21. VL Einführung in die Num. Mathematik 24.6.09
und 22. VL, 29.6.09

Verfahren höherer Ordnung

Die bisher besprochenen Methoden (Euler, Heun) haben wir weitestgehend intuitiv ermittelt. Um systematisch Einschrittverfahren höherer Ordnung zu konstruieren, betrachten wir die zum AWP $y' = f(t, y)$, $y(a) = y_0$ äquivalente Gleichung (nach Integration)

$$y(t) = y_0 + \int_a^t f(s, y(s))\, ds \qquad (6.16)$$

bzw. für eine Diskretisierung des Intervalls $[a, b]$

$$y(t_{k+1}) = y(t_k) + \int_{t_k}^{t_{k+1}} f(s, y(s))\, ds. \qquad (6.17)$$

Das letzte Integral aus (6.17) approximieren wir durch eine Quadraturformel

$$\int_{t_k}^{t_{k+1}} f(s, y(s))\, ds \approx h_k \sum_{\ell=1}^{m} \gamma_\ell f(s_\ell, y(s_\ell)), \qquad (6.18)$$

wobei die s_ℓ zu einer Zerlegung von $[t_k, t_{k+1}]$ gehören. (6.17) und (6.18) ergeben

$$y(t_{k+1}) \approx y(t_k) + h_k \sum_{\ell=1}^{m} \gamma_\ell f(s_\ell, y(s_\ell)), \qquad (6.19)$$

wobei wir die Werte $y(s_\ell)$ nicht kennen. Sie müssen näherungsweise aus $y(t_k)$ bestimmt werden, damit (6.19) als Integrationsverfahren benutzt werden kann.

Wählt man z.B. $m=2$ und $\gamma_1=\gamma_2=1/2$ sowie $s_1=t_k$ und $s_2=t_{k+1}$, dann bedeutet (6.19)

$$y(t_{k+1}) \approx y(t_k) + \frac{h_k}{2}\left[f(t_k,y(t_k)) + f(t_{k+1},y(t_{k+1}))\right]$$

und mit der Approximation

$$y(t_{k+1}) \approx y(t_k) + h_k f(t_k,y(t_k))$$

ergibt sich mit

$$y(t_{k+1}) \approx y(t_k) + \frac{h_k}{2}\left[f(t_k,y(t_k)) + f(t_{k+1},y(t_k)+h_k f(t_k,y(t_k)))\right]$$

die Grundlage für das Verfahren von Heun.

Im Weiteren wollen wir mit y_k die Verfahrenswerte zur Näherung der exakten Werte $y(t_k)$ betrachten und als Näherungen von $f(s_l,y(s_l))$

$$f(s_l,y(s_l)) \approx k_l(t_j,y_j)$$

verwenden. Mit

$$s_l = t_k + \alpha_l h_k, \quad \alpha_l = \sum_{r=1}^{l-1} \beta_{lr}$$

werden die k_l rekursiv definiert:

$$\begin{aligned}
k_1(t_k,y_k) &= f(t_k,y_k)\\
k_2(t_k,y_k) &= f(t_k+\alpha_2 h_k, y_k+h_k\beta_{21}k_1(t_k,y_k))\\
k_3(t_k,y_k) &= f(t_k+\alpha_3 h_k, y_k+h_k(\beta_{31}k_1+\beta_{32}k_2)) \qquad (6.20)\\
&\vdots\\
k_m(t_k,y_k) &= f(t_k+\alpha_m h_k, y_k+h_k(\beta_{m1}k_1+\dots+\beta_{m,m-1}k_{m-1})).
\end{aligned}$$

Ausgehend von (6.19) und (6.20) wird durch

$$y_{k+1} = y_k + h_k\big(\gamma_1 k_1(t_k,y_k) + \dots + \gamma_m k_m(t_k,y_k)\big) \qquad (6.21)$$

ein explizites numerisches Verfahren zur Lösung des AWP $y' = f(t,y)$, $y(a) = y_0$ definiert.

Definition 6.11

Das Verfahren (6.21) heißt <u>m-stufiges Runge-Kutta-Verfahren</u> mit k_ℓ aus (6.20) und die k_ℓ heißen Stufenwerte.

Bemerkung

Wir haben oben schon festgestellt, dass im Fall $m=2$ mit $\gamma_1 = \gamma_2 = \frac{1}{2}$, $\alpha_2 = 1$, $\beta_{21} = 1$ (6.21) gerade das Heun-Verfahren ergibt, also ein Verfahren mit der Konsistenzordnung $p=2$. Wir werden nun Bedingungen für die freien Parameter im Verfahren (6.21) formulieren, so dass einmal ein konsistentes Verfahren ($p \geq 1$) entsteht und andererseits eine möglichst große Konsistenzordnung erhalten wird.

Aus der Verwendung der Quadraturformel

$$h_k \sum_{\ell=1}^{m} \gamma_\ell f(s_\ell, y(s_\ell)) \approx \int_{t_k}^{t_{k+1}} f(s,y(s))\,ds$$

folgt die sinnvolle Forderung

$$1 = \gamma_1 + \gamma_2 + \dots + \gamma_m, \qquad (6.22)$$

also haben die γ_ℓ die Funktion von Gewichten

Fordert man vom Verfahren (6.21), dass die Dgl. $y' = 1$ (y linear) exakt integriert wird, ergibt sich die Bedingung

$$\alpha_\ell = \beta_{\ell 1} + \dots + \beta_{\ell \ell-1} \,. \qquad (6.23)$$

Es ist nämlich $f(t,y) \equiv 1$ und damit $k_\ell \equiv 1$ f.a. ℓ. Ausgangspunkt war

$$k_\ell(t_k, y_k) \approx f(s_\ell, y(s_\ell))$$

und

$$k_\ell \approx f(t_k + \alpha_\ell h_k, y(t_k) + h_k(\beta_{\ell 1} k_1 + \dots + \beta_{\ell \ell-1} k_{\ell-1})).$$

Also steht das y-Argument für $y(s_\ell) = y(t_k + \alpha_\ell h_k)$. Wir fordern, dass dies bei $f \equiv 1$ exakt ist, also

$$y(s_\ell) = y(t_k) + h_k(\beta_{\ell 1} + \dots + \beta_{\ell \ell-1}), \qquad (6.24)$$

da alle $k_r = 1$ sind. Andererseits ist y als exakte Lösung linear, d.h.

$$y(s_\ell) = y(t_k) + \alpha_\ell h_k, \qquad (6.25)$$

und aus dem Vergleich von (6.24), (6.25) folgt

$$\alpha_\ell = \beta_{\ell 1} + \dots + \beta_{\ell \ell-1}.$$

Definition 6.12

Die Tabelle mit den Koeffizienten $\alpha_\ell, \beta_{\ell r}, \gamma_r$ in der Form

$$
\begin{array}{c|cccc}
0 & & & & \\
\alpha_2 & \beta_{21} & & & \\
\alpha_3 & \beta_{31} & \beta_{32} & & \\
\vdots & \vdots & \vdots & \ddots & \\
\alpha_m & \beta_{m1} & \beta_{m2} & \cdots & \beta_{m\,m-1} \\
\hline
 & \gamma_1 & \gamma_2 & \gamma_{m-1} & \gamma_m
\end{array}
\tag{6.26}
$$

heißt Butcher-Tabelle und beschreibt das Verfahren (6.21). α_1 ist hier gleich 0, weil explizite Verfahren betrachtet werden.

Satz 6.13

Ein explizites Runge-Kutta-Verfahren (6.21), dessen Koeffizienten die Bedingungen (6.22) und (6.23) erfüllen, ist konsistent.

Beweis

Es ist zu zeigen, dass der lokale Diskretisierungsfehler die Ordnung $O(h_k^{p+1})$ mit $p \geq 1$ hat.

Wir setzen $h_k =: h$, da k fest fixiert ist.

$$
\begin{aligned}
|d_{k+1}| &= |y(t_{k+1}) - y(t_k) - h\phi(t_k, y(t_k), h)| \\
&= \Big|y(t_{k+1}) - y(t_k) - h\sum_{r=1}^{m} \gamma_r k_r(t_k, y(t_k))\Big| \\
&\overset{(6.22)}{=} \Big|y(t_{k+1}) - y(t_k) - hf(t_k, y(t_k)) - h\Big(\sum_{r=1}^{m} \gamma_r k_r(t_k, y(t_k)) - f(t_k, y(t_k))\Big)\Big| \\
&\leq \underbrace{|y(t_{k+1}) - y(t_k) - hy'(t_k)|}_{O(h^2)} + h\Big|\sum_{r=1}^{m} \gamma_r \underbrace{\big(k_r(t_k, y(t_k)) - f(t_k, y(t_k))\big)}_{O(h) \text{ aufgrund von (6.23), s.a. Satz 6.10}}\Big|
\end{aligned}
$$

also

$$|d_{k+1}| \leq Ch^2.$$

$\square$

Bemerkung

Butcher hat bewiesen, wie groß die maximale Ordnung ist, welche mit einem m-stufigen Runge-Kutta-Verfahren erreichbar ist, was in der folgenden Tabelle notiert ist:

m	1	2	3	4	5	6	7	8	9	für $m \geq 9$
p	1	2	3	4	4	5	6	6	7	$p < m-2$

Einige konkrete Runge-Kutta-Verfahren und deren Butcher Tabellen

i) Euler-Verfahren

$$\begin{array}{c|c} 0 & \\ \hline & 1 \end{array} \qquad m=1,\ \gamma_1 = 1$$

$$y_{k+1} = y_k + h_k \underbrace{f(t_k, y_k)}_{k_1} \qquad ,\ p=1$$

ii) Modifiziertes Euler-Verfahren

$$\begin{array}{c|cc} 0 & & \\ \frac{1}{2} & \frac{1}{2} & \\ \hline & 0 & 1 \end{array} \qquad m=2,\ \gamma_1 = 0,\ \gamma_2 = 1,\ \alpha_2 = \frac{1}{2},\ \beta_{21} = \frac{1}{2}$$

$$k_1 = f(t_k, y_k)$$

$$k_2 = f\left(t_k + \frac{h_k}{2}, y_k + \frac{h_k}{2} k_1\right)$$

$$y_{k+1} = y_k + h_k k_2 \qquad ,\ p=2$$

iii) Verfahren von Runge von 3. Ordnung

$$\begin{array}{c|ccc} 0 & & & \\ \frac{1}{2} & \frac{1}{2} & & \\ 1 & 0 & 1 & \\ \hline & 0 & 0 & 1 \end{array} \qquad m=3,\ \gamma_1=\gamma_2=0,\ \gamma_3=1,\ \alpha_2=\frac{1}{2},\ \alpha_3=1,\ \beta_{21}=\frac{1}{2},\ \beta_{31}=0,\ \beta_{32}=1$$

$$k_1 = f(t_k, y_k)$$
$$k_2 = f\left(t_k + \frac{h_k}{2}, y_k + \frac{h_k}{2} k_1\right)$$
$$k_3 = f(t_k + h_k, y_k + h_k k_2)$$

$$y_{k+1} = y_k + h_k k_3 \quad , \; p = 3$$

iv) Klassisches Runge-Kutta-Verfahren 4. Ordnung

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

$m = 4$

$\alpha_2 = \alpha_3 = \frac{1}{2}, \alpha_4 = 1$

$\beta_{21} = \frac{1}{2}, \; \beta_{31} = 0, \; \beta_{32} = \frac{1}{2}$

$\beta_{41} = \beta_{42} = 0, \; \beta_{43} = 1$

$\gamma_1 = \gamma_4 = \frac{1}{6}, \; \gamma_2 = \gamma_3 = \frac{1}{3}$

$$k_1 = f(t_k, y_k)$$
$$k_2 = f\left(t_k + \frac{h_k}{2}, y_k + \frac{h_k}{2} k_1\right)$$
$$k_3 = f\left(t_k + \frac{h_k}{2}, y_k + \frac{h_k}{2} k_2\right)$$
$$k_4 = f(t_k + h_k, y_k + h_k k_3)$$
$$y_{k+1} = y_k + h_k\left(\frac{1}{6}k_1 + \frac{1}{3}k_2 + \frac{1}{3}k_3 + \frac{1}{6}k_4\right), \; p = 4$$

Bemerkung

Die Ordnung eines konkreten Runge-Kutta-Verfahrens kann mit Hilfe von Taylor-Entwicklungen ermittelt werden, wobei man dabei von einer geeigneten Glattheit von $f(t,y)$ ausgeht.

Im Folgenden soll die Ordnung eines 3-stufigen expliziten Runge-Kutta-Verfahrens bestimmt werden. (8)

Satz 6.14

Sei f dreimal stetig partiell differenzierbar und gelte für die Parameter

$$\alpha_2 = \beta_{21}$$
$$\alpha_3 = \beta_{31} + \beta_{32}$$
$$\gamma_1 + \gamma_2 + \gamma_3 = 1$$

sowie

$$\alpha_2\gamma_2 + \alpha_3\gamma_3 = \frac{1}{2}$$
$$\alpha_2\gamma_3\beta_{32} = \frac{1}{6}$$
$$\alpha_2^2\gamma_2 + \alpha_3^2\gamma_3 = \frac{1}{3}.$$

Dann hat das Runge-Kutta-Verfahren (expl., 3-stufig) die Fehlerordnung $p=3$.

Beweis:

Grundlage für den Beweis ist die Taylorapproximation (6.27)

$$f(t+\Delta t, y+\Delta y) = f(t,y) + \begin{pmatrix} \frac{\partial f}{\partial t}(t,y) \\ \frac{\partial f}{\partial y}(t,y) \end{pmatrix} \cdot \begin{pmatrix} \Delta t \\ \Delta y \end{pmatrix} + \frac{1}{2}(\Delta t \;\; \Delta y) \begin{pmatrix} \frac{\partial^2 f}{\partial t^2}(t,y) & \frac{\partial^2 f}{\partial t \partial y}(t,y) \\ \frac{\partial^2 f}{\partial y \partial t}(t,y) & \frac{\partial^2 f}{\partial y^2}(t,y) \end{pmatrix} \begin{pmatrix} \Delta t \\ \Delta y \end{pmatrix} + O(\Delta t^3, \Delta t^2\Delta y, \Delta t\Delta y^2, \Delta y^3)$$

der Funktion f, wobei $\frac{\partial^2 f}{\partial t \partial y} = \frac{\partial^2 f}{\partial y \partial t}$ aufgrund der Glattheit von f gilt.

Mit $\bar{k}_1 = f(t_k, y(t_k))$, $\bar{k}_2 = f(t_k + \alpha_2 h, y(t_k) + \alpha_2 h \bar{k}_1)$,

$\bar{k}_3 = f(t_k + \alpha_3 h, y(t_k) + h(\beta_{31}\bar{k}_1 + \beta_{32}\bar{k}_2))$

gilt es, den lokalen Diskretisierungsfehler

$$d_{k+1} = y(t_{k+1}) - y(t_k) - h\{\gamma_1\bar{k}_1 + \gamma_2\bar{k}_2 + \gamma_3\bar{k}_3\}$$

abzuschätzen, wobei schon $\alpha_2 = \beta_{21}$ verwendet wurde ($h = h_k$).
Mit $\Delta t = \alpha_2 h$ und $\Delta y = \alpha_2 h f(t_k, y(t_k))$ ergibt (6.27) für $\bar{k}_2$

$$\bar{k}_2 = f(t_k + \Delta t, y(t_k) + \Delta y)$$

$$= f + \alpha_2 h f_t + \alpha_2 h f \cdot f_y + \frac{1}{2}\alpha_2^2 h^2 f_{tt} + \alpha_2^2 h^2 f f_{ty} + \frac{1}{2}\alpha_2^2 h^2 f^2 f_{yy} + O(h^3)$$

$$=: f + \alpha_2 h F + \frac{1}{2}\alpha_2^2 h^2 G + O(h^3). \qquad (6.28)$$

$f, f_t, f_{tt}, f_{ty}, f_{yy}$ sind dabei die Funktions- bzw. Ableitungswerte an der Stelle $(t_k, y(t_k))$. Für $\bar{k}_3$ erhält man unter Nutzung von (6.28) und (6.27)

$$\bar{k}_3 = f(t_k + \alpha_3 h, y(t_k) + h(\beta_{31}\bar{k}_1 + \beta_{32}\bar{k}_2))$$

$$= f + \alpha_3 h f_t + h(\beta_{31}\bar{k}_1 + \beta_{32}\bar{k}_2) f_y$$

$$+ \frac{1}{2}\alpha_3^2 h^2 f_{tt} + \alpha_3(\beta_{31}\bar{k}_1 + \beta_{32}\bar{k}_2) h^2 f_{ty} + \frac{1}{2}(\beta_{31}\bar{k}_1 + \beta_{32}\bar{k}_2)^2 h^2 f_{yy} + O(h^3)$$

$$= f + h\{\alpha_3 f_t + (\beta_{31} + \beta_{32}) f \cdot f_y\}$$

$$+ h^2\{\alpha_2\beta_{32} F f_y + \frac{1}{2}\alpha_3^2 f_{tt} + \alpha_3(\beta_{31} + \beta_{32}) f f_{ty} + \frac{1}{2}(\beta_{31} + \beta_{32})^2 f^2 f_{yy}\} + O(h^3)$$

$$= f + \alpha_3 h F + h^2\{\alpha_2\beta_{32} F f_y + \frac{1}{2}\alpha_3^2 G\} + O(h^3). \qquad (6.29)$$

Mit (6.28) und (6.29) folgt für den lokalen Diskretisierungsfehler

$$d_{k+1} = h f\{1 - \gamma_1 - \gamma_2 - \gamma_3\} + h^2 F\{\frac{1}{2} - \alpha_2\gamma_2 - \alpha_3\gamma_3\} \qquad (6.30)$$

$$+ h^3\{F f_y[\frac{1}{6} - \alpha_2\gamma_3\beta_{32}] + G[\frac{1}{6} - \frac{1}{2}\alpha_2^2\gamma_2 - \frac{1}{2}\alpha_3^2\gamma_3]\} + O(h^4).$$

Aufgrund der Voraussetzungen werden die Klammerausdrücke gleich Null und es gilt

$$d_{k+1} = O(h^4),$$

also hat das Verfahren die Fehlerordnung $p = 3$. $\square$

Folgerung

Mit Lösungen des Gleichungssystems

$$\begin{aligned} \gamma_1+\gamma_2+\gamma_3&=1\\ \alpha_2\gamma_2+\alpha_3\gamma_3&=\frac{1}{2}\\ \alpha_2\gamma_3\beta_{32}&=\frac{1}{6}\\ \alpha_2^2\gamma_2+\alpha_3^2\gamma_3&=\frac{1}{3} \end{aligned} \qquad (6.31)$$

hat das dazugehörige 3-stufige Runge-Kutta-Verfahren die Fehlerordnung $p=3$, wobei $\beta_{21}=\alpha_2$ ist.
(6.31) hat z.B. mit den Einschränkungen $\alpha_2\neq\alpha_3$ und $\alpha_2\neq\frac{2}{3}$ die ~~zweiparametrische~~ Lösungen~~menge~~

$$\gamma_2=\frac{3\alpha_3-2}{6\alpha_2(\alpha_3-\alpha_2)}, \quad \gamma_3=\frac{2-3\alpha_2}{6\alpha_3(\alpha_3-\alpha_2)},$$

$$\gamma_1=\frac{6\alpha_2\alpha_3+2-3(\alpha_2+\alpha_3)}{6\alpha_2\alpha_3}, \quad \beta_{32}=\frac{\alpha_3(\alpha_3-\alpha_2)}{\alpha_2(2-3\alpha_2)}, \qquad (6.32)$$

für $\alpha_2,\alpha_3\in\mathbb{R}$, also die zweiparametrische Lösungsmenge

$$\mathcal{M}=\{(\gamma_1,\gamma_2,\gamma_3,\alpha_2,\alpha_3,\beta_{32})\,|\,\gamma_1,\gamma_2,\gamma_3,\beta_{32} \text{ gemäß } (6.32), \alpha_2,\alpha_3\in\mathbb{R}, \alpha_2\neq\alpha_3, \alpha\neq 2/3\}$$

Die restlichen Parameter des Verfahrens ergeben sich aus

$$\beta_{21}=\alpha_2, \quad \beta_{31}=\alpha_3-\beta_{32}.$$

Implizite Runge-Kutta-Verfahren

Explizite Verfahren neigen zur Instabilität und damit besteht die Gefahr der Verstärkung von Rundungsfehlern.
Implizite Verfahren erweisen sich als stabil, speziell wenn es sich um die Lösung von AWP mit sogenannten steifen Differentialgleichungen handelt.

Im Unterschied zum Gleichungssystem (6.20) wird beim impliziten Runge-Kutta-Verfahren das Gleichungssystem

$$k_r(t_k, y_k) = f(t_k + \alpha_r h_k, y_k + h_k(\beta_{r1} k_1 + \dots + \beta_{rm} k_m))$$

$$r = 1, \dots, m \qquad (6.33)$$

zur Bestimmung der k_r zugrunde gelegt.
Mit (6.33) wird (6.21) zu einem impliziten Runge-Kutta-Verfahren. Aus (6.33) ergibt sich die Butcher-Tabelle

$$\begin{array}{c|ccc}
\alpha_1 & \beta_{11} & \dots & \beta_{1m} \\
\alpha_2 & \beta_{21} & \dots & \beta_{2m} \\
\vdots & \vdots & & \vdots \\
\alpha_m & \beta_{m1} & \dots & \beta_{mm} \\
\hline
 & \gamma_1 & \dots & \gamma_m
\end{array} \qquad (6.34)$$

Die Überlegung, die bei den expliziten Verfahren die Bedingung (6.23) für die Koeffizienten α_r, β_{rl} gerechtfertigt haben, ergeben analog bei den impliziten Runge-Kutta-Verfahren die Bedingung

$$\alpha_r = \beta_{r1} + \beta_{r2} + \ldots + \beta_{rm}, \quad r=1,\ldots,m. \quad (6.35)$$

Beispiele vom impliziten Runge-Kutta-Verfahren

i)
$$\begin{array}{c|c} \frac{1}{2} & \frac{1}{2} \\ \hline & 1 \end{array}$$

$m=1, \ p=2$

$$k_1 = f\left(t_k + \frac{h_k}{2}, y_k + \frac{h_k}{2} k_1\right)$$

$$y_{k+1} = y_k + h_k k_1$$

ii)
$$\begin{array}{c|cc} \frac{3-\sqrt{3}}{6} & \frac{1}{4} & \frac{3-2\sqrt{3}}{12} \\ \frac{3+\sqrt{3}}{6} & \frac{3+2\sqrt{3}}{12} & \frac{1}{4} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

$m=2, \ p=4$

$$k_1 = f\left(t_k + \frac{3-\sqrt{3}}{6} h_k, y_k + h_k\left(\frac{1}{4}k_1 + \frac{3-2\sqrt{3}}{12} k_2\right)\right)$$

$$k_2 = f\left(t_k + \frac{3+\sqrt{3}}{6} h_k, y_k + h_k\left(\frac{3+2\sqrt{3}}{12} k_1 + \frac{1}{4} k_2\right)\right)$$

$$y_{k+1} = y_k + \frac{h_k}{2}(k_1 + k_2)$$

Zur Lösbarkeit des Gleichungssystems (6.33) gilt der

Satz 6.15

f genüge auf $[a,b] \times \mathbb{R}$ der Lipschitz-Bedingung

$$|f(t,y_1) - f(t,y_2)| \leq L\,|y_1 - y_2|$$

und die Schrittweite $h = h_k$ genüge der Bedingung

$$q = h \cdot L \cdot \max_{1 \leq j \leq m} \left(\sum_{r=1}^{m} |\beta_{jr}|\right) < 1$$

Dann hat (6.33) zur Bestimmung von $k_1, \ldots, k_m$ genau eine Lösung

Beweis Aussage folgt aus dem Banachschen Fixpunktsatz. $\square$