

# Soluzione numerica di equazioni differenziali ordinarie

L. P.

5 Novembre 2010

## 1. Metodo di Eulero

Consideriamo un'equazione differenziale ordinaria della forma

$$\frac{dx}{dt} = f(x, t), \quad (1)$$

dove  $f(x, t)$  è una funzione definita su una regione opportuna  $D$  del piano  $(x, t)$ . Supponiamo di voler conoscere la soluzione  $x(t)$  di questa equazione, che soddisfi la condizione iniziale

$$x(t_0) = x_0, \quad (2)$$

dove  $(x_0, t_0)$  appartiene all'interno di  $D$ .

Vogliamo impostare un metodo numerico per la soluzione di questa equazione. Ovviamente non possiamo pretendere di avere un'espressione analitica come risultato. Quello che potremo avere sarà una tabella di valori  $t_n$  della variabile indipendente, e dei corrispondenti valori  $x_n$  di  $x(t_n)$ :

$$\begin{array}{ll} t_0 & x_0 \\ t_1 & x_1 \\ t_2 & x_2 \\ \vdots & \vdots \end{array}$$

Per valutare le quantità  $x_n$ , applicheremo successivamente dei metodi che permettano di stimare  $x(t+h)$ , dove  $h$  è un piccolo incremento di  $t$  che supporremo positivo, quando sia noto  $x(t)$  (ed eventualmente valori precedenti di  $x$ ).

Supponiamo di conoscere  $x(t_0) = 0$ . Consideriamo lo sviluppo di Taylor della soluzione  $x(t)$  della (1) attorno a  $t_0$ :

$$x(t_0 + h) = x_0 + hx_1 + O(h^2). \quad (3)$$

Si ha evidentemente

$$x_1 = \left. \frac{dx}{dt} \right|_{t=t_0}. \quad (4)$$

D'altra parte, poiché  $x(t)$  è per ipotesi la soluzione della (1) attorno a  $t = t_0$ , si ha

$$\left. \frac{dx}{dt} \right|_{t=t_0} = f(x_0, t_0). \quad (5)$$

Otteniamo così una stima di  $x(t_0 + h)$ :

$$\tilde{x}(t_0 + h) = x_0 + hf(x_0, t_0). \quad (6)$$

Vediamo che la differenza fra questa stima e la soluzione esatta  $x(t)$  è proporzionale a  $h^2$ .

Questo ragionamento suggerisce un primo metodo per risolvere l'equazione differenziale (1) lungo l'intervallo  $[t_0, t_0 + T]$  con la condizione iniziale  $x(t_0)$ . Si divide l'intervallo in  $N$  subintervalli  $[t_{n-1}, t_n]$ ,  $n = 1, \dots, N$ , con  $t_N = t_0 + T$ ,  $t_n - t_{n-1} = T/N = h$ . Si pone  $x(t_0) = x_0$  e si valutano successivamente  $x(t_n)$ ,  $n = 1, \dots, N$  mediante l'espressione

$$x(t_n) = x(t_{n-1}) + hf(x(t_{n-1}), t_{n-1}), \quad n = 1, \dots, N. \quad (7)$$

Otteniamo così una tabella  $(t_n, x_n)$ , che può essere raffinata mediante interpolazione. Questo metodo di soluzione numerica è noto come **metodo di Eulero**.

Per valutare l'errore in cui si incorre con questo metodo, supponiamo di conoscere la soluzione esatta  $x(t)$  nell'intervallo  $[t_0, t_0 + h]$ . Si ha allora

$$x(t_0 + h) = x(t_0) + \int_{t_0}^{t_0+h} dt f(x(t), t). \quad (8)$$

Per il teorema della media, esiste un valore  $\bar{t}$  di  $t$ , compreso fra  $t_0$  e  $t_0 + h$ , tale che

$$x(t_0 + h) = x(t_0) + h f(x(\bar{t}), \bar{t}). \quad (9)$$

Ora  $x(t)$  è derivabile, e se  $f(x, t)$  è anche essa derivabile rispetto ai suoi argomenti, e poiché  $|\bar{t} - t_0| < h$ , si ha

$$|f(x(\bar{t}), \bar{t}) - f(x_0, t_0)| < Kh, \quad (10)$$

per una qualche costante positiva  $K$ . Otteniamo così

$$|\delta x| < Kh^2. \quad (11)$$

D'altra parte, il numero di intervalli di lunghezza  $h$  in cui dobbiamo dividere un intervallo di lunghezza finita  $T$  è pari a  $T/h$ . Quindi l'errore sulla stima di  $x(t_0 + T)$  è proporzionale a  $h^1$ . Si dice che il metodo di Eulero è un metodo del **primo ordine**.

Consideriamo per esempio l'equazione

$$\frac{dx}{dt} = x, \quad (12)$$

con la condizione iniziale

$$x(0) = 1. \quad (13)$$

È noto che la soluzione di questa equazione differenziale con la relativa condizione iniziale è data da

$$x(t) = e^t. \quad (14)$$

Valutando la soluzione con il metodo di Eulero, è facile vedere che si ottiene

$$x(t + nh) = (1 + h)^n. \quad (15)$$

Quindi, se fissiamo  $T$  e  $N$  otteniamo l'approssimazione

$$x_N(T) = \left(1 + \frac{T}{N}\right)^N. \quad (16)$$

Si ha evidentemente

$$\lim_{N \rightarrow \infty} x_N(T) = x(T). \quad (17)$$

Vediamo in figura 1 che la soluzione converge effettivamente alla soluzione esatta, però abbastanza lentamente. In effetti, immaginiamo di voler valutare  $x(T)$  con  $T$  fissato. Se dividiamo l'intervallo  $[0, T]$  in  $N$  intervalli, il "passo"  $h$  dell'approssimazione è pari a  $1/N$ , e l'errore è proporzionale a  $h^2$  a ogni passo. Quindi l'errore su  $x(T)$ , dato che gli errori si accumulano, sarà proporzionale a  $Nh^2 \sim 1/N$ . Per avere una cifra significativa in più nella valutazione di  $x(T)$  occorrerà introdurre 10 volte tanti punti. Per intenderci, con  $N = 32$  si ha  $x_N(1) = 2.6770$  invece del risultato esatto  $x(1) = 2.7183$ , mentre con 320 punti si ha  $x_N(1) = 2.7140$ .

L'errore introdotto dal metodo di Eulero può avere conseguenze spiacevoli. Consideriamo per esempio il sistema di equazioni differenziali ordinarie

$$\frac{dx}{dt} = -y; \quad \frac{dy}{dt} = x. \quad (18)$$

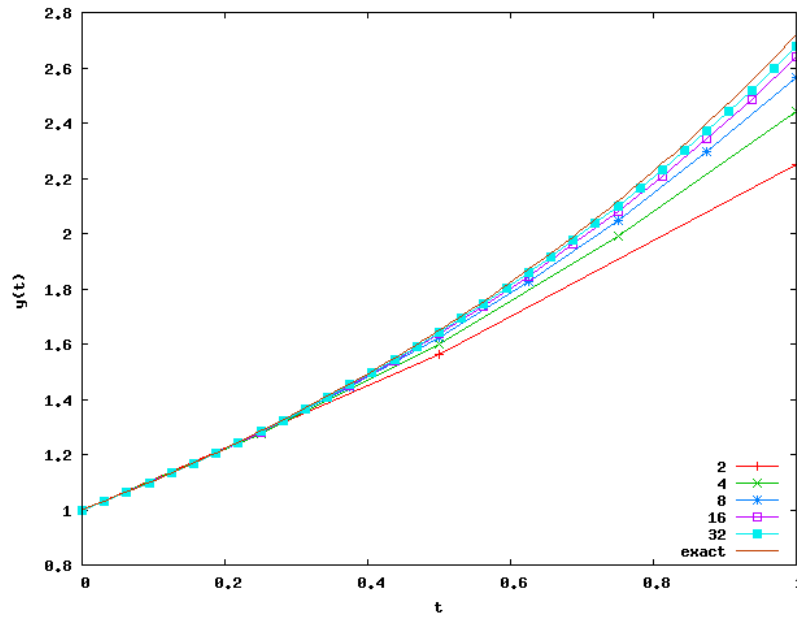
La soluzione esatta di questa equazione differenziale che soddisfa le condizioni iniziali ( $x(0) = 1, y(0) = 0$ ) è data dalla coppia  $\cos t, \sin t$ , per cui si ha ovviamente

$$x^2(t) + y^2(t) = x_0^2 + y_0^2; \quad \forall t. \quad (19)$$

Usando il metodo di Eulero, otteniamo la seguente relazione:

$$x(t_0 + h) = x(t_0) - hy(t_0); \quad (20)$$

$$y(t_0 + h) = y(t_0) + hx(t_0). \quad (21)$$



**Figura 1.** Approssimazioni successive a  $x(t) = \exp(x)$ , mediante il metodo di Eulero, con  $n = 2, 4, 8, 16, 32$  punti. Viene mostrata anche la soluzione esatta.

È utile scrivere questa relazione in forma matriciale:

$$\mathbf{X}(t_0 + h) = (\mathbf{1} + h\mathbf{F})\mathbf{X}(t_0), \quad (22)$$

dove

$$\mathbf{X}(t_0) = \begin{pmatrix} x(t_0) \\ y(t_0) \end{pmatrix}, \quad (23)$$

$\mathbf{1}$  è la matrice identità, e

$$\mathbf{F} = \begin{pmatrix} 0, & -1 \\ 1, & 0 \end{pmatrix}. \quad (24)$$

Otteniamo così la seguente stima di  $\mathbf{X}(T)$ :

$$\mathbf{X}_N(T) = \begin{pmatrix} 1, & -T/N \\ T/N, & 1 \end{pmatrix}^N \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \quad (25)$$

Ora, se calcoliamo  $x_N^2(T) + y_N^2(T)$ , otteniamo

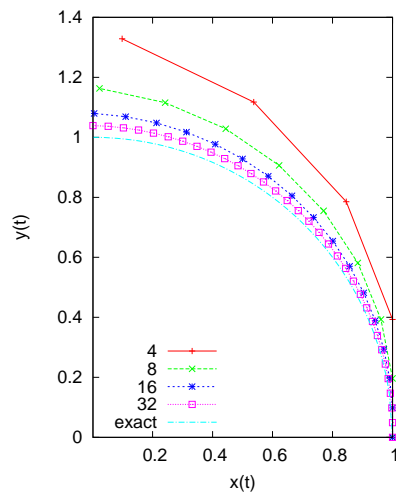
$$x_N^2(T) + y_N^2(T) = \left(1 + \left(\frac{T}{N}\right)^2\right)^N (x^2(0) + y^2(0)) \simeq e^{T^2/N} (x^2(0) + y^2(0)). \quad (26)$$

Si vede dalla figura 2 che la soluzione numerica tende ad allontanarsi dalla soluzione esatta al crescere di  $T$ , formando una spirale logaritmica, come si può vedere, ad esempio, in figura 3.

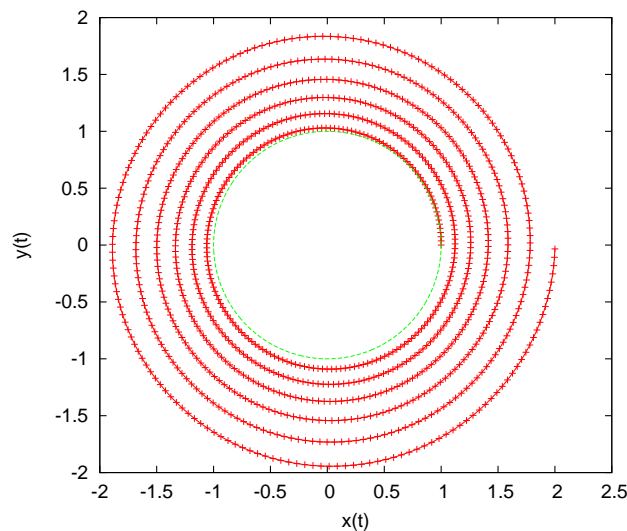
## 2. Metodo di Heun

Per avere una migliore approssimazione si potrebbe pensare di considerare dei termini successivi nello sviluppo di Taylor (3) e del secondo membro dell'equazione (1). Tuttavia questo approccio non è molto pratico, perché richiede di valutare esplicitamente le derivate successive della  $f(x, t)$ , che debbono prima essere espresse algebricamente e poi valutate numericamente. Vorremmo invece avere dei metodi in cui sia possibile limitarsi a valutare la  $f(x, t)$ , magari più volte.

È possibile trovare un metodo di questo tipo con le seguenti considerazioni. Consideriamo un'equazione differenziale  $x'(t) = f_x(x, y)$ , la cui soluzione esatta  $x(t)$  è descritta dalla curva in figura 4. Supponiamo che per un certo valore  $t_0$  di  $t$  la soluzione si trovi nel punto  $P = (t_0, x_0)$ . Allora  $f(x_0, t_0)$  è



**Figura 2.** Soluzione dell'equazione differenziale (18) con il metodo di Eulero, per  $0 \leq t \leq \pi/2$ , con condizione iniziale  $x(0) = 1$ ,  $y(0) = 0$ , e per  $N = 4, 8, 16, 32$ . Viene anche mostrata la soluzione esatta  $x(t) = \cos t$ ,  $y(t) = \sin t$ .

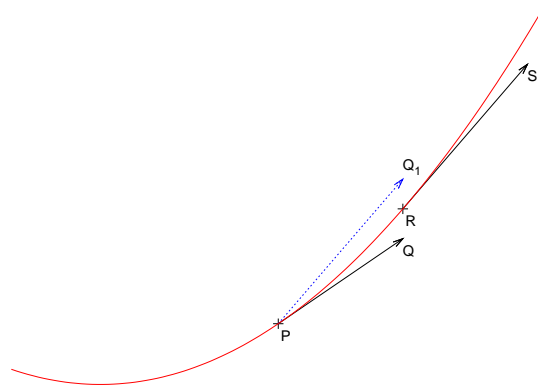


**Figura 3.** Soluzione dell'equazione differenziale (18) con il metodo di Eulero, per  $0 \leq t \leq 12\pi$ , con condizione iniziale  $x(0) = 1$ ,  $y(0) = 0$ , e per  $N = 1024$ . Viene anche mostrata la soluzione esatta  $x(t) = \cos t$ ,  $y(t) = \sin t$ .

la pendenza della tangente in P alla curva. Per  $t = t_0 + h$  la soluzione esatta si troverà, per esempio, in R, mentre la soluzione ottenuta con il metodo di Eulero dovrà trovarsi lungo la tangente, per esempio in Q. Nel nostro caso, in cui la soluzione esatta è concava, Q si trova al disotto della curva: in altri termini il metodo di Eulero sottostima la soluzione.

Supponiamo di conoscere il valore esatto della soluzione per  $t = t_0 + h$ , e quindi di  $f_1 = f(x(t_0 + h), t_0 + h)$ , che corrisponde alla pendenza del segmento RS. Se, invece di valutare  $x(t_0 + h)$  con il metodo di Eulero mediante  $f(x_0, t_0)$  usassimo  $f_1$ , otterremmo per  $x(t_0 + h)$  la stima rappresentata dal punto Q<sub>1</sub> (i vettori PQ<sub>1</sub> e RS sono uguali): in questo caso abbiamo sovrastimato  $x(t_0 + h)$ . Notiamo che se la curva esatta  $x(t)$  fosse stata convessa, piuttosto che concava, il metodo di Eulero avrebbe fornito una sovrastima, mentre il corrispondente vettore PQ<sub>1</sub> avrebbe fornito una sottostima.

Sembra evidente che si potrebbe ottenere una migliore stima di  $x(t_0 + h)$  prendendo la media di



**Figura 4.** Verso la definizione del metodo di Heun. Data la soluzione nota nel punto P, il punto Q rappresenta la stima della soluzione con il metodo di Eulero. La curva rossa rappresenta la soluzione esatta. Il punto  $Q_1$  rappresenta la soluzione ottenuta a partire da P, ma con la tangente valutata nel punto R corrispondente alla soluzione esatta. Il punto R è prossimo al punto medio del segmento  $QQ_1$ .

queste due stime:

$$x(t_0 + h) \simeq x(t_0) + \frac{1}{2} (f(x_0, t_0) + f(x(t_0 + h), t_0 + h)) h. \quad (27)$$

Il problema è che non conosciamo  $x(t_0 + h)$ ! Tuttavia possiamo utilizzare il metodo di Eulero per avere una prima stima di  $x(t_0 + h)$ , per poi utilizzare questa stima nell'equazione (27). Otteniamo così il **metodo di Heun**:

$$x_1 = x_0 + f(x_0, t_0)h; \quad (28)$$

$$x(t_0 + h) \simeq x(t_0) + \frac{h}{2} [f(x_0, t_0) + f(x_1, t_0 + h)]. \quad (29)$$

Proviamo a valutare la soluzione di  $x'(t) = x(t)$  con questo metodo. Abbiamo

$$x_1 = (1 + h) x(t_0); \quad (30)$$

$$x(t_0 + h) \simeq x(t_0) \left[ 1 + \frac{1}{2} (1 + 1 + h) h \right] = x(t_0) \left( 1 + h + \frac{h^2}{2} \right). \quad (31)$$

Otteniamo così i primi tre termini corretti dello sviluppo di  $x(t_0 + h)$ . In figura 5 riporto il confronto fra i risultati del metodo di Eulero e quello di Heun per l'equazione (12). Notiamo che spesso il costo computazionale più grande si ha per la valutazione della funzione  $f(x, t)$  a secondo membro dell'equazione differenziale. Ora, il metodo di Eulero richiede una valutazione ad ogni passo, mentre quello di Heun ne richiede due. È quindi corretto confrontare i risultati dei due metodi a parità del numero di valutazioni di  $f(x, t)$ , e quindi con un numero di punti doppio nel metodo di Eulero che in quello di Heun. Vediamo che su questa scala Heun con 8 punti dà risultati praticamente indistinguibili dalla soluzione esatta, mentre Eulero con 16 punti se ne discosta visibilmente.

Vediamo che cosa succede per l'equazione (18). Otteniamo

$$X(t + h) = GX(t), \quad (32)$$

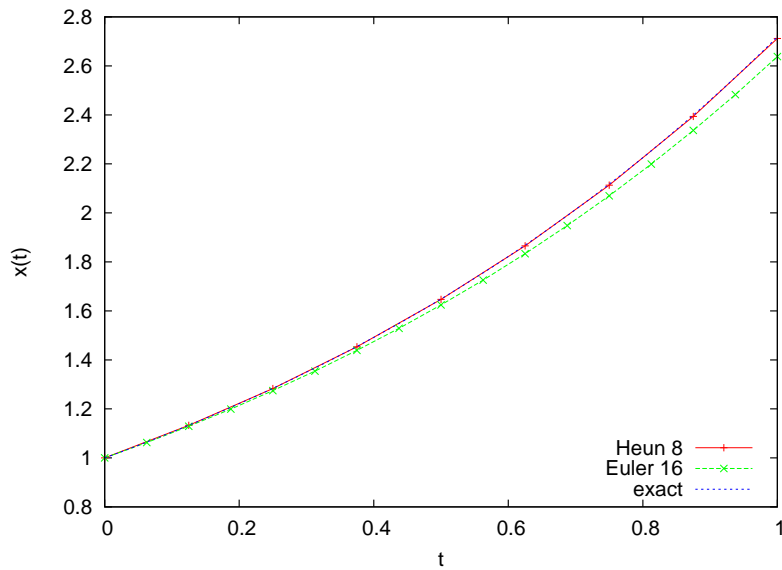
dove

$$G = \begin{pmatrix} 1 - h^2/2, & -h \\ h, & 1 - h^2/2 \end{pmatrix}. \quad (33)$$

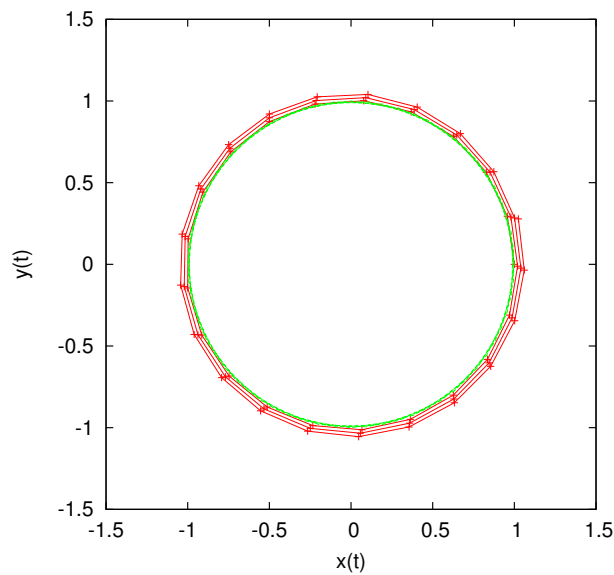
Si ha così

$$x^2(t + h) + y^2(t + h) = (1 + h^4/4) (x^2(t) + y^2(t)). \quad (34)$$

La conservazione di  $x^2 + y^2$  è molto meglio approssimata, ma rimane comunque vero che gli errori si accumulano, e che la soluzione numerica sarà una spirale logaritmica, piuttosto che un'orbita chiusa.



**Figura 5.** Confronto fra il metodo di Heun e quello di Eulero nella soluzione dell'equazione (12), con condizione iniziale  $x(0) = 1$ . Sono riportate le soluzioni del metodo di Eulero con 16 punti, quelle con il metodo di Heun con 8 punti (che richiede 16 valutazioni della funzione  $f(x, t)$ ), e la soluzione esatta.



**Figura 6.** Soluzione dell'equazione differenziale (18) con il metodo di Heun, per  $0 \leq t \leq 6\pi$  con condizione iniziale  $x(0) = 1$ ,  $y(0) = 0$ , e per  $N = 64$ . Viene mostrata anche la soluzione esatta.

### 3. Metodo del punto di mezzo implicito

Supponiamo di nuovo di conoscere esattamente la soluzione  $x(t)$  dell'equazione differenziale (1). Per il teorema della media, si ha, come abbiamo già visto,

$$x(t_0 + h) = x(t_0) + hf(x(\bar{t}), \bar{t}), \quad (35)$$

dove  $t_0 < \bar{t} < t_0 + h$ . Questa relazione suggerisce un'altra possibile approssimazione di  $x(t_0 + h)$ : con ogni probabilità il punto di mezzo dell'intervallo  $[t_0, t_0 + h]$  è più vicino a  $\bar{t}$  di ciascuno degli estremi.

Possiamo porre così

$$\tilde{x}(t_0 + h) = x(t_0) + hf(\bar{x}, \bar{t}), \quad (36)$$

dove  $\bar{x} = x(t_0 + h/2)$ . Il problema è, evidentemente, che non conosciamo  $x(t_0 + h/2)$ . Nel caso del metodo di Heun, avevamo risolto questo problema stimando  $x(t_0 + h)$  con il metodo di Eulero, e poi valutando l'incremento medio di  $x(t)$  nell'intervallo  $[t_0, t_0 + h]$  prendendo la media degli incrementi stimati all'inizio e alla fine dell'intervallo. Un'altra possibilità è invece di porre

$$\bar{x} = \frac{x(t_0 + h) + x(t_0)}{2}, \quad (37)$$

cioè pari alla media dei valori di  $x(t)$  agli estremi dell'intervallo. Otteniamo così la stima

$$\tilde{x}(t_0 + h) = x(t_0) + hf((\tilde{x}(t_0 + h) + x(t_0))/2, t_0 + h/2). \quad (38)$$

Questa espressione deve essere considerata come un'equazione in  $\tilde{x}(t_0 + h)$ . A differenza dei metodi che abbiamo visto finora, che fornivano un'espressione esplicita per la stima di  $x(t_0 + h)$ , questo metodo fornisce un'equazione (in generale non lineare) che deve essere risolta per ottenere la soluzione. Si parla quindi di **metodi impliciti**.

Qual è il vantaggio di questo metodo? Riscriviamo (38) nella forma

$$\tilde{x}(t_0 + h) - x(t_0) = hf((\tilde{x}(t_0 + h) + x(t_0))/2, t_0 + h/2). \quad (39)$$

Vediamo che in questa equazione  $x(t_0)$  e  $\tilde{x}(t_0 + h)$  hanno un ruolo perfettamente simmetrico: possiamo considerarla come un'equazione in  $\tilde{x}(t_0 + h)$ , con  $x(t_0)$  noto, oppure altrettanto bene come un'equazione in  $x(t_0)$ , se  $\tilde{x}(t_0 + h)$  è noto. Questa simmetria è un grande vantaggio se si ha a che fare con equazioni invarianti rispetto alla trasformazione  $t \rightarrow -t$ , come sono le equazioni della meccanica delle particelle.

Chiaramente il metodo richiede di avere a disposizione un algoritmo rapido e poco costoso per risolvere l'equazione (38). Dato che  $h$  è piccolo, non è difficile trovarlo. Consideriamo per esempio la successione definita da

$$x_0 = x(t_0); \quad x_{n+1} = x_0 + hf((x_n + x_0)/2, t_0 + h/2); \quad n = 1, 2, \dots \quad (40)$$

Allora questa successione tende (per  $h$  abbastanza piccolo!) a  $x^*$ , che soddisfa l'equazione

$$x^* = x_0 + hf((x^* + x_0)/2, t_0 + h/2). \quad (41)$$

Supponiamo infatti che, p.es.,  $x_n = x^* + \delta x$ . Si ha allora

$$x_{n+1} = x_0 + hf((x_n + x_0)/2, t_0 + h/2) \simeq x^* + h f'(x^*) \delta x. \quad (42)$$

Quindi, se  $|h f'(x^*)| < 1$ , si ha  $|x_{n+1} - x^*| < |\delta x| = |x_n - x^*|$ , e la successione  $(x_n)$  converge a  $x^*$ . Questo è un metodo abbastanza facile da programmare per risolvere l'equazione, non necessariamente il più efficiente.

Vediamo il comportamento di questo metodo con le nostre equazioni differenziali. Per l'equazione (12) si ha l'equazione

$$\tilde{x}(t_0 + h) = \tilde{x} = x_0 + h \left( \frac{x_0 + \tilde{x}}{2} \right). \quad (43)$$

Questa è un'equazione lineare che ha la soluzione

$$\tilde{x} = x_0 \frac{1 + h/2}{1 - h/2}. \quad (44)$$

La precisione di questa soluzione è comparabile con quella del metodo di Heun.

È più interessante considerare la soluzione dell'equazione (18). Posto

$$\mathbf{X} = \begin{pmatrix} x \\ y \end{pmatrix}; \quad \|\mathbf{X}\|^2 = x^2 + y^2, \quad (45)$$

e introdotte le notazioni  $\mathbf{X}_0 = \mathbf{X}(t_0)$  e  $\tilde{\mathbf{X}} = \mathbf{X}(t_0 + h)$ , si ha l'equazione

$$\mathbf{A}\tilde{\mathbf{X}} = \mathbf{B}\mathbf{X}_0, \quad (46)$$

dove

$$\mathbf{A} = \begin{pmatrix} 1, & -h/2 \\ h/2, & 1 \end{pmatrix}, \quad (47)$$

$$\mathbf{B} = \begin{pmatrix} 1, & h/2 \\ -h/2, & 1 \end{pmatrix}. \quad (48)$$

Otteniamo così

$$\tilde{\mathbf{X}} = \mathbf{A}^{-1} \mathbf{B} \mathbf{X}_0 = \mathbf{T} \mathbf{X}_0, \quad (49)$$

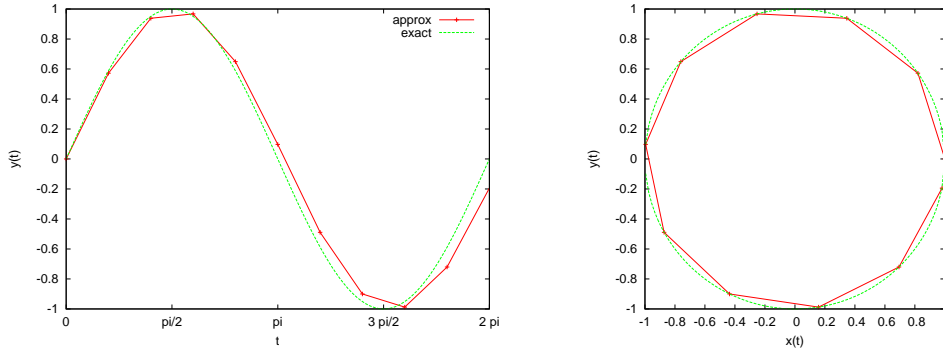
dove

$$\mathbf{T} = \frac{1}{1 + h^2/4} \begin{pmatrix} 1 - h^2/4, & h \\ -h, & 1 - h^2/4 \end{pmatrix}. \quad (50)$$

È facile verificare che si ha allora

$$\|\tilde{\mathbf{X}}\|^2 = \|\mathbf{X}_0\|^2. \quad (51)$$

Quindi la traiettoria della soluzione approssimata non si discosterà da quella della traiettoria esatta. Ovviamente questo non vuol dire che la soluzione sia esatta, come si può vedere, per esempio, dalla figura 7. Tuttavia l'errore sta più in un apprezzabile ritardo della soluzione approssimata da quella esatta, piuttosto che in un errore sulla traiettoria.



**Figura 7.** Metodo del punto di mezzo implicito nella soluzione dell'equazione (18). A sinistra è riportata la soluzione  $y(t)$  in funzione di  $t$  per  $0 \leq t \leq 2\pi$ , con  $N = 10$  punti, insieme con la soluzione esatta  $\sin t$ . A destra la corrispondente traiettoria nel piano  $(x, y)$ .

### Il metodo del punto di mezzo implicito come integratore symplettico

Questa proprietà del metodo del punto di sella implicito è una conseguenza di una proprietà più profonda. Consideriamo un'equazione differenziale di forma canonica per la coppia di variabili  $(x, p)$ :

$$\frac{dx}{dt} = \frac{\partial H}{\partial p}; \quad (52)$$

$$\frac{dp}{dt} = -\frac{\partial H}{\partial x}. \quad (53)$$

In questa equazione,  $H(x, p)$  è la hamiltoniana. Notiamo che l'equazione (18) è di questa forma, se si pone  $p = y$  e  $H = (p^2 + y^2)/2$ . Allora  $H(x, p)$  è conservata, nel senso che, se  $(x(t), p(t))$  è una soluzione delle (52,53) che soddisfa  $(x(t_0), p(t_0)) = (x_0, p_0)$ , si ha

$$H(x(t), p(t)) = H(x_0, p_0) = \text{const.} \quad (54)$$

Si ha in effetti

$$\frac{dH}{dt} = \frac{\partial H}{\partial x} \frac{dx}{dt} + \frac{\partial H}{\partial p} \frac{dp}{dt} = \frac{\partial H}{\partial x} \frac{\partial H}{\partial p} - \frac{\partial H}{\partial p} \frac{\partial H}{\partial x} = 0. \quad (55)$$

Valutiamo ora la variazione di  $H$  per la soluzione differenziale con il metodo del punto di mezzo implicito. Indichiamo con  $\mathbf{X} = (x, p)$  il generico punto del piano  $x, p$ , con  $\mathbf{X}_0 = (x_0, p_0)$  la soluzione all'istante  $t_0$ , e con  $\tilde{\mathbf{X}} = \mathbf{X}_0 + \delta\tilde{\mathbf{X}}$  la stima di  $\mathbf{X}(t_0 + h)$  ottenuta mediante il metodo del punto di mezzo implicito:

$$\tilde{\mathbf{X}} = (x_0 + \delta\tilde{x}, p_0 + \delta\tilde{p}), \quad (56)$$

dove  $(\delta\tilde{x}, \delta\tilde{p})$  sono soluzioni del sistema d'equazioni

$$\delta\tilde{x} = hH_x(x_0 + \delta\tilde{x}/2, p_0 + \delta\tilde{p}/2); \quad (57)$$

$$\delta\tilde{p} = -hH_p(x_0 + \delta\tilde{x}/2, p_0 + \delta\tilde{p}/2) \quad (58)$$



In questa equazione abbiamo introdotto la notazione  $H_x = \partial H / \partial x$ , ecc. Indichiamo con  $\bar{X} = (\bar{x}, \bar{p})$  il punto di mezzo implicito. Si ha allora

$$H(\tilde{X}) = H(\bar{X}) + H_p \delta \tilde{x} / 2 + H_x \delta \tilde{p} / 2 + O(\delta \tilde{X}^2); \quad (59)$$

$$H(X_0) = H(\bar{X}) - H_p \delta \tilde{x} / 2 - H_x \delta \tilde{p} / 2 + O(\delta \tilde{X}^2), \quad (60)$$

dove le derivate sono valutate in  $\bar{X}$ . Sottraendo membro a membro, e tenendo conto del fatto che i termini di secondo ordine in  $\delta \tilde{X}$  si cancellano, otteniamo

$$\begin{aligned} \Delta H &= H(\tilde{X}) - H(X_0) = H_p \delta \tilde{x} / 2 + H_x \delta \tilde{p} / 2 + O(\delta \tilde{X}^3) \\ &= h(H_x H_p - H_p H_x) + O(\delta \tilde{X}^3) = O(\delta \tilde{X}^3). \end{aligned} \quad (61)$$

La conservazione di  $H$  viene quindi verificata a meno di termini di ordine  $\delta \tilde{X}^3$ , e quindi proporzionali a  $h^3$ .

Per sistemi che retti da equazioni differenziali di forma canonica vale un'altra legge di conservazione più sottile. Supponiamo che all'istante  $t = t_0$  il sistema si trovi in  $X_0 = (x_0, p_0)$ . Consideriamo delle piccole perturbazioni dello stato iniziale,  $\delta X_1 = (\delta x_0, 0)$  e  $\delta X_2 = (0, \delta p_0)$ . Nello spazio  $(x, p)$  queste perturbazioni individuano un parallelogramma di area  $A = \delta X_1 \times \delta X_2 = \delta_0 \delta p_0$ , dove "×" indica il prodotto vettoriale. Seguiamo adesso la soluzione dell'equazione differenziale ad un istante generico  $t$ . Indichiamo con  $X(t)$  la soluzione che soddisfa la condizione iniziale  $X(t_0) = X_0$ , e rispettivamente con  $X_1(t)$  e  $X_2(t)$  quelle che soddisfano le condizioni iniziali  $X(t_0) = X_0 + \delta X_1$  e  $X(t_0) = X_0 + \delta X_2$ . Allora i vettori  $\delta X_i(t) = X_i(t) - X(t)$ , per  $i = 1, 2$ , identificano un parallelogramma di area  $A(t) = \delta X_1(t) \times \delta X_2(t)$ . Ora, si ha

$$A(t) = A, \quad \forall t. \quad (62)$$

In effetti, posto

$$\delta X_i(t) = (\delta x_i(t), \delta p_i(t)), \quad (63)$$

si ha, per  $i = 1, 2$ ,

$$\frac{d\delta x_i}{dt} = H_{px} \delta x_i + H_{pp} \delta p_i; \quad (64)$$

$$\frac{d\delta p_i}{dt} = -H_{xx} \delta x_i - H_{xp} \delta p_i. \quad (65)$$

In questa espressione, le derivate di  $H$  sono valutate in  $X(t)$ . D'altra parte si ha

$$\begin{aligned} \frac{dA}{dt} &= \frac{d}{dt} \det \begin{pmatrix} \delta x_1(t) & \delta p_1(t) \\ \delta x_2(t) & \delta p_2(t) \end{pmatrix} = \delta \dot{x}_1 \delta p_2 + \delta x_1 \delta \dot{p}_2 - (\delta \dot{x}_2 \delta p_1 + \delta x_2 \delta \dot{p}_1) \\ &= H_{px} \delta x_1 \delta p_2 + H_{pp} \delta p_1 \delta p_2 - H_{xx} \delta x_1 \delta x_2 - H_{xp} \delta x_1 \delta x_2 \\ &\quad - (H_{px} \delta x_2 \delta p_1 + H_{pp} \delta p_2 \delta p_1 - H_{xx} \delta x_2 \delta x_1 - H_{xp} \delta x_2 \delta p_1) = 0. \end{aligned} \quad (66)$$

È immediato generalizzare questo risultato a un sistema di  $2N$  equazioni differenziali, con  $N$  coppie di variabili coniugate  $(x_i, p_i)$  ( $i = 1, \dots, N$ ), rispetto alle quali le equazioni hanno la forma canonica (52,53). Per ciascuna di esse, l'area  $A_i(t)$  del parallelogramma costruito in maniera analoga si conserva. Come corollario, se consideriamo  $2N$  piccole perturbazioni  $\delta X_i$  ( $i = 1, \dots, N$ ), che racchiudono una piccola regione dello spazio  $(x_1, p_1, \dots, x_N, p_N)$ , e seguiamo l'evolversi di questa regione nel tempo, il volume corrispondente rimane costante. In meccanica, questo risultato è noto come **teorema di Liouville**.

Mostriamo adesso che il metodo del punto di mezzo implicito soddisfa la conservazione di  $A$ , beninteso se i  $\delta X_i$  sono abbastanza piccoli. Dato  $X_0 = X(t_0)$  e l'incremento  $h$  di  $t$ , l'approssimazione  $\tilde{X} = (\tilde{x}, \tilde{p})$  di  $X(t_0 + h)$  è la soluzione delle equazioni

$$\tilde{x} = x_0 + h H_p \left( \frac{x_0 + \tilde{x}}{2}, \frac{p_0 + \tilde{p}}{2} \right); \quad (67)$$

$$\tilde{p} = p_0 - h H_x \left( \frac{x_0 + \tilde{x}}{2}, \frac{p_0 + \tilde{p}}{2} \right). \quad (68)$$

Dato un incremento  $\delta\mathbf{X}_0 = (\delta x_0, \delta p_0)$  della condizione iniziale a  $t = 0$ , il corrispondente incremento  $\delta\tilde{\mathbf{X}}$  di  $\tilde{\mathbf{X}}$  è la soluzione delle equazioni

$$\delta\tilde{x} = \delta x_0 + \frac{h}{2} [H_{px}(\delta x_0 + \delta\tilde{x}) + H_{pp}(\delta p_0 + \delta\tilde{p})]; \quad (69)$$

$$\delta\tilde{p} = \delta p_0 - \frac{h}{2} [H_{xx}(\delta x_0 + \delta\tilde{x}) + H_{xp}(\delta p_0 + \delta\tilde{p})]. \quad (70)$$

Queste equazioni possono essere scritte nella forma seguente:

$$\mathbf{A}\delta\tilde{\mathbf{X}} = \mathbf{B}\delta\mathbf{X}_0, \quad (71)$$

dove le matrici  $\mathbf{A}$  e  $\mathbf{B}$  sono definite da

$$\mathbf{A} = \begin{pmatrix} \frac{h}{2}H_{px}, & -\frac{h}{2}H_{pp} \\ \frac{h}{2}H_{xx}, & 1 + \frac{h}{2}H_{xp} \end{pmatrix}; \quad \mathbf{B} = \begin{pmatrix} 1 + \frac{h}{2}H_{px}, & \frac{h}{2}H_{pp} \\ -\frac{h}{2}H_{xx}, & 1 - \frac{h}{2}H_{xp} \end{pmatrix}. \quad (72)$$

Esse ammettono quindi la soluzione

$$\delta\tilde{\mathbf{X}} = \mathbf{A}^{-1}\mathbf{B}\delta\mathbf{X}_0 = \mathbf{T}\delta\mathbf{X}_0, \quad (73)$$

dove, posto

$$\mathcal{H} = \det \begin{pmatrix} H_{xx}, & H_{xp} \\ H_{px}, & H_{pp} \end{pmatrix}, \quad (74)$$

si ha (tenendo conto dell'uguaglianza delle derivate miste)

$$\mathbf{T} = \frac{1}{|\mathbf{A}|} \begin{pmatrix} 1 + hH_{xp} - (h/2)^2\mathcal{H}, & -hH_{pp} \\ hH_{xx}, & 1 - hH_{px} - (h/2)^2\mathcal{H} \end{pmatrix}; \quad (75)$$

$$|\mathbf{A}| = 1 - (h/2)^2\mathcal{H}. \quad (76)$$

Consideriamo i due vettori iniziali  $\delta\mathbf{X}_1 = (\delta x_0, 0)$  e  $\delta\mathbf{X}_2 = (0, \delta p_0)$ . I corrispondenti incrementi  $\delta\tilde{\mathbf{X}}_1$ ,  $\delta\tilde{\mathbf{X}}_2$ , sono dati da

$$\delta\tilde{\mathbf{X}}_1 = \frac{1}{|\mathbf{A}|} \begin{pmatrix} (1 + hH_{xp} - (h/2)^2\mathcal{H}) \delta x_0 \\ hH_{xx}, \delta x_0 \end{pmatrix}; \quad (77)$$

$$\delta\tilde{\mathbf{X}}_2 = \frac{1}{|\mathbf{A}|} \begin{pmatrix} -hH_{pp} \delta p_0 \\ (1 - hH_{px} - (h/2)^2\mathcal{H}) \delta p_0 \end{pmatrix}. \quad (78)$$

Con questi vettori iniziali si ha  $A(t_0) = \delta x_0 \delta p_0$ . L'area fra i vettori  $\delta\tilde{\mathbf{X}}_1$  e  $\delta\tilde{\mathbf{X}}_2$  è pari al prodotto vettoriale dei due vettori, che è uguale al determinante della matrice che si ottiene mettendoli fianco a fianco. Un facile calcolo mostra che si ha

$$\delta\tilde{\mathbf{X}}_1 \times \delta\tilde{\mathbf{X}}_2 = |\mathbf{T}| \delta x_0 \delta p_0, \quad (79)$$

dove  $|\mathbf{T}| = \det \mathbf{T}$ . È facile inoltre verificare che

$$|\mathbf{T}| = 1. \quad (80)$$

Quindi il metodo del punto di mezzo implicito conserva le aree del tipo  $\delta\tilde{\mathbf{X}}_1 \times \delta\tilde{\mathbf{X}}_2$ . Queste forme sono dette **forme simplettiche**, e per questa ragione si parla di **proprietà simplettica** del punto di mezzo implicito. Questo metodo è quindi particolarmente adatto per l'integrazione delle equazioni differenziali di forma canonica. Nel caso particolare della dinamica delle particelle esistono però altri metodi più semplici (espliciti) che possiedono questa proprietà.

#### 4. Metodi di Runge-Kutta

Il problema del metodo del punto di mezzo implicito è che la soluzione dell'equazione per  $\bar{t} = x(t_0 + h/2)$  richiede un numero di chiamate alla funzione  $f(x, t)$  che non è prevedibile. Ci si può domandare che cosa succede se ci si limita a una sola chiamata per stimare  $\bar{x}$ , ponendo cioè

$$\bar{x} = x_0 + \frac{h}{2} f(x_0, t_0), \quad (81)$$

per poi calcolare la stima  $\tilde{x}$  mediante la (36). Otteniamo così un algoritmo che può essere scritto nella forma seguente:

$$k_1 = f(x_0, t_0); \quad (82)$$

$$k_2 = f(x_0 + (h/2)k_1, t_0 + h/2); \quad (83)$$

$$\tilde{x} = x_0 + hk_2. \quad (84)$$

Questo algoritmo è noto come **metodo del punto di mezzo esplicito**, o anche come **metodo di Runge-Kutta del secondo ordine**, o RK2. Si può mostrare che questo metodo è soggetto ad errori più grandi di quello di Heun, per uno sforzo computazionale (numero di chiamate alla funzione  $f(x, t)$ ) sostanzialmente equivalente. Tuttavia esso ci fornisce il pretesto per una discussione che ci permetterà di definire una classe di metodi d'integrazione noti come **metodi di Runge-Kutta**, alcuni dei quali sono fra i metodi più utilizzati nelle applicazioni.

Consideriamo la relazione (8). L'integrale a secondo membro può essere stimato mediante una qualunque delle discretizzazioni utilizzate per valutare numericamente gli integrali. Denotiamo  $f(x(t), t)$  con  $F(t)$ . Si ha allora, per esempio:

**La regola dei rettangoli:**  $I = \int_{t_0}^{t_0+h} dt F(t) \simeq h F(t_0 + h/2);$

**La regola dei trapezi:**  $I \simeq \frac{h}{2} (F(t_0) + F(t_0 + h));$

**La regola di Simpson:**  $I \simeq \frac{h}{6} (F(t_0) + 4F(t_0 + h/2) + F(t_0 + h)).$

A queste prescrizioni deve essere associato un metodo per stimare i valori di  $F(t^*) = f(x(t^*), t^*)$ , e quindi di  $x^* = x(t^*)$ , dove  $t^*$  è un valore di  $t$  compreso fra  $t_0$  e  $t_0 + h$ . Questo metodo può essere implicito (richiedere cioè la risoluzione di un'equazione) o esplicito (fornire un'espressione della stima di  $x^*$  in funzione di quantità già note). Da questo punto di vista, il metodo del punto di mezzo implicito corrisponde alla regola dei rettangoli implicita, il metodo RK2 alla regola dei rettangoli esplicita, mentre il metodo di Heun corrisponde alla regola dei trapezi esplicita. (La superiorità del metodo di Heun rispetto alla RK2 corrisponde alla migliore approssimazione che si ottiene dalla regola dei trapezi rispetto a quella dei rettangoli.)

In generale, metodi di questo tipo richiedono la valutazione successiva di quantità della forma

$$k_i = f(x_0 + h \sum_j a_{ij} k_j, t_0 + c_i h), \quad i = 1, \dots, s, \quad (85)$$

per poi produrre una stima di  $x(t_0 + h)$  mediante un'espressione del tipo

$$\tilde{x}(t_0 + h) = x_0 + h \sum_i b_i k_i. \quad (86)$$

Queste espressioni vengono riassunte convenientemente in una tabella, che è chiamata **tabella di Butcher**:

$$\begin{array}{c|cccc} c_1 & a_{11} & a_{12} & \cdots & a_{1s} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_s & a_{s1} & a_{s2} & \cdots & a_{ss} \\ \hline & b_1 & b_2 & \cdots & b_s \end{array} \quad (87)$$

In questo schema, abbiamo per esempio

- Per il metodo di Eulero:

$$\begin{array}{c|c} 0 & 0 \\ \hline & 1 \end{array} \quad (88)$$

- Per il metodo di Heun:

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1 & 0 \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array} \quad (89)$$

- Per il metodo RK2:

$$\begin{array}{c|cc} 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \hline & 0 & 1 \end{array} \quad (90)$$

- Per il metodo del punto di mezzo implicito:

$$\begin{array}{c|c} \frac{1}{2} & \frac{1}{2} \\ \hline & 1 \end{array} \quad (91)$$

Notiamo che nei metodi espliciti, gli elementi  $a_{ij}$  con  $j \geq i$  sono nulli. Se si conviene di trattare solo metodi espliciti, la tabella di Butcher può essere semplificata nella forma

$$\begin{array}{c|cccc} 0 & & & & \\ c_2 & a_{21} & & & \\ c_3 & a_{31} & a_{32} & & \\ \vdots & \vdots & \vdots & \ddots & \\ c_s & a_{s1} & a_{s2} & \cdots & a_{s,s-1} \\ \hline & b_1 & b_2 & \cdots & b_{s-1} & b_s \end{array} \quad (92)$$

Notiamo che dato che le  $k_i$  possono essere interpretate come la stima della pendenza della soluzione  $x(t)$  all'istante  $t_0 + c_i h$ , si richiede per coerenza che siano soddisfatte le relazioni

$$c_i = \sum_j a_{ij}, \quad (93)$$

per  $i = 1, \dots, s$ , e che si abbia inoltre

$$\sum_i b_i = 1. \quad (94)$$

Il metodo di Runge-Kutta più utilizzato è noto come RK4. Esso è un metodo esplicito, descritto dalla seguente tabella di Butcher semplificata:

$$\begin{array}{c|cccc} 0 & & & & \\ \frac{1}{2} & \frac{1}{2} & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & \\ 1 & 0 & 0 & 1 & \\ \hline & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array} \quad (95)$$

In pratica questo corrisponde alla seguente sequenza di valutazioni:

$$k_1 = f(x_0, t_0); \quad (96)$$

$$k_2 = f(x_0 + (h/2)k_1, t_0 + h/2); \quad (97)$$

$$k_3 = f(x_0 + (h/2)k_2, t_0 + h/2); \quad (98)$$

$$k_4 = f(x_0 + hk_3, t_0 + h); \quad (99)$$

$$\tilde{x}(t_0 + h) = x_0 + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4). \quad (100)$$

Si vede che il metodo corrisponde alla regola di Simpson, ma in cui si fanno due valutazioni successive per il valore di  $F(t_0 + h/2)$ , e l'ultima viene utilizzata per stimare  $F(t_0 + h)$ . L'errore su  $x(t + h)$  ad ogni passo è di ordine  $h^5$ , per cui l'errore accumulato su un intervallo finito è di ordine  $h^4$ . Per questo il metodo è detto del "quarto ordine", e viene indicato come RK4.

#### Errore di troncamento nei metodi di Runge-Kutta

Per stimare l'errore di troncamento di questi metodi, si può usare il metodo seguente. Consideriamo l'equazione differenziale

$$\frac{dx}{dt} = f(x, t). \quad (101)$$

Supponendo di voler conoscere la soluzione  $x(t)$  nel punto  $t = t_0 + h$ , scriviamo lo sviluppo di Taylor generico di  $x(t_0 + h)$  in potenze di  $h$ :

$$x(t_0 + h) = x_0 + \frac{h}{2}x_1 + \frac{h^2}{2}x_2 + \frac{h^3}{6}x_3 + O(h^4). \quad (102)$$

Sviluppiamo quindi il secondo membro dell'equazione (101) in potenze di  $h$ , tenendo conto dello sviluppo di  $x(t_0 + h)$ :

$$f(x(t_0 + h), t_0 + h) = f_0 + h(f_x x_1 + f_t) + \frac{h^2}{2}(f_{xx}(x_1^2 + x_2) + 2f_{xt}x_1 + f_{tt}) + O(h^3), \quad (103)$$

dove  $f_0 = f(x_0, t_0)$ ,  $f_x = \partial f / \partial x|_{x_0, t_0}$ , ecc. Integrando questa espressione fra  $t_0$  e  $h$ , e aggiungendo  $x(t_0) = x_0$  otteniamo lo sviluppo della soluzione della (101):

$$\begin{aligned} x(t_0 + h) &= x_0 + \int_0^h dh' f(x(t_0 + h'), h') \\ &= x_0 + hf_0 + \frac{h^2}{2}(x_1 f_x + f_t) + \frac{h^3}{6}((x_2 + x_1^2)f_{xx} + 2x_1 f_{xt} + f_{tt}) + O(h^4). \end{aligned} \quad (104)$$

Confrontando la (104) con la (102) possiamo ottenere le espressioni dei coefficienti di Taylor  $x_i$ :

$$x_1 = f_0; \quad (105)$$

$$x_2 = f_0 f_x + f_t; \quad (106)$$

$$x_3 = f_{xx} f_0^2 + 2f_{xt} f_0 + f_{tt} + f_x(f_t + f_0 f_x). \quad (107)$$

Valutiamo adesso un'approssimazione di tipo RK2:

$$k_1 = f(x_0, t_0); \quad (108)$$

$$k_2 = f(x_0 + ak_1, t_0 + ah); \quad (109)$$

$$\tilde{x}(t_0 + h) = x_0 + (1 - b)k_1 + bk_2. \quad (110)$$

Possiamo ora valutare lo sviluppo in serie di Taylor di  $h$  della stima  $\tilde{x}(t_0 + h)$ :

$$\begin{aligned} \tilde{x}(t_0 + h) &= x_0 + hf_0 + h^2 ab(f_x f_0 + f_t) \\ &\quad + \frac{h^3}{2} a^2 b(f_{xx} f_0^2 + 2f_{xt} f_0 + f_{tt}) + O(h^4). \end{aligned} \quad (111)$$

Vediamo che perché i due sviluppi coincidano fino all'ordine  $h^2$  si deve avere  $ab = \frac{1}{2}$ . Resta quindi la sola libertà di scegliere  $a$ . Possiamo valutare allora la differenza fra (104) e (111):

$$x(t_0 + h) - \tilde{x}(t_0 + h) = \frac{h^3}{12} [2f_t f_x + 2f_0 f_x^2 + (2 - 3a)(f_{tt} + 2f_{xt} f_0 + f_{xx} f_0^2)] + O(h^4). \quad (112)$$

Non possiamo annullare l'errore di ordine  $h^3$ . Tuttavia, scegliendo  $a = 2/3$ , possiamo semplificarlo:

$$x(t_0 + h) - \tilde{x}(t_0 + h) = \frac{h^3}{6} f_x(f_t + f_0 f_x) + O(h^4). \quad (113)$$

Conclusione: lo schema RK2, con  $b = 1/(2a)$ , ha un errore di troncamento proporzionale a  $h^3$  ad ogni passo, e quindi un errore totale di ordine  $h^2$ . In un certo senso, lo schema ottimale si ha per  $a = 2/3$ .

Il calcolo corrispondente per RK4 è analogo, ma estremamente tedioso. Si può trovare nel seguente notebook di *Mathematica*: rk4.nb. Il calcolo conferma che l'errore corrispondente è di ordine  $h^5$  per ogni passo, e quindi  $h^4$  globalmente. Notiamo che però l'assenza della proprietà simplettica per questo metodo non lo rende particolarmente adatto all'integrazione delle equazioni di tipo canonico. D'altra parte, esso è un metodo adatto all'integrazione di equazioni del moto di sistemi dissipativi.

## 5. Metodi di Verlet

Per l'integrazione delle equazioni del moto di un sistema di particelle che obbediscono alla legge di Newton (con forze che dipendono solo dalla posizione reciproca delle particelle) è conveniente usare i metodi di Verlet. Essi sono metodi del terzo ordine, che possiedono la proprietà simplettica e che sono molto semplici da programmare. Supponiamo di volere risolvere l'equazione differenziale

$$\frac{d^2 x}{dt^2} = f(x, t). \quad (114)$$

Consideriamo lo sviluppo di Taylor della soluzione  $x(t)$  attorno a  $t_0$ , dove  $x(t_0) = x_0$ :

$$x(t_0 + h) = x_0 + hx_1 + \frac{h^2}{2}x_2 + \frac{h^3}{3!}x_3 + O(h^4). \quad (115)$$

Valutando questo sviluppo in  $-h$ , otteniamo

$$x(t_0 - h) = x_0 - hx_1 + \frac{h^2}{2}x_2 - \frac{h^3}{3!}x_3 + O(h^4). \quad (116)$$

Sommando membro a membro otteniamo

$$x(t_0 + h) + x(t_0 - h) = 2x_0 + h^2x_2 + O(h^4). \quad (117)$$

D'altra parte, poichè  $x(t)$  è soluzione della (114), si ha

$$x_2 = \left. \frac{d^2x}{dt^2} \right|_{t=t_0} = f(x_0, t_0). \quad (118)$$

Si ha quindi

$$x(t_0 + h) = 2x(t_0) - x(t_0 - h) + h^2f(x_0, t_0) + O(h^4) = \tilde{x}(t_0 + h) + O(h^4). \quad (119)$$

Questa relazione definisce il metodo di Verlet.

Osserviamo che il metodo richiede la conoscenza ad ogni passo del valore  $x(t_0)$  della soluzione all'istante considerato e del suo valore  $x(t_0 - h)$  al *passo precedente*. D'altra parte, le condizioni iniziali dell'equazione (114) sono di solito note nella forma  $(x(t_0), v(t_0))$ , dove  $v_0 = dx/dt|_{t=t_0}$ . Quindi per potere iniziare a valutare la soluzione  $x(t)$  con il metodo di Verlet, è necessario utilizzare un altro metodo per ottenere una stima di  $x(t_0 - h)$ . In questo senso, il metodo di Verlet non è **autosufficiente**. D'altra parte, dato che l'uso di questo altro metodo è ristretto ad un solo passo, si può utilizzare anche un metodo abbastanza dispendioso in termini di calcolo, come il metodo RK4. Ottenuta la stima di  $x(t_0 - h)$ , il metodo di Verlet permette la valutazione della soluzione  $x(t)$  fino all'ordine  $h^3$  (globale) mediante *una sola* valutazione di  $f(x, t)$  a ogni passo.

Notiamo anche che la relazione (117) è perfettamente simmetrica fra  $x(t_0 + h)$  e  $x(t_0 - h)$ . Quindi il metodo di Verlet possiede, come il metodo del punto di mezzo implicito, la proprietà di essere invariante rispetto alla trasformazione  $t \rightarrow -t$ . In altri termini, se valutiamo la soluzione  $x(t)$  per  $t_0 \leq t \leq t_0 + T$ , possiamo ripercorrerla all'indietro risolvendo l'equazione con il metodo di Verlet per incrementi  $h$  negativi, a partire dalla condizione iniziale  $x(t_0 + T + h), x(t_0 + T)$ : questo, beninteso, a meno degli effetti degli errori di arrotondamento, legati al fatto che i numeri reali sono rappresentati con precisione finita entro il calcolatore.

È noto che per forze conservative l'energia meccanica si conserva. In pratica, se  $x = (x_i)$ , ( $i = 1, \dots, N$ ), e l'equazione (114) ha la forma

$$\frac{d^2x_i}{dt^2} = -\frac{1}{m_i} \frac{\partial U}{\partial x_i}, \quad (120)$$

dove  $U(x) = U(x_1, \dots, x_N)$  è una certa funzione, si ha

$$E = \sum_{i=1}^N \frac{1}{2} m_i \left( \frac{dx_i}{dt} \right)^2 + U(x) = \text{const.} \quad (121)$$

Per valutare questa quantità, abbiamo bisogno di una stima della velocità  $v = (dx_i/dt)$ . Sottraendo membro a membro le (118) e (119), si ha

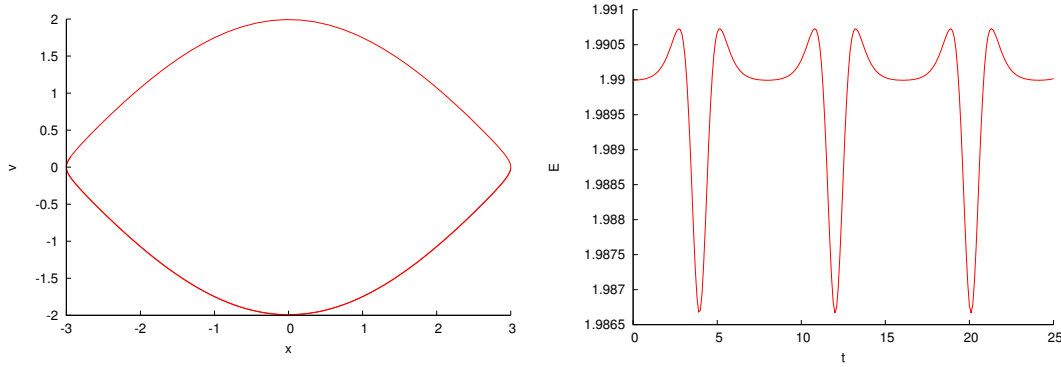
$$v = \left. \frac{dx}{dt} \right|_{t=t_0} = \frac{x(t_0 + h) - x(t_0 - h)}{2h} + O(h^2). \quad (122)$$

Notiamo inoltre che in questa espressione appare la differenza di quantità molto vicine, per cui ci saranno degli errori di arrotondamento importanti. Comunque, è possibile verificare che  $E$  rimane costante a meno di  $h^2$  a ogni passo, e quindi  $h$  globalmente. In realtà il metodo di Verlet ha un corportamento migliore di questa stima, perchè l'errore viene compiuto nella *stima* di  $v$ , mentre la soluzione approssima quella esatta a meno di  $h^3$  globalmente: il termine di ordine  $h^2$  che compare nella stima di  $v$  *non* contribuisce alla stima di  $x(t_0 + h)$ . In effetti,  $E$  rimane costante a meno di termini di ordine  $h^3$ , anche se la sua stima, istante per istante, fluttua di valori di ordine  $h^2$ . Tuttavia il fatto che la velocità venga stimata così male può suggerire di introdurre uno schema che, mantenendo i vantaggi del metodo di Verlet, produce anche una buona valutazione della velocità ad ogni passo. Questo schema è chiamato **velocity Verlet**.

## Il metodo “velocity Verlet”

Il metodo **velocity Verlet** richiede formalmente due valutazioni di  $f(x, t)$  per passo, ma si vede che ciascuna valutazione può essere utilizzata in due passi successivi, così che l’impegno computazionale è essenzialmente lo stesso che nel metodo di Verlet standard. Esso ha il vantaggio di fornire anche una buona stima della velocità ad ogni passo. Supponiamo di conoscere  $x(t_0) = x_0$  e  $v(t_0) = v_0$ . Stimiamo  $v(t_0 + h/2)$ :

$$v(t_0 + h/2) = v_0 + \frac{h}{2}f(x_0, t_0) + O(h^2). \quad (123)$$



**Figura 8.** Soluzione dell’equazione differenziale del pendolo  $d^2x/dt^2 = -\sin x$  con le condizioni iniziali  $x_0 = 3$ ,  $v_0 = 0$  con il metodo “velocity Verlet”. A sinistra: traiettoria nel piano  $(x, v)$ . A destra: valore dell’energia  $E = v^2/2 + (1 - \cos x)$ . Si vede che l’energia non rimane costante, ma oscilla periodicamente. In effetti la quantità conservata è una *hamiltoniana ombra* la cui differenza dalla hamiltoniana “vera” può essere stimata dal grafico.

Per valutare  $x(t_0 + h)$ , usiamo l’approssimazione

$$x(t_0 + h) = x(t_0) + hv(t_0 + h/2). \quad (124)$$

Per valutare  $v(t_0 + h)$ , usiamo ancora (123):

$$v(t_0 + h) = v(t_0 + h/2) + (h/2)f(x(t_0 + h), t_0 + h). \quad (125)$$

Eliminando  $v(t_0 + h/2)$  mediante l’equazione (123), otteniamo il seguente algoritmo:

$$x(t_0 + h) \simeq x(t_0) + hv(t_0) + (h^2/2)f(x(t_0), t_0); \quad (126)$$

$$v(t_0 + h) \simeq v_0 + \frac{h}{2}(f(x_0, t_0) + f(x(t_0 + h), t_0 + h)). \quad (127)$$

Questo metodo fornisce un aggiornamento della *coppia*  $(x, v)$  a ogni passo. Notare che la seconda valutazione della forza che appare nella seconda equazione può essere usata nel passo successivo per l’aggiornamento di  $x(t)$ .

Si può mostrare (mediante l’usuale e tedioso metodo) che l’errore commesso su  $v(t_0 + h)$  e  $x(t_0 + h)$  è di ordine  $h^3$ . Si può anche mostrare esplicitamente che l’algoritmo ha la proprietà *simplettica*. In effetti, considerando la perturbazione  $\delta\mathbf{X} = (\delta x_0, \delta v_0)$ , otteniamo la seguente espressione per la variazione  $\delta\tilde{\mathbf{X}}$ :

$$\delta\tilde{\mathbf{X}} = \mathbf{T}\delta\mathbf{X}, \quad (128)$$

dove la matrice  $\mathbf{T}$  è data da

$$\mathbf{T} = \begin{pmatrix} 1 + \frac{h^2}{2}f_x, & h \\ \frac{h}{2}\left[f_x + \left(1 + \frac{h^2}{2}\right)f_x(\tilde{x}(t_0 + h), t_0 + h)\right], & 1 + \frac{h^2}{2}f_x(\tilde{x}(t_0 + h), t_0 + h) \end{pmatrix}. \quad (129)$$

Si verifica facilmente che all’ordine valutato si ha

$$|\mathbf{T}| = 1. \quad (130)$$

L'algoritmo "velocity Verlet" è vicino al **metodo leapfrog**, in cui la velocità viene valutata ai tempi  $t_{k+1/2} = t_0 + (k + 1/2)h$ , mentre le posizioni sono valutate ai tempi  $t_k = t_0 + kh$ . Esplicitamente, questo metodo è definito dalle seguenti equazioni:

$$v(t_{k+1/2}) \simeq v(t_{k-1/2}) + hf(x(t_k), t_k); \quad (131)$$

$$x(t_{k+1}) \simeq x(t_k) + hv(t_{k+1/2}). \quad (132)$$

Notiamo che il metodo è esplicitamente invariante per inversione temporale. In effetti, supponiamo di avere ottenuto le stime di  $x(t_{k+1})$  e  $v(t_{k+1/2})$  mediante le espressioni scritte sopra, e valutiamo  $x(t_k)$  e  $v(t_{k-1/2})$ . Otterremo

$$x(t_k) = x(t_{k+1}) - hv(t_{k+1/2}) = x(t_k) + hv(t_{k+1/2}) - hv(t_{k+1/2}) = x(t_k); \quad (133)$$

$$\begin{aligned} v(t_{k-1/2}) &= v(t_{k+1/2}) - hf(x(t_k), t_k) = v(t_{k-1/2}) + hf(x(t_k), t_k) - hf(x(t_k), t_k) \\ &= v(t_{k-1/2}). \end{aligned} \quad (134)$$

Si può anche controllare che l'algoritmo "velocity Verlet" (126,127) possiede la stessa invarianza. Abbiamo in effetti, scrivendo  $x_0$  per  $x(t_0)$  ecc., e  $x_1$  per  $x(t_0 + h)$  ecc.,

$$\begin{aligned} x_0 &= x_1 - hv_1 + (h^2/2)f(x_1, t_1) \\ &= x_0 + hv_0 + (h^2/2)f(x_0, t_0) - h[v_0 + (h/2)(f(x_0, t_0) + f(x_1, t_1))] + (h^2/2)f(x_1, t_1) \\ &= x_0; \end{aligned} \quad (135)$$

$$\begin{aligned} v_0 &= v_1 - (h/2)[f(x_0, t_0) + f(x_1, t_1)] \\ &= v_0 + (h/2)[f(x_0, t_0) + f(x_1, t_1)] - (h/2)[f(x_0, t_0) + f(x_1, t_1)] \\ &= v_0. \end{aligned} \quad (136)$$

In effetti l'algoritmo "velocity Verlet" può essere ottenuto a partire dal metodo leapfrog introducendo la seguente stima della velocità all'istante  $t_k$ :

$$v(t_k) = \frac{v(t_{k+1/2}) - v(t_{k-1/2})}{h}, \quad (137)$$

ed eliminando la velocità  $v(t_{k+1/2})$  a ogni passo. Così, in un certo senso, il metodo "velocity Verlet" è solo un modo differente di valutare il metodo leapfrog.

## 6. Estrapolazione di Richardson

È possibile ottenere approssimazioni ancora migliori alla soluzione esatta  $x(t)$  utilizzando l'informazione contenuta nella dipendenza della stima di  $\tilde{x}(t_0 + h)$  dall'ampiezza del passo  $h$ . Supponiamo di avere un metodo che permetta di stimare  $x(t_0 + h)$ , noto  $x(t_0)$ , con un errore di ordine  $h^k$ , dove  $k$  è un certo intero positivo. Si avrà così

$$x(t_0 + h) = \tilde{x}_h + ah^k + \dots, \quad (138)$$

dove abbiamo trascurato termini di ordine superiore ad  $h^k$ . Supponiamo adesso di valutare  $x(t_0 + h)$  usando un passo  $h/2$  e iterando una volta l'algoritmo. Otteniamo una stima che indichiamo con  $\tilde{x}(h/2)$ , e si ha

$$x(t_0 + h) = \tilde{x}_{h/2} + \frac{a}{2^k}h^k + \dots \quad (139)$$

Moltiplichiamo questa seconda equazione per  $2^k$  e sottraiamo membro a membro la (138). Otteniamo

$$(2^k - 1)x(t_0 + h) = 2^k\tilde{x}_{h/2} - \tilde{x}_h + \dots \quad (140)$$

Quindi

$$x(t_0 + h) = \frac{2^k\tilde{x}_{h/2} - \tilde{x}_h}{2^k - 1} + \dots, \quad (141)$$

dove i termini trascurati sono di ordine superiore a  $h^k$ . Questa tecnica, che può essere facilmente generalizzata a fattori diversi da 2, è chiamata **estrapolazione di Richardson**.

Vediamo come funziona questo metodo per la soluzione dell'equazione differenziale

$$\frac{dx}{dt} = f(x, t) \quad (142)$$



con il metodo di Eulero. Si ha

$$\tilde{x}_h(t_0 + h) = x_0 + hf_0, \quad (143)$$

dove  $f_0 = f(x_0, t_0)$ , e

$$\tilde{x}_{h/2}(t_0 + h/2) = x_0 + \frac{h}{2}f_0; \quad (144)$$

$$\tilde{x}_{h/2}(t_0 + h) = x_0 + \frac{h}{2}f_0 + \frac{h}{2}f\left(x_0 + \frac{h}{2}f_0, t_0 + \frac{h}{2}\right). \quad (145)$$

PoichÈ l'errore *globale* del metodo di Eulero è di ordine  $h$ , poniamo  $k = 1$ . Abbiamo così la stima accelerata

$$\begin{aligned} \tilde{x}(t_0 + h) &= 2\tilde{x}_{h/2}(t_0 + h) - \tilde{x}_h(t_0 + h) \\ &= x_0 + hf\left(x_0 + \frac{h}{2}f_0, t_0 + \frac{h}{2}\right). \end{aligned} \quad (146)$$

Riotteniamo così il metodo del punto di mezzo esplicito.