Mario Carpino

Introduzione ai metodi di calcolo di effemeridi e determinazione orbitale

Versione preliminare (3 maggio 2010)

Indice

Capitolo 1

Le equazioni del moto e il problema dei due corpi

1.1 Le equazioni del moto

Una classe molto vasta di sistemi dinamici di interesse astronomico è descrivibile con un modello di forze molto semplice, in cui un insieme finito di punti massa si muove sotto l'effetto della sola attrazione reciproca secondo la legge dell'inverso del quadrato della distanza[(1)]. Se consideriamo un sistema di questo tipo, costituito da $N+1$ punti materiali, le equazioni dinamiche che ne descrivono il moto sono allora:

$$\frac{d^2\mathbf{R}_i}{dt^2} = G\sum_{j=0}^{N} m_j \frac{\mathbf{r}_{ij}}{r_{ij}^3} \quad (j \neq i,\ i = 0, 1, \ldots N)\,, \tag{1.1}$$

dove G è la costante universale di gravitazione, $\mathbf{R}_i$ sono le coordinate (vettori posizione) degli $N+1$ punti ($i = 0, 1, \ldots N$), m_i le loro masse e $\mathbf{r}_{ij} = \mathbf{R}_j - \mathbf{R}_i$ i vettori posizione mutua. È incredibile la complicazione e la varietà di configurazioni che possono nascere da un'equazione così semplice!

L'equazione (1.1) è ovviamente valida in un sistema di riferimento inerziale (i cui assi possono essere scelti a piacere); siccome tuttavia le osservazioni forniscono sempre misure relative, è più comune assumere come origine delle coordinate la posizione di uno dei corpi (diciamo m_0) e scrivere le equazioni del moto per i vettori posizione relativa $\mathbf{r}_i = \mathbf{R}_i - \mathbf{R}_0$. Nel caso del Sistema Solare usualmente m_0 è il Sole; sottraendo membro a membro l'espressione (1.1) si ottengono le equazioni del moto in *coordinate eliocentriche*

$$\frac{d^2\mathbf{r}_i}{dt^2} = -G(m_0 + m_i)\frac{\mathbf{r}_i}{r_i^3} + G\sum_{j=1}^{N} m_j \left(\frac{\mathbf{r}_j - \mathbf{r}_i}{r_{ij}^3} - \frac{\mathbf{r}_j}{r_j^3}\right) \quad (j \neq i,\ i = 1, \ldots N)\,. \tag{1.2}$$

[(1)] Le discrepanze dalla legge dell'inverso del quadrato della distanza dovute all'estensione finita dei corpi sono in genere molto piccole per due ragioni: a) è possibile dimostrare che un corpo esteso in cui le masse sono disposte con simmetria sferica produce lo stesso campo gravitazionale di un corpo puntiforme avente la stessa massa totale, posto nel centro di massa del corpo; quindi per produrre una deviazione dalla legge $1/r^2$ occorre non solo che il corpo abbia grandi dimensioni, ma che la sua forma si discosti significativamente dalla sfera (e, come si sa, i pianeti hanno forma quasi sferica); b) i termini della forza di attrazione dovuti alla forma finita del corpo diminuiscono con la distanza come $1/r^k$ con $k > 2$, e quindi a distanze sufficientemente grandi diventano trascurabili rispetto al termine principale $1/r^2$. Di fatto occorre tener conto di questi effetti solo nel caso di satelliti abbastanza vicini al pianeta attorno a cui ruotano (come nel caso della Luna o dei satelliti di Giove); nel moto dei pianeti principali del Sistema Solare questi termini sono del tutto trascurabili. Altrettanto trascurabili sono generalmente tutti i tipi di forza di origine non gravitazionale: solo in caso di asteroidi molto piccoli (pochi chilometri di diametro) è a volte rilevabile una perturbazione orbitale dovuta all'*effetto Yarkovski* (emissione anisotropa di radiazione infrarossa).

Le differenze rispetto all'equazione (1.1) sono due: 1) il termine centrale contiene la somma delle masse del Sole e del pianeta $m_0 + m_i$; 2) compare un nuovo termine perturbativo $-Gm_j\mathbf{r}_j/r_j^3$ (detto *perturbazione indiretta*) che rappresenta l'effetto dell'attrazione degli altri pianeti sul Sole (in altre parole è una *forza apparente* introdotta dalla non-inerzialità del sistema di riferimento). È noto che la forza kepleriana è conservativa e può essere espressa come il gradiente di un potenziale; ciò è vero anche per il secondo membro dell'equazione (1.2), che può essere scritta

$$\frac{d^2\mathbf{r}_i}{dt^2} = \nabla_i(U_i + R_i)\,, \tag{1.3}$$

dove ∇_i è il gradiente rispetto a $\mathbf{r}_i$ e

$$U_i = \frac{G(m_0 + m_i)}{r_i}, \qquad R_i = G\sum_{j=1}^{N} m_j \left(\frac{1}{r_{ij}} - \frac{\mathbf{r}_i \cdot \mathbf{r}_j}{r_j^3}\right) \quad (j \neq i) \tag{1.4}$$

sono rispettivamente il termine centrale del potenziale e la *funzione perturbatrice*. Anche per la funzione perturbatrice si può notare che, accanto al termine diretto (proporzionale al reciproco della distanza mutua) che esprime l'attrazione degli altri pianeti sul pianeta perturbato, compare un termine indiretto.

Le equazioni (1.1) possiedono dieci integrali del moto: di questi, sei sono collegati col moto del centro di massa $\mathbf{B}$ che, come per ogni sistema isolato, si muove di moto rettilineo uniforme:

$$\mathbf{B}(t) = \frac{\sum_{i=0}^{N} m_i\mathbf{R}_i(t)}{\sum_{i=0}^{N} m_i} = \dot{\mathbf{B}}t + \mathbf{B}_0 \quad (\dot{\mathbf{B}}, \mathbf{B}_0 \text{ costanti})\,; \tag{1.5}$$

gli altri quattro sono rappresentati dalle tre componenti del momento angolare totale e dall'energia totale. Si noti che nel passaggio dalla formulazione inerziale (1.1) a quella eliocentrica (1.2) delle equazioni del moto si eliminano dal problema le sei coordinate (posizione e velocità) del Sole ma si perdono anche i sei integrali del centro di massa, per cui il numero di gradi di libertà del sistema rimane invariato.

1.2 Il problema di Keplero

Si chiama *problema di Keplero* il caso particolare dell'equazione (1.1) in cui $N = 1$ (si hanno cioè solo due corpi, ad esempio il Sole e un pianeta). In coordinate eliocentriche il problema è descritto dall'equazione

$$\frac{d^2\mathbf{r}}{dt^2} = -\mu\frac{\mathbf{r}}{r^3}\,, \tag{1.6}$$

dove $\mu = G(m_0 + m_1)$ e dove è sottinteso l'indice $i = 1$. In altre parole *il moto di un sistema isolato di due particelle sottoposte alla mutua attrazione kepleriana è equivalente al moto di una sola particella in un campo centrale prodotto da una massa pari alla somma delle due masse.*

La soluzione dell'equazione (1.6) può essere ottenuta analiticamente. Cominciamo con il ridurre il numero di gradi di libertà sfruttando gli integrali primi; utilizziamo per primo il momento angolare

$$\mathbf{h} = \mathbf{r} \times \dot{\mathbf{r}}$$

che è una costante del moto, come è logico trattandosi di forze centrali e come è immediato dimostrare direttamente dalla (1.6). Per comodità di notazione introduciamo un sistema di riferimento rotante definito dai tre versori $\mathbf{u}_r$ (parallelo a $\mathbf{r}$), $\mathbf{u}_h$ (parallelo a $\mathbf{h}$) e $\mathbf{u}_\theta$ (tale che $\mathbf{u}_r \times \mathbf{u}_\theta = \mathbf{u}_h$). Naturalmente l'asse $\mathbf{u}_h$ è fisso nello spazio, perché $\mathbf{h}$ è costante; ne consegue che il moto avviene in un piano (passante per l'origine e ortogonale a $\mathbf{u}_h$) e può quindi essere descritto da due sole coordinate polari $r = |\mathbf{r}|$ e θ (angolo misurato da un asse

fisso, scelto a piacere nel piano); in questo sistema di riferimento i vettori posizione, velocità e momento angolare si scrivono

$$\begin{aligned} \mathbf{r} &= r\mathbf{u}_r\,, \\ \dot{\mathbf{r}} &= \dot{r}\mathbf{u}_r + r\dot{\theta}\mathbf{u}_\theta\,, \\ \mathbf{h} &= r\mathbf{u}_r \times (\dot{r}\mathbf{u}_r + r\dot{\theta}\mathbf{u}_\theta) = r^2\dot{\theta}\mathbf{u}_h\,. \end{aligned} \tag{1.7}$$

Le componenti $\mathbf{u}_r$ e $\mathbf{u}_\theta$ dell'equazione del moto (1.6) sono rispettivamente

$$\begin{aligned} \ddot{r} - r\dot{\theta}^2 &= -\frac{\mu}{r^2}\,, \\ \frac{1}{r}\frac{d}{dt}(r^2\dot{\theta}) &= 0\,. \end{aligned} \tag{1.8}$$

La seconda delle (1.8) esprime semplicemente la costanza del modulo del vettore momento angolare e porta direttamente all'integrale primo

$$r^2\dot{\theta} = h\,. \tag{1.9}$$

Poichè $r^2\dot{\theta}/2 = dA/dt$ è la *velocità areolare* del pianeta, la (1.9) è equivalente alla seconda legge di Keplero (*il raggio vettore Sole-pianeta spazza aree uguali in tempi uguali*). Si noti che questo risultato è conseguenza unicamente della conservazione del momento angolare e pertanto è valido per qualsiasi campo di forza centrale.

Il problema di Keplero si riduce quindi alla soluzione del sistema formato dalla prima delle (1.8) e dalla (1.9). Come passo intermedio si può ricavare la forma della traiettoria: operando il cambiamento di variabile $u = 1/r$ ed eliminando il tempo t tra le due equazioni, si ottiene

$$\frac{d^2u}{d\theta^2} + u = \frac{\mu}{h^2}\,,$$

che ha come soluzione generale

$$u = \frac{\mu}{h^2} + K\cos(\theta - \theta_0)$$

(K e θ_0 sono due costanti di integrazione); reintroducendo r e definendo

$$p = \frac{h^2}{\mu}\,, \qquad e = \frac{Kh^2}{\mu} \tag{1.10}$$

la soluzione si può scrivere

$$r = \frac{p}{1 + e\cos(\theta - \theta_0)}\,, \tag{1.11}$$

dove e e θ_0 sono due nuove costanti di integrazione[(2)]. Come è noto, la (1.11) è l'equazione di una conica (un'ellisse se $0 \le e < 1$, una parabola se $e = 1$ o un'iperbole se $e > 1$) rispetto a uno dei fuochi; θ_0 definisce la direzione nel piano orbitale (*linea degli apsidi*) in cui si ha il punto di massimo avvicinamento del pianeta al Sole (*perielio* o, più genericamente, *pericentro*); è usuale assumere questa direzione come origine della coordinata polare angolare e scrivere la (1.11) nella forma

$$r = \frac{p}{1 + e\cos f}\,, \tag{1.12}$$

dove l'angolo $f = \theta - \theta_0$ è chiamato *anomalia vera*. Dunque la prima legge di Keplero (*i pianeti descrivono ellissi di cui il Sole occupa uno dei fuochi*) descrive solo un caso particolare

[(2)] Si noti che il fatto che $r(\theta)$ sia una funzione periodica di periodo 2π implica che il pianeta ritorna allo stesso punto nello spazio dopo una rivoluzione, cioè che l'orbita è una *traiettoria chiusa*; ciò non è vero per un generico potenziale centrale, neppure della forma r^k, tranne che nel caso kepleriano $k = -1$ e nel caso armonico $k = 2$ (*teorema di Bertrand*).

della soluzione della (1.6). Per completezza notiamo anche che le *orbite rettilinee* (soluzioni caratterizzate da $h = 0$) sono escluse dalla presente trattazione.

Il parametro p che compare nella (1.12) è detto *semilatus rectum*, perché è la distanza Sole-pianeta quando il raggio vettore è ortogonale alla linea degli apsidi ($f = \pi/2,\ 3\pi/2$). Nel caso di un'orbita ellittica, le distanze al perielio r_p e all'afelio r_a sono date da

$$r_p = \frac{p}{1+e}\,, \qquad r_a = \frac{p}{1-e}\,, \tag{1.13}$$

mentre il semiasse maggiore dell'ellisse vale

$$a = \frac{r_p + r_a}{2} = \frac{p}{1-e^2}\,, \tag{1.14}$$

cioè

$$p = a(1-e^2)\,; \quad r_p = a(1-e)\,; \quad r_a = a(1+e)\,. \tag{1.15}$$

1.3 L'integrale dell'energia

Un altro integrale primo della (1.6) è l'energia totale[3]

$$C = \frac{1}{2}v^2 - \frac{\mu}{r} \tag{1.16}$$

(dove $v^2 = \dot{\mathbf{r}} \cdot \dot{\mathbf{r}}$), come si può verificare direttamente calcolandone la derivata temporale e tenendo conto della (1.6)

$$\frac{dC}{dt} = \dot{\mathbf{r}} \cdot \ddot{\mathbf{r}} - \nabla\left(\frac{\mu}{r}\right) \cdot \dot{\mathbf{r}} = \dot{\mathbf{r}} \cdot \left(\ddot{\mathbf{r}} + \frac{\mu}{r^3}\mathbf{r}\right) = 0\,.$$

L'energia C *dipende solamente dal semiasse maggiore dell'orbita* e non dagli altri parametri orbitali. Ricaviamo questa importante proprietà, anche se la dimostrazione risulta un po' tediosa. Utilizzando le (1.7) e (1.9) ricaviamo

$$v^2 = \dot{r}^2 + r^2\dot{\theta}^2 = \dot{r}^2 + \frac{h^2}{r^2}\,. \tag{1.17}$$

Cerchiamo ora delle espressioni per $\dot{r}^2$ e h^2/r^2 che dipendano solo da r ed f. Derivando la (1.12) rispetto al tempo si ottiene

$$\dot{r} = \frac{pe\dot{f}\sin f}{(1+e\cos f)^2} = \frac{r^2 e\dot{f}\sin f}{p} = \frac{h}{p}\,e\sin f\,, \tag{1.18}$$

dove si è sostituito $h = r^2\dot{f}$ (equazione (1.9)). Dalla (1.12) si ricava poi

$$\frac{h^2}{r^2} = \frac{h^2}{p^2}(1+e\cos f)^2\,. \tag{1.19}$$

Sostituendo le (1.18) e (1.19) nella (1.17) e tenendo conto ancora della (1.12) si ottiene

$$\begin{aligned} v^2 &= \frac{h^2}{p^2}\Big[1+e^2+2e\cos f\Big] = \frac{h^2}{p^2}\Big[2(1+e\cos f) - (1-e^2)\Big] \\ &= \frac{2h^2}{rp} - \frac{h^2}{p^2}(1-e^2) \end{aligned}$$

[3] A rigore l'espressione (1.16) non è l'energia del sistema di due corpi (1.1) ma quella del sistema fittizio (1.6) (particella di massa nulla in un campo centrale generato da una massa $m_0 + m_1$); si può comunque dimostrare che l'energia del sistema (1.1), calcolata nel sistema di riferimento baricentrico, coincide con C a meno di una costante di proporzionalità.

e, tenendo conto delle (1.10) e (1.14),

$$v^2 = \mu \left(\frac{2}{r} - \frac{1}{a} \right) . \tag{1.20}$$

Sostituendo la (1.20) nella (1.16) si ha infine

$$C = -\frac{\mu}{2a} \tag{1.21}$$

che è la relazione cercata.

1.4 La legge oraria (orbita ellittica)

L'equazione (1.12) fornisce la forma dell'orbita ma non dà indicazioni sui tempi in cui essa viene percorsa, cioè sulla legge oraria. Limitiamoci al caso di un'orbita ellittica (i casi parabolico e iperbolico richiederebbero una trattazione a parte). Il periodo orbitale T può essere ricavato direttamente dalla velocità areolare $dA/dt = h/2$ e dall'area dell'ellisse $A = \pi ab$

$$T = \frac{A}{dA/dt} = \frac{\pi ab}{h/2} ,$$

dove

$$b = a\sqrt{1 - e^2} \tag{1.22}$$

è il *semiasse minore* dell'ellisse. Esprimendo h e b in funzione del semiasse maggiore a (equazioni (1.10), (1.15) e (1.22)) si ottiene

$$T = \frac{2\pi}{\sqrt{\mu}} \, a^{3/2} , \tag{1.23}$$

che costituisce la terza legge di Keplero (*i quadrati dei periodi orbitali dei pianeti sono proporzionali ai cubi dei semiassi maggiori*)[(4)]. Indicando con $n = 2\pi/T$ la velocità angolare media del pianeta sull'orbita (detta *moto medio*), dalla (1.23) si ricava l'equazione fondamentale che lega il moto medio al semiasse maggiore

$$n^2 a^3 = \mu . \tag{1.24}$$

Per quanto riguarda la legge oraria, dalle (1.17) e (1.20) si ricava

$$r\frac{dr}{dt} = \sqrt{v^2 r^2 - h^2} = \sqrt{2\mu r - \frac{\mu r^2}{a} - h^2}$$

e, tenendo conto delle (1.10), (1.15) e (1.24),

$$\begin{aligned} r\frac{dr}{dt} &= \sqrt{2\mu r - \frac{\mu r^2}{a} - \mu a(1 - e^2)} = na\sqrt{2ar - r^2 - a^2(1 - e^2)} \\ &= na\sqrt{a^2 e^2 - (r - a)^2} . \end{aligned}$$

La determinazione di r in funzione del tempo è perciò ricondotta ad una integrazione diretta

$$n\, dt = \frac{r\, dr}{a\sqrt{a^2 e^2 - (r - a)^2}} \tag{1.25}$$

[(4)] Si noti tuttavia che, in coordinate eliocentriche, la legge così enunciata è valida solo approssimativamente, in quanto la costante μ contiene la somma delle masse del Sole e del pianeta, e quindi è leggermente diversa da pianeta a pianeta.

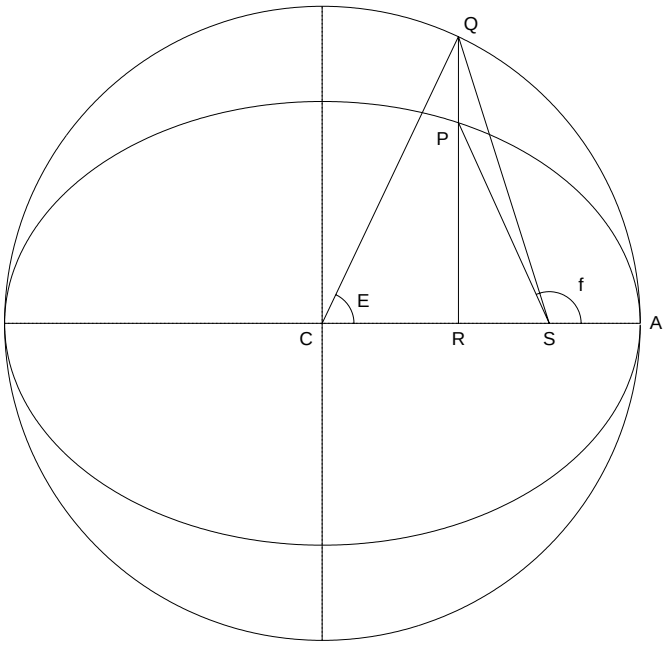

Figura 1: Costruzione geometrica per derivare l'equazione di Keplero.

che può essere eseguita facilmente introducendo una variabile ausiliaria E (detta *anomalia eccentrica*), definita da

$$r - a = -ae\cos E\,, \qquad dr = ae\sin E\,dE\,. \tag{1.26}$$

Con questa sostituzione la (1.25) diventa

$$n\,dt = (1 - e\cos E)\,dE$$

che ha per soluzione

$$n(t - t_0) = E - e\sin E \tag{1.27}$$

(relazione nota come *equazione di Keplero*), dove t_0 è una costante di integrazione. L'angolo definito da

$$M = n(t - t_0) \tag{1.28}$$

è detto *anomalia media*. Dalle equazioni (1.12), (1.26) e (1.28) si deduce che il valore delle tre anomalie (media, eccentrica e vera) coincide quando il pianeta passa al perielio o all'afelio

$$\begin{array}{rcccccll} M & = & E & = & f & = & 0 + 2k\pi\ (k = 0, 1, 2, \ldots) & \text{per } r = r_p = a(1-e)\,; \\ M & = & E & = & f & = & \pi + 2k\pi\ (k = 0, 1, 2, \ldots) & \text{per } r = r_a = a(1+e)\,; \end{array}$$

in particolare, i tre angoli compiono una rivoluzione nel medesimo tempo, con velocità angolare media pari al moto medio n. Ne consegue anche che la costante t_0 nella (1.27) definisce un istante di tempo in cui $M = 0$, cioè un'*epoca di passaggio al perielio*.

1.5 Derivazione geometrica della legge oraria

L'equazione di Keplero può essere dedotta anche per mezzo di un'elegante costruzione geometrica, che utilizza solamente le prime due leggi di Keplero e ha il vantaggio di mostrare

il significato dell'anomalia vera ed eccentrica. Nel piano orbitale scegliamo un sistema di riferimento (ξ, η) in cui l'asse ξ è diretto lungo la linea degli apsidi (perielio) e tracciamo un cerchio di raggio a circoscritto all'ellisse (Figura 1).

Indichiamo con S il fuoco dell'ellisse occupato dal Sole, con A la posizione del perielio e con P la posizione del pianeta sull'orbita. La seconda legge di Keplero dice che l'area del settore di ellisse SPA è proporzionale al tempo, cioè all'anomalia media; deve perciò essere

$$\text{area } SPA = \pi ab\,\frac{M}{2\pi} = \frac{1}{2}abM$$

(quando il pianeta ha compiuto una rivoluzione completa, l'area spazzata è uguale all'area totale dell'ellisse πab). Tracciamo ora una retta passante per P e ortogonale all'asse maggiore dell'ellisse (retta SA) e indichiamo rispettivamente con R e Q le intersezioni di tale retta con l'asse maggiore e il cerchio circoscritto. Si noti che il triangolo mistilineo SQA si può ottenere da SPA moltiplicando la dimensione ortogonale all'asse maggiore di SPA per un fattore a/b (l'ellisse è un "cerchio schiacciato"), per cui risulta anche

$$\text{area } SQA = \frac{a}{b}\,\text{area } SPA = \frac{1}{2}a^2 M\,.$$

Introduciamo ora l'anomalia eccentrica E come l'angolo (misurato dal *centro* dell'ellisse C) tra l'asse maggiore e Q. ACQ è un settore circolare e dunque

$$\text{area } ACQ = \frac{1}{2}a^2 E\,; \tag{1.29}$$

d'altra parte ACQ è l'unione di SQA con il triangolo CQS, da cui

$$\text{area } ACQ = \text{area } SQA + \text{area } CQS = \frac{1}{2}a^2 M + \frac{1}{2}a^2 e\sin E \tag{1.30}$$

(la distanza CS tra il centro e il fuoco di un'ellisse è uguale ad ae). Eguagliando le (1.29) e (1.30) si ottiene nuovamente l'equazione di Keplero

$$M = E - e\sin E\,. \tag{1.31}$$

Le coordinate del pianeta nel piano dell'orbita sono date da

$$\begin{array}{rclcl} \xi & = & \vec{SR} & = & \vec{SC} + \vec{CR} \;=\; -ae + a\cos E \\ \eta & = & \vec{RP} & = & \dfrac{b}{a}\,\vec{RQ} \;=\; b\sin E\,. \end{array} \tag{1.32}$$

Il confronto di queste relazioni con le coordinate polari

$$\xi = r\cos f\,, \qquad \eta = r\sin f \tag{1.33}$$

permette di stabilire una relazione diretta tra anomalia vera e anomalia eccentrica

$$\tan\left(\frac{f}{2}\right) = \sqrt{\frac{1+e}{1-e}}\,\tan\left(\frac{E}{2}\right)\,. \tag{1.34}$$

1.6 Il vettore di Laplace-Lenz

Un altro integrale primo dell'equazione (1.6) è il *vettore di Laplace-Lenz*, definito come

$$\mathbf{e} = \frac{1}{\mu}\dot{\mathbf{r}} \times \mathbf{h} - \mathbf{u}_r\,. \tag{1.35}$$

Per dimostrare che si tratta effettivamente di una costante del moto, calcoliamo

$$\mu\dot{\mathbf{e}} = \ddot{\mathbf{r}} \times \mathbf{h} - \mu\dot{\mathbf{u}}_r \,. \tag{1.36}$$

Introducendo nella (1.36) l'equazione del moto (1.6) e ponendo in evidenza i versori, e sostituendo

$$\dot{\mathbf{u}}_r = \dot{\theta}\mathbf{u}_\theta \,,$$

la (1.36) diventa

$$\mu\dot{\mathbf{e}} = -\frac{\mu h}{r^2}\mathbf{u}_r \times \mathbf{u}_h + \mu\dot{\theta}\mathbf{u}_r \,. \tag{1.37}$$

Sostituendo nella (1.37)

$$\mathbf{u}_r \times \mathbf{u}_h = -\mathbf{u}_\theta \,,$$

si ha

$$\mu\dot{\mathbf{e}} = -\mu\mathbf{u}_\theta \left[\frac{h}{r^2} - \dot{\theta}\right] , \tag{1.38}$$

che è uguale a zero in virtù della (1.9). Notiamo perciò che, mentre l'integrale dell'energia e del momento angolare valgono per vaste classi di sistemi dinamici (rispettivamente per campi di forza conservativi e centrali), il vettore di Laplace-Lenz è un integrale del moto caratteristico del problema dei due corpi. Ricaviamo ora una espressione alternativa di $\mathbf{e}$ che permette di comprenderne meglio il significato geometrico. Dalla definizione (1.35) segue

$$\mathbf{e} = \frac{r^2\dot{\theta}}{\mu}\Big[\dot{r}\mathbf{u}_r \times \mathbf{u}_h + r\dot{\theta}\mathbf{u}_\theta \times \mathbf{u}_h\Big] - \mathbf{u}_r = \left[\frac{r^3\dot{\theta}^2}{\mu} - 1\right]\mathbf{u}_r - \frac{\dot{r}r^2\dot{\theta}}{\mu}\mathbf{u}_\theta$$

e, tenendo conto delle (1.9), (1.10), (1.18) e (1.12),

$$\begin{aligned} \mathbf{e} &= \left[\frac{h^2}{\mu r} - 1\right]\mathbf{u}_r - \frac{\dot{r}h}{\mu}\mathbf{u}_\theta = \left[\frac{p}{r} - 1\right]\mathbf{u}_r - \frac{h^2}{p\mu}\, e\sin f\, \mathbf{u}_\theta \\ &= e\cos f\, \mathbf{u}_r - e\sin f\, \mathbf{u}_\theta \,. \end{aligned} \tag{1.39}$$

Questa espressione dimostra che il vettore di Laplace-Lenz ha modulo pari all'eccentricità orbitale e punta nella direzione del perielio (infatti l'angolo compreso tra $\mathbf{e}$ ed $\mathbf{r}$ è uguale all'anomalia vera f).

Ricapitolando, abbiamo ricavato sei integrali primi indipendenti dell'equazione del moto (1.6): l'energia totale, tre componenti del momento angolare totale e due componenti del vettore di Laplace-Lenz (la terza componente non è indipendente, perchè $\mathbf{e}$ giace nel piano orbitale e perciò è ortogonale al momento angolare $\mathbf{h}$, cioè è vincolato dalla condizione $\mathbf{e} \cdot \mathbf{h} = 0$).

1.7 Soluzione numerica dell'equazione di Keplero

Nella maggior parte delle applicazioni l'equazione di Keplero (1.31) deve essere risolta rispetto a E. Il caso più normale è quello del calcolo delle *effemeridi*[5] di un pianeta, di cui si conoscono gli elementi orbitali e si vuole determinare la posizione a un certo istante t; nella (1.31) sono dunque noti M (dalla (1.28)) ed e e si vuole determinare E.

[5] Il termine *effemeride* (dal greco *ephemeros*, "giornaliero") indica una tavola in cui si riportano in anticipo (solitamente da un anno per l'altro e, appunto, con periodocità giornaliera) le coordinate celesti del Sole, della Luna e dei pianeti principali. In un'epoca in cui non era disponibile a tutti *software* che permette di calcolare velocemente le posizioni dei corpi celesti (cioè, in realtà, fino a pochi anni fa) le effemeridi erano uno strumento essenziale sia per gli astronomi che dovevano pianificare le osservazioni, sia per i marinai che dovevano "fare il punto" della nave.

Si possono trovare in letteratura decine di metodi diversi per la soluzione di questo problema; la maggior parte di questi si basa su un processo iterativo, che è composto da due ingredienti:

1) una *formula di partenza* che fornisce un valore approssimato E_0;

2) una *formula iterativa* che corregge un valore approssimato E_n fornendo un nuovo valore E_{n+1} più vicino alla soluzione.

Con questi due ingredienti è possibile costruire una successione di valori E_0, E_1, ..., E_n, ... che converga verso la soluzione esatta: il procedimento viene arrestato dopo un numero finito di passi, quando l'errore è sceso al disotto di un valore prefissato. Senza entrare in particolari, citiamo due metodi iterativi molto semplici, consigliabili per valori di eccentricità non troppo elevati. Il primo è il metodo di Newton-Raphson, che cerca la soluzione di una equazione del tipo $F(E) = 0$ approssimando la funzione F con il suo sviluppo di Taylor troncato al primo termine

$$F(E_{n+1}) \simeq F(E_n) + F'(E_n)(E_{n+1} - E_n)\,;$$

imponendo $F(E_{n+1}) = 0$ si ottiene

$$E_{n+1} = E_n - \frac{F(E_n)}{F'(E_n)}$$

che, nel caso dell'equazione di Keplero

$$F(E) = E - e\sin E - M$$

diventa

$$E_{n+1} = E_n + \frac{M - E_n + e\sin E_n}{1 - e\cos E_n}\,. \tag{1.40}$$

Un altro metodo molto usato (anche se converge un po' più lentamente) consiste nel porre semplicemente

$$E_{n+1} = M + e\sin E_n\,; \tag{1.41}$$

esso si basa sul fatto che l'applicazione $E_n \to E_{n+1}$ così definita è una contrazione, perciò la successione $\{E_n\}$ tende a un limite che è il punto unito della (1.41), cioè la soluzione cercata.

Come valore di partenza per le iterazioni si può adottare per entrambi i metodi il valore[6]

$$E_0 = M\,. \tag{1.42}$$

1.8 L'orbita nello spazio

Ci siamo fino a ora occupati del moto del pianeta nel suo piano orbitale; in particolare le equazioni (1.32) forniscono la soluzione nel *sistema di riferimento orbitale* (ξ, η, ζ), in cui il piano (ξ, η) coincide con il piano dell'orbita (ζ è parallelo al momento angolare $\mathbf{h}$) e l'asse ξ è diretto verso il perielio. Per passare alla posizione del pianeta in un generico sistema di riferimento (x_1, x_2, x_3) occorre conoscere l'orientazione degli assi (ξ, η, ζ) rispetto agli assi (x_1, x_2, x_3); questa è descritta, come è usuale in geometria, da una terna di angoli di Eulero che, nel caso del problema dei due corpi, assumono nomi particolari (Figura 2):

[6] Bisogna però notare che, con il valore di partenza (1.42), il metodo di Newton-Raphson non è convergente per valori dell'eccentricità maggiori di circa 0.9733.

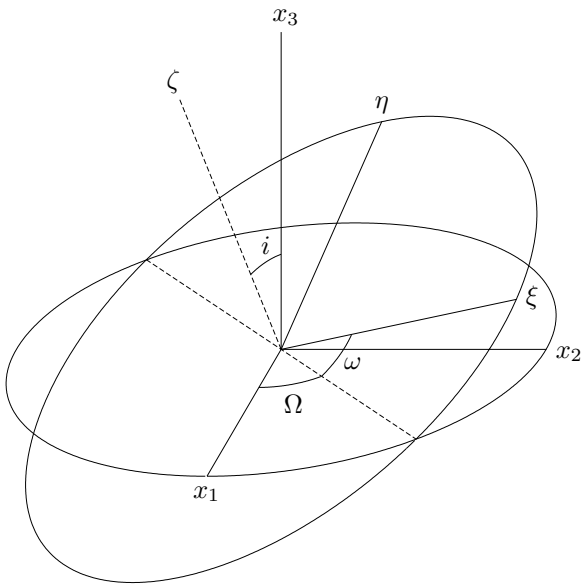

Figura 2: Elementi che definiscono l'orientazione dell'orbita nello spazio.

- l'*inclinazione* orbitale i è l'angolo tra il piano orbitale e il piano (x_1, x_2), cioè tra il vettore momento angolare $\mathbf{h}$ e l'asse x_3;

- la *longitudine del nodo ascendente* Ω è l'angolo tra l'asse x_1 e la linea dei nodi (intersezione del piano orbitale con il piano (x_1, x_2)), misurato sul piano (x_1, x_2) in senso diretto (antiorario), dall'asse x_1 al nodo ascendente (punto in cui il pianeta attraversa il piano (x_1, x_2) passando da valori negativi a valori positivi di x_3);

- l'*argomento del perielio* ω è l'angolo tra la linea dei nodi e la linea degli apsidi, misurato sul piano orbitale in senso diretto, dal nodo ascendente alla direzione del perielio.

In alternativa all'argomento del perielio ω viene a volte usata la *longitudine del perielio*[7], definita come $\varpi = \Omega + \omega$.

È comodo esprimere la trasformazione tra (ξ, η, ζ) e (x_1, x_2, x_3) in funzione delle matrici di rotazione; si chiama *matrice di rotazione* $\mathbf{R}_j(\alpha)$ la matrice che trasforma la terna di coordinate (x_1, x_2, x_3) nelle coordinate (x_1', x_2', x_3') associate ad un nuovo sistema di riferimento,

[7] Riguardo alla nomenclatura degli angoli bisogna osservare che, nell'uso tradizionale

- vengono chiamati *anomalie* gli angoli misurati nel piano orbitale a partire dalla linea degli apsidi (direzione del perielio);

- vengono chiamati *argomenti* gli angoli misurati nel piano orbitale a partire dalla linea dei nodi (nodo ascendente);

- vengono chiamati *longitudini* gli angoli misurati nel piano fondamentale (x_1, x_2) del sistema di riferimento inerziale a partire dall'asse x_1, anche quando (come nel caso della longitudine del perielio) l'angolo sia in realtà formato dalla somma di più termini, di cui solo il primo è misurato a partire dall'asse x_1.

ruotato di un angolo α attorno all'asse x_j (in senso diretto)

$$\begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} = \mathbf{R}_j(\alpha) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} .$$

La matrice di rotazione $\mathbf{R}_3(\alpha)$ attorno all'asse x_3 è data da

$$\mathbf{R}_3(\alpha) = \begin{pmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} ; \tag{1.43}$$

le matrici di rotazione attorno agli altri assi possono essere ricavate da $\mathbf{R}_3(\alpha)$ per permutazione ciclica delle linee e delle colonne:

$$\mathbf{R}_1(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{pmatrix} ; \qquad \mathbf{R}_2(\alpha) = \begin{pmatrix} \cos\alpha & 0 & -\sin\alpha \\ 0 & 1 & 0 \\ \sin\alpha & 0 & \cos\alpha \end{pmatrix} . \tag{1.44}$$

Come si può ricavare dalla Figura 2, la terna (x_1, x_2, x_3) può essere ottenuta dalla terna (ξ, η, ζ) per mezzo di tre rotazioni successive: 1) una rotazione di $-\omega$ attorno all'asse ζ, che porta l'asse ξ a coincidere con la linea dei nodi (nodo ascendente); 2) una rotazione di $-i$ attorno al nuovo asse ξ (risultato della precedente rotazione), che porta l'asse ζ a coincidere con l'asse x_3; 3) una rotazione di $-\Omega$ attorno al nuovo asse ζ (risultato delle due precedenti rotazioni). La trasformazione cercata è dunque

$$\begin{aligned} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} &= \mathbf{R}(\Omega, i, \omega) \begin{pmatrix} \xi \\ \eta \\ \zeta \end{pmatrix} , \\ \text{dove} \quad \mathbf{R}(\Omega, i, \omega) &= \mathbf{R}_3(-\Omega)\, \mathbf{R}_1(-i)\, \mathbf{R}_3(-\omega) ; \end{aligned} \tag{1.45}$$

sostituendo le espressioni esplicite delle matrici di rotazione si ottiene

$$\begin{aligned} R_{11} &= \cos\omega\cos\Omega - \sin\omega\sin\Omega\cos i \\ R_{12} &= -\sin\omega\cos\Omega - \cos\omega\sin\Omega\cos i \\ R_{13} &= \sin\Omega\sin i \\ R_{21} &= \cos\omega\sin\Omega + \sin\omega\cos\Omega\cos i \\ R_{22} &= -\sin\omega\sin\Omega + \cos\omega\cos\Omega\cos i \\ R_{23} &= -\cos\Omega\sin i \\ R_{31} &= \sin\omega\sin i \\ R_{32} &= \cos\omega\sin i \\ R_{33} &= \cos i . \end{aligned} \tag{1.46}$$

In definitiva l'orbita di un pianeta può essere specificata per mezzo dei sei *elementi kepleriani* a, e, i, ω, Ω ed M; di questi due (a ed e) descrivono la forma e le dimensioni dell'orbita; altri tre (i, ω, e Ω) danno l'orientazione del piano orbitale nel sistema di riferimento usato; l'ultimo (M) indica la posizione del pianeta lungo l'orbita a un certo istante di tempo. È conveniente abituarsi a pensare a questo insieme di parametri come a un altro modo di specificare le coordinate del pianeta, alternativo rispetto alle sei componenti dei vettori $\mathbf{r}$ e $\dot{\mathbf{r}}$ ma ad esse completamente equivalenti, tanto che tra le due rappresentazioni esiste una corrispondenza biunivoca[(8)] (si vedano le relative formule di trasformazione nelle Appendici).

[(8)] Per essere precisi, la corrispondenza tra gli elementi kepleriani così come sono stati qui definiti e i vettori posizione e velocità è biunivoca solo nel caso di orbite ellittiche (stati legati con energia $C < 0$) ma la trattazione può essere generalizzata o introducendo elementi orbitali definiti in modo differente per orbite paraboliche e iperboliche, o introducendo *elementi universali*, validi per tutti i tipi di orbita.

1.9 Le serie f e g

In alcune applicazioni (ad esempio, nella determinazione degli elementi orbitali a partire da tre osservazioni) è conveniente usare una soluzione approssimata dell'equazione del moto (1.6) che abbia la forma di una serie di Taylor, cioè un'espressione del tipo

$$\mathbf{r}(t) = \mathbf{r}_0 + \dot{\mathbf{r}}_0 \Delta t + \frac{1}{2!}\ddot{\mathbf{r}}_0 \Delta t^2 + \frac{1}{3!}\left.\frac{d^3\mathbf{r}}{dt^3}\right|_0 \Delta t^3 + \frac{1}{4!}\left.\frac{d^4\mathbf{r}}{dt^4}\right|_0 \Delta t^4 + \dots \,, \tag{1.47}$$

dove tutte le derivate sono calcolate al tempo t_0 e $\Delta t = t - t_0$. Il secondo membro della (1.47) può essere trasformato in modo da contenere solamente $\mathbf{r}_0$ e la sua derivata prima; infatti derivando successivamente l'equazione del moto (1.6) e sostituendo $\ddot{\mathbf{r}}_0$ con il suo valore $-\mu\mathbf{r}_0/r_0^3$ si ottengono espressioni sostitutive per le derivate di ordine superiore

$$\begin{aligned} \ddot{\mathbf{r}}_0 &= -\mu\frac{\mathbf{r}_0}{r_0^3} \\ \left.\frac{d^3\mathbf{r}}{dt^3}\right|_0 &= -\mu\left(\frac{1}{r_0^3}\dot{\mathbf{r}}_0 - \frac{3\mathbf{r}_0}{r_0^5}\,\mathbf{r}_0\cdot\dot{\mathbf{r}}_0\right) \\ &\dots \end{aligned} \tag{1.48}$$

Utilizzando le (1.48), la (1.47) si può scrivere

$$\mathbf{r}(t) = f\,\mathbf{r}_0 + g\,\dot{\mathbf{r}}_0\,, \tag{1.49}$$

dove f e g sono due serie

$$\begin{aligned} f &= 1 - \frac{1}{2}u\Delta t^2 + \frac{1}{2}us\Delta t^3 + \frac{1}{24}u(3w - 2u - 15s^2)\Delta t^4 + \\ &\qquad -\frac{1}{8}us(3w - 2u - 7s^2)\Delta t^5 + \dots \\ g &= \Delta t - \frac{1}{6}u\Delta t^3 + \frac{1}{4}us\Delta t^4 + \frac{1}{120}u(9w - 8u - 45s^2)\Delta t^5 + \dots \end{aligned} \tag{1.50}$$

in cui s, u e w sono tre costanti scalari, funzioni delle condizioni iniziali solamente

$$s = \frac{1}{r_0^2}\,\mathbf{r}_0\cdot\dot{\mathbf{r}}_0\,; \qquad u = \frac{\mu}{r_0^3}\,; \qquad w = \frac{1}{r_0^2}\dot{\mathbf{r}}_0\cdot\dot{\mathbf{r}}_0\,. \tag{1.51}$$

1.10 Elementi orbitali non singolari

I sei elementi kepleriani a, e, i, ω, Ω ed M sopra introdotti non sono ben definiti per orbite di eccentricità e/o inclinazione prossime a zero: in particolare, la longitudine del nodo Ω e l'argomento del perielio ω risultano indeterminati per $i = 0$; ancora l'argomento del perielio ω e l'anomalia media M sono indeterminati per $e = 0$. Ciò significa che la trasformazione di coordinate da elementi kepleriani a vettori posizione e velocità è localmente non invertibile nell'intorno di $i = 0$, $e = 0$, cioè la matrice jacobiana della trasformazione è singolare; questo fatto introduce instabilità numeriche in alcune applicazioni (ad esempio, nella determinazione degli elementi orbitali a partire da osservazioni con il metodo dei minimi quadrati), per cui è talvolta consigliabile utilizzare una parametrizzazione dell'orbita che non presenti questo inconveniente. Ciò può essere ottenuto usando due accorgimenti:

1) poiché le direzioni della linea dei nodi e del perielio sono indeterminate per $i = 0$, $e = 0$, occorre evitare di usare angoli che siano definiti a partire da tali direzioni; in

particolare, l'argomento del perielio ω e le anomalie (media M ed eccentrica E) vanno sostituiti dalle corrispondenti *longitudini*

$$\begin{aligned} &\text{longitudine del perielio :} && \varpi = \Omega + \omega\,, \\ &\text{longitudine media :} && L = M + \varpi\,, \\ &\text{longitudine eccentrica :} && F = E + \varpi\,; \end{aligned} \tag{1.52}$$

2) le coppie di variabili di tipo "polare" (e, ϖ) e (i, Ω) vanno sostituite con le corrispondenti variabili "cartesiane"

$$\begin{aligned} h &= e\sin\varpi\,, & k &= e\cos\varpi\,; \\ P &= \tan(i/2)\sin\Omega\,, & Q &= \tan(i/2)\cos\Omega\,. \end{aligned} \tag{1.53}$$

Con queste sostituzioni, l'*equazione di Keplero* (1.31) prende la forma

$$L = F + h\cos F - k\sin F\,; \tag{1.54}$$

i metodi numerici introdotti per la soluzione della (1.31) possono essere facilmente adattati alla (1.54).

Le *coordinate nel piano orbitale* (1.32) non possono più essere riferite alla direzione del perielio ξ; introduciamo perciò nel piano orbitale un nuovo sistema di assi (ξ', η') ruotato di un angolo $-\varpi$ rispetto a (ξ, η), cioè ruotato di un angolo $-\Omega$ rispetto alla direzione del nodo ascendente

$$\begin{pmatrix} \xi' \\ \eta' \\ 0 \end{pmatrix} = \mathbf{R}_3(-\varpi) \begin{pmatrix} \xi \\ \eta \\ 0 \end{pmatrix}\,; \tag{1.55}$$

in questo modo l'asse ξ' viene a coincidere con l'asse x_1 del sistema di riferimento inerziale quando l'inclinazione i tende a zero. Esplicitando la (1.55) e introducendo le nuove variabili la (1.32) diventa

$$\begin{aligned} \xi' &= \xi\cos\varpi - \eta\sin\varpi = a\big[(1-\gamma h^2)\cos F + \gamma hk\sin F - k\big]\,, \\ \eta' &= \xi\sin\varpi + \eta\cos\varpi = a\big[(1-\gamma k^2)\sin F + \gamma hk\cos F - h\big]\,, \end{aligned} \tag{1.56}$$

$$\text{dove} \quad \gamma = \frac{1-\sqrt{1-e^2}}{e^2} = \frac{1}{1+\sqrt{1-e^2}}\,.$$

La relazione tra il sistema di riferimento orbitale (ξ', η', ζ') e quello inerziale (x_1, x_2, x_3) diventa

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{R}(P, Q) \begin{pmatrix} \xi' \\ \eta' \\ \zeta' \end{pmatrix}\,, \tag{1.57}$$

dove la nuova matrice di rotazione $\mathbf{R}(P, Q)$ è definita come

$$\begin{aligned} \mathbf{R}(P,Q) &= \mathbf{R}_3(-\Omega)\,\mathbf{R}_1(-i)\,\mathbf{R}_3(-\omega)\,\mathbf{R}_3(\varpi) \\ &= \mathbf{R}_3(-\Omega)\,\mathbf{R}_1(-i)\,\mathbf{R}_3(\Omega)\,. \end{aligned} \tag{1.58}$$

Le funzioni trigonometriche di Ω e i che compaiono nelle matrici di rotazione della (1.58) si possono esprimere in funzione delle variabili non singolari P e Q

$$\cos\Omega = \frac{Q}{\sqrt{P^2+Q^2}}\,, \qquad \sin\Omega = \frac{P}{\sqrt{P^2+Q^2}}\,,$$

$$\sin i = \frac{2\sqrt{P^2+Q^2}}{1+P^2+Q^2}\,, \qquad \cos i = \frac{1-P^2-Q^2}{1+P^2+Q^2}\,,$$

da cui si ottiene

$$\mathbf{R}_3(-\Omega) = \begin{pmatrix} \cos\Omega & -\sin\Omega & 0 \\ \sin\Omega & \cos\Omega & 0 \\ 0 & 0 & 1 \end{pmatrix} = \frac{1}{\sqrt{P^2+Q^2}} \begin{pmatrix} Q & -P & 0 \\ P & Q & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{R}_1(-i) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos i & -\sin i \\ 0 & \sin i & \cos i \end{pmatrix} = \frac{1}{1+P^2+Q^2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1-P^2-Q^2 & -2\sqrt{P^2+Q^2} \\ 0 & 2\sqrt{P^2+Q^2} & 1-P^2-Q^2 \end{pmatrix}$$

$$\mathbf{R}_3(\Omega) = \begin{pmatrix} \cos\Omega & \sin\Omega & 0 \\ -\sin\Omega & \cos\Omega & 0 \\ 0 & 0 & 1 \end{pmatrix} = \frac{1}{\sqrt{P^2+Q^2}} \begin{pmatrix} Q & P & 0 \\ -P & Q & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Sostituendo queste espressioni nella (1.58) si ottiene l'espressione esplicita della matrice di rotazione

$$\mathbf{R}(P,Q) = \frac{1}{1+P^2+Q^2} \begin{pmatrix} 1-P^2+Q^2 & 2PQ & 2P \\ 2PQ & 1+P^2-Q^2 & -2Q \\ -2P & 2Q & 1-P^2-Q^2 \end{pmatrix}. \qquad (1.59)$$

Capitolo 2

Determinazione degli elementi orbitali di un pianeta da osservazioni angolari

2.1 Gli osservabili

Nel capitolo precedente si è visto come, a partire dagli elementi orbitali di un pianeta che percorre un'orbita kepleriana, è possibile calcolarne la posizione in funzione del tempo. Questa a sua volta può essere usata per produrre una *effemeride*, cioè una tavola che riporta, a intervalli prefissati di tempo, la *posizione angolare apparente* del pianeta rispetto all'osservatore, cioè il versore $\mathbf{s}$ del vettore posizione relativa

$$\Delta\mathbf{r} = \rho\mathbf{s} = \mathbf{r} - \mathbf{R}$$

(dove $\mathbf{r}$ e $\mathbf{R}$ sono rispettivamente le posizioni eliocentriche del pianeta osservato e dell'osservatore); questo perché le misure astrometriche classiche sono misure solamente angolari, mentre il modulo ρ del vettore posizione apparente (cioè la distanza tra il pianeta e l'osservatore) non è osservabile. La direzione di osservazione $\mathbf{s}$ è usualmente espressa in coordinate polari (*ascensione retta* α e *declinazione* δ)

$$\begin{cases} s_1 &= \cos\delta\cos\alpha \\ s_2 &= \cos\delta\sin\alpha \\ s_3 &= \sin\delta\,. \end{cases} \tag{2.1}$$

Nel presente capitolo ci occupiamo del corrispondente *problema inverso*: calcolare gli elementi orbitali di un pianeta del Sistema Solare a partire da un certo insieme di osservazioni angolari. Notiamo innanzitutto che, poiché ogni osservazione fornisce la misura di due quantità scalari indipendenti (ascensione retta e declinazione), ci aspettiamo di aver bisogno di tre osservazioni (eseguite a tre tempi diversi t_k, $k = 1, 2, 3$) per poter ricavare i sei elementi kepleriani dell'orbita; perciò il metodo che studieremo (dovuto a Gauss) è anche chiamato *determinazione dell'orbita da tre osservazioni*. La relazione da cui partiamo è quindi

$$\mathbf{r}_k = \rho_k\mathbf{s}_k + \mathbf{R}_k \qquad (k = 1, 2, 3)\,, \tag{2.2}$$

dove l'indice k si riferisce ai tre diversi istanti di tempo t_k. Nella (2.2) supporremo note la posizione dell'osservatore $\mathbf{R}_k$ e la posizione angolare apparente del pianeta osservato $\mathbf{s}_k$, mentre la posizione eliocentrica del pianeta $\mathbf{r}_k$ e la sua distanza dall'osservatore ρ_k sono incognite. Assumeremo inoltre che l'orbita del pianeta osservato sia ellittica.

2.2 Il metodo di Gauss

Cerchiamo di ridurre il numero di incognite che compaiono nell'equazione (2.2). A questo scopo osserviamo che, poichè il pianeta osservato si muove su un'orbita kepleriana, i tre

vettori $\mathbf{r}_k$ devono essere complanari; ciò significa che, a meno di casi particolari degeneri, esistono tre scalari c_k (non nulli) tali che

$$c_1\mathbf{r}_1 + c_2\mathbf{r}_2 + c_3\mathbf{r}_3 = 0\,. \tag{2.3}$$

I coefficienti c_k sono ovviamente determinati a meno di una costante moltiplicativa arbitraria, per cui possiamo imporre, ad esempio, $c_2 = -1$ e trascrivere la (2.3) nella forma

$$\mathbf{r}_2 = c_1\mathbf{r}_1 + c_3\mathbf{r}_3\,. \tag{2.4}$$

Eseguendo i prodotti vettoriali tra i vettori $\mathbf{r}_k$ e tenendo conto della (2.4)

$$\begin{aligned} \mathbf{r}_1 \times \mathbf{r}_2 &= c_3\,\mathbf{r}_1 \times \mathbf{r}_3 \\ \mathbf{r}_3 \times \mathbf{r}_2 &= c_1\,\mathbf{r}_3 \times \mathbf{r}_1 \end{aligned}$$

otteniamo per c_1 e c_3 le seguenti espressioni

$$c_1 = \frac{\mathbf{r}_2 \times \mathbf{r}_3}{\mathbf{r}_1 \times \mathbf{r}_3}\,, \qquad c_3 = \frac{\mathbf{r}_1 \times \mathbf{r}_2}{\mathbf{r}_1 \times \mathbf{r}_3}\,. \tag{2.5}$$

Per semplificare queste equazioni esprimiamo $\mathbf{r}_1$ ed $\mathbf{r}_3$ in funzione di $\mathbf{r}_2$ ed $\dot{\mathbf{r}}_2$, usando le serie f e g centrate al tempo t_2; per semplificare la notazione scegliamo l'origine dell'asse dei tempi al tempo t_2, cioè assumiamo $t_2 = 0$; la (1.49) si può riscrivere

$$\mathbf{r}_1 = f_1\mathbf{r}_2 + g_1\dot{\mathbf{r}}_2\,, \qquad \mathbf{r}_3 = f_3\mathbf{r}_2 + g_3\dot{\mathbf{r}}_2\,, \tag{2.6}$$

dove si è posto per brevità $f_k = f(t_k)$, $g_k = g(t_k)$. Usando le (2.6), i prodotti vettoriali che compaiono nelle (2.5) diventano

$$\begin{aligned} \mathbf{r}_2 \times \mathbf{r}_3 &= \mathbf{r}_2 \times \big(f_3\mathbf{r}_2 + g_3\dot{\mathbf{r}}_2\big) &&= g_3\mathbf{h} \\ \mathbf{r}_1 \times \mathbf{r}_2 &= \big(f_1\mathbf{r}_2 + g_1\dot{\mathbf{r}}_2\big) \times \mathbf{r}_2 &&= -g_1\mathbf{h} \\ \mathbf{r}_1 \times \mathbf{r}_3 &= \big(f_1\mathbf{r}_2 + g_1\dot{\mathbf{r}}_2\big) \times \big(f_3\mathbf{r}_2 + g_3\dot{\mathbf{r}}_2\big) &&= \big(f_1g_3 - f_3g_1\big)\mathbf{h} \end{aligned}$$

(dove $\mathbf{h} = \mathbf{r}_2 \times \dot{\mathbf{r}}_2$ è il vettore momento angolare) che, sostituite nelle (2.5), forniscono

$$c_1 = \frac{g_3}{f_1g_3 - f_3g_1}\,, \qquad c_3 = \frac{-g_1}{f_1g_3 - f_3g_1}\,. \tag{2.7}$$

Sostituendo le (2.2) nella (2.3) si ottiene

$$c_1\rho_1\mathbf{s}_1 + c_2\rho_2\mathbf{s}_2 + c_3\rho_3\mathbf{s}_3 + c_1\mathbf{R}_1 + c_2\mathbf{R}_2 + c_3\mathbf{R}_3 = 0\,, \tag{2.8}$$

che è un sistema di tre equazioni nelle tre incognite ρ_k; inoltre, in esse compaiono le quantità c_k, che contengono un numero di incognite che dipende dal punto in cui decidiamo di troncare le serie f e g. Infatti, se negli sviluppi (1.52) tenessimo solo il primo termine (cioè ponessimo $f_k = 1$, $g_k = t_k$), i coefficienti c_k sarebbero funzioni di quantità note e quindi immediatamente calcolabili: questa sarebbe però un'approssimazione veramente rozza, equivalente alla sostituzione dell'orbita ellittica con un moto rettilineo uniforme! Al contrario, usando un numero arbitrario di termini, si introducono nell'espressione di c_k le quantità s, u e w (equazione (1.53)), con l'effetto di introdurre nuovamente nella (2.8) i vettori $\mathbf{r}_2$ e $\dot{\mathbf{r}}_2$ che siamo appena riusciti a eliminare. Nel metodo di Gauss si fa una scelta intermedia, usando solo i primi due termini delle (1.52), che contengono solamente il modulo del vettore $\mathbf{r}_2$ tramite il coefficiente $u = \mu/r_2^3$. Così, ponendo

$$f_k \simeq 1 - \frac{\mu t_k^2}{2r_2^3}\,, \qquad g_k \simeq t_k - \frac{\mu t_k^3}{6r_2^3}\,, \tag{2.9}$$

si introduce nelle (2.8) solo un'incognita aggiuntiva (r_2). Occorre dunque un'altra equazione da aggiungere al sistema: questa può essere ottenuta elevando al quadrato la (2.2) nel caso $k = 2$

$$r_2^2 = \rho_2^2 + 2\rho_2\, \mathbf{s}_2 \cdot \mathbf{R}_2 + R_2^2\,; \tag{2.10}$$

le (2.8) e (2.10) costituiscono quindi un sistema di quattro equazioni nelle quattro incognite ρ_1, ρ_2, ρ_3 ed r_2, che deve essere risolto per poter ottenere i vettori $\mathbf{r}_k$ (dalla (2.2)) ed $\dot{\mathbf{r}}_2$ (per mezzo delle (2.6)); i due vettori $\mathbf{r}_2$ ed $\dot{\mathbf{r}}_2$ definiscono le condizioni iniziali del moto, cioè costituiscono la soluzione cercata del problema; gli elementi orbitali possono essere ricavati dai vettori $\mathbf{r}_2$ ed $\dot{\mathbf{r}}_2$ con il metodo esposto nell'Appendice A.2.

Prima di procedere con la soluzione notiamo che il fatto di aver introdotto le approssimazioni (2.9) produce due effetti:

1) la soluzione trovata dalle equazioni (2.8) e (2.10) non sarà una soluzione esatta del problema posto, cioè l'orbita corrispondente alle condizioni iniziali $(\mathbf{r}_2, \dot{\mathbf{r}}_2)$ non sarà esattamente compatibile con le osservazioni $\mathbf{s}_k$; si rende perciò necessario un secondo stadio di calcolo (basato su un procedimento iterativo), in cui la precisione della soluzione è migliorata;

2) siccome le serie troncate (2.9) sono in grado di rappresentare approssimativamente l'orbita solo per un tempo abbastanza corto (una frazione non troppo grande del periodo orbitale), ne consegue che il metodo di Gauss è applicabile solo quando le tre osservazioni disponibili sono abbastanza ravvicinate nel tempo. Questa non è di solito una grave limitazione, in quanto il metodo è utile soprattutto nel caso di oggetti appena scoperti (tipicamente asteroidi), di cui non si conosce ancora l'orbita e che perciò devono essere "seguiti" notte per notte per non essere "persi"; di solito quindi le prime misurazioni astrometriche disponibili sono per forza di cose ravvicinate. In seguito, quando le osservazioni si accumulano, il problema diventa quello di migliorare e correggere la determinazione iniziale dell'orbita tenendo conto delle nuove misurazioni; questo problema però richiede procedimenti di altro tipo, basati sul metodo dei minimi quadrati (vedi Capitolo 4).

2.3 La soluzione preliminare

Per risolvere il sistema costituito dalle (2.8) e (2.10), calcoliamo dapprima l'espressione esplicita dei coefficienti c_1 e c_3: dalle (2.9) si ricava

$$f_1 g_3 - f_3 g_1 \simeq \big(t_3 - t_1\big)\left[1 - \frac{\mu(t_3 - t_1)^2}{6r_2^3} + \frac{\mu^2 t_1^2 t_3^2}{12 r_2^6}\right],$$

che si può approssimare (trascurando le potenze di t_k superiori alla terza come infinitesimi di ordine superiore, analogamente a quanto fatto nelle (2.9)) come

$$f_1 g_3 - f_3 g_1 \simeq t_{13}\left[1 - \frac{t_{13}^2}{6r_2^3}\right], \qquad \big(\text{dove } t_{13} = t_3 - t_1\big).$$

Per $|\alpha|, |\beta| \ll 1$ si può porre $(1+\alpha)^{-1} \simeq 1 - \alpha$ e $(1+\alpha)(1+\beta) \simeq 1 + \alpha + \beta$ (a meno di infinitesimi di ordine superiore); mantenendo lo stesso grado di approssimazione delle formule precedenti si può dunque scrivere

$$\frac{1}{f_1 g_3 - f_3 g_1} \simeq \frac{1}{t_{13}}\left[1 + \frac{\mu t_{13}^2}{6r_2^3}\right],$$

e le (2.7) diventano

$$\begin{aligned} c_1 &\simeq \frac{t_3}{t_{13}}\left[1+\frac{\mu}{6r_2^3}\left(t_{13}^2-t_3^2\right)\right] = A_1+\frac{B_1}{r_2^3} \\ c_3 &\simeq -\frac{t_1}{t_{13}}\left[1+\frac{\mu}{6r_2^3}\left(t_{13}^2-t_1^2\right)\right] = A_3+\frac{B_3}{r_2^3}\,, \end{aligned} \tag{2.11}$$

dove si è posto per brevità

$$\begin{aligned} A_1 &= \frac{t_3}{t_{13}}\,, \qquad B_1 = \frac{\mu t_3}{6t_{13}}\left(t_{13}^2-t_3^2\right), \\ A_3 &= -\frac{t_1}{t_{13}}\,, \qquad B_3 = -\frac{\mu t_1}{6t_{13}}\left(t_{13}^2-t_1^2\right). \end{aligned} \tag{2.12}$$

La (2.8) può essere espressa in forma matriciale

$$\begin{pmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{pmatrix} \begin{pmatrix} c_1\rho_1 \\ c_2\rho_2 \\ c_3\rho_3 \end{pmatrix} = - \begin{pmatrix} G_1 \\ G_2 \\ G_3 \end{pmatrix} \tag{2.13}$$

con

$$G_i = \sum_{k=1}^{3} R_{ik}c_k\,, \tag{2.14}$$

dove si sono indicate con S_{ik} ed R_{ik} le componenti $i$esime dei vettori $\mathbf{s}_k$ ed $\mathbf{R}_k$. Invertendo la matrice S_{ik} si ottiene

$$c_k\rho_k = -\sum_{i=1}^{3} S_{ki}^{-1}G_i \tag{2.15}$$

cioè, nel caso di $k=2$,

$$\rho_2 = \sum_{i=1}^{3} S_{2i}^{-1}G_i\,.$$

Sostituendo in questa espressione le (2.14) e (2.11) si ottiene

$$\rho_2 = \sum_{i=1}^{3} S_{2i}^{-1}\sum_{k=1}^{3} R_{ik}c_k = A_2^* + \frac{B_2^*}{r_2^3}\,, \tag{2.16}$$

dove

$$A_2^* = \sum_{i=1}^{3} S_{2i}^{-1}\sum_{k=1}^{3} R_{ik}A_k\,, \qquad B_2^* = \sum_{i=1}^{3} S_{2i}^{-1}\sum_{k=1}^{3} R_{ik}B_k\,. \tag{2.17}$$

Sostituendo la (2.16) nella (2.10) si ottiene infine

$$r_2^8 - \left[A_2^{*2} + 2A_2^*\mathbf{s}_2\cdot\mathbf{R}_2 + R_2^2\right]r_2^6 - \left[2A_2^*B_2^* + 2B_2^*\mathbf{s}_2\cdot\mathbf{R}_2\right]r_2^3 - B_2^{*2} = 0\,, \tag{2.18}$$

che è un'equazione di ottavo grado nell'unica incognita r_2: la sua soluzione[(1)] fornisce quindi il valore di r_2. Da r_2 si possono poi ricavare i valori di f_k e g_k (dalle (2.9)) e i coefficienti c_1

[(1)] Non esiste un metodo algebrico per trovare la soluzione di un'equazione di ottavo grado del tipo della (2.18); esistono però diversi metodi numerici (su cui qui non ci soffermiamo) per calcolare gli zeri di un polinomio di grado arbitrario; in una implementazione del metodo su calcolatore, questi possono usualmente essere utilizzati "a scatola chiusa" come chiamate a sottoprogrammi di libreria.

e c_3 (dalle (2.11)), quindi i valori ρ_k delle distanze pianeta-osservatore al tempo delle misure (dalla (2.15)). Dalle (2.2) si ottengono quindi le tre posizioni eliocentriche del pianeta $\mathbf{r}_k$; le (2.6) permettono poi di ottenere $\dot{\mathbf{r}}_2$, ad esempio

$$\dot{\mathbf{r}}_2 = \frac{\mathbf{r}_1 - f_1 \mathbf{r}_2}{g_1} \,. \tag{2.19}$$

I due vettori $(\mathbf{r}_2, \dot{\mathbf{r}}_2)$ così trovati costituiscono la prima approssimazione alle condizioni iniziali cercate.

Bisogna notare che la (2.18) può ammettere più di una soluzione reale positiva. Infatti un polinomio di ottavo grado ha in generale otto zeri nel campo complesso; ovviamente nel nostro caso hanno significato fisico solo le soluzioni della (2.18) che sono reali e positive; il numero di queste soluzioni sarà variabile da caso a caso, in funzione del valore dei coefficienti dell'equazione (2.18). Si ha quindi il problema di scegliere tra le varie soluzioni "a priori possibili" quella che corrisponde all'orbita reale del pianeta osservato. A volte alcune soluzioni possono essere scartate per considerazioni di verosimiglianza: ad esempio, alcune radici possono portare a orbite iperboliche, caso che per semplicità abbiamo escluso dalla presente trattazione e che si considera non realistico per un asteroide (ma attenzione: alcune comete si muovono su orbite iperboliche, e non è sempre facile distinguere una cometa da un asteroide!). In generale però la scelta tra le diverse soluzioni può essere fatta con sicurezza solo continuando a seguire l'oggetto e acquisendo nuove osservazioni, che saranno compatibili solo con una delle orbite calcolate.

2.4 La correzione della soluzione preliminare

La correzione della soluzione preliminare ottenuta nel paragrafo precedente al fine di ottenere condizioni iniziali che siano esattamente compatibili con le osservazioni può essere effettuata con diversi metodi, che hanno però in comune lo stesso procedimento iterativo. Questo si basa sul fatto che, utilizzando i valori preliminari di $(\mathbf{r}_2, \dot{\mathbf{r}}_2)$, è possibile calcolare le serie f e g (e quindi i coefficienti c_1 e c_3) con una precisione superiore a quella consentita dalle espressioni (2.9); con questi nuovi valori di c_k è possibile ricalcolare G_i (equazione (2.14)) e ρ_k (equazione (2.15)); dalle (2.2) si ottengono dunque $\mathbf{r}_k$ e da questi (per mezzo della (2.19)) i nuovi valori di $(\mathbf{r}_2, \dot{\mathbf{r}}_2)$. A questo punto il procedimento può essere ripetuto un numero arbitrario di volte, fino al raggiungimento della convergenza. I diversi metodi si differenziano solamente per il modo in cui i valori di f_k e g_k sono calcolati a partire da $(\mathbf{r}_2, \dot{\mathbf{r}}_2)$.

Il procedimento più ovvio sembrerebbe quello di utilizzare direttamente le definizioni (1.52) delle serie f e g. Questo metodo però non è molto adatto alla implementazione in un programma di calcolo automatico: infatti i termini delle serie (1.52) hanno espressioni diverse gli uni dagli altri e vanno quindi programmati separatamente. Inoltre sarebbe comunque inevitabile troncare la serie a una potenza prefissata di t: ciò introduce ancora una approssimazione, mentre in questo stadio vorremmo usare espressioni esatte. Un'espressione esatta e in forma chiusa delle serie f e g è ricavata nell'Appendice B; nel nostro caso, le equazioni (B.7) e (B.8) prendono la forma

$$\begin{aligned} f_k &= 1 - \left[\frac{1 - \cos(E_k - E_2)}{1 - e \cos E_2} \right] , \\ g_k &= t_k - \frac{1}{n} \Big[(E_k - E_2) - \sin(E_k - E_2) \Big] . \end{aligned} \tag{2.20}$$

Nelle espressioni (2.20) compaiono l'eccentricità orbitale e, il moto medio n e i valori dell'anomalia eccentrica E_k ai tempi delle osservazioni t_k; questi possono essere ricavati trasformando le condizioni iniziali del moto $(\mathbf{r}_2, \dot{\mathbf{r}}_2)$ in elementi orbitali, come spiegato nell'Appendice A.2. Riassumendo, i passi da eseguire per la correzione dei valori di $(\mathbf{r}_2, \dot{\mathbf{r}}_2)$ sono:

1) ricavare da $(\mathbf{r}_2, \dot{\mathbf{r}}_2)$ gli elementi orbitali (seguendo il procedimento indicato nell'Appendice A.2); in particolare occorre ottenere e, n ed M_2 (valore dell'anomalia media al tempo t_2);

2) calcolare i valori dell'anomalia media ai tempi delle misure

$$M_k = M_2 + n(t_k - t_2)\,;$$

3) calcolare i valori E_k dell'anomalia eccentrica ai tempi delle misure, risolvendo l'equazione di Keplero

$$M_k = E_k - e \sin E_k\,;$$

4) calcolare i valori di f_k e g_k usando l'equazione (2.20);

5) calcolare i valori di c_1 e c_3 usando l'equazione (2.7);

6) calcolare G_i dalla (2.14);

7) ricavare ρ_k dalla (2.15);

8) ricalcolare $\mathbf{r}_k$ dalla (2.2);

9) ricalcolare $\dot{\mathbf{r}}_2$ dalla (2.19).

Con i nuovi valori di $(\mathbf{r}_2, \dot{\mathbf{r}}_2)$ così ottenuti si può ritornare al punto 1) e iterare il procedimento quante volte è necessario. Siccome in questo algoritmo non è stata introdotta nessuna approssimazione, il valore delle condizioni iniziali che si ottiene a convergenza raggiunta corrisponde all'orbita che passa esattamente per le tre osservazioni date.

2.5 Limitazioni del metodo di Gauss

Il metodo di Gauss, pur utilissimo nelle fasi iniziali della determinazione orbitale, ha alcune limitazioni che è bene conoscere:

1) poiché il calcolo della soluzione preliminare per la distanza eliocentrica del pianeta r_2 si basa sull'uso delle serie f e g troncate a una potenza piuttosto bassa dell'intervallo di tempo Δt, il metodo funziona solo se la distanza temporale tra le osservazioni usate è piccola rispetto al periodo orbitale; pur non essendo possibile stabilire un limite preciso, in pratica si osserva che la convergenza del metodo diventa difficile quando $n\Delta t \approx 1$, cioè le osservazioni coprono un arco di orbita dell'asteroide osservato uguale o maggiore di circa 1 radiante. In pratica ciò costituisce raramente una limitazione grave in quanto, nelle fasi iniziali della scoperta di un oggetto, questo viene seguito e osservato più volte, cosicché solitamente sono disponibili almeno tre osservazioni a distanza ravvicinata;

2) si è visto che per ottenere l'equazione (2.18) che fornisce il valore preliminare della distanza r_2 è necessario usare l'inversa della matrice S_{ik}, e si è assunto implicitamente che questa sia invertibile. In realtà esiste un caso in cui essa *non è* invertibile: poiché gli elementi della matrice S_{ik} sono le componenti cartesiane dei tre versori $\mathbf{s}_k$, quando essi sono complanari le componenti S_{ik} non sono linearmente indipendenti e la matrice è singolare[(2)]; in questo caso il metodo di Gauss non è applicabile[(3)]. Questo problema

[(2)] Detto in altre parole, il determinante Δ della matrice S_{ik} è uguale al volume del parallelepipedo che ha per spigoli i tre vettori $\mathbf{s}_k$; quando questi sono complanari il parallelepipedo degenera in una figura piana, il determinante si annulla e la matrice è singolare.

[(3)] Si noti che questa limitazione non riguarda solo il calcolo della radice preliminare dell'equazione (2.18), ma anche i passi successivi della sua correzione, che fanno comunque tutti uso dell'equazione (2.15). Si tratta dunque di un problema che non può essere risolto scegliendo per le iterazioni descritte nella sezione 2.4 un punto di partenza diverso da quello ottenibile dalla soluzione dell'equazione (2.18).

ha natura essenziale solo nel caso in cui l'orbita dell'asteroide osservato sia esattamente complanare a quella della Terra; in tal caso risulteranno complanari anche i vettori posizione relativa $\mathbf{s}_k$, comunque si scelgano i tempi di osservazione, e il metodo di Gauss è inapplicabile. In un caso simile (che in pratica è molto raro, almeno se si considerano corpi celesti naturali) bisogna ricorrere a metodi di determinazione orbitale espressamente rivolti alle orbite complanari. Se invece le orbite della Terra e dell'asteroide non sono complanari, può capitare per caso che le tre osservazioni scelte siano complanari, ma tale condizione è del tutto accidentale, e sarà possibile scegliere (o eseguire) una nuova osservazione in modo che ciò non accada. L'unico caso di importanza pratica è quello in cui la distanza temporale tra le osservazioni utilizzate sia così piccola che le osservazioni risultino *quasi complanari*; in questo caso, anche se la matrice S_{ik} risulta a rigore invertibile, può verificarsi che la sua inversione risulti estremamente instabile, nel senso che piccolissime variazioni dei valori di S_{ik} si traducano in grandi variazioni del risultato[(4)]; questo effetto è uno degli aspetti dell'instabilità del metodo di Gauss nel caso di osservazioni molto ravvicinate che è descritto nel prossimo punto;

3) in generale non esiste nessuna garanzia che il procedimento iterativo descritto nella sezione 2.4 converga, anche nei casi in cui il problema in sé ammette una soluzione. Ciò ha portato a sviluppare metodi alternativi alla soluzione dell'equazione (2.18) per trovare un punto di partenza per le iterazioni (incluso scegliere casualmente un certo numero di valori iniziali di r_2, distribuiti su un intervallo abbastanza ampio da comprendere tutti i casi ragionevolmente possibili);

4) il fatto che il metodo di Gauss, quando il procedimento iterativo descritto nella sezione 2.4 ha raggiunto la convergenza, fornisca una soluzione *esatta* del problema posto, cioè porti alla determinazione di un insieme di elementi orbitali che riproducono esattamente le osservazioni di partenza (nel senso che una effemeride, calcolata a partire da queste condizioni iniziali, ritrova esattamente i valori angolari delle osservazioni) non deve trarre in inganno e portare a credere che gli elementi orbitali così calcolati siano privi di errore. Infatti la soluzione del problema di Gauss è una funzione delle osservazioni di partenza, che inevitabilmente sono affette da errori di misura: gli elementi orbitali calcolati conterranno l'effetto di questi errori, più o meno amplificato da coefficienti che sono gli elementi della matrice jacobiana della trasformazione, che descrivono la *sensibilità* della soluzione rispetto a variazioni dei dati di partenza[(5)]. È anche chiaro che il problema, posto in questi termini, non riguarda il metodo di Gauss in sé: stiamo qui parlando piuttosto di proprietà generali della soluzione del problema inverso di Keplero (il metodo di Gauss è semplicemente un algoritmo per calcolare questa soluzione). Benché non esista una trattazione generale e rigorosa di come tale sensibilità vari al variare della configurazione geometrica Terra-Sole-asteroide e dell'intervallo di tempo Δt compreso tra le tre osservazioni, l'esperienza mostra che tale sensibilità cresce molto quando Δt tende a zero, fino a rendere del tutto impossibile la determinazione orbitale (l'algoritmo di Gauss non converge, o converge su soluzioni

[(4)] Ciò avviene perché i coefficienti dell'inversa S_{ik}^{-1} sono proporzionali al reciproco del determinante Δ di S_{ik}.

[(5)] Non è pratica comune investigare gli errori da cui è affetta la soluzione ricavata dal metodo di Gauss; ciò è dovuto al fatto che gli elementi orbitali così ottenuti sono quasi sempre considerati solo come valori provvisori, utili per poter rintracciare l'asteroide nei giorni immediatamente successivi alla scoperta o per servire come punto di partenza per una determinazione orbitale più accurata, basata su un insieme più numeroso di osservazioni astrometriche; solitamente l'analisi degli errori orbitali è perciò rimandata alle fasi successive dell'elaborazione, in cui si utilizza il metodo dei minimi quadrati (vedi capitolo seguente). Non è comunque difficile ottenere una stima numerica approssimata degli errori dell'orbita calcolata con il metodo di Gauss, ad esempio variando i valori delle osservazioni astrometriche (α_i, δ_i) di piccole quantità, compatibili con gli errori osservativi, e osservando come varia di conseguenza la soluzione orbitale.

patentemente prive di senso fisico). Probabilmente ciò deriva dalla concomitanza di più fattori:

(a) l'annullarsi del determinante della matrice S_{ik}, descritto sopra. In termini molto approssimativi, quando Δt tende a zero il *moto apparente* dell'asteroide sulla volta celeste tende a essere indistinguibile da un moto rettilineo uniforme e quindi definisce solo 4 parametri indipendenti (α, δ, $\dot{\alpha}$ e $\dot{\delta}$), insufficienti a determinare 6 elementi orbitali;

(b) come già osservato, le osservazioni astrometriche non contengono alcuna informazione sulla distanza ρ dell'asteroide e, quindi, sul valore del semiasse maggiore della sua orbita. Il processo di determinazione orbitale supplisce a questa carenza di informazione sfruttando due effetti: 1) la determinazione della velocità angolare dell'oggetto, che è legata al suo semiasse maggiore dalla terza legge di Keplero; 2) il fatto che la posizione dell'osservatore (della Terra) cambi nel tempo e quindi che osservazioni eseguite a tempi diversi contengano informazioni sulla distanza a causa dell'effetto di parallasse. Il modo in cui questi due effetti agiscono sul processo di determinazione orbitale è in qualche modo nascosto nelle formule del metodo di Gauss, ma è innegabile che essi in qualche modo siano presenti. Entrambi questi effetti dipendono in modo essenziale dallo scorrere del tempo, ed è quindi comprensibile che diventino sempre meno efficaci quando Δt tende a zero; di fatto si osserva empiricamente che, quando $\Delta t \to 0$, l'incertezza sul valore del semiasse maggiore cresce molto più rapidamente di quella, ad esempio, degli elementi orbitali che definiscono l'orientazione del piano orbitale nello spazio (Ω, ω e i).

A causa di questi fattori, quando $\Delta t \to 0$ l'errore nella stima degli elementi orbitali peggiora rapidamente, finché il metodo di Gauss diventa inapplicabile perché non riesce a convergere.

Le considerazioni precedenti giustificano in qualche modo quanto trovato empiricamente da generazioni di astronomi, e cioè che il metodo di Gauss funziona bene se l'intervallo Δt su cui sono distribuite le osservazioni non è né troppo piccolo né troppo grande. Pur non essendo possibile dare limiti precisi (che ovviamente dipendono dall'orbita, dalla geometria delle osservazioni, ecc.), si può dire molto indicativamente che, per un asteroide di fascia principale[(6)], Δt deve essere compreso tra 1–2 giorni e poche settimane.

[(6)] Semiasse maggiore compreso tra circa 2 e 3.3 AU.

Capitolo 3

Cenni sul metodo dei minimi quadrati

Abbiamo visto nel capitolo precedente come è possibile calcolare gli elementi orbitali di un pianeta del Sistema Solare usando *tre osservazioni astrometriche* solamente angolari (cioè che forniscono solo la posizione angolare del corpo sulla volta celeste e non la sua distanza dall'osservatore). Questo procedimento può essere visto come la soluzione del *problema inverso* rispetto al problema di Keplero studiato nel Capitolo 1, ed è molto utile soprattutto nelle fasi iniziali della scoperta di un nuovo pianeta o asteroide, quando ancora non si ha alcuna informazione sulla sua orbita. In seguito, quando l'oggetto continua a essere osservato e quindi si viene accumulando un numero sempre maggiore di osservazioni, il problema diventa quello di *migliorare la conoscenza dell'orbita* sfruttando tutte le osservazioni disponibili. Un modo molto comune di affrontare questo problema consiste nell'usare il *metodo dei minimi quadrati*, un metodo per la soluzione di problemi inversi che ha un campo di applicazione estremamente vasto. In questo capitolo descriveremo pertanto il metodo da un punto di vista del tutto generale, rimandando ai capitoli seguenti la descrizione della sua applicazione al problema della determinazione orbitale.

3.1 Modello dell'esperimento e osservabili

Un tipo di problema molto comune in fisica sperimentale è quello della *stima di parametri fisici non direttamente osservabili* attraverso la misura di grandezze che siano a essi collegate per mezzo di una espressione matematica nota (legge fisica) che si assume come modello dell'esperimento. In termini del tutto generali la formulazione matematica del problema è di questo tipo: si vuole determinare il valore di un certo insieme di M *parametri* β_k $(k = 1, 2, \ldots, M)$ attraverso la misurazione di un certo numero N di *osservabili* x_i $(i = 1, 2, \ldots, N)$; le osservabili sono funzioni dei parametri attraverso funzioni *note* f_i dei parametri, che si assumono come modello matematico in grado di descrivere l'esperimento

$$x_i = f_i(\beta_k)\,. \tag{3.1}$$

Le equazioni (3.1) sono chiamate *equazioni di osservazione*. I motivi per cui le funzioni $f_i(\beta)$ sono differenti per le diverse osservabili x_i possono essere di varia natura:

- potrebbe essere che si stia osservando ripetutamente la stessa osservabile a differenti istanti di tempo e che la funzione f che lega i parametri β all'osservabile x dipenda esplicitamente dal tempo; in altre parole si avrebbe[(1)]

$$x_i = x(t_i) = f(\beta, t_i)\,;$$

[(1)] Il problema della determinazione orbitale a partire da osservazioni angolari è *quasi* di questa specie, a parte per il fatto che esistono due differenti tipi di osservabili (ascensione retta e declinazione).

- potrebbe essere che si stia osservando osservabili omogenei, cioè tali che la relazione tra β e x sia data sempre dalla stessa funzione f che però dipende (oltre che da β) anche da altri parametri γ, il cui valore (supposto noto) varia al variare dell'indice i; si avrebbe cioè[(2)]
$$x_i = f(\beta, \gamma_i)\,;$$
- oppure potrebbe essere che gli osservabili x_i siano veramente quantità fisiche differenti, legate ai parametri β da funzioni diverse[(3)].

Indipendentemente dalla forma specifica delle funzioni $f_i(\beta_k)$, assumeremo comunque che esse siano note e che esista un algoritmo (almeno numerico) per calcolarle. Per quanto riguarda il numero totale di osservabili N e di parametri M, si possono avere tre casi[(4)]:

1) $N < M$: il numero di osservabili è minore del numero di parametri da determinare;

2) $N = M$: il numero di osservabili è uguale al numero di parametri da determinare;

3) $N > M$: il numero di osservabili è maggiore del numero di parametri da determinare.

Nel primo caso il problema è indeterminato (ammette infinite soluzioni); si potrebbe dire che l'esperimento è progettato male perché nelle osservabili non c'è abbastanza informazione per ricavare tutti i parametri. Al massimo si potrebbero usare le equazioni di osservazione come *condizioni di vincolo* tra i valori dei parametri, ma in questa sede tale possibilità non ci interessa, e non considereremo oltre questo caso. Nel secondo caso la soluzione è (almeno teoricamente) banale; se la funzione $f(\beta)$ è invertibile, le equazioni (3.1) definiscono univocamente il valore dei parametri β (naturalmente, come si è visto nel capitolo precedente nel caso del metodo di Gauss, ciò non significa necessariamente che sia semplice trovare un algoritmo che fornisca la soluzione). Il caso che ci interessa maggiormente è quindi il terzo, in cui l'informazione fornita dall'esperimento è sovrabbondante; questo fa sì che in generale le equazioni di osservazione non siano risolubili esattamente. Infatti il modello adottato per l'esperimento ha sempre un certo grado di approssimazione (a seconda dei punti di vista si può dire che la teoria è incompleta o che le misure sono soggette a errore); in altre parole le equazioni di osservazione vanno riscritte più esattamente

$$x_i = f_i(\beta_k) + \varepsilon_i \tag{3.2}$$

dove ε_i sono gli *scarti* (o errori di misura). Si noti il diverso *status* che hanno le variabili che compaiono nella (3.2): le x_i sono i risultati delle misure e quindi sono note; i parametri β_k sono ignoti e devono essere determinati; gli scarti ε_i sono ignoti e, quasi per definizione, non possono essere conosciuti. Se si conoscessero i meccanismi fisici che producono gli scarti ε_i, questi potrebbero essere inclusi nel modello deterministico f_i, eventualmente introducendo nuovi parametri β_k; ma in generale, in ogni esperimento, è necessario arrestare

[(2)] Un esempio di questo tipo è il seguente: si supponga di voler determinare le coordinate cartesiane (x, y, z) di un punto dello spazio misurando la sua distanza d_i da una serie di punti (X_i, Y_i, Z_i) $(i = 1, \ldots, N)$ di coordinate note; in questo caso le equazioni di osservazione si scrivono
$$d_i = f(x, y, z, X_i, Y_i, Z_i) = \sqrt{(x - X_i)^2 + (y - Y_i)^2 + (z - Z_i)^2}\,,$$
dove $\beta = (x, y, z)$ sono i parametri da determinare e $\gamma_i = (X_i, Y_i, Z_i)$ sono quantità note.

[(3)] La determinazione dell'orbita di un asteroide potrebbe ad esempio utilizzare contemporaneamente osservazioni astrometriche angolari, misure di distanza e di velocità radiale (per effetto Doppler) ottenute con un radar, osservazioni da sonde spaziali, osservazioni dei tempi di transito o di occultazione con un altro corpo celeste, ecc.

[(4)] Le considerazioni seguenti assumono ovviamente che nel computo dei numeri N ed M si tenga conto separatamente di tutte le *componenti scalari* che formano le grandezze, per cui ad esempio per un vettore nello spazio occorre contare separatamente le tre coordinate cartesiane, per una osservazione astrometrica angolare occorre contare separatamente le misure di ascensione retta e declinazione, ecc.

la modellizzazione dell'osservabile a un certo grado di approssimazione, accettando come inevitabile errore tutto ciò che va oltre questa descrizione[5]. Notiamo anche che è proprio il fatto che gli scarti ε_i non seguano leggi note (e in particolare non siano rappresentabili dalle funzioni $f_i(\beta_k)$) che fa sí che le equazioni (3.1) non abbiano soluzione. Da un punto di vista puramente astratto possiamo cioè rappresentarci la situazione in questo modo: se la nostra conoscenza della realtà fosse perfetta (se non esistessero errori di misura), le misure rispecchierebbero direttamente il "vero" valore dei parametri $\beta_k^{(0)}$, cioè il valore delle misure

$$x_i^{(0)} = f_i(\beta_k^{(0)})\,, \tag{3.3}$$

pur essendo un vettore dello spazio $\Re^N$, apparterrebbe a una varietà di dimensione M, immagine della funzione $f : \Re^M \to \Re^N$. Se ora introduciamo gli scarti $\varepsilon_i^{(0)}$, la (3.3) diventa

$$x_i = x_i^{(0)} + \varepsilon_i^{(0)} = f_i(\beta_k^{(0)}) + \varepsilon_i^{(0)}\,; \tag{3.4}$$

poiché (per quanto ne possiamo sapere) $\varepsilon_i^{(0)}$ è un vettore generico di $\Re^N$, x_i *non* appartiene all'immagine $f(\Re^M)$, e quindi non esisterà alcun valore dei parametri β_k che soddisfi la (3.1). Ovviamente la situazione descritta dall'equazione (3.4) è del tutto astratta, in quanto non è in alcun modo possibile risalire ai valori "veri" dei parametri $\beta_k^{(0)}$ e ai corrispondenti valori "veri" degli scarti (che abbiamo indicato con $\varepsilon_i^{(0)}$). Vale appena la pena di notare che ovviamente non è possibile cercare di determinare contemporaneamente i valori di $\beta_k^{(0)}$ e di $\varepsilon_i^{(0)}$ dalle (3.4) che, così considerate, diventano un sistema di N equazioni in $N + M$ incognite. Pertanto, il massimo che si può cercare di ottenere è una *stima* $\hat{\beta}_k$ del valore dei parametri β_k, a cui corrisponde una stima degli scarti $\hat{\varepsilon}_i$, secondo una scomposizione del valore delle osservazioni

$$x_i = f_i(\hat{\beta}_k) + \hat{\varepsilon}_i$$

che sarà differente dalla (3.4) e che può essere definita univocamente solo introducendo ipotesi aggiuntive sul comportamento degli scarti; in particolare, una richiesta che sembra ragionevole è che gli scarti siano il più possibile piccoli. Il metodo dei minimi quadrati si basa sul principio che la stima più attendibile dei parametri $\hat{\beta}_k$ sia quella che si ottiene imponendo che la somma dei quadrati degli scarti

$$\phi = \sum_{i=1}^{N} \varepsilon_i^2 \tag{3.5}$$

sia minima. Accettando questo principio come un assioma, cerchiamo ora di ricavarne un algoritmo per il calcolo dei parametri $\hat{\beta}_k$.

3.2 Minimi quadrati (caso lineare)

Semplifichiamo la matematica del problema supponendo che le funzioni f_i siano *lineari* nei parametri β_k; in altre parole supponiamo che la (3.2) sia della forma[6]

$$x_i = A_{ik}\beta_k + \varepsilon_i \tag{3.6}$$

[5] Ad esempio nella misura astrometrica della posizione di un asteroide non è possibile tener conto nel dettaglio della complessa dinamica degli strati di atmosfera attraverso cui passa la luce che proviene dall'asteroide stesso e dalle stelle usate come riferimento, che produce deflessioni del raggio luminoso, rapidamente variabili nel tempo, che si traducono in variazioni della posizione apparente degli oggetti.

[6] Usiamo qui la notazione (convenzione di Einstein) per cui si sottintende la sommatoria su indici ripetuti: $A_{ik}\beta_k = \sum_{k=1}^{M} A_{ik}\beta_k$.

dove A_{ik} è una matrice rettangolare $N \times M$ generica (ma nota), chiamata *matrice disegno* del problema. Nella funzione da minimizzare

$$\phi = \varepsilon_i \varepsilon_i \tag{3.7}$$

gli scarti ε_i vanno considerati come funzioni dei parametri da determinare e delle misure

$$\varepsilon_i = x_i - A_{ik}\beta_k \,. \tag{3.8}$$

La soluzione che risponde al criterio dei minimi quadrati si ottiene quindi imponendo

$$\frac{\partial \phi}{\partial \beta_k} = 0$$

cioè

$$\frac{1}{2}\frac{\partial \phi}{\partial \beta_k} = \frac{\partial \varepsilon_i}{\partial \beta_k}\varepsilon_i = -A_{ik}\varepsilon_i = -A_{ik}(x_i - A_{ij}\beta_j) = 0 \tag{3.9}$$

che si può scrivere

$$A_{ik}A_{ij}\beta_j = A_{ik}x_i$$

o, usando una notazione matriciale più compatta,

$$A'A\beta = A'x \,, \tag{3.10}$$

dove con A' si è indicata la trasposta della matrice A. La matrice $A'A$ è una matrice quadrata $N \times N$ ed è chiamata *matrice normale* del problema; se essa è invertibile, la (3.10) fornisce un'unica soluzione

$$\hat{\beta} = (A'A)^{-1}A'x \,. \tag{3.11}$$

Che cosa si può dire a proposito dell'invertibilità della matrice normale? Poiché è immediato dimostrare che *il nucleo della matrice $A'A$ coincide con il nucleo della matrice A*[7], ne consegue che le matrici $A'A$ e A hanno rango uguale; quindi la matrice normale $A'A$ è invertibile (cioè ha rango M) se e solo se la matrice disegno A ha rango M, cioè il suo nucleo si riduce al nucleo banale (contiene solo il vettore nullo). A prima vista questa sembra una condizione abbastanza ragionevole per un esperimento ben progettato; infatti, se la matrice A (cioè l'applicazione $\beta \mapsto x$) ammettesse un nucleo non banale, vorrebbe dire che *esistono combinazioni lineari non banali dei parametri β che hanno effetto nullo sulle osservabili*: è quindi chiaro che non esisterebbe alcuna possibilità di determinare tali combinazioni lineari sulla base del risultato dell'esperimento. D'altra parte ciò vorrebbe anche dire che *alcuni dei parametri β_k sono inutili alla descrizione dell'esperimento* e possono essere eliminati, riparametrizzando l'esperimento e riducendo le dimensioni delle matrici $A'A$ e A fino a che M coincida con il rango di A.

Nella pratica nei casi in cui la matrice normale risulta singolare non è sempre facile capire quali siano i parametri responsabili e come sia possibile riparametrizzare l'esperimento in modo da eliminare la singolarità, e nello stesso tempo mantenere per i parametri un significato fisico comprensibile. Spesso la deficienza di rango della matrice normale è conseguenza di una *simmetria* del problema, cioè di una proprietà di invarianza della soluzione, che deve essere compresa e rimossa per poter ottenere un sistema risolubile. In casi simili è a volte utile diagonalizzare la matrice normale e studiarne gli autovalori e autovettori.

---

[7] Dimostrazione:

1) se $x \in \ker(A) \Rightarrow Ax = 0 \Rightarrow A'Ax = 0 \Rightarrow x \in \ker(A'A)$, quindi $\ker(A) \subseteq \ker(A'A)$;
2) se $x \in \ker(A'A) \Rightarrow A'Ax = 0 \Rightarrow x'A'Ax = 0$; ponendo $y = Ax$, ciò vuol dire che $y'y = \sum_i y_i^2 = 0$, il che implica $y = 0 \Rightarrow Ax = 0 \Rightarrow x \in \ker(A)$; quindi $\ker(A'A) \subseteq \ker(A)$;

le due relazioni precedenti dimostrano perciò che $\ker(A'A) \equiv \ker(A)$.

La matrice normale $A'A$ è, per costruzione, una matrice simmetrica e quindi, in base a un noto teorema di algebra lineare, è diagonalizzabile per mezzo di una matrice ortogonale; in altre parole esiste sempre una matrice ortogonale P tale che la matrice

$$D = P^{-1}\,(A'A)\,P = P'\,(A'A)\,P$$

è diagonale; i vettori colonna della matrice P sono gli autovettori della matrice normale $A'A$. Sempre per costruzione, la matrice normale è *semi-definita positiva*[8], quindi tutti i suoi autovalori sono non negativi; in altre parole la forma diagonale di $A'A$ può essere scritta come

$$D = P'\,(A'A)\,P = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N^2 \end{pmatrix}.$$

È immediato osservare che gli autovettori che corrispondono ad autovalori nulli appartengono al nucleo della matrice, di cui anzi costituiscono una base; quindi il numero di autovalori nulli è pari alla dimensione del nucleo della matrice, cioè alla sua deficienza di rango. La cosa più interessante è però che gli autovettori associati agli autovalori nulli definiscono le direzioni nello spazio dei parametri β che corrispondono a variazioni che non hanno alcun effetto sul valore delle osservabili, e quindi danno un suggerimento preciso su come modificare la formulazione dell'esperimento in modo da rimuovere la deficienza di rango (vedi un esempio di questo tipo di analisi nella sezione 3.7.3). Nel seguito supporremo comunque di non avere parametri sovrabbondanti, e che quindi la soluzione del problema dei minimi quadrati sia unica e sia data dalla (3.11).

Una notevole proprietà della soluzione (3.11) è espressa dalla (3.9), che riscriviamo come

$$\varepsilon_i A_{ik} = 0\,; \tag{3.12}$$

questa è una relazione di ortogonalità tra il vettore degli scarti e le colonne della matrice A. Poiché le colonne di A costituiscono i vettori di base della varietà di tutti i risultati conformi alla teoria, questa proprietà si può esprimere dicendo che *gli scarti sono ortogonali alla varietà dei risultati ammissibili.*

Si può quindi dare una semplice interpretazione geometrica del metodo dei minimi quadrati (Figura 3). Il risultato della misura x è un punto in uno spazio Ω di dimensione N; la teoria prevede che i risultati ammissibili dell'esperimento non siano punti qualsiasi di Ω, ma appartengano a una varietà lineare $\omega \subset \Omega$ avente dimensione $M < N$. Anche il valore "vero" x_0 del risultato dell'esperimento (il valore che si otterrebbe se si potessero eliminare tutte le cause di errore) appartiene a tale varietà: $x_0 \in \omega$. Il risultato sperimentale x però non coincide con il valore ideale x_0 perché contiene anche un errore di misura: $x = x_0 + \varepsilon_0$; siccome lo scarto ε_0 è un vettore generico di Ω, ne consegue che in generale x non appartiene alla varietà ω. Il problema consiste allora nell'associare a x un punto $\hat{x} \in \omega$ che costituisca la nostra stima di x_0: il metodo dei minimi quadrati sceglie come $\hat{x}$ il punto di ω più vicino a x secondo la metrica euclidea (3.7), cioè la proiezione ortogonale di x sulla varietà ω. In questo modo si ottiene anche una stima degli errori di misura, data da $\hat{\varepsilon} = x - \hat{x}$.

[8] Una matrice quadrata B di dimensioni $N \times N$ si dice *semi-definita positiva* quando, per qualsiasi vettore $x \in \Re^N$ non nullo, risulta $x'Bx \geq 0$; quando la diseguaglianza è soddisfatta in senso stretto ($x'Bx > 0$) la matrice si dice *definita positiva*. La matrice normale $A'A$ è sempre almeno semi-definita positiva in quanto l'espressione $x'A'Ax$ è la norma euclidea del vettore $y = Ax$

$$x'A'Ax = y'y = \sum_{i=1}^{N} y_i^2$$

e quindi non può essere negativa. Come si è visto nella nota precedente, il caso $x'A'Ax = 0$ si verifica solo quando il vettore x appartiene al nucleo di $A'A$; ne consegue che, quando la matrice normale non ha deficienze di rango (è invertibile), è sempre *definita positiva*.

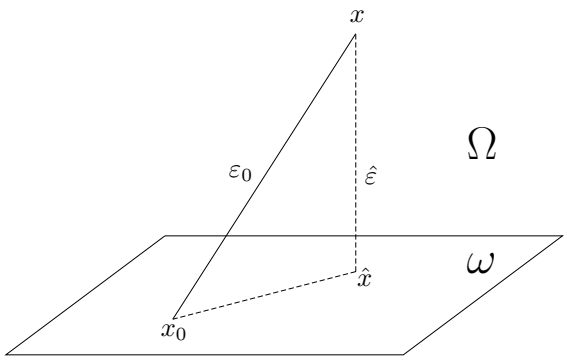

Figura 3: Interpretazione geometrica del metodo dei minimi quadrati.

3.3 La propagazione della covarianza

Il metodo dei minimi quadrati acquista pieno significato se lo si considera nell'ambito di una interpretazione probabilistica dell'esperimento, in cui si assume che gli scarti siano variabili stocastiche, cioè quantità che dipendono da fenomeni fisici che sono al di fuori del controllo e della capacità di modellizzazione dello sperimentatore, cosicché i valori di ε_i variano da una misurazione all'altra in modo imprevedibile.

Consideriamo ad esempio il caso in cui l'esperimento consista nell'osservazione astrometrica di un corpo celeste: l'osservabile (la posizione apparente dell'oggetto, cioè l'angolazione con cui la luce che proviene dall'oggetto raggiunge il rivelatore dello strumento di misura) dipende da una serie di fenomeni geometrici e fisici di cui bisogna tener conto nella riduzione della misura: la posizione geometrica del corpo celeste rispetto all'osservatore, la rifrazione atmosferica, la deflessione del raggio luminoso dovuta alla velocità propria dell'osservatore (aberrazione), ecc. Questi fenomeni obbediscono a leggi fisiche note: la loro descrizione in termini quantitativi dipende da una serie di parametri fisici, alcuni dei quali hanno valori noti[(9)], altri invece sono ignoti e devono essere introdotti nel formalismo dei minimi quadrati tra i parametri β da determinare. Si può assumere che appartengano alla prima categoria tutti i parametri necessari alla modellizzazione della rifrazione atmosferica (variazione dell'indice di rifrazione in funzione dell'altezza) e dell'aberrazione (velocità della Terra sulla sua orbita), mentre appartengono alla seconda categoria le variabili angolari che descrivono la posizione relativa dell'astro rispetto all'osservatore (ascensione retta e declinazione): lo scopo dell'esperimento è appunto la determinazione di questi due parametri.

Tuttavia, accanto a questi meccanismi, ci sono anche altri effetti che influiscono sul risultato della misura in modo imprevedibile: la radiazione luminosa, prima di raggiungere

[(9)] Si deve qui intendere *noto* nel senso di *conosciuto con una precisione sufficiente da poter calcolare l'effetto sull'osservabile con un errore trascurabile rispetto all'errore di misura.*

l'osservatore, attraversa una serie di strati d'aria di densità e temperatura differenti e rapidamente variabili a causa del moto turbolento dell'atmosfera; le ottiche del telescopio sono soggette a piccole deformazioni in conseguenza delle variazioni di temperatura e di assetto; la formazione stessa dell'immagine è un processo intrinsecamente stocastico, prodotto dell'accumularsi sul fotorecettore dei singoli fotoni. Tutti questi effetti fanno sì che, quando l'osservazione viene ripetuta più volte nelle stesse condizioni[(10)], il valore misurato sia ogni volta differente. Tutte queste componenti della misura vengono inglobate negli scarti ε_i i cui valori, riferiti a una certa realizzazione dell'esperimento, sono imprevedibili, ma tuttavia possiedono alcune regolarità di comportamento che possono essere evidenziate attraverso uno studio statistico basato sulla ripetizione dell'esperimento. In questo modo, pur senza poter predire il valore esatto assunto dagli scarti, è possibile definire la probabilità che tale valore appartenga a un certo intervallo.

In termini matematici una variabile stocastica è descritta da una *misura di probabilità*, cioè da un numero reale P, compreso tra 0 e 1, assegnato a ogni *evento*; qui per *evento* si intende un sottoinsieme S dello spazio Ω di tutti i valori possibili per la variabile (cioè di tutti i possibili esiti dell'esperimento). La misura deve essere tale che la probabilità assegnata a tutto lo spazio Ω (evento certo) sia $P(\Omega) = 1$ e la probabilità assegnata all'insieme vuoto $\emptyset$ (evento impossibile) sia $P(\emptyset) = 0$. Nelle applicazioni fisiche hanno particolare interesse le variabili stocastiche reali x (in cui lo spazio dei valori possibili Ω coincide con la retta reale: $x \in \Re$); in questi casi è comodo assegnare le probabilità dei diversi eventi per mezzo di una funzione non negativa $p(x)$ (densità di probabilità) tale che la probabilità di un qualsiasi evento $S \subseteq \Re$ sia uguale all'integrale di $p(x)$ esteso ad S

$$P(S) = \int_S p(x)\, dx\,;$$

ovviamente deve valere la condizione di normalizzazione

$$P(\Omega) = \int_{-\infty}^{+\infty} p(x)\, dx = 1\,.$$

Il *valore atteso* o di aspettazione di una qualsiasi funzione $y(x)$ della variabile stocastica x è definito da

$$E[y] = \int_{\Re} y(x)\, p(x)\, dx\,. \tag{3.13}$$

A partire da questa espressione si definiscono poi alcune quantità integrali della variabile stocastica chiamate *momenti*: si dice *momento semplice* di ordine k della variabile stocastica x il valore di aspettazione di x^k, cioè la quantità

$$m_k = E\left[x^k\right] = \int_{\Re} x^k\, p(x)\, dx\,; \tag{3.14}$$

in particolare il momento semplice del primo ordine $m_1 = \overline{x}$ è il *valore medio* (o *valore di aspettazione*) di x. Analogamente si dice *momento centrato* di ordine k la quantità

$$\mu_k = E\left[(x - \overline{x})^k\right] = \int_{\Re} (x - \overline{x})^k\, p(x)\, dx\,; \tag{3.15}$$

(10) L'espresssione *nelle stesse condizioni* indica qui, in modo abbastanza tautologico, l'esatta eguaglianza della parte deterministica del modello dell'esperimento (cioè l'eguaglianza sia della matrice disegno A, sia dei parametri β_k). Nella maggior parte dei casi reali una ripetizione dell'osservazione rigorosamente *nelle stesse condizioni* è impossibile, perché ripetere l'esperimento significa eseguirlo a tempi diversi, e lo scorrere del tempo inevitabilmente modifica alcune delle variabili (nel caso dell'osservazione astrometrica, la posizione della Terra, l'altezza dell'astro sull'orizzonte, ecc.). Tuttavia spesso in pratica è ugualmente possibile distinguere tra la parte deterministica e la parte stocastica della misura; ad esempio, nel caso di un'osservazione astrometrica, la variazione delle condizioni tra osservazioni ripetute a distanza di pochi minuti è così piccola che la variazione della parte deterministica è, con grande approssimazione, lineare nel tempo, mentre la componente stocastica ha variazioni a frequenza molto più elevata.

il momento centrato del secondo ordine $\mu_2 = \mathrm{var}(x) = \sigma_x^2$ è la *varianza* della variabile x; la sua radice quadrata σ_x è la *deviazione standard* di x.

L'importanza dei momenti consiste nel fatto che essi (e soprattutto i momenti di ordine più basso) forniscono alcune indicazioni molto generali sulle proprietà della variabile stocastica: il valore medio $\overline{x}$ indica il valore tipico attorno a cui si concentrano i risultati dell'esperimento, mentre la deviazione standard σ_x dà una misura approssimata della deviazione dei singoli risultati dal valore medio. In molti casi di interesse pratico non si hanno sufficienti informazioni sul comportamento della variabile stocastica tali da permettere di stimare in modo attendibile la forma della funzione $p(x)$ e ci si limita quindi a descrivere la variabile attraverso i suoi momenti del primo e secondo ordine $\overline{x}$ e σ_x. Tale descrizione è utile soprattutto perché risulta *chiusa* rispetto a trasformazioni lineari della variabile stessa, nel senso che è possibile calcolare valore medio e varianza di una qualsiasi funzione lineare di x conoscendo solo valore medio e varianza di x; è immediato infatti verificare che, se $y = ax + b$ (dove a e b sono valori arbitrari ma costanti, cioè *non* variabili stocastiche), risulta[(11)]

$$\overline{y} = \overline{x} + b\,; \qquad \sigma_y^2 = a^2\sigma_x^2$$

In modo analogo si possono definire le probabilità quando la variabile stocastica è multidimensionale, cioè assume valori $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ in $\Re^n$; in tal caso la densità di probabilità è una funzione di più variabili $p(x_1, x_2, \ldots, x_n)$ e la probabilità di un evento $S \subseteq \Re^n$ è definita dall'integrale multiplo

$$P(S) = \int_S p(x_1, x_2, \ldots, x_n)\, d^n x\,; \tag{3.16}$$

analogamente il valore di aspettazione di una qualsiasi funzione $y(x_1, x_2, \ldots, x_n)$ è dato da

$$E[y] = \int_{\Re^n} y(x_1, x_2, \ldots, x_n)\, p(x_1, x_2, \ldots, x_n)\, d^n x\,.$$

Il valore medio di $\mathbf{x}$ è un vettore che ha per componenti i valori medi delle componenti di $\mathbf{x}$:

$$\overline{\mathbf{x}} = (\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_n)\,,$$

dove

$$\overline{x}_i = E[x_i] = \int_{\Re^n} x_i\, p(x_1, x_2, \ldots, x_n)\, d^n x\,,$$

mentre il concetto di varianza deve essere esteso, nel caso multidimensionale, a quello di *matrice di covarianza*, una matrice simmetrica $n \times n$ definita (in notazione matriciale) come

$$\mathrm{Cov}(\mathbf{x}) = E[(\mathbf{x} - \overline{\mathbf{x}})(\mathbf{x} - \overline{\mathbf{x}})']\,;$$

le componenti della matrice di covarianza $\mathrm{Cov}(\mathbf{x}) = (c_{ij})$ sono cioè definite come

$$c_{ij} = E[(x_i - \overline{x}_i)(x_j - \overline{x}_j)] = \int_{\Re^n} (x_i - \overline{x}_i)(x_j - \overline{x}_j)\, p(x_1, x_2, \ldots, x_n)\, d^n x\,.$$

Gli elementi diagonali $c_{ii} = E[(x_i - \overline{x}_i)^2]$ della matrice di covarianza sono le varianze $\sigma_{x_i}^2$ dei singoli elementi x_i del vettore $\mathbf{x}$, mentre gli elementi fuori diagonale $c_{ij} = E[(x_i - \overline{x}_i)(x_j - \overline{x}_j)]$ sono chiamati *covarianze* tra gli elementi x_i e x_j. Questa estensione del concetto di varianza a matrice di covarianza, oltre a essere in qualche modo naturale, è anche richiesta dalla necessità di mantenere la proprietà di chiusura della descrizione della variabile stocastica basata sui momenti del primo e del secondo ordine rispetto a trasformazioni lineari anche

[(11)] Tale chiusura della descrizione basata sui momenti del primo e del secondo ordine non sussiste più nel caso di trasformazioni generiche: se cioè $y(x)$ è una funzione qualsiasi di x, il calcolo dei suoi momenti richiede la conoscenza della densità di probabilità $p(x)$, cioè il calcolo di un integrale analogo all'espressione (3.13).

al caso multidimensionale; infatti se, analogamente a quanto fatto nel caso di una variabile monodimensionale, definiamo $y = \sum_{i=1}^{n} a_i x_i + b$ (dove a_i e b sono costanti), il valore medio di y risulta (tenendo conto del fatto che il valore di aspettazione è un funzionale lineare)

$$\overline{y} = E[y] = E\left[\sum_{i=1}^{n} a_i x_i + b\right] = \sum_{i=1}^{n} a_i E\left[x_i\right] + b = \sum_{i=1}^{n} a_i \overline{x}_i + b\,,$$

mentre la sua varianza vale

$$\begin{aligned} \sigma_y^2 &= E\left[(y-\overline{y})^2\right] = E\left[\textstyle\sum_{i=1}^{n} a_i (x_i - \overline{x}_i) \sum_{j=1}^{n} a_j (x_j - \overline{x}_j)\right] \\ &= E\left[\textstyle\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (x_i - \overline{x}_i)(x_j - \overline{x}_j)\right] = \\ &= \textstyle\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j E\left[(x_i - \overline{x}_i)(x_j - \overline{x}_j)\right] \\ &= \textstyle\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j c_{ij}\,, \end{aligned} \tag{3.17}$$

cioè il calcolo della varianza di y richiede la conoscenza di tutte le componenti della matrice di covarianza di $\mathbf{x}$ (non sono sufficienti le sole varianze delle componenti $\sigma^2_{x_i}$)[12]. Più in generale se consideriamo il caso di una funzione lineare vettoriale, cioè definiamo $\mathbf{y} = (y_1, y_2, \ldots, y_m) \in \Re^m$ come una funzione lineare di $\mathbf{x}$

$$\mathbf{y} = B\,\mathbf{x}\,,$$

dove B è una matrice $m \times n$, risulta

$$\overline{\mathbf{y}} = E[\mathbf{y}] = E[B\mathbf{x}] = B\,E[\mathbf{x}] = B\overline{\mathbf{x}}\,, \tag{3.18}$$

mentre la matrice di covarianza si trasforma secondo una relazione di equivalenza

$$\begin{aligned} \mathrm{Cov}(\mathbf{y}) &= E[(\mathbf{y}-\overline{\mathbf{y}})(\mathbf{y}-\overline{\mathbf{y}})'] = E[(B\mathbf{x} - B\overline{\mathbf{x}})(B\mathbf{x} - B\overline{\mathbf{x}})'] \\ &= E[B(\mathbf{x}-\overline{\mathbf{x}})(\mathbf{x}-\overline{\mathbf{x}})'B'] = B\,E[(\mathbf{x}-\overline{\mathbf{x}})(\mathbf{x}-\overline{\mathbf{x}})']\,B' \\ &= B\,\mathrm{Cov}(\mathbf{x})\,B' \end{aligned} \tag{3.19}$$

(*legge di propagazione della covarianza*).

La covarianza tra due grandezze ha dimensioni fisiche pari al prodotto delle dimensioni delle grandezze corrispondenti (ad esempio la covarianza tra due lunghezze ha le dimensioni di un'area). È comune descrivere la covarianza c_{ij} per mezzo del *coefficiente di correlazione* ρ_{ij}, definito come il rapporto tra la covarianza e il prodotto delle deviazioni standard $\sigma_i\,\sigma_j$

$$\rho_{ij} = \frac{c_{ij}}{\sigma_i\,\sigma_j} = \frac{E[(x_i - \overline{x}_i)(x_j - \overline{x}_j)]}{\sqrt{E\left[(x_i - \overline{x}_i)^2\right]\,E\left[(x_j - \overline{x}_j)^2\right]}}$$

Il coefficiente di correlazione ρ_{ij} è una quantità adimensionale, il cui valore è sempre compreso tra -1 e $+1$, cioè $|\rho_{ij}| \leq 1$[13].

[12] L'equazione (3.17) dimostra che la matrice di covarianza è sempre una matrice *definita positiva* (o almeno *semi-definita positiva*). Una matrice quadrata C è detta *definita positiva* (risp. *semi-definita positiva*) quando per ogni vettore $a \in \Re^n$ diverso dal vettore nullo vale $a'\,C\,a > 0$ (risp. ≥ 0). L'equazione (3.17) mostra che la forma quadratica $a'\,\mathrm{Cov}(\mathbf{x})\,a$ può sempre essere interpretata come la varianza di una opportuna combinazione lineare delle variabili x_i, e dunque non può essere negativa.

[13] Dimostrazione: indicando per brevità di notazione $\Delta x_i = x_i - \overline{x}_i$, consideriamo la varianza σ^2 dell'espressione $\alpha\Delta x_i - \Delta x_j$

$$\begin{aligned} \sigma^2 &= E\left[(\alpha\Delta x_i - \Delta x_j)^2\right] = \alpha^2\,E\left[\Delta x_i^2\right] - 2\alpha\,E\left[\Delta x_i \Delta x_j\right] + E\left[\Delta x_j^2\right] \\ &= \alpha^2\sigma_i^2 - 2\alpha\,c_{ij} + \sigma_j^2\,. \end{aligned}$$

Poiché deve risultare $\sigma^2 \geq 0$ per ogni valore di α, ne segue che il discriminante della forma quadratica $\alpha^2\sigma_i^2 - 2\alpha\,c_{ij} + \sigma_j^2$ non può essere positivo

$$c_{ij}^2 - \sigma_i^2\sigma_j^2 \leq 0\,,$$

da cui $c_{ij}^2/(\sigma_i^2\sigma_j^2) \leq 1$.

Una semplificazione notevole del calcolo delle probabilità di eventi per una variabile multidimensionale si verifica quando la densità di probabilità congiunta $p(x_1, x_2, \ldots, x_n)$ si riduce al prodotto di funzioni di una sola variabile

$$p(x_1, x_2, \ldots, x_n) = p_1(x_1)\, p_2(x_2) \cdots p_n(x_n)\,. \tag{3.20}$$

In tal caso la probabilità di un evento definito solo attraverso limiti su una delle variabili (diciamo x_1), ad esempio $P\{a \le x_1 \le b\}$, risulta

$$P\{a \le x_1 \le b\} = \int_a^b p_1(x_1)\, dx_1 \int_{-\infty}^{+\infty} p_2(x_2)\, dx_2 \cdots \int_{-\infty}^{+\infty} p_n(x_n)\, dx_n\,,$$

cioè, tenendo conto della condizione di normalizzazione $\int_{-\infty}^{+\infty} p_k(x_k)\, dx_k = 1$,

$$P\{a \le x_1 \le b\} = \int_a^b p_1(x_1)\, dx_1\,.$$

In altre parole in questo caso le diverse componenti x_k della variabile multidimensionale si comportano in modo del tutto indipendente; l'esperimento si riduce a un insieme di esperimenti indipendenti, ciascuno descritto separatamente da una propria densità di probabilità $p_k(x_k)$. Per questo motivo variabili stocastiche la cui densità di probabilità congiunta è fattorizzabile come dall'equazione (3.20) sono chiamate *variabili indipendenti*.

È immediato verificare che variabili stocastiche *indipendenti* sono anche *non correlate*, nel senso che il loro coefficiente di correlazione è uguale a zero. Infatti, se $p(x_i, x_j) = p_i(x_i)\, p_j(x_j)$, risulta

$$\begin{aligned} c_{ij} &= \int_{\Re^2} (x_i - \overline{x}_i)(x_j - \overline{x}_j)\, p(x_i, x_j)\, dx_i\, dx_j \\ &= \int_{-\infty}^{+\infty} (x_i - \overline{x}_i)\, p(x_i)\, dx_i \int_{-\infty}^{+\infty} (x_j - \overline{x}_j)\, p(x_j)\, dx_j \\ &= 0\,, \end{aligned}$$

dato che, per definizione, $\int_{-\infty}^{+\infty} (x_i - \overline{x}_i)\, p(x_i)\, dx_i = 0$. L'implicazione inversa, ovviamente, non vale: variabili non correlate non sono necessariamente indipendenti (non è possibile dedurre una proprietà puntuale della funzione densità di probabilità da una sua proprietà integrale). Tuttavia nell'ottica, spiegata in precedenza, in cui si rinuncia alla descrizione puntuale della densità di probabilità e ci si limita a utilizzare i suoi momenti fino al secondo ordine, il concetto di non-correlazione è quello che più si avvicina a quello di indipendenza e in qualche modo (molto impreciso) ne prende il posto[(14)].

Il significato della correlazione può essere forse meglio compreso con un semplice esempio. Supponiamo di avere una coppia di variabili stocastiche x e y, aventi rispettivamente varianza σ_x^2 e σ_y^2 e correlazione $c_{xy} = \rho_{xy}\sigma_x\sigma_y$. Cerchiamo ora di *prevedere* il valore di y dalla misura di x cioè, più precisamente, di costruire una funzione di x che sia uno *stimatore* di y. Cerchiamo lo stimatore $\hat{y}$ nella forma di una funzione lineare di x; definiamo cioè

$$\hat{y} = ax + b\,,$$

[(14)] Ad esempio è immediato dimostrare che, se $y = ax + b$ (y è una funzione lineare di x, caso che in qualche modo si può considerare come opposto a quello in cui x e y sono indipendenti), risulta $\sigma_y^2 = a^2\sigma_x^2$ e

$$c_{xy} = E[(x - \overline{x})(y - \overline{y})] = E[(x - \overline{x})(ax - a\overline{x})] = a\, E[(x - \overline{x})(x - \overline{x})] = a\, \sigma_x^2\,;$$

quindi il coefficiente di correlazione assume il massimo valore possibile

$$\rho_{xy} = \frac{c_{xy}}{\sqrt{\sigma_x^2 \sigma_y^2}} = \frac{a\, \sigma_x^2}{\sqrt{a^2 \sigma_x^4}} = \frac{a}{|a|} = \pm 1\,.$$

dove a e b sono due costanti, i cui valori devono essere determinati in modo che lo stimatore sia il più preciso possibile. In particolare richiediamo:

1) che il valore di aspettazione dello stimatore sia uguale al valore di aspettazione della quantità da stimare: $E[\hat{y}] = E[y]$. Se definiamo l'*errore di stima* δ come

$$\delta = \hat{y} - y\,,$$

questa condizione equivale a richiedere che il valore medio dell'errore di stima sia nullo

$$E[\delta] = 0\,;$$

2) che la varianza dell'errore di stima δ sia la più piccola possibile.

La prima condizione si può scrivere

$$E[\delta] = E[\hat{y}] - E[y] = a\overline{x} + b - \overline{y} = 0\,,$$

da cui segue immediatamente

$$b = \overline{y} - a\overline{x}\,;$$

in altre parole la costante additiva b deve essere scelta in modo da compensare la differenza tra i valori medi di y e di ax; per semplicità di notazione nel seguito possiamo assumere, senza perdita di generalità, che $\overline{x} = \overline{y} = 0$ e quindi $b = 0$.

La varianza dell'errore di stima può essere calcolata come caso particolare della legge di propagazione della covarianza (3.17):

$$\sigma_\delta^2 = a^2\sigma_x^2 + \sigma_y^2 - 2ac_{xy} = a^2\sigma_x^2 + \sigma_y^2 - 2a\sigma_x\sigma_y\rho_{xy}\,.$$

Per trovare il valore di a che rende minimo σ_δ^2 imponiamo

$$\frac{d\sigma_\delta^2}{da} = 2a\sigma_x^2 - 2\sigma_x\sigma_y\rho_{xy} = 0\,,$$

da cui

$$a = \frac{\sigma_y}{\sigma_x}\,\rho_{xy}\,; \tag{3.21}$$

adottando questo valore ottimale di a, la varianza dell'errore di stima risulta

$$\sigma_\delta^2 = \left(1 - \rho_{xy}^2\right)\,\sigma_y^2\,. \tag{3.22}$$

Consideriamo ora due casi particolari estremi. Se x e y sono non correlate ($\rho_{xy} = 0$), la formula (3.21) fornisce $a = 0$, cioè $\hat{y} = 0$: il modo migliore di prevedere y è usare semplicemente il suo valore medio! (naturalmente la varianza di questo stimatore banale è uguale alla varianza di y, come risulta dalla (3.22)). In altre parole, la misura di x non fornisce alcuna informazione aggiuntiva sul valore di y. All'estremo opposto, se la correlazione tra x e y è massima ($\rho_{xy} = \pm 1$), lo stimatore migliore si ottiene ponendo $a = \pm\sigma_y/\sigma_x$. In questo caso la varianza dell'errore di stima è nulla: la conoscenza di x permette di prevedere esattamente y, cioè y è una funzione lineare di x[15]. Naturalmente nei casi intermedi (in cui $0 < |\rho_{xy}| < 1$) si ha un risultato intermedio: la conoscenza di x permette di prevedere il valore di y con una precisione tanto maggiore quanto più grande è il valore assoluto della correlazione.

Ritornando al metodo dei minimi quadrati, spesso si assume (per semplicità o per mancanza di informazioni più precise) che gli scarti ε_i siano variabili stocastiche non correlate

[15] Combinando questo risultato con quello della nota precedente possiamo affermare che due variabili stocastiche hanno correlazione uguale a ± 1 se e solo se sono legate da una relazione lineare del tipo $y = ax + b$.

e aventi la stessa varianza σ_0^2; si assume cioè che la matrice di covarianza degli scarti sia proporzionale alla matrice unità I

$$C_{\varepsilon\varepsilon} = \sigma_0^2 \, I \, ; \tag{3.23}$$

applicando la legge di propagazione della covarianza (3.19) allo stimatore (3.11) si ottiene allora l'espressione della matrice di covarianza dei parametri stimati

$$\begin{array}{rcl} C_{\hat{\beta}\hat{\beta}} & = & (A'A)^{-1} A' C_{\varepsilon\varepsilon} A (A'A)^{-1} = \sigma_0^2 (A'A)^{-1} A' I A (A'A)^{-1} \\ & = & \sigma_0^2 (A'A)^{-1} \end{array} \tag{3.24}$$

3.4 Minimi quadrati pesati

La scelta di minimizzare la funzione definita dalla (3.5) è naturale quando la matrice di covarianza degli scarti è della forma (3.23) (se tutti gli scarti hanno uguale varianza e sono non correlati, non c'è ragione di pesarli in modo diverso nella forma quadratica ϕ). Più rigorosamente, è possibile dimostrare che, nell'ipotesi (3.23), lo stimatore che deriva dalla (3.5) gode di alcune proprietà di ottimalità: in particolare, è lo stimatore *lineare* di minore varianza. Se inoltre si assume che la distribuzione degli scarti sia *normale*, si ha un risultato ancora più forte: lo stimatore (3.11) è quello di minore varianza tra tutti gli stimatori (lineari e non) *centrati*, cioè che conservano il valore medio.

Consideriamo ora il caso in cui la matrice di covarianza degli scarti $C_{\varepsilon\varepsilon}$ sia piena; questo caso si può sempre ricondurre al caso (3.23) (scarti non correlati e varianze uguali) attraverso una opportuna trasformazione lineare

$$\eta = P\varepsilon \, , \tag{3.25}$$

tale che la nuova matrice di covarianza $C_{\eta\eta}$ sia la matrice unità[16]

$$C_{\eta\eta} = P C_{\varepsilon\varepsilon} P' = I \, . \tag{3.26}$$

Applicando la trasformazione P, il problema originario (3.6) può essere formalmente trasformato in un nuovo problema ai minimi quadrati

$$y = B\beta + \eta \qquad (\text{dove } y = Px, \; B = PA)$$

in cui la matrice di covarianza degli scarti η è uguale alla matrice unità. Dalla (3.26) risulta

$$C_{\varepsilon\varepsilon} = P^{-1} P'^{-1} = (P'P)^{-1}$$

e quindi la funzione da minimizzare (3.5) diventa

$$\phi = \varepsilon' P' P \varepsilon = \varepsilon' W \varepsilon \, ; \tag{3.27}$$

il criterio dei minimi quadrati è stato modificato introducendo una *matrice dei pesi* $W = P'P$ uguale all'inversa della matrice di covarianza degli scarti

$$W = C_{\varepsilon\varepsilon}^{-1} \, . \tag{3.28}$$

[16] Dimostrazione: siccome $C_{\varepsilon\varepsilon}$ è una matrice definita positiva, può sempre essere diagonalizzata con una matrice ortogonale; esiste cioè una matrice ortogonale Q tale che

$$Q C_{\varepsilon\varepsilon} Q' = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2) \, ,$$

e gli autovalori σ_i^2 sono positivi; perciò la matrice P definita da

$$P = \mathrm{diag}(1/\sigma_1, 1/\sigma_2, \ldots, 1/\sigma_n) \, Q$$

soddisfa la (3.26).

La soluzione, applicando la (3.11), risulta

$$\hat{\beta} = (B'B)^{-1}B'y = (A'P'PA)^{-1}A'P'Px = (A'WA)^{-1}A'Wx\,; \tag{3.29}$$

la matrice di covarianza dei parametri stimati $\hat{\beta}$ è data da

$$C_{\hat{\beta}\hat{\beta}} = (A'WA)^{-1}\,. \tag{3.30}$$

Geometricamente, il criterio (3.27) può ancora essere interpretato come la ricerca del punto x_0 dell'ipersuperficie delle soluzioni ammissibili più vicino al valore osservato x, in uno spazio in cui la metrica non è più quella euclidea, ma è data dalla forma quadratica (3.27).

3.5 La stima di σ_0^2

Spesso l'informazione *a priori* che si ha sulla matrice di covarianza delle osservabili (cioè degli errori di misura) è incompleta; in tali casi spesso si assume come matrice di covarianza un'espressione del tipo (3.23), in cui la costante di proporzionalità σ_0^2 non è nota e deve essere stimata *a posteriori* dai risultati della stima. Uno stimatore di σ_0^2 sarebbe $\varepsilon'\varepsilon/N$, ma dobbiamo notare che non conosciamo i valori "veri" degli scarti ε, ma solo una loro stima $\hat{\varepsilon}$, dipendente dalla stima dei parametri (3.11); le due quantità sono legate dalla relazione

$$\varepsilon = x - A\beta = (x - A\hat{\beta}) + A(\hat{\beta} - \beta) = \hat{\varepsilon} + A(\hat{\beta} - \beta)\,. \tag{3.31}$$

In questa equazione, $A(\hat{\beta} - \beta)$ è un vettore appartenente all'immagine di A, mentre $\hat{\varepsilon}$ è ortogonale all'immagine di A (equazione (3.12)). Possiamo quindi scegliere in $\Re^N$ una nuova base ortonormale tale che il vettore ε si trasformi in un vettore $\delta = T\varepsilon$ (dove T è una matrice ortogonale) i cui primi M elementi costituiscono una base per la varietà delle soluzioni possibili, e i rimanenti $N - M$ elementi costituiscono una base per la varietà degli scarti stimati $\hat{\varepsilon}$; in questa nuova base il vettore $\hat{\varepsilon}$ assume dunque la forma[17]

$$\hat{\varepsilon} = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_N) \mapsto (0, 0, \ldots, 0, \delta_{M+1}, \delta_{M+2}, \ldots, \delta_N)\,. \tag{3.32}$$

Notiamo a questo punto che una trasformazione ortogonale T lascia invariata la forma (3.23) della matrice di covarianza

$$C_{\delta\delta} = \sigma_0^2 TIT' = \sigma_0^2 TT^{-1} = \sigma_0^2 I\,,$$

per cui il calcolo della varianza di $\hat{\varepsilon}$ risulta immediato

$$E[\hat{\varepsilon}'\hat{\varepsilon}] = \sum_{i=M+1}^{N} E[\delta_i^2] = (N - M)\sigma_0^2\,,$$

da cui si ricava l'espressione di uno stimatore per σ_0^2

$$\hat{\sigma}_0^2 = \frac{1}{N - M}\hat{\varepsilon}'\hat{\varepsilon}\,. \tag{3.33}$$

In modo analogo si può procedere se si assume per la matrice di covarianza degli errori un modello del tipo

$$C_{\varepsilon\varepsilon} = \sigma_0^2\Sigma\,, \tag{3.34}$$

dove Σ è una matrice nota e σ_0^2 è un parametro da determinare[18]; in questo caso l'espressione dello stimatore per σ_0^2 è

$$\hat{\sigma}_0^2 = \frac{1}{N - M}\hat{\varepsilon}'\Sigma^{-1}\hat{\varepsilon}\,. \tag{3.35}$$

[17] Riprendendo l'analogia geometrica della Figura 3, si può dire che M componenti del vettore degli scarti sono proiettate sulla varietà ω e determinano l'errore di stima $\hat{\beta} - \beta$; solo le rimanenti $N - M$ componenti rimangono nel vettore dei residui $\hat{\varepsilon}$, la cui "lunghezza" è perciò ridotta da $\sqrt{N}\sigma_0$ a $\sqrt{N - M}\sigma_0$.

[18] Si noti che il valore di σ_0^2 è ininfluente sulla soluzione (3.29), che è invariante rispetto a un fattore moltiplicativo arbitrario in W, per cui si può porre ad esempio $W = \Sigma^{-1}$.

3.6 Linearizzazione di un problema non lineare

Ritorniamo ora al caso di un problema di stima *non lineare*, descritto da un'equazione generica del tipo (3.1); questo problema si può risolvere approssimativamente con un metodo di linearizzazione e approssimazioni successive, che si compone dei seguenti passi:

1) si parte da una stima iniziale dei parametri $\beta_k^{(0)}$, sufficientemente vicina al valore "vero", a cui corrisponde una soluzione di tentativo $x_i^{(0)} = f_i(\beta_k^{(0)})$;

2) si linearizza l'equazione (3.1) nell'intorno di $\beta_k^{(0)}$ con uno sviluppo di Taylor troncato al primo termine
$$x_i^{(0)} + \Delta x_i = f_i(\beta_k^{(0)}) + \frac{\partial f_i}{\partial \beta_k}\Delta\beta_k \,; \tag{3.36}$$
le equazioni di osservazione per gli incrementi Δx, $\Delta\beta$ sono ora del tipo della (3.6), dove il ruolo della matrice A è svolto dalla matrice jacobiana della funzione f calcolata nel punto $\beta^{(0)}$;

3) si risolve il problema così linearizzato secondo il procedimento solito (equazione (3.11) o (3.29)) e si trova una nuova stima dei parametri $\hat{\beta} = \beta^{(0)} + \Delta\hat{\beta}$;

4) si assume la soluzione trovata come nuovo punto di linearizzazione $\beta^{(0)}$ e si itera il procedimento fino a convergenza, cioè finché la variazione dei parametri $\Delta\beta$ tra un passo e il successivo è molto più piccola della loro incertezza statistica (data dalla (3.24) o (3.30)).

3.7 Esempi di applicazione dei minimi quadrati

3.7.1 Misurazione diretta di una grandezza fisica

Come primo esempio, estremamente semplice, di applicazione del metodo dei minimi quadrati, consideriamo il caso in cui si misuri direttamente e ripetutamente il parametro che si vuole determinare. In questo caso le equazioni di osservazione sono semplicemente

$$x_i = \beta_1 + \varepsilon_i \qquad (i = 1, \ldots, N)$$

(dove N è il numero di ripetizioni della misura), cioè le osservazioni coincidono con il parametro da determinare, a meno degli errori di misura. La matrice A è quindi costituita da una sola colonna

$$A = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

e la soluzione (3.11) prende la forma

$$\hat{\beta} = \frac{1}{N}\sum_{i=1}^{N} x_i \,,$$

cioè la stima migliore del valore della quantità misurata è data dalla *media aritmetica* dei valori delle diverse misurazioni. La varianza dello stimatore $\hat{\beta}$ è data da

$$\sigma_{\hat{\beta}}^2 = \frac{\sigma_0^2}{N} \,,$$

dove σ_0^2 è la varianza delle singole misure x_i nel senso della (3.23); cioè ripetendo una misura N volte e prendendo la media dei risultati l'errore quadratico medio decresce come $1/\sqrt{N}$. Naturalmente ciò vale solo se gli errori delle singole misure sono effettivamente scorrelati: ogni possibile componente di errore sistematico *non* si riduce come $1/\sqrt{N}$!

Nel caso in cui gli errori di misura ε_i siano scorrelati ma abbiano varianze diverse (ad esempio perché le corrispondenti misure sono state ottenute con strumenti differenti), cioè la matrice di covarianza degli errori è diagonale

$$C_{\varepsilon\varepsilon} = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N^2 \end{pmatrix},$$

anche la matrice dei pesi risulta diagonale

$$W = C_{\varepsilon\varepsilon}^{-1} = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & w_N \end{pmatrix}$$

con $w_i = 1/\sigma_i^2$, e la soluzione (3.29) si scrive

$$\hat{\beta} = \frac{\sum_{i=1}^N w_i x_i}{\sum_{i=1}^N w_i},$$

cioè si ottiene dalla *media pesata* delle misure, con pesi pari al reciproco delle varianze degli errori corrispondenti. In questo caso la varianza dello stimatore $\hat{\beta}$ vale

$$\sigma_{\hat{\beta}}^2 = \left(\sum_{i=1}^N \frac{1}{\sigma_i^2} \right)^{-1}.$$

3.7.2 Interpolazione di una funzione assegnata per punti

Come secondo esempio consideriamo il problema, di interesse estremamente generale, dell'interpolazione di un insieme di dati con una funzione che appartenga a una certa famiglia. Supponiamo di avere una serie di coppie di valori (x_i, t_i) $(i = 1, \ldots, N)$ che consideriamo come il risultato del campionamento di una funzione $x(t)$ in corrispondenza dell'insieme dei valori $\{t_i\}$ della variabile dipendente, e supponiamo di voler rappresentare (approssimare) la funzione $x(t)$ come combinazione lineare di un certo numero M di *funzioni di base* (supposte note) $f_k(t)$ $(k = 1, \ldots, M)$

$$\hat{x}(t) = \sum_{k=1}^M \beta_k f_k(t). \tag{3.37}$$

Scegliamo come rappresentazione migliore di $x(t)$ la funzione $\hat{x}(t)$ che minimizza la somma degli scarti quadratici calcolati nei punti di base $\{t_i\}$, cioè quella che rende minima la quantità

$$\phi = \sum_{i=1}^N \left[\hat{x}(t_i) - x(t_i) \right]^2.$$

Con queste assunzioni il problema si traduce in una stima ai minimi quadrati dei coefficienti β_k della (3.37) a partire dalle "misure" x_i, con una matrice disegno data da

$$A_{ik} = f_k(t_i).$$

Come esempio concreto consideriamo il caso dell'*interpolazione lineare*, in cui si cerca una funzione interpolante della forma

$$\hat{x}(t) = \beta_1 + \beta_2 t\,;$$

risulta quindi

$$A = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \\ 1 & t_N \end{pmatrix}, \quad A'A = \begin{pmatrix} N & \sum t_i \\ \sum t_i & \sum t_i^2 \end{pmatrix}, \quad A'x = \begin{pmatrix} \sum x_i \\ \sum x_i t_i \end{pmatrix},$$

da cui deriva

$$(A'A)^{-1} = \frac{1}{\Delta} \begin{pmatrix} \sum t_i^2 & -\sum t_i \\ -\sum t_i & N \end{pmatrix}, \qquad \text{con} \quad \Delta = N \sum t_i^2 - \left(\sum t_i\right)^2$$

e

$$\hat{\beta}_1 = \frac{\sum x_i \sum t_i^2 - \sum x_i t_i \sum t_i}{\Delta}, \qquad \hat{\beta}_2 = \frac{N \sum x_i t_i - \sum x_i \sum t_i}{\Delta}.$$

3.7.3 Esempio di deficienza di rango

Come semplice esempio di deficienza di rango consideriamo il caso in cui si abbiano un certo numero M di punti lungo una retta e si vogliano determinare le loro ascisse ξ_i ($i = 1, \ldots, M$) misurando le loro differenze di posizione $\delta_{ik} = \xi_i - \xi_k$. Nel caso $M = 4$ la relazione tra parametri e osservabili ha la forma[19]

$$\begin{pmatrix} \delta_{12} \\ \delta_{13} \\ \delta_{14} \\ \delta_{23} \\ \delta_{24} \\ \delta_{34} \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{pmatrix};$$

la corrispondente matrice normale

$$A'A = \begin{pmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix} \tag{3.38}$$

ha determinante nullo e quindi non è invertibile. Analizziamo quindi gli autovalori e autovettori della matrice normale. In generale la diagonalizzazione per via analitica di una matrice $M \times M$ passa attraverso la soluzione di una equazione algebrica di grado M e quindi non è banale; tuttavia nei casi pratici è possibile utilizzare uno dei tanti algoritmi numerici che fanno parte delle librerie di algebra matriciale. Nel caso in esame utilizzando uno di questi metodi si trova che la matrice normale (3.38) possiede un autovalore triplo $\lambda_{1,2,3} = 4$ (non ci interessa qui specificare i tre autovettori a cui esso è associato) e un autovalore singolo $\lambda_4 = 0$, a cui è associato l'autovettore

$$v_4 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

[19] Il valore $M = 4$ è stato scelto appositamente come valore minimo di M per cui il numero delle osservabili $N = M(M-1)/2$ risulta maggiore di M.

Questo è il nucleo non banale delle matrici $A'A$ e A, responsabile della deficienza di rango[20]. L'espressione dell'autovettore indica qual'è il problema: sommando alle ascisse di tutti i punti una stessa quantità (cioè applicando la trasformazione $\xi_i \mapsto \xi_i + \Delta\xi$, dove $\Delta\xi$ è lo stesso per ogni i) i valori delle osservabili non cambiano. Questa proprietà della matrice disegno è legata a una evidente proprietà di invarianza del problema geometrico di partenza: cambiando la posizione del punto di origine dell'asse ξ, tutte le differenze di coordinate rimangono uguali[21]. Questa osservazione dà anche un suggerimento su come rimuovere la degenerazione. In generale ciò può essere ottenuto in diversi modi; elenchiamo alcune delle soluzioni possibili nel caso in esame:

1) il metodo più semplice consiste nell'assumere come origine degli assi uno dei punti, ad esempio il primo, imponendo per definizione $\xi_1 = 0$ ed escludendo ξ_1 dall'insieme dei parametri da determinare; in questo modo il numero delle incognite è ridotto a $M = 3$, l'invarianza del problema è rimossa e la matrice normale risultante non è più singolare;

2) un metodo più elegante consiste nell'introdurre un vincolo sulla combinazione lineare di parametri che non ha effetto sulle osservabili, nel nostro caso la somma delle ascisse dei punti, che è lo stesso che dire il loro valore medio, cioè l'ascissa del loro baricentro. A questo scopo esiste una variante del metodo dei minimi quadrati (basata sul metodo dei moltiplicatori di Lagrange, e che non trattiamo qui) che consente di trovare il valore minimo della funzione (3.5) vincolato da una o più relazioni lineari assegnate tra i parametri;

3) un metodo alternativo per imporre un vincolo ai parametri consiste nell'introdurlo come una *pseudo-osservazione* aggiuntiva alle osservabili vere e proprie; nel nostro caso dovremmo includere un'osservazione della quantità $\xi_1 + \xi_2 + \xi_3 + \xi_4$ che abbia come risultato un valore scelto arbitrariamente (che corrisponde al quadruplo del valore medio dell'ascissa del baricentro dell'insieme di punti), ad esempio $\xi_1+\xi_2+\xi_3+\xi_4 = 0$; con queste modifiche la matrice disegno del problema diventa

$$A = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

e la matrice normale risultante

$$A'A = \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

[20] A posteriori è immediato verificare che

$$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{e} \quad \begin{pmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} .$$

[21] Naturalmente questa proprietà di invarianza avrebbe dovuto essere chiara fin dall'inizio, senza bisogno di calcolare il determinante della matrice normale; abbiamo scelto appositamente un caso molto semplice per illustrare l'utilizzo dell'informazione contenuta negli autovettori della matrice normale.

è ovviamente invertibile[22].

3.7.4 Riduzione astrometrica di un'immagine

Studiamo ora come si possono determinare le coordinate astrometriche (*ascensione retta* α e *declinazione* δ) di un corpo celeste che compare in una immagine presa al telescopio. Il primo passo consiste nell'eseguire la *riduzione astrometrica dell'immagine*, cioè trovare la corrispondenza tra le *coordinate di lastra*[23] (ξ, η) misurate sul piano dell'immagine (espresse in millimetri nel caso di una lastra fotografica o come coordinate di *pixel* nel caso di un'immagine CCD[24]) e le coordinate astrometriche; quindi, avendo misurato le coordinate di lastra dell'oggetto che interessa, si possono calcolare le sue coordinate astrometriche.

La riduzione astrometrica dell'immagine viene eseguita identificando su di essa un certo numero n di stelle fisse, misurando le loro coordinate di lastra (ξ_i, η_i) $(i = 1, \ldots, n)$ e confrontandole con le corrispondenti coordinate astrometriche (α_i, δ_i) ottenute da un catalogo stellare. L'immagine sul piano focale del telescopio è una proiezione della volta celeste che, nella nomenclatura della cartografia, viene chiamata *proiezione gnomonica*[25]; si ottiene geometricamente proiettando le posizioni delle stelle su un piano tangente alla sfera celeste nel punto corrispondente al centro del campo di vista del telescopio (α_0, δ_0), utilizzando come punto di proiezione il centro della sfera, e assegnando alla sfera un raggio pari alla lunghezza focale del telescopio. Analiticamente la proiezione può essere eseguita utilizzando il seguente algoritmo:

1) si trasformano le coordinate astrometriche (polari) della stella (α, δ) nelle coordinate cartesiane del corrispondente versore:

$$\begin{cases} x_1 &= \cos\delta\cos\alpha \\ x_2 &= \cos\delta\sin\alpha \\ x_3 &= \sin\delta\,; \end{cases}$$

2) si trasformano le coordinate cartesiane in un nuovo sistema di riferimento (x'_1, x'_2, x'_3) in cui l'asse (x'_1) punta nella direzione (α_0, δ_0) e l'asse (x'_3), ortogonale a (x'_1), giace nel piano (x'_1, x_3); ciò si ottiene applicando alle coordinate (x_1, x_2, x_3) le due matrici di rotazione

$$\begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = \mathbf{R}_2(-\delta_0)\,\mathbf{R}_3(\alpha_0) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\,;$$

3) a questo punto le coordinate (x'_2, x'_3) sono proporzionali alle coordinate di lastra (ξ, η) a meno di un fattore di proiezione, dato dal rapporto tra la lunghezza focale F del

[22] Notiamo che questo secondo metodo di imporre un vincolo al valore dei parametri, pur essendo di più immediata implementazione, è meno rigoroso del metodo dei moltiplicatori di Lagrange in quanto il vincolo, essendo introdotto come pseudo-osservazione, non è imposto esattamente ma con un certo errore di misura: la condizione di vincolo ha un residuo che concorre a formare la funzione da minimizzare (3.5) e che quindi in generale non sarà esattamente nullo. Questo effetto può essere reso trascurabile lavorando nell'ambito del metodo dei minimi quadrati pesati e assegnando alla pseudo-osservazione di vincolo una varianza a priori molto più piccola (cioè un peso molto più grande) di quella delle osservazioni vere e proprie.

[23] Per ragioni storiche si continua a usare il termine *coordinate di lastra* (*plate coordinates*) anche nel caso di immagini CCD.

[24] Il CCD (acronimo dall'inglese *charge coupled device*) è un circuito integrato formato da una matrice di elementi semiconduttori fotosensibili, in grado di trasformare la radiazione luminosa incidente in una carica elettrica che viene successivamente amplificata e digitalizzata per mezzo di appositi circuiti elettronici, producendo un'immagine in forma digitale.

[25] Questo perché si tratta dello stesso tipo di proiezione che regola la geometria della produzione dell'ombra dello gnomone di una meridiana solare.

telescopio e la coordinata x'_1, cioè[26]

$$\begin{pmatrix} \xi \\ \eta \end{pmatrix} = \frac{F}{x'_1} \begin{pmatrix} x'_2 \\ x'_3 \end{pmatrix} .$$

Partendo da un valore approssimato delle coordinate di centro lastra (α_0, δ_0) (fornite dagli *encoder* del telescopio o da un'analisi approssimativa dell'immagine) si può quindi calcolare un primo valore approssimato delle coordinate di lastra (ξ_i, η_i) delle stelle di riferimento, che chiameremo coordinate *nominali* di lastra (perché determinate in base a un valore nominale del puntamento del telescopio) e che indicheremo con la notazione $(\xi_i^{(0)}, \eta_i^{(0)})$. Tali coordinate nominali saranno differenti da quelle effettivamente misurate sull'immagine (ξ_i, η_i) non solo per l'imperfetta conoscenza di (α_0, δ_0), ma per una serie di altre ragioni: il valore della lunghezza focale del telescopio potrebbe non essere conosciuto con precisione sufficiente (e comunque è soggetto a variazioni al variare della temperatura e della messa a fuoco dello strumento); l'asse η del CCD potrebbe non essere allineato esattamente con l'asse polare; l'ottica del telescopio potrebbe non essere geometricamente perfetta e introdurre piccole deformazioni rispetto alla legge di proiezione gnomonica (ad esempio è da prevedersi una certa deformazione delle ottiche dovuta alle flessioni meccaniche e quindi variabile con l'inclinazione del telescopio e differente da un'immagine all'altra). È perciò abbastanza naturale cercare di descrivere la relazione tra coordinate nominali e coordinate misurate attraverso un'espressione del tipo

$$\begin{pmatrix} \xi_i - \xi_i^{(0)} \\ \eta_i - \eta_i^{(0)} \end{pmatrix} = \begin{pmatrix} \Delta\xi \\ \Delta\eta \end{pmatrix} + B \begin{pmatrix} \xi_i \\ \eta_i \end{pmatrix} , \tag{3.39}$$

dove il vettore $(\Delta\xi, \Delta\eta)$ rappresenta una traslazione dell'origine delle coordinate e la matrice generica B può modellare sia rotazioni di assi, sia variazioni del fattore di scala, sia eventualmente anche una non perfetta ortogonalità degli assi. In molti casi si aggiungono alla (3.39) anche termini quadratici nelle coordinate (ξ, η), che possono modellare effetti non lineari dovuti alle ottiche del telescopio o ad altre cause[27]. Oltre a ciò ci sono ovvia-

[26] Questa definizione degli assi (ξ, η), del tutto convenzionale, corrispondere a scegliere l'asse ξ in modo che sia orientato nel senso della crescita dell'ascensione retta e l'asse η nel senso della crescita della declinazione.

[27] Ad esempio la rifrazione atmosferica introduce nel valore osservato della distanza zenitale z di un astro una variazione pari a

$$R(z) = K \tan z \, , \tag{3.40}$$

dove K (costante di rifrazione) vale $K = 60''.4$ (alla pressione di 1 atmosfera e alla temperatura di 0 °C). Ciò che conta nella riduzione astrometrica di un'immagine è solo l'effetto differenziale della (3.40) per piccole variazioni Δz della distanza zenitale z attorno al valore $z = z_0$ del centro lastra, che può essere scritto come

$$R(z_0 + \Delta z) = K \tan(z_0 + \Delta z) = K \, \frac{\tan z_0 + \tan \Delta z}{1 - \tan z_0 \tan \Delta z}$$

e che, ponendo per brevità di notazione $\delta = \tan \Delta z$, può essere approssimato come

$$\begin{aligned} & R(z_0 + \Delta z) - R(z_0) = \\ & \quad = K \left[(1 + \tan^2 z_0)\, \delta + (\tan z_0 + \tan^3 z_0)\, \delta^2 + (\tan^2 z_0 + \tan^4 z_0)\, \delta^3 + 0(\delta^4) \right] . \end{aligned} \tag{3.41}$$

Per valutare gli ordini di grandezza dei diversi termini che compongono la (3.41) possiamo assumere come valore massimo per la distanza zenitale $z_0 \approx 75°$ (e quindi $\tan z_0 \approx 3.7$) e come un valore ragionevole per la dimensione dell'immagine $\Delta z \approx 0°.5$ (quindi $\delta \approx 0.01$), che corrisponde a un campo inquadrato di $1° \times 1°$, da cui si ricavano le seguenti stime

$$\begin{aligned} K\,(1 + \tan^2 z_0)\,\delta &\approx 7''.9 \\ K\,(\tan z_0 + \tan^3 z_0)\,\delta^2 &\approx 0''.26 \\ K\,(\tan^2 z_0 + \tan^4 z_0)\,\delta^3 &\approx 0''.008 \, ; \end{aligned}$$

si vede dunque che la deformazione dell'immagine introdotta dall'effetto differenziale della rifrazione può essere modellata adeguatamente da termini lineari e quadratici nelle coordinate di lastra, mentre i termini cubici sono trascurabili.

mente gli errori: errori di misura delle coordinate di lastra (ξ_i, η_i) ed errori nelle coordinate astrometriche di catalogo (α_i, δ_i) (perché anche le coordinate riportate dai cataloghi sono state ottenute da misure astrometriche). In definitiva si arriva a una rappresentazione della differenza tra coordinate misurate e coordinate nominali del tipo

$$\begin{aligned} \xi_i - \xi_i^{(0)} &= c_1 + c_2\xi_i + c_3\eta_i + c_4\xi_i^2 + c_5\eta_i^2 + c_6\xi_i\eta_i + \varepsilon_i^{(\xi)} \\ \eta_i - \eta_i^{(0)} &= c_7 + c_8\xi_i + c_9\eta_i + c_{10}\xi_i^2 + c_{11}\eta_i^2 + c_{12}\xi_i\eta_i + \varepsilon_i^{(\eta)} \,. \end{aligned} \tag{3.42}$$

I parametri c_i, noti con il nome di *costanti di lastra*, possono essere determinati con il metodo dei minimi quadrati. Infatti la (3.42) ha la forma tipica di un problema di stima lineare dei parametri (3.6) di dimensioni $N = 2n$ e $M = 12$, con $\beta_i = c_i$,

$$x = \begin{pmatrix} \xi_1 - \xi_1^{(0)} \\ \eta_1 - \eta_1^{(0)} \\ \xi_2 - \xi_2^{(0)} \\ \eta_2 - \eta_2^{(0)} \\ \vdots \\ \xi_n - \xi_n^{(0)} \\ \eta_n - \eta_n^{(0)} \end{pmatrix}, \qquad \varepsilon = \begin{pmatrix} \varepsilon_1^{(\xi)} \\ \varepsilon_1^{(\eta)} \\ \varepsilon_2^{(\xi)} \\ \varepsilon_2^{(\eta)} \\ \vdots \\ \varepsilon_n^{(\xi)} \\ \varepsilon_n^{(\eta)} \end{pmatrix},$$

$$A = \begin{pmatrix} 1 & \xi_1 & \eta_1 & \xi_1^2 & \eta_1^2 & \xi_1\eta_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \xi_1 & \eta_1 & \xi_1^2 & \eta_1^2 & \xi_1\eta_1 \\ 1 & \xi_2 & \eta_2 & \xi_2^2 & \eta_2^2 & \xi_2\eta_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \xi_2 & \eta_2 & \xi_2^2 & \eta_2^2 & \xi_2\eta_2 \\ \vdots & & & & & & & & & & & \vdots \\ 1 & \xi_n & \eta_n & \xi_n^2 & \eta_n^2 & \xi_n\eta_n & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \xi_n & \eta_n & \xi_n^2 & \eta_n^2 & \xi_n\eta_n \end{pmatrix}.$$

Se sull'immagine si riesce a identificare un numero sufficiente di stelle di riferimento[28], è allora possibile applicare la (3.11) e ricavare una stima delle costanti di lastra, che possono successivamente essere usate per determinare le coordinate astrometriche di qualsiasi oggetto che compare nell'immagine (ad esempio un asteroide). Infatti, partendo dalle sue coordinate di lastra misurate (ξ, η), utilizzando le (3.42) si possono ricavare le corrispondenti coordinate nominali

$$\begin{aligned} \xi^{(0)} &= \xi - c_1 - c_2\xi - c_3\eta - c_4\xi^2 - c_5\eta^2 - c_6\xi\eta \\ \eta^{(0)} &= \eta - c_7 - c_8\xi - c_9\eta - c_{10}\xi^2 - c_{11}\eta^2 - c_{12}\xi\eta \end{aligned}$$

e da queste le coordinate astrometriche

$$\begin{pmatrix} \cos\delta\cos\alpha \\ \cos\delta\sin\alpha \\ \sin\delta \end{pmatrix} = \frac{1}{\sqrt{F^2 + \xi^{(0)2} + \eta^{(0)2}}}\, \mathbf{R}_3(-\alpha_0)\, \mathbf{R}_2(\delta_0) \begin{pmatrix} F \\ \xi^{(0)} \\ \eta^{(0)} \end{pmatrix}$$

[28] In teoria, al fine di avere $N = 2n \geq M = 12$, sarebbero sufficienti $n = 6$ stelle; in pratica è consigliabile averne una quantità almeno doppia, in modo da poter avere una stima attendibile degli errori ed eliminare eventuali identificazioni errate o misure mal riuscite.

Capitolo 4

Correzione differenziale di un'orbita kepleriana

4.1 Definizione delle osservabili

Affrontiamo ora il problema della correzione differenziale dei parametri orbitali di un'orbita kepleriana; supponiamo cioè di avere una serie di osservazioni della posizione di un corpo celeste a tempi diversi e di voler determinare i parametri orbitali che meglio si adattano a esse, assumendo come modello dinamico un'orbita kepleriana (trascurando cioè ogni possibile perturbazione al problema dei due corpi). Questo è un tipico problema di minimi quadrati e, siccome le equazioni che danno la soluzione del problema dei due corpi sono non lineari, dobbiamo linearizzarle, riducendoci a un sistema di equazioni di osservazione del tipo (3.36). Ciò significa tra l'altro che il metodo che andiamo a descrivere presuppone la conoscenza di un valore approssimato degli elementi orbitali da cui far partire il procedimento iterativo di linearizzazione e correzione; tale valore di partenza può essere ottenuto utilizzando uno dei metodi di determinazione dell'orbita a partire da tre osservazioni, applicato a tre osservazioni opportunamente scelte tra quelle disponibili.

Riguardo alla natura delle osservabili, consideriamo due casi: 1) osservazione di un sistema stellare binario (visuale); 2) osservazione di un pianeta o asteroide del Sistema Solare, cioè in orbita attorno al Sole.

4.2 Sistema binario visuale

Nel caso di di un sistema stellare binario la riduzione della lastra o dell'immagine CCD o la misurazione micrometrica visuale forniscono direttamente (a meno di un fattore di scala introdotto dalla distanza) la proiezione della posizione relativa dei due astri su di un piano ortogonale alla linea di vista; assumendo per convenzione tale piano come piano (x_1, x_2), si può dire che l'osservazione fornisce direttamente le coordinate inerziali x_1 e x_2 (o più precisamente la differenza tra le coordinate inerziali dei due astri); la coordinata x_3 si traduce in una variazione della distanza tra la stella e l'osservatore e non è perciò direttamente misurabile. Ciò implica che l'orientazione dell'orbita nello spazio non è completamente determinabile: infatti se l'orbita viene trasformata nella sua immagine riflessa sul piano (x_1, x_2), la sua proiezione su tale piano non cambia. Tale riflessione corrisponde a uno scambio tra nodo ascendente e discendente, cioè a una trasformazione $\Omega \to \Omega + \pi$, $\omega \to \omega + \pi$, e quindi i parametri orbitali sono determinabili dalle osservazioni a meno di una

trasformazione simile[1]. È convenzione assumere l'asse x_1 orientato nella direzione del polo nord celeste e l'asse x_2 orientato nella direzione della ascensione retta crescente.

Poiché la distanza del sistema binario dalla Terra non è mai nota con precisione sufficiente, le osservazioni di x_1 e x_2 sono mantenute nella forma di angoli (in secondi d'arco); ciò significa che anche il valore del semiasse maggiore a risulterà espresso in secondi d'arco. Inoltre la costante di massa del sistema $\mu = G(m_0 + m_1)$ è sconosciuta, quindi anche il moto medio n (o il periodo orbitale T) deve essere determinato *indipendentemente dal semiasse maggiore*. Se d'altra parte è nota la distanza del sistema dalla Terra, dagli elementi orbitali ottenuti è possibile ottenere una stima della somma delle masse delle due stelle.

In conclusione, per poter applicare il metodo dei minimi quadrati a misure di posizione relativa di stelle binarie è necessario ricavare le derivate parziali

$$\frac{\partial x_1}{\partial \beta_k} \quad \text{e} \quad \frac{\partial x_2}{\partial \beta_k} \qquad (k = 1, 2, \ldots, 7)$$

delle coordinate x_1 e x_2 rispetto ai sei elementi kepleriani e al moto medio.

4.3 Pianeta del Sistema Solare

Nel caso di di un pianeta (o asteroide) del Sistema Solare le osservazioni astrometriche forniscono la posizione apparente del corpo sulla volta celeste, cioè la sua ascensione retta α e declinazione δ, che sono le coordinate polari angolari del vettore posizione relativa del pianeta rispetto all'osservatore $\mathbf{r} - \mathbf{r}_{oss}$

$$\begin{cases} x_1 - x_{1oss} = \rho \cos\delta \cos\alpha \\ x_2 - x_{2oss} = \rho \cos\delta \sin\alpha \\ x_3 - x_{3oss} = \rho \sin\delta \,. \end{cases} \tag{4.1}$$

Le derivate parziali degli osservabili α e δ rispetto agli elementi orbitali β_k sono gli elementi della matrice jacobiana $\partial(\rho, \alpha, \delta)/\partial(\beta_k)$, che può essere ottenuta come prodotto delle due matrici jacobiane

$$\frac{\partial(\rho, \alpha, \delta)}{\partial(\beta_k)} = \frac{\partial(\rho, \alpha, \delta)}{\partial(x_1, x_2, x_3)} \, \frac{\partial(x_1, x_2, x_3)}{\partial(\beta_k)} \,. \tag{4.2}$$

Derivando la (4.1) si ottiene la matrice

$$\frac{\partial(x_1, x_2, x_3)}{\partial(\rho, \alpha, \delta)} = \begin{pmatrix} \cos\alpha\cos\delta & -\rho\sin\alpha\cos\delta & -\rho\cos\alpha\sin\delta \\ \sin\alpha\cos\delta & \rho\cos\alpha\cos\delta & -\rho\sin\alpha\sin\delta \\ \sin\delta & 0 & \rho\cos\delta \end{pmatrix} ,$$

invertendo la quale si ottiene la prima matrice jacobiana che compare nella (4.2)

$$\frac{\partial(\rho, \alpha, \delta)}{\partial(x_1, x_2, x_3)} = \frac{1}{\rho} \begin{pmatrix} \rho\cos\alpha\cos\delta & \rho\sin\alpha\cos\delta & \rho\sin\delta \\ -\sin\alpha/\cos\delta & \cos\alpha/\cos\delta & 0 \\ -\cos\alpha\sin\delta & -\sin\alpha\sin\delta & \cos\delta \end{pmatrix} . \tag{4.3}$$

Rimangono quindi da calcolare le derivate parziali $\partial x_i/\partial\beta_k$ delle coordinate del pianeta rispetto agli elementi orbitali; dobbiamo però osservare che, a differenza del caso di un sistema stellare binario, per un pianeta del Sistema Solare di solito si può supporre di conoscere la costante di massa del sistema[2]. In questo caso quindi, nel calcolare le derivate parziali il moto medio non va considerato come una variabile indipendente ma come una funzione del semiasse maggiore.

[1] Questa proprietà si può dimostrare anche analiticamente: infatti la trasformazione $\Omega \to \Omega+\pi$, $\omega \to \omega+\pi$ lascia invariati gli elementi R_{11}, R_{12}, R_{21} ed R_{22} della matrice di rotazione $\mathbf{R}(\Omega, i, \omega)$ (equazioni (1.45) e (1.46)), che sono gli unici che influenzano gli osservabili x_1 e x_2.

[2] La massa del Sole m_S e dei pianeti maggiori del Sistema Solare m_p è conosciuta, cosicché $\mu = G(m_S + m_p)$ è noto; la massa degli asteroidi è di solito scarsamente conosciuta, ma è così piccola che si può porre con grande approssimazione $\mu = Gm_S$.

4.4 Derivate parziali del vettore posizione

Per calcolare le derivate del vettore posizione rispetto agli elementi orbitali usiamo la formulazione delle leggi del moto in variabili non singolari (1.57):

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{R}(P,Q) \begin{pmatrix} \xi' \\ \eta' \\ \zeta' \end{pmatrix} .$$

Notiamo subito che in questa espressione gli elementi P e Q compaiono solo nella matrice di rotazione $\mathbf{R}(P,Q)$, per cui

$$\frac{\partial}{\partial(P,Q)} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \frac{\partial \mathbf{R}(P,Q)}{\partial(P,Q)} \begin{pmatrix} \xi' \\ \eta' \\ \zeta' \end{pmatrix} , \tag{4.4}$$

mentre i rimanenti elementi orbitali β_k compaiono solo nell'espressione delle coordinate nel sistema di riferimento orbitale (ξ', η', ζ'), quindi

$$\frac{\partial}{\partial(\beta_k)} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{R}(P,Q) \frac{\partial}{\partial(\beta_k)} \begin{pmatrix} \xi' \\ \eta' \\ \zeta' \end{pmatrix} \quad (\text{per } \beta_k \neq P, Q) . \tag{4.5}$$

Le derivate parziali della matrice di rotazione $\mathbf{R}(P,Q)$ possono essere ottenute direttamente derivando la (1.59)

$$\frac{\partial \mathbf{R}(P,Q)}{\partial P} = \frac{1}{(1+P^2+Q^2)^2} \begin{pmatrix} -4P(1+Q^2) & 2Q(1-P^2+Q^2) & 2(1-P^2+Q^2) \\ 2Q(1-P^2+Q^2) & 4PQ^2 & 4PQ \\ -2(1-P^2+Q^2) & -4PQ & -4P \end{pmatrix}$$

$$\frac{\partial \mathbf{R}(P,Q)}{\partial Q} = \frac{1}{(1+P^2+Q^2)^2} \begin{pmatrix} 4P^2Q & 2P(1+P^2-Q^2) & -4PQ \\ 2P(1+P^2-Q^2) & -4Q(1+P^2) & -2(1+P^2-Q^2) \\ 4PQ & 2(1+P^2-Q^2) & -4Q \end{pmatrix} \tag{4.6}$$

4.5 Derivate parziali nel sistema di riferimento orbitale

Le derivate parziali di ξ' ed η' rispetto ad h e k si ottengono direttamente dalle espressioni (1.56), tenendo conto che queste dipendono da h e k sia esplicitamente, sia attraverso γ (che è funzione di $e^2 = h^2 + k^2$), sia attraverso la longitudine eccentrica F (che è funzione della longitudine media L e di h e k attraverso l'equazione di Keplero non singolare (1.54)):

$$\begin{aligned}
\frac{\partial \xi'}{\partial h} &= a\gamma\big[-2h\cos F + k\sin F\big] + \frac{\partial \xi'}{\partial \gamma}\frac{\partial \gamma}{\partial h} + \frac{\partial \xi'}{\partial F}\frac{\partial F}{\partial h} \\
\frac{\partial \xi'}{\partial k} &= a\big[\gamma h\sin F - 1\big] + \frac{\partial \xi'}{\partial \gamma}\frac{\partial \gamma}{\partial k} + \frac{\partial \xi'}{\partial F}\frac{\partial F}{\partial k} \\
\frac{\partial \eta'}{\partial h} &= a\big[\gamma k\cos F - 1\big] + \frac{\partial \eta'}{\partial \gamma}\frac{\partial \gamma}{\partial h} + \frac{\partial \eta'}{\partial F}\frac{\partial F}{\partial h} \\
\frac{\partial \eta'}{\partial k} &= a\gamma\big[-2k\sin F + h\cos F\big] + \frac{\partial \eta'}{\partial \gamma}\frac{\partial \gamma}{\partial k} + \frac{\partial \eta'}{\partial F}\frac{\partial F}{\partial k} ,
\end{aligned} \tag{4.7}$$

dove

$$\begin{aligned} \frac{\partial \xi'}{\partial \gamma} &= ah\big[-h\cos F + k\sin F\big] \\ \frac{\partial \eta'}{\partial \gamma} &= ak\big[-k\sin F + h\cos F\big] \end{aligned} \tag{4.8}$$

e

$$\begin{aligned} \frac{\partial \xi'}{\partial F} &= a\big[-(1-\gamma h^2)\sin F + \gamma hk\cos F\big] \\ \frac{\partial \eta'}{\partial F} &= a\big[(1-\gamma k^2)\cos F - \gamma hk\sin F\big] \end{aligned} \tag{4.9}$$

e inoltre

$$\frac{\partial \gamma}{\partial h} = \frac{h\gamma^2}{\sqrt{1-e^2}}\,, \qquad \frac{\partial \gamma}{\partial k} = \frac{k\gamma^2}{\sqrt{1-e^2}}\,. \tag{4.10}$$

Le derivate della longitudine eccentrica F si ottengono differenziando l'equazione di Keplero (1.54)

$$dL = (1 - h\sin F - k\cos F)\,dF + \cos F\,dh - \sin F\,dk\,, \tag{4.11}$$

da cui

$$\frac{\partial F}{\partial h} = -\frac{\cos F}{1 - h\sin F - k\cos F}\,, \qquad \frac{\partial F}{\partial k} = \frac{\sin F}{1 - h\sin F - k\cos F}\,. \tag{4.12}$$

Le derivate rispetto al semiasse maggiore a hanno forma diversa a seconda che il moto medio n sia considerato una variabile indipendente (da determinare separatamente) oppure una funzione del semiasse maggiore stesso attraverso la terza legge di Keplero (1.24); nel primo caso è sufficiente considerare la dipendenza esplicita delle (1.56) da a

$$\begin{aligned} \frac{\partial \xi'}{\partial a} &= (1-\gamma h^2)\cos F + \gamma hk\sin F - k \\ \frac{\partial \eta'}{\partial a} &= (1-\gamma k^2)\sin F + \gamma hk\cos F - h\,; \end{aligned} \tag{4.13}$$

nel secondo caso è necessario tenere conto anche della dipendenza implicita attraverso la longitudine eccentrica F che è funzione della longitudine media L, a sua volta funzione del moto medio

$$L = L_0 + n(t - t_0)\,. \tag{4.14}$$

In tal caso si ha quindi

$$\begin{aligned} \frac{\partial \xi'}{\partial a} &= (1-\gamma h^2)\cos F + \gamma hk\sin F - k + \frac{\partial \xi'}{\partial F}\frac{\partial F}{\partial L}\frac{\partial L}{\partial n}\frac{\partial n}{\partial a} \\ \frac{\partial \eta'}{\partial a} &= (1-\gamma k^2)\sin F + \gamma hk\cos F - h + \frac{\partial \eta'}{\partial F}\frac{\partial F}{\partial L}\frac{\partial L}{\partial n}\frac{\partial n}{\partial a}\,; \end{aligned} \tag{4.15}$$

dalla (4.11) si ricava

$$\frac{\partial F}{\partial L} = \frac{1}{1 - h\sin F - k\cos F}\,; \tag{4.16}$$

dalla (4.14)

$$\frac{\partial L}{\partial n} = t - t_0\,, \tag{4.17}$$

mentre differenziando la (1.24) si ha

$$\frac{\partial n}{\partial a} = -\frac{3a^2}{2n}\,. \tag{4.18}$$

Nel caso che il moto medio sia un parametro indipendente da determinare occorrono ancora le derivate parziali

$$\frac{\partial \xi'}{\partial n} = \frac{\partial \xi'}{\partial F}\frac{\partial F}{\partial L}\frac{\partial L}{\partial n}\,, \qquad \frac{\partial \eta'}{\partial n} = \frac{\partial \eta'}{\partial F}\frac{\partial F}{\partial L}\frac{\partial L}{\partial n}\,. \tag{4.19}$$

Infine, le derivate rispetto al valore iniziale della longitudine media L_0 sono

$$\frac{\partial \xi'}{\partial L_0} = \frac{\partial \xi'}{\partial F}\frac{\partial F}{\partial L_0} \qquad \frac{\partial \eta'}{\partial L_0} = \frac{\partial \eta'}{\partial F}\frac{\partial F}{\partial L_0}\,, \tag{4.20}$$

dove

$$\frac{\partial F}{\partial L_0} = \frac{\partial F}{\partial L} = \frac{1}{1 - h\sin F - k\cos F}\,. \tag{4.21}$$

Appendice A

Formule di trasformazione tra elementi kepleriani e vettori posizione e velocità

A.1 Passaggio da elementi kepleriani a coordinate cartesiane

Riassumiamo qui i passaggi che permettono di ottenere le coordinate cartesiane $\mathbf{r}$, $\mathbf{v}$ a partire dai sei elementi kepleriani a, e, i, ω, Ω ed M (limitatamente al caso di un'orbita ellittica, cioè per $e < 1$):

1) si ricava l'anomalia eccentrica E risolvendo l'equazione di Keplero:

$$M = E - e \sin E \,; \tag{A.1}$$

2) si calcolano le componenti del vettore posizione $\mathbf{r}'$ nel sistema di riferimento orbitale (asse ξ diretto lungo la linea degli apsidi)

$$\begin{aligned} r'_1 &= \xi &&= a(\cos E - e) \\ r'_2 &= \eta &&= a\sqrt{1-e^2}\,\sin E \\ r'_3 &= \zeta &&= 0\,; \end{aligned} \tag{A.2}$$

le formule per le componenti della velocità $\mathbf{v}'$ nello stesso sistema di riferimento si ottengono derivando le (A.2) rispetto al tempo

$$\begin{aligned} v'_1 &= \frac{d\xi}{dt} &&= -a \sin E \,\frac{dE}{dt} \\ v'_2 &= \frac{d\eta}{dt} &&= a\sqrt{1-e^2}\cos E \,\frac{dE}{dt} \\ v'_3 &= \frac{d\zeta}{dt} &&= 0\,; \end{aligned}$$

la derivata dell'anomalia eccentrica E si ottiene derivando la (A.1)

$$\frac{dE}{dt} = \frac{n}{1 - e\cos E}\,,$$

dove

$$n = \frac{dM}{dt} = \sqrt{\frac{\mu}{a^3}}\,, \tag{A.3}$$

da cui

$$\begin{aligned} v'_1 &= \frac{d\xi}{dt} = \frac{-na\sin E}{1-e\cos E} \\ v'_2 &= \frac{d\eta}{dt} = \frac{na\sqrt{1-e^2}\cos E}{1-e\cos E} \\ v'_3 &= \frac{d\zeta}{dt} = 0\,; \end{aligned} \tag{A.4}$$

3) si calcola la matrice di rotazione $\mathbf{R}(\Omega, i, \omega)$ che trasforma il sistema di riferimento orbitale nel sistema di riferimento inerziale; le sue componenti sono date da

$$\begin{aligned} R_{11} &= \cos\omega\cos\Omega - \sin\omega\sin\Omega\cos i \\ R_{12} &= -\sin\omega\cos\Omega - \cos\omega\sin\Omega\cos i \\ R_{13} &= \sin\Omega\sin i \\ R_{21} &= \cos\omega\sin\Omega + \sin\omega\cos\Omega\cos i \\ R_{22} &= -\sin\omega\sin\Omega + \cos\omega\cos\Omega\cos i \\ R_{23} &= -\cos\Omega\sin i \\ R_{31} &= \sin\omega\sin i \\ R_{32} &= \cos\omega\sin i \\ R_{33} &= \cos i\,; \end{aligned} \tag{A.5}$$

4) si ruotano i vettori $\mathbf{r}'$ e $\mathbf{v}'$ nel sistema di riferimento inerziale

$$\mathbf{r} = \mathbf{R}(\Omega, i, \omega)\,\mathbf{r}'\,; \qquad \mathbf{v} = \mathbf{R}(\Omega, i, \omega)\,\mathbf{v}'\,. \tag{A.6}$$

A.2 Passaggio da coordinate cartesiane a elementi kepleriani

I passaggi per operare la trasformazione inversa (da coordinate cartesiane $\mathbf{r}$, $\mathbf{v}$ a elementi kepleriani a, e, i, ω, Ω ed M, sempre limitandosi al caso di un'orbita ellittica) sono:

1) si calcola il vettore momento angolare $\mathbf{h}$

$$\mathbf{h} = \mathbf{r}\times\mathbf{v}\,, \tag{A.7}$$

le cui componenti nel sistema di riferimento inerziale sono

$$\begin{aligned} h_1 &= h\sin\Omega\sin i \\ h_2 &= -h\cos\Omega\sin i \\ h_1 &= h\cos i\,; \end{aligned}$$

l'inclinazione orbitale i e la longitudine del nodo Ω possono quindi essere calcolate dalle seguenti formule

$$\cos i = \frac{h_3}{h}\,, \qquad \sin i = \frac{\sqrt{h_1^2+h_2^2}}{h} \tag{A.8}$$

e

$$\cos\Omega = \frac{-h_2}{\sqrt{h_1^2+h_2^2}}\,, \qquad \sin\Omega = \frac{h_1}{\sqrt{h_1^2+h_2^2}}\,; \tag{A.9}$$

2) si calcola il semiasse maggiore a dalla formula

$$\frac{1}{a} = \frac{2}{r} - \frac{v^2}{\mu}\,; \tag{A.10}$$

3) si calcolano il vettore di Laplace-Lenz

$$\mathbf{e} = \frac{\mathbf{v} \times \mathbf{h}}{\mu} - \frac{\mathbf{r}}{r} \tag{A.11}$$

(il cui modulo è uguale all'eccentricità orbitale e) e il versore della linea dei nodi

$$\mathbf{k} = \frac{\mathbf{x}_3 \times \mathbf{h}}{h} = \begin{pmatrix} \cos\Omega \\ \sin\Omega \\ 0 \end{pmatrix} \tag{A.12}$$

(dove $\mathbf{x}_3$ è il versore dell'asse x_3); l'argomento del perielio ω è l'angolo compreso tra i vettori $\mathbf{k}$ ed $\mathbf{e}$, quindi il suo coseno è uguale al prodotto scalare dei due versori corrispondenti, mentre il suo seno è uguale alla proiezione del loro prodotto vettoriale lungo la direzione del momento angolare

$$\cos\omega = \frac{\mathbf{k} \cdot \mathbf{e}}{e}\,, \qquad \sin\omega = \left(\frac{\mathbf{k} \times \mathbf{e}}{e}\right) \cdot \left(\frac{\mathbf{h}}{h}\right)\,; \tag{A.13}$$

4) analogamente, l'anomalia vera f è l'angolo compreso tra la linea degli apsidi $\mathbf{e}$ e il vettore posizione $\mathbf{r}$, quindi

$$\cos f = \frac{\mathbf{e} \cdot \mathbf{r}}{er}\,, \qquad \sin f = \left(\frac{\mathbf{e} \times \mathbf{r}}{er}\right) \cdot \left(\frac{\mathbf{h}}{h}\right)\,; \tag{A.14}$$

dall'anomalia vera f si ricava l'anomalia eccentrica E

$$\tan\left(\frac{E}{2}\right) = \sqrt{\frac{1-e}{1+e}}\,\tan\left(\frac{f}{2}\right) \tag{A.15}$$

e da questa l'anomalia media M, utilizzando l'equazione di Keplero (A.1).

Appendice B

Forma chiusa delle serie f e g

Riportiamo qui la derivazione delle espressioni in forma chiusa per le serie f e g che sono necessarie nel procedimento di correzione della soluzione preliminare del metodo di Gauss (paragrafo 2.4). Riscriviamo la definizione delle serie f e g (1.49)

$$\mathbf{r} = f\,\mathbf{r}_0 + g\,\dot{\mathbf{r}}_0\,, \tag{B.1}$$

dove $\mathbf{r} = \mathbf{r}(t)$, $\mathbf{r}_0 = \mathbf{r}(0)$ e $\dot{\mathbf{r}}_0 = \dot{\mathbf{r}}(0)$. Eseguendo il prodotto vettoriale della (B.1) con $\dot{\mathbf{r}}_0$ e $\mathbf{r}_0$ si ottiene

$$\begin{array}{rcccc} \mathbf{r}\times\dot{\mathbf{r}}_0 & = & f\mathbf{r}_0\times\dot{\mathbf{r}}_0 & = & f\mathbf{h} \\ \mathbf{r}\times\mathbf{r}_0 & = & g\dot{\mathbf{r}}_0\times\mathbf{r}_0 & = & -g\mathbf{h} \end{array}$$

(dove $\mathbf{h} = \mathbf{r}_0\times\dot{\mathbf{r}}_0 = \mathbf{r}(t)\times\dot{\mathbf{r}}(t)$ è il vettore momento angolare, costante del moto nel problema dei due corpi), cioè

$$f = \frac{\mathbf{r}\times\dot{\mathbf{r}}_0}{\mathbf{h}}\,, \qquad g = \frac{-\mathbf{r}\times\mathbf{r}_0}{\mathbf{h}}\,. \tag{B.2}$$

I prodotti vettoriali che compaiono nelle (B.2) possono essere calcolati facilmente nel sistema di riferimento orbitale (asse ξ diretto lungo la linea degli apsidi, asse η nel piano orbitale ed ortogonale all'asse ξ, asse ζ parallelo al momento angolare); in tale riferimento le (B.2) diventano

$$f = \frac{\xi\dot{\eta}_0 - \eta\dot{\xi}_0}{h}\,, \qquad g = \frac{\eta\xi_0 - \xi\eta_0}{h}\,. \tag{B.3}$$

Introducendo nelle (B.3) le espressioni esplicite delle componenti dei vettori posizione e velocità (equazioni (A.2) e (A.4)) si ottiene

$$\begin{aligned} f &= \frac{a(\cos E - e)\,na\sqrt{1-e^2}\cos E_0 + a\sqrt{1-e^2}\sin E\,na\sin E_0}{h(1-e\cos E_0)} \\ &= \frac{na^2\sqrt{1-e^2}}{h}\left[\frac{\cos(E-E_0) - e\cos E_0}{1-e\cos E_0}\right] \end{aligned} \tag{B.4}$$

e

$$\begin{aligned} g &= \frac{a\sqrt{1-e^2}\sin E\,a(\cos E_0 - e) - a(\cos E - e)\,a\sqrt{1-e^2}\sin E_0}{h} \\ &= \frac{a^2\sqrt{1-e^2}}{h}\big[\sin(E-E_0) + e\sin E_0 - e\sin E\big]\,. \end{aligned} \tag{B.5}$$

La (B.4) può essere semplificata introducendo l'espressione di h che si può ricavare dalle (1.10), (1.15) e (1.24)

$$h = \sqrt{\mu p} = \sqrt{\mu a}\sqrt{1-e^2} = na^2\sqrt{1-e^2}\,, \tag{B.6}$$

ottenendo

$$f = \frac{\cos(E - E_0) - e\cos E_0}{1 - e\cos E_0} = 1 - \left[\frac{1 - \cos(E - E_0)}{1 - e\cos E_0}\right] ; \tag{B.7}$$

nella (B.5), oltre alla (B.6), si può sostituire

$$e\sin E = E - M = E - nt$$

(e un'espressione analoga per $e\sin E_0$), conseguenza dell'equazione di Keplero (1.31), ottenendo

$$g = t - \frac{1}{n}\Big[(E - E_0) - \sin(E - E_0)\Big] . \tag{B.8}$$