

9 Numerische Verfahren für Anfangswertaufgaben

Inhalt

9.1 Einige einfache Verfahren

9.2 Einschrittverfahren – Definition und Eigenschaften

9.3 Runge-Kutta-Verfahren

9.4 Lineare Mehrschrittverfahren

9.5 Konvergenz von Einschrittverfahren

9.6 Nullstabilität linearer Mehrschrittverfahren

9.7 Schrittweitensteuerung

9.1 Einige einfache Verfahren

Wir betrachten die AWA

$$\mathbf{y}' = \mathbf{f}(t, \mathbf{y}), \quad \mathbf{y}(t_0) = \mathbf{y}_0. \quad (\text{AWA})$$

Unter den Voraussetzungen von Satz 8.1 besitzt es eine eindeutige Lösung, sagen wir über dem Intervall I . Wir wollen diese Lösung $\mathbf{y}(t)$ für $t \in [t_0, t_{\text{end}}] \subseteq I$ durch das Euler-Verfahren^a (auch Euler-Cauchy-Verfahren^b oder Polygonzug-Verfahren), den Prototyp eines numerischen Verfahrens zur Lösung von AWAs, approximieren.

Wie alle numerischen Methoden, die hier besprochen werden, basiert das Euler-Verfahren auf der Idee der Diskretisierung: Statt (AWA) über $[t_0, t_{\text{end}}]$ zu lösen, geben wir uns damit zufrieden, Näherungswerte für die Lösung auf einer diskreten Teilmenge $\{t_n : n = 0, 1, \dots, N\}$ zu bestimmen.

^aLEONHARD EULER (1707–1783)

^bAUGUSTIN LOUIS CAUCHY (1789–1857)

Wähle $N \in \mathbb{N}$, $h = (t_{\text{end}} - t_0)/N$ und definiere $t_n := t_0 + nh$, $n = 0, 1, \dots, N$, d.h. $t_0 < t_1 < \dots < t_{N-1} < t_N = t_{\text{end}}$. Die Zahl $h > 0$ heißt **Schrittweite** (ausser Bequemlichkeit gibt es vorerst keinen Grund, die Schrittweite h **konstant** zu wählen).

Bezeichnet nun \mathbf{y}_n einen Näherungswert für $\mathbf{y}(t_n)$, $n = 0, 1, \dots, N - 1$, dann ist (falls $\mathbf{y} \in C^2[t_0, t_{\text{end}}]$)

$$\begin{aligned} \mathbf{y}(t_{n+1}) &= \mathbf{y}(t_n + h) = \mathbf{y}(t_n) + h\mathbf{y}'(t_n) + \frac{1}{2}h^2\mathbf{y}''(\xi) \\ &\approx \mathbf{y}(t_n) + h\mathbf{y}'(t_n) = \mathbf{y}(t_n) + h\mathbf{f}(t_n, \mathbf{y}(t_n)) \approx \mathbf{y}_n + h\mathbf{f}(t_n, \mathbf{y}_n). \end{aligned}$$

(Explizites) Euler-Verfahren:

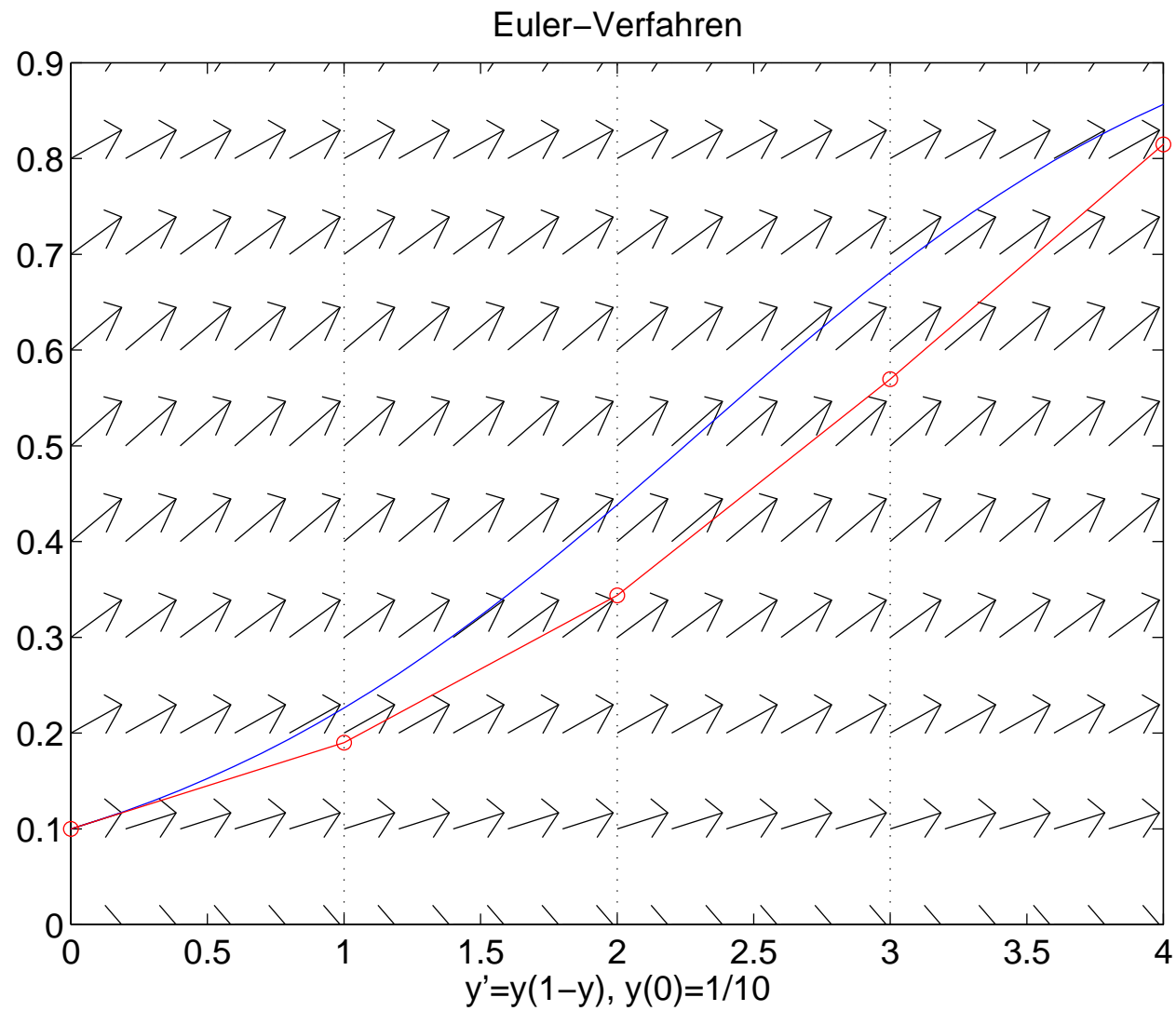
\mathbf{y}_0 gegeben,

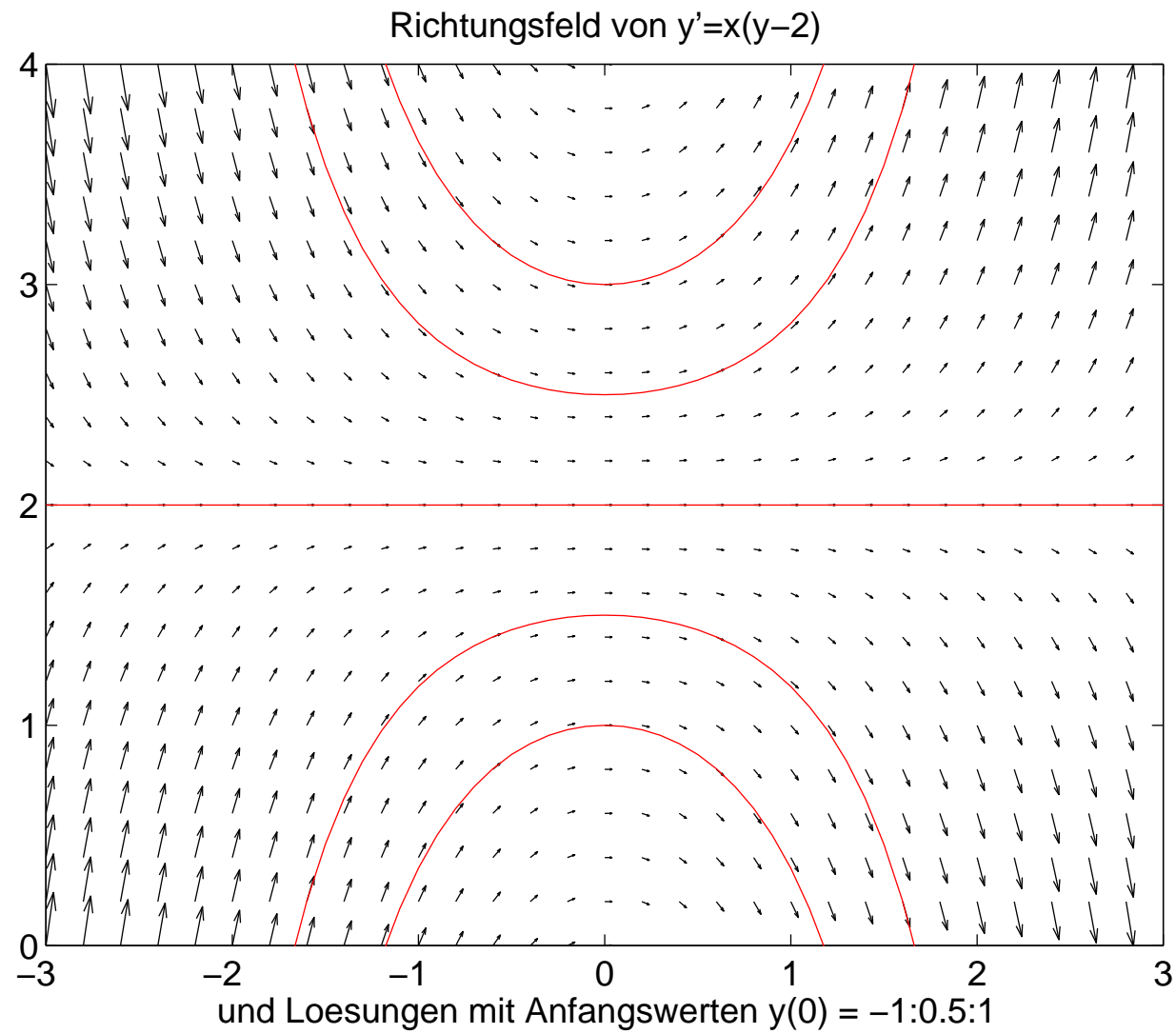
$$\mathbf{y}_{n+1} := \mathbf{y}_n + h\mathbf{f}(t_n, \mathbf{y}_n) \quad (n = 0, 1, \dots, N - 1).$$

Zur Veranschaulichung einer GDG 1. Ordnung $y' = f(t, y)$ wird oft das assoziierte **Richtungsfeld** herangezogen: In jedem Punkt $(t, y) \in \mathbb{R}^2$ wird ein Pfeil gezeichnet, der in die Richtung $y' = f(t, y)$ weist.

Der Graph der Lösung der AWA $y' = f(t, y), y(t_0) = y_0$, muss dann einerseits das Richtungsfeld respektieren (d.h. die Tangenten an den Graphen sind Elemente des Richtungsfelds), andererseits den Punkt (t_0, y_0) enthalten.

Die nächste Abbildung zeigt, wie das Euler-Verfahren die **logistische Gleichung** $y' = y(1 - y)$ mit AB $y(0) = 1/10$ löst. Statt der exakten Trajektorie zu folgen (was natürlich unmöglich ist), produziert das Euler-Verfahren eine „stückweise lineare Lösung“ (einen Polygonzug). An der Stelle $t_0 = 0$ arbeitet das Euler-Verfahren mit der richtigen Steigung $f(t_0, y_0) = 9/100$, bereits an der Stelle $t = 1$ ist die Steigung „falsch“. In späteren Schritten entfernt man sich (potentiell) immer weiter von der exakten Lösung.





Konvergiert das Euler-Verfahren, d.h. strebt die Näherungslösung für $h \rightarrow 0$ gegen die exakte Lösung $\mathbf{y}(t)$?

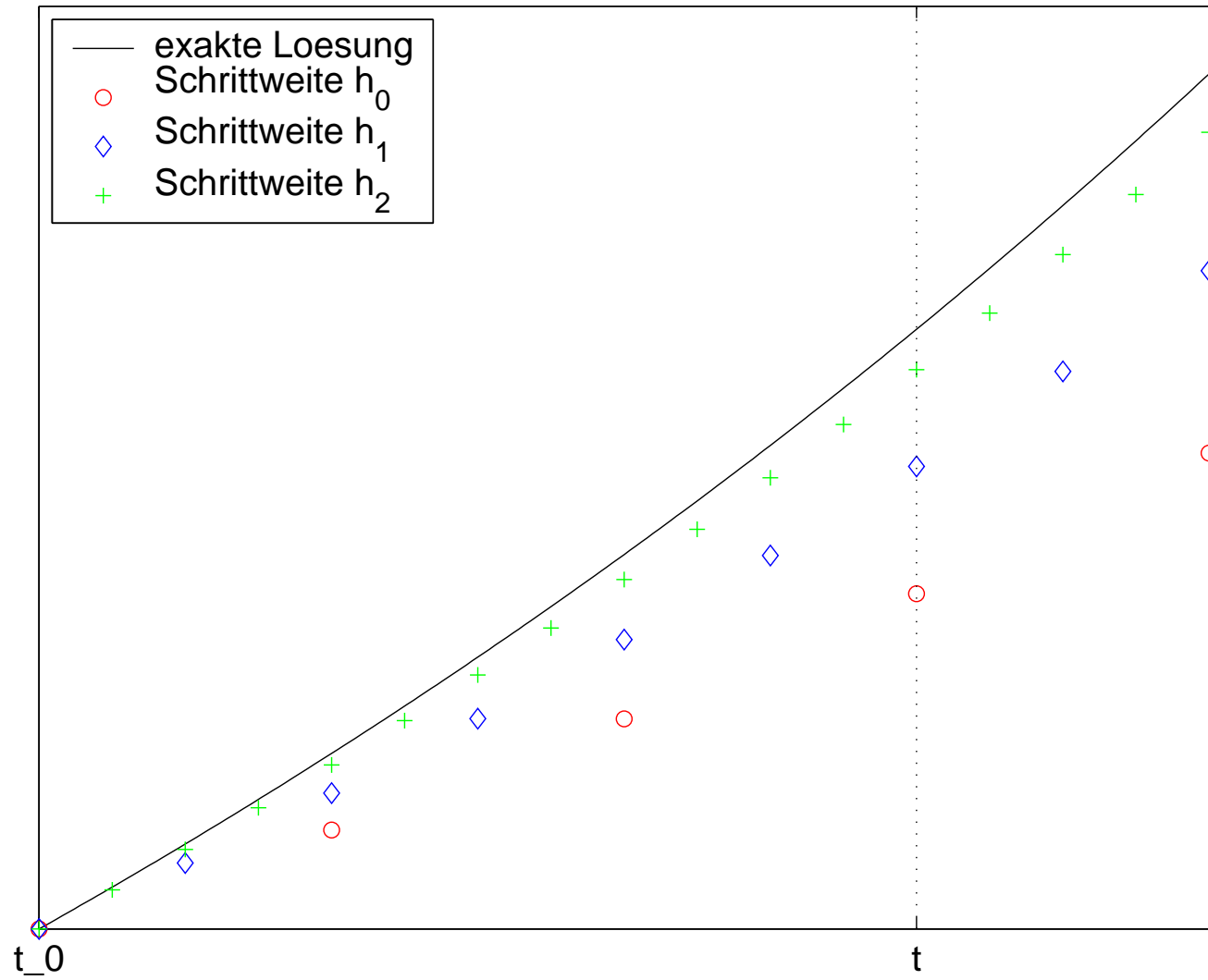
Formal: Zu jedem Wert von $h > 0$ gehört eine Folge von Näherungen $\mathbf{y}_n = \mathbf{y}_n(h)$, $n = 0, 1, \dots, N(h) := \lfloor (t_{\text{end}} - t_0)/h \rfloor$.

Das Verfahren heißt **konvergent** (in $[t_0, t_{\text{end}}]$), wenn gilt

$$\lim_{h \rightarrow 0+} \max_{0 \leq n \leq N(h)} \|\mathbf{y}_n(h) - \mathbf{y}(t_n)\| = 0. \quad (\text{Konv})$$

Satz 9.1 (Konvergenz des Euler-Verfahrens) *Unter den Voraussetzungen von Satz 8.1 konvergiert das Euler-Verfahren, genauer:*

$$\max_{0 \leq n \leq N(h)} \|\mathbf{y}_n(h) - \mathbf{y}(t_n)\| = O(h) \quad \text{für } h \rightarrow 0.$$



Modifiziertes Euler-Verfahren:

y_0 gegeben,

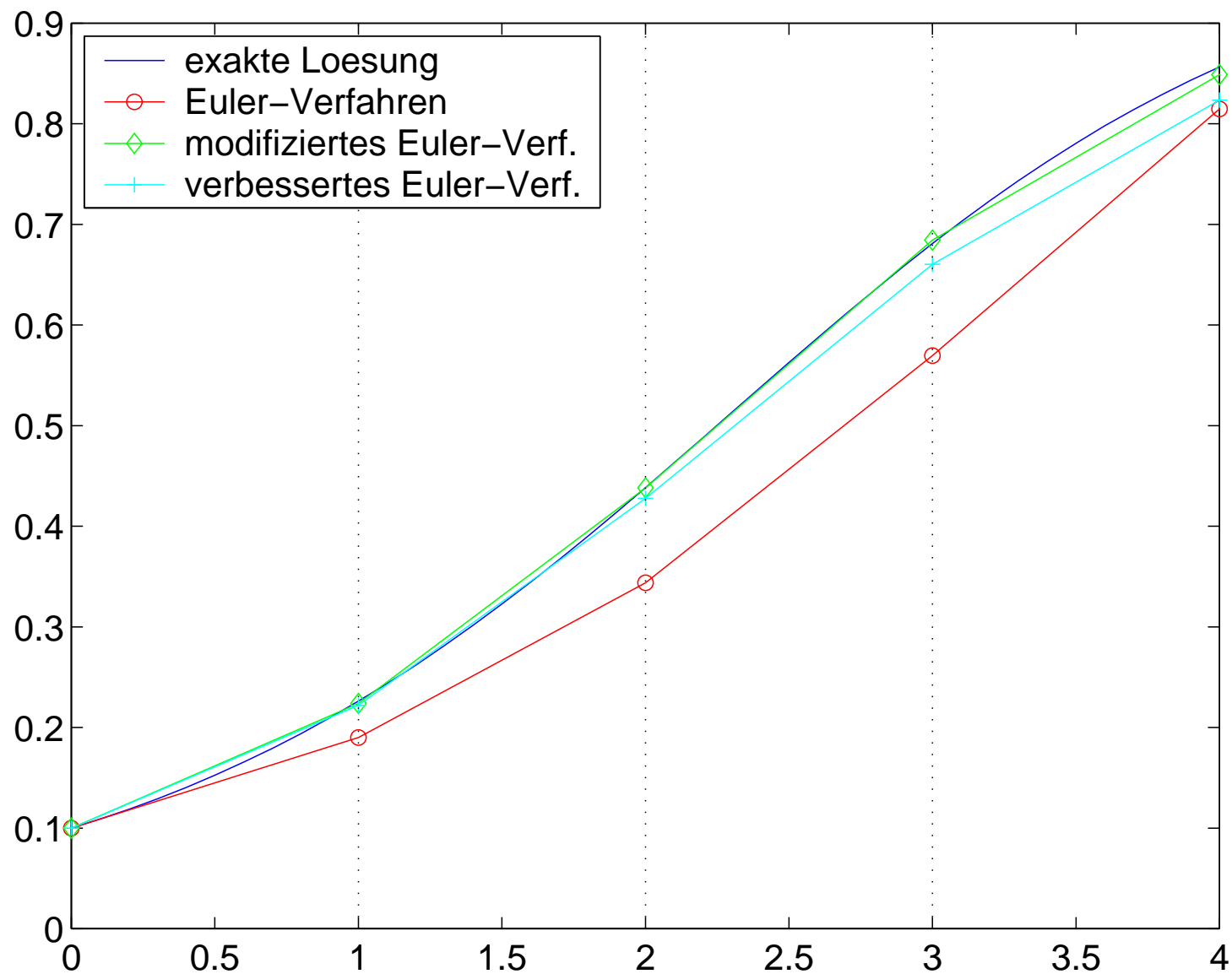
$$y_{n+1} := y_n + h f\left(t_n + \frac{1}{2}h, y_n + \frac{1}{2}h f(t_n, y_n)\right) \quad (n = 0, 1, \dots, N-1).$$

Verbessertes Euler-Verfahren:

y_0 gegeben,

$$y_{n+1} := y_n + \frac{1}{2}h[f(t_n, y_n) + f(t_n + h, y_n + hf(t_n, y_n))]$$
$$(n = 0, 1, \dots, N-1).$$

Die Idee dieser beiden Verfahren macht man sich leicht im Richtungsfeld klar. Auch das modifizierte und das verbesserte Euler-Verfahren konvergieren: Eine triviale Modifikation des Beweises von Satz 9.1 zeigt, dass für diese Verfahren sogar $\max_{0 \leq n \leq N(h)} \|y_n(h) - y(t_n)\| = O(h^2)$ gilt.



Implizites Euler-Verfahren:

y_0 gegeben,

$$y_{n+1} := y_n + hf(t_n + h, y_{n+1}) \quad (n = 0, 1, \dots, N-1).$$

Im Gegensatz zu allen bisher erwähnten Verfahren ist dieses Verfahren **implizit**, d.h. um y_{n+1} zu bestimmen, muss in jedem Schritt ein i.Allg. nichtlineares Gleichungssystem gelöst werden.

Wie in Satz 9.1 zeigt man auch für das implizite Euler-Verfahren

$$\max_{0 \leq n \leq N(h)} \|y_n(h) - y(t_n)\| = O(h) \quad \text{für } h \rightarrow 0.$$

Bisher traten in der Verfahrensgleichung ausschließlich Approximationen an die Lösung an **zwei** aufeinanderfolgenden Zeitpunkten auf, sog. **Ein-schrittverfahren**.

Mehrschrittverfahren dagegen beruhen auf mehrere Stellen umfassenden Differenzenformeln, etwa

$$\frac{y(t+h) - y(t-h)}{2h} = y'(t) + \frac{h^2}{6} y'''(t) + O(h^4),$$
$$\frac{3y(t) - 4y(t-h) + y(t-2h)}{2h} = y'(t) - \frac{h^2}{3} y'''(t) + O(h^3).$$

Diese Formeln führen, bei $t = t_{n+1}$ ausgewertet, zur expliziten

Mittelpunktsregel: (midpoint-rule, leapfrog method)

$$\begin{aligned} & \mathbf{y}_0, \mathbf{y}_1 \text{ gegeben,} \\ & \mathbf{y}_{n+1} = \mathbf{y}_{n-1} + 2h\mathbf{f}(t_n, \mathbf{y}_n) \quad (n = 1, 2, \dots, N-1). \end{aligned} \tag{9.1}$$

sowie einem impliziten Verfahren aus der sog. **BDF-Familie** (Backward Differentiation Methods)

Ein BDF-Verfahren:

$$\begin{aligned} & \mathbf{y}_0, \mathbf{y}_1 \text{ gegeben,} \\ & 3\mathbf{y}_{n+1} - 4\mathbf{y}_n + \mathbf{y}_{n-1} = 2h\mathbf{f}(t_n + h, \mathbf{y}_{n+1}) \quad (n = 1, 2, \dots, N-1). \end{aligned}$$

Wie man sieht kann hier die Verfahrensgleichung erst dann ausgewertet werden, wenn neben den Anfangswerten \mathbf{y}_0 auch eine Approximation \mathbf{y}_1 vorliegt. Eine solche **Anlaufrechnung** erfordern alle Mehrschrittverfahren.

9.2 Einschrittverfahren – Definition und Eigenschaften

Ein **Einschrittverfahren** (ESV) hat die Form

$$\mathbf{y}_{n+1} = \mathbf{y}_n + h\Phi_f(\mathbf{y}_{n+1}, \mathbf{y}_n, t_n; h). \quad (\text{ESV})$$

Wir werden ausschließlich Verfahren untersuchen, bei denen die **Verfahrensfunktion** Φ_f die folgenden beiden Eigenschaften besitzt:

$$\Phi_{f \equiv 0}(\mathbf{y}_{n+1}, \mathbf{y}_n, t_n; h) \equiv \mathbf{0} \quad (\text{V}_1)$$

und

$$\|\Phi_f(\mathbf{y}_{n+1}, \mathbf{y}_n, t_n; h) - \Phi_f(\mathbf{y}_{n+1}^*, \mathbf{y}_n^*, t_n; h)\| \leq M \sum_{j=0}^1 \|\mathbf{y}_{n+j} - \mathbf{y}_{n+j}^*\|. \quad (\text{V}_2)$$

Bei „vernünftigen“ ESV ist (V_2) eine Folge der Lipschitz-Stetigkeit von f (vgl. Satz 8.1), die immer vorausgesetzt wird.

Beispiele für etwas kompliziertere ESV:

$$y_{n+1} - y_n = \frac{1}{4}h(k_1 + 3k_3)$$

mit $k_1 = f(t_n, y_n)$, $k_2 = f(t_n + \frac{1}{3}h, y_n + \frac{1}{3}hk_1)$, (Beispiel 1)

$$k_3 = f(t_n + \frac{2}{3}h, y_n + \frac{2}{3}hk_2),$$

ein explizites ESV, das zur Klasse der **Runge-Kutta-Verfahren**^{a b} (vgl. §9.3) gehört und ein implizites ESV der Runge-Kutta-Klasse,

$$y_{n+1} - y_n = \frac{1}{2}h(k_1 + k_2)$$

mit $k_1 = f(t_n, y_n)$, $k_2 = f(t_n + h, y_n + \frac{1}{2}hk_1 + \frac{1}{2}hk_2)$. (Beispiel 2)

^aCARLE DAVID TOLMÉ RUNGE (1856–1927)

^bMARTIN WILHELM KUTTA (1867–1944)

Das Verfahren (**ESV**) heißt **konvergent**, wenn

$$\lim_{\substack{h \rightarrow 0 \\ t=t_0+nh}} \mathbf{y}_n = \lim_{\substack{h \rightarrow 0 \\ t=t_0+nh}} \mathbf{y}_n(h) = \mathbf{y}(t)$$

gilt – und zwar

- für alle AWPes, die den Voraussetzungen von Satz **8.1** genügen ($\mathbf{y}(t)$ bezeichnet die Lösung eines solchen AWPes),
- gleichmäßig für alle $t \in [t_0, t_{\text{end}}]$,
- für alle Lösungen $\{\mathbf{y}_n(h)\} = \{\mathbf{y}_n\}$ von (**ESV**) mit Anfangswert $\mathbf{y}_0(h)$, der $\lim_{h \rightarrow 0} \mathbf{y}_0(h) = \mathbf{y}_0$ erfüllt.

Äquivalent: Der **globale Diskretisierungsfehler** $e_n = e_n(h) := \mathbf{y}(t_n) - \mathbf{y}_n(h)$ strebt gleichmäßig gegen 0 (für $h \rightarrow 0$):

$$\lim_{h \rightarrow 0} \max_{0 \leq n \leq N} \|\mathbf{e}_n(h)\| = \lim_{h \rightarrow 0} \max_{0 \leq n \leq N} \|\mathbf{y}(t_n) - \mathbf{y}_n(h)\| = 0.$$

Wir definieren den **lokalen Diskretisierungsfehler** (lokaler Abbruchfehler) eines Verfahrens $T_n = T_n(h)$ im n -ten Schritt als die Differenz von linker Seite und rechter Seite der Verfahrensgleichung (so skaliert, dass für $h \rightarrow 0$ die Differentialgleichung approximiert wird) wenn anstelle der Näherungslösung die exakte Lösung eingesetzt wird.

Ein Verfahren heißt **konsistent von der Ordnung p** , falls

$$T_n(h) = O(h^p) \quad \text{für } h \rightarrow 0.$$

Beispiel: Mittelpunktsregel

$$\begin{aligned} T_n &= \frac{y(t_{n+1}) - y(t_{n-1}))}{2h} - f(t_n, y(t_n)) \\ &= \left[y'(t_n) + \frac{h^2}{6} y'''(t_n) + O(h^4) \right] - y'(t_n) = \frac{h^2}{6} y'''(t_n) + O(h^4), \end{aligned}$$

d.h. konsistent von der Ordnung $p = 2$.

In der Literatur wird oft auch eine abweichende Definition des lokalen Diskretisierungsfehlers verwandt, und zwar basierend auf der Verfahrensgleichung in der Form (**ESV**), was zur Erhöhung um eine h -Potenz führt.

Motivation: Beschreibung des Fehlers in einem Schritt unter der **Lokalisierungsannahme** $\mathbf{y}_n = \mathbf{y}(t_n)$. In diesem Fall liefert 1 Schritt von (**ESV**) im Vgl. zur exakten Lösung

$$\begin{aligned}\hat{\mathbf{y}}_{n+1}(h) &= \mathbf{y}(t_n) + h\Phi_f(\hat{\mathbf{y}}_{n+1}, \mathbf{y}(t_n), t_n; h), \\ \mathbf{y}(t_{n+1}) &= \mathbf{y}(t_n) + h\Phi_f(\mathbf{y}(t_{n+1}), \mathbf{y}(t_n), t_n; h) + h\mathbf{T}_{n+1}(h).\end{aligned}$$

Wir bezeichnen die Differenz $\mathbf{S}_{n+1}(h) := \mathbf{y}(t_{n+1}) - \hat{\mathbf{y}}_{n+1}$ als **Schrittfehler**. Es gilt (mit (**V₂**))

$$\begin{aligned}\|\mathbf{S}_{n+1}\| &\leq h\|\Phi_f(\mathbf{y}(t_{n+1}), \mathbf{y}(t_n), t_n; h) - \Phi_f(\hat{\mathbf{y}}_{n+1}, \mathbf{y}(t_n), t_n; h)\| + h\|\mathbf{T}_{n+1}\| \\ &\leq hM\|\mathbf{S}_{n+1}\| + h\|\mathbf{T}_{n+1}\|,\end{aligned}$$

d.h. wegen $(1 - hM)\|\mathbf{S}_{n+1}\| \leq h\|\mathbf{T}_{n+1}\|$ verhält sich \mathbf{S}_n für $h \rightarrow 0$ wie $h\mathbf{T}_n$.

Relevanz von S_n : Bei der Integration von (AWA) von t_0 bis t_{end} sind $(t_{\text{end}} - t_0)/h$ Schritte erforderlich. Wird im n -ten Schritt der Schrittfehler S_n begangen, so ergibt sich – unter der vereinfachenden Annahme, dass diese Fehler sich nicht gegenseitig beeinflussen – insgesamt ein akkumulierter (globaler) Fehler von

$$\frac{t_{\text{end}} - t_0}{h} S(h) \sim T(h).$$

Die entscheidende Voraussetzung für diese vereinfachende Annahme ist, wie wir sehen werden, die **Stabilität** des Verfahrens.

9.3 Runge-Kutta-Verfahren

9.3.1 Herleitung mittels Quadraturverfahren

Ausgangspunkt wie immer

$$\mathbf{y}(t+h) = \mathbf{y}(t) + [\mathbf{y}(t+h) - \mathbf{y}(t)] = \mathbf{y}(t) + \int_t^{t+h} \mathbf{y}'(s) ds$$

Substitution: $s = t + \tau h$, $0 \leq \tau \leq 1$

$$= \mathbf{y}(t) + h \int_0^1 \mathbf{y}'(t + \tau h) d\tau.$$

Approximiere durch Quadraturformel

$$\int_0^1 g(\tau) d\tau \approx \sum_{j=1}^m \beta_j g(\gamma_j). \quad (*)$$

Damit zumindest $g \equiv 1$ exakt integriert wird, fordern wir $\sum_{j=1}^m \beta_j = 1$.

Daraus folgt

$$\begin{aligned} \mathbf{y}(t+h) &\approx \mathbf{y}(t) + h \sum_{j=1}^m \beta_j \mathbf{y}'(t + \gamma_j h) \\ &= \mathbf{y}(t) + h \sum_{j=1}^m \beta_j \mathbf{f}(t + \gamma_j h, \mathbf{y}(t + \gamma_j h)) \end{aligned} \quad (\text{RK-1})$$

Problem: $\mathbf{y}(t + \gamma_j h) = \mathbf{y}(t) + h \int_0^{\gamma_j} \mathbf{y}'(t + \tau h) d\tau$ sind unbekannt. Näherungen wieder durch Quadraturformeln, aber mit den alten Knoten γ_j ($j = 1, \dots, m$) aus (*) (sonst würden sich neue 'Unbekannte' $\mathbf{y}(t + \text{Knoten} \cdot h)$ ergeben).

$$\int_0^{\gamma_j} g(\tau) d\tau \approx \sum_{\ell=1}^m \alpha_{j,\ell} g(\gamma_\ell) \quad (j = 1, \dots, m). \quad (**)$$

Damit zumindest $g \equiv 1$ exakt integriert wird, fordern wir $\sum_{\ell=1}^m \alpha_{j,\ell} = \gamma_j$
 $\forall j = 1, \dots, m$.

Damit ergibt sich

$$\begin{aligned}\mathbf{y}(t + \gamma_j h) &\approx \mathbf{y}(t) + h \sum_{\ell=1}^m \alpha_{j,\ell} \mathbf{y}'(t + \gamma_\ell h) \\ &= \mathbf{y}(t) + h \sum_{\ell=1}^m \alpha_{j,\ell} \mathbf{f}(t + \gamma_\ell h, \mathbf{y}(t + \gamma_\ell h))\end{aligned}\quad (\text{RK-2})$$

Abkürzung: $\tilde{\mathbf{k}}_j := \mathbf{f}(t + \gamma_j h, \mathbf{y}(t + \gamma_j h)) \quad (j = 1, \dots, m).$

(RK-2): $\tilde{\mathbf{k}}_j \approx \mathbf{f}\left(t + \gamma_j h, \mathbf{y}(t) + h \sum_{\ell=1}^m \alpha_{j,\ell} \tilde{\mathbf{k}}_\ell\right) \quad (j = 1, \dots, m).$

(RK-1): $\mathbf{y}(t + h) \approx \mathbf{y}(t) + h \sum_{j=1}^m \beta_j \tilde{\mathbf{k}}_j.$

m -stufiges Runge-Kutta-Verfahren (RKV):

$$\mathbf{y}_{n+1} = \mathbf{y}_n + h \sum_{j=1}^m \beta_j \mathbf{k}_j \quad \text{mit} \quad (RKV)$$

$$\mathbf{k}_j = \mathbf{f} \left(t_n + \gamma_j h, \mathbf{y}_n + h \sum_{\ell=1}^m \alpha_{j,\ell} \mathbf{k}_\ell \right) \quad (j = 1, \dots, m).$$

Übliche Darstellung durch
Butcher-Tableau^a

γ_1	$\alpha_{1,1}$	\cdots	$\alpha_{1,m}$
\vdots	\vdots		\vdots
γ_m	$\alpha_{m,1}$	\cdots	$\alpha_{m,m}$
	β_1	\cdots	β_m

^aJOHN CHARLES BUTCHER (*1933)

Beispiele.

0	0	0
1	1	0
	1/2	1/2

symbolisiert ein zweistufiges explizites RKV, nämlich das verbesserte Euler-Verfahren:

$$\mathbf{k}_1 = f(t_n, \mathbf{y}_n),$$

$$\mathbf{k}_2 = f(t_n + h, \mathbf{y}_n + h\mathbf{k}_1),$$

$$\mathbf{y}_{n+1} = \mathbf{y}_n + \frac{h}{2}(\mathbf{k}_1 + \mathbf{k}_2)$$

$$= \mathbf{y}_n + \frac{h}{2} \left(f(t_n, \mathbf{y}_n) + f(t_n + h, \mathbf{y}_n + hf(t_n, \mathbf{y}_n)) \right)$$

Beachte: ein RKV ist **explizit**, wenn $\alpha_{j,\ell} = 0 \ \forall j \leq \ell$ gilt, d.h. wenn die Koeffizienten $\alpha_{j,\ell}$ eine echte untere Dreiecksmatrix bilden.

0	1/4	-1/4	symbolisiert ein zweistufiges implizites RKV:
2/3	1/4	5/12	
	1/4	3/4	

$$k_1 = f \left(t_n, y_n + h \left(\frac{1}{4} k_1 - \frac{1}{4} k_2 \right) \right),$$

$$k_2 = f \left(t_n + \frac{2}{3} h, y_n + h \left(\frac{1}{4} k_1 + \frac{5}{12} k_2 \right) \right),$$

(„zwei“ i.A. nichtlineare Gleichungen für k_1 und k_2)

$$y_{n+1} = y_n + \frac{h}{4} (k_1 + 3k_2).$$

(Beispiel 2 aus §9.2 ist ein weiteres implizites zweistufiges RKV.)

0	0	0	0
1/2	1/2	0	0
1	-1	2	0
<hr/>			
	1/6	4/6	1/6

symbolisiert ein dreistufiges explizites RKV

Verfahren dritter Ordnung von Kutta:

$$\mathbf{k}_1 = \mathbf{f}(t_n, \mathbf{y}_n),$$

$$\mathbf{k}_2 = \mathbf{f}\left(t_n + \frac{1}{2}h, \mathbf{y}_n + \frac{1}{2}h\mathbf{k}_1\right),$$

$$\mathbf{k}_3 = \mathbf{f}\left(t_n + h, \mathbf{y}_n + h(-\mathbf{k}_1 + 2\mathbf{k}_2)\right),$$

$$\mathbf{y}_{n+1} = \mathbf{y}_n + \frac{1}{6}h(\mathbf{k}_1 + 4\mathbf{k}_2 + \mathbf{k}_3).$$

0	0	0	0
1/3	1/3	0	0
2/3	0	2/3	0
	1/4	0	3/4

symbolisiert ein dreistufiges explizites RKV

Verfahren dritter Ordnung von Heun:

$$\begin{aligned}
 k_1 &= f(t_n, y_n), \\
 k_2 &= f\left(t_n + \frac{1}{3}h, y_n + \frac{1}{3}hk_1\right), \\
 k_3 &= f\left(t_n + \frac{2}{3}h, y_n + \frac{2}{3}hk_2\right), \\
 y_{n+1} &= y_n + \frac{h}{4}(k_1 + 3k_3)
 \end{aligned}$$

(vgl. Beispiel 1 aus §9.2).

0	0	0	0	0
1/2	1/2	0	0	0
1/2	0	1/2	0	0
1	0	0	1	0
	1/6	2/6	2/6	1/6

symbolisiert ein vierstufiges explizites RKV

Klassisches Runge-Kutta-Verfahren:

$$k_1 = f(t_n, y_n),$$

$$k_2 = f(t_n + \frac{1}{2}h, y_n + \frac{1}{2}hk_1),$$

$$k_3 = f(t_n + \frac{1}{2}h, y_n + \frac{1}{2}hk_2),$$

$$k_4 = f(t_n + h, y_n + hk_3),$$

$$y_{n+1} = y_n + \frac{1}{6}h(k_1 + 2k_2 + 2k_3 + k_4).$$

Eine alternative Form von (RKV) ist

$$\mathbf{y}_{n+1} = \mathbf{y}_n + h \sum_{j=1}^m \beta_j \mathbf{f}(t_n + \gamma_j h, \tilde{\mathbf{y}}_j) \quad (\text{RKV}^*)$$

mit $\tilde{\mathbf{y}}_j = \mathbf{y}_n + h \sum_{\ell=1}^m \alpha_{j,\ell} \mathbf{f}(t_n + \gamma_\ell h, \tilde{\mathbf{y}}_\ell) \quad (j = 1, \dots, m)$

(setze $\mathbf{k}_j = \mathbf{f}(t_n + \gamma_j h, \tilde{\mathbf{y}}_j)$).

Interpretation:

$\tilde{\mathbf{y}}_j$: Approximationen an die Lösung \mathbf{y} zur Zeit $t_n + \gamma_j h$

\mathbf{k}_j : Approximationen an die Steigungen $\mathbf{y}'(t_n + \gamma_j h)$

9.3.2 Konsistenzordnung bei Runge-Kutta-Verfahren

Jedes RKV hat die Form

$$\mathbf{y}_{n+1} = \mathbf{y}_n + h\Phi_f(\mathbf{y}_{n+1}, \mathbf{y}_n, t_n; h)$$

$$\text{mit } \Phi_f(\mathbf{y}_{n+1}, \mathbf{y}_n, t_n; h) = \sum_{j=1}^m \beta_j \mathbf{k}_j.$$

Ein RKV ist genau dann konsistent (und damit konvergent), wenn

$$\sum_{j=1}^m \beta_j = 1.$$

Um die Konsistenzordnung eines RKVs zu bestimmen (oder um m -stufige RKV mit möglichst hoher Konsistenzordnung zu konstruieren), sind komplizierte Rechnungen erforderlich.

Wir untersuchen als Beispiel explizite dreistufige RKV,

$$\begin{array}{c|ccc}
 0 & 0 & 0 & 0 \\
 \gamma_2 & \gamma_2 & 0 & 0 \\
 \gamma_3 & \gamma_3 - \alpha_{3,2} & \alpha_{3,2} & 0 \\
 \hline
 & \beta_1 & \beta_2 & \beta_3
 \end{array},$$

und entwickeln

$$\begin{aligned}
 \mathbf{T}_{n+1}(h) &= \frac{\mathbf{y}(t_{n+1}) - \mathbf{y}(t_n)}{h} - \Phi_f(\mathbf{y}(t_n), t_n; h) \\
 &= \frac{\mathbf{y}(t_{n+1}) - \mathbf{y}(t_n)}{h} - \sum_{j=1}^3 \beta_j \mathbf{k}_j
 \end{aligned}$$

nach Potenzen von h (unter der Voraussetzung, dass \mathbf{y} bzw. f genügend oft differenzierbar sind).

Für skalare AWA ergibt sich mit den Abkürzungen

$$F := f_t + f_y f \quad \text{und} \quad G := f_{tt} + 2f_{ty}f + f_{yy}f^2$$

(alle Ableitungen von f werden an der Stelle $(t_n, y(t_n))$ ausgewertet) die Beziehung

$$\frac{y(t_{n+1}) - y(t_n)}{h} = f + \frac{1}{2}Fh + \frac{1}{6}(G + f_y F)h^2 + O(h^3).$$

Andererseits ist

$$k_1 = f(t_n, y(t_n)) = f,$$

$$k_2 = f(t_n + h\gamma_2, y(t_n) + h\gamma_2 k_1) = f + h\gamma_2 F + \frac{1}{2}h^2\gamma_2^2 G + O(h^3),$$

$$\begin{aligned} k_3 &= f(t_n + h\gamma_3, y(t_n) + h(\gamma_3 - \alpha_{3,2})k_1 + h\alpha_{3,2}k_2) \\ &= f + h\gamma_3 F + h^2(\gamma_2\alpha_{3,2}Ff_y + \frac{1}{2}\gamma_3^2 G) + O(h^3). \end{aligned}$$

Das bedeutet:

$$\begin{aligned} \mathbf{T}_{n+1}(h) = & \left[1 - \sum_{j=1}^3 \beta_j\right] f + \left[\frac{1}{2} - \beta_2 \gamma_2 - \beta_3 \gamma_3\right] Fh \\ & + \left[\left(\frac{1}{3} - \beta_2 \gamma_2^2 - \beta_3 \gamma_3^2\right) \frac{1}{2} G + \left(\frac{1}{6} - \beta_3 \gamma_2 \alpha_{3,2}\right) F f_y\right] h^2 + O(h^3) \end{aligned}$$

Folgerungen:

1. Das Euler-Verfahren ist das einzige einstufige explizite RKV der Ordnung 1 ($\beta_1 = 1$). Es gibt kein einstufiges explizites RKV höherer Ordnung.
2. Die zweistufigen expliziten RKV der Ordnung 2 sind durch

$$\beta_1 + \beta_2 = 1 \quad \text{und} \quad \beta_2 \gamma_2 = \frac{1}{2}$$

charakterisiert. Beispiele sind das modifizierte ($\beta_1 = 0, \beta_2 = 1, \gamma_2 = \frac{1}{2}$) und das verbesserte Euler-Verfahren ($\beta_1 = \beta_2 = \frac{1}{2}, \gamma_2 = 1$). Kein explizites zweistufiges RKV besitzt die Ordnung 3.

3. Explizite dreistufige RKV der Ordnung 3 sind durch die vier Gleichungen

$$\begin{aligned}\beta_1 + \beta_2 + \beta_3 &= 1, & \beta_2\gamma_2^2 + \beta_3\gamma_3^2 &= \frac{1}{3}, \\ \beta_2\gamma_2 + \beta_3\gamma_3 &= \frac{1}{2}, & \beta_3\gamma_2\alpha_{3,2} &= \frac{1}{6}\end{aligned}$$

charakterisiert. (Man kann zeigen, dass keine dieser Methoden die Ordnung 4 besitzt.) Beispiele sind das Verfahren von Heun ($\beta_1 = \frac{1}{4}$, $\beta_2 = 0$, $\beta_3 = \frac{3}{4}$, $\gamma_2 = \frac{1}{3}$, $\gamma_3 = \alpha_{3,2} = \frac{2}{3}$) und das Verfahren von Kutta ($\beta_1 = \frac{1}{6}$, $\beta_2 = \frac{2}{3}$, $\beta_3 = \frac{1}{6}$, $\gamma_2 = \frac{1}{2}$, $\gamma_3 = 1$, $\alpha_{3,2} = 2$).

4. Ähnliche (kompliziertere) Rechnungen zeigen, dass es eine zweiparametrische Familie expliziter vierstufiger RKV der Ordnung 4 gibt, von denen keines die Ordnung 5 besitzt. Ein Beispiel ist das klassische Runge-Kutta-Verfahren. Weitere Beispiele sind

0	0	0	0	0	
$1/3$	$1/3$	0	0	0	
$2/3$	$-1/3$	1	0	0	(3/8-Regel)
1	1	-1	1	0	
<hr/>					
	$1/8$	$3/8$	$3/8$	$1/8$	

0	0	0	0	0	
$2/5$	$2/5$	0	0	0	
$3/5$	$-3/20$	$3/4$	0	0	(Formel von Kuntzmann).
1	$19/44$	$-15/44$	$40/44$	0	
<hr/>					
	$55/360$	$125/360$	$125/360$	$55/360$	

Die oben beschriebene Methode, die Ordnung eines RKVs zu bestimmen, wird für Verfahren höherer Ordnung schnell unübersichtlich: Die Koeffizienten eines expliziten Verfahrens der Ordnung 3 müssen 4 Gleichungen erfüllen (s.o.), während bei einem Verfahren der Ordnung 8 bereits 200 nicht-lineare Gleichungen überprüft werden müssen. Die sog. Butcher-Theorie (vgl. J. C. Butcher, *The Numerical Analysis of Ordinary Differential Equations. Runge-Kutta and General Linear Methods*. John Wiley & Sons, Chichester 1987) erleichtert mit Hilfe graphentheoretischer Bäume die Buchhaltung bei den partiellen Ableitungen von f und erlaubt eine elegante Berechnung der Ordnung eines gegebenen RKVs (sie liefert aber keine Methode, ein Verfahren mit gewünschter Ordnung zu konstruieren).

Wir beschränken uns hier darauf, *notwendige Ordnungsbedingungen* abzuleiten, die sich aus den speziellen AWPen $y' = y + t^{\ell-1}$, $y(0) = 0$, ($\ell \in \mathbb{N}$) ergeben.

Satz 9.2 (Notwendige Ordnungsbedingungen für RKV) *Das durch das Butcher-Tableau*

$$\begin{array}{c|c} c & A \\ \hline & b^\top \end{array}$$

definierte RKV besitze die Ordnung p . Dann gelten

$$b^\top A^k C^{\ell-1} e = \frac{(\ell-1)!}{(\ell+k)!} = \frac{1}{\ell(\ell+1)\dots(\ell+k)}$$

für $\ell = 1, 2, \dots, p$ und $k = 0, 1, \dots, p - \ell$.

Dabei sind

$$b := [\beta_1, \beta_2, \dots, \beta_m]^\top,$$

$$A := [\alpha_{j,\nu}]_{1 \leq j, \nu \leq m},$$

$$C := \text{diag}(\gamma_1, \gamma_2, \dots, \gamma_m)$$

$$\text{und } e := [1, 1, \dots, 1]^\top \in \mathbb{R}^m.$$

Spezialfälle der notwendigen Bedingungen aus Satz 9.2 sind (für $k = 0$)

$$\mathbf{b}^\top C^{\ell-1} \mathbf{e} = \sum_{j=1}^m \beta_j \gamma_j^{\ell-1} = \frac{1}{\ell} \quad \text{für } \ell = 1, 2, \dots, p$$

sowie (für $\ell = 1$ mit $k \leftarrow k + 1$)

$$\mathbf{b}^\top A^{k-1} \mathbf{e} = \frac{1}{k!} \quad \text{für } k = 1, 2, \dots, p.$$

Bemerkung. Ein explizites m -stufiges RKV besitzt höchstens die Konsistenzordnung m , denn hier ist $A^m = O$ (A ist echte untere Dreiecksmatrix). Für die optimale Ordnung $p(m)$ eines expliziten m -stufigen RKVs gilt sogar $p(m) \leq m - 1$ falls $m \geq 5$, genauer:

m	1	2	3	4	5	6	7	8	9	10	11	12
$p(m)$	1	2	3	4	4	5	6	6	7	7	8	9

9.4 Lineare Mehrschrittverfahren

Bei **Mehrschrittverfahren** gehen, im Gegensatz zu den Einschrittverfahren, bei der Berechnung der Näherung \mathbf{y}_{n+1} neben \mathbf{y}_n noch weiter zurückliegende Näherungen $\mathbf{y}_{n-1}, \mathbf{y}_{n-2}, \dots$ mit ein. Die wichtigste Klasse der **linearen Mehrschrittverfahren** (LMV) besitzt die allgemeine Form

$$\sum_{j=0}^r \alpha_j \mathbf{y}_{n+j} = h \sum_{j=0}^r \beta_j \mathbf{f}(t_{n+j}, \mathbf{y}_{n+j}), \quad n = 0, 1, 2, \dots \quad (\text{LMV})$$

mit Konstanten $\{\alpha_j\}_{j=0}^r$ sowie $\{\beta_j\}_{j=0}^r$.

Da diese nur bis auf einen (von Null verschiedenen) Skalierungsfaktor eindeutig sind, wird oft die Normierung $\alpha_r = 1$ vorgenommen.

Das Verfahren (**LMV**) ist **explizit** falls $\beta_r = 0$, ansonsten **implizit**.

Wir gehen zunächst auf spezielle Unterfamilien LMV ein.

9.4.1 Adams-Verfahren

Adams-Verfahren sind von der speziellen Gestalt

$$\mathbf{y}_{n+r} - \mathbf{y}_{n+r-1} = h \sum_{j=0}^r \beta_j \mathbf{f}(t_{n+j}, \mathbf{y}_{n+j}), \quad (9.2)$$

d.h. sie sind charakterisiert durch

$$\alpha_r = 1, \quad \alpha_{r-1} = -1, \quad \text{sowie } \alpha_0 = \alpha_1 = \cdots = \alpha_{r-2} = 0.$$

Die Koeffizienten β_j ergeben sich aus der Forderung maximaler Ordnung für gegebenes r . Diese beträgt im expliziten Fall r , im impliziten $r + 1$. Erstere Verfahrensklasse werden als **Adams-Bashforth-Verfahren** bezeichnet, letztere als **Adams-Moulton-Verfahren**.

Alternative Herleitung der Adams-Verfahren:

Aus der Dgl. ergibt sich

$$\mathbf{y}(t_{n+r}) = \mathbf{y}(t_{n+r-1}) + \int_{t_{n+r-1}}^{t_{n+r}} \mathbf{y}'(\tau) d\tau = \mathbf{y}(t_{n+r-1}) + \int_{t_{n+r-1}}^{t_{n+r}} \mathbf{f}(\tau, \mathbf{y}(\tau)) d\tau.$$

Wählt man $\{\beta_j\}_{j=0}^r$ nun so, dass

$$\int_{t_{n+r-1}}^{t_{n+r}} g(\tau) d\tau \approx h \sum_{j=0}^r \beta_j g(t_{n+j})$$

eine interpolatorische Quadraturformel darstellt^a., so ergeben sich ebenso die Koeffizienten des jeweiligen (expliziten bzw. impliziten) Adams-Verfahrens.

^ad.h. man ersetze g durch das Interpolationspolynom durch die Wertepaare $\{(t_n, g(t_n)), \dots, (t_{n+r-1}, g(t_{n+r-1}))\}$ vom Grad $r - 1$ (expliziter Fall) bzw. durch $\{(t_n, g(t_n)), \dots, (t_{n+r}, g(t_{n+r}))\}$ vom Grad r (impliziter Fall)

Für $r = 1$ bis 4 ergeben sich die expliziten **Adams-Bashforth-Verfahren** wie folgt:

$$r = 1 : \quad \mathbf{y}_{n+1} = \mathbf{y}_n + h\mathbf{f}(t_n, \mathbf{y}_n),$$

$$r = 2 : \quad \mathbf{y}_{n+2} = \mathbf{y}_{n+1} + \frac{h}{2} (-\mathbf{f}(t_n, \mathbf{y}_n) + 3\mathbf{f}(t_{n+1}, \mathbf{y}_{n+1})),$$

$$r = 3 : \quad \mathbf{y}_{n+3} = \mathbf{y}_{n+2} + \frac{h}{12} (5\mathbf{f}(t_n, \mathbf{y}_n) - 16\mathbf{f}(t_{n+1}, \mathbf{y}_{n+1}) \\ + 23\mathbf{f}(t_{n+2}, \mathbf{y}_{n+2})),$$

$$r = 4 : \quad \mathbf{y}_{n+4} = \mathbf{y}_{n+3} + \frac{h}{24} (-9\mathbf{f}(t_n, \mathbf{y}_n) + 37\mathbf{f}(t_{n+1}, \mathbf{y}_{n+1}) \\ - 59\mathbf{f}(t_{n+2}, \mathbf{y}_{n+2}) + 55\mathbf{f}(t_{n+3}, \mathbf{y}_{n+3})).$$

Insbesondere erhalten wir für $r = 1$ das **explizite Euler-Verfahren**.

Die impliziten **Adams-Moulton-Verfahren** für $r = 1$ bis 4 lauten:

$$r = 1 : \quad \mathbf{y}_{n+1} = \mathbf{y}_n + \frac{h}{2} (\mathbf{f}(t_n, \mathbf{y}_n) + \mathbf{f}(t_{n+1}, \mathbf{y}_{n+1})),$$

$$r = 2 : \quad \mathbf{y}_{n+2} = \mathbf{y}_{n+1} + \frac{h}{12} (-\mathbf{f}(t_n, \mathbf{y}_n) + 8\mathbf{f}(t_{n+1}, \mathbf{y}_{n+1}) \\ + 5\mathbf{f}(t_{n+2}, \mathbf{y}_{n+2})),$$

$$r = 3 : \quad \mathbf{y}_{n+3} = \mathbf{y}_{n+2} + \frac{h}{24} (\mathbf{f}(t_n, \mathbf{y}_n) - 5\mathbf{f}(t_{n+1}, \mathbf{y}_{n+1}) \\ + 19\mathbf{f}(t_{n+2}, \mathbf{y}_{n+2}) + 9\mathbf{f}(t_{n+3}, \mathbf{y}_{n+3})),$$

$$r = 4 : \quad \mathbf{y}_{n+4} = \mathbf{y}_{n+3} + \frac{h}{720} (-19\mathbf{f}(t_n, \mathbf{y}_n) + 106\mathbf{f}(t_{n+1}, \mathbf{y}_{n+1}) \\ - 264\mathbf{f}(t_{n+2}, \mathbf{y}_{n+2}) + 646\mathbf{f}(t_{n+3}, \mathbf{y}_{n+3}) \\ + 251\mathbf{f}(t_{n+4}, \mathbf{y}_{n+4})).$$

Für $r = 1$ erhalten wir die **Trapezregel**.

9.4.2 Nyström-Verfahren

Nyström-Verfahren besitzen die Form

$$\mathbf{y}_{n+r} - \mathbf{y}_{n+r-2} = h \sum_{j=0}^r \beta_j \mathbf{f}(t_{n+j}, \mathbf{y}_{n+j}). \quad (9.3)$$

Als explizites Nyström-Verfahren mit $r = 2$ erhalten wir die bereits bekannte Mittelpunktsregel (9.1).

Als implizites Nyström-Verfahren mit $r = 2$ ergibt sich die **Simpson-Regel**

$$\mathbf{y}_{n+2} - \mathbf{y}_n = \frac{h}{3} \left(\mathbf{f}(t_n, \mathbf{y}_n) + 4\mathbf{f}(t_{n+1}, \mathbf{y}_{n+1}) + \mathbf{f}(t_{n+2}, \mathbf{y}_{n+2}) \right).$$

9.4.3 BDF-Verfahren

Im Gegensatz zu den Runge-Kutta- und Adams-Verfahren, die mit Hilfe von Quadratur hergeleitet wurden, beruht die Idee der **BDF-Verfahren** (backward differentiation formulas) auf **numerischer Differentiation**.

Zur Konstruktion eines linearen r -Schritt-Verfahrens seien die Näherungen $y_n, y_{n-1}, \dots, y_{n-r+1}$ bekannt und die nächste Näherung y_{n+1} zu berechnen. Hierzu wird folgende Konstruktionsidee verwandt:

- (1) Lege durch die $r + 1$ Wertepaare $\{(t_j, y_j)\}_{j=n-r+1}^{n+1}$ das eindeutig bestimmte Interpolationspolynom $p \in \mathcal{P}_r$.
- (2) Die fehlende Interpolationsvorschrift $p(t_{n+1}) = y_{n+1}$ (y_{n+1} ist noch zu berechnen) wird ersetzt durch die Vorschrift

$$p'(t_{n+1}) = f(t_{n+1}, y_{n+1}), \quad (9.4)$$

d.h. dass p an der Stelle t_{n+1} die Differentialgleichung erfüllen soll.

Die Newton-Darstellung des gesuchten Interpolationspolynoms ist gegeben durch

$$p(t) = f_{n+1} + f_{n+1,n}(t - t_{n+1}) + f_{n+1,n,n-1}(t - t_{n+1})(t - t_n) \\ + \cdots + f_{n+1,n,\dots,n-r+2,n-r+1}(t - t_{n+1}) \cdots (t - t_{n-r+2})$$

Unter der vereinfachenden Annahme konstanter Schrittweite h erhält man aus der Newton-Darstellung von p

$$p'(t_{n+1}) = \sum_{j=1}^r f_{n+1,n,\dots,n+1-j} h^{j-1} (j-1)! . \quad (9.5)$$

Führt man die Notation der **Rückwärtsdifferenzen** ein:

$$\nabla^0 y_{n+1} := y_{n+1}$$

$$\nabla y_{n+1} := y_{n+1} - y_n,$$

$$\nabla^j y_{n+1} := \nabla(\nabla^{j-1} y_{n+1}) = \nabla^{j-1} y_{n+1} - \nabla^{j-1} y_n, \quad j = 2, 3, \dots$$

so läßt sich $p'(t_{n+1})$ ausdrücken als

$$p'(t_{n+1}) = \frac{1}{h} \sum_{j=1}^r \frac{\nabla^j y_{n+1}}{j},$$

und zusammen mit der Forderung (9.4) die Verfahrensvorschrift

$$\sum_{j=1}^r \frac{1}{j} \nabla^j y_{n+1} = h f(t_{n+1}, y_{n+1}) \quad (9.6)$$

des BDF-Verfahrens der Ordnung r , kurz **BDF $_r$** genannt.

Durch die Indexverschiebung $n + 1 \mapsto n + r$ erhalten wir die BDF-Verfahren in unserer Standardnotation für lineare Mehrschrittverfahren.

Die **BDF-Verfahren** für $r = 1$ bis 6 lauten:

$$r = 1 : \quad \mathbf{y}_{n+1} - \mathbf{y}_n = h\mathbf{f}(t_{n+1}, \mathbf{y}_{n+1}),$$

$$r = 2 : \quad \frac{3}{2}\mathbf{y}_{n+2} - 2\mathbf{y}_{n+1} + \frac{1}{2}\mathbf{y}_n = h\mathbf{f}(t_{n+2}, \mathbf{y}_{n+2})$$

$$r = 3 : \quad \frac{11}{6}\mathbf{y}_{n+3} - 3\mathbf{y}_{n+2} + \frac{3}{2}\mathbf{y}_{n+1} - \frac{1}{3}\mathbf{y}_n = h\mathbf{f}(t_{n+3}, \mathbf{y}_{n+3}),$$

$$r = 4 : \quad \frac{25}{12}\mathbf{y}_{n+4} - 4\mathbf{y}_{n+3} + 3\mathbf{y}_{n+2} - \frac{4}{3}\mathbf{y}_{n+1} + \frac{1}{4}\mathbf{y}_n = h\mathbf{f}(t_{n+4}, \mathbf{y}_{n+4}),$$

$$r = 5 : \quad \frac{137}{60}\mathbf{y}_{n+5} - 5\mathbf{y}_{n+4} + 5\mathbf{y}_{n+3} - \frac{10}{3}\mathbf{y}_{n+2} + \frac{5}{4}\mathbf{y}_{n+1} - \frac{1}{5}\mathbf{y}_n \\ = h\mathbf{f}(t_{n+5}, \mathbf{y}_{n+5}),$$

$$r = 6 : \quad \frac{147}{60}\mathbf{y}_{n+6} - 6\mathbf{y}_{n+5} - \frac{15}{2}\mathbf{y}_{n+4} - \frac{20}{3}\mathbf{y}_{n+3} + \frac{15}{4}\mathbf{y}_{n+2} - \frac{6}{5}\mathbf{y}_{n+1} + \frac{1}{6}\mathbf{y}_n \\ = h\mathbf{f}(t_{n+6}, \mathbf{y}_{n+6}).$$

Für $r = 1$ erhalten wir das **implizite Euler-Verfahren**.

9.4.4 Der lokale Diskretisierungsfehler bei LMV

Setzt man in (LMV) die exakte Lösung ein und beachtet, dass diese $\mathbf{y}'(t_n) = \mathbf{f}(t_n, \mathbf{y}(t_n))$ erfüllt, so erhalten wir (nach Division durch h)

$$\mathbf{T}_{n+r}(h) = \frac{1}{h} \left(\sum_{j=0}^r \alpha_j \mathbf{y}(t_{n+j}) - h \sum_{j=0}^r \beta_j \mathbf{y}'(t_{n+j}) \right).$$

Entwickelt man alle Auswertungen von \mathbf{y} bzw. \mathbf{y}' an der Stelle $t = t_n$

$$\mathbf{y}(t_{n+j}) = \mathbf{y}(t_n) + jh\mathbf{y}'(t_n) + \frac{(jh)^2}{2}\mathbf{y}''(t_n) + \frac{(jh)^3}{6}\mathbf{y}'''(t_n) + \dots$$

$$\mathbf{y}'(t_{n+j}) = \mathbf{y}'(t_n) + jh\mathbf{y}''(t_n) + \frac{(jh)^2}{2}\mathbf{y}'''(t_n) + \frac{(jh)^3}{6}\mathbf{y}''''(t_n) + \dots,$$

und sortiert Terme mit gleicher h -Potenz, so erhält man die Darstellung ...

$$\begin{aligned}
\mathbf{T}_{n+r}(h) = & \frac{1}{h} \left[\sum_{j=0}^r \alpha_j \right] \mathbf{y}(t_n) + \left[\sum_{j=0}^r (j\alpha_j - \beta_j) \right] \mathbf{y}'(t_n) \\
& + h \left[\sum_{j=0}^r \left(\frac{j^2}{2} \alpha_j - j\beta_j \right) \right] \mathbf{y}''(t_n) \\
& + \dots + h^q \left[\sum_{j=0}^r \left(\frac{j^{q+1}}{(q+1)!} \alpha_j - \frac{j^q}{q!} \beta_j \right) \right] \mathbf{y}^{(q+1)}(t_n) + O(h^{q+1}).
\end{aligned}$$

Konsistenz des LMV, also $\mathbf{T}_n(h) \rightarrow 0$ für $h \rightarrow 0$, liegt somit vor, sofern

$$\sum_{j=0}^r \alpha_j = 0 \quad \text{sowie} \quad \sum_{j=0}^r j\alpha_j = \sum_{j=0}^r \beta_j. \quad (9.7)$$

Konsistenzordnung $p \in \mathbb{N}$ liegt vor, wenn die ersten $p + 1$ Terme in eckigen Klammern verschwinden.

9.4.5 Charakteristische Polynome

Bei der Analyse LMV spielen die mit den Koeffizienten $\{\alpha_j\}_{j=0}^r$ und $\{\beta_j\}_{j=0}^r$ gebildeten **charakteristischen Polynome**

$$\rho(\zeta) := \sum_{j=0}^r \alpha_j \zeta^j \quad \text{und} \quad \sigma(\zeta) := \sum_{j=0}^r \beta_j \zeta^j \quad (9.8)$$

eine zentrale Rolle.

Bei einem r -Schrittverfahren gilt $\rho \in \mathcal{P}_r$ und, bei impliziten r -Schrittverfahren, $\sigma \in \mathcal{P}_r$. Bei expliziten LMV ist der Grad von σ kleiner als r .

Mit Hilfe der charakteristischen Polynome lassen sich die beiden Konsistenzbedingungen (9.7) formulieren als

$$\rho(1) = 0 \quad \text{und} \quad \rho'(1) = \sigma(1). \quad (9.9)$$

Beispiel: Beim Adams-Moulton-Verfahren für $r = 2$

$$\mathbf{y}_{n+2} = \mathbf{y}_{n+1} + \frac{h}{12} \left(-\mathbf{f}(t_n, \mathbf{y}_n) + 8\mathbf{f}(t_{n+1}, \mathbf{y}_{n+1}) + 5\mathbf{f}(t_{n+2}, \mathbf{y}_{n+2}) \right),$$

gilt

$$\alpha_0 = 0, \alpha_1 = -1, \alpha_2 = 1, \quad \beta_0 = \frac{-1}{12}, \beta_1 = \frac{8}{12}, \beta_2 = \frac{5}{12}$$

und man erhält die charakteristischen Polynome

$$\rho(\zeta) = \zeta^2 - \zeta, \quad \sigma(\zeta) = \frac{1}{12} (5\zeta^2 + 8\zeta - 1).$$

Wegen

$$\alpha_0 + \alpha_1 + \alpha_2 = 0,$$

$$0 \cdot \alpha_0 + 1 \cdot \alpha_1 + 2 \cdot \alpha_2 - (\beta_0 + \beta_1 + \beta_2) = 0,$$

$$0 \cdot \alpha_0 + \frac{1}{2} \cdot \alpha_1 + \frac{4}{2} \alpha_2 - (0 \cdot \beta_0 + 1 \cdot \beta_1 + 2 \cdot \beta_2) = 0,$$

$$\frac{1}{6}(0 \cdot \alpha_0 + 1 \cdot \alpha_1 + 8\alpha_2) - \frac{1}{2}(0 \cdot \beta_0 + 1 \cdot \beta_1 + 4 \cdot \beta_2) = 0,$$

$$\frac{1}{24}(0 \cdot \alpha_0 + 1 \cdot \alpha_1 + 16\alpha_2) - \frac{1}{6}(0 \cdot \beta_0 + 1 \cdot \beta_1 + 8 \cdot \beta_2) = \frac{5}{8} - \frac{2}{3} \neq 0$$

besitzt dieses Verfahren die Konsistenzordnung $p = 3$.

9.4.6 Anlaufrechnung

Bei jedem Start (oder Neustart) eines r -Schrittverfahrens ($r > 1$) müssen zunächst neben den Anfangswerten y_0 die Approximationen y_1, y_2, \dots, y_{r-1} berechnet werden.

Um die Ordnung p dieses Verfahrens aufrechtzuerhalten genügt es, die Anfangswerte mit einem Verfahren der Ordnung $p - 1$ zu berechnen.

Grund: da dies nur $r - 1$ Mal erfolgt – unabhängig von n – verliert man im Grenzwert $h \rightarrow 0$ keine h -Potenz durch die Akkumulation der Schrittfehler bei der Anlaufrechnung; somit ist die Konvergenzordnung des ab $n = r$ benutzten Verfahrens ausschlaggebend.

Beispiel: Zur Berechnung von y_1 bei der Anlaufrechnung für die Mittelpunktsregel (9.1), welche konsistent ist von der Ordnung $p = 2$, könnte man ohne Ordnungsverlust das explizite Euler-Verfahren heranziehen.

9.4.7 Prädiktor-Korrektor-Verfahren

Eine beliebte Technik, das aufwändige Gleichungslösen in jedem Schritt eines impliziten LMV zu umgehen besteht darin, zunächst mittels eines **expliziten** Verfahrens gleicher Schrittlänge eine Näherung \hat{y}_{n+r} zu berechnen (**Prädiktor**) und diesen Wert dann im impliziten Verfahren (**Korrektor**) anstelle von y_{n+r} einzusetzen.

Beispiel: Kombination des Adams-Bashforth-Verfahrens mit $r = 1$ (explizites Euler-Verfahren) mit dem Adams-Moulton-Verfahren der gleichen Schrittzahl (Trapezregel):

$$\hat{y}_{n+1} = y_n + h f(t_n, y_n), \quad (\text{Prädiktorschritt}),$$

$$y_{n+1} = y_n + \frac{h}{2} (f(t_n, y_n) + f(t_{n+1}, \hat{y}_{n+1})), \quad (\text{Korrektorschritt}).$$

Man kann zeigen, dass dieses Verfahren die Konsistenzordnung $p = 2$ besitzt.

9.4.8 Vergleich Ein- und Mehrschrittverfahren

Vorteile von Einschrittverfahren:

- Einschrittverfahren sind **selbststartend**.
- Eine Änderung der Schrittweite h ohne Mehraufwand möglich.
- Integration über Unstetigkeitsstellen der Lösung bei Einschrittverfahren möglich ohne Ordnungsverlust, sofern diese nur Gitterpunkte sind.

Vorteil von Mehrschrittverfahren: wenige Auswertungen der rechten Seite bei gleicher Konsistenzordnung.

9.5 Konvergenz von Einschrittverfahren

Satz 9.3 (Beziehung lokaler-globaler Diskretisierungsfehler) *Unter den gegebenen Voraussetzungen (vgl. Satz 8.1 sowie (V_2)) gilt*

$$\|e_n\| \leq \left((\|e_0\| + (t_n - t_0) \max_{1 \leq j \leq n} \|T_j\|) \right) \exp(M(t_n - t_0)).$$

Insbesondere ist ein ESV genau dann konvergent, wenn es konsistent ist.

Satz 9.4 (Gesamtfehler bei expliziten ESV) Das **explizite** **ESV** (**ESV**) werde auf einem Rechner durch

$$\tilde{\mathbf{y}}_{n+1} = \tilde{\mathbf{y}}_n + h\Phi_f(\tilde{\mathbf{y}}_n, t_n; h) + \boldsymbol{\varepsilon}_{n+1}, \quad \tilde{\mathbf{y}}_0 = \mathbf{y}_0 + \boldsymbol{\varepsilon}_0,$$

realisiert. Ist $\|\boldsymbol{\varepsilon}_n\| \leq \varepsilon$ und $\|\mathbf{T}_n\| \leq T$ für alle $n = 0, 1, \dots$, so folgt

$$\|\mathbf{y}(t_n) - \tilde{\mathbf{y}}_n\| \leq (\|\boldsymbol{\varepsilon}_0\| + (t_n - t_0)(T + \frac{\varepsilon}{h})) \exp(M(t_n - t_0)).$$

Dilemma: Bei „großer“ Schrittweite h ist (üblicherweise) T groß; wählt man h sehr „klein“, so ist zwar T klein, aber $\frac{\varepsilon}{h}$ groß, dh. der Anteil der Rundungsfehler dominiert den Gesamtfehler.

Wichtig sind also Verfahren hoher Ordnung, weil dort T schon bei „moderater“ Größe von h klein wird.

9.6 Nullstabilität linearer Mehrschrittverfahren

Konvergenz bei Einschrittverfahren: Beitrag des n -ten Schrittfehlers S_n beschränkt durch $e^{M(t_{\text{end}} - t_0)} \|S_n\|$, was für $h \rightarrow 0$ noch gegen Null geht.

Diese Eigenschaft nennt man auch **Nullstabilität**, und sie ist für ESV immer gegeben. Bei LMV ist dies nicht der Fall.

Beispiel: Beim LMV $y_{n+2} - 3y_{n+1} + 2y_n = -hf(t_n, y_n)$ beträgt der lokale Diskretisierungsfehler

$$T_{n+2}(h) = \frac{h}{2} y''(t_n) + O(h^2).$$

Wir wenden es an auf die AWA

$$y'(t) = 0, \quad y(0) = 0 \quad \text{mit Lösung } y(t) \equiv 0. \quad (9.10)$$

Mit den Startwerten $y_0 = y_1 = 0$ liefert es tatsächlich die exakte Lösung.

Setzt man jedoch $y_1 = h$ um einen kleinen Fehler der Ordnung h bei der Approximation von $y(t_1)$ zu simulieren, so erhält man für die Approximation der Lösung an der Stelle $t = 1$ folgende Werte ($h = 1/N$):

N	y_N
5	6.2
10	102.3
20	5.2×10^4

Wegen $y_1 \rightarrow 0$ für $h \rightarrow 0$ sind diese Anfangsdaten zulässig. Wie man sieht konvergiert y_N aber nicht gegen $y(t_N) = y(1) = 0$.

Die Approximationen erfüllen die Differenzengleichung

$$y_{n+2} - 3y_{n+1} + 2y_n = 0, \quad n = 0, 1, 2, \dots \quad (9.11)$$

mit der Lösung

$$y_n = (2^n - 1)y_1 - (2^n - 2)y_0 = 2y_0 - y_1 + 2^n(y_1 - y_0), \quad n \geq 2.$$

9.6.1 Lösung linearer Differenzengleichungen

Zur Lösung der homogenen linearen Differenzengleichung der Ordnung r

$$\alpha_r y_{n+r} + \alpha_{r-1} y_{n+r-1} + \cdots + \alpha_0 y_n = \sum_{j=0}^r \alpha_j y_{n+j} = 0 \quad (9.12)$$

macht man den Ansatz $y_n = \zeta^n$. Einsetzen in (9.12) führt nach Division durch ζ^n auf die Bedingung

$$\sum_{j=0}^r \alpha_j \zeta^j = 0,$$

d.h. $y_n = \zeta^n$ löst (9.12) falls ζ Nullstelle des ersten charakteristischen Polynoms $\rho(\zeta)$ ist (vgl. (9.8)).

Ist nun

$$\rho(\zeta) = \alpha_r (\zeta - \zeta_1)(\zeta - \zeta_2) \cdots (\zeta - \zeta_r)$$

und sind die Nullstellen paarweise verschieden, so stellen $\{\zeta_i^n\}_{i=1}^r$ insgesamt r linear unabhängige Lösungen von (9.12) dar. Aufgrund der Linearität und Homogenität von (9.12) ist jede Linearkombination

$$y_n = c_1 \zeta_1^n + c_2 \zeta_2^n + \cdots + c_r \zeta_r^n, \quad c_1, \dots, c_r \text{ beliebig} \quad (9.13)$$

ebenfalls Lösung. Zu gegebenen Anfangsbedingungen y_0, \dots, y_{r-1} werden die Koeffizienten c_j bestimmt als Lösung des LGS

$$\begin{aligned} c_1 + c_2 + \cdots + c_r &= y_0, \\ c_1 \zeta_1 + c_2 \zeta_2 + \cdots + c_r \zeta_r &= y_1, \\ &\vdots \\ c_1 \zeta_1^{r-1} + c_2 \zeta_2^{r-1} + \cdots + c_r \zeta_r^{r-1} &= y_{r-1}. \end{aligned}$$

Beispiel: Das charakteristische Polynom der Differenzengleichung (9.11) ist

$$\rho(\zeta) = \zeta^2 - 3\zeta + 2 = (\zeta - 1)(\zeta - 2).$$

Fazit: Besitzt $\rho(\zeta)$ Nullstellen vom Betrag > 1 , so kann das zugehörige LMV nicht konvergent sein.

Fallen zwei Nullstellen zusammen, d.h. $\zeta_i = \zeta_j$, so sind ζ_i^n und ζ_j^n linear abhängig. Eine weitere linear unabhängige Lösung ist $n\zeta_i^{n-1}$.

Bei einer dreifachen Nullstelle ζ_i ist $n^2\zeta_i^{n-2}$ eine weitere etc.

Beispiel: Das (konsistente) LMV

$$\mathbf{y}_{n+2} - 2\mathbf{y}_{n+1} + \mathbf{y}_n = \frac{h}{2} (\mathbf{f}(t_{n+2}, \mathbf{y}_{n+2}) - \mathbf{f}(t_n, \mathbf{y}_n)).$$

angewandt auf $y'(t) = 0$.

Fazit: Besitzt $\rho(\zeta)$ mehrfache Nullstellen von Betrag 1, so kann ebenfalls keine Konvergenz vorliegen.

Beispiel: Das konsistente LMV

$$\mathbf{y}_{n+3} - 2\mathbf{y}_{n+2} + \frac{5}{4}\mathbf{y}_{n+1} - \frac{1}{4}\mathbf{y}_n = \frac{h}{4}f(t_n, \mathbf{y}_n)$$

angewandt auf (9.10).

Fazit: Bei mehrfachen Nullstellen mit Betrag < 1 liegt Konvergenz vor (lineares Wachstum wird übertroffen durch exponentielles Abklingen).

Ein LMV heißt **nullstabil**, falls für die Nullstellen dessen ersten charakteristischen Polynoms $\rho(\zeta)$ gilt

- (a) $|\zeta_i| \leq 1$ für alle $i = 1, 2, \dots, r$
- (b) Ist ζ_i mehrfache Nullstelle, so gilt $|\zeta_i| < 1$.

Beispiele:

- Die Adams-Verfahren sind nullstabil.
- Die BDF-Verfahren sind nullstabil für $1 \leq r \leq 6$.

Unsere Beispiele zeigen, dass Nullstabilität notwendig für die Konvergenz ist. Tatsächlich ist sie auch hinreichend, damit ein konsistentes LMV konvergiert.

Satz 9.5 (Dahlquist, 1956) *Für LMV angewandt auf (AWA) gilt*

$$\text{Konsistenz} + \text{Nullstabilität} \Leftrightarrow \text{Konvergenz} .$$

Bemerkungen 9.6

- (a) *Ein konsistentes LMV besitzt stets die Nullstelle $\zeta = 1$ (vgl. (9.9)).*
- (b) *Wir werden sehen, dass außer Nullstabilität auch andere Stabilitätsbegriffe bei AWAen von Bedeutung sind.*
- (c) *Einschrittverfahren sind stets nullstabil.*

9.7 Schrittweitensteuerung

Kein Verfahren zur Lösung von AWAen arbeitet in der Praxis mit einer **konstanten** Schrittweite. Man wird vielmehr versuchen, die Schrittweite an das Verhalten der Lösung y anzupassen (ändert sich y in einem Bereich schnell, so ist dort eine kleine Schrittweite angebracht; in Bereichen, in denen y kaum variiert, ist eine größere Schrittweite ausreichend). Wir werden hier eine **Schrittweitensteuerung** vorstellen, die zum Ziel hat, den lokalen Diskretisierungsfehler T_n zu kontrollieren:

$$\|T_n\| \sim \text{tol}, \quad n = 1, 2, \dots,$$

mit einer vorgegebenen Toleranz tol . Bei Systemen von DGen (insbesondere dann, wenn die Lösungskomponenten von unterschiedlicher Größenordnung sind) wird man für jede Komponente eine eigene absolute Fehlertoleranz und global eine relative Fehlertoleranz festsetzen.

Satz 9.3 besagt, dass mit dem lokalen auch der (eigentlich interessierende) globale Diskretisierungsfehler kontrolliert wird.

Um den lokalen Diskretisierungsfehler zu schätzen verwendet man **zwei Methoden unterschiedlicher Konsistenzordnungen** (sagen wir p und q mit $p < q$), um \mathbf{y}_n aus \mathbf{y}_{n-1} zu berechnen:

$$\mathbf{y}_n = \mathbf{y}_{n-1} + h\Phi_f(\mathbf{y}_{n-1}, t_{n-1}; h) \quad \text{bzw.} \quad \hat{\mathbf{y}}_n = \mathbf{y}_{n-1} + h\hat{\Phi}_f(\mathbf{y}_{n-1}, t_{n-1}; h)$$

Für die zugehörigen lokalen Diskretisierungsfehler gelten:

$$\mathbf{T}_n = \frac{\mathbf{y}(t_n) - \mathbf{y}(t_{n-1})}{h} - \Phi_f(\mathbf{y}(t_{n-1}), t_{n-1}; h) = O(h^p),$$

$$\hat{\mathbf{T}}_n = \frac{\mathbf{y}(t_n) - \mathbf{y}(t_{n-1})}{h} - \hat{\Phi}_f(\mathbf{y}(t_{n-1}), t_{n-1}; h) = O(h^q).$$

Daraus folgt

$$\mathbf{T}_n - \hat{\mathbf{T}}_n = \hat{\Phi}_f(\mathbf{y}_{n-1}, t_{n-1}; h) - \Phi_f(\mathbf{y}_{n-1}, t_{n-1}; h) = \frac{1}{h}(\hat{\mathbf{y}}_n - \mathbf{y}_n).$$

Wegen $\mathbf{T}_n - \hat{\mathbf{T}}_n = \mathbf{T}_n(1 + O(h^{q-p})) \sim \mathbf{T}_n$ erhalten wir aus

$$\frac{1}{h} \|\mathbf{y}_n - \hat{\mathbf{y}}_n\| \sim \|\mathbf{T}_n\|$$

eine (grobe) Schätzung für $\|\mathbf{T}_n\|$.

Ist $\frac{1}{h} \|\mathbf{y}_n - \hat{\mathbf{y}}_n\| > \text{tol}$, so wird die Schrittweite h verworfen und mit

$$\left(\frac{\tilde{h}}{h}\right)^p = \alpha \frac{h \text{ tol}}{\|\mathbf{y}_n - \hat{\mathbf{y}}_n\|} \quad (*)$$

eine neue Schrittweite \tilde{h} bestimmt.

Die Wahl von \tilde{h} motiviert sich folgendermaßen:

benutzte Schrittweite h : $\frac{1}{h} \|\mathbf{y}_n - \hat{\mathbf{y}}_n\| \sim \|\mathbf{T}_n\| = ch^p + O(h^{p+1}) \sim ch^p,$

erwünschte Schrittweite \tilde{h} : $\text{tol} = \|\mathbf{T}_n\| = c\tilde{h}^p + O(\tilde{h}^{p+1}) \sim c\tilde{h}^p.$

(α ist hier ein Sicherheitsfaktor, etwa $\alpha = 0.9$).

Ausgehend von y_{n-1} werden jetzt neue Näherungen y_n und \hat{y}_n (an der Stelle $t_{n-1} + \tilde{h}$) berechnet. Diesen Prozess wiederholt man solange, bis $\frac{1}{h} \|y_n - \hat{y}_n\| \leq \text{tol}$ erfüllt ist. Dann wird $(*)$ verwendet, um eine neue (größere) Schrittweite für den nächsten Schritt ($n \rightarrow n + 1$) vorzuschlagen.

Um den Aufwand in Grenzen zu halten, verwendet man zur Berechnung von y_n und \hat{y}_n zwei RKV (verschiedener Ordnungen), deren Butcher-Tableaus sich nur in b unterscheiden (d.h. A und c sind gleich, so dass die k_j nur einmal berechnet werden müssen). Man spricht von **eingebetteten RKV** und schreibt

$$\begin{array}{c|c} c & A \\ \hline & b^\top \\ \hline & \hat{b}^\top \end{array} \quad \text{z.B.} \quad \begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 0 & 0 \\ \hline & 1 & 0 \\ \hline & 1/2 & 1/2 \end{array}.$$

Hier wird das Euler-Verfahren ($p = 1$) in das verbesserte Euler-Verfahren ($p = 2$) eingebettet.

Ein populäres Beispiel ist die **Fehlberg 4(5)-Formel**:

0	0	0	0	0	0	0
$\frac{1}{4}$	$\frac{1}{4}$	0	0	0	0	0
$\frac{3}{8}$	$\frac{3}{32}$	$\frac{9}{32}$	0	0	0	0
$\frac{12}{13}$	$\frac{1932}{2197}$	$-\frac{7200}{2197}$	$\frac{7296}{2197}$	0	0	0
1	$\frac{439}{216}$	-8	$\frac{3680}{513}$	$-\frac{845}{4104}$	0	0
$\frac{1}{2}$	$-\frac{8}{27}$	2	$-\frac{3544}{2565}$	$\frac{1859}{4104}$	$-\frac{11}{40}$	0
	$\frac{25}{216}$	0	$\frac{1408}{2565}$	$\frac{2197}{4104}$	$-\frac{1}{5}$	0
	$\frac{16}{135}$	0	$\frac{6656}{12825}$	$\frac{28561}{56430}$	$-\frac{9}{50}$	$\frac{2}{55}$

Hier werden zwei sechststufige RKV der Ordnungen 4 bzw. 5 kombiniert.

Ein weiteres Beispiel ist die **Dormand-Prince 4(5)-Formel**:

0	0	0	0	0	0	0	0
$\frac{1}{5}$	$\frac{1}{5}$	0	0	0	0	0	0
$\frac{3}{10}$	$\frac{3}{40}$	$\frac{9}{40}$	0	0	0	0	0
$\frac{4}{5}$	$\frac{44}{45}$	$-\frac{56}{15}$	$\frac{32}{9}$	0	0	0	0
$\frac{8}{9}$	$\frac{19372}{6561}$	$-\frac{25360}{2187}$	$\frac{64448}{6561}$	$-\frac{212}{729}$	0	0	0
1	$\frac{9017}{3168}$	$-\frac{355}{33}$	$\frac{46732}{5247}$	$\frac{49}{176}$	$-\frac{5103}{18656}$	0	0
1	$\frac{35}{384}$	0	$\frac{500}{1113}$	$\frac{125}{192}$	$-\frac{2187}{6784}$	$\frac{11}{84}$	0
	$\frac{35}{384}$	0	$\frac{500}{1113}$	$\frac{125}{192}$	$-\frac{2187}{6784}$	$\frac{11}{84}$	0
	$\frac{5179}{57600}$	0	$\frac{7571}{16695}$	$\frac{393}{640}$	$-\frac{92097}{339200}$	$\frac{187}{2100}$	$\frac{1}{40}$

Hier wird ein sechsstufiges RKV der Ordnung 4 in ein siebenstufiges der Ordnung 5 eingebettet.