classificazione delle varianti del filtro di Kalman.

2.5 Sviluppi recenti

Lo studio del filtro di Kalman resta un settore vivo, a più di 60 anni dalla sua concezione. Si continuano a identificare nuovi settori di applicazione, nuove tecniche di miglioramento dell'accuratezza, della stabilità numerica e della performance computazionale. Nuove varianti, sia generiche che specializzate, si aggiungono a quelle già consolidate. Solo negli ultimi mesi si evidenziano una tecnica di “Iteratively Saturated Kalman Filtering” (ISKF) [8], che mira a ridurre l'impatto di eventuali misurazioni anomale, e la possibilità di affiancare una rete neurale profonda (DNN) ad un filtro classico nei cosiddetti “AI-Aided Kalman Filters” [9].

2.6 Limitazioni

Sebbene oggi assuma un ruolo fondamentale nell'ingegneria aerospaziale e in diversi altri settori, il filtro di Kalman presenta alcune limitazioni intrinseche:

- Per sistemi non lineari con distribuzioni non Gaussiane, non è garantita, in generale, l'ottimalità.
- Benché la conoscenza del sistema reale sia la base dell'efficacia del filtro, questa può anche costituire un limite sulla precisione, qualora modelli sufficientemente accurati non siano disponibili.

Capitolo 3

Modello astronomico

In questo capitolo si presentano le formule e i modelli teorici che governano le osservazioni di corpi celesti e il tracciamento del loro moto. Si delinea un sistema dinamico altamente non lineare, che dimostra la necessità di tecniche e algoritmi avanzati per lo studio di questo caso d'uso reale.

3.1 Coordinate celesti

Le osservazioni astrometriche fanno uso di una coppia di angoli (α, δ) per identificare la posizione dei corpi nella sfera celeste, o, più precisamente, per identificare la direzione orientata dall'osservatore verso la posizione apparente del corpo considerato. I due angoli sono definiti con riferimento al piano equatoriale e al punto vernale dell'epoca J2000.0 (mezzogiorno del 1° gennaio 2000, calendario gregoriano e tempo terrestre). Nella pratica si considerano assi orientati secondo il sistema ICRS, adottato dalla International Astronomical Union dal 1° gennaio 1998, spostando l'origine dal baricentro del sistema solare alla posizione dell'osservatore (osservazione topocentrica). In questo contesto, l'ascensione retta $\alpha \in [0, 2\pi)$ rappresenta la rotazione attorno all'asse z , mentre la declinazione $\delta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ rappresenta la distanza angolare dal piano xy .

Dunque, indicando con Δ la distanza apparente tra l'osservatore e il corpo al momento dell'osservazione, la posizione apparente del corpo osservato, in coordinate cartesiane topocentriche, è data da:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = R_z(\alpha) R_y(-\delta) \begin{pmatrix} \Delta \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \Delta \cos(\delta) \cos(\alpha) \\ \Delta \cos(\delta) \sin(\alpha) \\ \Delta \sin(\delta) \end{pmatrix} \quad (3.1)$$

La conversione a coordinate baricentriche richiede semplicemente di sommare la posizione dell'osservatore rispetto al baricentro del sistema solare.

Si definiscono le distanze apparenti R tra l'osservatore e il Sole al momento dell'osservazione e d (solitamente indicata con r , ma da non confondersi con la distanza dal fuoco occupato dell'orbita) tra il corpo osservato e il Sole nel momento in cui la luce osservata viene riflessa sulla superficie del corpo. Si definiscono gli angoli di separazione $\theta \in [0, \pi]$, detta elongazione, tra il corpo e il Sole come appaiono all'osservatore al momento dell'osservazione e $\phi \in [0, \pi]$, detto angolo di fase, tra l'osservatore e il Sole come appaiono dal corpo nel momento in cui la luce osservata viene riflessa. Seppure si tratti di una definizione analoga, l'angolo di fase viene definito, in genere, come l'angolo tra la luce incidente e la luce riflessa verso l'osservatore, per via del suo impatto sul modello ottico descritto nella sezione 3.2.

Si è parlato finora di quantità apparenti. Questa precisazione è necessaria per via della velocità finita della luce e ad altri fenomeni come aberrazioni, rifrazioni ed effetti relativistici. Da questo punto in poi si considereranno le quantità apparenti come coincidenti alle controparti “reali” (geometriche), con la consapevolezza che ciò introdurrà imprecisioni nel modello, le quali dovranno essere considerate nell'applicazione del filtro di Kalman. Queste imprecisioni sono accettabili per il tracciamento di asteroidi, ma possono diventare problematiche per oggetti più veloci.

Con questa semplificazione diventa possibile considerare il triangolo tra il corpo osservato, l'osservatore e il Sole e applicare il teorema dei seni e il teorema dei coseni

per determinare relazioni fra gli angoli e le distanze. In particolare:

$$\frac{d}{\sin(\theta)} = \frac{R}{\sin(\phi)} \quad (3.2)$$

$$\frac{\Delta}{\sin(\pi - \theta - \phi)} = \frac{R}{\sin(\phi)} \Rightarrow \frac{\Delta}{\sin(\theta + \phi)} = \frac{R}{\sin(\phi)} \quad (3.3)$$

$$r^2 + \Delta^2 - 2r\Delta \cos(\phi) = R^2 \Rightarrow \phi = \arccos\left(\frac{r^2 + \Delta^2 - R^2}{2r\Delta}\right) \quad (3.4)$$

3.2 Fotometria e il sistema H-G

Nello studio degli asteroidi, l'intensità della luce osservata è espressa dalla magnitudine apparente e può essere utilizzata per stimare la distanza del corpo. Intuitivamente, questa quantità dipende dall'angolo di fase ($\phi \approx 0$ indica un corpo perfettamente illuminato, mentre $\phi \approx 1$ comporta che non venga riflessa luce verso l'osservatore), dalle distanze d e Δ e da alcune caratteristiche fisiche del corpo. Tuttavia non esiste un modello macroscopico esatto.

Una delle prime applicazioni delle nozioni di fotometria allo studio degli asteroidi ha prodotto il modello H-G [10] [11], usato tuttora quando non si dispone di informazioni sufficienti per impiegare modelli più precisi. Gli asteroidi vengono caratterizzati da due parametri: H , detta magnitudine assoluta, che equivale alla magnitudine apparente per $d = \Delta = 1$ AU e $\phi = 0$ e G , detto *slope parameter*, che quantifica un picco di luminosità che spesso si verifica nella curva di fase vicino all'opposizione. Se si esprimono le distanze in unità astronomiche, la magnitudine apparente è descritta da:

$$V = H + 5 \log_{10}(d\Delta) - 2.5 \log_{10}(\Phi(\phi)) \quad (3.5)$$

$$\Phi(\phi) = (1 - G) \exp\left(-3.33\left(\tan\left(\frac{\phi}{2}\right)\right)^{0.63}\right) + G \exp\left(-1.87\left(\tan\left(\frac{\phi}{2}\right)\right)^{1.22}\right) \quad (3.6)$$

3.3 Orbite ellittiche

Il moto dei pianeti, degli asteroidi e delle comete è ben approssimato da orbite ellittiche, con un fuoco occupato dal baricentro del sistema solare.

Tali orbite possono essere caratterizzate tramite i parametri orbitali kepleriani:

- il semiasse maggiore a ;
- l'eccentricità e , dove $e = 0$ indica un'orbita circolare, mentre $e \rightarrow 1$ descrive un'orbita progressivamente più schiacciata;
- l'inclinazione $i \in [0, \pi]$ del piano dell'orbita, in genere espressa rispetto al piano dell'orbita della Terra, ma in questo caso espressa rispetto al piano xy del sistema di riferimento, con $i < \frac{\pi}{2}$ indicante un'orbita progradata e $i > \frac{\pi}{2}$ indicante un'orbita retrogradata;
- la longitudine del nodo ascendente Ω ;
- l'argomento del periapside ω .

La posizione del corpo nell'orbita è identificata dall'anomalia. Più precisamente, la distanza angolare dal periapside è detta anomalia vera e indicata con ν .

Si definisce la distanza radiale r dal fuoco occupato e l'argomento di latitudine $u = \omega + \nu$, cosicché le coordinate cartesiane del corpo siano date da:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = R_z(\Omega)R_x(i)R_z(u) \begin{pmatrix} r \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} r(\cos(\Omega)\cos(u) - \sin(\Omega)\sin(u)\cos(i)) \\ r(\sin(\Omega)\cos(u) + \cos(\Omega)\sin(u)\cos(i)) \\ r\sin(u)\sin(i) \end{pmatrix} \quad (3.7)$$

Per studiare l'evoluzione della posizione nel tempo si introducono i concetti di anomalia media M e anomalia eccentrica E .

La prima avanza linearmente nel tempo in base al moto medio n e al tempo di periapside τ :

$$M = n(t - \tau) \quad (3.8)$$

Il moto medio è definito dal periodo dell'orbita come $n = \frac{2\pi}{T}$ ed è ricavabile dal semiasse maggiore e dal parametro gravitazionale $\mu = GM$ del corpo primario (in questo caso G è la costante di gravitazione universale e M è la massa del corpo primario):

$$n = \sqrt{\frac{\mu}{a^3}} \quad (3.9)$$

Una volta determinata l'anomalia media, si risale all'anomalia eccentrica esprimendo gli angoli in radianti e risolvendo l'equazione di Keplero tramite metodi numerici:

$$M = E - e \sin(E) \quad (3.10)$$

In questo caso si è scelto di utilizzare il metodo di Newton, ponendo $f(x) = x - e \sin(x) - M$, $f'(x) = 1 - e \cos(x)$ e $x_0 = M$.

Dall'anomalia eccentrica è possibile passare all'anomalia vera. Una formula numericamente stabile richiede di calcolare il valore intermedio β :

$$\beta = \frac{e}{1 + \sqrt{1 - e^2}} \quad (3.11)$$

$$\nu = E + 2 \arctan \left(\frac{\beta \sin(E)}{1 - \beta \cos(E)} \right) \quad (3.12)$$

Infine, la distanza radiale è legata all'anomalia eccentrica da:

$$r = a(1 - e \cos(E)) \quad (3.13)$$

Capitolo 4

Contributo

In questo capitolo si illustra come sono state messe in pratica le nozioni teoriche in modo da sperimentare concretamente l'utilizzo del filtro di Kalman, come sono state adattate le formule astronomiche per ottenere le equazioni del sistema, come sono state ricavate le matrici di covarianza e le sfide affrontate durante il percorso. Si presentano, infine, i risultati ottenuti dapprima su dati simulati e successivamente su dati reali.

4.1 Approccio diretto

Prima ancora di considerare strategie di filtraggio o di tracciamento del moto orbitale, si potrebbe pensare di risalire a una posizione direttamente dai dati osservati. In effetti, se si conoscesse la distanza del corpo Δ , la si potrebbe sfruttare, insieme agli angoli α, δ , nella formula 3.1. Tuttavia, risalire a Δ dalla magnitudine visuale V richiede particolare attenzione.

Idealmente, si vorrebbe sfruttare direttamente la formula 3.5, ma non si dispone della distanza dal Sole d e dell'angolo di fase ϕ . Una possibile soluzione prevede di risolvere in ϕ la seguente equazione tramite metodi numerici:

$$R^2 \sin(\theta) \sin(\theta + \phi) - 10^{0.2(V-H)} \sin^2(\phi) \sqrt{\Phi(\phi)} = 0 \quad (4.1)$$

Tale equazione deriva dai seguenti passaggi:

$$\begin{aligned} V &= H + 5 \log_{10}(d\Delta) - 2.5 \log_{10}(\Phi(\phi)) \\ \rightarrow 0.2(V - H) + \log_{10}((\Phi(\phi))^{0.5}) &= \log_{10}(d\Delta) \\ \rightarrow d\Delta &= 10^{0.2(V-H)} \sqrt{\Phi(\phi)} \end{aligned}$$

Utilizzando le equazioni 3.2 e 3.3, si trova che:

$$d\Delta = \frac{R^2 \sin(\theta) \sin(\theta + \phi)}{\sin^2(\phi)}$$

Sostituendo e moltiplicando entrambi i lati per $\sin^2(\phi)$ si ottiene l'espressione desiderata. Si osservi che per $\phi = 0$ e $\phi = \pi$ sarebbe impossibile effettuare osservazioni valide, per cui si può assumere $\sin(\phi) \neq 0$. Lo stesso vale per l'elongazione θ .

Il vantaggio di questa specifica scrittura sta nel fatto che l'espressione ha valore positivo finito per $\phi = 0$, ovvero $R^2 \sin^2(\theta)$, e ha valore negativo finito per $\phi \rightarrow \pi$, ovvero $-R^2 \sin^2(\theta)$, il che permette di utilizzare il metodo di Brent nell'intervallo $[0, \pi]$.

Una volta determinata una radice, la distanza è data da:

$$\Delta = \frac{R \sin(\theta + \phi)}{\sin(\phi)} \quad (4.2)$$

Si osservi che in generale può esistere più di una radice, specie se l'elongazione θ è piccola; tuttavia questo non si verifica nei dataset considerati.

In ogni caso, questo processo è estremamente sensibile all'errore sui dati iniziali e la magnitudine risulta essere una misura piuttosto imprecisa. Tutto ciò evidenzia ulteriormente la necessità di strategie più sofisticate.

Questo approccio iniziale resta comunque un utile punto di riferimento e verrà utilizzato per quantificare l'efficacia del filtro di Kalman in questo contesto.

Per il calcolo di R e θ si può sfruttare un servizio di effemeridi che fornisca le posizioni dell'osservatore e del Sole. In questo caso si è utilizzato il servizio JPL, sia attraverso le funzioni `astropy.coordinates.get_body_barycentric` e `astropy.coordinates.EarthLocation.get_gcrs_posvel`, che dalla piattaforma web <https://ssd.jpl.nasa.gov/horizons/app.html>.

Indicando con \hat{R} il versore dall'osservatore verso il sole e con $\hat{\Delta}$ il versore dall'osservatore verso il corpo, l'elongazione è data da:

$$\theta = \arccos(\hat{\Delta} \cdot \hat{R}) \quad (4.3)$$

Con $\hat{\Delta}$ dato da:

$$\hat{\Delta} = \begin{pmatrix} \cos(\delta) \cos(\alpha) \\ \cos(\delta) \sin(\alpha) \\ \sin(\delta) \end{pmatrix} \quad (4.4)$$

Si procede ora con l'impostazione del problema sotto forma di sistema dinamico, come richiesto dal filtro di Kalman.

4.2 Sfide sul tracciamento di angoli

Per via della natura del problema, si vuole memorizzare quantità angolari sia nel vettore di stato che nel vettore di misurazione. Tuttavia, fare ciò senza dovuti accorgimenti introdurrebbe forti discontinuità nelle equazioni che descrivono il sistema. Si noti, ad esempio, che all'avanzare dell'anomalia, questa passa da $\nu \approx 2\pi$ a $\nu \approx 0$. In aggiunta, una stima $\hat{\nu} \approx 2\pi$ affiancata ad una misurazione $\nu \approx 0$ è in realtà da considerarsi valida, ma dal punto di vista del filtro apparirebbe come una grossa discrepanza.

Per affrontare questo problema, si decide di non memorizzare direttamente gli angoli, ma bensì di rappresentarli nelle loro componenti sulla circonferenza unitaria, ossia la coppia $(\cos(\cdot), \sin(\cdot))$. Questo approccio è analogo alla rappresentazione di rotazioni in tre dimensioni attraverso l'uso di quaternioni e permette di rimuovere le discontinuità di cui si è parlato.

In figura 4.1 si mostra l'efficacia di questa soluzione applicata alla simulazione di un corpo in moto circolare uniforme.

Tuttavia restano delle limitazioni, dovute al fatto che questa trasformazione è pressoché lineare solo per piccoli errori sugli angoli. Ne segue, ad esempio, che la media fra due coppie di componenti possa non corrispondere esattamente alla media dei due rispettivi angoli, oppure possa non risiedere precisamente sulla circonferenza unitaria.

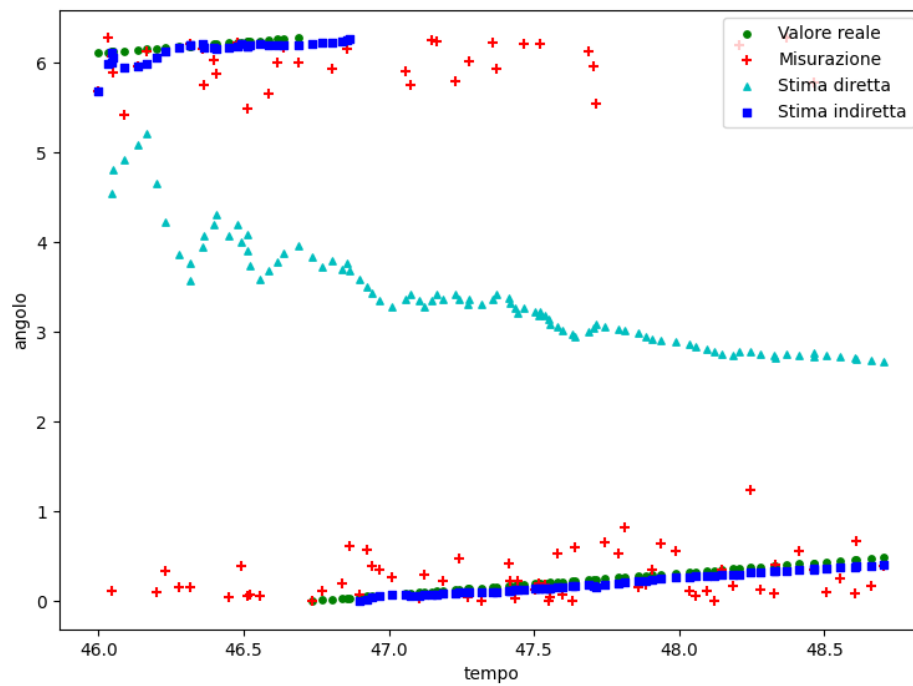


Figura 4.1: Confronto dei metodi per la stima di angoli.

4.3 Scelta dei vettori di stato e di misurazione

Un passo fondamentale è la scelta delle quantità da considerare nel vettore di stato. I parametri orbitali a, e, i, Ω, ω sono pressoché costanti, ma averne stime accurate nel tempo può essere vantaggioso, per cui si sceglie di tracciarli. Questo apre anche la possibilità di partire con stime grossolane e lasciare al filtro il compito di determinare i valori più probabili.

Il moto medio n può sempre essere determinato dal semiasse maggiore, ma un approccio che consenta piccole variazioni dovute a fattori esterni può produrre risultati migliori.

I parametri H e G sono costanti, ma soggetti ad incertezza, per cui si cerca di trovare stime più accurate.

La decisione più complessa riguarda la posizione del corpo. L'obiettivo finale è tracciarne la posizione cartesiana, per cui un'opzione naturale sarebbe quella di aggiungere al vettore le componenti x, y, z . Tuttavia, per questa tesi si è scelto di tracciare l'anomalia media M e, solo in un secondo momento, risalire alla posizione cartesiana. Questo permette di avere una legge di evoluzione molto più semplice e numericamente stabile. Sarebbe comunque consigliabile uno studio approfondito per mettere a confronto le diverse alternative.

In conclusione, esprimendo le quantità angolari nelle loro componenti, come descritto nella sezione 4.2, il vettore di stato diventa:

$$x = \begin{pmatrix} \cos(M) \\ \sin(M) \\ a \\ e \\ \cos(i) \\ \sin(i) \\ \cos(\Omega) \\ \sin(\Omega) \\ \cos(\omega) \\ \sin(\omega) \\ n \\ H \\ G \end{pmatrix} \quad (4.5)$$

La scelta del vettore di misurazione è più diretta, poiché le quantità misurate sono α , δ e V :

$$z = \begin{pmatrix} \cos(\alpha) \\ \sin(\alpha) \\ \cos(\delta) \\ \sin(\delta) \\ V \end{pmatrix} \quad (4.6)$$

4.4 Equazioni del sistema

La legge di evoluzione f agisce esclusivamente sull'anomalia media, secondo la formula 3.8:

$$x_{k-1} = \begin{pmatrix} \cos(M_{k-1}) \\ \sin(M_{k-1}) \\ \dots \\ n \\ \dots \end{pmatrix} \rightarrow M_{k-1} = \arctan\left(\frac{\sin(M_{k-1})}{\cos(M_{k-1})}\right)$$

$$\rightarrow M_k = M_{k-1} + n(t_k - t_{k-1}) \rightarrow x_k = \begin{pmatrix} \cos(M_k) \\ \sin(M_k) \\ \dots \end{pmatrix}$$

La trasformazione di uscita h , invece, prevede una fase iniziale di calcolo della posizione cartesiana. Dall'anomalia media si passa all'anomalia eccentrica, secondo la formula 3.10, poi dall'anomalia eccentrica all'anomalia vera, secondo la formula 3.12, e alla distanza radiale r , secondo la formula 3.13. Dunque, la posizione è data dalla formula 3.7.

Successivamente, dalla posizione baricentrica si ottiene la posizione topocentrica e si determinano i valori osservati. Anche in questo caso si fa uso di un servizio di effemeridi. Trovato $\hat{\Delta}$ come vettore normalizzato della posizione topocentrica e

Δ come il suo modulo, si sfrutta l'inverso della formula 4.4:

$$\sin(\delta) = z \quad (4.7)$$

$$\cos(\delta) = \sqrt{1 - \sin^2(\delta)} \quad (4.8)$$

$$\cos(\alpha) = \frac{x}{\cos(\delta)} \quad (4.9)$$

$$\sin(\alpha) = \frac{y}{\cos(\delta)} \quad (4.10)$$

Infine, la magnitudine visuale V è data dalla formula 3.5, utilizzando l'angolo di fase ricavato dalla formula 3.4.

4.5 Matrici di covarianza

La precisione del filtro di Kalman dipende fortemente dall'accuratezza delle matrici di covarianza Q e R utilizzate. Ovviamente non se ne conoscono i valori reali, ma analisi sia formali che statistiche possono fornire stime pressoché ottimali.

Per modelli semplici, Q ed R sono spesso matrici diagonali, ma non è questo il caso per il sistema considerato. In effetti, se si considera una generica quantità angolare θ con distribuzione di von Mises avente valore medio μ e parametro di concentrazione κ , approssimata da una distribuzione normale lungo la retta tangente avente varianza $\sigma^2 = \frac{1}{\kappa}$, si trova che la coppia $(\cos(\theta), \sin(\theta))$ ha la seguente matrice di auto-covarianza:

$$\begin{pmatrix} \sigma^2 \sin^2(\mu) & -\sigma^2 \sin(\mu) \cos(\mu) \\ -\sigma^2 \sin(\mu) \cos(\mu) & \sigma^2 \cos^2(\mu) \end{pmatrix} \quad (4.11)$$

Dunque, in prima approssimazione, Q è data da (prosegue):

$$Q = \begin{pmatrix} \sigma_M^2 \sin^2(M) & -\sigma_M^2 \sin(M) \cos(M) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\sigma_M^2 \sin(M) \cos(M) & \sigma_M^2 \cos^2(M) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_a^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_e^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_i^2 \sin^2(i) & -\sigma_i^2 \sin(i) \cos(i) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\sigma_i^2 \sin(i) \cos(i) & \sigma_i^2 \cos^2(i) & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \quad (4.12)$$

$$Q = \begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_\Omega^2 \sin^2(\Omega) & -\sigma_\Omega^2 \sin(\Omega) \cos(\Omega) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\sigma_\Omega^2 \sin(\Omega) \cos(\Omega) & \sigma_\Omega^2 \cos^2(\Omega) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_\omega^2 \sin^2(\omega) & -\sigma_\omega^2 \sin(\omega) \cos(\omega) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\sigma_\omega^2 \sin(\omega) \cos(\omega) & \sigma_\omega^2 \cos^2(\omega) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_n^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_H^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_G^2 \end{pmatrix} \quad (4.13)$$

La matrice R richiede ulteriori passaggi. Se si considera il versore di osservazione $\hat{\Delta}$ con distribuzione di Fisher avente parametro di concentrazione $\kappa_{\alpha,\delta}$, allora gli angoli α e δ sono approssimati da distribuzioni di von Mises con parametri di concentrazione $\kappa_{\alpha,\delta} \cos^2(\delta)$ e $\kappa_{\alpha,\delta}$ rispettivamente. Dunque, in prima approssimazione, R è data da:

$$R = \begin{pmatrix} \sigma_{\alpha,\delta}^2 \frac{\sin^2(\alpha)}{\cos^2(\delta)} & -\sigma_{\alpha,\delta}^2 \frac{\sin(\alpha) \cos(\alpha)}{\cos^2(\delta)} & 0 & 0 & 0 \\ -\sigma_{\alpha,\delta}^2 \frac{\sin(\alpha) \cos(\alpha)}{\cos^2(\delta)} & \sigma_{\alpha,\delta}^2 \frac{\cos^2(\alpha)}{\cos^2(\delta)} & 0 & 0 & 0 \\ 0 & 0 & \sigma_{\alpha,\delta}^2 \sin^2(\delta) & -\sigma_{\alpha,\delta}^2 \sin(\delta) \cos(\delta) & 0 \\ 0 & 0 & -\sigma_{\alpha,\delta}^2 \sin(\delta) \cos(\delta) & \sigma_{\alpha,\delta}^2 \cos^2(\delta) & 0 \\ 0 & 0 & 0 & 0 & \sigma_V^2 \end{pmatrix} \quad (4.14)$$

4.6 Simulazione

Inizialmente si è scelto di verificare l'efficacia del filtro utilizzando dati simulati. In primo luogo, il file `simulation/ideal_model_test.py` si occupa di simulare il moto orbitale, sia del corpo osservato, che del Sole e dell'osservatore, secondo le formule descritte nella sezione 3.3. Per garantire dati realistici, sono stati usati parametri orbitali reali, seppur successivamente considerati costanti, ottenuti dalla piattaforma JPL Horizons, in particolare accurati all'epoca J2000.0 tempo dinamico baricentrico. Per il corpo osservato si è considerato Bennu (1999 RQ36) e per l'osservatore si è considerato il baricentro Terra-Luna, garantendo sufficiente diversificazione nei dati.

Dalle posizioni dei tre soggetti, si determinano, poi, i valori osservati. Gli angoli α, δ sono ottenuti dalla direzione $\hat{\Delta}$, dopo che essa è stata perturbata secondo una distribuzione di Fisher (`scipy.stats.vonmises_fisher`). La magnitudine visuale V è ricavata dall'angolo di fase, a sua volta ricavato trigonometricamente, e perturbata secondo una distribuzione normale (`numpy.random.normal`).

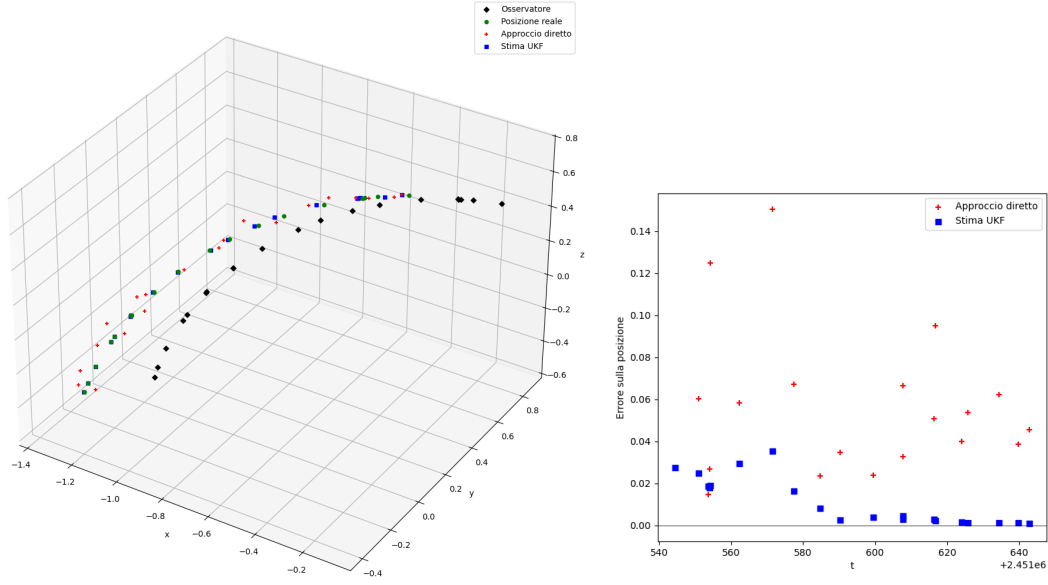


Figura 4.2: Visualizzazione delle stime sui dati simulati (primi 20 passi), confrontate con l’approccio diretto.

Ovviamente questa simulazione non modella la velocità finita della luce.

Una volta ottenute le osservazioni, separate da un lasso di tempo variabile, si crea un’istanza UKF con $\alpha = 10^{-3}$, $\beta = 2$, $\kappa = 0$ e vengono fornite la legge di evoluzione f e la trasformazione di uscita h ricavate nella sezione 4.4 e le matrici di covarianza R e Q della sezione 4.5. Le stime sono poi confrontate con le posizioni ottenute dall’approccio diretto discusso nella sezione 4.1.

In figura 4.2 si riportano i risultati sui primi 20 passi di simulazione, con una rappresentazione 3D delle posizioni e un grafico di andamento dell’errore, calcolato come $\left\| \begin{pmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{pmatrix} - \begin{pmatrix} x \\ y \\ z \end{pmatrix} \right\|$. Si evidenzia una graduale convergenza.

4.7 Applicazione a dati reali

Il Minor Planet Center raccoglie dati su osservazioni ottiche astrometriche di corpi minori, comete e satelliti naturali, effettuate da osservatori in tutto il mondo. I dati sono forniti senza alcun tipo di correzione, come specificato nelle linee guida (<https://minorplanetcenter.net/iau/info/Astrometry.html#corrs>).

Moduli Python esistenti come `astroquery` consentono di richiedere automaticamente le osservazioni; tuttavia si possono scaricare archivi più completi direttamente dalla piattaforma web https://www.minorplanetcenter.net/db_search.

Questi archivi sono file di testo in classico stile Fortran a 80 colonne e seguono il formato descritto alla pagina <https://www.minorplanetcenter.net/iau/info/OpticalObs.html>. In particolare, contengono le seguenti informazioni di nostro interesse:

- l'ascensione retta, espressa come angolo orario, minuti e secondi
- la declinazione, espressa in gradi, minuti e secondi
- la magnitudine osservata
- la banda utilizzata, ossia un'indicazione della fascia di frequenze elettromagnetiche considerate
- il codice dell'osservatorio

Il file `parse.py` si occupa dell'interpretazione dei dati e della trasformazione in oggetti `MinorPlanetObservation`. Questo passaggio richiede la conversione degli angoli in radianti e della magnitudine osservata in magnitudine visuale. In effetti, i modelli fotometrici si basano sulla magnitudine relativa alla luce visibile, per cui può essere necessario utilizzare la tabella di conversione fornita alla pagina <https://www.minorplanetcenter.net/iau/info/BandConversion.txt>.

Infine, il codice dell'osservatorio viene confrontato con la lista alla pagina <https://www.minorplanetcenter.net/iau/lists/ObsCodes.html>, per risalire alle coordinate geodetiche dell'osservatore. Anche in questo caso, un servizio di effemeridi può essere utilizzato per risalire alla posizione baricentrica dell'osservatore in un determinato istante date le coordinate geodetiche.

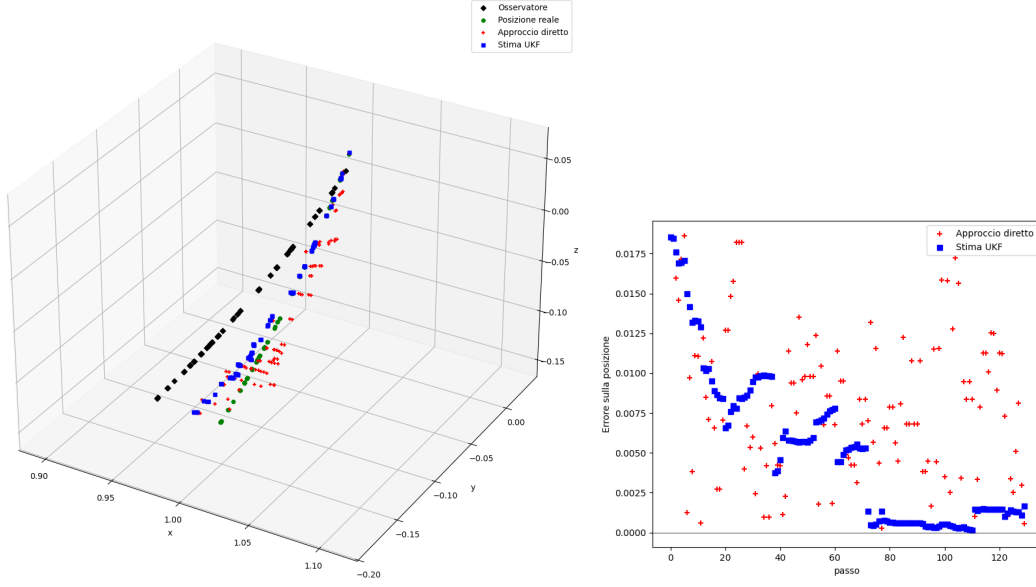


Figura 4.3: Visualizzazione delle stime sui dati reali dell'asteroide Bennu (prime 130 osservazioni), confrontate con l'approccio diretto.

Si procede, poi, come per i dati simulati, con un'istanza UKF con $\alpha = 10^{-3}$, $\beta = 2$, $\kappa = 0$. In figura 4.3 si riportano i risultati sulle prime 130 osservazioni. Anche in questo caso si osserva una generale convergenza, seppur decisamente più lenta rispetto alla simulazione. Inoltre, si è osservato che il processo non risulta perfettamente stabile; su altre porzioni di dati, lievi discostamenti improvvisi portano a rapida divergenza. Tutto ciò potrebbe essere dovuto a diversi motivi, possibilmente legati a problemi di stabilità numerica, per cui studi ulteriori sono necessari. In ogni caso, si è mostrato come il filtro permetta di ottenere risultati dalle 10 alle 100 volte più accurati rispetto all'approccio diretto, nonostante le diverse semplificazioni che sono state fatte, dando prova concreta dell'utilità del filtro di Kalman in un contesto reale.

Capitolo 5

Conclusioni

TODO

Bibliografia

- [1] R. Kalman, “A new approach to linear filtering and prediction problems,” *Journal of Basic Engineering*, vol. 82, pp. 35–45, Mar. 1960.
- [2] G. Welch and G. Bishop, “An introduction to the kalman filter,” Tech. Rep. 95-041, University of North Carolina at Chapel Hill, 1995.
- [3] Y. Pei, S. Biswas, D. Fussell, and K. Pingali, “An elementary introduction to kalman filtering,” *Communications of the ACM*, vol. 62, p. 122–133, Oct. 2019.
- [4] A. Gelb and T. A. S. Corporation, *Applied Optimal Estimation*. The MIT Press, 1974.
- [5] H. Masnadi-Shirazi, A. Masnadi-Shirazi, and M. Dastgheib, “A step by step mathematical derivation and tutorial on kalman filters.” ArXiv Preprint ArXiv:1910.03558, Oct. 2019.
- [6] G. Smith, S. Schmidt, and L. McGee, “Application of statistical filter theory to the optimal estimation of position and velocity on board a circumlunar vehicle,” Tech. Rep. R-135, National Aeronautics and Space Administration, 1962.
- [7] E. Wan and R. Van Der Merwe, “The unscented kalman filter for nonlinear estimation,” in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium*, pp. 153–158, 2000.
- [8] A. Yang and S. Boyd, “Iteratively saturated kalman filtering,” 2025.

- [9] N. Shlezinger, G. Revach, A. Ghosh, S. Chatterjee, S. Tang, T. Imbiriba, J. Dunik, O. Straka, P. Closas, and Y. C. Eldar, “Ai-aided kalman filters,” 2025.
- [10] E. Bowell, B. Hapke, D. Domingue, K. Lumme, J. Peltoniemi, and A. Harris, “Application of photometric models to asteroids,” in *Asteroids II*, pp. 524–556, The University of Arizona Press, 1989.
- [11] R. Dymock, “The H and G magnitude system for asteroids,” *Journal of the British Astronomical Association*, vol. 117, pp. 342–343, Nov. 2007.