

I metodi di Runge-Kutta sono metodi ad un passo in cui il valore u_i viene calcolato combinando un prefissato numero s di valori della funzione f calcolati in punti opportunamente scelti. L'ordine del metodo dipende da s : per valori più bassi di s , cresce con s , per valori più alti di s la crescita dell'ordine, come ha dimostrato Butcher, è un po' meno regolare.

Per quanto riguarda la stabilità, va fatta una distinzione fra metodi espliciti, che in generale hanno regioni di stabilità assoluta piuttosto piccole, e metodi impliciti, molto più stabili.

I metodi di Runge-Kutta espliciti, ed in particolare quello del quarto ordine, sono probabilmente fra le formule più usate per risolvere le equazioni differenziali ordinarie. I motivi di questa preferenza vanno ricercati in una grande semplicità e generalità di uso: le formule sono facilmente programmabili, richiedono poche righe di codice e si prestano ad essere generalizzate senza difficoltà ai sistemi. Inoltre il fatto che si tratti di formule ad un passo consente un controllo adattivo della lunghezza del passo, fornendo uno strumento molto flessibile anche in presenza di situazioni di quasi singolarità della soluzione.

C'è però un rovescio della medaglia: calcolare la soluzione con il metodo di Runge-Kutta in generale richiede, a parità di precisione, molte più valutazioni della f che con i metodi che vedremo nei capitoli successivi.

Fissato un intero s , un *metodo di Runge-Kutta esplicito ad s stadi* è individuato da una matrice A di ordine s strettamente triangolare inferiore e da un vettore \mathbf{b} di ordine s ed ha la forma

$$u_i = u_{i-1} + h \sum_{j=1}^s b_j F_j, \quad (1)$$

dove
$$F_j = f(x_{i-1} + c_j h, u_{i-1} + h \sum_{r=1}^{j-1} a_{jr} F_r), \quad \text{e} \quad c_j = \sum_{r=1}^{j-1} a_{jr}. \quad (2)$$

Il numero s rappresenta la *complessità* del metodo, misurata per mezzo del numero di valutazioni della funzione f richiesto ad ogni passo. Il metodo di Runge-Kutta per $s = 1$ coincide con il metodo di Eulero.

Lo studio della convergenza di un metodo della forma (1) è semplice, infatti per il teorema 4.3 un metodo ad un passo è convergente se è consistente, e quindi, nel caso di (1), se per ogni punto (x, y) prefissato è

$$\lim_{h \rightarrow 0} \sum_{j=1}^s b_j F_j(x, y(x)) = f(x, y).$$

Poiché

$$\lim_{h \rightarrow 0} F_j(x, y(x)) = f(x, y), \quad \text{per } j = 1, \dots, s,$$

la condizione di consistenza dei metodi di Runge-Kutta è

$$\sum_{j=1}^s b_j = 1.$$

Gli elementi A e \mathbf{b} che individuano un particolare metodo sono ricavati imponendo che il metodo abbia un certo ordine (di solito il più elevato possibile). Per ottenere questo si scrive lo sviluppo di Taylor dell'errore locale del metodo in potenze di h e si impone che siano nulli i coefficienti dei primi termini, per ogni indice i e ogni funzione f . Per chiarire questo modo di procedere, ricaviamo una formula a due stadi.

Per $s = 2$ si pone

$$A = \begin{bmatrix} 0 & 0 \\ a_{21} & 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad c_1 = 0, \quad c_2 = a_{21},$$

e la formula è della forma

$$u_i = u_{i-1} + h[b_1 F_1 + b_2 F_2] \\ F_1 = f(x_{i-1}, u_{i-1}), \quad F_2 = f(x_{i-1} + c_2 h, u_{i-1} + a_{21} h F_1).$$

Sostituendo risulta

$$u_i = u_{i-1} + h\phi(x_{i-1}, u_{i-1}; h),$$

dove

$$\phi(x_{i-1}, u_{i-1}; h) = b_1 f(x_{i-1}, u_{i-1}) + b_2 f(x_{i-1} + c_2 h, u_{i-1} + a_{21} h f(x_{i-1}, u_{i-1})).$$

Con la formula di Taylor si ha (la funzione f e le sue derivate si intendono calcolate nel punto (x_{i-1}, y_{i-1})):

$$\begin{aligned} \phi(x_{i-1}, y_{i-1}; h) = b_1 f + b_2 \left[f + c_2 h f_x + a_{21} h f_y f + \frac{c_2^2 h^2}{2} f_{xx} \right. \\ \left. + c_2 a_{21} h^2 f_{xy} f + \frac{a_{21}^2 h^2}{2} f_{yy} f^2 \right] + O(h^3). \end{aligned}$$

D'altra parte è

$$\begin{aligned} \frac{y_i - y_{i-1}}{h} &= y'(x_{i-1}) + \frac{h}{2} y''(x_{i-1}) + \frac{h^2}{6} y'''(x_{i-1}) + O(h^3) \\ &= f + \frac{h}{2} (f_x + f_y f) + \frac{h^2}{6} \left[f_{xx} + 2f_{xy} f + f_{yy} f^2 + f_y (f_x + f_y f) \right] + O(h^3), \end{aligned}$$

per cui, dalla (2) del cap. 4 si ha

$$\begin{aligned} \tau_i &= (1 - b_1 - b_2) f + h \left[\left(\frac{1}{2} - b_2 c_2 \right) f_x + \left(\frac{1}{2} - b_2 a_{21} \right) f_y f \right] \\ &+ h^2 \left[\left(\frac{1}{6} - \frac{b_2 c_2^2}{2} \right) f_{xx} + \left(\frac{1}{3} - b_2 c_2 a_{21} \right) f_{xy} f + \left(\frac{1}{6} - \frac{b_2 a_{21}^2}{2} \right) f_{yy} f^2 \right. \\ &\left. + \frac{1}{6} f_y (f_x + f_y f) \right] + O(h^3). \end{aligned}$$

Imponendo che i coefficienti di h^k per $k = 0$ e $k = 1$ siano nulli per ogni f si ottengono le relazioni

$$\begin{cases} b_1 + b_2 = 1, \\ b_2 c_2 = \frac{1}{2}, \\ b_2 a_{21} = \frac{1}{2}, \end{cases} \quad (3)$$

Non è invece possibile annullare il coefficiente di h^2 e quindi non esiste alcuna scelta dei parametri che dia un metodo di Runge-Kutta esplicito a due stadi di ordine 3. Dal sistema (3), essendo $c_2 = a_{21}$, si ricavano due parametri in funzione del terzo e quindi infiniti metodi di ordine 2. Ad

esempio si ottengono i seguenti metodi*:

$$b_1 = 0, \quad b_2 = 1, \quad c_2 = a_{21} = \frac{1}{2}, \quad \text{metodo di Eulero modificato}$$

$$u_i = u_{i-1} + hf(x_{i-1} + \frac{h}{2}, u_{i-1} + \frac{h}{2}f(x_{i-1}, u_{i-1})),$$

$$b_1 = b_2 = \frac{1}{2}, \quad c_2 = a_{21} = 1, \quad \text{metodo di Heun}$$

$$u_i = u_{i-1} + \frac{h}{2} \left[f(x_{i-1}, u_{i-1}) + f(x_{i-1} + h, u_{i-1} + hf(x_{i-1}, u_{i-1})) \right].$$

Per la prima l'errore locale risulta

$$\tau_i = \frac{h^2}{6} \left[\frac{1}{4} (f_{xx} + 2f_{xy}f + f_{yy}f^2) + f_y(f_x + f_yf) \right] + O(h^3),$$

e per la seconda

$$\tau_i = \frac{h^2}{6} \left[-\frac{1}{2} (f_{xx} + 2f_{xy}f + f_{yy}f^2) + f_y(f_x + f_yf) \right] + O(h^3).$$

In modo analogo e con fatica crescente al crescere di s si possono ricavare metodi di Runge-Kutta di ordine più elevato. Ad esempio, per $s = 3$ si ottengono 4 relazioni per determinare 6 parametri, per $s = 4$ si ottengono 8 relazioni per determinare 10 parametri e così via. Vi è quindi un certo grado di libertà nella risoluzione dei sistemi ottenuti: il criterio seguito è in genere quello della massima semplicità delle formule che si ottengono. Naturalmente si possono seguire anche altri criteri, come quello di minimizzare l'errore locale o l'ingombro di memoria richiesto.

I metodi vengono di solito descritti in modo sintetico tramite la matrice A e i vettori \mathbf{b} e \mathbf{c} nel modo seguente:

| | | | | | |
|-------|----------|----------|-----|-------------|-------|
| 0 | | | | | |
| c_2 | a_{21} | | | | |
| c_3 | a_{31} | a_{32} | | | |
| ... | ... | ... | | | |
| c_s | a_{s1} | a_{s2} | ... | $a_{s,s-1}$ | |
| | b_1 | b_2 | ... | ... | b_s |

Ne elenchiamo alcuni per $s \leq 4$.

* Ai vari metodi di Runge-Kutta vengono dati nomi differenti dai diversi autori. Questi sono i nomi dati da Henrici [1962].

Per $s = 1$

| | |
|---|---|
| 0 | |
| | 1 |

Per $s = 2$

| | | | |
|---------------|---------------|---------------|---------------|
| 0 | | 0 | |
| $\frac{1}{2}$ | $\frac{1}{2}$ | 1 | 1 |
| | | 0 | 1 |
| | | $\frac{1}{2}$ | $\frac{1}{2}$ |

Per $s = 3$

| | | | | | |
|---------------|---------------|---------------|---------------|---------------|---|
| 0 | | 0 | | 0 | |
| $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | 1 | 2 |
| $\frac{2}{3}$ | 0 | $\frac{2}{3}$ | | | |
| | | $\frac{1}{4}$ | 0 | $\frac{3}{4}$ | |
| | | $\frac{1}{6}$ | $\frac{2}{3}$ | $\frac{1}{6}$ | |

Per $s = 4$

| | | | | | | | |
|---------------|---------------|---------------|--------------------------|---------------------------|--------------------------|-----------------------|--------------------------|
| 0 | | 0 | | 0 | | 0 | |
| $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ |
| $\frac{1}{2}$ | 0 | $\frac{1}{2}$ | | $\frac{-1 + \sqrt{2}}{2}$ | $\frac{2 - \sqrt{2}}{2}$ | $\frac{-\sqrt{2}}{2}$ | $\frac{2 + \sqrt{2}}{2}$ |
| 1 | 0 | 0 | 1 | 1 | 0 | | |
| | | $\frac{1}{6}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{2 - \sqrt{2}}{6}$ |
| | | $\frac{1}{6}$ | $\frac{2 - \sqrt{2}}{6}$ | $\frac{2 + \sqrt{2}}{6}$ | $\frac{1}{6}$ | | |

Il primo dei due metodi per $s = 4$ è il *classico metodo di Runge-Kutta* di ordine 4, il secondo quello di *Gill*, usato perché più efficiente nella gestione della memoria, caratteristica questa che lo rende particolarmente utile quando si devono risolvere sistemi di equazioni (si veda [Butcher]).

Per $s \leq 4$ i metodi di Runge-Kutta espliciti hanno ordine s e complessità s , cioè hanno un errore locale che tende a zero con h^s e richiedono ad ogni passo s valutazioni di f . Fissato un certo errore massimo di troncamento, più s è grande, minore è il numero dei passi occorrenti su un dato intervallo, e anche se il numero di valutazioni di f cresce con s , è di solito conveniente scegliere metodi con s elevato.

5.1 Esempio. Si applicano i metodi sopra elencati al problema test B per $\eta = 1$, scegliendo all'inizio un passo h in modo che l'errore locale non superi $\epsilon = 10^{-4}$. L'errore viene valutato con la stima di Richardson, partendo da un valore iniziale di $h = 0.1$ e utilizzando la (16) del cap. 3. Per un metodo ad s stadi sono richieste $3s - 1$ valutazioni della f (infatti il valore $f(x_{i-1}, u_{i-1})$ è comune a \bar{u}_1 e u_1).

Per $s = 1$, metodo di Eulero, come si è visto nell'esempio 3.9 la condizione che l'errore locale sia minore di 10^{-4} è soddisfatta inizialmente quando $h \sim 0.16 \cdot 10^{-4}$, e se si mantiene questo passo per tutto l'intervallo $[1, 2]$ si devono fare più di 60 mila valutazioni della f . Se invece si applica il metodo in modo adattivo la soluzione viene calcolata in 14050 passi con 28100 valutazioni della f . Se infine si applica un metodo parzialmente adattivo, controllando l'errore locale ogni dieci passi, la soluzione viene calcolata in 14067 passi con 15346 valutazioni della f .

Per $s = 2$ si usa il metodo di Eulero modificato che ha ordine 2. L'errore locale viene stimato con la relazione:

$$\tau_1 \sim \frac{4(\bar{u}_1 - u_1)}{3h}.$$

Per $h = 0.1$ si ha $\tau_1 = -0.805418 \cdot 10^{-1}$. Poiché il metodo è del secondo ordine, l'errore locale risulta dell'ordine di ϵ se si fissa

$$h = \left(\frac{\epsilon}{|\tau_1|} \right)^{1/2} \cdot 0.1 = 0.0035 \sim \frac{1}{284}.$$

Se se si mantiene questo passo per tutto l'intervallo $[1, 2]$ la soluzione viene calcolata in 284 passi e quindi con 573 valutazioni della f . Se invece si applica il metodo in modo adattivo la soluzione viene calcolata in 110 passi con 551 valutazioni della f . Se infine si applica un metodo parzialmente adattivo, controllando l'errore locale ogni dieci passi, la soluzione viene calcolata in 118 passi con 270 valutazioni della f .

Per $s = 3$ si usa il metodo $\mathbf{b}^T = \left(\frac{1}{4}, 0, \frac{3}{4} \right)$ (quello della tabella di sinistra), che ha ordine 3. L'errore locale viene stimato con la relazione:

$$\tau_1 \sim \frac{8(\bar{u}_1 - u_1)}{7h}.$$

Per $h = 0.1$ si ha $\tau_1 = 0.852531 \cdot 10^{-2}$. Poiché il metodo è del terzo ordine, l'errore locale risulta dell'ordine di ϵ se si fissa

$$h = \left(\frac{\epsilon}{|\tau_1|} \right)^{1/3} \cdot 0.1 = 0.0227 \sim \frac{1}{45}.$$

Se se si mantiene questo passo per tutto l'intervallo $[1, 2]$ la soluzione viene calcolata in 45 passi e quindi con 143 valutazioni della f . Se invece si applica il metodo in modo adattivo la soluzione viene calcolata in 23 passi con 188 valutazioni della f . Se infine si applica un metodo parzialmente adattivo, controllando l'errore locale ogni dieci passi, la soluzione viene calcolata in 28 passi con 101 valutazioni della f .

Per $s = 4$ si usa il classico metodo di Runge-Kutta di ordine 4. L'errore locale viene stimato con la relazione:

$$\tau_1 \sim \frac{16(\bar{u}_1 - u_1)}{15h}.$$

Per $h = 0.1$ si ha $\tau_1 = -0.185125 \cdot 10^{-3}$. Poiché il metodo è del quarto ordine, l'errore locale risulta dell'ordine di ϵ se si fissa

$$h = \left(\frac{\epsilon}{|\tau_1|} \right)^{1/4} \cdot 0.1 = 0.0857 \sim \frac{1}{12}.$$

Se se si mantiene questo passo per tutto l'intervallo $[1, 2]$ la soluzione viene calcolata in 12 passi e quindi con 59 valutazioni della f . Se invece si applica il metodo in modo adattivo la soluzione viene calcolata in 9 passi con 102 valutazioni della f . Se infine si applica un metodo parzialmente adattivo, controllando l'errore locale ogni dieci passi, la soluzione viene calcolata in 11 passi con 66 valutazioni della f .

La seguente tabella riporta il passo minimo e massimo, il numero di valutazioni della f e il massimo modulo degli errori effettivamente prodotti con il metodo parzialmente adattivo.

| s | h_{min} | h_{max} | no. val. di f | $\max e_i $ |
|-----|-----------|-----------|-----------------|-------------------------|
| 2 | 0.0035 | 0.0239 | 270 | $0.32136 \cdot 10^{-4}$ |
| 3 | 0.0227 | 0.0820 | 101 | $0.19386 \cdot 10^{-4}$ |
| 4 | 0.0857 | 0.0857 | 66 | $0.58357 \cdot 10^{-5}$ |

In questo caso particolare, in cui la soluzione è decrescente, l'errore globale si mantiene sempre più basso di quello locale. Ad esempio, per $s = 4$ il grafico del modulo dell'errore effettivamente prodotto con il metodo classico di Runge-Kutta del quarto ordine parzialmente adattivo risulta come in figura 5.1.

Fig. 5.1 - Errore della soluzione calcolata con il metodo classico di Runge-Kutta del quarto ordine parzialmente adattivo

■

Per le formule sopra riportate il numero s di stadi è pari all'ordine. Questa proprietà vale però solo fino a $s = 4$. Per $s > 4$ non è possibile ricavare metodi di ordine s con solo s stadi: ad esempio, per ottenere un metodo di ordine 5 sono necessari 6 stadi. Per i metodi di ordine maggiore di 4 vale la seguente tabella (per la cui costruzione si veda [Butcher]):

| | | | | | | | | |
|--------|---|---|---|---|---|---|---|----|
| ordine | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| stadi | 1 | 2 | 3 | 4 | 6 | 7 | 9 | 11 |

Oltre che la tecnica di Richardson, usata nell'esempio 5.1, ai metodi di Runge-Kutta si può applicare anche un'altra tecnica di stima dell'errore locale, che si basa sul confronto fra i valori ottenuti con due metodi diversi, scelti opportunamente. Indichiamo con

$$u_i = u_{i-1} + h\phi(x_{i-1}, u_{i-1}; h) \quad \text{e} \quad \bar{u}_i = \bar{u}_{i-1} + h\bar{\phi}(x_{i-1}, \bar{u}_{i-1}; h)$$

i due metodi e supponiamo che il primo sia di ordine p e il secondo di ordine $p + 1$. Quindi si ha

$$\begin{aligned} y_i &= y_{i-1} + h\phi(x_{i-1}, y_{i-1}; h) + h\tau_i, \\ \bar{y}_i &= y_{i-1} + h\bar{\phi}(x_{i-1}, y_{i-1}; h) + h\bar{\tau}_i, \end{aligned}$$

in cui $\tau_i \sim Mh^p$ e $\bar{\tau}_i \sim \bar{M}h^{p+1}$. Applicando i due metodi a partire dallo stesso punto u_0 si ottengono i valori u_1 e \bar{u}_1 . La differenza $u_1 - \bar{u}_1$ risulta

quindi calcolata con un metodo di ordine p e fornisce una stima di $h\tau_1$, cioè relativa al metodo di ordine minore. Il calcolo prosegue poi con il valore \bar{u}_1 , che si suppone sia un'approssimazione di y_1 migliore di u_1 .

Perché un simile procedimento sia efficace però è necessario che i due metodi che si usano sfruttino una parte comune di risultati intermedi in modo da ridurre la complessità. Sono state costruite varie famiglie di metodi di Runge-Kutta che soddisfano al requisito di avere la minima complessità possibile. In pratica un metodo è costituito da due formule, la prima di ordine p e la seconda di ordine $p + 1$, in cui la matrice A della seconda è ottenuta orlando la matrice A della prima con una riga.

Un metodo molto usato è il seguente metodo di Fehlberg, costituito da una formula a 5 stadi di ordine 4 e una a 6 stadi di ordine 5. Nella prima riga sotto la linea sono riportati i coefficienti \mathbf{b} da usare per calcolare u_1 utilizzando gli F_j ottenuti dalle prime 5 righe, nella seconda riga sotto la linea sono riportati i coefficienti \mathbf{b} da usare per calcolare \bar{u}_1 utilizzando tutti gli F_j : in pratica sono richieste 6 valutazioni della f per passo, la metà di quelle richieste dalla tecnica di Richardson per il metodo del quarto ordine.

| | | | | | | |
|---------------|------------------|--------------------|------------------|------------------|-----------------|----------------|
| 0 | | | | | | |
| $\frac{2}{9}$ | $\frac{2}{9}$ | | | | | |
| $\frac{1}{3}$ | $\frac{1}{12}$ | $\frac{1}{4}$ | | | | |
| $\frac{3}{4}$ | $\frac{69}{128}$ | $-\frac{243}{128}$ | $\frac{135}{64}$ | | | |
| 1 | $-\frac{17}{12}$ | $\frac{27}{4}$ | $-\frac{27}{5}$ | $\frac{16}{15}$ | | |
| $\frac{5}{6}$ | $\frac{65}{432}$ | $-\frac{5}{16}$ | $\frac{13}{16}$ | $\frac{4}{27}$ | $\frac{5}{144}$ | |
| | $\frac{1}{9}$ | 0 | $\frac{9}{20}$ | $\frac{16}{45}$ | $\frac{1}{12}$ | |
| | $\frac{47}{450}$ | 0 | $\frac{12}{25}$ | $\frac{32}{225}$ | $\frac{1}{30}$ | $\frac{6}{25}$ |

5.2 Esempio. Si applica il metodo di Fehlberg al problema test B per $\eta = 1$. Posto $h = 0.1$ si ottiene

$$\text{in } x_1 = 1.1, \quad u_1 = 0.754522, \quad \bar{u}_1 = 0.754531, \quad \frac{u_1 - \bar{u}_1}{h} = 9.40908 \cdot 10^{-5}.$$

L'errore locale risulta quindi dell'ordine di $\epsilon = 10^{-4}$. Se si applica il metodo in modo adattivo, imponendo che ad ogni passo l'errore locale risulti dell'ordine di ϵ , si ottengono passi di lunghezza crescente: all'inizio

$$h = \left(\frac{\epsilon}{9.40908 \cdot 10^{-5}} \right)^{1/4} \cdot 0.1 = 0.1015,$$

fino ad arrivare ad $h = 0.347$ alla fine dell'intervallo. In tutto sono richieste 42 valutazioni della f e il massimo errore effettivo risulta $0.154383 \cdot 10^{-5}$. ■

Per stimare l'errore locale dei metodi di Runge-Kutta vi sono anche altre tecniche, meno costose di quelle illustrate, che sfruttano i risultati ottenuti in vari passi precedenti (per i dettagli si veda [Butcher]).

Per studiare la stabilità assoluta dei metodi di Runge-Kutta espliciti, applichiamo la (1) al problema test A. Si ha

$$u_i = u_{i-1} + h \sum_{j=1}^s b_j F_j, \quad F_j = \lambda \left[u_{i-1} + h \sum_{r=1}^{j-1} a_{jr} F_r \right], \quad (4)$$

che in forma matriciale si scrive

$$u_i = u_{i-1} + h \mathbf{b}^T \mathbf{F}, \quad \mathbf{F} = \lambda (u_{i-1} \mathbf{e} + h A \mathbf{F}),$$

dove \mathbf{e} è il vettore di s componenti tutte uguali a 1 e \mathbf{F} è il vettore di componenti F_j . Posto $z = h\lambda$ si ha

$$(I - zA) \mathbf{F} = \lambda u_{i-1} \mathbf{e},$$

da cui

$$u_i = u_{i-1} [1 + z \mathbf{b}^T (I - zA)^{-1} \mathbf{e}]. \quad (5)$$

Poiché A è strettamente triangolare inferiore è $A^s = O$ e risulta

$$(I - zA)^{-1} = I + zA + z^2 A^2 + \dots + z^{s-1} A^{s-1},$$

da cui

$$u_i = u_{i-1} p_s(z), \quad \text{dove} \quad p_s(z) = 1 + \sum_{r=1}^s (\mathbf{b}^T A^{r-1} \mathbf{e}) z^r. \quad (6)$$

Quindi $p_s(z)$ è un polinomio di grado s in z .

La regione di stabilità assoluta del metodo (1) è allora costituita dai punti z del piano complesso per cui

$$|p_s(z)| < 1.$$

Nel caso dei metodi con $s \leq 4$ i coefficienti del polinomio $p_s(z)$ si trovano subito notando che la soluzione del problema test A è

$$y(x) = y_0 e^{\lambda(x-x_0)},$$

per cui

$$y_i = y_{i-1} e^z = y_{i-1} \sum_{r=0}^{\infty} \frac{z^r}{r!}. \quad (7)$$

D'altra parte, per definizione di errore locale, dalla (6) risulta

$$y_i = y_{i-1} p_s(z) + h\tau_i,$$

e poiché per $s \leq 4$ il metodo ad s stadi ha ordine s , ne segue che

$$y_i = y_{i-1} p_s(z) + O(h^{s+1}).$$

Confrontando con la (7) ne segue che

$$p_s(z) = \sum_{r=0}^s \frac{z^r}{r!},$$

per cui la regione di stabilità assoluta di un metodo di Runge-Kutta esplicito ad s stadi, con $s \leq 4$, è costituito dai punti z del piano complesso tali che

$$\left| \sum_{r=0}^s \frac{z^r}{r!} \right| < 1.$$

Nella figura 5.2 sono rappresentate le regioni di stabilità assoluta* dei metodi di Runge-Kutta espliciti con $s \leq 4$. I metodi sono assolutamente stabili nelle regioni rappresentate.

* Per risolvere l'equazione $|p_s(z)| = 1$ nel campo complesso, conviene prima determinare la soluzione \bar{x} reale negativa dell'equazione $p_s(x) = 1$ (per s pari) o $p_s(x) = -1$ (per s dispari), usando il metodo delle tangenti. Si consideri poi una griglia di punti x_k equidistanti in $[\bar{x}, 0]$ e per ogni k si risolva l'equazione $|p_s(x_k + iy)|^2 = 1$. L'equazione risulta di grado s in y^2 . Si consiglia di usare il metodo delle tangenti utilizzando come approssimazione iniziale il valore ottenuto per il k precedente. In modo simile si può procedere per quella parte della regione di stabilità nel primo quadrante.

Fig. 5.2 - Regioni di stabilità assoluta dei metodi di Runge-Kutta espliciti con $s \leq 4$

Se $\lambda < 0$ è reale, le limitazioni per la stabilità assoluta sono

$$h|\lambda| < \begin{cases} 2 & \text{per } s = 1, 2, \\ 2.51275 & \text{per } s = 3, \\ 2.78529 & \text{per } s = 4. \end{cases}$$

Per $s > 4$ la determinazione dei coefficienti di $p_s(z)$ è più complicata, perché l'ordine del metodo è minore di s , per cui dipendono da h^s sia $p_s(z)$ che $h\tau_i$. Quindi $p_s(z)$ deve essere calcolato esplicitamente per sostituzione nella (4). Ne segue che la regione di stabilità assoluta risulta diversa anche nell'ambito di metodi dello stesso ordine.

Per il metodo di Runge-Kutta-Fehlberg per $\lambda < 0$ reale si ha:

per la prima formula, di ordine 4 e $s = 5$

$$p_5(z) = 1 + z + \frac{z^2}{2} + \frac{z^3}{6} + \frac{z^4}{24} + \frac{z^5}{96}$$

e $|p_5(z)| < 1$ per $h|\lambda| < 2.92581$;

per la seconda formula, di ordine 5 e $s = 6$

$$p_6(z) = 1 + 2z + z^2 + \frac{z^3}{3} + \frac{z^4}{12} + \frac{3z^5}{160} + \frac{z^6}{960}$$

e $|p_6(z)| < 1$ per $h|\lambda| < 2.19291$.

Poiché $p_s(x)$ è un polinomio, per x reale risulta

$$\lim_{x \rightarrow -\infty} |p_s(x)| = \infty,$$

e quindi esiste $\bar{x} < 0$ tale che

$$|p_s(x)| \geq 1 \quad \text{per ogni } x \leq \bar{x}.$$

Ne segue che nessun metodo di Runge-Kutta esplicito è A-stabile.

La regione di stabilità assoluta dei metodi di Runge-Kutta espliciti è troppo ridotta se il problema che si deve risolvere è stiff. Anzi, la situazione risulta tanto più grave quanto più è alto l'ordine del metodo usato, perché i metodi di ordine più alto consentirebbero, considerando solo l'errore locale, un passo abbastanza lungo.

5.3 Esempio. Applichiamo il metodo classico di Runge-Kutta del quarto ordine al problema test C dell'esempio 3.7, con un procedimento adattivo, in modo che il modulo l'errore locale non superi $\epsilon = 10^{-2}$. Inizialmente viene posto $h = 0.0038$ e ad ogni passo h cresce. Dopo 10 passi, per $x_{10} = 0.06667$ risulta $h = 0.01$. Dopo altri 5 passi, per $x_{15} = 0.17$ risulta $h = 0.0257$. Da questo punto fino al termine il passo non cresce più, anche se l'errore locale decresce vertiginosamente: la derivata quinta della $y(x)$ passa infatti da -21 per $x = 0.2$ a -10^{-34} per $x = 1$. Il valore $h = 0.0257$ del passo è determinato solo dalla condizione di stabilità assoluta. Aumentando h oltre questo valore insorgerebbe instabilità. ■

Se nella (2) modifichiamo il secondo estremo della sommatoria facendo assumere ad r valori maggiori o uguali a j e in corrispondenza poniamo $a_{jr} \neq 0$, il calcolo di F_j viene a coinvolgere qualche F_r con indice maggiore di j o anche F_j stesso. Le formule risultano in tal modo implicite.

I metodi di Runge-Kutta impliciti hanno, a parità di numero di stadi, un ordine molto maggiore di quelli espliciti. Purtroppo questo non si riflette in un reale vantaggio dal punto di vista della complessità, a causa del fatto che per calcolare u_i si deve fare ricorso ad un metodo iterativo. In compenso i metodi impliciti hanno delle regioni di stabilità assoluta molto maggiori di quelli espliciti ed è solo per considerazioni di stabilità che si giustifica il loro uso.

Formalmente, fissato un intero s , un *metodo di Runge-Kutta implicito ad s stadi* è individuato da una matrice A di ordine s e da un vettore \mathbf{b} di ordine s ed ha la forma

$$u_i = u_{i-1} + h \sum_{j=1}^s b_j F_j,$$

dove
$$F_j = f(x_{i-1} + c_j h, u_{i-1} + h \sum_{r=1}^s a_{jr} F_r), \quad \text{e} \quad c_j = \sum_{r=1}^s a_{jr}. \quad (8)$$

Una sottoclasse dei metodi impliciti è costituita dai metodi semi-impliciti, in cui la matrice A è triangolare inferiore con qualche elemento $a_{jj} \neq 0$, per cui in alcune delle somme in (8) l'indice r va fino a j .

Per costruire metodi di Runge-Kutta impliciti ci si può servire di una tecnica analoga a quella vista per i metodi espliciti.

5.4 Esempio. Per $s = 1$ si pone

$$A = [a_{11}], \quad \mathbf{b} = [b_1], \quad c_1 = a_{11},$$

e il metodo è della forma

$$u_i = u_{i-1} + b_1 h F_1, \quad F_1 = f(x_{i-1} + c_1 h, u_{i-1} + a_{11} h F_1).$$

Procedendo come nel paragrafo 2, si ha con la formula di Taylor (la funzione f e le sue derivate si intendono calcolate nel punto (x_{i-1}, y_{i-1})):

$$F_1 = f + c_1 h f_x + a_{11} h f_y f + O(h^2),$$

da cui

$$u_i = u_{i-1} + b_1 h f + b_1 h^2 (c_1 f_x + a_{11} f_y f) + O(h^3),$$

e quindi

$$\tau_i = (1 - b_1) f + h \left[\left(\frac{1}{2} - b_1 c_1 \right) f_x + \left(\frac{1}{2} - b_1 a_{11} \right) f_y f \right] + O(h^2).$$

Imponendo che i coefficienti di h^0 e di h^1 siano nulli per ogni f , si ottiene

$$b_1 = 1, \quad c_1 = a_{11} = \frac{1}{2},$$

a cui corrisponde il metodo di Runge-Kutta implicito di ordine 2

$$u_i = u_{i-1} + h F_1, \quad F_1 = f\left(x_{i-1} + \frac{h}{2}, u_{i-1} + \frac{h}{2} F_1\right).$$

Per $s = 2$ si pone

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad c_1 = a_{11} + a_{12}, \quad c_2 = a_{21} + a_{22},$$

e il metodo è della forma

$$\begin{aligned} u_i &= u_{i-1} + h [b_1 F_1 + b_2 F_2] \\ F_1 &= f(x_{i-1} + c_1 h, u_{i-1} + h [a_{11} F_1 + a_{12} F_2]), \\ F_2 &= f(x_{i-1} + c_2 h, u_{i-1} + h [a_{21} F_1 + a_{22} F_2]). \end{aligned}$$

Procedendo come nel paragrafo 2 e imponendo che siano nulli i coefficienti di h^k per $k = 0, \dots, 3$, nello sviluppo di τ_i si ottengono le relazioni

$$\left\{ \begin{array}{l} \sum_{i=1}^2 b_i = 1, \\ \sum_{i=1}^2 b_i c_i = \frac{1}{2}, \\ \sum_{i=1}^2 b_i c_i^2 = \frac{1}{3}, \\ \sum_{i=1}^2 b_i c_i^3 = \frac{1}{4}, \end{array} \right. \quad \left\{ \begin{array}{l} \sum_{i=1}^2 b_i \sum_{j=1}^2 a_{ij} c_j = \frac{1}{6}, \\ \sum_{i=1}^2 b_i c_i \sum_{j=1}^2 a_{ij} c_j = \frac{1}{8}, \\ \sum_{i=1}^2 b_i \sum_{j=1}^2 a_{ij} c_j^2 = \frac{1}{12}, \\ \sum_{i=1}^2 b_i \sum_{j=1}^2 c_j \sum_{k=1}^2 a_{ik} a_{kj} = \frac{1}{24}. \end{array} \right. \quad (9)$$

Le quattro relazioni di sinistra sono indipendenti fra di loro e consentono di calcolare i b_i e i c_i . I c_i risultano le soluzioni dell'equazione

$$6c^2 - 6c + 1 = 0,$$

e quindi

$$c_1 = \frac{3 - \sqrt{3}}{6} \quad \text{e} \quad c_2 = \frac{3 + \sqrt{3}}{6}, \quad \text{da cui} \quad b_1 = b_2 = \frac{1}{2}.$$

Sostituendo questi valori nelle prime due relazioni di destra (le altre sono dipendenti da queste), risulta

$$a_{11} = a_{22} = \frac{1}{4}, \quad a_{12} = \frac{3 - 2\sqrt{3}}{12}, \quad a_{21} = \frac{3 + 2\sqrt{3}}{12}.$$

I valori dei parametri così trovati sono gli unici che soddisfano il sistema (9). Quindi vi è un solo metodo di Runge-Kutta implicito a 2 stadi di ordine 4.

| | | |
|--------------------------|----------------------------|----------------------------|
| $\frac{3 - \sqrt{3}}{6}$ | $\frac{1}{4}$ | $\frac{3 - 2\sqrt{3}}{12}$ |
| $\frac{3 + \sqrt{3}}{6}$ | $\frac{3 + 2\sqrt{3}}{12}$ | $\frac{1}{4}$ |
| | $\frac{1}{2}$ | $\frac{1}{2}$ |

Per ricavare un metodo semi-implicito basta porre

$$A = \begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad c_1 = a_{11}, \quad c_2 = a_{21} + a_{22}.$$

Il metodo è della forma

$$\begin{aligned} u_i &= u_{i-1} + h[b_1 F_1 + b_2 F_2] \\ F_1 &= f(x_{i-1} + c_1 h, u_{i-1} + h a_{11} F_1), \\ F_2 &= f(x_{i-1} + c_2 h, u_{i-1} + h[a_{21} F_1 + a_{22} F_2]), \end{aligned}$$

e si impone che abbia ordine 3. Si ottengono le quattro relazioni

$$\begin{cases} \sum_{i=1}^2 b_i c_i^k = \frac{1}{k+1}, & \text{per } k = 0, 1, 2, \\ b_1 c_1 a_{11} + b_2 c_1 a_{21} + b_2 c_2 a_{22} = \frac{1}{2}, \end{cases}$$

da cui si possono ottenere i parametri con un grado di libertà. I seguenti due metodi si ottengono imponendo che $c_1 = a_{11} = 0$ e che $a_{11} = a_{22}$.

| | | | | |
|---------------|---------------|--------------------------|--------------------------|--------------------------|
| 0 | 0 | $\frac{3 + \sqrt{3}}{6}$ | $\frac{3 + \sqrt{3}}{6}$ | |
| $\frac{2}{3}$ | $\frac{1}{3}$ | $\frac{3 - \sqrt{3}}{6}$ | $\frac{-\sqrt{3}}{3}$ | $\frac{3 + \sqrt{3}}{6}$ |
| $\frac{3}{3}$ | $\frac{1}{3}$ | $\frac{3 - \sqrt{3}}{6}$ | $\frac{3}{3}$ | $\frac{3 + \sqrt{3}}{6}$ |
| | $\frac{1}{4}$ | $\frac{3}{4}$ | $\frac{1}{2}$ | $\frac{1}{2}$ |

entrambi di ordine 3. ■

Dal punto di vista del calcolo di F_j , i metodi semi-impliciti sono molto più semplici di quelli impliciti in generale. Infatti ciascun F_j può essere calcolato risolvendo una sola equazione del tipo (8), ad esempio con il metodo iterativo

$$F_j^{(k)} = f\left(x_{i-1} + c_j h, u_{i-1} + h \sum_{r=1}^{j-1} a_{jr} F_r + h a_{jj} F_j^{(k-1)}\right), \quad (10)$$

Se la funzione f è lipschitziana rispetto ad y di costante L , si può scegliere h in modo che il metodo sia convergente. Infatti

$$\begin{aligned} |F_j^{(k)} - F_j| &= \left| f\left(x_{i-1} + c_j h, u_{i-1} + h \sum_{r=1}^{j-1} a_{jr} F_r + h a_{jj} F_j^{(k-1)}\right) \right. \\ &\quad \left. - f\left(x_{i-1} + c_j h, u_{i-1} + h \sum_{r=1}^{j-1} a_{jr} F_r + h a_{jj} F_j\right) \right| \\ &\leq hL|a_{jj}| |F_j^{(k-1)} - F_j|. \end{aligned}$$

Se h è tale che

$$hL|a_{jj}| < 1,$$

il metodo iterativo (10) è convergente per ogni scelta iniziale $F_j^{(0)}$.

Invece nel caso di un metodo implicito in generale, è necessario calcolare gli F_j risolvendo un sistema del tipo

$$\mathbf{F} = \mathbf{g}(\mathbf{F}), \quad (11)$$

in cui \mathbf{F} è il vettore di componenti F_j , $j = 1, \dots, s$ e \mathbf{g} è un vettore di s funzioni. Si può scegliere h in modo che il metodo iterativo

$$\mathbf{F}^{(k)} = \mathbf{g}(\mathbf{F}^{(k-1)}), \quad (12)$$

sia convergente. Infatti procedendo come sopra si ha

$$|F_j^{(k)} - F_j| \leq hL \sum_{r=1}^s |a_{jr}| |F_r^{(k-1)} - F_r|,$$

e passando alle norme si ha

$$\|\mathbf{F}^{(k)} - \mathbf{F}\| \leq hL\|A\| \|\mathbf{F}^{(k-1)} - \mathbf{F}\|.$$

Se h è tale che

$$hL\|A\| < 1, \quad (13)$$

il metodo iterativo (12) è convergente per ogni scelta del punto iniziale $\mathbf{F}^{(0)}$ e quindi il sistema (11) ha soluzione unica.

5.5 Esempio. Si applica il metodo implicito di ordine 4 ricavato nell'esempio 5.4 al problema test B per $\eta = 1$. Essendo

$$\|A\|_\infty = \frac{3 + \sqrt{3}}{6} \sim 0.78868, \quad L = |f_y(x_0, y_0)| = 4,$$

la condizione (13) è verificata per $h < 0.3$. Assumendo ad esempio $h = 0.1$, il sistema (11) viene ad avere una e una sola soluzione che si può calcolare con il metodo (12). Nel punto (x_0, y_0) si pone $F_1^{(0)} = F_2^{(0)} = 0$ e si itera la (12) ottenendo la successione

| k | $F_1^{(k)}$ | $F_2^{(k)}$ |
|-----|-------------|-------------|
| 1 | -2.97974 | -2.93266 |
| 2 | -2.73115 | -2.05364 |
| 3 | -2.74206 | -2.17902 |
| 4 | -2.74288 | -2.16570 |
| 5 | -2.74260 | -2.16673 |
| 6 | -2.74265 | -2.16669 |

Il metodo iterativo viene interrotto quando

$$\|\mathbf{F}^{(k)} - \mathbf{F}^{(k-1)}\| < 10^{-4},$$

e si assume $\mathbf{F} = \mathbf{F}^{(6)}$; poiché ogni iterazione richiede due valutazioni della f , il vettore \mathbf{F} è stato ottenuto con 12 valutazioni della f . Si ricava quindi $u_1 = 0.754533$ con un errore effettivo dell'ordine di 10^{-6} . Nei passi successivi si assumono per $F_1^{(0)}$ e $F_2^{(0)}$ i valori ottenuti al passo precedente e poiché la f_y diminuisce in modulo, anche il numero delle iterazioni richieste diminuisce. All'ultimo passo, per $x_i = 0.9$, bastano 4 iterazioni. L'errore effettivo si mantiene sempre dell'ordine di 10^{-6} . ■

Il numero di valutazioni della f richiesto in questo esempio è piuttosto elevato, perché il metodo (12) converge lentamente. Il metodo convergerebbe più rapidamente se si scegliesse h più piccolo, ma questo ci obbligherebbe a fare un maggior numero di passi del metodo di Runge-Kutta. Inoltre la condizione (13) può essere troppo restrittiva se L è grande. In tal caso conviene applicare al sistema (11) il metodo di Newton-Raphson. Posto $J = I - M$, in cui M è la matrice il cui (j, l) -esimo elemento è

$$m_{jl} = \frac{\partial g_j}{\partial F_l}(\mathbf{F}^{(k-1)}) = ha_{jl}f_y(x_{i-1} + c_j h, u_{i-1} + h \sum_{r=1}^s a_{jr} F_r^{(k-1)}),$$

si calcola il vettore \mathbf{d} , soluzione del sistema lineare

$$(I - M)\mathbf{d} = \mathbf{g}(\mathbf{F}^{(k-1)}) - \mathbf{F}^{(k-1)},$$

e si costruisce il vettore

$$\mathbf{F}^{(k)} = \mathbf{F}^{(k-1)} + \mathbf{d}.$$

Lo studio della convergenza del metodo di Newton-Raphson è più complicato che per il metodo (12), e la scelta del vettore $\mathbf{F}^{(0)}$ può essere determinante, specialmente quando il problema è stiff, per L elevato. Però se $\mathbf{F}^{(0)}$ è abbastanza vicino a \mathbf{F} , il metodo di Newton-Raphson converge molto più rapidamente. Per la effettiva convergenza di questo metodo e per le varianti che ne diminuiscono la complessità (a scapito della velocità), occorre fare riferimento alla teoria del metodo di Newton-Raphson.

5.6 Esempio. Si applica il metodo implicito di ordine 4 ricavato nell'esempio 5.4 al problema test B per $\eta = 1$, procedendo come nell'esempio 5.5, ma utilizzando il metodo di Newton-Raphson per risolvere il sistema (11). Si ottiene la successione

| k | $F_1^{(k)}$ | $F_2^{(k)}$ |
|-----|-------------|-------------|
| 1 | -2.73873 | -2.12749 |
| 2 | -2.74264 | -2.16669 |
| 3 | -2.74264 | -2.16669 |

■

Come risulta anche dall'esempio 5.4, la costruzione diretta di metodi di Runge-Kutta impliciti è assai onerosa. Studieremo perciò una sottoclasse dei metodi ricavati da formule di quadratura. Poiché $y' = f(x, y)$, si ha

$$y_i - y_{i-1} = \int_{x_{i-1}}^{x_i} y'(x) dx = \int_{x_{i-1}}^{x_i} f(x, y(x)) dx. \quad (14)$$

Con il cambiamento di variabile

$$x = x_{i-1} + th, \quad 0 \leq t \leq 1$$

questo integrale diventa uguale a

$$h \int_0^1 g(t) dt, \quad \text{dove } g(t) = f(x(t), y(x(t))). \quad (15)$$

Supponiamo di approssimare questo integrale con una formula di quadratura gaussiana su s nodi, che sarà della forma

$$h \sum_{j=1}^s b_j g(c_j), \quad x_{i-1} < c_j < x_i. \quad (16)$$

Utilizzando la (16) per approssimare l'integrale (14) si ha

$$y_i - y_{i-1} = h \sum_{j=1}^s b_j g(c_j) + O(h^{p+1}), \quad (17)$$

dove p dipende dal grado di precisione della formula di quadratura. Per questa quadratura vengono scelte formule gaussiane che a parità di numero s di valutazioni della funzione g sono molto più precise di quelle newtoniane: ad esempio per la formula di Gauss-Legendre è $p = 2s$. Sono molto usate anche le formule di Radau e di Lobatto per le quali è rispettivamente $p = 2s - 1$ e $p = 2s - 2$.

La (17) ci suggerisce il metodo

$$u_i = u_{i-1} + h \sum_{j=1}^s b_j g(c_j), \quad (18)$$

in cui i nodi c_j e i pesi b_j , $j = 1, \dots, s$, sono noti dalla teoria ma i valori $g(c_j)$ non sono direttamente calcolabili e devono a loro volta essere approssimati. Poiché

$$g(c_j) = f(x_{i-1} + c_j h, y(x_{i-1} + c_j h)),$$

si esprime (in modo simile a quanto fatto sopra)

$$y(x_{i-1} + c_j h) = y(x_{i-1}) + h \int_0^{c_j} g(t) dt, \quad j = 1, \dots, s.$$

Questi integrali vengono quindi approssimati con formule di quadratura della forma

$$h \sum_{r=1}^s a_{jr} g(c_r), \quad (19)$$

in modo da ottenere per $g(c_j)$ l'approssimazione

$$F_j = f(x_{i-1} + c_j h, u_{i-1} + h \sum_{r=1}^s a_{jr} F_r), \quad j = 1, \dots, s. \quad (20)$$

Nella formula (19) i nodi sono prefissati: sono gli stessi c_r , $r = 1, \dots, s$ della formula (16), ma gli estremi di integrazione adesso non sono più 0 e 1 bensì 0 e c_j . In questo caso perciò non è possibile servirsi di pesi a_{jr} già calcolati nella teoria dell'integrazione numerica, ma occorre calcolarli specificatamente per queste formule, imponendo la condizione che la (19) abbia grado di precisione quanto più elevato possibile, cioè che integri esattamente i polinomi t^k per $k = 0, \dots, s-1$. Queste condizioni si traducono nel sistema lineare

$$\sum_{r=1}^s a_{jr} c_r^k = \frac{c_j^{k+1}}{k+1}, \quad \text{dove } k = 0, \dots, s-1. \quad (21)$$

Riassumendo, un metodo di Runge-Kutta implicito ad s stadi ricavato dalle formule di quadratura di Gauss-Legendre viene costruito in questo modo:

- a) i b_i e i c_i , $i = 1, \dots, s$, sono i pesi e i nodi della formula di Gauss-Legendre su s nodi, riferita all'intervallo di integrazione $[0, 1]$ (attenzione: di solito le formule gaussiane sono riferite all'intervallo $[-1, 1]$);
- b) gli a_{jr} sono calcolati risolvendo gli s sistemi lineari (21).

Il metodo ottenuto viene detto di *Runge-Kutta-Gauss* ad s stadi.

Poiché il grado di precisione delle formule di quadratura è almeno 0, dalla (16) risulta che se $g(t) \equiv 1$, l'integrale (15) è calcolato esattamente dalla (16), cioè che

$$\sum_{i=1}^s b_i = 1.$$

Perciò i metodi ricavati dalle formule di quadratura sono consistenti e per il teorema 4.3 sono convergenti.

5.7 Esempio. Ricaviamo il metodo di Runge-Kutta-Gauss a 2 stadi usando la formula di quadratura di Gauss-Legendre a due punti (19) dell'Appendice:

$$\int_{-1}^1 f(x) dx = f\left(-\frac{\sqrt{3}}{3}\right) + f\left(\frac{\sqrt{3}}{3}\right) + O(h^5).$$

Per prima cosa si deve trasformare l'intervallo $[-1, 1]$ nell'intervallo $[0, 1]$, ponendo $t = \frac{x+1}{2}$; si ottiene la formula gaussiana

$$\int_0^1 g(t) dt = \frac{1}{2} \left[g\left(\frac{3-\sqrt{3}}{6}\right) + g\left(\frac{3+\sqrt{3}}{6}\right) \right],$$

da cui si ricavano $b_1 = b_2 = \frac{1}{2}$, $c_1 = \frac{3-\sqrt{3}}{6}$, $c_2 = \frac{3+\sqrt{3}}{6}$. Gli a_{jr} si determinano risolvendo il sistema (21). Per $j = 1$ si ha

$$\begin{cases} a_{11} + a_{12} = \frac{3-\sqrt{3}}{6}, \\ \frac{3-\sqrt{3}}{6} a_{11} + \frac{3+\sqrt{3}}{6} a_{12} = \frac{1}{2} \left(\frac{3-\sqrt{3}}{6}\right)^2; \end{cases}$$

per $j = 2$ si ha

$$\begin{cases} a_{21} + a_{22} = \frac{3+\sqrt{3}}{6}, \\ \frac{3-\sqrt{3}}{6} a_{21} + \frac{3+\sqrt{3}}{6} a_{22} = \frac{1}{2} \left(\frac{3+\sqrt{3}}{6}\right)^2, \end{cases}$$

da cui si ricava

$$a_{11} = a_{22} = \frac{1}{4}, \quad a_{12} = \frac{3-2\sqrt{3}}{12}, \quad a_{21} = \frac{3+2\sqrt{3}}{12}.$$

Si è così riottenuto il metodo di ordine 4 già ricavato nell'esempio 5.4. Il primo metodo semi-implicito ottenuto nell'esempio 5.4 si può ricavare in modo analogo usando per la quadratura una formula di Radau a due punti. ■

Da questo esempio risulta evidente che, a differenza dei metodi espliciti, i metodi di Runge-Kutta impliciti sono facili da costruire, dato che le formule di Gauss-Legendre sono disponibili per qualunque ordine. Inoltre i metodi così ottenuti hanno il massimo ordine possibile, tenuto conto del numero di parametri. Vale infatti il seguente teorema [Butcher].

5.8 Teorema. Per ogni intero $s > 0$ esiste ed è unico il metodo di Runge-Kutta implicito ad s stadi di ordine $2s$. Tale metodo è quello ricavato dalla formula di quadratura di Gauss-Legendre su s nodi. ■

Per i metodi semi-impliciti è stato dimostrato che con s stadi si possono ottenere metodi di ordine $s + 1$.

Per quanto riguarda la stima dell'errore locale, purtroppo per i metodi impliciti e semi-impliciti non è possibile adottare una tecnica analoga a quella di Fehlberg vista nel par. 2, per cui il controllo sulla lunghezza del passo deve essere fatto con la tecnica di Richardson.

Applicando le (8) al problema test A si ottiene ancora la (5), in cui però adesso A non è più strettamente triangolare inferiore. Per $|z|$ sufficientemente piccolo la matrice $I - zA$ è non singolare e gli elementi di $(I - zA)^{-1}$ sono funzioni razionali di z . Nel caso di un metodo di Runge-Kutta-Gauss la matrice A è piena e gli elementi di $(I - zA)^{-1}$ sono uguali al rapporto di un polinomio di grado $s - 1$ in z e del polinomio $\det(I - zA)$ di grado s in z . Risulta perciò

$$u_i = u_{i-1} \frac{n(z)}{d(z)}, \quad (22)$$

dove $n(z)$ e $d(z)$ sono polinomi di grado s in z . D'altra parte sappiamo che

$$y_i = y_{i-1} e^z.$$

Scrivendo la formula di Padé di e^z risulta

$$e^z = \frac{p_s(z)}{q_s(z)} + O(z^{2s+1}),$$

dove $p_s(z)$ e $q_s(z)$ sono due polinomi di grado s . Quindi $\frac{n(z)}{d(z)}$ risulta proprio l'approssimante di Padé di grado (s, s) . Ad esempio dalla tabella di fig.2.1 risulta

$$\frac{p_1(z)}{q_1(z)} = \frac{1 + \frac{1}{2} z}{1 - \frac{1}{2} z}, \quad \frac{p_2(z)}{q_2(z)} = \frac{1 + \frac{1}{2} z + \frac{1}{12} z^2}{1 - \frac{1}{2} z + \frac{1}{12} z^2},$$

$$\frac{p_3(z)}{q_3(z)} = \frac{1 + \frac{1}{2}z + \frac{1}{10}z^2 + \frac{1}{120}z^3}{1 - \frac{1}{2}z + \frac{1}{10}z^2 - \frac{1}{120}z^3}.$$

Queste approssimanti di Padé possono essere scritte nella forma

$$\frac{p_i(z)}{q_i(z)} = \frac{a(z) + zb(z)}{a(z) - zb(z)},$$

dove $a(z)$ e $b(z)$ sono polinomi pari in z . Posto $w = \frac{a(z)}{zb(z)}$, risulta

$$\frac{p_i(z)}{q_i(z)} = \frac{w + 1}{w - 1}.$$

Per studiare la stabilità assoluta dei metodi di Runge-Kutta-Gauss basta perciò vedere dove

$$\left| \frac{w + 1}{w - 1} \right| < 1, \quad \text{cioè} \quad |w + 1| < |w - 1|.$$

Si può dimostrare [Butcher] che per z tale che $Re(z) < 0$ risulta $Re(w) < 0$, e quindi $|w + 1| < |w - 1|$. Ne segue che se $Re(\lambda) < 0$ questi metodi sono assolutamente stabili per ogni h , cioè sono A-stabili.

5.9 Esempio. Si applica il metodo implicito di ordine 4 ricavato nell'esempio 5.4 al problema test C, già risolto nell'esempio 5.3. Inizialmente si fissa $h = 0.01$; l'errore effettivo della soluzione calcolata scende da $5.4 \cdot 10^{-4}$ al primo passo a $6.7 \cdot 10^{-7}$ al decimo passo per $x_i = 0.1$. A questo punto si aumenta il passo a $h = 0.05$; l'errore effettivo si mantiene più o meno costante dell'ordine di 10^{-6} nei passi successivi. Però la scelta di $\mathbf{F}^{(0)}$ è critica: assumendo per $\mathbf{F}^{(0)}$ i valori calcolati al passo precedente sono richieste fino a 90 iterazioni. ■

Naturalmente non tutti i metodi impliciti sono A-stabili. Anche per i metodi semi-impliciti la situazione varia. Applicando un metodo semi-implicito al problema test A si ottiene ancora la (22), in cui però i gradi di $n(z)$ e $d(z)$ possono essere diversi. Se il grado di $n(z)$ è maggiore del grado di $d(z)$ (analogamente a quanto accade per i metodi impliciti) il metodo non è A-stabile.

Per il primo metodo semi-implicito dell'esempio 5.4 si ha

$$u_i = u_{i-1} r_1(z), \quad r_1(z) = \frac{1 + \frac{2}{3}z + \frac{1}{6}z^2}{1 - \frac{1}{3}z},$$

($r_1(z)$ è l'approssimante di Padé di grado (2,1) di e^z), e quindi il metodo non è A-stabile. Se $\lambda < 0$ è reale, il metodo è assolutamente stabile per

$$h|\lambda| < 6.$$

Il secondo metodo semi-implicito dell'esempio 5.4 è invece A-stabile. Infatti risulta

$$u_i = u_{i-1} r_2(z), \quad r_2(z) = \frac{1 - \frac{\sqrt{3}}{3}z - \frac{1 + \sqrt{3}}{6}z^2}{1 - \frac{3 + \sqrt{3}}{3}z + \frac{2 + \sqrt{3}}{6}z^2},$$

e, posto $t = \frac{3 + \sqrt{3}}{6}z$, si ha $r_2(t) = \frac{1 - (\sqrt{3} - 1)(t + t^2)}{(1 - t)^2}$.

In questa frazione il modulo del denominatore è maggiore del modulo del numeratore per $Re(t) < 0$, come si vede ponendo $t = x + iy$

$$\begin{aligned} |(1 - t)^2|^2 - |1 - (\sqrt{3} - 1)(t + t^2)|^2 &= (2\sqrt{3} - 3)[(x^2 + y^2)^2 \\ &\quad + 4(2 + \sqrt{3})x^2 - 2(1 + \sqrt{3})x(1 + 2x^2 + 2y^2)], \end{aligned}$$

che è maggiore di zero per $x < 0$. Quindi $|r_2(z)| < 1$ per ogni z la cui parte reale è negativa.

5.1 Si applichino metodi di Runge-Kutta di vario ordine ai problemi dell'esercizio 3.1 e si confrontino i risultati ottenuti.

5.2 Si dia una interpretazione geometrica del metodo di Eulero modificato e del metodo di Heun.

5.3 Si verifichi che se la funzione f non dipende esplicitamente da y il metodo classico di Runge-Kutta del quarto ordine fornisce lo stesso risultato

che si otterrebbe integrando la f con la formula di Newton-Cotes dei tre punti.

5.4 Si dica qual è l'ordine del metodo semi-implicito

| | | | |
|---------------|---------------|---------------|---------------|
| 0 | 0 | | |
| $\frac{1}{2}$ | $\frac{1}{4}$ | $\frac{1}{4}$ | |
| 1 | 0 | 1 | 1 |
| | $\frac{1}{6}$ | $\frac{4}{6}$ | $\frac{1}{6}$ |

e se ne studi la stabilità.

(Traccia: l'ordine è 4; applicando il metodo al problema test A si ottiene

$$u_i = u_{i-1}r(x), \quad \text{dove} \quad r(x) = \frac{1 + \frac{3}{4}z + \frac{1}{4}z^2 + \frac{1}{24}z^3}{1 - \frac{1}{4}z}, \quad z = h\lambda.$$

Per $\lambda < 0$ reale, il metodo è assolutamente stabile se $h|\lambda| < 5.42$.)

5.5 Si scrivano le relazioni che legano i coefficienti di un metodo di Runge-Kutta di ordine 3 e di ordine 4.

(Traccia: per l'ordine $p = 3, 4$ si ha

$$\sum_{i=1}^p b_i c_i^r = \frac{1}{r+1}, \quad r = 0, \dots, p-1,$$

$$\sum_{i=1}^p b_i c_i^r \sum_{j=1}^p a_{ij} c_j = \frac{1}{2(r+3)}, \quad r = 0 \text{ se } p = 3, \quad r = 0, 1 \text{ se } p = 4.$$

Inoltre se $p = 4$ anche

$$\sum_{i=1}^4 b_i \sum_{j=1}^4 a_{ij} c_j^2 = \frac{1}{12}, \quad \sum_{i=1}^4 b_i \sum_{j=1}^4 c_j \sum_{k=1}^4 a_{ik} a_{kj} = \frac{1}{24}.)$$

5.6 Si applichi il metodo di Runge-Kutta del quarto ordine al problema

$$\begin{cases} y' = -100y \\ y(0) = 1. \end{cases}$$

Si dica quante operazioni occorrono per ottenere un'approssimazione di $y(1)$. Si applichi poi il metodo di Runge-Kutta-Gauss a due stadi e si confronti la complessità con il caso precedente.