Fondamenti di Calcolo Numerico

——————— ∼ ♢ ∼ ———————

Appunti relativi alla soluzione numerica di un problema di Cauchy

——————— ∼ ♢ ∼ ———————

Claudia Fassino
(fassino@dima.unige.it)

Premessa

Queste dispense riassumono le mie lezioni relative alla soluzione numerica di un problema di Cauchy, svolte nell'ambito del corso di Fondamenti di Calcolo Numerico del Prof. Piana per il corso di Laurea Triennale in Matematica (Università di Genova). Tali appunti sono stati scritti con l'intento di fornire agli studenti una traccia delle lezioni svolte e non vogliono e non possono in alcun modo sostituire i tanti libri sull'argomento presenti in letteratura, quale ad esempio il libro Matematica Numerica di A. Quarteroni, R. Sasso, F. Saleri [1] che ho utilizzato come traccia.

Ringrazio in anticipo tutti coloro (studenti e non) che mi segnaleranno sviste ed errori e daranno suggerimenti per rendere migliori queste dispense.

1 Il problema di Cauchy

Nel seguito vengono descritti alcune classi di metodi numerici per risolvere il seguente problema di Cauchy, consistente nel cercare una funzione y continua e derivabile su un intervallo $I \subset \mathbb{R}$, contenente un punto x_0, tale che

$$\begin{cases} y'(x) &= f(x, y(x)) \ \forall x \in I \\ y(x_0) &= y_0 \end{cases} \tag{1}$$

dove f è una funzione definita e continua su $I \times \mathbb{R}$ a valori in $\mathbb{R}$. Nel seguito si considera $I = [x_0, x_0 + T]$, $T > 0$. Il problema di Cauchy (1) è del

primo ordine in quanto interviene solo la derivata prima di y e richiede una condizione iniziale per evitare di avere più di una soluzione.

Condizione sufficiente per l'esistenza e unicità della soluzione del problema di Cauchy (1) è la Lipschtzianità di f rispetto al secondo argomento.

Definizione 1.1. *Una funzione f definita su $D \subset \mathbb{R}^2$ a valori in $\mathbb{R}$ è Lipschitziana rispetto al secondo argomento se esiste $K > 0$ tale che*

$$|f(x, z_1) - f(x, z_2)| < K\,|z_1 - z_2| \qquad \forall (x, z_1), (x, z_2) \in D$$

Il seguente teorema stabilisce condizioni sufficienti per l'esistenza e l'unicità della soluzione del problema di Cauchy (1).

Teorema 1.1. *Sia $R = \{(x, z) \in \mathbb{R}^2 \,|\, |x - x_0| < a_1,\ |z - y_0| < a_2\}$ e sia f Lipschitziana su R rispetto al secondo argomento.*
Siano $M = \max_{(x,z)\in R} |f(x, z)|$ e $\delta = \min\{a_1, a_2/M\}$.

Allora esiste unica una funzione $y = y(x)$ soluzione del problema (1) su $[x_0 - \delta, x_0 + \delta]$.

Nel seguito si suppone, quindi, che la **funzione f sia Lipschitziana rispetto al secondo argomento** per garantire l'esistenza e l'unicità della soluzione del problema di Cauchy (1).

Si noti che il fatto di considerare un'equazione differenziale del primo ordine non è un'ipotesi troppo restrittiva in quanto un'equazione differenziale di ordine superiore al primo può essere trasformata in un sistema di equazioni differenziali del primo ordine nel modo seguente. A partire da

$$y^{(n)} + \sum_{i=1}^{n-1} \alpha_i y^{(i)} = f(x, y(x))$$

si introducono le seguenti funzioni

$$Y_0(x) = y(x) \quad Y_1(x) = y'(x), \ldots, Y_{n-1}(x) = y^{(n-1)}(x)$$

e si ottiene il sistema differenziale del primo ordine

$$\begin{cases} Y_0'(x) &= Y_1(x) \\ Y_1'(x) &= Y_2(x) \\ &\vdots \\ Y_{n-1}'(x) &= Y_n(x) \\ Y_{n-1}'(x) &= f(x, Y_0(x)) - \sum_{i=1}^{n-1} \alpha_i Y_i \end{cases}$$

Tale sistema può essere risolto con metodi derivanti dalla generalizzazione dei metodi usati per le equazioni.

2 Discretizzazione del problema di Cauchy

I metodi descritti nel seguito non calcolano esplicitamente la funzione soluzione $y(x)$, ma approssimano i valori della funzione y sui punti $\{x_i\}_{i=1,\ldots,N}$ di una griglia, cioè approssimano $y(x_i)$, denotata nel seguito, per brevità, con y_i. Una volta note le approssimazioni di y_i, è possibile ottenere una funzione che approssimi $y(x)$ utilizzando tecniche quali l'interpolazione.

I metodi descritti utilizzano una **griglia** di N punti equidistanti appartenenti all'intervallo $[a,\, b]$, con $a = x_0$, tali che $x_j = x_0 + j*h$, con $h = (b-a)/N$, $j = 0, \ldots, N$.

I metodi costruiscono una successione di valori u_j, $j = 0, \ldots, N$, tali che $u_0 = y_0$ oppure u_0 è uguale ad una approssimazione di y_0, e u_j, $j = 1, \ldots, N$, è un'approssimazione di y_j funzione di h e dei valori $u_0, \ldots, u_{j-1}$.

Un metodo viene detto a **un passo** se per ogni j il termine u_{j+1} dipende solo da u_j, altrimenti il metodo viene detto **a più passi** (multistep). Inoltre un metodo viene detto **esplicito** se u_{j+1} si ricava direttamente da u_j ed **implicito** se u_{j+1} dipende implicitamente da se stesso tramite f.

Valgono le seguenti definizioni.

Definizione 2.1. *Si definisce* **errore totale** *(commesso dall'approssimazione al passo i) la differenza $e_i = u_i - y_i$. Un metodo si dice* **convergente** *se, posto $u_0 = y_0$, si ha*

$$\lim_{h\to 0} e_i = 0 \quad \forall i = 0, \ldots, N.$$

Se $\max |e_i|$ è infinitesimo di ordine p allora si dice che il metodo ha ordine di convergenza p.

Se un metodo è **convergente** allora, al decrescere del passo h l'approssimazione di y_i con u_i migliora ed è possibile, in assenza di errori algoritmici, trovare un passo h tale che l'errore e_i è minore, in valore assoluto, di una soglia prefissata.

Definizione 2.2. *Denotato con u_j^* il valore calcolato dal metodo al passo j utilizzando i valori esatti y_i, $i = 0, \ldots, (j-1)$, si definisce* **errore di troncamento locale** *il rapporto $\tau_j = (y_{j+1} - u_{j+1}^*)/h$. Un metodo si dice* **consistente** *se*

$$\lim_{h\to 0} \tau = 0,$$

*dove $\tau = \max_{i=1\ldots,N} |\tau_i|$ è l'***errore di troncamento globale***. Se τ è infinitesimo di ordine p allora si dice che il metodo ha ordine di consistenza p.*

Se un metodo è **consistente** allora, supponendo di non aver commesso errori ai passi precedenti, l'errore nell'approssimare y_i con u_i tende a zero per h che tende a zero. Ciò significa che il "metodo di approssimazione" utilizzato per costruire la successione $\{u_i\}_{i=1,\dots,N}$ migliora al decrescere del passo h.

Definizione 2.3. *Siano* $\{u_i^{(h)}\}_{i=1,\dots,N}$ *la successione costruita a partire da* $u_0^{(h)} = y_0$ *e* $\{z_i^{(h)}\}_{i=1,\dots,N}$ *la successione costruita a partire da* $z_0^{(h)} = y_0 + \delta$ *ottenuta lo stesso metodo usato per calcolare* $\{u_i^{(h)}\}_{i=1\dots,N}$. *Un metodo si dice* **zero-stabile** *se esistono* $h_0 > 0$ *e* $C > 0$ *tali che*

$$\forall h \in (0, h_0], \qquad |u_i^{(h)} - z_i^{(h)}| \leq C\delta, \qquad i = 1, \dots, N.$$

Se un metodo è **zero-stabile** allora risente poco delle perturbazioni sul valore iniziale y_0, che potrebbe essere noto a meno di una perturbazione δ. La zero-stabilità assicura che è possibile scegliere un passo h tale che la successione ottenuta a partire da un dato iniziale perturbato differisca da quella ottenuta con il valore iniziale esatto per meno di $O(\delta)$.

3 I metodi a un passo

In questo paragrafo vengono descritti alcuni classici **metodi a un passo** e alcune proprietà dei metodi a un passo espliciti. Denoteremo con $f_j = f(x_j, u_j)$.

Tra i metodi a un passo si possono ricordare

- Metodo di Eulero esplicito:

$$u_{j+1} = u_j + hf_j \qquad j = 0, \ldots, (N-1)$$

- Metodo di Eulero implicito:

$$u_{j+1} = u_j + hf_{j+1} \qquad j = 0, \ldots, (N-1)$$

- Metodo del Trapezio (implicito):

$$u_{j+1} = u_j + \frac{h}{2}(f_j + f_{j+1}) \qquad j = 0, \ldots, (N-1)$$

- Metodi θ:

$$u_{j+1} = u_j + h\left(\theta f_j + (1-\theta) f_{j+1}\right) \qquad \theta \in [0,1], \quad j = 0, \ldots, (N-1)$$

  Se $\theta \neq 1$ si ottengono metodi impliciti. Se $\theta = 1$ si ottiene il metodo di Eulero esplicito, se $\theta = 0$ si ottiene il metodo di Eulero implicito, se $\theta = 1/2$ si ottiene il metodo del Trapezio.

- Metodo di Heun (esplicito):

$$u_{j+1} = u_j + \frac{h}{2}\left(f_j + f(x_{j+1}, u_j + hf_j)\right) \qquad j = 0, \ldots, (N-1)$$

  ottenuto dal metodo del Trapezio approssimando u_{j+1} in f con il valore calcolato dal metodo di Eulero esplicito.

3.1 Studio del metodo di Eulero esplicito

Al primo passo, poiché $y'(x_0) = f(x_0, y_0)$, approssimando la derivata con il rapporto incrementale, si ha che

$$f(x_0, y_0) = y'(x_0) \simeq \frac{y_1 - y_0}{h}, \quad \text{cioè} \quad y_1 \simeq y_0 + hf(x_0, y_0).$$

Denotando $u_1 = y_0 + hf(x_0, y_0)$, si ha che u_1 è una approssimazione di y_1. In questo caso, l'unico errore introdotto è dovuto alla scelta del modello in cui la derivata prima è approssimata dal rapporto incrementale.

Dal secondo passo si introduce, oltre all'errore del modello, anche l'errore dovuto al fatto che nessun valore di y sulla griglia è noto esattamente. Infatti si ha che, per ogni $i = 1, \ldots, N$,

$$f(x_i, y_i) = y'(x_i) \simeq \frac{y_{i+1} - y_i}{h} \text{ cioè } y_{i+1} \simeq y_i + hf(x_i, y_i) \simeq u_i + hf(x_i, u_i).$$

Denotando $u_{i+1} = u_i + hf(x_i, u_i)$, si ha che u_i è una approssimazione di y_i.

Si consideri l'errore di troncamento locale per studiare la consistenza del metodo. Innanzi tutto dallo sviluppo di Taylor di y centrato in x_i si ha che

$$y_{i+1} = y_i + hy'(x_i) + \frac{h^2}{2}y''(x_i + \theta_i) = y_i + hf(x_i, y_i) + \frac{h^2}{2}y''(x_i + \theta_i) \text{ con } \theta_i \in [0, 1].$$

Indicato con $u^*_{i+1} = y_i + hf(x_i, y_i)$, cioè supponendo di non aver fatto nessun errore ai passi precedenti, si ha che

$$\tau_i = \frac{y_{i+1} - u^*_{i+1}}{h} = \frac{y_i + hf(x_i, y_i) - y_i - hf(x_i, y_i) + \frac{h^2}{2}y''(x_i + \theta_i)}{h} = \frac{h}{2}y''(x_i + \theta_i).$$

Se y ha le derivate continue fino alla seconda, allora $\tau = \max |\tau_i| \leq h \max_{x \in [a,b]} |y''(x)|$ e quindi $\lim_{h \to 0} \tau = 0$ e il metodo è **consistente di ordine** 1.

Per studiare la zero-stabilità, si consideri la successione $z_{i+1} = z_i + hf(x_i, z_i)$, $i = 0, \ldots, N-1$, tale che $z_0 = y_0 + \delta$. Si ha che $z_{i+1} - u_{i+1} = z_i - u_i + h\left(f(x_i, z_i) - f(x_i, u_i)\right)$ e quindi, poiché f è Lipschitziana rispetto alla costante L,

$$\begin{aligned} |z_{i+1} - u_{i+1}| &\leq |z_i - u_i| + h\,|f(x_i, z_i) - f(x_i, u_i)| \leq |z_i - u_i| + hL|z_i - u_i| \\ &= (1 + hL)|z_i - u_i| \leq (1 + hL)^{i+1}|z_0 - u_0| \leq (1 + hL)^{i+1}\delta. \end{aligned}$$

Poiché $t > 0$, $(1 + t) < \sum_{k=0}^{\infty} \frac{t^k}{k!} = e^t$, e poiché $h(i+1) < hN = a + hN - a = b - a$, si ha che

$$|z_{i+1} - u_{i+1}| \leq e^{Lh(i+1)}\delta \leq e^{(b-a)L}\delta$$

e quindi il metodo è **zero-stabile**.

Per lo studio della convergenza, si consideri l'errore totale $e_{i+1} = y_{i+1} - u_{i+1}$. Dallo studio della consistenza del metodo si ha che lo sviluppo di Taylor di y centrato in x_i può essere riscritto come $y_{i+1} = y_i + hf(x_i, y_i) + h\tau_i$ e quindi

$$\begin{aligned} |e_{i+1}| &= |y_{i+1} - u_{i+1}| = |y_i + hf(x_i, y_i) + h\tau_i - u_i + hf(x_i, u_i)| \\ &\leq |e_i| + hL|e_i| + h|\tau_i| \leq (1 + hL)|e_i| + h\tau \\ &\leq (1 + hL)\left((1 + hL)|e_{i-1}| + h\tau\right) + h\tau = (1 + hL)^2|e_{i-1}| + h\tau(1 + (1 + hL)) \\ &\leq \cdots \leq (1 + hL)^{i+1}|e_0| + h\tau(1 + (1 + hL) + \cdots + (1 + hL)^i). \end{aligned}$$

Poiché $\sum_{k=0}^{p} \alpha^k = \frac{\alpha^{p+1}-1}{\alpha-1}$, si ottiene

$$|e_{i+1}| \leq (1+hL)^{i+1}|e_0| + h\tau\frac{(1+hL)^{i+1}-1}{hL} \leq e^{L(b-a)}|e_0| + \frac{\tau}{L}e^{L(b-a)}$$

Se $e_0 = 0$ oppure se $\lim_{h\to 0} e_0 = 0$ di ordine 1, poiché anche τ è un infinitesimo di ordine 1, si ha che il metodo è **convergente** con ordine di convergenza 1.

3.2 Generalità metodi a un passo espliciti

Si illustrano innanzi tutto alcune proprietà dei metodi ad un passo **espliciti**, che possono essere descritti nel modo seguente [1].

Data una funzione $\Phi = \Phi(x, z, h, f(x, z))$ e fissato il valore u_0, coincidente con y_0 o con una sua approssimazione, un metodo ad un passo esplicito è descritto da

$$u_{i+1} = u_i + h\Phi(x_i, u_i, h, f_i), \qquad i = 0, \ldots, N-1., \quad \text{dove} \quad f_i = f(x_i, u_i) \tag{2}$$

Dallo sviluppo di Taylor della funzione $y(x)$ centrato in x_j e dall'equazione differenziale del problema di Cauchy (1) si ha che

$$y_{j+1} = y_j + hy'(x_j) + h^2y''(\eta_j)/2 = y_j + hf(x_j, y_j) + h^2y''(\eta_j)/2 \tag{3}$$

con $\eta_j \in [x_j, x_{j+1}]$. Dalla definizione 2.2 si ha che l'errore di troncamento locale $\tau_{j+1} = \frac{y_{j+1}-u^*_{j+1}}{h}$, con $u^*_{j+1} = y_j + h\Phi(x_j, y_j, h, f(x_j, y_j))$. Si ottiene quindi, per sostituzione,

$$\tau_{j+1} = f(x_j, y_j) - \Phi(x_j, y_j, h, f) + hy''(\eta_j)/2.$$

L'errore di troncamento valuta quindi la bontà del modello utilizzato per approssimare l'equazione differenziale. Se il metodo è consistente allora il $\lim_{h\to 0} \tau_j = 0$ e quindi $\lim_{h\to 0} \Phi(x_j, y_j, h, f) = f(x_j, y_j)$. Questo significa che Φ può considerarsi una buona approssimazione di f perché, al decrescere di h, Φ tende a f.

Inoltre, utilizzando la definizione di errore di troncamento locale, si ha che $y_{i+1} = u^*_{i+1} + h\tau_i$, cioè

$$y_{i+1} = y_i + h\Phi(x_i, y_i, h, f(x_i, y_i)) + h\tau_i \tag{4}$$

Nel seguito si suppone che Φ sia **Lipschitziana rispetto al secondo argomento**, cioè esistono $h_0 > 0$ e $\Lambda > 0$ tali che $\forall h \in (0, h_0]$ si ha

$$|\Phi(x, u, h, f(x, u)) - \Phi(x, z, h, f(x, z))| \leq \Lambda|u - z|.$$

Dimostriamo il legame tra la zero-stabilità, la consistenza e la convergenza di un metodo esplicito a un passo. Vale il seguente teorema.

Teorema 3.1. *Dato il metodo ad un passo esplicito definito da (2), se* $\Phi = \Phi(x, z, h, f)$ *è Lipschitziana rispetto al secondo argomento* z*, allora*

1. *il metodo è* **zero-stabile***;*
2. *se il metodo è consistente e se* $\lim_{h\to 0} |y_0 - u_0| = 0$*, allora il metodo è* **convergente***. Inoltre, se* $|y_0 - u_0| = O(h^p)$ *e se il metodo ha ordine di consistenza* p*, allora il metodo ha ordine di convergenza* p*.*

Dimostrazione.

1. Si considerino le due successioni $u_{i+1} = u_i + h\Phi(x_i, u_i, h, f_i)$, tale che $u_0 = y_0$ e $v_{i+1} = v_i + h\Phi(x_i, v_i, h, f(x_i, v_i))$, tale che $z_0 = y_0 + \delta$, $i = 0, \ldots, N-1$. Si ha che $v_{i+1} - u_{i+1} = v_i - u_i + h\left(\Phi(x_i, v_i, h, f(x_i, v_i)) - \Phi(x_i, u_i, h, f(x_i, u_i))\right)$ e quindi, poiché Φ è Lipschitziana rispetto al secondo argomento allora esistono h_0 e Λ tali che

$$\begin{aligned} |v_{i+1} - u_{i+1}| &\leq |v_i - u_i| + h\Lambda|v_i - u_i| = (1 + h\Lambda)|v_i - u_i| \\ &\leq (1 + h\Lambda)^{i+1}|v_0 - u_0| \leq (1 + h\Lambda)^{i+1}\delta \leq e^{(b-a)\Lambda}\delta \end{aligned}$$

   per ogni $h \in (0, h_0]$ e quindi il metodo è zero-stabile.

2. Per lo studio della convergenza, per l'equazione (4) si ha che $y_{i+1} = y_i + h\Phi(x_i, y_i, h, f(x_i, y_i)) + h\tau_i$. Poiché $u_{i+1} = u_i + h\Phi(x_i, u_i, h, f(x_i, u_i))$, si ottiene che l'errore totale può essere maggiorato come segue

$$\begin{aligned} |e_{i+1}| &= |y_{i+1} - u_{i+1}| \\ &= |y_i + h\Phi(x_i, y_i, h, f(x_i, y_i)) + h\tau_i - u_i - h\Phi(x_i, u_i, h, f(x_i, u_i))| \\ &\leq |y_i - u_i| + h|\Phi(x_i, y_i, h, f(x_i, y_i)) - \Phi(x_i, u_i, h, f(x_i, u_i))| + h\tau \\ &\leq |y_i - u_i| + h\Lambda|y_i - u_i| = (1 + h\Lambda)|e_i| + h\tau. \end{aligned}$$

   Procedendo analogamente a quanto fatto per il metodo di Eulero esplicito si ottiene

$$|e_{i+1}| = e^{\Lambda(b-a)}\left(|e_0| + \frac{\tau}{\Lambda}\right),$$

   da cui, se $\lim_{h\to 0} |y_0 - u_0| = 0$ e se il metodo è consistente, cioè se $\lim_{h\to 0} \tau = 0$, allora $\lim_{h\to 0} |e_j| = 0$ e il metodo è convergente. Ovviamente l'ordine di infinitesimo di τ e di $y_0 - u_0$ si ripercuote sull'ordine di infinitesimo dell'errore globale. □

Si noti che, nel caso del metodo di Eulero esplicito, essendo $\Phi = f$ con f Lipschitziana, il teorema precedente garantisce la zero-stabilità del metodo.

3.3 Metodo di Heun

Il metodo di Heun è un metodo a un passo esplicito, definito da

$$u_{j+1} = u_j + \frac{h}{2}\left(f_j + f(x_{j+1}, u_j + hf_j)\right) \qquad j = 0, \ldots, (N-1)$$

In questo caso la funzione $\Phi(x, z, h, f(x, z)) = \frac{1}{2}\left(f(x, z) + f(x + h, z + hf(x, z))\right)$ che risulta Lipschitziana rispetto al secondo argomento poiché f è Lipschitziana rispetto al secondo argomento. Infatti si ha che

$$\begin{aligned}
&|\Phi(x, z, h, f(x, z)) - \Phi(x, v, h, f(x, v))| = \\
&\frac{1}{2}|(f(x, z) + f(x + h, z + hf(x, z)) - f(x, v) - f(x + h, v + hf(x, v))| \\
&\leq \frac{1}{2}\left(|(f(x, z) - f(x, v)| + |f(x + h, z + hf(x, z)) + f(x + h, v + hf(x, v))|\right) \\
&\leq \frac{1}{2}\left(L|z - v| + L|z + hf(x, z) - v + hf(x, v)|\right) \\
&\leq L|z - v| + \frac{1}{2}Lh|f(x, z) - f(x, v)| \leq L|z - v| + \frac{1}{2}L^2h|z - v| = (L + \frac{1}{2}L^2h)|z - v|
\end{aligned}$$

Perciò, fissato h_0 esiste $\Lambda = (L + \frac{1}{2}L^2h_0)$ tale che, $\forall h \in (0, h_0]$ si ha

$$|\Phi(x, z, h, f(x, z)) - \Phi(x, v, h, f(x, v))| \leq \Lambda|z - v|$$

cioè Φ è Lipschitziana rispetto al secondo argomento. Da ciò segue che il metodo di Heun è **zero-stabile**.

Inoltre si dimostra che il metodo è consistente di ordine 1 e quindi è anche **convergente** di ordine 1. Per la dimostrazione della consistenza si consideri l'errore di troncamento locale

$$\begin{aligned}
\tau_i &= \frac{y_{i+1} - u^*_{i+1}}{h} = \frac{y_i + hf(x_i, y_i) + \frac{h^2}{2}y''(x_i) + \frac{h^3}{6}y^{(3)}(\eta_i) - y_i - h\Phi(x_{i+1}, y_i + hf(x_i, y_i))}{h} \\
&= f(x_i, y_i) + \frac{h}{2}y''(x_i) + \frac{h^2}{6}y^{(3)}(\eta_i) - \frac{1}{2}\left(f(x_i, y_i) + f(x_i + h, y_i + hf(x_i, y_i))\right) \\
&= \frac{1}{2}\left(hy''(x_i) + \frac{h^2}{3}y^{(3)}(\eta_i) + f(x_i, y_i) - f(x_i + h, y_i + hf(x_i, y_i)) \pm f(x_i + h, y_{i+1})\right)
\end{aligned}$$

Poiché, per ogni k, $f(x_k, y_k) = y'(x_k)$, si ha che

$$f(x_{i+1}, y_{i+1}) - f(x_i, y_i) = y'(x_{i+1}) - y'(x_i) = hy''(x_i) + \frac{h^2}{2}y^{(3)}(\mu_i)$$

e quindi

$$\begin{aligned}
|\tau_i| &\leq \frac{1}{2}\left|hy''(x_i) + \frac{h^2}{3}y^{(3)}(\eta_i) + f(x_i, y_i) - f(x_i + h, y_{i+1})\right. \\
&+ \left. f(x_i + h, y_{i+1}) - f(x_i + h, y_i + hf(x_i, y_i)|\right. \\
&\leq \frac{1}{2}\left|\frac{h^2}{3}y^{(3)}(\eta_i) - \frac{h^2}{2}y^{(3)}(\mu_i)\right| + L\left|y_{i+1} - y_i - hf(x_i, y_i))|\right. \\
&= \frac{1}{2}\left|\frac{h^2}{3}y^{(3)}(\eta_i) - \frac{h^2}{2}y^{(3)}(\mu_i)\right| + L\frac{h^2}{2}\left|y''(\eta_i)|\right.
\end{aligned}$$

Ne segue che τ è un infinitesimo di ordine 2 e quindi il metodo di Heun è **consistente** di **ordine 2**.

4 Metodi di Runge-Kutta espliciti

I metodi di Runge-Kutta sono metodi a un passo che raggiungono un ordine maggiore di 1 utilizzando più valutazioni della funzione f ad ogni passo. Per ricavare i metodi di Runge-Kutta si passa dalla forma differenziale alla forma integrale del problema di Cauchy, cioè si esprime la soluzione nella forma

$$y(x) = y_0 + \int_{x_0}^{x} y'(s)ds = y_0 + \int_{x_0}^{x} f(s, y(s))ds$$

e si applicano tecniche numeriche, quali le formule di quadratura, per approssimare l'integrale.

In generale i metodi di Runge-Kutta sono descritti dalla formula

$$u_{j+1} = u_j + hF(x_j, u_j, h, f) \qquad \text{con} \tag{5}$$

$$F = \sum_{i=1}^{s} b_i K_i \qquad \text{e} \tag{6}$$

$$K_i = f(x_j + c_i h, u_j + h \sum_{r=1}^{s} a_{ir} K_r) \qquad i = 1, \ldots, s \tag{7}$$

dove s indica il numero degli **stadi** del metodo.

Si noti che i metodi di Runge-Kutta sono tutti metodi a un passo espliciti, nel senso che $F(x_j, u_j, h, f)$ non dipende da u_{j+1}. In questo ambito si usa il termite esplicito per indicare un'altra proprietà del metodo e cioè la proprietà che ogni K_i dipende solo dai valori $K_1, \ldots, K_{i-1}$. Il termine "esplicito" riguarda quindi il calcolo di ogni K_i e non il calcolo di U_{j+1} (essendo sottinteso che per ogni metodo di Runge-Kutta tale calcolo è esplicito).

Nel seguito si considerano solo metodi di Runge-Kutta espliciti e tali che $c_i = \sum_{r=1}^{s} a_{ir}$.

Per i metodi Runge-Kutta valgono le proprietà seguenti.

1. se f è Lipschitziana rispetto alla seconda variabile si dimostra per induzione su i che anche ogni K_i è Lipschitziana rispetto alla seconda variabile. Ne segue che tale proprietà è soddisfatta anche dalla funzione F.

2. Poiché i metodi di Runge-Kutta sono tutti metodi a un passo espliciti, se f è Lipschitziana rispetto alla seconda variabile, allora i metodi sono zero-stabili (vedi teorema 3.1).

3. I metodi di Runge-Kutta sono consistenti se e solo se $\sum_{i=1}^{s} b_i = 1$. Infatti, posto $u^*_{j+1} = y_j + hF(x_j, y_j, h, f)$ si ha che

$$\tau_j = \frac{u^*_{j+1} - y_{j+1}}{h} = F(x_j, y_j, h, f) - f(x_j, y_j) - \frac{h}{2} y''(\eta_j)$$

Poiché, per la continuità delle funzioni, $\lim_{h\to 0} K_i = f(x_j, y_j)$, per ogni i, si ha che $\lim_{h\to 0} F(x_j, y_j, h, f) = \sum_{i=1}^{s} b_i f(x_j, y_j)$. Si conclude che $\lim_{h\to 0} \max|\tau_j| = 0$ se e solo se $\sum_{i=1}^{s} b_i = 1$.

4. Se f è Lipschitziana rispetto alla seconda variabile ogni metodo consistente è anche convergente (in quanto zero-stabile, vedi teorema 3.1).

5. Un metodo di Runge-Kutta esplicito a s stadi non può avere ordine maggiore di s. Inoltre non esistono metodi Runge-Kutta espliciti a s stadi di ordine s se $s \geq 5$.

4.1 Esempi di metodi di Runge-Kutta espliciti

In questo paragrafo vengono brevemente descritti alcuni metodi di Runge-Kutta nel caso in cui la funzione f del problema di Cauchy (1) sia Lipschitziana rispetto alla seconda variabile.

Innanzi tutto si noti che il metodo di **Eulero esplicito** può essere visto come metodo Runge-Kutta con $u_{j+1} = u_j + hK_1$, con $K_1 = f(x_j, u_j)$. È un metodo a uno stadio di ordine 1.

Il metodo di **Eulero modificato** è un metodo di Runge-Kutta a 2 stadi, definito da

$$u_{j+1} = u_j + hK_2 \qquad \text{con}$$

$$K_1 = f(x_j, u_j) \qquad \text{e} \qquad K_2 = f(x_j + \frac{h}{2}, u_j + \frac{h}{2}K_1)$$

Poichè $\sum_{i=1}^{2} b_i = 0 + 1 = 1$ il metodo è consistente e quindi convergente. Tale schema risulta essere, inoltre, di ordine 2.

Il metodo di **Heun** è un metodo di Runge-Kutta a 2 stadi, definito da

$$u_{j+1} = u_j + \frac{h}{2}(K_1 + K_2)) \qquad \text{con}$$

$$K_1 = f(x_j, u_j) \qquad \text{e} \qquad K_2 = f(x_j + h, u_j + hK_1)$$

Poichè $\sum_{i=1}^{2} b_i = 0.5 + 0.5 = 1$ il metodo è consistente e quindi convergente. Tale schema risulta essere, inoltre, di ordine 2.

Infine si consideri metodo di Runge-Kutta 4 stadi descritto da

$$u_{j+1} = u_j + \frac{h}{6}(K_1 + 2K_2 + 2K_3 + K_4)$$

con

$$\begin{aligned} K_1 &= f(x_j, u_j) \\ K_2 &= f(x_j + \frac{h}{2}, u_j + \frac{h}{2}K_1) \\ K_3 &= f(x_j + \frac{h}{2}, u_j + \frac{h}{2}K_2) \\ K_4 &= f(x_j + h, u_j + hK_3) \end{aligned}$$

Poichè $\sum_{i=1}^{4} b_i = 1$ il metodo è consistente e quindi convergente. Tale schema risulta essere, inoltre, di ordine 4.

5 I metodi a più passi (multistep) lineari

Un metodo si dice a più passi (lineare) se è definito nel modo seguente:

$$u_{n+1} = \sum_{j=0}^{p} a_j u_{n-j} + h \sum_{j=-1}^{p} b_j f(x_{n-j}, u_{n-j}) \tag{8}$$

$$\tag{9}$$

In tal caso il metodo si dice a $p+1$ passi in quanto ogni u_{n+1} dipende dai $p+1$ valori precedenti $u_n, \ldots, u_{n-p}$. Se $b_{-1} = 0$ il metodo è esplicito, altrimenti è implicito.

Per "inizializzare" il metodo sono necessari $p+1$ valori iniziali $u_0, \ldots, u_p$. Come ovvio, si fissa $u_0 = y_0$ e si possono approssimare gli altri valori, utilizzando, ad esempio, metodi a un passo.

Si noti che, se $p = 0$, si ritrovano i metodi a un passo:

- con $b_{-1} = 0$, $a_0 = b_0 = 1$ Eulero esplicito $u_{n+1} = u_n + hf_n$;
- con $b_{-1} = a_0 = 1$, $b_0 = 0$ Eulero implicito $u_{n+1} = u_n + hf_{n+1}$;
- con $b_{-1} = b_0 = 0.5$, $a_0 = 1$ il metodo del Trapezio $u_{n+1} = u_n + \frac{h}{2}(f_n + f_{n+1})$.

5.1 Consistenza

Per lo studio della consistenza si viste come nel caso dei metodi a un passo e si definisce l'errore di troncamento locale

$$\tau_{n+1} = \frac{y(x_{n+1}) - u^*_{n+1}}{h},$$

dove u^*_{n+1} è calcolato, mediante il metodo in esame, a partire dai valori esatti $y(x_n), \ldots, y(x_{n-p})$.

Teorema 5.1. *Un metodo multistep (8) è consistente se e solo se*

$$\sum_{j=0}^{p} a_j = 1 \qquad \text{e} \qquad \sum_{j=-1}^{p} b_j - \sum_{j=0}^{p} j a_j = 1. \tag{10}$$

Inoltre, se $y \in \mathcal{C}^{q+1}([a,b])$, e se valgono, oltre la (10), anche

$$\sum_{j=0}^{p} (-j)^i a_j + i \sum_{j=-1}^{p} (-j)^{i-1} b_j = 1 \qquad i = 2, \ldots, q$$

il metodo ha ordine di consistenza q.

Dimostrazione. Si ha che

$$\tau_{n+1} = \frac{y(x_{n+1}) - \sum_{j=0}^{p} a_j y(x_{n-j}) - h\sum_{j=-1}^{p} b_j f(x_{n-j}, y(x_{n-j}))}{h} \quad (11)$$

$$= \frac{y(x_{n+1}) - \sum_{j=0}^{p} a_j y(x_{n-j}) - h\sum_{j=-1}^{p} b_j y'(x_{n-j})}{h} \quad (12)$$

Si considerino gli sviluppi di Taylor delle funzioni $y(x)$ e $y'(x)$ centrati nel punto x_{n-p} e valutati in x_{n-j}, $j = -1, \ldots, p$

$$y(x_{n-j}) = y(x_{n-p}) + h(p-j)y'(x_{n-p}) + \frac{h^2}{2}(p-j)^2 y''(\eta_j)$$

$$y'(x_{n-j}) = y'(x_{n-p}) + h(p-j)y''(x_{n-p}) + \frac{h^2}{2}(p-j)^2 y^{(3)}(\mu_j)$$

Sostituendo tali sviluppi nell'equazione (**??**) si ottiene

$$h\tau_{n+1} = y(x_{n-p}) + h(p+1)y'(x_{n-p}) + \frac{h^2}{2}(p+1)^2 y''(\eta_{-1})$$
$$- \sum_{j=0}^{p} a_j \left(y(x_{n-p}) + h(p-j)y'(x_{n-p}) + \frac{h^2}{2}(p-j)^2 y''(\eta_j) \right)$$
$$- h\sum_{j=-1}^{p} b_j \left(y'(x_{n-p}) + h(p-j)y''(x_{n-p}) + \frac{h^2}{2}(p-j)^2 y^{(3)}(\mu_j) \right)$$

e quindi

$$h\tau_{n+1} = y(x_{n-p}) \left(1 - \sum_{j=0}^{p} a_j \right) + hy'(x_{n-p}) \left((p+1) - \sum_{j=0}^{p} a_j(p-j) - \sum_{j=-1}^{p} b_j \right)$$
$$+ h^2 (\ldots) .$$

Si conclude che $\max \tau_n$ tende a 0 per $h \to 0$ se e solo se $1 - \sum_{j=0}^{p} a_j = 0$ e $(p+1) - \sum_{j=0}^{p} a_j(p-j) - \sum_{j=-1}^{p} b_j = 0$, cioè se e solo se $\sum_{j=0}^{p} a_j = 1$ e se $(p+1) - p\sum_{j=0}^{p} a_j + \sum_{j=0}^{p} j a_j - \sum_{j=-1}^{p} b_j = 0$. La seconda relazione può essere riscritta sfruttando la condizione $\sum_{j=0}^{p} a_j = 1$ e si ottiene $1 + \sum_{j=0}^{p} j a_j - \sum_{j=-1}^{p} b_j = 0$ e si conclude.

La dimostrazione relativa all'ordine q è analoga, e si ottiene considerando lo sviluppo di y e y' fino all'ordine $q+1$. □

Osservazione. Per quanto riguarda il metodi a un passo, poiché in tal caso $p = 0$, le relazioni (10) sono più semplici. Si ha che un metodo lineare a un passo è consistente se e solo se $a_0 = 1$ e $b_{-1} + b_0 = 1$. Inoltre il metodo è dell'ordine 1 se $2b_{-1} \neq 1$ e di ordine 2 altrimenti. Da tali relazioni segue che i metodi di Eulero esplicito ed implicito sono consistenti di ordine 1 e che il metodo del Trapezio è consistente di ordine 2.

5.2 Zero-stabilità

Il concetto di zero-stabilità per i metodi a $p+1$ passi coincide sostanzialmente con quello per i metodi a 1 passo.

Definizione 5.1. *Il metodo (8) a $p+1$ passi è zero-stabile se $\exists h_0 > 0$ ed $\exists C > 0$ tali che $\forall h \in (0, h_0]$ $|z_n - u_n| < C\varepsilon$, dove z_n è calcolata partendo dai valori $z_k = u_k + \delta_k$ e $|\delta_k| < \varepsilon$, $k = 0, \ldots, p$.*

La zero-stabilità per i metodi a $p+1$ passi può essere determinata usando il seguente teorema (che non viene dimostrato in questo ambito).

Teorema 5.2. *Un metodo (8) a $p+1$ passi consistente è zero-stabile se e solo se viene soddisfatta la condizione sulle radici, cioè se e solo se il polinomio $\rho(r) = r^{p+1} - \sum_{j=0}^{p} a_j r^{p-j}$ ha radici r tali che $|r| < 1$ oppure $|r| = 1$ e $\rho'(r) \neq 0$, cioè r ha molteplicità 1.*

Nel caso dei metodi a un passo ($p = 0$), si ha che $\rho(z) = r - 1$ e quindi poichè viene soddisfatta la condizione sulle radici un metodo a un passo (lineare) consistente è sempre zero-stabile.

Sono quindi zero-stabili i metodi di Eulero esplicito ed implicito e il metodo del Trapezio.

5.3 Convergenza

Teorema 5.3. *Il metodo (8) a $p+1$ passi consistente è convergente se e solo se è zero-stabile l'errore sui dati iniziali è un infinitesimo rispetto ad h. L'ordine di convergenza del metodo è q se l'ordine di consistenza è q e se l'errore sui dati iniziali è un infinitesimo di ordine q rispetto ad h.*

Si noti che la consistenza del metodo equivale al fatto che vengano soddisfatte le condizioni (10) e la zero-stabilità equivale alla condizione sulle radici.

5.4 Esempi di metodi a 2 passi

Metodo Mid-Point (esplicito) con $p = 1$:

$$u_{n+1} = u_{n-1} + 2hf_n.$$

Il metodo è definito scegliendo $a_0 = 0$, $a_1 = 1$, $b_{-1} = 0$, $b_0 = 2$ e $b_1 = 0$ ed è tale che:

- esplicito perché $b_{-1} = 0$;

- consistente perché $\sum_{j=0}^{1} a_j = a_0 + a_1 = 1$ e $\sum_{j=-1}^{1} b_j - \sum_{j=0}^{1} j a_j = b_{-1} + b_0 + b_1 - a_1 = 2 - 1 = 1$; inoltre, se $y \in \mathcal{C}^3([a,b])$, ha ordine di consistenza 2 perché $\sum_{j=0}^{1}(-j)^2 a_j + 2\sum_{j=-1}^{1}(-j)b_j = a_1 = 1$ e la stessa relazione non vale per $i = 3$.

- zero-stabile perchè $\rho(r) = r^2 - a_1 = r^2 - 1$ che soddisfa la condizione delle radici.

- convergente di ordine 2 perché è consistente di ordine 2 e zero-stabile.

Metodo di Adams-Bashforth (esplicito) con $p = 1$:

$$u_{n+1} = u_n + \frac{h}{2}(3f_n - f_{n-1}).$$

Il metodo è definito scegliendo $a_0 = 1$, $a_1 = 0$, $b_{-1} = 0$, $b_0 = \frac{3}{2}$ e $b_1 = -\frac{1}{2}$ ed è tale che:

- esplicito perché $b_{-1} = 0$;

- consistente perché $\sum_{j=0}^{1} a_j = a_0 + a_1 = 1$ e $\sum_{j=-1}^{1} b_j - \sum_{j=0}^{1} j a_j = b_{-1} + b_0 + b_1 - a_1 = \frac{3}{2} - \frac{1}{2} = 1$; inoltre, se $y \in \mathcal{C}^3([a,b])$, ha ordine di consistenza 2 perché $\sum_{j=0}^{1}(-j)^2 a_j + 2\sum_{j=-1}^{1}(-j)b_j = a_1 + 2(b_{-1} - b_1) = 1$ e la stessa relazione non vale per $i = 3$.

- zero-stabile perchè $\rho(r) = r^2 - a_0 r - a_1 = r^2 - r$ che soddisfa la condizione delle radici.

- convergente di ordine 2 perché è consistente di ordine 2 e zero-stabile.

Metodo di Adams-Moulton (implicito) con $p = 1$:

$$u_{n+1} = u_n + \frac{h}{12}(5f_{n+1} + 8f_n - f_{n-1}).$$

Il metodo è definito scegliendo $a_0 = 1$, $a_1 = 0$, $b_{-1} = \frac{5}{12}$, $b_0 = \frac{8}{12}$ e $b_1 = -\frac{1}{12}$ ed è tale che:

- implicito perché $b_{-1} \neq 0$;

- consistente perché $\sum_{j=0}^{1} a_j = a_0 + a_1 = 1$ e $\sum_{j=-1}^{1} b_j - \sum_{j=0}^{1} j a_j = b_{-1} + b_0 + b_1 - a_1 = 1$; inoltre, se $y \in \mathcal{C}^4([a,b])$, ha ordine di consistenza 3 perché

$$\sum_{j=0}^{1} (-j)^2 a_j + 2 \sum_{j=-1}^{1} (-j) b_j = a_1 + 2(b_{-1} - b_1) = 1 \qquad \text{per } i = 2$$

$$\sum_{j=0}^{1} (-j)^3 a_j + 3 \sum_{j=-1}^{1} (-j)^2 b_j = -a_1 + 3(b_{-1} + b_1) = 1 \qquad \text{per } i = 3$$

- zero-stabile perchè $\rho(r) = r^2 - a_0 r - a_1 = r^2 - r$ che soddisfa la condizione delle radici.

- convergente di ordine 3 perché è consistente di ordine 3 e zero-stabile.

6 Metodi predictor-corrector

Noti i valori $u_0, u_1, \ldots, u_n$, per calcolare il valore u_{n+1} con un metodo multistep lineare implicito, è necessario risolvere un problema di punto fisso del tipo $u_{n+1} = \Psi(u_{n+1})$ dove

$$\Psi(v) = \sum_{j=0}^{p} a_j u_{n-j} + h \sum_{j=0}^{p} b_j f(u_{n-j}) + h b_{-1} f(v). \tag{13}$$

Un generico problema di punto fisso $x = g(x)$ ammette unica soluzione se la funzione g è, in un intorno del punto fisso, una contrazione, cioè se $|g(x) - g(y)| < K|x - y|$ con $K < 1$ in tale intorno.

Supponendo che f sia una funzione lipschitziana di costante L su $[a, b]$, per quanto riguarda i metodi multistep lineari impliciti si ha

$$\begin{aligned}
|\Psi(u_{n+1}) - \Psi(v_{n+1})| &= \Bigg| \sum_{j=0}^{p} a_j u_{n-j} + h \sum_{j=0}^{p} b_j f(u_{n-j}) + h b_{-1} f(u_{n+1}) \\
&- \sum_{j=0}^{p} a_j u_{n-j} - h \sum_{j=0}^{p} b_j f(u_{n-j}) - h b_{-1} f(v_{n+1}) \Bigg| = h|b_{-1}||f(u_{n+1}) - f(v_{n+1})| \\
&\leq Lh|b_{-1}||u_{n+1} - v_{n+1}|
\end{aligned}$$

da cui si ottiene una contrazione scegliendo h tale che $Lh|b_{-1}| < 1$, cioè $h < \frac{1}{L|b_{-1}|}$.

Un metodo per approssimare la soluzione di un problema di punto fisso è il seguente. Scelto un valore iniziale x_0 si costruisce una successione $x_{n+1} = g(x_n)$. Per il problema associato ai metodi multistep, scelto un valore iniziale $u_{n+1}^{(0)}$, si approssima u_{n+1} mediante la successione $u_{n+1}^{(i+1)} = \Psi(u_{n+1}^{(i)})$, supponendo di aver già calcolato $u_0, u_1, \ldots, u_n$.

Per il calcolo di ogni valore u_{n+1} il metodo **Predictor-Corrector** si basa sui seguenti passi:

- calcolo del valore $u_{n+1}^{(0)}$ mediante un metodo multistep esplicito a partire da valori $u_0, u_1, \ldots, u_n$ (il metodo, in tale ambito, viene chiamato **predictor** e il passo indicato con P);
- calcolo di $f(u_{n+1}^{(0)})$ (il passo vene indicato con E);
- calcolo di $u_{n+1}^{(1)}$ mediante mediante la formula (13) (il metodo, in tale ambito, viene chiamato **corrector** e il passo indicato con C);
- calcolo $f(u_{n+1}^{(1)})$ che verrà utilizzato per valutare u_{n+2} (passo E).

Il metodo precedente viene brevemente indicato come $PECE$. In realtà le fasi E e C possono essere calcolate piú volte in sequenza, se vengono applicati più passi del metodo iterativo per approssimare il punto fisso. In tal caso il metodo Predictor-Corrector viene schematizzato come $P(EC)^m E$.

Vale il seguente teorema [1].

Teorema 6.1. *Se il metodo predictor ha ordine di convergenza $\widetilde{q}$ e il metodo corrector ha ordine di convergenza q, allora*

se $\widetilde{q} \geq q$ allora il metodo Predictor-Corrector ha ordine q;

se $\widetilde{q} \leq q$ e $m \geq q - \widetilde{q}$ allora il metodo Predictor-Corrector ha ordine q;

se $\widetilde{q} \leq q$ e $m < q - \widetilde{q}$ allora il metodo Predictor-Corrector ha ordine $\widetilde{q} + m < q$.

6.1 Esempi di metodi Predictor-Corrector

Il metodo di Heun presentato nei paragrafi precedenti è un Predictor-Corrector derivante dal metodo di Eulero esplicito e dal metodo implicito del Trapezio.

Un altro esempio di metodo Predictor-Corrector è dato dall'uso del metodo esplicito di Adams-Bashforth (di ordine $\widetilde{q} = 2$) e dal metodo implicito di Adams-Moulton (di ordine $q = 3$). Usando il teorema precedente si ha che il metodo Predictor-Corrector ha ordine q se si sceglie $m \geq q - \widetilde{q} = 1$.

I passi del Predictor-Corrector sono dati da:

[P]: $u_{n+1}^{(0)} = u_n + \frac{h}{2}(3f_n - f_{n-1})$.

[E]: valutazione di $f(x_{n+1}, u_{n+1}^{(0)})$.

[C]: $u_{n+1}^{(1)} = u_n + \frac{h}{12}(5f(x_{n+1}, u_{n+1}^{(0)}) + 8f_n - f_{n-1})$.

7 Assoluta stabilità

Dallo studio della convergenza e della zero-stabilità si ricava che, a patto di scegliere h sufficientemente piccolo, l'errore dell'approssimazione di y_n con u_n è minore di una soglia prefissata piccola a piacere. Tuttavia, nella pratica, non è possibile scegliere un passo h troppo piccolo, perchè non si possono eseguire troppi passi, sia per motivi di tempo richiesto dal calcolo che per motivi legati agli errori algoritmici. D'altra parte, valori di h troppo grandi introducono errori di approssimazione troppo elevati. Si vuole allora scegliere valori di h non troppo piccoli, per limitare il numero di passi, e non troppo grandi per evitare errori elevati. Per testare un passo h è utile quindi affiancare allo studio della zero-stabilità, lo studio della assoluta stabilità, concetto legato al comportamento asintotico delle approssimazioni u_n per h fissato e $x_n \to \infty$, contrariamente a quanto fatto per la zero-stabilità, analizzata con l'intervallo $[x_0, b]$ fissato e h che tende a zero. Si richiede che, per h fissato, u_n sia asintoticamente limitata, cioè non si presenta un accumulo "anomalo" di errori. Più formalmente viene data la seguente definizione.

Definizione 7.1. *Si consideri il* **problema modello**

$$\begin{cases} y'(x) &= \lambda y(x), \qquad \lambda \in \mathbb{C},\ Re(\lambda) < 0 \\ y(0) &= 1 \end{cases}$$

dove $Re(\lambda)$ è la parte reale del numero complesso λ. Un metodo numerico si dice **assolutamente stabile** *se, applicato al problema modello con λ e h fissati, permette di calcolare la successione $\{u_n\}$ tale che $\lim_{n\to\infty} u_n = 0$.*

L'insieme $\mathcal{A} = \{z = h\lambda \in \mathbb{C} \,|\, $ il metodo è assolutamente stabile rispetto a h e $\lambda\}$ si dice **regione di assoluta stabilità**.

Osservazione.

- Poiché la soluzione esatta del problema modello $y(x) = e^{\lambda x}$ tende a 0 per $x \to \infty$, richiedere che anche $\lim_{n\to\infty} u_n = 0$ significa richiedere che gli errori introdotti siano tali da non perturbare il comportamento asintotico della soluzione approssimata rispetto alla soluzione esatta.

- Uno dei criteri con cui scegliere un metodo numerico rispetto ad un altro è l'ampiezza della regione di assoluta stabilità. Si preferiscono i metodi che, a parità di λ permettono la scelta di valori di h più elevati.

- Se si esamina un problema modello si poò scegliere un corretto passo h che garantisca l'assoluta stabilità. Tuttavia, la scelta è di h è necessaria

quando si deve risolvere un generico problema di Cauchy (1) che non coinvolge il valore di λ. In tal caso un buon candidato per λ è un parametro μ tale che $\mu \simeq \frac{\partial f(x,y)}{\partial y}(x, y(x))$, $x \in [a, b]$, se tale parametro esiste.

7.1 Assoluta stabilità per i metodi a un passo: un esempio

Si consideri il metodo di Heun, esplicito a un passo, applicato al problema modello, cioè si consideri il caso in cui $f(x, y) = \lambda y$. Poiché $f_n = \lambda u_n$ e $f(x_n, u_n + hf_n) = \lambda(u_n + hf_n)$, si ottiene

$$u_{n+1} = u_n + \frac{h}{2}(\lambda u_n + \lambda(u_n + h\lambda u_n)) = \left(1 + h\lambda + \frac{h^2\lambda^2}{2}\right) u_n = \left(1 + h\lambda + \frac{h^2\lambda^2}{2}\right)^{n+1} u_0$$

da cui $u_n \to 0$ per $n \to \infty$ se e solo se $|1 + h\lambda + \frac{h^2\lambda^2}{2}| < 1$.

7.2 Assoluta stabilità per i metodi multistep lineari

Un generico metodo multistep lineare applicato al problema test genera la successione

$$u_{n+1} = \sum_{j=0}^{p} a_j u_{n-j} + h\lambda \sum_{j=-1}^{p} b_j u_{n-j} = \sum_{j=0}^{p} (a_j + h\lambda b_j) u_{n-j} + h\lambda b_{-1} u_{n+1}$$

da cui

$$(1 - h\lambda b_{-1}) u_{n+1} - \sum_{j=0}^{p} (a_j + h\lambda b_j) u_{n-j} = 0$$

Tale equazione è un'equazione alle differenze la cui generica soluzione u_j è data da $u_j = \sum_{k=0}^{p} c_k r_k^j$, dove r_k è la k-esima radice del polinomio

$$\sigma(r) = (1 - h\lambda b_{-1}) r^{p+1} - \sum_{j=0}^{p} (a_j + h\lambda b_j) r^{p-j}.$$

Si ha che $u_n \to 0$ per $n \to \infty$ se e solo se $|r_k| < 1$ per ogni k.

Si studiano le regioni di assoluta stabilità dei metodi lineari a uno o più passi presentati nei paragrafi precedenti. Si ricordi che, nel problema modello si ha che $Re(\lambda) < 0$.

- Eulero esplicito: $u_{n+1} = u_n + hf_n$, con $a_0 = 1$, $b_{-1} = 0$ e $b_0 = 1$. Si ha $\sigma(r) = r - (1 + h\lambda) = 0$ e quindi $|r| < 1$ implica $|1 + h\lambda| < 1$, cioè la regione di assoluta stabilità nel piano complesso è costituita dall'interno della circonferenza di centro $(-1, 0)$ e raggio 1.

- Eulero implicito: $u_{n+1} = u_n + hf_{n+1}$, con $a_0 = 1$, $b_{-1} = 1$ e $b_0 = 0$. Si ha $\sigma(r) = (1 - h\lambda)r - 1 = 0$ e quindi $|r| < 1$ implica $\frac{1}{|1-h\lambda|} < 1$, cioè la regione di assoluta stabilità nel piano complesso è costituita dal semipiano contenente i numeri complessi con parte reale negativa.

- Metodo del Trapezio: $u_{n+1} = u_n + \frac{h}{2}(f_n + f_{n+1})$, con $a_0 = 1$, $b_{-1} = b_0 = 0.5$. Si ha $\sigma(r) = (1 - \frac{h}{2}\lambda)r - (1 + \frac{h}{2}\lambda) = 0$ e quindi $|r| < 1$ implica $\frac{|1+\frac{h}{2}\lambda|}{|1-h\lambda|} < 1$, cioè la regione di assoluta stabilità nel piano complesso è costituita dal semipiano contenente i numeri complessi con parte reale negativa.

- Metodo Midpoint: $u_{n+1} = u_{n-1} + 2hf_n$, con $a_1 = 1$, $b_0 = 2$. Si ha $\sigma(r) = (1 - \frac{h}{2}\lambda)r^2 - 2h\lambda - 1 = 0$. Anche solo considerando il caso con $\lambda \in \mathbb{R}$, $\lambda < 0$, si ha che almeno una radice è, in modulo, maggiore di 1. Infatti la radice $r = h\lambda - \sqrt{(}h^2\lambda^2 + 1)$ è tale che $|r| > 1$, in quanto

  $$|r| = -h\lambda + \sqrt{(}h^2\lambda^2 + 1) > \sqrt{(}h^2\lambda^2 + 1) > 1$$

  in quanto $\lambda < 0$. Si ha quindi che la regione di Assoluta stabilità è vuota, perché nessuna scelta di $h\lambda$ permette di ottenere entrambe le radici di σ minori di 1 in modulo.

8 Esempio di applicazione

Si consideri il problema di Cauchy

$$\begin{cases} y' &= f(x,y) \qquad \text{con} \qquad f(x,y) = \frac{y}{1+x} + 3 \\ y(0) &= 0 \end{cases} \tag{14}$$

la cui soluzione è data da $y(x) = 3(1+x)ln(1+x)$. Si noti che la funzione $f(x,y)$ è Lipschitziana rispetto alla seconda variabile se $x \in [0, \infty)$.

Si calcolano le approssimazioni della soluzione sui punti di una griglia contenuta in [0, 1] di passo h mediante il metodo di Eulero esplicito (ordine 1), il metodo di Heun (ordine 2), un metodo Predictor-Corrector formato da Adams-Bashforth e Adams-Moulton (ordine 3), un metodo di Runge-Kutta a 4 stadi (ordine 4).

La tabella 1 presenta i risultati ottenuti applicando i metodi precedenti al problema di Cauchy (14) con $h = 0.25$. Le prime 4 righe riportano i valori delle approssimazioni u_j, $j = 0, \ldots, 4$, sui punti della griglia, mentre l'ultima riga presenta i valori esatti della soluzione sulla griglia. Si noti come, a parità di h, al crescere dell'ordine del metodo si ottengono valori più vicini alla soluzione esatta.

x	0	0.25	0.5	0.75	1
Eulero	0	0.75000	1.65000	2.67500	3.80714
Heun	0	0.82500	1.80250	2.90648	4.11857
P.C.	0	0.83678	1.82598	2.94045	4.16223
R.K.	0	0.83672	1.82449	2.93785	4.15872
y	0	0.83678	1.82459	2.93798	4.15888

Table 1: Approssimazioni della soluzione sulla griglia di passo $h = 0.25$.

Le tabelle 2 e 3 sono relative all'applicazione dei metodi precedenti al problema di Cauchy (14) con differenti passi $h = 10^{-i}$, $i = 1, \ldots, 4$. Il valore di h è riportato nella prima colonna.

Le colonne successive alla prima nella tabella 2 riportano, al variare di h, il massimo errore commesso dai vari metodi nell'approssimare $y(x_j)$ con u_j, $j = 0, \ldots, N$. Si noti che, applicando un metodo di ordine p e dividendo per 10 il passo h, si ottiene un errore diviso per 10^p, poiché l'errore si comporta come h^p.

L'ordine di un metodo viene ulteriormente evidenziato dalle colonne, successive alla prima, nella tabella 3 dove vengono riportati, al variare di h, i rapporti tra il massimo errore commesso dai vari metodi nell'approssimare

h	Eulero	Heun	Pred-Corr	R.K.
0.1000	0.146254	0.00706812606	0.0003323723030	0.000004424302289
0.0100	0.014962	0.00007456297	0.0000004359796	0.000000000466304
0.0010	0.001499	0.00000074956	0.0000000004486	0.000000000000046
0.0001	0.000149	0.00000000749	0.0000000000004	0.000000000000007

Table 2: Massimo errore totale al decrescere di h.

$y(x_j)$ con u_j, $j = 0, \ldots, N$ e h^p, se p è l'ordine del metodo. Si noti che tale rapporto resta praticamente costante.

h	Eulero $/h$	Heun $/h^2$	Pred-Corr $/h^3$	R.K. $/h^4$
0.1000	1.4625	0.7068	0.3323	0.0442
0.0100	1.4962	0.7456	0.4359	0.0466
0.0010	1.4996	0.7495	0.4482	0.0461
0.0001	1.4999	0.7499	0.4467	0.0754

Table 3: Rapporto tra il massimo errore totale e il passo h elevato all'ordine del metodo, al variare di h.

Infine, si noti che si trovano alcune anomalie nel comportamento quando $h = 10^{-4}$, poiché l'elevato numero di passi implica un elevato numero di elementi u_j da calcolare e quindi un grande accumulo di errore algoritmico.

References

[1] A. Quarteroni, R. Sasso, F. Saleri. *Matematica Numerica*, Springer, 2008