CAPITOLO 8

EQUAZIONI DIFFERENZIALI

Consideriamo il seguente problema di Cauchy per i sistemi di equazioni differenziali del primo ordine :

$$y'(t) = f(t,y(t))$$
$$y(t_0)=y_0 \qquad (8.1)$$

dove f(t,y): $[t_0, t_f]\times R^m \to R^m$. La funzione f(t,y) è supposta continua rispetto a t , e lipschitziana rispetto ad y nella striscia illimitata $[t_0, t_f]\times R^m$,

$$||f(t,u)-f(t,v)||<L||u-v||, \quad \forall\, t\in[t_0, t_f] \text{ e } \forall\, u,v \in\times R^m.$$

In tali condizioni e garantita l'esistenza e l'unicità della soluzione nell'intero intervallo di integrazione $[t_0, t_f]$.

Consideriamo una discretizzazione dell'intervallo $[t_0, t_f]$ che, per semplicità di esposizione, supporremo uniforme:

$$t_0< t_1<...< t_N (=t_f) \qquad h=\frac{t_f - t_0}{N}.$$

e, per ogni intervallo $[t_n, t_{n+1}]$, consideriamo l'identità:

$$\int_{t_n}^{t_{n+1}} y'(t)dt = \int_{t_n}^{t_{n+1}} f(t, y(t))dt$$

$$y(t_{n+1}) = y(t_n) + \int_{t_n}^{t_{n+1}} f(t, y(t))dt .$$

Ogni formula di quadratura che fa uso dei valori nodali di y(t) può essere utilizzata per creare una formula di integrazione numerica per il problema (8.1). Limitiamoci a considerare alcune formule, trattate nel capitolo precedente, che fanno uso di uno o di entrambi gli estremi dell'integrale. In particolare:

1) $$\int_{t_n}^{t_{n+1}} f(t, y(t))dt = hf(t_n,y(t_n)) +\sigma_1(t_n,h)$$

$$\sigma_1(t_n,h) = \tfrac{1}{2}\frac{\partial}{\partial t} f(\xi_n,y(\xi_n))h^2 = \tfrac{1}{2} y''(\xi_n)h^2 \qquad \xi_n\in (t_n, t_{n+1})$$

2) $\int_{t_n}^{t_{n+1}} f(t,y(t))dt = hf(t_{n+1},y(t_{n+1})) + \sigma_2(t_n,h)$

$$\sigma_2(t_n,h) = -\tfrac{1}{2}\frac{\partial}{\partial t}f(\xi_n,y(\xi_n))h^2 = -\tfrac{1}{2}y''(\xi_n)h^2 \qquad \xi_n \in (t_n,\ t_{n+1})$$

3) $\int_{t_n}^{t_{n+1}} f(t,y(t))dt = \tfrac{1}{2}h\big(f(t_n,y(t_n))+f(t_{n+1},y(t_{n+1}))\big) + \sigma_3(t_n,h)$

$$\sigma_3(t_n,h) = -\tfrac{1}{12}\frac{\partial^2}{\partial t^2}f(\xi_n,y(\xi_n))h^3 = -\tfrac{1}{12}y'''(\xi_n)h^3 \qquad \xi_n \in (t_n,\ t_{n+1})$$

Otteniamo così le relazioni:

1') $y(t_{n+1}) = y(t_n) + hf(t_n,y(t_n)) + \sigma_1(t_n,h)$

2') $y(t_{n+1}) = y(t_n) + hf(t_{n+1},y(t_{n+1})) + \sigma_2(t_n,h)$

3') $y(t_{n+1}) = y(t_n) + \tfrac{1}{2}h\big(f(t_n,y(t_n))+f(t_{n+1},y(t_{n+1}))\big) + \sigma_3(t_n,h)$.

Trascurando ad ogni passo l'errore $\sigma(t_n,h)$, detto **errore locale di troncamento**, si ottengono le formule ricorsive:

formula di **Eulero Esplicita** $y_{n+1} = y_n + hf(t_n,y_n)$

formula di **Eulero Implicita** $y_{n+1} = y_n + hf(t_{n+1},y_{n+1})$

formula dei **Trapezi** $y_{n+1} = y_n+\tfrac{1}{2}h(f(t_n,y_n)+f(t_{n+1},y_{n+1}))$.

dove, per ogni n, y_n è l'approssimazione della soluzione nel punto $t_n= t_0+nh$ ed y_0 è il valore iniziale assegnato nel problema (8.1).
Si osservi che le formule di Eulero Implicito e dei trapezi presentano una maggiore complessità computazionale, rispetto alla formula di Eulero esplicito, poiche l'incognita y_{n+1} si presenta come la risoluzione di un sistema di equazioni, in generale, non lineari.
Per comodità di trattazione, esprimiamo le precedenti formule nella forma generale:

$$y_{n+1}= y_n+h\Phi(t_n, y_n, y_{n+1}) \qquad (8.2)$$

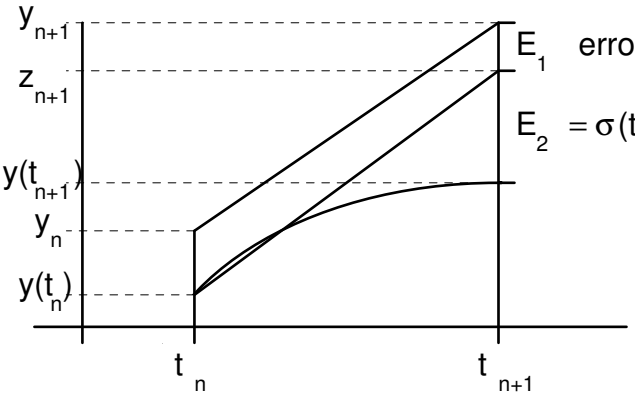
per ciacuna delle quali è immediato verificare la lipschitzianità

$$|\Phi(t,u,v)-\Phi(t,w,z)|<M(|u-w|+|v-z|)$$

come conseguenza della lispchitzianità di f.

Detto $e_n := y_n - y(t_n)$ l'errore accumulato fino al passo n-esimo di integrazione, analizziamo come esso si propaga nel passo successivo. A tale scopo indichiamo con z_{n+1} il valore fornito dalla formula (8.2) qualora essa fosse applicata al punto $y(t_n)$ della traiettoria esatta:

$$z_{n+1} = y(t_n) + h\Phi(t_n, y(t_n), z_{n+1})$$



L'errore totale al passo n+1 è quindi dato da:

$$e_{n+1} = y_{n+1} - y(t_{n+1}) = y_{n+1} - z_{n+1} + z_{n+1} - y(t_{n+1}) = E_1 + E_2 = (y_{n+1} - z_{n+1}) + \sigma(t_n,h)$$

$$\| e_{n+1} \| \leq \| y_{n+1} - z_{n+1} \| + \| \sigma(t_n,h) \| \qquad (8.3)$$

dove:

$$y_{n+1} - z_{n+1} = y_n + h\Phi(t_n, y_n, y_{n+1}) - \left(y(t_n) + h\Phi(t_n, y(t_n), z_{n+1}) \right)$$

$$y_{n+1} - z_{n+1} = e_n + h\left(\Phi(t_n, y_n, y_{n+1}) - \Phi(t_n, y(t_n), z_{n+1})\right)$$

$$\| y_{n+1} - z_{n+1} \| \leq \| e_n \| + hM\left(\| e_n \| + \| y_{n+1} - z_{n+1} \| \right)$$

$$(1-hM) \| y_{n+1} - z_{n+1} \| < (1+hM) \| e_n \|$$

e, per h sufficientemente piccolo,

$$\| y_{n+1} - z_{n+1} \| < \frac{1+hM}{1-hM} \| e_n \| .$$

Tornando alla (8.3) si ottiene quindi:

$$||e_{n+1}|| < \frac{1+hM}{1-hM} ||e_n|| + ||\sigma(t_n,h)||.$$

Osservato inoltre che $\frac{1+hM}{1-hM} < 1+3hM$, si ha:

$$||e_{n+1}|| < (1+ 3hM) ||e_n|| + ||\sigma(t_n,h)||$$

Maggiorando infine l'errore locale di troncamento $||\sigma(t_n,h)||$ in modo uniforme sull'intervallo di integrazione $[t_0, t_f]$

$$\sigma(h) := \max_{t\in[t_0,t_n]} ||\sigma(t,h)||$$

si ottiene la seguente relazione ricorsiva per l'errore:

$$||e_{n+1}|| < (1+ 3hM) ||e_n|| + \sigma(h), \quad n=0,1,...,N-1. \qquad (8.4)$$

Lemma: *Se la successione* $\{a_n\}$, $a_n>0$, *soddisfa la relazione ricorsiva*

$$a_{n+1} < (1+hQ)a_n + c(h) \qquad n=0,1,2,...,N$$

con $(1+hQ)>0$, *allora vale la maggiorazione:*

$$a_m < (1+hQ)^N a_0 + c(h)\frac{(1+hQ)^N - 1}{hQ} \qquad \forall m\le N.$$

(La dimostrazione è lasciata come esercizio).

Applicando il lemma alla relazione ricorsiva (8.4), tenendo conto che $e_0=0$, si ottiene la maggiorazione:

$$||e_m|| < \sigma(h)\frac{(1+3hM)^N - 1}{3hM} \qquad \forall m\le N$$

e, tenuto conto della disuguaglianza $(1+ 3hM)<e^{3hM}$,

$$||e_m|| < \sigma(h)\frac{e^{3MhN}-1}{3hM} \qquad \forall m \le N$$

Poichè il numero totale di passi N e l'ampiezza del passo h sono legati dalla relazione $Nh = (t_f - t_0)$, si ottiene

$$||e_m|| < \sigma(h)\frac{e^{3M(t_f - t_0)}-1}{3hM} \qquad \forall m \le N. \qquad (8.5)$$

L'ultima relazione è fondamentale per l'analisi della convergenza del metodo.

Si dirà che il metodo (8.2) è **convergente** nell'intervallo d'integrazione $[t_0, t_f]$, se

$$\max_{m \le N} ||e_m|| \to 0 \quad \text{per } N \to \infty \text{ e } h \to 0$$

ferma restando la relazione $Nh = (t_f - t_0)$. Si dirà inoltre che il metodo ha **ordine di convergenza** uguale a p se il termine $\max_{m \le N} ||e_m||$ è infinitesimo di ordine p.

Dalla relazione (8.5) si deduce immediatamente il seguente teorema di convergenza:

Teorema di convergenza. *Affinchè il metodo*

$$y_{n+1} = y_n + h\Phi(t_n, y_n, y_{n+1})$$

sia convergente di ordine p *nell'intervallo* $[t_0 - t_f]$ *è sufficiente che la funzione* $\Phi(t, u, v)$ *sia lipchitziana rispetto a* u *e* v *per ogni* $t \in [t_0 - t_f]$, *e che il rapporto* $\frac{\sigma(h)}{h}$ *sia infinitesimo di ordine* p.

Dalle espressioni dell'errore locale di troncamento si deduce che i metodi di Eulero esplicito ed implicito convergono con ordine p=1, mentre il metodo dei trapezi converge con ordine p=2.

Esistono molte altre formule del tipo considerato che sono convergenti con vari ordini. A titolo di esempio, una formula esplicita di ordine p=2 è la seguente

formula di **Eulero generalizzata**.

$$y_{n+1} = y_n + hf\left(t_n + \tfrac{1}{2}h,\ y_n + \tfrac{1}{2}hf(t_n, y_n)\right).$$

Essa deriva dalla relazione:

$$y(t_{n+1}) = y(t_n) + hf\left(t_n+\tfrac{1}{2}h,\ y(t_n)+\tfrac{1}{2}hf(t_n,y(t_n))\right) + \sigma(t_n,h) \qquad (8.6)$$

nella quale l'errore di troncamento

$$\sigma(t_n,h)= y(t_{n+1}) - y(t_n) - hf\left(t_n+\tfrac{1}{2}h,\ y(t_n)+\tfrac{1}{2}hf(t_n,y(t_n))\right)$$

risulta essere un infinitesimo di ordine 3 rispetto ad h. Ciò si verifica facilmente sviluppando la funzione f(t,y) in un intorno del punto $(t_n,y(t_n))$:

$$f\left(t_n+\tfrac{1}{2}h,\ y(t_n)+\tfrac{1}{2}hf(t_n,y(t_n))\right)=$$
$$=f(t_n,y(t_n)) + \tfrac{1}{2}h\, f_t(t_n,y(t_n)) + \tfrac{1}{2}hf(t_n,y(t_n))\, f_y(t_n,y(t_n)) + O(h^2)$$

e sviluppando pure y(t) in un intorno di t_n

$$y(t_{n+1}) = y(t_n+h) = y(t_n) + hy'(t_n)+\tfrac{1}{2}h^2y''(t_n) + O(h^3).$$

Tenuto infine conto che $y'(t_n)=f(t_n,y(t_n))$, e che

$$y''(t_n)=\frac{\partial}{\partial t}f(t_n,y(t_n))=f_t(t_n,y(t_n))+f_y(t_n,y(t_n))y'(t_n),$$

dalla (8.6) si ricava $\sigma(t_n,h)=O(h^3)$.

Propagazione dell'errore e stabilità.

Abbiamo visto in precedenza per il problema iniziale (8.1), che ad ogni passo l'errore e_n si compone di due parti: l'errore propagato e l'errore di troncamento. Abbiamo altresì visto che gli errori si accumulano durante il processo di integrazione e la stima (8.5) ne rappresenta una limitazione uniforme su tutto l'intervallo $[t_0 - t_f]$. Di fatto, ad ogni passo, applichiamo la formula per risolvere l'equazione data con valore iniziale perturbato y_n anzichè $y(t_n)$.
Se, in particolare, ad ogni passo l'errore propagato $E_2 := ||y_{n+1} - z_{n+1}||$ risulta minore dell'errore accumulato fino al passo precedente, cioè se:

$$||y_{n+1} - z_{n+1}|| \le ||e_n||, \qquad (8.7)$$

allora si dice che il metodo è **stabile**.
Per i metodi stabili la relazione (8.3) si può sviluppare nel seguente modo:

$$|| e_{n+1} || \leq ||y_{n+1} - z_{n+1}|| + ||\sigma(t_n,h)|| < ||e_n|| + ||\sigma(t_n,h)||$$

$$< ||e_{n-1}|| + ||\sigma(t_{n-1},h)|| + ||\sigma(t_n,h)|| < ...$$

$$< ||e_0|| + ||\sigma(t_0,h)|| + ||\sigma(t_1,h)|| + ... + ||\sigma(t_n,h)||$$

$$= ||\sigma(t_0,h)|| + ||\sigma(t_1,h)|| + ... + ||\sigma(t_n,h)||.$$

Poichè, come abbiamo visto, $\sigma(t_k,h)<\sigma(h)$ per ogni k, allora si ottiene :

$$|| e_m || < m\,\sigma(h) < N\,\sigma(h) = \sigma(h)\,\frac{t_f - t_0}{h} \qquad \forall m \leq N.$$

Ciò significa che, per i metodi stabili, la crescita dell'errore è limitata in modo lineare rispetto all'intervallo $[t_0 - t_f]$ anziche in modo esponenziale come indicato dalla (8.5) per un metodo qualunque. Vediamo, a questo proposito, un esempio numerico istruttivo:

Consideriamo il problema scalare:

$$y' = -100y + 100\,\mathrm{sen}(t)$$
$$y(0)=0$$

la cui soluzione esatta è:

$$y(t) = \frac{\mathrm{sen}(t) \; - \; 0.01\cos(t) \; + \; 0.01\,e^{-100t}}{1.0001}$$

Supponiamo di integrare il problema nell'intervallo [0,3] con il metodo di RK esplicito di ordine 4. In corrispondenza a vari valori del passo h troviamo le seguenti approssimazioni nel punto finale $t_f=3$:

h	0.015	0.020	0.025	0.030
N	200	150	120	100
y(3)	0.151004	0.150996	0.150943	$6.7\cdot10^{11}$

Cosa è successo nel passare dal passo 0.025 al passo 0.030 ? Siamo passati da una situazione in cui la condizione (8.11) è verificata ad una in cui non lo è più. In altre parole siamo passati da una propagazione lineare ad una proagazione esponenziale dell'errore. Come vedremo tra poco, l'insorgenza del fenomeno di propagazione esponanziale dell'errore dipende sia dal problema trattato che dal metodo impiegato.
Naturalmente sarebbe preferibile utilizzare un metodo per il quale la condizione di stabilità (8.11) fosse verificata *per ogni scelta dal passo*.

In generale è difficile verificare la stabilità dei metodi per equazioni qualunque, e pertanto ci limiteremo a studiare la stabilità per una classe molto particolare di equazioni test.

Consideriamo dapprima la seguente equazione scalare:

$$y'(t)=\lambda y(t)$$
$$y(0)=1 \qquad (8.8)$$

dove, per ragioni che vedremo tra poco, il coefficiente λ e la funzione y sono *complessi*.

E' noto che la soluzione è data della funzione $y(t)=e^{\lambda t}$.

Detto $\lambda=\alpha+i\beta$, si ottiene:

$$y(t)=e^{\lambda t}=e^{(\alpha+i\beta)t}=e^{\alpha t}(\cos\beta t + i\,\mathrm{sen}\beta t)$$

e per i moduli:

$$|y(t)|=|y_0|e^{\alpha t}$$

Per quanto riguarda la stabilità del problema (8.8) rispetto alle variazioni sul dato iniziale, sia z(t) la soluzione di (8.8) con dato iniziale z_0 . Per la linearità dell'equazione si ha

$$|y(t)-z(t)|=|y_0-z_0|e^{\alpha t}$$

e quindi la condizione $\alpha\leq 0$ è necessaria e sufficiente per avere

$$|y(t)-z(t)|\leq|y_0-z_0|e^{\alpha t} \qquad \text{per ogni } t>0.$$

In questo caso diremo che (8.8) è un **problema stabile**.

Analizziamo ora la stabilità dei metodi numerici per il problema (8.8) nell'ipotesi che il problema stesso sia stabile, cioè che sia $\alpha = \mathrm{Re}(\lambda) < 0$.

Il metodo di Eulero esplicito, applicato all'equazione test, è:

$$y_{n+1} = y_n + h\lambda y_n = (1+h\lambda) y_n$$

ed il corrispondente valore z_{n+1} è dato da:

$$z_{n+1} = y(t_n) + h\lambda y(t_n) = (1+h\lambda) y(t_n).$$

Si ha quindi, per l'errore propagato:

$$y_{n+1} - z_{n+1} = (1+h\lambda)\, e_n.$$

$$|| y_{n+1} - z_{n+1} || = |(1+h\lambda)|\, ||e_n||.$$

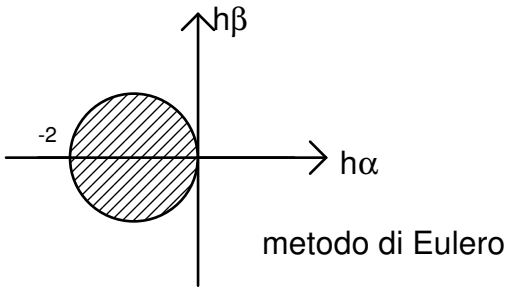
In base alla definizione precedente, si osserva che il metodo è stabile per quei valori complessi del prodotto $h\lambda$ per i quali si ha:

$$|(1+h\lambda)| \leq 1.$$

La funzione $\varphi(h\lambda) := (1+h\lambda)$ è detta **funzione di stabilità** e varia da metodo a metodo. La regione del piano complesso nella quale si ha:

$$|\varphi(h\lambda)| \leq 1$$

è detta **regione di assoluta stabilità** del metodo.
Per il metodo di Eulero la regione di assoluta stabilità è tratteggiata nella seguente figura:



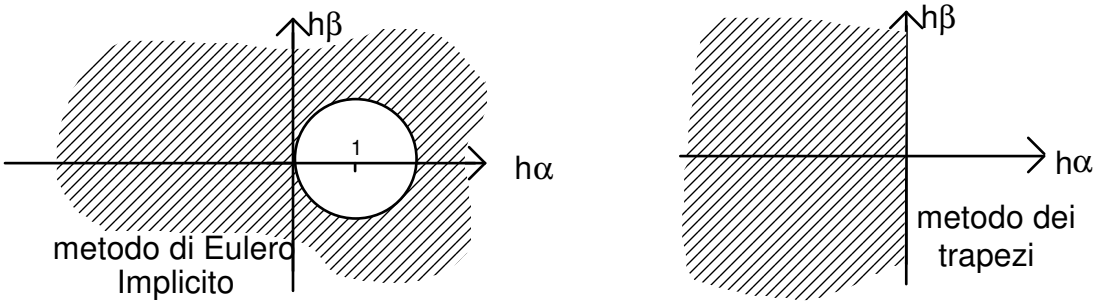
In maniera analoga si trovano le funzioni di stabilità:

$$\varphi(h\lambda) = \frac{1}{1-h\lambda}$$ per il metodo di Eulero implicito

e

$$\varphi(h\lambda)=\frac{1+\frac{h\lambda}{2}}{1-\frac{h\lambda}{2}}$$ per il metodo dei trapezi

alle quali corrispondono le seguenti regioni di assoluta stabilità:



Se la regione di assoluta stabilità di un metodo include il semipiano negativo C^-, diremo che il metodo è **assolutamente stabile** o **incondizionatamente stabile** inquanto risulta stabile per tutte le equazioni (8.8) stabili e *per ogni passo* h.

Equazioni "stiff"

Consideriamo la seguente classe di equazioni differenziali:

$$y'=\lambda\,(y-F(t)) + F'(t)$$

con $\lambda<<-1$ (negativo e grande in modulo). Assegnato il valore iniziale $y(t_0)=y_0$, la soluzione è:

$$y(t) = (y_0 - F(t_0))e^{\lambda(t-t_0)} +F(t)$$

Per ogni $y_0 \neq F(t_0)$, la soluzione y(t) è una funzione che, quando t si allontana da t_0, precipita sulla funzione F(t).

Finchè il termine $(y_0 - F(t_0))e^{\lambda(t-t_0)}$ non è trascurabile rispetto a F(t), si è nella **fase transitoria**, altrimenti si è nella **fase stazionaria**, nella quale la soluzione è praticamente uguale a F(t). Evidentemente la fase transitoria è tanto più breve quanto più grande è il modulo di λ. Si osservi però che, anche nella fase stazionaria, se consideriamo un punto t_n ed un valore perturbato della soluzione y_n, la traiettoria uscente dal punto (t_n, y_n) è

$$x(t) = (y_n - F(t_n))e^{\lambda(t-t_n)} + F(t)$$

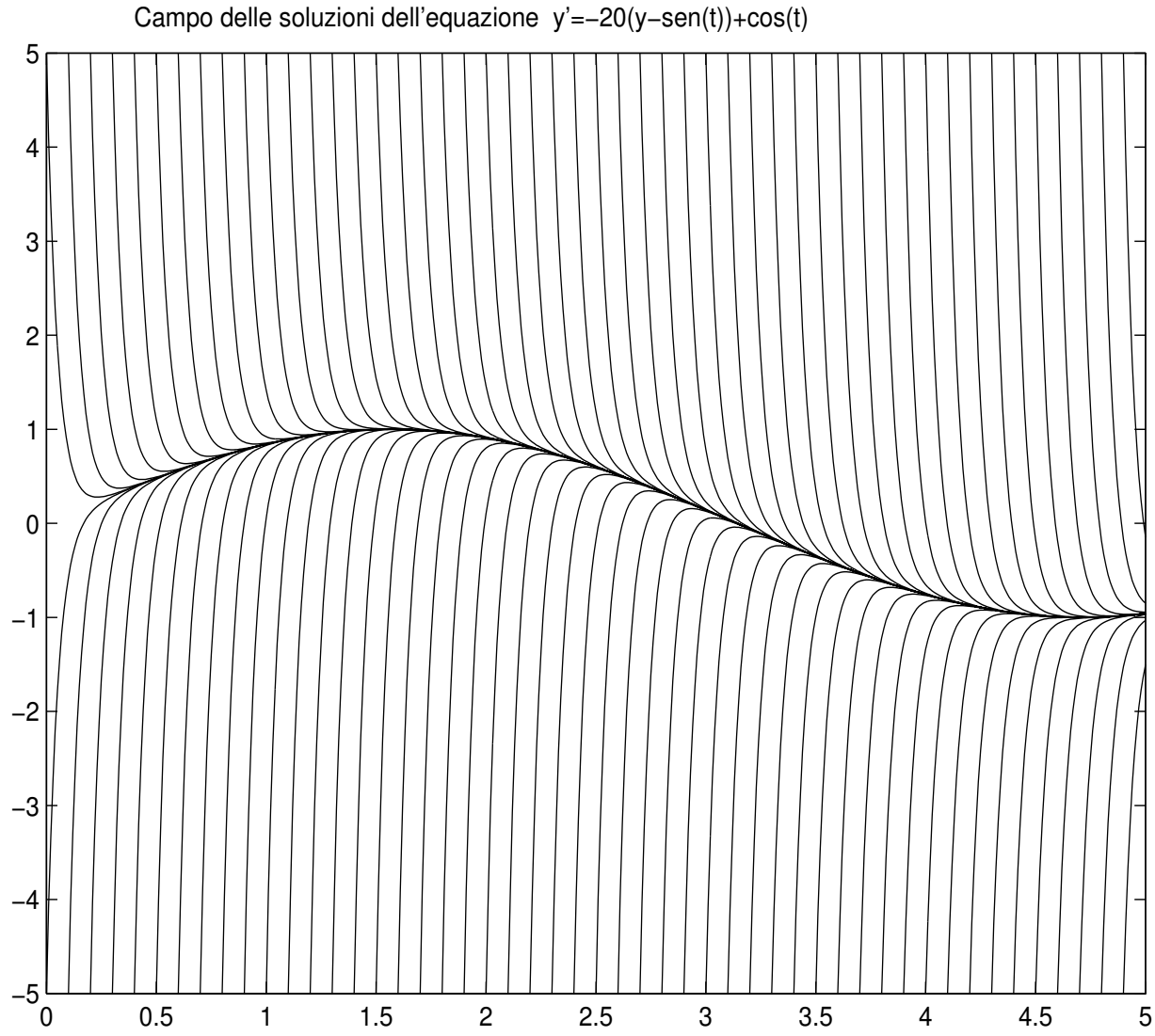
che a sua volta precipita sulla funzione F(t) ed è tale che la sua derivata in t_n si discosta sensibilmente da quella della soluzione esatta anche se siamo lontani da t_0. In altre parole, nella fase stazionaria le altre curve integrali sono sensibilmente diverse dalla soluzione esatta.

Nel seguente grafico si vede il campo delle soluzioni che escono da punti esterni alla soluzione esatta per il problema

$$y' = -20\,(y - \mathrm{sen}(t)) + \cos(t)$$
$$y(0) = 5$$

la cui soluzione e'

$$y(t) = 5\,e^{\lambda t} + \mathrm{sen}(t)$$

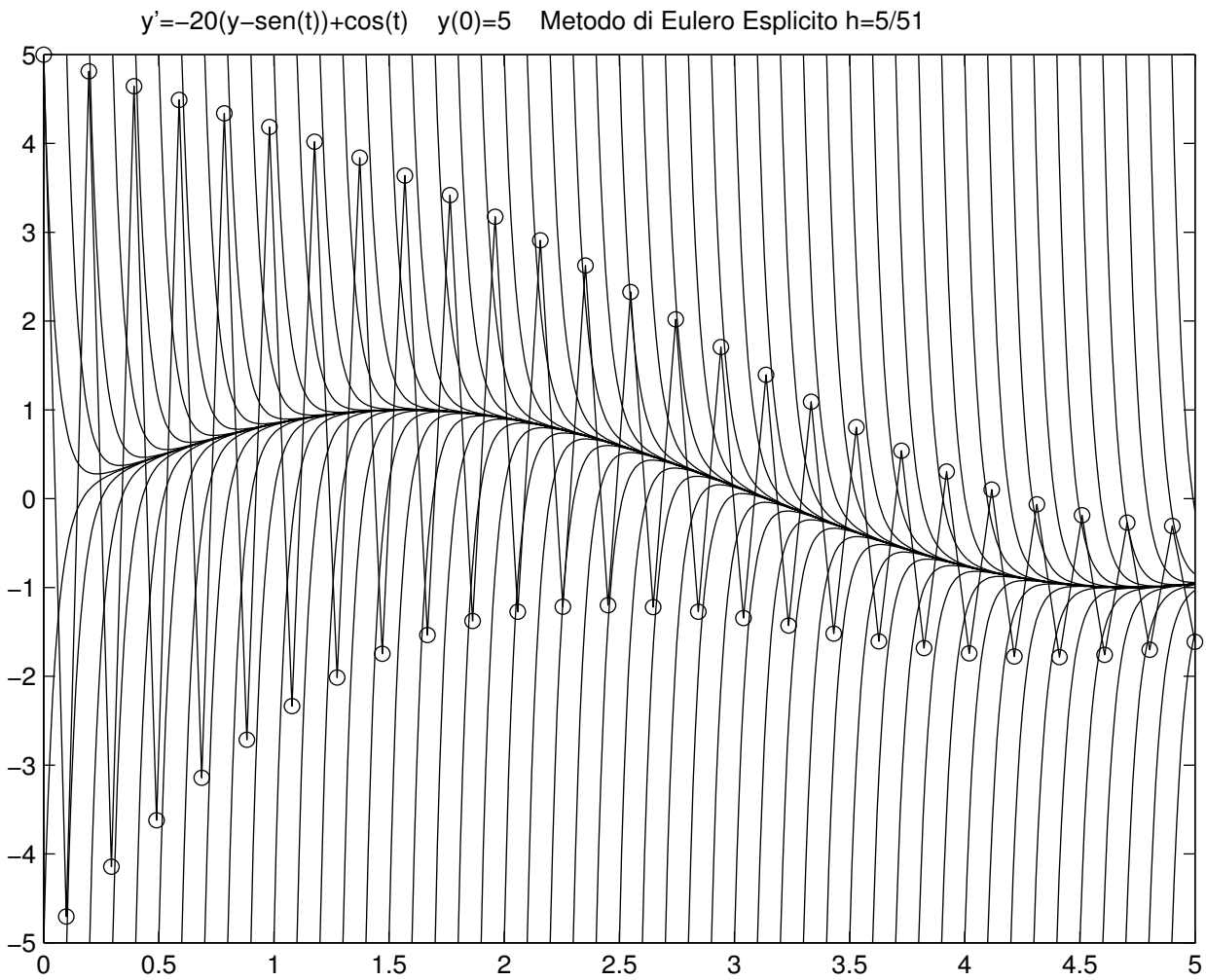


La figura successiva mostra l'approssimazione numerica ottenuta col metodo di Eulero Esplicito. Per il valore assegnato del passo d'integrazione h, l'errore di propagazione E_1 del metodo numerico e' molto grande rispetto all'errore locale di troncamento. Più precisamente, come per l'equazione test (8.8), l'errore propagato dal metodo è:

$$| y_{n+1} - z_{n+1}| = |\varphi(h\lambda)|\, |e_n|.$$

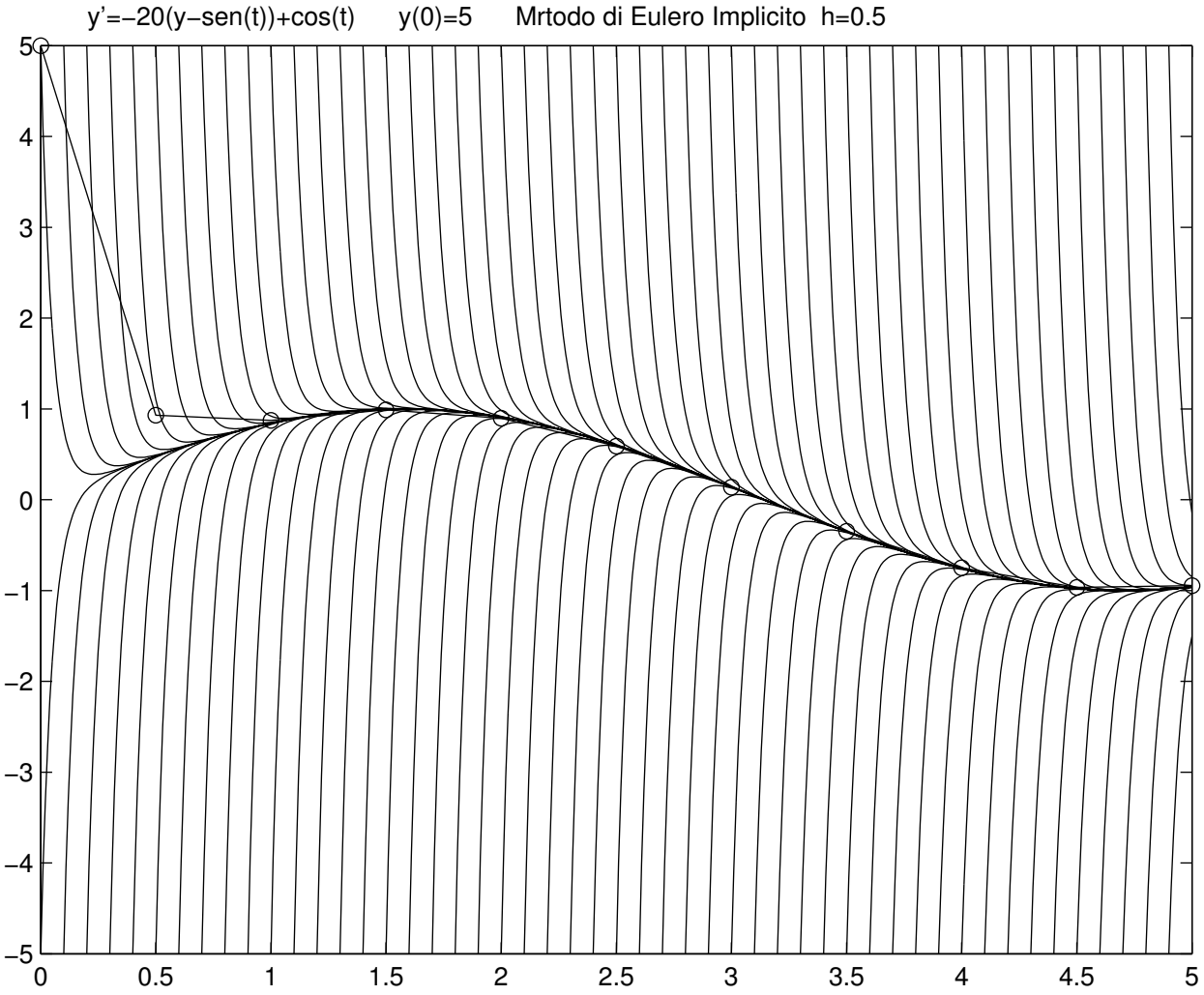
Poiche' il metodo ha una regione finita di assoluta stabilita',bisogna procedere con un passo tale che $|\varphi(h\lambda)|\leq 1$. Nel nostro caso $|1-20h|\leq 1$ e quindi $h\leq 0.1$.

Nella figura si osserva che con un passo di poco inferiore a 0.1 l'errore e' dovuto essenzialmente alla componente di propagazione anche nella fase stazionaria, dove invece l'errore di troncamento e' molto piccolo. Per un valore appena superiore ad 1 si avrebbe una crescita esponenziale dell'errore.



Viceversa, per il metodo di Eulero Implicito, che e' assolutamente stabile, la condizione $|\varphi(h\lambda)|\leq 1$ è verificata per ogni valore del passo.

Nella figura successiva e' illustrato il caso di un passo h=0.5 per il quale, nella fase stazionaria, l'errore propagato e' trascurabile e l'errore globale e' dovuto essenzialmente all'errore di troncamento .



Equazioni differenziali le cui soluzioni hanno un comportamento simile a questo sono dette **equazioni stiff** (rigide) . Nella fase transitoria esse potranno essere integrate indifferentemente con metodi espliciti o impliciti in quanto l'errore locale sara' dominato dall'errore di troncamento ed il passo sara' modulato su di esso. Invece nella fase stazionaria, se il metodo non e' stabile, l'errore locale sara' dominato dall'errore di propagazione che imporra' un passo forzatamente piccolo, molto piu' piccolo di quanto sarebbe richiesto dall'errore di troncamento. Un metodo stabile consentira', viceversa, di procedere con passi di integrazione piu' ampi vincolati essenzialmente dall'errore di troncamento.

Per l'analisi della *stabilità dei sistemi* consideriamo ora, come equazione test, il sistema lineare:

$$y'(t) = Ay(t)$$
$$y(0)= u \qquad (8.9)$$

dove $A \in R^{m\times m}$ e $u=(1,1,...,1)\in R^m$. In questo caso il sistema è stabile se e solo se $\mathbb{R}e(\lambda)\leq 0$ per ogni autovalore λ di A

Il metodo di Eulero esplicito applicato al sistema (8.9) è :

$$y_{n+1}= y_n+hAy_n= (I+hA)y_n$$

mentre per il metodo di Eulero Implicito si ha:

$$y_{n+1}= y_n+hAy_{n+1}$$

e quindi

$$y_{n+1}=(I-hA)^{-1}y_n$$

Per il metodo dei trapezi si ha:

$$y_{n+1} = \left(I-\frac{hA}{2}\right)^{-1}\left(I+\frac{hA}{2}\right)y_n$$

e non è difficile vedere, più in generale, che per ogni metodo si ha:

$$y_{n+1}=\varphi(hA)y_n$$

dove la funzione φ, che ora trasforma matrici in matrici, è proprio la funzione di stabilità precedentemente definita per il caso scalare.
Come per il caso scalare, l'errore di propagazione è

$$||y_{n+1}-z_{n+1}||\leq||\varphi(hA)||\ ||e_n||$$

ed il metodo è stabile se, in qualche norma, $||\varphi(hA)||\leq 1$, cioè se il raggio spettrale, e quindi il modulo di ogni autovalore della matrice $\varphi(hA)$ è <1 (Attenzione: devo escludere $\varphi(hA) =1$ perche' in tal caso potrebbe non esistere una norma $||\varphi(hA)||\leq 1$).
Ricordando ora che se λ è autovalore di A allora $\varphi(h\lambda)$ è autovalore di $\varphi(hA)$, è sufficiente che sia $|\varphi(h\lambda)|<1$ per ogni λ autovalore di A. Poichè λ è, in generale, un numero complesso, lo studio delle regioni di stabilità per l' equazione test scalare (8.8) è

sufficiente anche per il caso vettoriale. Infatti un metodo risulta stabile se $h\lambda$ è incluso (in senso stretto) nella regione di assoluta stabilità per ogni λ autovalore di A.
In particolare se il metodo è assolutamente stabile, allora la condizione di stabilità

$$||y_{n+1}-z_{n+1}|| \leq ||e_n||$$

è verificata, *indipendentemente dal passo* h, per tutte le equazioni test stabili (cioè con autovalori di A a parte reale negativa).

Stima dell'errore locale ed algoritmi a passo variabile:

E' chiaro che la soluzione di una equazione differenziale puo' avere comportamenti qualitativi molto diversi lungo l'intervallo di integrazione. L'esempio piu' evidente e' quello delle equazioni stiff che, dopo un tratto transitorio nel quale la soluzione subisce una variazione molto rapida, passano al regime stazionario dove la soluzione e' liscia e potrebbe essere integrata con un passo molto piu' grande aumentando l'efficienza dell'algoritmo.

In questo paragrafo si propone una procedura ***empirica*** di integrazione a passo variabile che, ad ogni passo, adatta la lunghezza del passo stesso alle caratteristiche qualitative dell'equazione e della soluzione basandosi su una stima locale dell'errore.

Supponiamo di voler integrare l'equazione differenziale con un metodo di ordine locale p+1. Al passo n-esimo disponiamo del valore approssimato y_n e, utilizzando la formula approssimata con passo h_{n+1}, calcoliamo y_{n+1}.

Contrariamente a quanto fatto per l'analisi della convergenza, indichiamo con z_{n+1} la soluzione esatta uscente dal punto y_n nel punto t_{n+1} e indichiamo con

$$\sigma_{n+1} = ||y_{n+1} - z_{n+1}||$$

l'errore commesso (si noti che questo non e' l'errore locale di troncamento come e' stato definito in precedenza!). Di questo errore possiamo avere una stima utilizzando un metodo di ordine superiore, diciamo p+2, che fornisce il valore approssimato $\bar{y}_{n+1}$ da considerarsi "esatto" rispetto all'approssimazione fornita dal metodo di ordine p+1. Quindi possiamo concludere che, utilizzando il metodo di ordine p+1, si e' commesso un errore che, a meno di un infinitesimo di ordine p+2, vale

$$\sigma_{n+1} \approx ||y_{n+1} - \bar{y}_{n+1}|| .$$

A questo punto sottoponiamo l'errore σ_{n+1} cosi' stimato, al ***test di tolleranza***

$$\sigma_{n+1} \leq TOL \cdot h_{n+1}$$

dove TOL e' la ***tolleranza per unita' di passo***, cioe' il massimo errore che intendo accettare per un passo di integrazione di ampiezza $h_{n+1}=1$.

Passo rifiutato: Se il test non viene superato**,** il valore y_{n+1} viene *rifiutato* e la formula d'integrazione viene ricalcolata con un nuovo passo d'integrazione h_{new} , inferiore ad h_{n+1}.

La riduzione del passo non viene fatta in maniera arbitraria, per esempio dimezzando la lunghezza del passo rifiutato, ma viene valutata in maniera "ottimale" utilizzando i calcoli gia' eseguiti. A tale scopo si osservi che l'errore σ_{n+1} ha la forma $K \cdot h^{p+1}$ per qualche valore di K che non conosco ma posso stimare dall'uguaglianza

$$\sigma_{n+1} = ||y_{n+1} - \bar{y}_{n+1}|| = k \cdot h^{p+1}.$$

Ottengo cosi' una stima di K

$$K = \frac{\sigma_{n+1}}{h_{n+1}^{p+1}}$$

che ritengo valida anche per piccole variazioni del passo. A questo punto posso dire che, per il nuovo passo h_{new}, commettero' un errore stimabile, a priori, in $K \cdot (h_{new})^{p+1}$.

Per passare il test di tolleranza con il nuovo passo, richiedero' che tale errore soddisfi

$$K \cdot (h_{new})^{p+1} \leq TOL \cdot h_{new}.$$

Per evitare che il test fallisca a causa dei termini trascurati (che, sebbene di ordine superiore, possono compromettere la stima del nuovo passo) il nuovo passo viene calcolato sulla base della richiesta piu' stringente

$$K \cdot (h_{new})^{p+1} = \tfrac{1}{2} TOL \cdot h_{new}. \qquad (1.14)$$

Da quest'ultima posso quindi ricavare una stima per h_{new}:

$$h_{new} = \sqrt[p]{\frac{TOL}{2K}} = \sqrt[p]{\frac{TOL \cdot h_{n+1}^{p+1}}{2\sigma_{n+1}}} = h_{n+1} \sqrt[p]{\frac{TOL \cdot h_{n+1}}{2\sigma_{n+1}}}$$

Poiche' in prossimita' di forti variazioni della soluzione il fattore di riduzione

$$R= \sqrt[p]{\frac{TOL \cdot h_{n+1}}{2\sigma_{n+1}}}$$

puo' risultare molto piccolo, quando $\sigma_{n+1}>>1$, allora, per evitare una riduzione eccessiva del passo , si definisce a priori una riduzione massima del passo, diciamo *non meno della meta'*, e quindi si definisce l'ampiezza del nuovo passo di tentativo

$$h_{new} = h_{n+1} \cdot \max\{½ , R\}$$

Con tale passo, rinominato h_{n+1} ($\leftarrow h_{new}$), si ripete l'intera procedura finche' il test di tolleranza viene superato. Quando cio' accade, il valore y_{n+1} viene accettato e si passa al passo successivo.

Passo accettato. Quando il valore y_{n+1} viene accettato e si passa al passo successivo, la procedura puo' essere ottimizzata utilizzando un passo di tentativo

$$h_{n+2} = R\, h_{n+1}$$

dove, per la (1,14), sara' R>1. Come nel caso della riduzione, anche nel caso di espansione del passo si pone una *protezione* del tipo

$$h_{n+2} = h_{n+1} \cdot \min\{2 , R\}$$

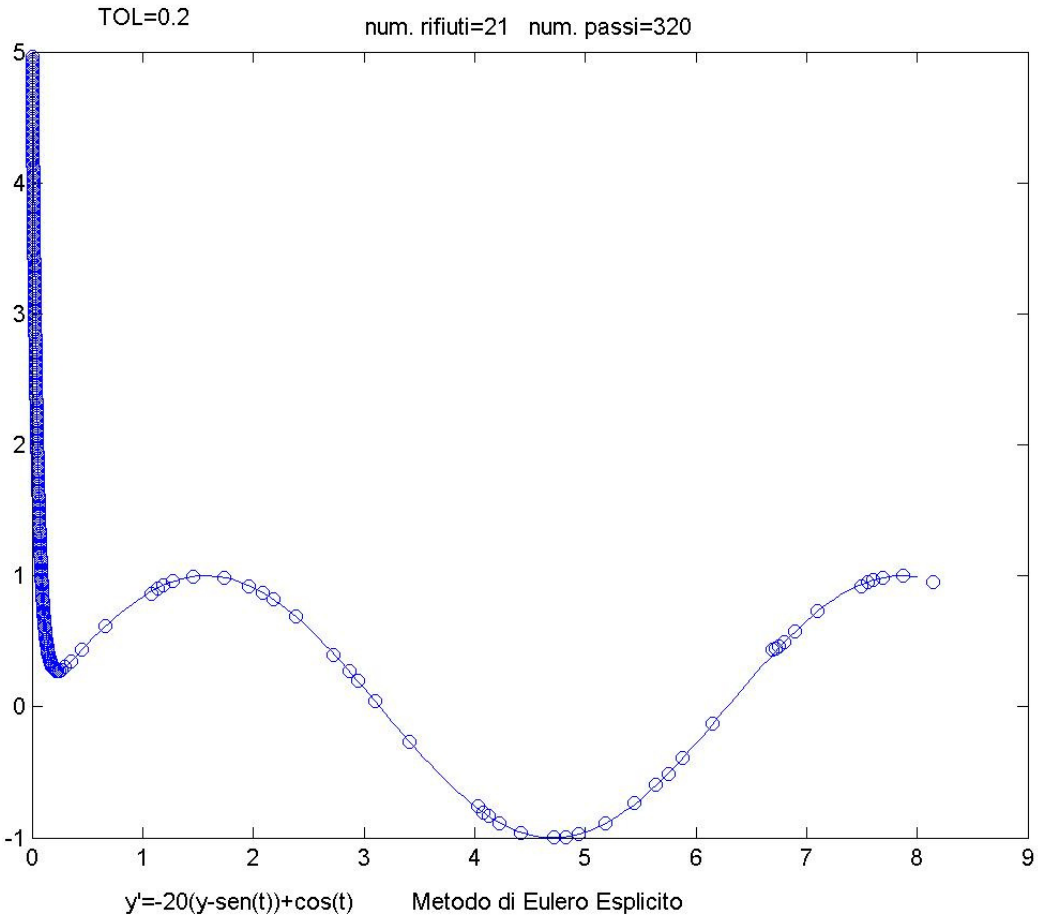
che ne limita l'allungamento.

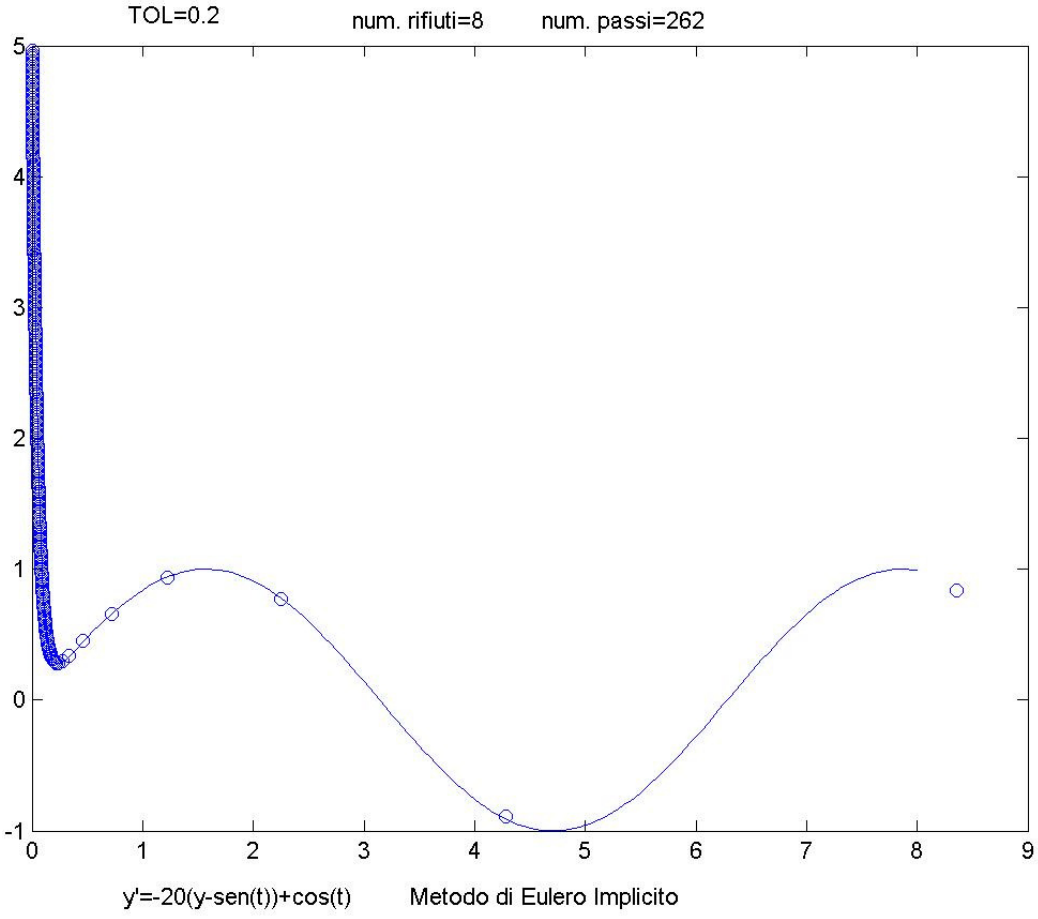
Con questo passo di tentativo, si applica la procedura descritta e si procede fino all'esaurimento dell'intervallo di integrazione.

Un'attenzione particolare va dedicata al primo passo per il quale si deve partire da un valore h_1 di tentativo e accorciarlo o allungarlo fino al primo passaggio o, rispettivamente, al primo rifiuto del test di tolleranza.

Per quanto riguarda la stima dell'errore, ci sono vari metodi. Negli esempi che seguono, si e' usato il metodo di estrapolazione di Richardson, del quale saltiamo la descrizione, ed i metodi di Runge-Kutta-Fehlberg che sono descritti nel paragrafo successivo.

Si noti che nell'esempio riportato i metodi di EE ed EI sono sostanzialmente equivalenti nella fase transitoria, mentre hanno un comportamento ben diverso nella fase stazionaria.


TOL=0.2
num. rifiuti=21 num. passi=320
y'=-20(y-sen(t))+cos(t) Metodo di Eulero Esplicito


TOL=0.2
num. rifiuti=8 num. passi=262
y'=-20(y-sen(t))+cos(t) Metodo di Eulero Implicito

Analisi asintotica.

Per quanto riguarda l'andamento asintotico della soluzione, osserviamo che, se $\alpha<0$, allora la soluzione

$$y(t)=e^{\alpha t}(\cos\beta t + i\,\mathrm{sen}\beta t)$$

tende a zero per t che tende a infinito. In altre parole la componente reale e immaginaria di y(t) tendono entrambe a zero. In questo caso si dice che la soluzione è **asintoticamente stabile**.
Se invece $\alpha=0$, allora la soluzione ha modulo costante uguale ad 1, mentre le componenti oscillano periodicamente. Infine, se $\alpha>0$ la soluzione diverge.

Abbiamo visto che i vari metodi numerici, applicati all'equazione test scalare, assumono la forma:

$$y_{n+1}= \varphi(h\lambda)y_n$$

per cui

$$|y_{n+1}|= |\varphi(h\lambda)|\,|y_n| = |\varphi(h\lambda)|^2\,|y_{n-1}| = \ldots = |\varphi(h\lambda)|^{n+1}\,|y_0|.$$

Tale relazione dice che la soluzione numerica ottenuta con passo h costante ha un comportamento asintotico che dipende da $|\varphi(h\lambda)|$ nel seguente modo:

$|\varphi(h\lambda)| < 1 \quad \Rightarrow \quad |y_n|\to 0 \qquad$ per $n\to\infty$

$|\varphi(h\lambda)| = 1 \quad \Rightarrow \quad |y_n| = |y_0| \qquad \forall n$

$|\varphi(h\lambda)| > 1 \quad \Rightarrow \quad |y_n|\to\infty \qquad$ per $n\to\infty$.

Sono interessanti i metodi che risultano asintoticamente stabili per tutte le equazioni che hanno soluzioni asintoticamente stabili, cioè $\alpha<0$ o, equivalentemente, λ nel semipiano negativo.

Dalle considerazioni precedenti risulta che il metodo di Eulero esplicito è asintoticamente stabile solo per quei valori del passo h, tali che $h\lambda$ rimane incluso strettamente nella regione di assoluta stabilità. Viceversa i metodi di Eulero implicito e dei trapezi risultano asintoticamente stabili per ogni valore del passo h, poichè le loro regioni di assoluta stabilità includono l'intero semipiano negativo.

Si osservi infine che il metodo di Eulero implicito ha una regione di assoluta stabilità più ampia del semipiano negativo. Ciò causa, per certi valori del passo, un andamento asintoticamente stabile del metodo anche per equazioni che hanno $\alpha>0$, le cui soluzioni esatte divergono. Questa proprietà, nota come *smorzamento numerico* (numerical damping), è un aspetto negativo del metodo.

Un metodo perfetto, da questo punto di vista, è il metodo dei trapezi la cui soluzione ha, in ogni caso, lo stesso andamento qualitativo della soluzione esatta per ogni passo h.

In particolare, consideriamo l'equazione test con $\lambda=i$ (unita' immaginaria)

$$y'=i\,y$$

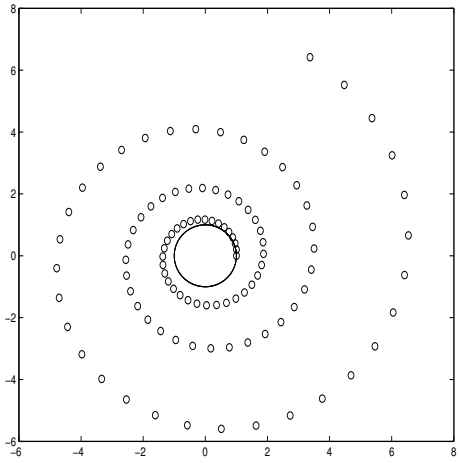
$$y(0)=1$$

la cui soluzione

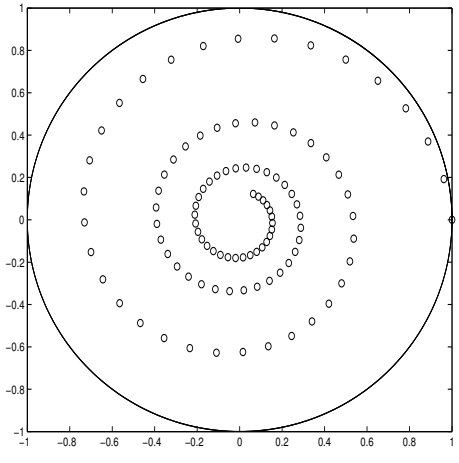
$$y(t)= \cos(t) + i\,\mathrm{sen}(t)$$

descrive un cerchio unitario $|y(t)|=1$ nel piano complesso C.

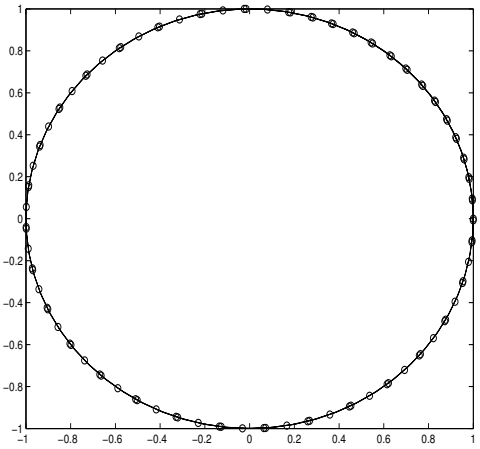
Per ogni metodo numerico, la soluzione e' data dalla sequenza di punti $y_{n+1}=\varphi(i\,h)y_n$. In figura sono riportate le traiettorie per i tre metodi di Eulero esplicito, implicito e dei trapezi. Si osservi che, dei tre, solo il metodo dei trapezi e' capace di conservare il modulo unitario della soluzione numerica, e fornire quindi una traiettoria chiusa ***per ogni scelta del passo*** h. Al contrario, gli altri due metodi forniscono, per ogni passo, una spirale che implode o esplode. Perche?



Eulero Esplicito



Eulero Implicito



Trapezi

Visto che la traiettoria ricavata col metodo dei trapezi descrive un cerchio dello stesso raggio della soluzione esatta, e' corretto dire che il metodo dei trapezi fornisce, in questo caso, la soluzione esatta?