



Università degli Studi di Padova

DIPARTIMENTO DI FISICA E ASTRONOMIA GALILEO GALILEI

Corso di Laurea Triennale in Astronomia

**Soluzioni numeriche dell'equazione di Kepler
nel problema dei due corpi**

Relatore:

Prof. Marco Favretti

Candidato:

Alessio Gabriele Soggiu

Matricola 1102770

Indice

1	Moto sotto l'azione della forza gravitazionale	1
1.1	Forze centrali e forza gravitazionale	1
2	Forma e descrizione delle orbite	5
2.1	Forma delle orbite	5
2.2	Orientazione nello spazio	7
3	L'equazione di Kepler	11
3.1	Orbite chiuse	13
3.2	Orbite aperte	16
3.2.1	Caso particolare: orbite paraboliche	17
4	Metodi risolutivi per l'equazione di Kepler	19
4.1	Metodo della falsa posizione (regula falsi)	21
4.2	Metodo delle sostituzioni successive	22
4.3	Metodo di Newton-Raphson	25
5	Implementazione dei metodi	29
5.1	Implementazione <i>regula falsi</i>	30
5.2	Implementazione sostituzioni successive	32
5.3	Implementazione Newton-Raphson	34
5.4	Confronto dei metodi	36
A	Codici Matlab	37
A.1	Codice <i>regula falsi</i>	37
A.2	Codice sostituzioni successive	39
A.3	Codice Newton-Raphson	40

Sommario

A partire dalla definizione di forza centrale e dalla Legge di Gravitazione Universale di Newton, ricaviamo l'equazione delle coniche in coordinate polari, definendo così le possibili orbite di un corpo soggetto all'azione della forza gravitazionale; inoltre definiamo l'orientazione dell'orbita nello spazio tramite gli angoli di Eulero.

Successivamente, con l'intento di ricavare la posizione di un corpo in moto sotto l'azione della forza gravitazionale in funzione del tempo, definiamo l'anomalia eccentrica E da una relazione con l'anomalia vera v ; partendo dall'equazione delle coniche ricaviamo l'equazione di Kepler per le orbite chiuse $E - e \sin E = M$, in cui M (anomalia media) è una funzione del tempo. Mostriamo inoltre come cambia l'equazione nel caso di orbite aperte ($\sinh H - H = M_{hyp}$) e, in particolare, nel caso di orbite paraboliche ($e = 1$), per le quali è possibile ricavarne una soluzione analitica.

Al fine di risolvere numericamente l'equazione trascendente $E - e \sin E = M$, introduciamo il metodo della falsa posizione, il metodo delle sostituzioni successive ed il metodo di Newton-Raphson. Infine implementiamo i metodi e facciamo un confronto tra essi in termini di costo computazionale e robustezza, dimostrando la maggior efficienza del metodo di Newton rispetto agli altri due.

Capitolo 1

Moto sotto l'azione della forza gravitazionale

1.1 Forze centrali e forza gravitazionale

La forza gravitazionale è una delle quattro interazioni fondamentali della fisica. Si tratta di una forza *centrale*, ovvero quel tipo di forza che soddisfa le seguenti proprietà:

- è sempre diretta verso un punto O detto *centro della forza*;
- il modulo della forza dipende dalla sola distanza r dal punto O , per cui $F = F(r)$;
- detto \mathbf{r} il raggio vettore che unisce O al corpo su cui agisce la forza, vale sempre la relazione

$$\mathbf{r}(t) \times \ddot{\mathbf{r}}(t) = 0 . \quad (1.1)$$

L'interazione gravitazionale si verifica tra due o più corpi dotati di massa; è una forza di tipo attrattivo, direttamente proporzionale alle masse in gioco e inversamente proporzionale al quadrato della distanza tra esse.

Consideriamo due corpi di massa m_1 e m_2 individuati dai vettori \mathbf{r}_1 e \mathbf{r}_2 in un sistema di riferimento inerziale con origine O . Definiamo il vettore $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$ che individua la posizione del corpo di massa m_2 rispetto al corpo di massa m_1 . In meccanica classica vale la seconda legge della dinamica di Newton $\mathbf{F} = m\mathbf{a} = m\ddot{\mathbf{r}}$, che mette in relazione la forza agente su un corpo con la massa del punto e l'accelerazione conferita ad esso, dovuta all'azione della forza stessa. Nel caso della forza gravitazionale, l'espressione della forza

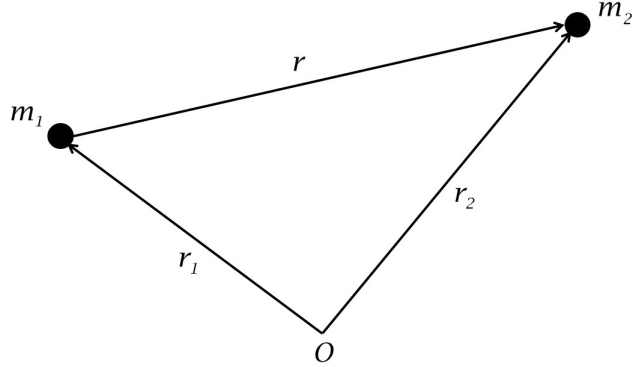


Figura 1.1: Rappresentazione del moto a due corpi. A partire dal centro O del sistema di riferimento, il corpo di massa m_1 è individuato dal vettore r_1 , mentre il corpo di massa m_2 è individuato dal vettore r_2 ; quest'ultimo ha posizione relativa al corpo di massa m_1 definita dal vettore $r = r_2 - r_1$

è data dalla legge di gravitazione universale di Newton, per cui le equazioni del moto dei due corpi rispetto all'origine O saranno

$$m_1 \ddot{\mathbf{r}}_1 = G \frac{m_1 m_2}{r^3} \mathbf{r} ; \quad (1.2)$$

$$m_2 \ddot{\mathbf{r}}_2 = -G \frac{m_1 m_2}{r^3} \mathbf{r} ; \quad (1.3)$$

in cui $G = 6.67384 \times 10^{-11} \frac{N \cdot m^2}{kg^2}$ è la costante di gravitazione universale. Sommando membro a membro otteniamo la relazione

$$m_1 \ddot{\mathbf{r}}_1 + m_2 \ddot{\mathbf{r}}_2 = 0 . \quad (1.4)$$

Integrando (1.4) due volte rispetto al tempo otteniamo

$$m_1 \dot{\mathbf{r}}_1 + m_2 \dot{\mathbf{r}}_2 = \mathbf{A} \quad (1.5)$$

$$m_1 \mathbf{r}_1 + m_2 \mathbf{r}_2 = \mathbf{A}t + \mathbf{B} \quad (1.6)$$

in cui \mathbf{A} e \mathbf{B} sono costanti di integrazione. Ricordando la definizione di centro di massa $\mathbf{r}_{\text{CM}} = \frac{m_1 \mathbf{r}_1 + m_2 \mathbf{r}_2}{m_1 + m_2}$, è facile notare che questo, per quanto detto prima, si muova di moto rettilineo uniforme. Infatti

$$\mathbf{r}_{\text{CM}} = \frac{\mathbf{A}t + \mathbf{B}}{m_1 + m_2} , \quad \dot{\mathbf{r}}_{\text{CM}} = \frac{\mathbf{A}}{m_1 + m_2} . \quad (1.7)$$

Queste due equazioni sono integrali del moto, quantità che rimangono costanti al variare del tempo e possono essere utilizzate per ridurre l'ordine delle

equazioni differenziali (1.2) e (1.3). A tal fine, poniamo l'origine del sistema di riferimento in corrispondenza del centro di massa del sistema, ovvero scegliamo $\mathbf{A} = 0$ e $\mathbf{B} = 0$. Quindi, da (1.6), $m_1\mathbf{r}_1 + m_2\mathbf{r}_2 = 0$. Ricordando che $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$ possiamo ricavare

$$\mathbf{r}_1 = \frac{m_2}{M}\mathbf{r}, \quad \mathbf{r}_2 = -\frac{m_1}{M}\mathbf{r} \quad (1.8)$$

dove $M = m_1 + m_2$.

Ora, semplifichiamo i termini di massa nelle equazioni (1.2) e (1.3)

$$\ddot{\mathbf{r}}_1 = G\frac{m_2}{r^3}\mathbf{r}, \quad \ddot{\mathbf{r}}_2 = -G\frac{m_1}{r^3}\mathbf{r}$$

e, sottraendo membro a membro, troviamo

$$\ddot{\mathbf{r}} + G(m_1 + m_2)\frac{\mathbf{r}}{r^3} = 0. \quad (1.9)$$

Applichiamo questa relazione ad entrambi i corpi singolarmente; ricordando (1.8) ricaviamo

$$\ddot{\mathbf{r}}_1 + G\frac{m_2^3}{M^2}\frac{\mathbf{r}_1}{r_1^3} = 0, \quad \ddot{\mathbf{r}}_2 + G\frac{m_1^3}{M^2}\frac{\mathbf{r}_2}{r_2^3} = 0. \quad (1.10)$$

A questo punto risulta evidente che, per qualsiasi caso considerato, le equazioni del moto sono accomunate dalla stessa forma

$$\ddot{\mathbf{r}} + \kappa^2\frac{\mathbf{r}}{r^3} = 0 \quad (1.11)$$

dove $\kappa > 0$ è una costante dipendente dalle masse in gioco.

La (1.11) rappresenta quindi un'equazione generale del moto sotto l'azione della forza gravitazionale. Operando il prodotto vettoriale tra \mathbf{r} e questa, possiamo ricavare direttamente la (1.1), $\mathbf{r} \times \ddot{\mathbf{r}} = 0$, che rappresenta, come detto in precedenza, la condizione caratteristica del moto sotto l'azione di forze centrali. La (1.1) può essere integrata

$$\mathbf{r} \times \dot{\mathbf{r}} = \mathbf{c} \quad (1.12)$$

con \mathbf{c} costante: questa equazione è detta *integrale del momento angolare* e mostra che il moto si svolge sul piano ortogonale al vettore \mathbf{c} ; la costanza di \mathbf{c} assicura che l'orientazione del piano non cambi. La (1.12) ci dà anche un'altra informazione: la velocità areolare, ovvero l'area spazzata dal raggio vettore \mathbf{r} nell'unità di tempo, è costante. Infatti, detta \mathbf{r} la posizione del

corpo in moto al tempo t e $\mathbf{r} + \dot{\mathbf{r}} dt$ quella al tempo $t + dt$, l'area infinitesima compresa tra i due vettori è data da

$$dA = \frac{1}{2} |\mathbf{r} \times (\mathbf{r} + \dot{\mathbf{r}} dt)| = \frac{1}{2} |\mathbf{r} \times \dot{\mathbf{r}}| dt = \frac{1}{2} c dt$$

per cui

$$\frac{dA}{dt} = \frac{c}{2} = \text{cost} . \quad (1.13)$$

Il fatto che il moto si svolga su un piano ci permette di semplificare il nostro problema: possiamo passare dalle generiche coordinate (x, y, z) a (ξ, η) in modo tale che la terza coordinata risulti sempre nulla. Ora possiamo riscrivere nelle nuove coordinate l'integrale del momento angolare (1.12)

$$\xi \dot{\eta} - \dot{\xi} \eta = c \quad (1.14)$$

e le componenti dell'equazione del moto (1.11)

$$\ddot{\xi} + \kappa^2 \frac{\xi}{r^3} = 0 , \quad \ddot{\eta} + \kappa^2 \frac{\eta}{r^3} = 0 \quad (1.15)$$

con $r = \sqrt{\xi^2 + \eta^2}$.

Ora, delle (1.15), moltiplichiamo la prima per $2\dot{\xi}$ e la seconda per $2\dot{\eta}$ e successivamente le sommiamo, ricavando

$$2\dot{\xi}\ddot{\xi} + 2\dot{\eta}\ddot{\eta} = -\kappa^2 \frac{2\dot{\xi}\xi + 2\dot{\eta}\eta}{r^3} .$$

Integrando una volta rispetto al tempo ricaviamo

$$\dot{\xi}^2 + \dot{\eta}^2 = \frac{2\kappa^2}{r} + h ; \quad (1.16)$$

la parte sinistra dell'equazione rappresenta due volte l'energia cinetica del corpo in moto per unità di massa, mentre la parte destra rappresenta meno due volte l'energia potenziale del corpo più una costante h detta *integrale dell'energia*. A questo punto è conveniente passare in coordinate polari tramite il cambio di variabili

$$\begin{cases} \xi = r \cos u \\ \eta = r \sin u \end{cases} \Rightarrow \begin{cases} \dot{\xi} = \dot{r} \cos u - r \sin u \dot{u} \\ \dot{\eta} = \dot{r} \sin u + r \cos u \dot{u} . \end{cases}$$

In questo modo l'equazione (1.14) diventa

$$r^2 \dot{u} = c ; \quad (1.17)$$

l'equazione integrale dell'energia (1.16) invece diventa

$$\dot{r}^2 + r^2 \dot{u}^2 = \frac{2\kappa^2}{r} + h . \quad (1.18)$$

Capitolo 2

Forma e descrizione delle orbite

2.1 Forma delle orbite

La traiettoria di un oggetto in moto sotto l'azione della forza gravitazionale ha la forma di una conica, ovvero una curva data dall'intersezione tra la superficie di un cono ed un piano; a seconda dell'inclinazione del piano rispetto al cono si ottiene una circonferenza, un'ellisse, una parabola o un'iperbole, come illustrato in Figura 2.1.

Per verificarlo analiticamente, rendiamo le equazioni ricavate in precedenza indipendenti dalla variabile t . Da (1.17) sappiamo che $\dot{u} = \frac{c}{r^2}$, per cui $\frac{dr}{dt} = \frac{dr}{du} \frac{du}{dt} = \frac{dr}{du} \frac{c}{r^2}$. Sostituendo questo risultato nell'equazione (1.18), troviamo

$$\left(-\frac{c}{r^2}\right)^2 \left(\frac{dr}{du}\right)^2 = \frac{2\kappa^2}{r} + h - \frac{c^2}{r^2}. \quad (2.1)$$

Ora poniamo $\sigma = \frac{c}{r} - \frac{\kappa^2}{c}$ e $A = \left(h + \frac{\kappa^4}{c^2}\right)^{1/2}$ ¹. Dal momento che $d\sigma/dr = -c/r^2$, possiamo ricavare l'equazione differenziale per le orbite

$$\left(\frac{d\sigma}{du}\right)^2 = A^2 - \sigma^2. \quad (2.2)$$

Operando la radice quadrata troviamo $d\sigma/du = \pm\sqrt{A^2 - \sigma^2}$; scegliamo il segno $-$. La soluzione dell'equazione quindi sar\`a

$$u = \arccos \frac{\sigma}{A} + \omega \quad \Rightarrow \quad \sigma = A \cos(u - \omega) \quad (2.3)$$

¹Sicuramente A \`e ben definito poich\`e la (2.1) pu\`o essere riscritta come $\left(\frac{d\sigma}{du}\right)^2 = h + \frac{\kappa^4}{c^2} - \sigma^2 \geq 0$, per cui $h + \frac{\kappa^4}{c^2} = \sigma^2 \geq 0$.

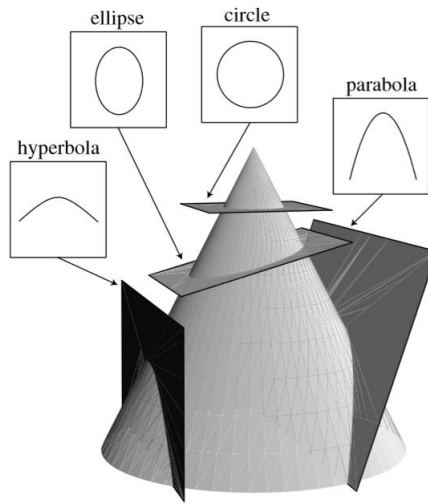


Figura 2.1: *Illustrazione di piani che intersecano una superficie conica con quattro differenti inclinazioni, dando origine alla nota famiglia di curve delle coniche.*

dove ω è una costante di integrazione. Se sostituiamo ora nuovamente $\sigma = c/r - \kappa^2/c$ e $A = (h + \kappa^4/c^2)^{1/2}$, possiamo esplicitare

$$r = \frac{c}{\frac{\kappa^2}{c} + \left(h + \frac{\kappa^4}{c^2}\right)^{1/2} \cos(u - \omega)}$$

Infine, raccogliendo un fattore κ^2/c al denominatore e definendo i parametri $p = \frac{\kappa^2}{c^2}$ ed $e = \left(1 + h \frac{c^2}{\kappa^4}\right)^{1/2}$ ricaviamo

$$r = \frac{p}{1 + e \cos(u - \omega)} \quad (2.4)$$

che rappresenta proprio l'equazione di una conica in coordinate polari. Il parametro p è detto *semilato retto* mentre e rappresenta l'*eccentricità*. A seconda dei valori assunti da e avremo una forma diversa dell'orbita:

- se $e = 0$ l'orbita è *circolare*;
- se $0 < e < 1$ l'orbita è *ellittica*;
- se $e = 1$ l'orbita è *parabolica*;
- se $e > 1$ l'orbita è *iperbolica*.

L'origine del sistema di riferimento corrisponderà al centro nel caso della circonferenza, ad uno dei due fuochi nel caso dell'ellisse, all'unico fuoco nel caso della parabola e al fuoco interno alla concavità del ramo su cui si svolge il moto nel caso dell'iperbole.

Il punto dell'orbita che corrisponde alla minima distanza dall'origine è detto *pericentro* e corrisponde al minimo valore assunto da r . Questo si ottiene quando $\cos(u - \omega) = 1$, ovvero quando $u = \omega$; per cui

$$r_{min} = \frac{p}{1 + e} . \quad (2.5)$$

La costante ω è perciò detta *argomento* o *longitudine del pericentro*

Nel caso di orbite chiuse è anche possibile ricavare il punto di massima distanza, detto *apocentro*. Questo si ottiene quando $\cos(u - \omega) = -1$, ovvero quando $u = \omega + \pi$; per cui

$$r_{max} = \frac{p}{1 - e} . \quad (2.6)$$

Calcolando la media aritmetica di r_{min} e r_{max} definiamo il *semiasse maggiore* dell'orbita (ovviamente solo nel caso delle orbite chiuse, per le quali esiste l'apocentro):

$$a = \frac{1}{2}(r_{min} + r_{max}) = \frac{p}{1 - e^2} ; \quad (2.7)$$

in questo modo possiamo riscrivere

$$p = a(1 - e^2) . \quad (2.8)$$

2.2 Orientazione nello spazio

Nel Capitolo 1 abbiamo detto che il moto di un corpo sotto l'azione della forza gravitazionale si svolge su un piano la cui orientazione rimane costante nel tempo. Ora il nostro obiettivo è di definire l'orientazione dell'orbita e del piano orbitale nello spazio; in particolare vogliamo ricavare dei parametri che definiscano l'orbita in maniera generale.

Consideriamo due sistemi di riferimento cartesiani (x, y, z) e (X, Y, Z) che abbiano origini coincidenti e orientazioni degli assi differenti ma fisse nel tempo. Il sistema di riferimento (x, y, z) deve essere inerziale, ovvero, per la descrizione del moto di un corpo rispetto a tale sistema di riferimento deve valere legge della dinamica $\mathbf{F} = m\mathbf{a}$; per cui anche la (1.11) è valida. Il sistema (X, Y, Z) è orientato in modo tale che l'asse Z sia parallelo al vettore \mathbf{c}

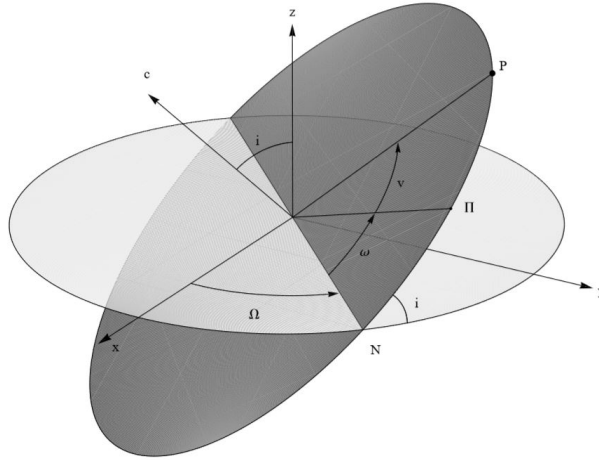


Figura 2.2: *Rappresentazione di una generica orbita nello spazio. Sono ben visibili gli angoli di Eulero Ω , i e ω che definiscono l'orbita. Il pericentro è indicato dalla lettera Π , il nodo ascendente con N , mentre P rappresenta il corpo in moto. L'angolo v è detto anomalia vera.*

della (1.12) (per cui il piano $(X, Y, 0)$ contiene il piano dell'orbita) e il pericentro r_{min} stia sull'asse X . L'intersezione tra l'orbita e il piano $(x, y, 0)$ dà luogo a due punti: il punto su cui giunge il corpo in moto da valori negativi a positivi di z è detto *nodo ascendente*, mentre l'altro è detto *nodo discendente*. Ora, la strategia per portarci dal sistema di riferimento (x, y, z) a (X, Y, Z) è quella di effettuare tre rotazioni consecutive, passando dunque attraverso due sistemi di riferimento intermedi (x_1, y_1, z_1) e (x_2, y_2, z_2) . Gli angoli che utilizzeremo per effettuare queste rotazioni saranno proprio i parametri orbitali che stiamo cercando.

Procediamo con quest'ordine:

1. $(x, y, z) \mapsto (x_1, y_1, z_1)$
 (x_1, y_1, z_1) è un sistema di riferimento tale che $z = z_1$ e x_1 contenga il nodo ascendente; per cui, in pratica, stiamo operando una rotazione attorno all'asse z . L'angolo di rotazione Ω è detto *longitudine del nodo ascendente*;
2. $(x_1, y_1, z_1) \mapsto (x_2, y_2, z_2)$
 (x_2, y_2, z_2) è un sistema di riferimento tale che $x_1 = x_2$ e z_2 sia parallelo al vettore \mathbf{c} ; si tratta quindi di effettuare una rotazione attorno all'asse x_1 di un angolo i detto *inclinazione*;
3. $(x_2, y_2, z_2) \mapsto (X, Y, Z)$
quest'ultimo passaggio consiste in una rotazione attorno all'asse z_2

(quindi attorno a \mathbf{c}) dell'angolo ω , la longitudine del pericentro, precedentemente introdotto. In questo modo abbiamo definito la posizione dell'orbita sul piano orbitale.

I tre angoli Ω , i e ω sono detti *angoli di Eulero*; questi, assieme all'eccentricità e , al semiasse maggiore (nel caso di orbite chiuse) a , al periodo P e al tempo di passaggio al pericentro T , costituiscono i sette parametri orbitali necessari a definire un'orbita (gli ultimi due verranno introdotti e utilizzati nel Capitolo 3). In Figura 2.2 è presente anche l'angolo v che corrisponde alla distanza angolare percorsa dal corpo, posizionato nel punto P , a partire dal pericentro, ovvero $v = u - \omega$. Questo è detto *anomalia vera* e sarà il punto di partenza per la trattazione del problema di Kepler.

Capitolo 3

L'equazione di Kepler

Finora abbiamo visto che le traiettorie dei corpi in moto sotto l'azione della forza gravitazionale sono delle coniche e abbiamo introdotto gli angoli di Eulero per identificare l'orientazione delle orbite nello spazio. Attraverso l'equazione

$$r = \frac{p}{1 + e \cos v} , \quad (3.1)$$

con $v = u - \omega$ anomalia vera, siamo quindi in grado di ricavare la posizione di un corpo in funzione dell'angolo v . A questo punto il nostro obiettivo è di ricavare un'equazione che ci dia la posizione del corpo in funzione del tempo.

Nel Capitolo 1 avevamo riscritto l'equazione integrale del momento angolare (1.12) nelle coordinate (ξ, η) e da queste siamo poi passati in coordinate polari, ricavando l'equazione (1.17), $r^2 \dot{u} = c$. Ora, $v = u - \omega$, per cui $\dot{u} = \dot{v}$ (poiché abbiamo visto che ω è un parametro fisso dell'orbita, indipendente dal tempo); ricordando che $p = c^2/\kappa^2$, la (1.17) può essere riscritta come

$$\frac{dv}{dt} = \frac{c}{r^2} = \frac{\kappa}{r^2} \left(\frac{c^2}{\kappa^2} \right)^{1/2} = \frac{\kappa}{r^2} p^{1/2} .$$

Eleviamo la (3.1) al quadrato e moltiplichiamo entrambi i membri per dv/dt ; per quanto scritto sopra

$$r^2 \frac{dv}{dt} = \frac{p^2}{(1 + e \cos v)^2} \frac{dv}{dt} = \kappa p^{1/2} ,$$

per cui

$$\frac{dv}{(1 + e \cos v)^2} = \frac{\kappa}{p^{3/2}} dt . \quad (3.2)$$

Integrando su v otteniamo l'equazione

$$\frac{\kappa}{p^{3/2}} t = \int \frac{dv}{(1 + e \cos v)^2} + cost. \quad (3.3)$$

che rappresenta l'*equazione di Kepler* in funzione dell'anomalia vera v . Attraverso la risoluzione dell'integrale nella parte destra dell'equazione potremmo ottenere la dipendenza dell'anomalia vera v dal tempo; tuttavia l'integrale non è risolvibile analiticamente, per cui è necessario prendere un'altra strada.

A questo punto possiamo dividere il problema nei due casi di *orbite chiuse* ($0 \leq e < 1$) e *orbite aperte* ($e \geq 1$). È utile introdurre una nuova variabile, l'*anomalia eccentrica* E , definita dalla relazione

$$\cos v = \frac{\cos E - e}{1 - e \cos E} . \quad (3.4)$$

Da questa definizione, ricordando che $\sin^2 v = 1 - \cos^2 v$, possiamo facilmente ricavare anche

$$\sin v = \frac{\sqrt{1 - e^2} \sin E}{1 - e \cos E} . \quad (3.5)$$

Sempre dalla (3.4) esplicitiamo

$$\cos E = \frac{\cos v + e}{1 + e \cos v} , \quad (3.6)$$

da cui ricaviamo

$$1 - e \cos E = \frac{1 - e^2}{1 + e \cos v} . \quad (3.7)$$

Infine, sfruttando la (3.5), troviamo

$$\sin E = \frac{\sqrt{1 - e^2} \sin v}{1 + e \cos v} . \quad (3.8)$$

Un'altra importante relazione analitica tra anomalia vera v ed anomalia eccentrica E si può ricavare sfruttando le regole di bisezione delle funzioni trigonometriche; in particolare troviamo che

$$\tan \frac{v}{2} = \sqrt{\frac{1 + e}{1 - e}} \tan \frac{E}{2} . \quad (3.9)$$

Facendo riferimento alla Figura 3.1, notiamo che, nel caso di orbite chiuse, E corrisponde alla distanza angolare, calcolata a partire da r_{min} , percorsa da un punto Q , che immaginiamo in moto lungo una circonferenza di raggio a che circoscrive l'orbita ellittica, tale che questo abbia sempre ascissa \vec{x} pari a quella del corpo in P che si muove lungo l'ellisse.

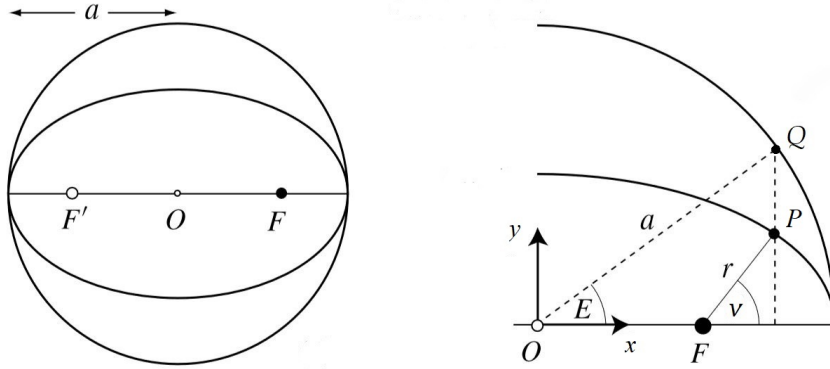


Figura 3.1: *Orbita ellittica circoscritta da una circonferenza; è evidente che, per un corpo situato nel punto P, anomalia vera v ed eccentrica E coincidano per angoli $0, \pi$.*

3.1 Orbite chiuse

Per questo primo caso è conveniente riscrivere l'equazione di Kepler in funzione dell'anomalia eccentrica E .

Per prima cosa deriviamo la (3.7) rispetto ad E tenendo conto della dipendenza di v da E

$$\frac{d}{dE}(1 - e \cos E) = \frac{d}{dE} \left(\frac{1 - e^2}{1 + e \cos v} \right) = \frac{d}{dv} \left(\frac{1 - e^2}{1 + e \cos v} \right) \frac{dv}{dE} ;$$

risulta

$$\sin E \, dE = \frac{(1 - e^2) \sin v}{(1 + e \cos v)^2} \, dv = \frac{(1 - e^2)^{3/2} \sin E}{(1 + e \cos v)^2 (1 - e \cos E)} \, dv ,$$

avendo sostituito nell'ultima parte $\sin v$ come in (3.5). Da questa possiamo esplicitare $\frac{dv}{(1 + e \cos v)^2}$ ed eguagliarla alla (3.2); ricordando la definizione di $p = a(1 - e^2)$, in cui a è il semiasse maggiore dell'orbita, troviamo

$$(1 - e \cos E) \, dE = \frac{\kappa}{a^{3/2}} \, dt .$$

Infine poniamo $n = \kappa/a^{3/2}$ e integriamo; il risultato è l'equazione trascendente

$$E - e \sin E = M \tag{3.10}$$

in cui $M = n(t - T)$ è detta *anomalia media* e T è una costante di integrazione che indica l'istante in cui $M = 0$. La (3.10) è l'*equazione di Kepler* nella forma

più nota; risolvendo per $E = E(M)$ possiamo ricavare la dipendenza di E (e, di conseguenza, di v) dal tempo. Notiamo che, quando $M = 0$, $E = e \sin E$, equazione verificata per $E = 0$; di conseguenza, dalla (3.5), anche $v = 0$; per cui T rappresenta l'istante in cui il corpo si trova nel pericentro. Nel caso di orbite chiuse, il passaggio al pericentro si verifica periodicamente per $v = E = M = 2k\pi$ (con $k = 0, 1, 2, \dots$). Il *periodo dell'orbita* P può essere definito come il tempo che intercorre tra due passaggi successivi al pericentro, ovvero tra $M_{t_0} = 0$ e $M_{t_1} = 2\pi$. Di conseguenza $M_{t_1} - M_{t_0} = 2\pi = nP$, da cui ricaviamo $P = \frac{2\pi}{n}$: n rappresenta quindi il termine di pulsazione del moto periodico e viene detto *moto medio*. Eguagliando questo risultato con la definizione precedentemente data di n , troviamo

$$n = \frac{2\pi}{P} = \frac{\kappa}{a^{3/2}} ;$$

Elevando al quadrato e riordinando l'equazione, ritroviamo la terza legge di Kepler: il quadrato del periodo è direttamente proporzionale al cubo del semiasse maggiore dell'orbita

$$\frac{P^2}{4\pi^2} = \frac{a^3}{\kappa^2} . \quad (3.11)$$

Ora abbiamo quindi tutte le relazioni che ci servono per ricavare la posizione di un corpo in funzione del tempo; dato un certo istante \tilde{t} , possiamo:

1. calcolare $M(\tilde{t})$ dalla definizione $M = n(\tilde{t} - T)$;
2. risolvere numericamente l'equazione di Kepler trovando $E(M(\tilde{t}))$;
3. ricavare v dalla (3.4) e, infine, ricavare r dalla (3.1).

Per quest'ultimo passaggio si può anche utilizzare la relazione che lega direttamente r con l'anomalia eccentrica E . Si ricava semplicemente sostituendo la (3.4) nell'equazione delle coniche (3.1); ricordando che $p = a(1 - e^2)$, otteniamo

$$r = \frac{p}{1 + e \frac{\cos E - e}{1 - e \cos E}} = \frac{p(1 - e \cos E)}{1 - e^2} = a(1 - e \cos E) \quad (3.12)$$

Inoltre, se consideriamo la decomposizione di r lungo gli assi X e Y del sistema di riferimento (X, Y, Z) introdotto nel Capitolo 2, possiamo scrivere

$$X = r \cos v = a(\cos E - e) ; \quad (3.13)$$

$$Y = r \sin v = a\sqrt{1 - e^2} \sin E . \quad (3.14)$$

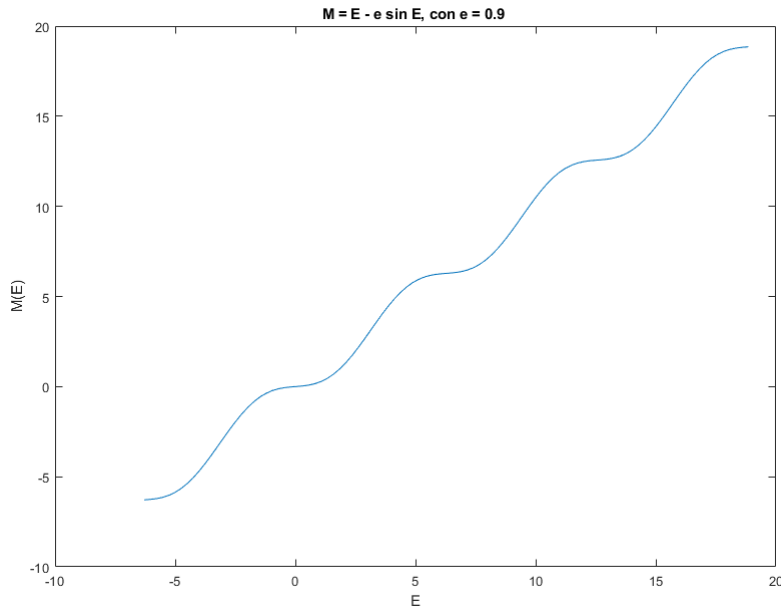


Figura 3.2: Grafico della funzione $M = E - e \sin E$, con $e = 0.9$.

Per quanto riguarda la risoluzione dell'equazione di Kepler, invece, si possono utilizzare diversi metodi numerici che vedremo in dettaglio nel Capitolo 4. Per ora ci limitiamo a fare qualche considerazione sulle possibili soluzioni.

L'equazione di Kepler $E - e \sin E = M$ può essere vista sia come una funzione esplicita $M = M(E)$ che come una funzione $E = E(M)$ implicitamente definita da $F_e(E, M) = E - e \sin E - M$. A noi interessa trovare soluzioni per quest'ultima, per cui facciamo le seguenti considerazioni:

1. $M = M(E)$ è una funzione monotona crescente, infatti $\frac{dM}{dE} = 1 - e \cos E > 0 \forall E$; di conseguenza $M = M(E)$ è invertibile con inversa $E = E(M)$;
2. poiché $\frac{dE}{dM} = \left(\frac{dM}{dE}\right)^{-1} > 0$, anche $E = E(M)$ è una funzione continua e monotona; inoltre $\lim_{M \rightarrow \pm\infty} E(M) = \pm\infty$;
3. $|E - M| = e|\sin E| < 1$, per cui $\lim_{E \rightarrow \pm\infty} M(E) = \pm\infty$.

Possiamo dunque concludere che, dato un certo M , esiste almeno una soluzione $E(M)$ che soddisfi l'equazione e che questa sia unica.

3.2 Orbite aperte

Il secondo caso da analizzare è quello di orbite paraboliche ed iperboliche. Per queste, l'eccentricità assume valori $e \geq 1$. Nel Capitolo 2 abbiamo definito $e = (1 + hc^2/\kappa^4)^{1/2}$, in cui h è detto integrale dell'energia; il fatto che $e \geq 1$ implica che $h \geq 0$, ovvero, l'energia totale del sistema a due corpi è positiva o nulla (al contrario di come accade nel caso di $e < 1$, per cui $h < 0$). Ora, come nel caso delle orbite chiuse, il nostro obiettivo è di riscrivere l'equazione di Kepler in una forma conveniente per essere risolta. Ricordando le relazioni (3.6) $\cos E = \frac{\cos v + e}{1 + e \cos v}$ e (3.8) $\sin E = \frac{\sqrt{1-e^2} \sin v}{1 + e \cos v}$, notiamo che, se $e \geq 1$, $\cos E$ può assumere valori maggiori di 1, mentre $\sin E$ assume valori immaginari. Per cui è preferibile trovare una nuova variabile che ci permetta di lavorare con equazioni reali.

Una buona idea è quella di utilizzare le funzioni iperboliche. Dalla forma complessa delle funzioni trigonometriche è facile ricavare le seguenti relazioni

$$\cos i\varphi = \cosh \varphi ;$$

$$\sin i\varphi = -i \sinh \varphi .$$

Per cui, operando il cambio di variabili $E = iH$, troviamo

$$\cos E = \cosh H ; \tag{3.15}$$

$$\sin E = -i \sinh H . \tag{3.16}$$

Sostituiamo ora $\cos E$ nell'equazione (3.12)

$$r = a(1 - e \cos E) = |a|(e \cosh H - 1) . \tag{3.17}$$

Nella sostituzione abbiamo tenuto conto del fatto che il semiasse maggiore $a < 0$ per $e \geq 1$: infatti $p = a(1 - e^2) = c^2/\kappa^2 \geq 0$, per cui, esplicitando a , otteniamo

$$a = \frac{c^2}{\kappa^2} \frac{1}{1 - (1 + h \frac{c^2}{\kappa^4})} = -\frac{\kappa^2}{h} ; \tag{3.18}$$

ma dato che $h \geq 0$, allora $a < 0$. Allo stesso modo possiamo sostituire $\cos E$ e $\sin E$ nelle espressioni (3.13), (3.14) delle componenti $X(E)$ e $Y(E)$

$$X = a(\cos E - e) = |a|(e - \cosh H) ; \tag{3.19}$$

$$Y = a\sqrt{1 - e^2} \sin E = |a|\sqrt{e^2 - 1} \sinh H . \tag{3.20}$$

Inoltre, possiamo anche riscrivere l'equazione di Kepler $E - \sin E = M$; definita l'anomalia eccentrica per il moto iperbolico (il moto parabolico verrà trattato a parte in seguito) $M_{hyp} = \frac{\kappa}{|a|^{3/2}}(t - T)$, troviamo

$$e \sinh H - H = M_{hyp} ; \tag{3.21}$$

i segni fanno sì che il passaggio al pericentro avvenga con il corpo in moto nel verso positivo delle Y . Per questa equazione è assicurata l'unicità delle soluzioni per ogni valore di M_{hyp} dato, per le stesse considerazioni fatte per l'equazione $E - e \sin E = M$.

3.2.1 Caso particolare: orbite paraboliche

Nel caso di orbite paraboliche, l'eccentricità assume valore $e = 1$. Ciò implica che $h = 0$, ovvero, l'energia totale del sistema è nulla. L'equazione delle coniche (3.1) nel caso parabolico diventa quindi

$$r = \frac{p}{1 + \cos v} . \quad (3.22)$$

Poniamo $q = r(0) = p/2$ la distanza del pericentro e $\sigma = \tan \frac{v}{2}$; ricordando la formula di bisezione della tangente $\tan \frac{v}{2} = \sqrt{\frac{1 - \cos v}{1 + \cos v}}$, possiamo riscrivere la (3.22) come

$$r = q(1 + \sigma^2) . \quad (3.23)$$

Direttamente dalla definizione di σ è facile ricavare $\cos v = \frac{1 - \sigma^2}{1 + \sigma^2}$ e $\sin v = \frac{2\sigma}{1 + \sigma^2}$; possiamo ora scrivere le componenti X e Y

$$X = r \cos v = q(1 - \sigma^2) ; \quad (3.24)$$

$$Y = r \sin v = 2q\sigma . \quad (3.25)$$

Nel caso del moto parabolico è possibile ricavare un'equazione di Kepler che sia risolvibile analiticamente. Vediamo come ottenerla. Per prima cosa, dalla definizione di $\sigma = \tan \frac{v}{2}$, derivando rispetto a v , troviamo

$$d\sigma = \frac{1}{2}(1 + \sigma^2) dv \quad (3.26)$$

Nel Capitolo 1 abbiamo scritto l'integrale del momento angolare (1.17) come $r^2 \dot{v} = c$; esplicitando \dot{v} e sostituendo r come nella (3.22) troviamo

$$\dot{v} = \frac{dv}{dt} = \frac{c}{q^2(1 + \sigma^2)^2} .$$

Ora, $p = 2q = c^2/\kappa^2$, per cui $c = \kappa\sqrt{2}q^{1/2}$; sostituendo nell'equazione di sopra risulta quindi

$$dv = \frac{k\sqrt{2}}{q^{3/2}(1 + \sigma^2)^2} dt .$$

A questo punto sostituiamo dv nell'equazione (3.26), la riordiniamo e troviamo

$$(1 + \sigma^2) d\sigma = \frac{\sqrt{2}}{2} \frac{\kappa}{q^{3/2}} dt,$$

che, integrata, ci dà l'equazione di Kepler nel caso di orbite paraboliche

$$\frac{1}{3}\sigma^3 + \sigma = M_{par}, \quad (3.27)$$

in cui $M_{par} = \frac{\sqrt{2}}{2} \frac{\kappa}{q^{3/2}} (t - T)$ è l'anomalia media per il moto parabolico. Come anticipato, questa equazione cubica può essere risolta analiticamente utilizzando le formule di Cardano; in particolare, dato un certo M_{par} , avremo un'unica soluzione reale

$$\sigma(M_{par}) = \frac{1}{2} Q^{1/3} - 2 Q^{-1/3} \quad (3.28)$$

in cui $Q = 12M_{par} + 4\sqrt{4 + 9M_{par}^2}$ è sempre positivo.

Capitolo 4

Metodi risolutivi per l'equazione di Kepler

L'equazione di Kepler è un'equazione trascendente in E , per cui non è possibile scrivere una soluzione analitica esplicita $E = E(M)$. Tuttavia, numerosi metodi numerici sono stati sviluppati e applicati a questa equazione nel corso dei secoli per risolvere il problema.

Per prima cosa, ricordiamo quanto detto nel Capitolo 3 per dimostrare esistenza e unicità della soluzione $E = E(M)$:

1. $M = M(E)$ è una funzione monotona crescente, infatti $\frac{dM}{dE} = 1 - e \cos E > 0 \forall E$; di conseguenza $M = M(E)$ è invertibile con inversa $E = E(M)$;
2. poiché $\frac{dE}{dM} = \left(\frac{dM}{dE}\right)^{-1} > 0$, anche $E = E(M)$ è una funzione continua e monotona; inoltre $\lim_{M \rightarrow \pm\infty} E(M) = \pm\infty$;
3. $|E - M| = e|\sin E| < 1$, per cui $\lim_{E \rightarrow \pm\infty} M(E) = \pm\infty$.

Inoltre abbiamo detto che questo vale anche nel caso si considerino moti iperbolici, per i quali l'equazione di Kepler diventa $e \sinh H - H = M_{hyp}$. In questo capitolo ci concentreremo esclusivamente sul caso delle orbite ellittiche, ma è importante far presente che i metodi di risoluzione qui analizzati possono essere estesi anche al caso di orbite aperte.

In molti metodi risolutivi l'equazione di Kepler viene riscritta nella forma

$$F_e(E) = E - e \sin E - M = 0, \quad (4.1)$$

per cui il problema si traduce nella ricerca dell'unico zero della funzione. In questo modo, oltre all'eccentricità e , anche l'anomalia media M viene considerata come un parametro. Come risulta evidente da $M = E - e \sin E$, i valori

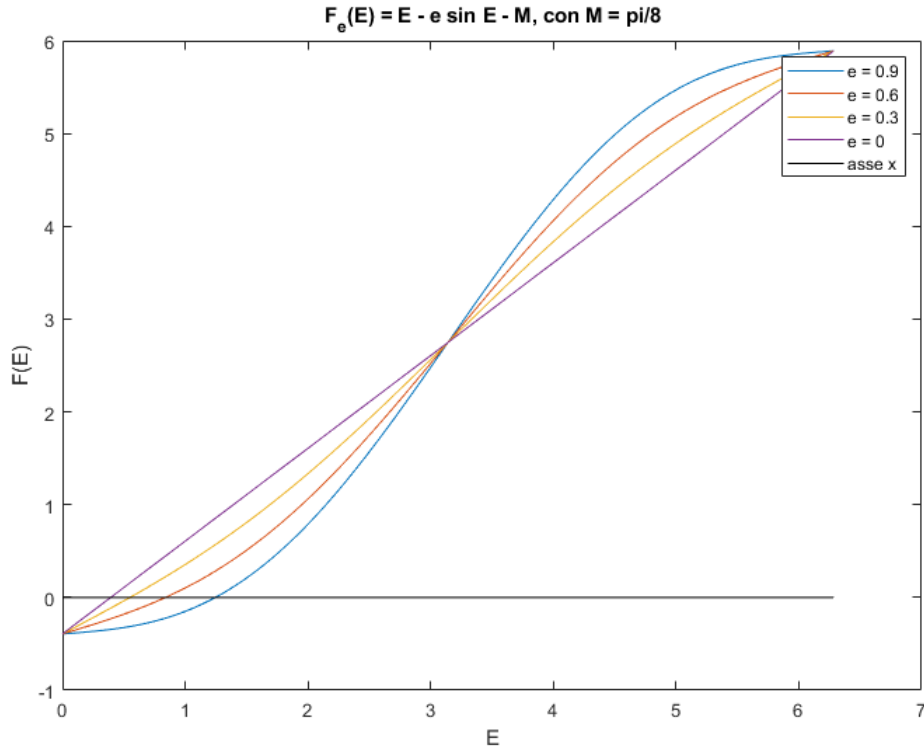


Figura 4.1: Grafico della funzione $F_e(E) = E - e \sin E - M$ nell'intervallo $[0, 2\pi]$ per quattro diversi valori dell'eccentricità e e anomalia media fissata a $M = \pi/8$.

che possono essere assegnati ad M dipendono strettamente dall'anomalia eccentrica E e, dal momento che le orbite ellittiche sono periodiche, possiamo limitarci a studiare la funzione per $E \in [0, 2\pi]$, limitando allo stesso intervallo i possibili valori di M . In realtà ha senso ridurre ulteriormente il dominio di E all'intervallo $[0, \pi]$; infatti, notiamo che $F_e(-E) = -E + e \sin E - M = 0$, per cui, ponendo $M' = -M(E)$ otteniamo un'equazione analoga all'equazione di Kepler $F_e(E) = 0$.

I metodi che analizzeremo sono detti *iterativi*, ovvero, si basano sulla ripetizione di un calcolo fino a quando non si arriva al risultato cercato. In particolare, a partire da un valore iniziale x_0 possiamo ottenere una successione di valori $x_{k+1} = g(x_k)$ (in cui g è una funzione che regola il calcolo dei termini successivi) che, dopo un certo numero di iterazioni, converga al risultato cercato (che indichiamo con \bar{x}); chiaramente questo avviene solo se $\lim_{k \rightarrow +\infty} x_k = \bar{x}$. Alcuni elementi fondamentali di un metodo iterativo sono:

- l'errore di troncamento al k -esimo passo $\epsilon_k = \bar{x} - x_k$;

- la velocità di convergenza p definita dal $\lim_{k \rightarrow +\infty} \frac{|\epsilon_{k+1}|}{|\epsilon_k|^p} = C$, dove C è detto fattore di convergenza. Se $p = 1$ la convergenza è lineare, mentre se $p = 2$ la convergenza è quadratica.

È inoltre fondamentale stabilire un criterio di arresto basato sul numero di cifre significative desiderato per il risultato finale e sul numero massimo di iterazioni da effettuare.

4.1 Metodo della falsa posizione (regula falsi)

Il metodo *regula falsi* è un classico metodo numerico che permette di stimare la soluzione di un'equazione scritta nella forma $f(x) = 0$ tramite l'intersezione con l'asse delle ascisse di una retta passante per due punti del piano appartenenti alla curva $(x, f(x))$.

In generale: consideriamo una funzione $y = f(x)$ della quale vogliamo trovare \bar{x} tale che $f(\bar{x}) = 0$. Prendiamo due punti sull'asse delle ascisse a_0 e b_0 tali che $f(a_0)f(b_0) < 0$ e calcoliamo la retta congiungente i punti del piano $(a_0, f(a_0))$ e $(b_0, f(b_0))$ tramite la formula

$$\frac{y - f(a_0)}{f(b_0) - f(a_0)} = \frac{x - a_0}{b_0 - a_0} . \quad (4.2)$$

A questo punto poniamo $y = 0$ e ricaviamo il punto di intersezione della retta con l'asse x

$$x_1 = a_0 - f(a_0) \frac{b_0 - a_0}{f(b_0) - f(a_0)} ; \quad (4.3)$$

questa rappresenta un primo valore approssimato della soluzione. A questo punto il processo può essere iterato prendendo x_1 come nuovo a_0 o b_0 a seconda che, rispettivamente, $f(x_1)f(b_0) < 0$ o $f(x_1)f(a_0) < 0$. Per cui, in generale, alla $(k + 1)$ -esima iterazione

$$x_{k+1} = a_k - f(a_k) \frac{b_k - a_k}{f(b_k) - f(a_k)} ; \quad (4.4)$$

se $f(x_k)f(b_k) < 0$ teniamo fisso $a_{k+1} = a_k$ e poniamo $b_{k+1} = x_k$, mentre se $f(x_k)f(a_k) < 0$ poniamo $a_{k+1} = x_k$ e teniamo fisso $b_{k+1} = b_k$. Dal momento che ad ogni iterazione vengono considerati intervalli contenenti la soluzione \bar{x} (in quanto appartenente all'asse \mathbf{x}) è garantita la convergenza globale del metodo. L'equazione (4.4) rappresenta quindi una successione di punti che converge alla soluzione.

Applichiamo ora il metodo all'equazione di Kepler, che riscriviamo come

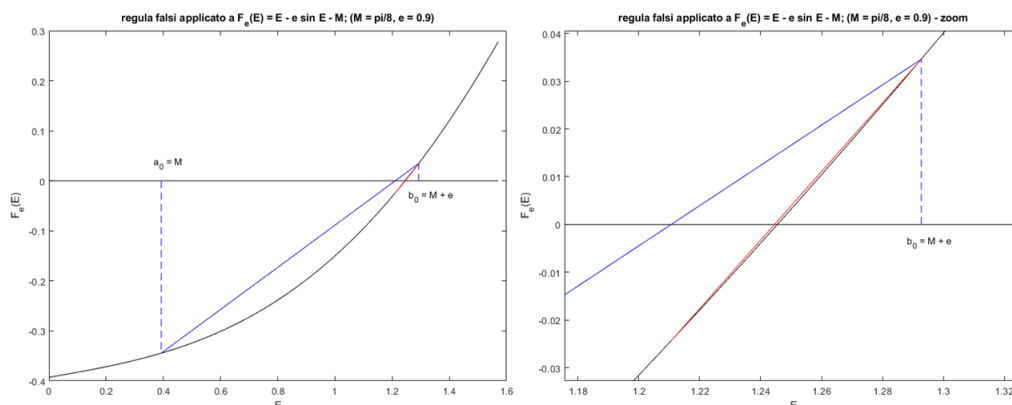


Figura 4.2: Metodo della falsa posizione applicato all'equazione di Keplero nella forma F_e con $e = 0.9$ ed $M = \pi/8$.

$F_e(E) = E - e \sin E - M$. Per prima cosa dobbiamo trovare dei valori iniziali a_0 e b_0 che soddisfino l'ipotesi $F_e(a_0)F_e(b_0) < 0$. Restringendoci al caso in cui $M \in [0, \pi]$, notiamo che $F_e(M) = -e \sin M \leq 0$, mentre $F_e(M + e) = e(1 - \sin(M + e)) \geq 0$, per cui possiamo prendere come valori iniziali $a_0 = M$ e $b_0 = M + e$. Svolgendo i calcoli della (4.4), troviamo per la prima iterazione

$$E_1 = M + \frac{e \sin M}{1 - \sin(M + e) + \sin M} ; \quad (4.5)$$

per l'iterazione successiva, se $F_e(E_1) < 0$ allora ripetiamo il procedimento con $a_1 = E_1$ e $b_1 = M + e$, mentre se $F_e(E_1) > 0$ teniamo fisso $a_1 = M$ e poniamo $b_1 = E_1$. In Figura 4.1 troviamo una rappresentazione delle prime due iterazioni del metodo applicato all'equazione di Keplero con valori di eccentricità e anomalia media fissati ad $e = 0.9$ ed $M = \pi/8$. Osservando la differenza tra le scale dell'asse x delle due figure come, alla seconda iterazione, la soluzione abbia precisione maggiore di almeno un ordine di grandezza rispetto alla prima. Con l'implementazione del metodo (che mostreremo nel Capitolo 4) siamo in grado di ottenere il profilo di convergenza della successione.

4.2 Metodo delle sostituzioni successive

Il metodo delle sostituzioni successive (anche noto come metodo delle iterazioni) è un altro metodo numerico che permette di risolvere un'equazione

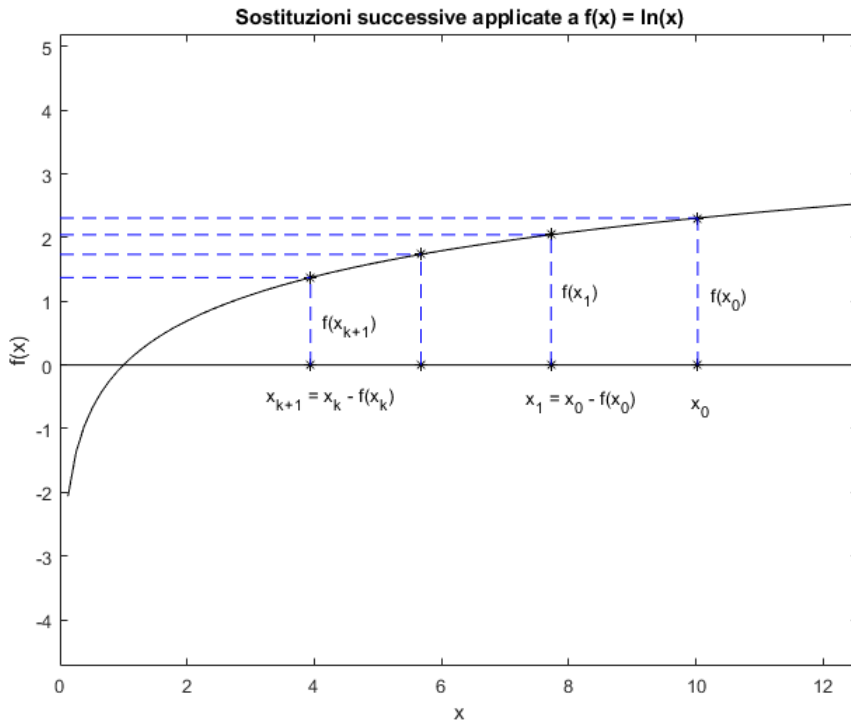


Figura 4.3: Metodo delle sostituzioni successive applicato a $f(x) = \ln x$.

nella forma $f(x) = 0$ tramite la generazione di una successione

$$\begin{aligned}
 x_1 &= x_0 - f(x_0) \\
 x_2 &= x_1 - f(x_1) \\
 &\dots \\
 x_{k+1} &= x_k - f(x_k)
 \end{aligned}
 \tag{4.6}$$

convergente al punto \bar{x} tale che $f(\bar{x}) = 0$. Fissato un valore di tolleranza $\varepsilon \in \mathbb{R}$, il processo viene iterato fino a quando $|x_{k+1} - x_k| < \varepsilon$, per cui x_{k+1} rappresenterà il valore approssimato della soluzione che stavamo cercando. Un esempio grafico è rappresentato in Figura 4.2.

Applichiamo ora il metodo al nostro caso. Per prima cosa, riscriviamo l'equazione di Kepler nella forma $F_e = E - E \sin E - M = 0$ e prendiamo come valore iniziale $E_0 = 0$. Seguendo la regola definita nella (4.6) ricaviamo i termini della successione

$$\begin{aligned}
 E_1 &= E_0 - F_e(E_0) = E_0 - E_0 + e \sin E_0 + M = M \\
 E_2 &= E_1 - E_1 + e \sin E_1 + M = M + e \sin M ;
 \end{aligned}$$

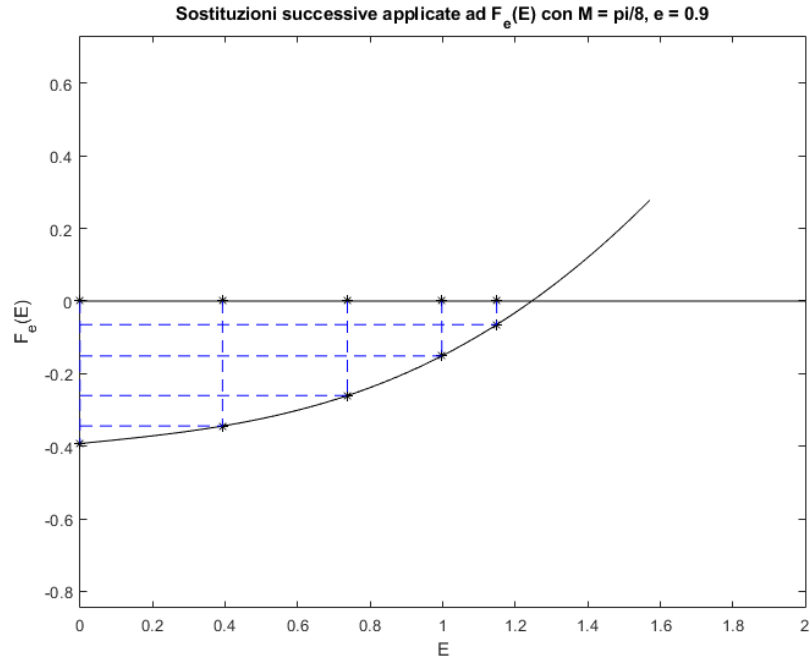


Figura 4.4: Metodo delle sostituzioni successive applicato all'equazione di Kepler con $e = 0.9$ ed $M = \pi/8$ fissati.

notiamo che, ad ogni iterazione, i termini lineari E_k si elidono tra loro, per cui la successione generale diventa

$$E_{k+1} = M + e \sin E_k . \quad (4.7)$$

La convergenza del metodo discende direttamente dal criterio di Cauchy¹ per la convergenza delle successioni reali una volta dimostrato che per ogni $\varepsilon > 0$ esiste k tale per cui

$$|E_{k+n} - E_k| < \varepsilon \quad (4.8)$$

con $n \in \mathbb{N}$. Vediamo come: per prima cosa notiamo che

$$\begin{aligned} E_{k+n} - E_k &= E_{k+1} - E_k + E_{k+2} - E_{k+1} + \cdots + E_{k+n} - E_{k+n-1} \\ &= \sum_{m=k}^{k+n-1} E_{m+1} - E_m \end{aligned} . \quad (4.9)$$

¹**Teorema** (Criterio di Cauchy). Una successione di numeri reali x_n converge ad un numero reale $l \Leftrightarrow \forall \varepsilon > 0 \exists N_\varepsilon$ intero positivo t.c. $|x_n - x_m| < \varepsilon \forall n, m > N_\varepsilon$.

Ora scriviamo esplicitamente il termine generale della sommatoria

$$\begin{aligned} E_{m+1} - E_m &= e(\sin E_m - \sin E_{m-1}) \\ &= 2e \sin\left(\frac{E_m - E_{m-1}}{2}\right) \cos\left(\frac{E_m + E_{m-1}}{2}\right) \\ &= e(E_m - E_{m-1}) \frac{\sin\left(\frac{E_m - E_{m-1}}{2}\right)}{\frac{E_m - E_{m-1}}{2}} \cos\left(\frac{E_m + E_{m-1}}{2}\right) ; \end{aligned}$$

nel secondo passaggio abbiamo utilizzato le formule di prostaferesi. Ora, ricordando che $\frac{\sin x}{x} \rightarrow 1$ per $x \rightarrow 0$ e $\cos x \leq 1$, effettuiamo le seguenti maggiorazioni

$$|E_{m+1} - E_m| \leq e|E_m - E_{m-1}| \leq \dots \leq e^m |E_1 - E_0| = Me^m . \quad (4.10)$$

Ora, tornando alla relazione (4.9), troviamo

$$\begin{aligned} |E_{k+n} - E_k| &= \sum_{m=k}^{k+n-1} |E_{m+1} - E_m| \leq M \sum_{m=k}^{k+n-1} e^m \\ &= Me^k \frac{1 - e^n}{1 - e} \leq M \frac{e^k}{1 - e} \end{aligned} \quad (4.11)$$

in cui, nel terzo passaggio, abbiamo utilizzato la formula per le somme geometriche parziali e nell'ultimo abbiamo sfruttato il fatto che $0 < e < 1$. Per cui, la successione (4.7) convergerà per valori di k maggiori di

$$k > \frac{\log \varepsilon \frac{1-e}{M}}{\log e} . \quad (4.12)$$

4.3 Metodo di Newton-Raphson

Il metodo di Newton-Raphson (o semplicemente di Newton) ci permette di calcolare la soluzione approssimata di un'equazione nella forma $f(x) = 0$. In generale: sia $f : [a, b] \rightarrow \mathbb{R}$ una funzione derivabile in $[a, b]$ tale che $f(a)f(b) < 0$ e $f'(x) \neq 0 \forall x \in [a, b]$. Sia inoltre $\bar{x} \in [a, b]$ tale che $f(\bar{x}) = 0$. Consideriamo un punto iniziale $x_0 \in [a, b]$; la retta tangente alla funzione in x_0 è data dallo sviluppo in serie di Taylor di $f(x)$ attorno a x_0 stoppato al primo ordine $y - f(x_0) = f'(x_0)(x - x_0)$, per cui la sua intersezione con l'asse delle ascisse $y = 0$ sarà data da $x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$. Iterando il procedimento $k + 1$ volte, con $k \in \mathbb{N}$, troviamo la successione

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \quad (4.13)$$

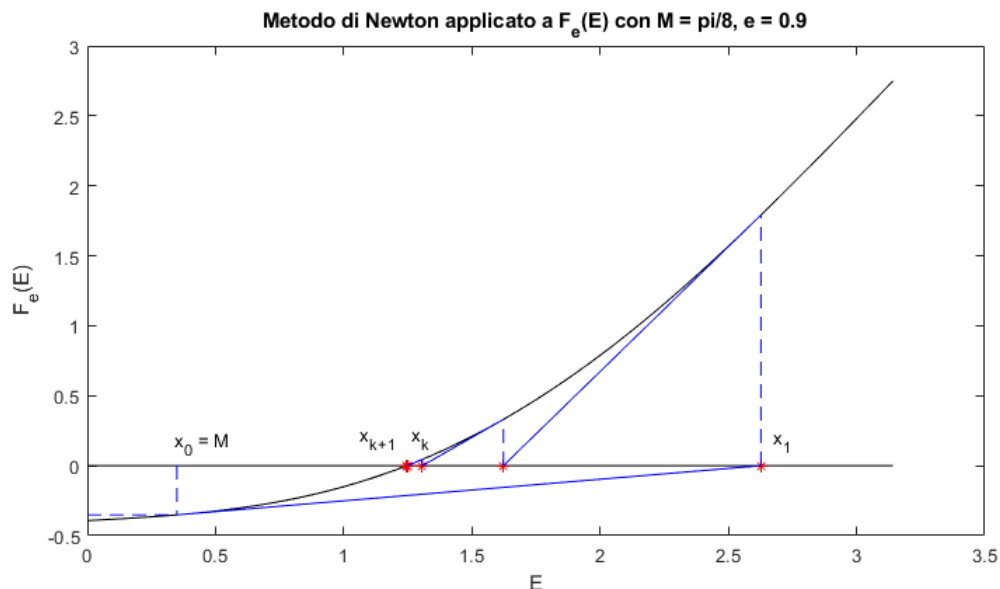


Figura 4.5: Metodo di Newton applicato all'equazione di Kepler $F_e(M)$ con eccentricità ed anomalia media fissate ad $e = 0.9$ ed $M = \pi/8$. Abbiamo preso $E_0 = M$ come valore iniziale.

che fornisce un valore approssimato della soluzione \bar{x} , ovvero $\lim_{k \rightarrow \infty} x_k = \bar{x}$. La convergenza globale del metodo è garantita quando sono verificate le ipotesi del seguente

Teorema 1. Sia $f \in C^2([a, b])$ tale che $f(a)f(b) < 0$ con f'' di segno costante. Se $f(a)f'(a) < b - a$ e $f(b)f'(b) < b - a$, allora il metodo di Newton converge all'unico zero \bar{x} di f in $[a, b]$.

Inoltre, se, come da ipotesi, $f'(\bar{x}) \neq 0$, ovvero \bar{x} è uno zero semplice, è garantito l'andamento quadratico della convergenza. Lo dimostriamo brevemente: dopo k iterazioni, preso a tale che $\bar{x} < a < x_k$, possiamo scrivere lo sviluppo in serie di Taylor attorno a x_k con resto in forma di Lagrange, valutato in \bar{x}

$$0 = f(\bar{x}) = f(x_k) + f'(x_k)(\bar{x} - x_k) + \frac{1}{2}f''(a)(\bar{x} - x_k)^2, \quad (4.14)$$

da cui esplicitiamo

$$\frac{f(x_k)}{f'(x_k)} = x_k - \bar{x} - \frac{1}{2}f''(a)(\bar{x} - x_k)^2; \quad (4.15)$$

per cui

$$x_{k+1} - \bar{x} = \frac{1}{2} \frac{f''(a)}{f'(x_k)} (x_k - \bar{x})^2, \quad (4.16)$$

ovvero, l'errore alla $(k + 1)$ -esima iterazione è proporzionale al quadrato di quello associato alla k -esima iterazione.

Nel nostro caso, riscrivendo l'equazione di Kepler come

$$F_e(E) = E - e \sin E - M = 0 \quad (4.17)$$

la successione (4.13) diventa

$$E_{k+1} = E_k + \frac{M - M_k}{1 - e \cos E_k}, \quad (4.18)$$

in cui $M_k = E_k - e \sin E_k$. L'errore associato alla $(k + 1)$ -esima iterazione sarà invece dato dalla relazione (4.16), per cui

$$|E_{k+1} - \bar{E}| = \frac{e |\sin A|}{2(1 - e |\cos E_k|)} |E_k - \bar{E}|^2 \leq \frac{e}{2(1 - e)} |E_k - \bar{E}|^2, \quad (4.19)$$

in cui \bar{E} tale che $F(\bar{E}) = 0$ ed $\bar{E} < A < E_k$.

Capitolo 5

Implementazione dei metodi

Nel Capitolo 4 abbiamo introdotto tre diversi metodi numerici permettono la risoluzione di equazioni scritte nella forma $f(x) = 0$ e che possono essere utilizzati per risolvere l'equazione di Kepler.

In questo Capitolo invece mostriamo i risultati dell'implementazione dei tre metodi applicati all'equazione di Kepler (tramite software di calcolo numerico MATLAB) ed il confronto tra di essi in termini di robustezza, intesa come applicabilità del metodo a tutte le configurazioni in cui può apparire l'equazione (in questo caso determinate dalla scelta della coppia di parametri e, M) e di costo computazionale, ovvero il numero di iterazioni necessarie per ottenere il risultato con la precisione richiesta.

Come abbiamo detto in precedenza, quando si adoperano dei metodi iterativi per la risoluzione di equazioni è necessario stabilire dei criteri di arresto basati sul numero massimo di iterazioni che si vogliono impiegare e sulla precisione della soluzione che si vuole ottenere. La scelta di quest'ultima, nel nostro caso, dipenderà dalla precisione con cui vogliamo determinare la posizione del corpo, quindi dal valore dell'anomalia media v o del raggio r .¹ In ogni caso, il nostro scopo è quello di mostrare l'andamento generale del numero di iterazioni necessarie ad ogni metodo per trovare la soluzione dell'equazione di Kepler, perciò, in ognuno dei programmi abbiamo imposto l'arresto delle iterazioni quando:

- viene raggiunto il numero massimo di iterazioni consentite, fissato a $k_{max} = 250$;

¹La relazione tra la precisione dE e v o r è data dalla derivazione dell'errore relativo dv/v e dr/r . Ricordando che $\tan \frac{v}{2} = \sqrt{\frac{1+e}{1-e}} \tan \frac{E}{2}$, troviamo $\frac{dv}{v} = \frac{\sqrt{1-e^2}}{1-e \cos E} \frac{1}{2 \arctan\left(\sqrt{\frac{1+e}{1-e}} \tan \frac{E}{2}\right)} dE$; per il raggio invece, ricordando che $r = a(1 - e \cos E)$, troviamo $\frac{dr}{r} = \frac{E-M}{1-e \cos E} dE$.

- il risultato dell'iterazione ha una precisione di 10^{-12} .

Quest'ultima scelta è stata fatta sulla base di quanto fatto da Conway nel 1986 in una ricerca simile. Abbiamo comunque appurato che, pur cambiando i valori di tolleranza e iterazioni massime, l'andamento del numero di iterazioni rimane qualitativamente lo stesso raffigurato nei grafici seguenti, non smentendo quindi la validità delle nostre considerazioni finali.

La compilazione del codice per una data coppia dei parametri e, M restituisce il numero di iterazioni, il valore x_k trovato ad ogni iterazione e lo scarto $E_k - E_{k-1}$; inoltre crea un grafico in cui è visibile il profilo di convergenza della successione, ovvero l'andamento dello scarto all'aumentare delle iterazioni. Qui sotto vediamo un esempio dell'*output* dei programmi

```

iterazione n 0      [Ek]: 1.371200955031613      [Ek-Ek-1]  1.42e-02
iterazione n 1      [Ek]: 1.385062228855144      [Ek-Ek-1]  1.39e-02
iterazione n 2      [Ek]: 1.384734243200663      [Ek-Ek-1]  1.35e-02
iterazione n 3      [Ek]: 1.384414018347311      [Ek-Ek-1]  5.33e-10
...
iterazione n 249    [Ek]: 1.373792635039864      [Ek-Ek-1]  5.20e-10
iterazione n 250    [Ek]: 1.373792635027560      [Ek-Ek-1]  5.08e-10
raggiunto il numero massimo di iterazioni

```

I codici dei programmi sono riportati in Appendice A. Procediamo ora con la disamina dei tre metodi.

5.1 Implementazione *regula falsi*

L'algoritmo che abbiamo ricavato per il metodo della falsa posizione è dato dalla relazione (4.4)

$$x_{k+1} = a_k - f(a_k) \frac{b_k - a_k}{f(b_k) - f(a_k)}.$$

Come anticipato nel Capitolo 4, per questo metodo abbiamo preso valori iniziali $a_0 = M$ e $b_0 = M + e$ dato che nel caso dell'equazione di Kepler ristretta al dominio $[0, \pi]$ si verifica che $F_e(M)F_e(M + e) < 0$. Visto che per ipotesi la soluzione è sempre contenuta all'interno dell'intervallo $[a_k, b_k]$, in questo caso abbiamo imposto il criterio di arresto in base alla lunghezza dello stesso (per cui quando lo scarto $b_k - a_k < 10^{-12}$) piuttosto che sulla differenza tra due iterate successive, rendendo così il calcolo più preciso. A seconda dei valori di eccentricità ed anomalia media considerati, il programma impiega più o meno iterazioni per raggiungere il risultato, o, addirittura, supera il

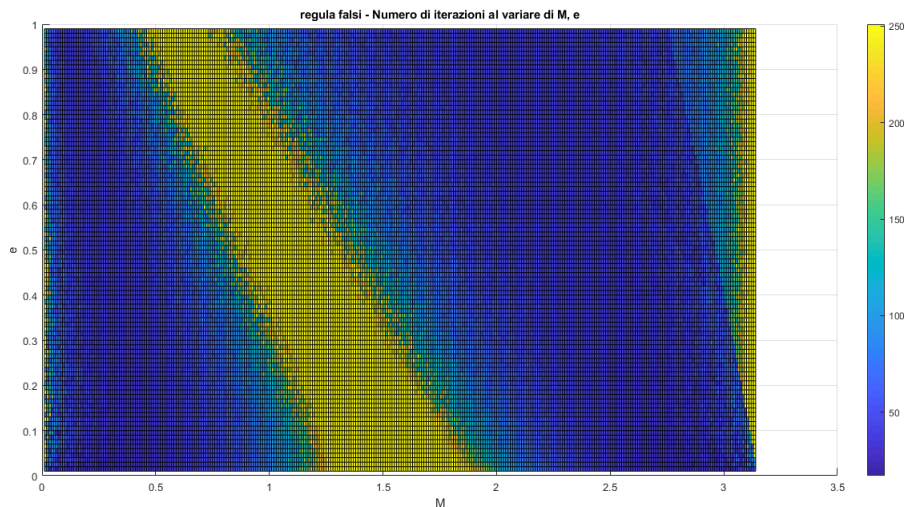


Figura 5.1: Metodo regula falsi. Proiezione sul piano M, e di un grafico 3D raffigurante il numero di iterazioni (legenda colorata) effettuate per ogni coppia di valori M, e .

limite massimo di iterazioni stabilito e non fornisce la soluzione cercata. Per avere una visione generale di questo fatto osserviamo il grafico in Figura (5.1) che mostra il numero di iterazioni necessarie al raggiungimento del risultato con la precisione richiesta per ogni coppia di valori e, M ($0 < M < \pi$, $0 < e < 0.99$ con passo 0.01). Da questo, seguendo la legenda colorata, notiamo che per le coppie di punti comprese nelle regioni gialle del grafico e nel loro intorno è necessario un maggior numero di iterazioni per raggiungere il risultato cercato. La regione di *non* convergenza è molto estesa nella parte centrale del grafico corrispondente a valori dell'anomalia media compresi tra 0.5 e 2 ed è presente anche per valori di M tendenti a *zero* e a π . Notiamo inoltre che il numero di iterazioni minime richieste dal metodo nei casi più favorevoli (zone blu del primo grafico) è circa 30.

Per concludere, in Figura 5.1 troviamo i profili di convergenza del metodo della falsa posizione applicato all'equazione di Kepler con $M = \pi/4$ ed eccentricità una volta $e = 0.2$ e l'altra $e = 0.6$: nel primo caso notiamo che il metodo converge dopo 61 iterazioni (come ci potevamo aspettare dal grafico in Figura 5.1), mentre, nel secondo caso, il metodo raggiunge il numero massimo di iterazioni prima di arrivare al risultato (il valore dello scarto all'ultima iterazione è dell'ordine di 10^{-10} , superiore di due ordini di grandezza rispetto a quanto richiesto).

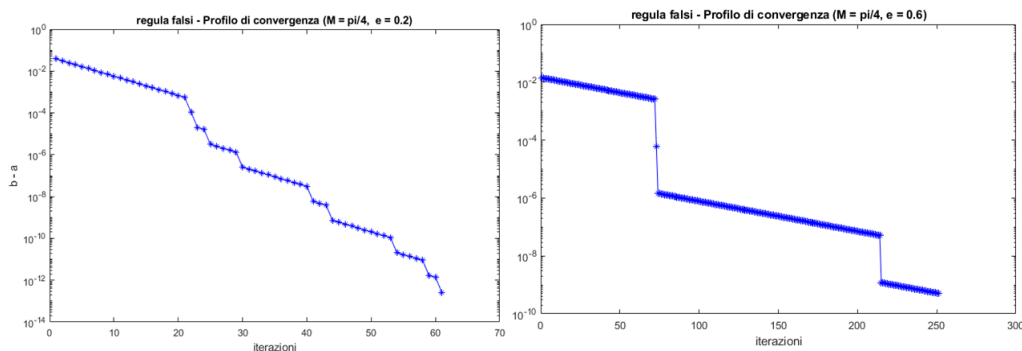


Figura 5.2: *Profili di convergenza del metodo della falsa posizione applicato all'equazione di Kepler con $M = \pi/4$ ed, a sinistra, $e = 0.2$, a destra, $e = 0.6$.*

5.2 Implementazione sostituzioni successive

L'algoritmo del metodo delle sostituzioni successive è dato dalla relazione 4.6

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

Come anticipato nel Capitolo 4, il valore iniziale che abbiamo considerato per la risoluzione dell'equazione di Kepler è $E_0 = 0$. Nel grafico in Figura 5.3 possiamo osservare il numero di iterazioni per tutte le combinazioni dei parametri e, M , come fatto per il metodo *regula falsi*. A differenza di quanto abbiamo visto per quest'ultimo, notiamo che nel metodo delle sostituzioni successive la successione converge al valore cercato entro il numero massimo di iterazioni per la maggior parte delle coppie e, M , con un costo crescente man mano che si raggiungono valori alti dell'eccentricità assieme a valori bassi dell'anomalia media ed anche quando queste crescono insieme; le regioni gialle, in cui i risultati non sono soddisfacenti, sono confinate nella zona in cui entrambi i parametri raggiungono i loro massimi valori contemporaneamente e per M tendente a zero.

Consideriamo ora un punto nella regione *blu* e uno nella regione *gialla* del grafico in Figura 5.2; i profili di convergenza relativi alle coppie di valori dei punti considerati sono riportati in Figura 5.4 (in questo metodo lo scarto è dato dalla differenza tra due iterate successive $E_k - E_{k-1}$). Come possiamo vedere, i profili di convergenza mostrano in entrambi i casi un andamento lineare nella scala logaritmica. Notiamo inoltre che nel primo caso sono state necessarie 16 iterazioni, sensibilmente meno rispetto al valore minimo trovato per la convergenza del metodo della falsa posizione.

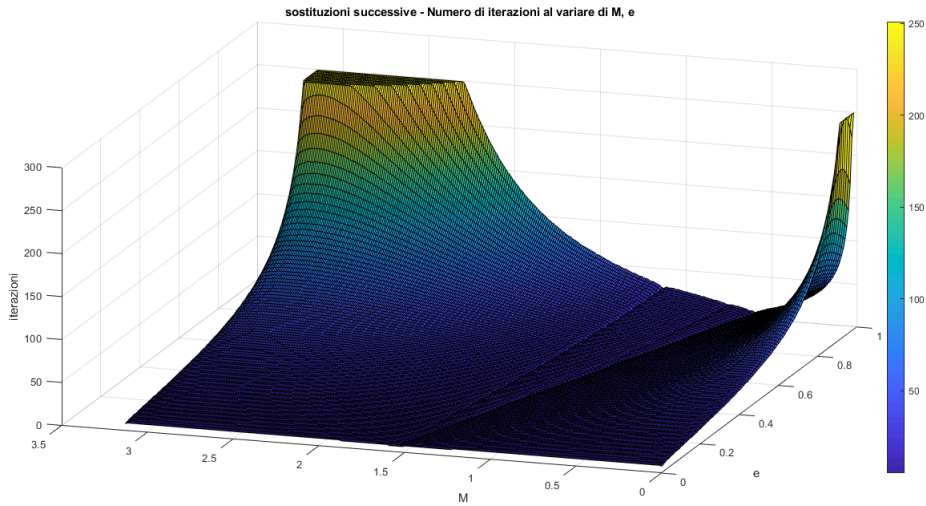


Figura 5.3: Metodo delle sostituzioni successive. Grafico 3D raffigurante il numero di iterazioni effettuate per ogni coppia di valori M, e .

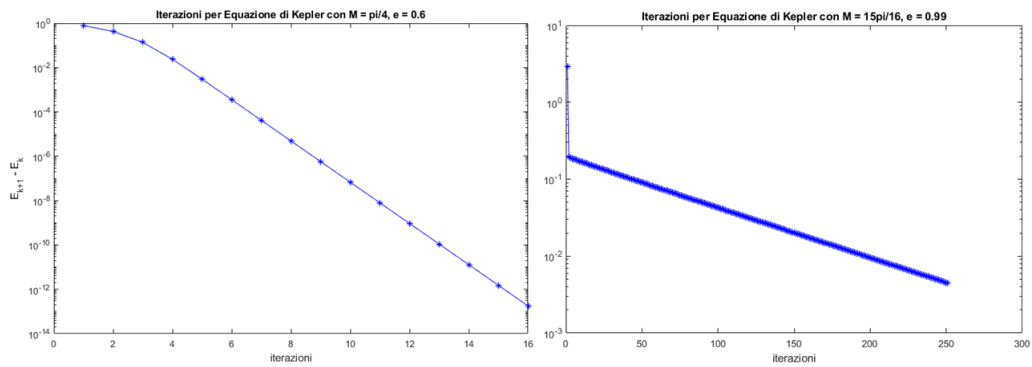


Figura 5.4: Profilo di convergenza del metodo delle sostituzioni successive applicato all'equazione di Kepler con, a sinistra, $M = \pi/4, e = 0.6$, a destra, $M = 3\pi/4, e = 0.99$

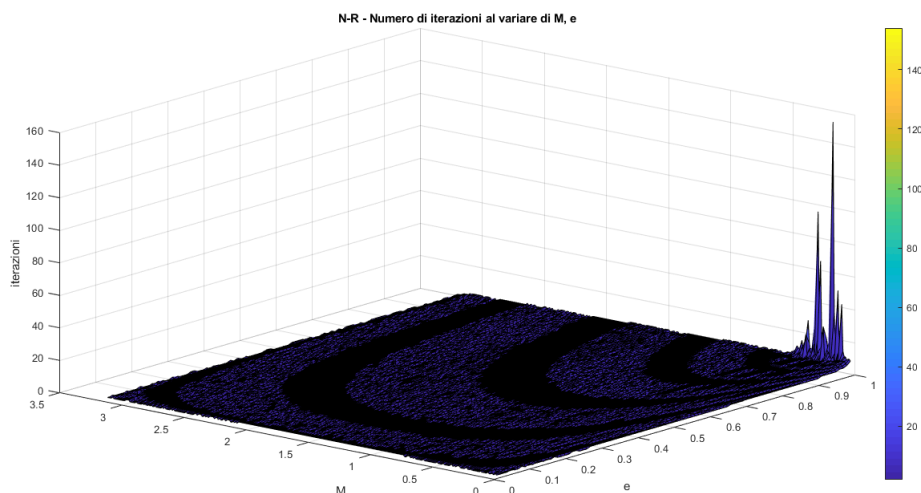


Figura 5.5: Metodo di Newton. Grafico 3D raffigurante il numero di iterazioni effettuate per ogni coppia di valori M, e .

5.3 Implementazione Newton-Raphson

L'algoritmo del metodo di Newton è dato dalla relazione (4.13)

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}.$$

Per la risoluzione dell'equazione di Kepler possiamo prendere come valore iniziale $E_0 = M$, come suggerito da molti testi in letteratura.

Come fatto per gli altri casi, consideriamo il grafico in Figura 5.3 che ci riferisce il numero di iterazioni effettuate per raggiungere il risultato cercato al variare della coppia di parametri e, M . Due fatti risultano evidenti: il primo è che per la quasi totalità delle combinazioni eccentricità-anomalia media, il numero di iterazioni necessario per la convergenza al risultato cercato è molto basso; il secondo è che in nessun caso viene superato il tetto delle 250 iterazioni, per cui, per ogni coppia e, M , il metodo di Newton converge ad un risultato che abbia una precisione di 10^{-12} . È però fondamentale notare che, per come abbiamo costruito il grafico ($0 < M < \pi$, $0 < e < 0.99$ con passo 0.01) questo vale solo per valori di e ed M definiti fino alla seconda cifra decimale. Infatti, se andiamo ad infittire il grafico prendendo valori con passo 0.001, troviamo che il problema si ripresenta (Figura 5.5, posizione degli assi invertita rispetto alla precedente). A questo punto andiamo a vedere i profili di convergenza. Nel primo caso, in cui abbiamo considerato i valori $M = \pi/4$ ed $e = 0.3$, sono bastate 4 iterazioni per raggiungere il risultato con precisione 10^{-12} e inoltre è evidente l'andamento quadratico dello scarto

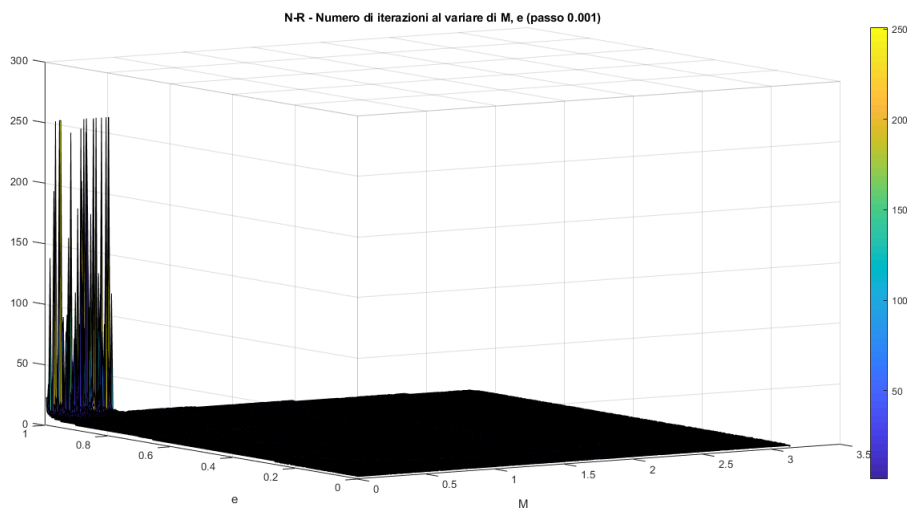


Figura 5.6: Metodo di Newton. Grafico 3D raffigurante il numero di iterazioni effettuate per ogni coppia di valori M, e e con passo 0.001.

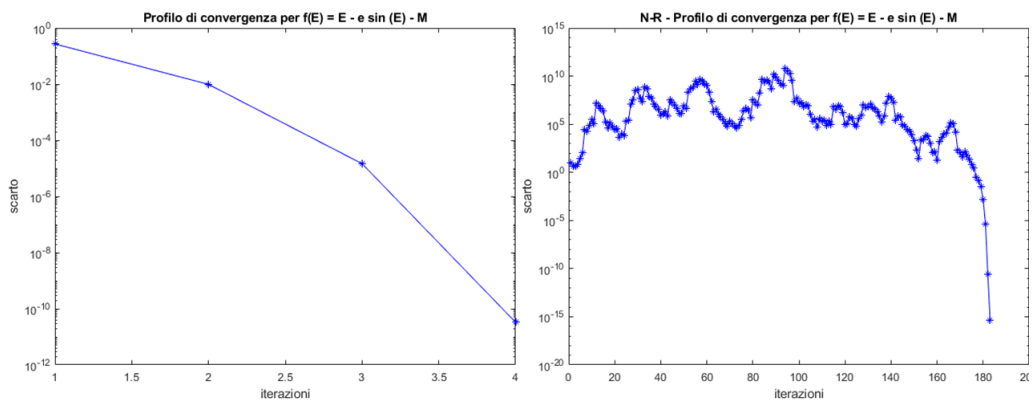


Figura 5.7: Profilo di convergenza del metodo di Newton applicato all'equazione di Kepler con, a sinistra, $M = \pi/4$, $e = 0.3$, a destra, $M = \pi/100$, $e = 0.997$

$E_k - E_{k-1}$ (come previsto dalle considerazioni teoriche fatte nel Capitolo 4). Nel secondo caso, invece, sono state necessarie 183 iterazioni ed è evidente un andamento estremamente irregolare dello scarto.

5.4 Confronto dei metodi

Per ognuno dei metodi abbiamo mostrato qual è il numero di iterazioni necessario affinché si raggiunga la precisione desiderata di 10^{-12} e la variazione di questo al variare dei parametri e, M . Mettendo a confronto i risultati ottenuti per i tre metodi notiamo che:

- in generale, il metodo di Newton, per ogni coppia di valori e, M , ha un costo computazionale sensibilmente minore rispetto agli altri metodi; a sua volta, il metodo delle sostituzioni successive ha un costo minore rispetto al metodo della falsa posizione;
- a parità di iterazioni massime, osserviamo che il metodo della falsa posizione è quello che presenta regioni di non convergenza più ampie; a seguire, il metodo delle sostituzioni successive presenta due regioni in cui il costo computazionale aumenta rispetto all'andamento medio, una per valori molto bassi di M e molto alti di e e una per valori alti di entrambi i parametri; nel terzo ed ultimo metodo, invece, il problema persiste solo per bassi valori di M ed alti valori di E .

Da queste considerazioni possiamo concludere che, tra i tre metodi, quello di Newton è il più efficiente sia per quanto riguarda il costo computazionale medio che per quanto riguarda la robustezza, poiché è possibile ottenere un risultato che rispetti le condizioni stabilite per un maggior numero di coppie e, M rispetto agli altri casi.

Appendice A

Codici Matlab

A.1 Codice *regula falsi*

```
function [E,iter,scarti]=falsapos(f,a,b,tol,itmax)

fa=feval(f,a);
fb=feval(f,b);
if fa*fb>0
disp('L intervallo iniziale non è accettabile')
return
end
E1=a-fa*(b-a)/(fb-fa);
E=[E1]; iter=0; scarti=[];
fE=feval(f,E1);
while abs(fE)>tol || (b-a)>tol
if fE*fa<0
b=E1;
else
a=E1;
end
fprintf('\n iterazione n %2.0f \t ...
[Ek]: %10.15f \t [b-a] %10.2e ', iter, E1, b-a)
E=[E, E1]; iter=iter+1;
scarti=[scarti, b-a];
E1=a-fa*(b-a)/(fb-fa);
fE=feval(f,E1);
if (iter >=itmax)
disp('raggiunto il numero massimo di iterazioni')
return
end
```

```

end

Codice di chiamata alla function

% valore parametri
M = pi/4; e=0.2;

% definizione della funzione
f=@(x) x-e*sin(x)-M; a=0; b=pi;

% tolleranza e numero massimo di iterazioni
tol=1e-12; itmax=251;

% grafico funzione e asse x
figure(1)
ezplot(f,[a b]);hold on; plot([a b],[0 0],'k-');hold off;

% punti iniziali
a=M;
b=M+e;

% Chiamata alla function
[E,iter,scarti]=falsapos(f,a,b,tol,itmax);

% profilo di convergenza
figure(2)
semilogy(1:iter,abs(scarti),'b-*')
title('Profilo di convergenza per f(E) = E - e sin (E) - M');
ylabel('b-a');

Codice grafico delle iterazioni

M=[0.01:0.01:pi];
e=[0.01:0.01:0.99];
tol=1.e-12;
itmax=251;
for j = [1:length(M)]
for k = [1:length(e)]
a=M(j);
b=M(j)+e(k);
f=@(x) x-e(k)*sin(x)-M(j);
[E,iter(j,k),scarti]=falsapos(f,a,b,tol,itmax);
end
end
[X,Y]=meshgrid(e,M);
surf(X,Y,iter)

```

A.2 Codice sostituzioni successive

```
function [x,iter,scarti]=iterazioni(f,x0,tol,itmax)
xold=x0;
x=[x0];
scarti=[];
dif=tol+1;
iter=0;
while (abs(dif) > tol) & (iter < itmax)
fx=feval(f,xold);
if fx==0
disp('soluzione esatta')
return;
end
iter=iter+1;
xnew=xold-fx;
dif=fx;
fprintf('\n iterazione n: %d \t ...
[E_%d]: %10.15f \t [Ek+1-Ek] %10.2e '...
,iter, iter, xnew, dif)
scarti=[scarti, dif];
x=[x, xnew];
xold=xnew;
if iter>itmax
disp('raggiunto il numero massimo di iterazioni')
return;
end
end
```

Codice di chiamata alla *function*

```
% valore parametri
M = 3*pi/4; e=0.99;

% definizione della funzione
f=@(x) x-e*sin(x)-M; a=0; b=pi;

% tolleranza e numero massimo di iterazioni
tol=1e-12;
itmax=251;

%grafico funzione e asse x
ezplot(f,[a b]);
hold on;
```

```

plot([a b],[0 0], 'k-');
hold off;

%valore iniziale
x0=0;

%chiamata alla function
[x,iter,scarti]=iterazioni(f,x0,tol,itmax);
% profilo di convergenza
figure(2)
semilogy(1:iter,abs(scarti),'b-*)
title('Iterazioni per Equazione di Kepler');
xlabel('iterazioni');
ylabel('E_k+_1 - E_k');

```

Codice grafico delle iterazioni

```

M=[0.01:0.01:pi];
e=[0.01:0.01:0.99];
tol=1.e-12;
itmax=251;
for j = [1:length(M)]
for k = [1:length(e)]
x0=0
f=@(x) x-e(k)*sin(x)-M(j);
[E,iter(j,k),scarti]=iterazioni(f,x0,tol,itmax);
end
end
[X,Y]=meshgrid(e,M);
surf(X,Y,iter)

```

A.3 Codice Newton-Raphson

```

function [x,iter,scarti]=newton(f,df,x0,tol,itmax)
xold=x0;
x=[x0];
scarti=[];
dif=tol+1;
iter=0;
while (abs(dif) > tol) && (iter < itmax)
fx=feval(f,xold);
if fx == 0
disp('soluzione esatta')

```

```

return;
end
dfx=feval(df,xold);
if dfx == 0
disp('ERRORE: derivata nulla.');
```

```

return;
end
iter=iter+1;
dif=-fx/dfx;
xnew=xold+dif;
fprintf('\n iterazione n: %2.0f \t ...
[Ek]: %10.15f \t [dif] %10.2e ',iter,...
xnew, dif)
scarti=[scarti;dif];
x=[x; xnew];
xold=xnew;
end
if (iter >=itmax)
disp('raggiunto il numero massimo di iterazioni')
end
fprintf('\n');
```

Codice di chiamata alla *function*

```

% valore parametri
M = pi/4; e=0.3;

% definizione della funzione
f=@(x) x-e*sin(x)-M; a=0; b=pi;

%definizione della derivata prima
df=@(x) 1-e*cos(x);

% tolleranza e numero massimo di iterazioni
tol=1e-12;
itmax=251;

ezplot(f,[a b]);
hold on;
plot([a b],[0 0],'k-');
hold off;
%valore iniziale
x0=M;
```

```

% chiamata alla function
[x,iter,scarti]=newton(f,df,x0,tol,itmax);

% profilo di convergenza
figure(2)
semilogy(1:iter,abs(scarti),'b-*')
title('Profilo di convergenza per f(E) = E - e sin (E) - M');
xlabel('iterazioni');
ylabel('scarto');

Codice grafico delle iterazioni

M=[0.01:0.01:pi];
e=[0.01:0.01:0.99];
tol=1.e-12;
itmax=251;
for j = [1:length(M)]
for k = [1:length(e)]
x0=M(j);
f=@(x) x-e(k)*sin(x)-M(j);
df=@(x) 1-e(k)*cos(x);
[x,iter(j,k),scarti]=newton(f,df,x0,tol,itmax);
end
end
[X,Y]=meshgrid(e,M);
surf(X,Y,iter)

```


Bibliografia

- [1] Sergei A. Klioner (2011), *Lecture Notes on Basic Celestial Mechanic*.
- [2] Carl D. Murray, Stanley F. Dermott (1999), *Solar System dynamics*, Cambridge University Press.
- [3] Leonida Rosino (1979), *Lezioni di Astronomia*, Edizioni Cedam - Padova.
- [4] Richard H. Battin (1999) *An introduction to the mathematics and methods of astrodynamics*, American Institute of Aeronautics and Astronautics.
- [5] Bruce A. Conway (1986), *An improved algorithm due to Laguerre for the solution of Kepler's equation*, Celestial Mechanics, Volume 39, Issue 2, pp.199-211.
- [6] Andrea Baruffolo (2014-2015), *Materiale didattico per il corso di Sperimentazioni di Fisica 1*, Corso di Laurea Triennale in Astronomia presso l'Università degli Studi di Padova.