Numerische Methoden zur Approximation von Lösungen gewöhnlicher Differentialgleichungen

Oliver Ebner, Matthias Puhr,
Andreas Reinhart, Matthias Weissenbacher

15. Februar 2006

Projektarbeit im Rahmen des Proseminares aus Differentialgleichungen, WS 05/06

Inhaltsverzeichnis

1 Explizite Einschrittverfahren zur zur numerischen Lösung von allgemeinen AWP

1.1 Einführung

Definition 1 (Anfangswertproblem) *Seien* $N \in \mathbb{N}, a, b \in \mathbb{R} : a < b, y_0 \in \mathbb{R}^N$ *und* $f : [a,b] \times \mathbb{R}^N \to \mathbb{R}^N$*,dann heißt*

$$y'(t) = f(t, y(t)),\ t \in [a,b]\,,\ y(a) = y_0 \tag{1}$$

ein Anfangswertproblem für ein System von N Differentialgleichungen 1.Ordnung (kurz: AWP). Eine auf dem Intervall $[a,b]$ *differenzierbare Abbildung y in den* R^N *mit* (1) *heißt „Lösung" des AWP Die Ableitung* $y'(t)$ *sei durch komponentenweise Differentiation erklärt.*

Definition 2 (Lipschitzstetigkeit) *Seien a,b,N,f wie in Def.1 gegeben,* $\|\cdot\| : \mathbb{R}^N \to \mathbb{R}_{\geq 0}$ *sei eine Norm auf dem* $\mathbb{R}^N$*, dann heißt f lipschitzstetig im 2. Teilargument* $\Leftrightarrow$

$$\exists L_f \geq 0 : \forall\,(t,u)\,,(t,v) \in [a,b] \times \mathbb{R}^N : ||f\,(t,u) - f\,(t,v)\,|| \leq L_f ||u - v|| \tag{2}$$

Lemma 1.1 *Seien f und y wie in Def.1 gegeben und es sei* $p \in \mathbb{N}$*. Ist f p-mal stetig partiell differenzierbar,dann ist y mindestens (p+1)-mal stetig differenzierbar.*

Beweis(Für den Fall N=1): Es reicht zu zeigen : $\forall p \in \mathbb{N}$: f p-mal stetig partiell differenzierbar $\Rightarrow \forall t \in [a,b]$: $y^{p+1}(t)$ ist eine Summe von Produkten höchstens p-facher partieller Ableitungen von f an der Stelle (t,y(t)). Für $p = 0$ ist $y'(t) = f(t, y(t))$, also ist y'(t) eine 0-fache partielle Ableitung von f an der Stelle (t,y(t)),somit insbesondere von der behaupteten Form. Sei nun $p \in \mathbb{N}$ und gelte die Behauptung für p,dann reicht es nach der Summen- und Produktregel zu zeigen dass die Ableitung eines jeden Faktors eines jeden Produktes,Summe von Produkten höchstens (p+1)-facher partieller Ableitungen von f an der Stelle $(t, y(t))$ ist. Es sei $g : [a,b] \times \mathbb{R}^N \to \mathbb{R}^N$ mit g eine höchtens p-fache partielle Ableitung von f,dann gilt nach der Kettenregel $\forall t \in [a,b] : (g(t, y(t)))' = g_t(t, y(t)) + g_y(t, (y(t))y'(t) = g_t(t, y(t)) + g_y(t, (y(t))f(t, y(t)), g_t, g_y$ und f sind höchstens (p+1)-fache partielle Ableitungen und daher folgt die Behauptung.

Satz 1.1 (von Picard/Lindelöf)
Seien a,N,f wie in Def.1 gegeben,weiters sei f stetig und erfülle (2)*. Dann existiert ein* $b > a$ *und genau ein* $y : [a,b] \to \mathbb{R}^N$ *das stetig differenzierbar ist und* (1) *erfüllt.*

Beweis : Siehe Literatur z.B. [3].

Bemerkung 1
Im Folgenden sei f wie in Def.1 gegeben und erfülle zusätzlich Bedingungen(z.B.von Satz 1) die die Existenz und Eindeutigkeit der Lösung y in (1) garantieren, weiters sei mit y diese Lösung im folgenden bezeichnet. Zusätzlich bezeichne für eine beliebige Menge M und $n \in \mathbb{N}$ $M^{[0,n]} := \mathrm{Abb}([0,n] \cap \mathbb{N}, M)$.

1.1.1 Definition eines Einschrittverfahrens und damit verbundener Begriffe

Definition 3 (Verfahren und Verfahrensfunktion) *Seien a,b,N wie in Def.1 gegeben, $\|\cdot\| : \mathbb{R}^N \to \mathbb{R}_{\geq 0}$ sei eine Norm auf dem $\mathbb{R}^N$, $D_{a,b,N} := \{(t,u,h) \in [a,b] \times \mathbb{R}^N \times \mathbb{R}_{>0} \mid h \leq b - t\}$*

- *Eine Verfahrensfunktion φ ist eine Abbildung $\varphi : D_{a,b,N} \to \mathbb{R}^N$*

- *Sei φ eine Verfahrensfunktion,dann heißt φ lipschitzstetig im 2. Teilargument genau dann, wenn*

$$\exists_{L_\varphi \geq 0} \forall_{(t,u,h) \in D_{a,b,N}} : \|\varphi(t,u,h) - \varphi(t,v,h)\| \leq L_\varphi \|u - v\| \tag{3}$$

- *Sei $n \in \mathbb{N}$,$\Gamma_{a,b,n} := \{(t_l)_{l=0}^n \in [a,b]^{[0,n]} \mid t_0 = a, t_n = b, \forall_{0 \leq l \leq n-1} t_l \leq t_{l+1}\}$.*
  Dann heißt $\Gamma_{a,b,n}$ die Menge aller Gitter des Intervalls $[a,b]$ der Ordnung n.
  Sei $\Delta \in \Gamma_{a,b,n}$ und $0 \leq l \leq n-1$,dann heißt $h_{\Delta,l} := t_{l+1} - t_l$ die Schrittweite des Gitters Δ an der Stelle l und $h_{\Delta,max} := \max\{h_{\Delta,l} \mid 0 \leq l \leq n-1\}$ heißt die größte Schrittweite des Gitters

- *Sei φ eine Verfahrensfunktion, $\Delta \in \Gamma_{a,b,n}$ ein Gitter der Ordnung n, $(u_l^{\varphi,\Delta})_{l=0}^n \in (\mathbb{R}^N)^{[0,n]}$ sei wie folgend rekursiv definiert :*

$$u_0^{\varphi,\Delta} := y_0,$$
$$\forall_{0 \leq l \leq n-1} : u_{l+1}^{\varphi,\Delta} := u_l^{\varphi,\Delta} + h_{\Delta,l}\varphi(t_l, u_l^{\varphi,\Delta}, h_{\Delta,l})$$

  ,dann heißt $(u_l^{\varphi,\Delta})_{l=0}^n$ das durch das Gitter Δ und die Verfahrensfunktion φ gegebene explizite Einschrittverfahren.

- *$V_\varphi := \bigcup_{n \in \mathbb{N}} \{(u_l^{\varphi,\Delta})_{l=0}^n \in (\mathbb{R}^N)^{[0,n]} \mid \Delta \in \Gamma_{a,b,n}\}$ heißt das durch die Verfahrensfunktion φ bestimmte explizite Einschrittverfahren.*

Bemerkung 2 (Simplifizierung der Notation) *Sollte aus dem Zusammenhang sowohl klar hervorgehen, welche Funktion φ dem Verfahren zugrunde liegt, als auch das spezielle Gitter Δ feststehen, so kann $(u_l^{\varphi,\Delta})_{l=0}^n$ durch $(u_l)_{l=0}^n$ und $h_{\Delta,l}$ durch h_l ersetzt werden. Hierauf wird in folgenden Kapiteln oft zurückgegriffen.*

Definition 4 (lokaler und globaler Verfahrensfehler)
Seien a,b wie in Def.1 gegeben. $\|\cdot\| : \mathbb{R}^N \to \mathbb{R}_{\geq 0}$ *sei eine Norm auf dem* $\mathbb{R}^N$.
$D_{a,b} := \{((t,h) \in [a,b] \times \mathbb{R}_{>0} \mid h \leq b - t\}$

- *Sei* φ *eine Verfahrensfunktion und sei* $(t,h) \in D_{a,b}$*,dann heißt*

$$\eta(t,h) := y(t) + h\varphi(t, y(t), h) - y(t+h)$$

der lokale Verfahrensfehler im Punkt $(t+h, y(t+h))$ *bezüglich der Schrittweite h.*

- *Sei* $(u_l^{\varphi,\Delta})_{l=0}^n \in (\mathbb{R}^N)^{[0,n]}$ *ein durch die Verfahrensfunktion* φ *und das Gitter* $\Delta \in \Gamma_{a,b,n}$ *bestimmtes Verfahren,dann heißt*

$$G_\varphi(\Delta) := \max_{0 \leq l \leq n-1} ||u_l^{\varphi,\Delta} - y(t_l)||$$

der globale Verfahrensfehler des Verfahrens $(u_l^{\varphi,\Delta})_{l=0}^n$.

Die quantitaive Beschreibung der Güte eines Verfahrens bei gegebener Verfahrensfunktion wird durch die Einführung nachfolgender Begriffe ermöglicht:

Definition 5 (Konsistenzordung und Konvergenzordnung)
Sei V_φ *ein durch die Verfahrensfunktion* φ *bestimmtes Verfahren.*

- *Man sagt* V_φ *besitzt Konsistenzordnung* $p \geq 1 \Leftrightarrow$

$$\exists_{C \geq 0} \forall_{(t,h) \in D_{a,b}} : \|\eta(t,h)\| \leq Ch^{p+1} \tag{4}$$

- *Man sagt* V_φ *besitzt Konvergenzordnung* $p \geq 1 \Leftrightarrow$

$$\exists_{K \geq 0} \forall_{n \in \mathbb{N}} \forall_{\Delta \in \Gamma_{a,b,n}} : G_\varphi(\Delta) \leq K h_{\Delta,max}^p \tag{5}$$

1.1.2 Ein Konvergenzkriterium für explizite ESV

Lemma 1.2 (Fehlerakkumulation) *Seien* $n \in \mathbb{N}, \quad L > 0, \quad (a_l)_{l=0}^n, (h_l)_{l=0}^n \in \mathbb{R}_{>0}^{[0,n]}$, $b \geq 0$ *und*

$$\forall_{0 \leq l \leq n-1} \quad a_{l+1} \leq (1 + h_l L) a_l + h_l b$$

$\forall_{0 \leq l \leq n}$ *sei* $x_l := \sum_{j=0}^{l-1} h_j$. *Dann gilt:*

$$\forall_{0 \leq l \leq n} \quad a_l \leq \frac{e^{Lx_l} - 1}{L} b + e^{Lx_l} a_0$$

Beweis : Durch Induktion nach l:

$$l = 0 \Rightarrow x_l = 0 \Rightarrow \frac{e^{Lx_l} - 1}{L} b + e^{Lx_l} a_0 = a_0$$

$$\begin{aligned} l \to l+1 : a_{l+1} &\leq \overbrace{(1 + h_l L)}^{\leq e^{h_l L}} a_l + h_l b \leq e^{h_l L} \left(\frac{e^{Lx_l} - 1}{L} b + e^{Lx_l} a_0 \right) + h_l b \\ &\leq \left(\frac{e^{L(x_l + h_l)} - 1 - h_l b}{L} + h_l \right) b + e^{L(x_l + h_l)} a_0 \\ &= \frac{e^{Lx_{l+1}} - 1}{L} b + e^{Lx_{l+1}} a_0 \end{aligned}$$

Satz 1.2 *Seien a,b,N wie in Def.1 gegeben und es sei $p \in \mathbb{N}$.Sei V_φ ein durch die Verfahrensfunktion φ bestimmtes Verfahren.Dann gilt : Wenn V_φ Konsistenzordnung p besitzt und (3) erfüllt,dann besitzt V_φ Konvergenzordnung p.*

Beweis : Seien L_φ, C die nach Vorraussetzung existierenden Konstanten in (2),(4). Sei $K := \frac{C}{L_\varphi}(e^{L_\varphi(b-a)} - 1)$, $n \in \mathbb{N}$ und $\Delta \in \Gamma_{a,b,n}$. Es gilt :

$$\begin{aligned} \forall_{0 \leq l \leq n-1} : & y(t_{l+1}) = y(t_l) + h_{t,l}\varphi(t_l, y_l, h_{t,l}) - \eta(t_l, h_{t,l}), \\ & u_{l+1}^{\varphi,\Delta} = u_l^{\varphi,\Delta} + h_{\Delta,l}\varphi(t_l, u_l^{\varphi,\Delta}, h_{\Delta,l}) \quad \Rightarrow \end{aligned}$$

$$\begin{aligned} \forall_{0 \leq l \leq n-1} : u_{l+1}^{\varphi,\Delta} - y(t_{l+1}) &= u_l^{\varphi,\Delta} + h_{\Delta,l}\varphi(t_l, u_l^{\varphi,\Delta}, h_{\Delta,l}) - y(t_l) - h_{t,l}\varphi(t_l, y(t_l), h_{\Delta,l}) \\ &\quad + \eta(t_l, h_{\Delta,l}) \quad \Rightarrow \end{aligned}$$

$$\begin{aligned} \forall 0 \leq l \leq n-1 : & \|u_{l+1}^{\varphi,\Delta} - y(t_{l+1})\| \\ &\leq \|u_l^{\varphi,\Delta} - y(t_l)\| + h_{\Delta,l}\|\varphi(t_l, u_l^{\varphi,\Delta}, h_{\Delta,l}) - \varphi(t_l, y(t_l), h_{\Delta,l})\| + \|\eta(t_l, h_{\Delta,l})\| \\ &\leq \|u_l^{\varphi,\Delta} - y(t_l)|| + h_{\Delta,l}L_\varphi\|u_l^{\varphi,\Delta} - y(t_l)|| + Ch_{\Delta,l}^{p+1} \\ &\leq (1 + h_{\Delta,l}L_\varphi)\|u_l^{\varphi,\Delta} - y(t_l)\| + h_{\Delta,l}Ch_{\Delta,max}^p \end{aligned}$$

Mit Lemma 1.2(wobei $a_l := \|u_l^{\varphi,\Delta} - y(t_l)\|, h_l := h_{\Delta,l}, L := L_\varphi, b := Ch_{\Delta,max}^p$, $x_l = t_l - a, \quad a_0 = 0$) folgt somit:

$$\forall_{0 \leq l \leq n} : \|u_l^{\varphi,\Delta} - y(t_l)\| \leq \frac{e^{L_\varphi(t_l - a)} - 1}{L_\varphi} Ch_{\Delta,max}^p \leq Kh_{\Delta,max}^p$$

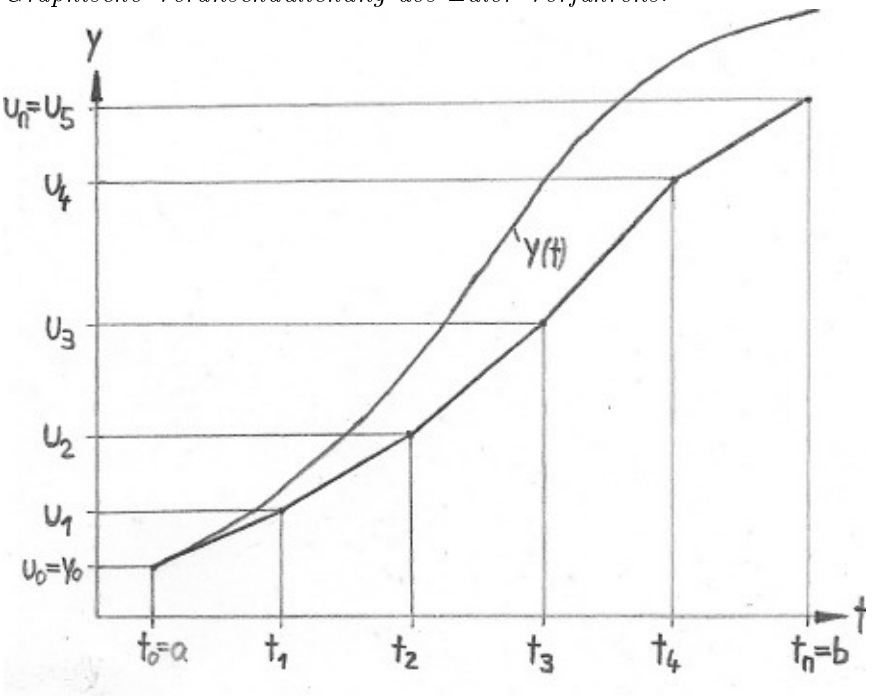
1.2 Spezielle Einschrittverfahren

Bemerkung 3
Im folgenden seien f_t, f_u die partiellen Ableitungen von f(im Fall N=1) nach der ersten bzw.zweiten Komponente.$||\cdot|| : \mathbb{R}^N \to \mathbb{R}_{\geq 0}$ sei eine Norm auf dem $\mathbb{R}^N$.

Beispiel 1 (Euler-Verfahren)
Das explizite Einschrittverfahren E_φ mit der Verfahrensfunktion

$$\begin{aligned}\varphi : D_{a,b,N} &\to \mathbb{R}^N \\ (t,u,h) &\mapsto f(t,u)\end{aligned}$$

wird als Euler-Verfahren(auch:Eulersches Polygonzugverfahren oder vorwärtsgerichtete Euler-Formel) bezeichnet.Ist $\Delta \in \Gamma_{a,b,n}$ ein Gitter der Ordnung n,so ist das Euler-Verfahren gegeben durch : $u_0^{\varphi,\Delta} = y_0$ und $\forall 0 \leq l \leq n-1 : u_{l+1}^{\varphi,\Delta} = u_l^{\varphi,\Delta} + h_{\Delta,l} f(t, u_l^{\varphi,\Delta})$
Graphische Veranschaulichung des Euler-Verfahrens:



Satz 1.3
Ist f 1-mal stetig partiell differenzierbar,so besitzt das Eulerverfahren E_φ Konsistenzordnung 1. Ist f zusätzlich lipschitzstetig im 2. Teilargument so besitzt E_φ Konvergenzordnung 1.
Beweis :
Wegen Lemma 1 ist y 2-mal stetig differenzierbar.Sei $C := \frac{1}{2}\max\{||y''(\tau)||_\infty \mid \tau \in [a,b]\}$,sei $(t,h) \in D_{a,b}$. Nach dem Satz von Taylor gilt $\forall_{1\leq l\leq N}\exists_{\tau_l\in[t,t+h]} : y(t+h) = y(t) + hy'(t) + \frac{h^2}{2}(y_l''(\tau_l))_{l=1}^N \Rightarrow$

$$\begin{aligned}||\eta(t,h)||_\infty &= ||y(t) + h\overbrace{f(t,y(t))}^{y'(t)} - (y(t) + hy'(t) + \frac{h^2}{2}\left(y_l''(\tau_l)\right)_{l=1}^N ||_\infty \\ &= \frac{h^2}{2}||\left(y_l''(\tau_l)\right)_{l=1}^N||_\infty \leq Ch^2\end{aligned}$$

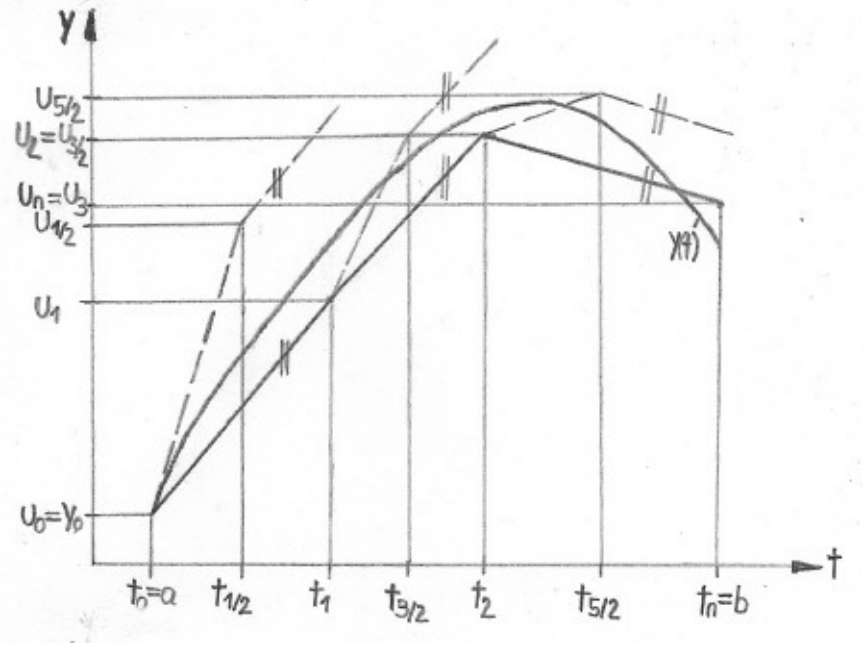
Aufgrund der Äquivalenz der Normen auf dem $\mathbb{R}^N$ ergibt sich die Behauptung für jede beliebige Norm. Sei nun f zusätzlich lipschitzstetig im 2. Teilargument. Wegen Satz 2 reicht es zu zeigen dass φ (3) erfüllt: Sei $L_\varphi := L_f$,seien $(t,u,h),(t,v,h) \in D_{a,b,N} \Rightarrow$

$$||\varphi(t,u,h) - \varphi(t,v,h)|| = ||f(t,u) - f(t,v)|| \leq L_\varphi ||u - v||$$

Beispiel 2 (modifiziertes Euler-Verfahren)
Das explizite Einschrittverfahren ME_φ mit der Verfahrensfunktion

$$\begin{aligned} \varphi : D_{a,b,N} &\to \mathbb{R}^N \\ (t,u,h) &\mapsto f\left(t + \frac{h}{2}, u + \frac{h}{2} f(t,u)\right) \end{aligned}$$

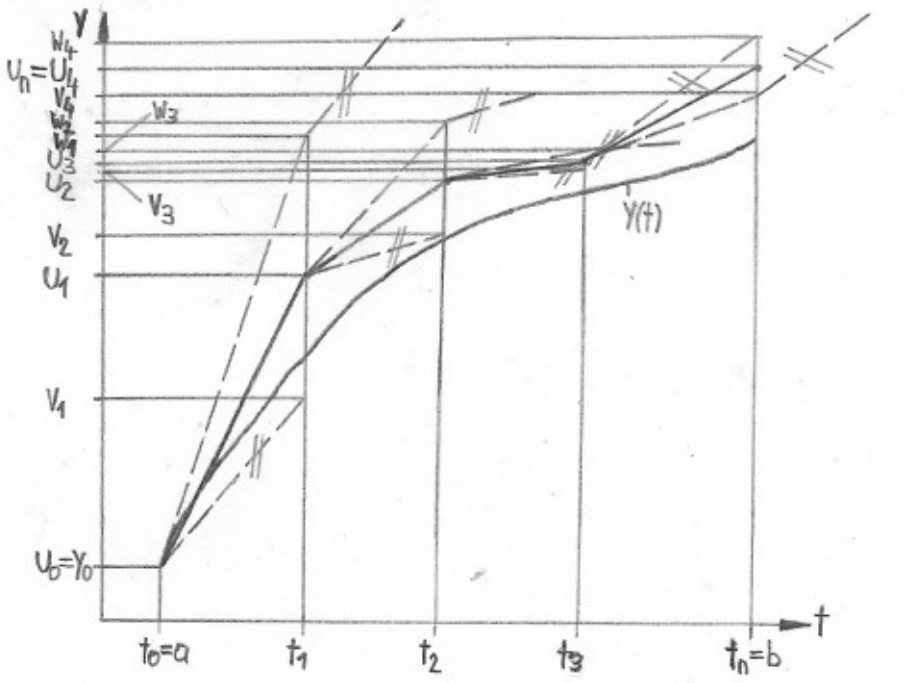
wird als modifiziertes Euler-Verfahren bezeichnet.
Ist $\Delta \in \Gamma_{a,b,n}$ ein Gitter der Ordnung n,so lässt sich das Verfahren $(u_l^{\varphi,\Delta})_{l=0}^n$ folgendermaßen darstellen : $u_0^{\varphi,\Delta} = y_0, \forall_{0 \leq l \leq n-1} : t_{l+\frac{1}{2}} := t_l + \frac{h_{\Delta,l}}{2}, u_{l+\frac{1}{2}}^{\varphi,\Delta} := u_l^{\varphi,\Delta} + \frac{h_{\Delta,l}}{2} f(t_l, u_l^{\varphi,\Delta}), u_{l+1}^{\varphi,\Delta} = u_l^{\varphi,\Delta} + h_{\Delta,l} f(t_{l+\frac{1}{2}}, u_{l+\frac{1}{2}}^{\varphi,\Delta})$
Graphische Veranschaulichung des modifizierten Euler-Verfahrens:



Beispiel 3 (Verfahren von Heun)
Das explizite Einschrittverfahren H_φ mit der Verfahrensfunktion

$$\begin{aligned} \varphi : D_{a,b,N} &\to \mathbb{R}^N \\ (t,u,h) &\mapsto \frac{1}{2} f(t,u) + \frac{1}{2} f(t+h, u + h f(t,u)) \end{aligned}$$

wird als Verfahren von Heun bezeichnet.
Ist $\Delta \in \Gamma_{a,b,n}$ *ein Gitter der Ordnung n,so lässt sich das Verfahren* $(u_l^{\varphi,\Delta})_{l=0}^n$ *folgendermaßen darstellen :* $\forall_{0\le l\le n-1} : v_l^{\varphi,\Delta} := u_l^{\varphi,\Delta} + h_{\Delta,l} f(t_l, u_l^{\varphi,\Delta}), w_l^{\varphi,\Delta} := u_l^{\varphi,\Delta} + h_{\Delta,l} f(t_l, v_l^{\varphi,\Delta}), u_{l+1}^{\varphi,\Delta} = \frac{1}{2}(v_l^{\varphi,\Delta} + w_l^{\varphi,\Delta})$
Graphische Veranschaulichung des Verfahrens von Heun:



Satz 1.4
Seien $a_1, a_2, b_1, b_2 \in \mathbb{R}$ *mit* $a_1 + a_2 = 1, a_2 b_1 = \frac{1}{2}, a_2 b_2 = \frac{1}{2}$, *Ist f 2-mal stetig partiell differenzierbar,so besitzt das Verfahren* $V_\varphi^{a_1,a_2,b_1,b_2}$ *das durch die Verfahrensfunktion* $\varphi : D_{a,b,N} \to \mathbb{R}^N$ *mit* $\varphi(t,u,h) = a_1 f(t,u) + a_2 f(t + b_1 h, u + b_2 h f(t,u))$ *bestimmt wird Konsistenzordnung 2. Ist f zusätzlich lipschitzstetig im 2. Teilargument so besitzt* $V_\varphi^{a_1,a_2,b_1,b_2}$ *Konvergenzordnung 2. Insbesondere ergibt sich die Aussage für das modifizierte Euler-Verfahren sowie das Verfahren von Heun.*
Beweis :(Nur für den Fall N=1)
Wegen Lemma 1 ist y 3-mal stetig differenzierbar.Nach dem Satz von Taylor gilt somit:

$$\forall (t,h) \in D_{a,b} : y(t+h) = y(t) + h y'(t) + \frac{h^2}{2} y''(t) + \mathcal{O}(h^3).$$

Weiters gilt nach dem Satz von Taylor :

$$\begin{aligned}\forall (t,h) \in D_{a,b} : \varphi(t,y(t),h) &= (a_1+a_2)f(t,y(t)) + h(a_2 b_1 f_t(t,y(t)) \\ &\quad + a_2 b_2 f_u(t,y(t))f(t,y(t))) + \mathcal{O}(h^2) \\ &= f(t,y(t)) + \frac{h}{2}(f_t(t,y(t)) \\ &\quad + f_u(t,y(t))f(t,y(t))) + \mathcal{O}(h^2). \quad \Rightarrow\end{aligned}$$

$$\begin{aligned}\forall (t,h) \in D_{a,b} : \eta(t,h) &= y(t) + h\varphi(t,y(t),h) - y(t+h) \\ &= y(t) + hf(t,y(t)) + \frac{h^2}{2}(f_t(t,y(t)) + f_u(t,y(t))f(t,y(t))) \\ &\quad - (y(t) + hf(t,y(t)) + \frac{h^2}{2}(f_t(t,y(t)) + f_u(t,y(t))f(t,y(t)))) + \mathcal{O}(h^3) \\ &= \mathcal{O}(h^3)\end{aligned}$$

Sei nun f zusätzlich lipschitzstetig im 2. Teilargument. Wegen Satz 2 reicht es zu zeigen das φ (3) erfüllt: Sei $L_\varphi := |a_1|L_f + |a_2|L_f + |a_2 b_2|(b-a)L_f^2$,seien $(t,u,h),(t,v,h) \in D_{a,b,N} \Rightarrow$

$$\begin{aligned}||\varphi(t,u,h) - \varphi(t,v,h)|| &= ||a_1 f(t,u) + a_2 f(t+b_1 h, u + b_2 h f(t,u)) \\ &\quad - a_1 f(t,v) - a_2 f(t+b_1 h, v + b_2 h f(t,v))|| \\ &\leq |a_1|L_f||u-v|| + |a_2|L_f||u + b_2 h f(t,u) - v - b_2 h f(t,v)|| \\ &\leq \left(|a_1|L_f + |a_2|L_f + |a_2 b_2| h L_f^2\right)||u-v|| \leq L_\varphi ||u-v||\end{aligned}$$

Da $ME_\varphi = V_\varphi^{1,0,\frac{1}{2},\frac{1}{2}}$ und $H_\varphi = V_\varphi^{\frac{1}{2},\frac{1}{2},1,1}$ gilt ergibt sich die Zusatzbehauptung.

Beispiel 4 (einfaches Kutta-Verfahren)
Das explizite Einschrittverfahren K_φ mit der Verfahrensfunktion

$$\begin{aligned}\varphi : D_{a,b,N} &\to \mathbb{R}^N \\ (t,u,h) &\mapsto \frac{1}{6}\left(k_1^{t,u} + 4k_2^{t,u,h} + k_3^{t,u,h}\right)\end{aligned}$$

wobei $k_1^{t,u} := f(t,u), k_2^{t,u,h} := f(t+\frac{h}{2}, u + \frac{h}{2}k_1^{t,u}), k_3^{t,u,h} := f(t+h, u + h(-k_1^{t,u} + 2k_2^{t,u,h}))$ wird als einfaches Kutta-Verfahren bezeichnet.

Satz 1.5
Ist f 3-mal stetig partiell differenzierbar,so besitzt das einfache Kutta-Verfahren K_φ Konsistenzordnung 3. Ist f zusätzlich lipschitzstetig im 2. Teilargument so besitzt K_φ Konvergenzordnung 3.

Beweis :(Nur für den Fall N=1)
Wegen Lemma 1 ist y 4-mal stetig differenzierbar.Dann gilt nach dem Satz von Taylor:

$$\begin{aligned}
\forall(t,h)\in D_{a,b}: y(t+h) &= y(t)+hy'(t)+\frac{h^2}{2}y''(t)+\frac{h^3}{6}y'''(t)+\mathcal{O}(h^4)\\
&= y(t)+hf(t,y(t))+\frac{h^2}{2}(f_t(t,y(t))+f_u(t,y(t))f(t,y(t)))\\
&+\frac{h^3}{6}(f_{tt}(t,y(t))+2f_{tu}(t,y(t))f(t,y(t))+f_{uu}(t,y(t))f^2(t,y(t))\\
&+f_t(t,y(t))f_u(t,y(t))+f_u^2(t,y(t))f(t,y(t)))+\mathcal{O}(h^4)
\end{aligned}$$

$$\begin{aligned}
\forall(t,h)\in D_{a,b}: k_2^{t,y(t),h} &= f(t,y(t))+\frac{h}{2}(f_t(t,y(t))+f_u(t,y(t))f(t,y(t)))+\frac{h^2}{8}(f_{tt}(t,y(t))\\
&+2f_{tu}(t,y(t))f(t,y(t))+f_{uu}(t,y(t))f^2(t,y(t)))+\mathcal{O}(h^3)
\end{aligned}$$

$$\begin{aligned}
\forall(t,h)\in D_{a,b}: k_3^{t,y(t),h} &= f(t,y(t))+hf_t(t,y(t))+h(-k_1^{t,y(t)}+2k_2^{t,y(t),h})f_u(t,y(t))\\
&+\frac{1}{2}(h^2f_{tt}(t,y(t))+2h^2(-k_1^{t,y(t)}+2k_2^{t,y(t),h})f_{tu}(t,y(t))\\
&+h^2(-k_1^{t,y(t)}+2k_2^{t,y(t),h})^2f_{uu}(t,y(t))))+\mathcal{O}(h^3)\\
&= f(t,y(t))+h(f_t(t,y(t))+f_u(t,y(t))f(t,y(t)))\\
&+\frac{h^2}{2}(2f_t(t,y(t))f_u(t,y(t))+2f_u^2(t,y(t))f(t,y(t))+f_{tt}(t,y(t))\\
&+2f(t,y(t))f_{tu}(t,y(t))+f^2(t,y(t))+f_{uu}(t,y(t)))+\mathcal{O}(h^3)
\end{aligned}$$

$$\begin{aligned}
\forall(t,h)\in D_{a,b}: \varphi(t,y(t),h) &= f(t,y(t))+\frac{h}{2}(f_t(t,y(t))+f_u(t,y(t))f(t,y(t)))\\
&+\frac{h^2}{6}(f_{tt}(t,y(t))+2f_{tu}(t,y(t))f(t,y(t))+f_{uu}(t,y(t))f^2(t,y(t))\\
&+f_t(t,y(t))f_u(t,y(t))+f(t,y(t))f_u^2(t,y(t)))+\mathcal{O}(h^3)\quad\Rightarrow
\end{aligned}$$

$$\begin{aligned}
\forall(t,h)\in D_{a,b}: \eta(t,h) &= y(t)+h\varphi(t,y(t),h)-y(t+h)\\
&= y(t)+hf(t,y(t))+\frac{h^2}{2}(f_t(t,y(t))+f_u(t,y(t))f(t,y(t)))\\
&+\frac{h^3}{6}(f_{tt}(t,y(t))+2f_{tu}(t,y(t))f(t,y(t))+f_{uu}(t,y(t))f^2(t,y(t))\\
&+f_t(t,y(t))f_u(t,y(t))+f(t,y(t))f_u^2(t,y(t)))\\
&-(y(t)+hf(t,y(t))+\frac{h^2}{2}(f_t(t,y(t))\\
&+f_u(t,y(t))f(t,y(t)))+\frac{h^3}{6}(f_{tt}(t,y(t))\\
&+2f_{tu}(t,y(t))f(t,y(t))+f_{uu}(t,y(t))f^2(t,y(t))\\
&+f_t(t,y(t))f_u(t,y(t))+f_u^2(t,y(t))f(t,y(t))))+\mathcal{O}(h^4)\\
&= \mathcal{O}(h^4)
\end{aligned}$$

Sei nun f zusätzlich lipschitzstetig in der 2.Komponente.Wegen Satz 2 reicht es zu zeigen das φ (3) erfüllt: Sei $L_\varphi := L_f + \frac{5}{6}L_f^2(b-a) + \frac{1}{6}L_f^3(b-a)^2$ *,seien* $(t,u,h),(t,v,h) \in D_{a,b,N} \Rightarrow$

$$\begin{aligned}
||\varphi(t,u,h) - \varphi(t,v,h)|| &= ||\frac{1}{6}\left(k_1^{t,u} + 4k_2^{t,u,h} + k_3^{t,u,h}\right) - \frac{1}{6}\left(k_1^{t,v} + 4k_2^{t,v,h} + k_3^{t,v,h}\right)|| \\
&\leq \frac{1}{6}\left(||k_1^{t,u} - k_1^{t,v}|| + 4||k_2^{t,u,h} - k_2^{t,v,h}|| + ||k_3^{t,u,h} - k_3^{t,v,h}||\right) \\
&\leq \frac{1}{6}\left(2L_f||u-v|| + 4L_f\left(||u-v|| + \frac{h}{2}L_f||u-v||\right)\right) \\
&+ \frac{1}{6}L_f h\left(L_f||u-v|| + 2L_f\left(||u-v|| + \frac{h}{2}L_f||u-v||\right)\right) \\
&= \left(L_f + \frac{5}{6}hL_f^2 + \frac{1}{6}h^2L_f^3\right)||u-v|| \\
&\leq L_\varphi||u-v||
\end{aligned}$$

Beispiel 5 (klassisches Runge-Kutta-Verfahren)
Das explizite Einschrittverfahren RK_φ *mit der Verfahrensfunktion*

$$\begin{aligned}
\varphi : D_{a,b,N} &\to \mathbb{R}^N \\
(t,u,h) &\mapsto \frac{1}{6}\left(k_1^{t,u} + 2k_2^{t,u,h} + 2k_3^{t,u,h} + k_4^{t,u,h}\right)
\end{aligned}$$

wobei $k_1^{t,u} := f(t,u), k_2^{t,u,h} := f(t+\frac{h}{2}, u+\frac{h}{2}k_1), k_3^{t,u,h} := f(t+\frac{h}{2}, u+\frac{h}{2}k_2)$,
$k_4^{t,u,h} := f(t+h, u+hk_3)$ *wird als klassisches Runge-Kutta-Verfahren bezeichnet.*

Satz 1.6
Ist f 4-mal stetig partiell differenzierbar,so besitzt das klassische Runge-Kutta-Verfahren RK_φ *Konsistenzordnung 4. Ist f zusätzlich lipschitzstetig in der 2.Komponente so besitzt* RK_φ *Konvergenzordnung 4.*
Beweis :(Nur für den Fall N=1)
Wegen Lemma 1 ist y 5-mal stetig differenzierbar.Dann ergibt sich die Behauptung ähnlich wie in Satz 5 durch Reihenentwicklung nach Taylor. Sei nun f zusätzlich lipschitzstetig im 2. Teilargument. Wegen Satz 2 reicht es zu zeigen das φ (3) erfüllt: Sei $L_\varphi := \frac{1}{6}(5L_f +$

$(b-a)L_f^2 + 2(b-a)L_f^3 + \frac{(b-a)^2}{2}L_f^4)$,*seien* $(t,u,h),(t,v,h) \in D_{a,b,N} \Rightarrow$

$$\begin{aligned}
||\varphi(t,u,h) - \varphi(t,v,h)|| &= ||\frac{1}{6}\left(k_1^{t,u} + 2k_2^{t,u,h} + 2k_3^{t,u,h} + k_4^{t,u,h}\right) \\
&\quad - \frac{1}{6}\left(k_1^{t,v} + 2k_2^{t,v,h} + 2k_3^{t,v,h} + k_4^{t,u,h}\right)|| \\
&\leq \frac{1}{6}\left(||k_1^{t,u} - k_1^{t,v}|| + 2||k_2^{t,u,h} - k_2^{t,v,h}||\right) \\
&\quad + \frac{1}{6}\left(2||k_3^{t,u,h} - k_3^{t,v,h}|| + ||k_4^{t,u,h} - k_4^{t,v,h}||\right) \\
&\leq \frac{1}{6}\left(3L_f||u-v|| + hL_f^2||u-v||\right) \\
&\quad + \frac{1}{6}\left(2L_f||u-v|| + hL_f^3||u-v||\right) \\
&\quad + \frac{1}{6}\left(L_f + hL_f^3 + \frac{h^2}{2}L_f^4\right)||u-v|| \\
&\leq \frac{1}{6}\left(5L_f + hL_f^2 + 2hL_f^3 + \frac{h^2}{2}L_f^4\right)||u-v|| \\
&\leq L_\varphi||u-v||
\end{aligned}$$

1.3 Rundungsfehleranalyse

Eine fehlerbehaftete Verfahrensvorschrift sei von der Form:

$$\begin{array}{lll} v_{\ell+1} = v_\ell + h * \varphi(t_\ell, v_\ell, h) + \rho_\ell & \ell = 0,\ldots,n & v_0 := y_0 + e_0 \\ \|\rho_\ell\| \leq \delta & \|e_0\| \leq \varepsilon & \end{array} \tag{6}$$

mit geeigneten Vektoren $e_0, p_\ell \in \mathbb{R}^n$ und einer Vektornorm $\|\cdot\|$.

Satz 1.7 *Zur Lösung des AWP:* $y' = f(t,y) \quad y(a) = y_0 \quad t \in [a,b]$
sei eine fehlerbehaftete Verfahrensfunktion wie in (6) *mit Konsistenzordnung* $p \geq 1$ *gegeben. Die Verfahrensfunktion* φ *soll einer Lipschitzbedingung in* v *genügen. Dann gelten folgende Abschätzungen für die durch* (6) *gewonnenen Approximationen:*

$$\max_{\ell \in \{0,\ldots,n\}} \|v_\ell - y(t_\ell)\| \leq K(h^p + \frac{\delta}{h}) + \varepsilon e^{L_\varphi(b-a)} \tag{7}$$

$$K := \frac{\max\{C,1\}}{L_\varphi}(e^{L_\varphi(b-a)} - 1)^1$$

[1] Für C siehe Definition der Konsistenzordnung

Beweis: Mit

$$e_\ell = v_\ell - y_\ell, \quad y_\ell = y(t_\ell), \quad \ell = 0, 1, \ldots, n,$$

$$\eta_\ell = \eta(t_\ell, h), \qquad \ell = 0, 1, \ldots, n-1,$$

gilt für $\ell = 0, 1, \ldots, n-1$

$$y_{\ell+1} = y_\ell + h\varphi(t_\ell, y_\ell; h) - \eta_\ell,$$

$$v_{\ell+1} = v_\ell + h\varphi(t_\ell, v_\ell; h) + \rho_\ell,$$

und daher

$$e_{\ell+1} = e_\ell + h\left[\varphi(t_\ell, v_\ell; h) - \varphi(t_\ell, y_\ell; h)\right] + \rho_\ell + \eta_\ell$$

beziehungsweise

$$\begin{aligned} \|e_{\ell+1}\| &\leq \|e_\ell\| + h\,\|\varphi(t_\ell, v_\ell; h) - \varphi(t_\ell, y_\ell; h)\| + \|\eta_\ell\| + \|\rho_\ell\| \\ &\leq (1 + hL_\varphi)\,\|e_\ell\| + h(Ch^p + \tfrac{\delta}{h}), \end{aligned}$$

Die Abschätzung $\|e_\ell\| \leq \varepsilon$ und Lemma 1.2 liefern zusammen die Aussage des obigen Satzes.

Bemerkungen

Die rechte Seite der Abschätzung (7) setzt sich aus 3 Teilen zusammmen:

Summand der oberen Schanke	**Abschätzung von**
$K \cdot h^p$	globalem Verfahrensfehler des Einschrittverfahrens
$K \cdot \frac{\delta}{h}$	aufgehäuften Rundungsfehlern
$\varepsilon e^{L_\varphi(b-a)}$	fehlerbehaftetem Anfangswert

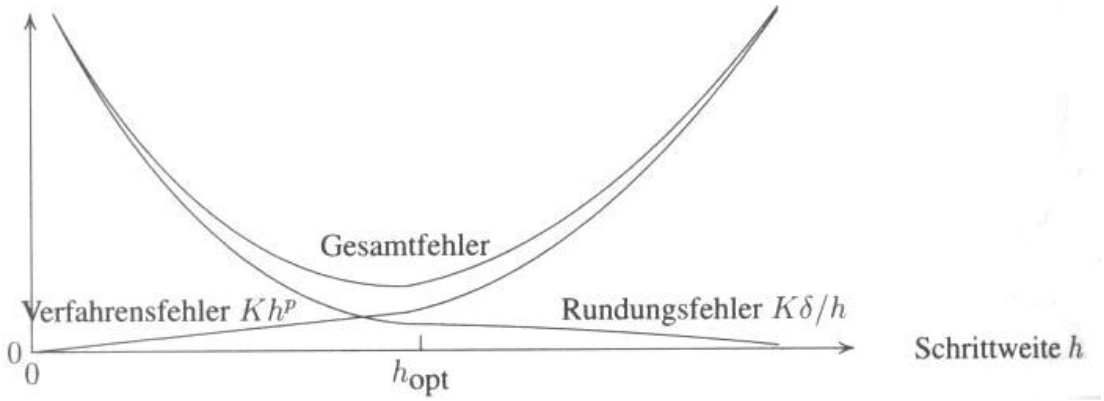
Lemma 1.3 (optimale Wahl von h:) *Es sei $\varepsilon = 0$, d.h. $v_0 = y_0$*
Dann gilt mit der Konstante K (wie oben) die Fehlerabschätzung:

$$\max_{\ell \in \{0,\ldots,n\}} \|v_\ell - y(t_\ell)\| \leq K(h^p + \frac{\delta}{h})$$

Mit der Wahl von $h = h_{opt} = \left(\frac{\delta}{p}\right)^{\frac{1}{p+1}}$ erhält man:

$$\max_{\ell \in \{0,\ldots,n\}} \|v_\ell - y(t_\ell)\| \leq \left(\frac{2K}{p^{\frac{p}{p+1}}}\right) \delta^{\frac{p}{p+1}}$$

Skizze



Der Gesamtfehler in Abhängigkeit von Verfahrensfehler und Rundungsfehler[2]

1.4 Asymptotische Entwicklung der Approximation

Vorbemerkungen

Es sei $\forall_{t\in[a,b]} \quad \mathbb{H}_t := \left\{\frac{t-a}{m} : m \in \mathbb{N}\right\}$. Eine hilfreiche Schreibweise im Rahmen der Betrachtung der Schrittweitenabhängigkeit von Approximationen sei nun angeführt:

$$\forall_{t\in[a,b]}\forall_{h=\frac{(t-a)}{n}\in\mathbb{H}_t} \qquad \begin{array}{l} u_h(a) := y_0, \\ u_h(a+(\ell+1)h) := u_h(a+\ell h) + h\varphi(a+\ell h, u_h(a+\ell h), h) \\ \ell = 0,\ldots,n-1 \end{array} \tag{8}$$

u_h heißt **Gitterfunktion** von φ.
Besitzt das zugrunde liegende Einschrittverfahren die Konsistenzordnung $p \geq 1$ und erfüllt φ eine Lipschitz-Bedingung (2) so gilt nach Satz 1.2:

$$u_h(t) = y(t) + \mathcal{O}(h^p), \qquad h \to 0, \ ,$$

was im Grunde genommen eine Kurzschreibweise für (5) darstellt. In dem Bemühen, aus bereits bestehenden konsistenten Verfahren, deren Verfahrensfunktion gewisse Differenzierbarkeitseigenschaften erfüllt, Approximationen höherer Ordnung zu gewinnen, erweist es sich als nützlich, das Verhalten von $u_h(t) - y(t); \quad t \in [a,b], h \in \mathbb{H}_t$ einer genaueren Untersuchung zu unterziehen. Das Hauptresultat einer solchen Untersuchung soll im folgenden Satz präsentiert werden.

[2]Quelle: Plato - Numerische Mathemetik kompakt; 2. Auflage, Vieweg 2004

Satz 1.8 (asymptotische Entwicklung des globalen Verfahrensfehlers) *Eine Verfahrensfunktion φ besitze die Konsistenzordnung $p \geq 1$ und sei lipschitzstetig im Sinne von (2). f und φ seien $(p+r)$-mal stetig partiell differenzierbar.*
Dann gibt es Funktionen $c_{p+j} \in \mathcal{C}^{r+1-j}([a,b],\mathbb{R}^n) \quad j = 1,\ldots,r-1$ mit $c_{p+j}(a) = 0$, sodass:

$$u_h(t) = y(t) + \sum_{j=0}^{r-1} c_{p+j}(t)h^{p+j} + \mathcal{O}(h^{p+r}), \quad t \in [a,b], \quad h \in \mathbb{H}_t := \left\{ \frac{t-a}{m} | m \in \mathbb{N} \right\} \quad (9)$$

Den **Beweis** hierzu findet man z.B. in [1].
Für den lokalen Verfahrensfehler erhält man folgenden

Satz 1.9 *(asymptotische Entwicklung des lokalen Verfahrensfehlers) Unter den Voraussetzungen des vorherigen Satzes gilt für jedes fixe $\ell \in \mathbb{N}$ die folgende Entwicklung des lokalen Verfahrensfehlers:*

$$u_h(a+\ell h) = y(a+\ell h) + b_{p+1}h^{p+1} + b_{p+2}h^{p+2} + \cdots + b_{p+r-1}h^{p+r-1} + \mathcal{O}(h^{p+r}) \quad (10)$$

mit $h > 0$ und von ℓ abhängigen Koeffizienten $b_{p+1},\ldots,b_{p+r} \in \mathbb{R}^n$

Beweis von (10):

$$c_{p+j}(a+\ell h) = \sum_{k=1}^{r-j-1} c_{p+j}^k(a)\tfrac{(\ell h)^k}{k!} + \mathcal{O}(h^{r-j}) \quad j = 0,1,\ldots,r-1$$

$$\Longrightarrow u_h(a+\ell h) = y(a+\ell h) + \sum_{j=0}^{r-1} c_{p+j}(a+\ell h)^{p+j} + \mathcal{O}(h^{p+r})$$

$$\Longrightarrow u_h(a+\ell h) = y(a+\ell h) + \sum_{s=1}^{r-1} \underbrace{\left[\sum_{k=1}^{s} c_{p+s-k}^{(k)}(a)\frac{\ell^k}{k!} \right]}_{=:b_{p+s}} h^{p+s} + \mathcal{O}(h^{p+r})$$

1.5 Extrapolation von Verfahren höherer Ordnung

Aus der Darstellung des Verfahrensfehlers

$$u_h(t) = y(t) + \sum_{j=0}^{r-1} c_{p+j}(t)h^{p+j} + \mathcal{O}(h^{p+r}), \quad t \in [a,b], \quad h \in \mathbb{H}_t := \{\frac{t-a}{n} | n \in \mathbb{N}\} \quad (11)$$

für ein Verfahren mit den Voraussetzungen aus Satz 1.8 wollen wir in unserem Bestreben, y zu approximieren, Nutzen ziehen. Zu diesem Zweck betrachten für ein fixiertes $t \in [a,b]$ und sukzessive verkleinerte $h_k \in \mathbb{H}_t$ $h_k < h_{k-1}$, $1 \leq k \leq m+1$; $\quad m \leq r$ das Polynom

$$\mathcal{P}(\xi) = d_0 + \sum_{j=0}^{m-1} d_{p+j}\xi^{p+j},$$

wobei die d_0, d_{p+j}, $0 \leq j \leq m-1$ durch die Interpolationsbedingungen $\mathcal{P}(h_k) = u_{h_k}(t)$ gegeben seien (zur Eindeutigkeit und Existenz eines solchen Polynoms siehe z.B. [2]). Es stellt sich heraus, dass unter genannten Bedingungen der konstante Term des Polynoms $\mathcal{P}$ als geeignete Approximation an y fungiert. Im Folgenden nehmen wir $h_k = \frac{h}{n_k}$, $0 \leq k \leq m$ zu festgelegtem $h \in \mathbb{H}_t$ und $n_k \in \mathbb{N}$ mit $n_m > n_{m-1} > \cdots > n_0$ an und benennen das aus den entsprechenden Interpolationsbedingungen resultierende Polynom $\mathcal{P}_h$.

Satz 1.10 *Es seien die Voraussetzungen von Satz 1.8 erfüllt. Für das Verfahren V_φ und für $t \in [a,b]$, $h \in \mathbb{H}_t$, $n_k \in \mathbb{N}$, $0 \leq k \leq m \leq r$ sei $\mathcal{P}_h$ wie oben gegeben. Dann gilt:*

$$\exists_{\Lambda_m, \Lambda_{m+1}, \ldots, \Lambda_{r-1} \in \mathbb{R}^{N \times N}} \quad \mathcal{P}_h(0) = y(t) + \sum_{\kappa=m}^{r-1} \left(\Lambda_\kappa c_{p+\kappa}(t)h^{p+\kappa}\right) + \mathcal{O}(h^{p+r}),$$

wobei die $c_{p+\kappa}$; $m \leq \kappa \leq r-1$ aus (9) stammen.

Beweis: Zunächst führen wir ihn für $N = 1$ und schreiben die Interpolationsbedingung in Matrixform:

$$\overbrace{\begin{pmatrix} 1 & \frac{1}{n_o^p} & \frac{1}{n_o^{p+1}} & \cdots & \frac{1}{n_o^{p+m-1}} \\ 1 & \frac{1}{n_1^p} & \frac{1}{n_1^{p+1}} & \cdots & \frac{1}{n_1^{p+m-1}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \frac{1}{n_m^p} & \frac{1}{n_m^{p+1}} & \cdots & \frac{1}{n_m^{p+m-1}} \end{pmatrix}}^{=:\mathrm{A}} \begin{pmatrix} d_0 \\ d_p h^p \\ d_{p+1}h^{p+1} \\ \vdots \\ d_{p+m-1}h^{p+m-1} \end{pmatrix} = \begin{pmatrix} u_{h_0}(t) \\ u_{h_1}(t) \\ \vdots \\ u_{h_m}(t) \end{pmatrix} \quad (12)$$

Aufgrund der Eindeutigkeit des interpolierenden Polynoms ist A invertierbar. Aus der asymptotischen Entwicklung (11) folgt

$$\begin{pmatrix} u_{h_0}(t) \\ u_{h_1}(t) \\ \vdots \\ u_{h_m}(t) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{n_o^p} & \frac{1}{n_o^{p+1}} & \cdots & \frac{1}{n_o^{p+m-1}} \\ 1 & \frac{1}{n_1^p} & \frac{1}{n_1^{p+1}} & \cdots & \frac{1}{n_1^{p+m-1}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \frac{1}{n_m^p} & \frac{1}{n_m^{p+1}} & \cdots & \frac{1}{n_m^{p+m-1}} \end{pmatrix} \begin{pmatrix} y(t) \\ c_p(t)h^p \\ \vdots \\ c_{p+m-1}(t)h^{p+m-1} \end{pmatrix} + \omega(h) \quad (13)$$

mit

$$\omega(h) = \sum_{k=m}^{r-1} c_{p+k}(t)h^{p+k} \begin{pmatrix} \frac{1}{n_o^{p+k}} \\ \frac{1}{n_1^{p+k}} \\ \vdots \\ \frac{1}{n_m^{p+k}} \end{pmatrix} + \mathcal{O}(h^{p+r}) \quad (14)$$

(12) und (13) ergeben

$$\begin{pmatrix} 1 & \frac{1}{n_o^p} & \frac{1}{n_o^{p+1}} & \cdots & \frac{1}{n_o^{p+m-1}} \\ 1 & \frac{1}{n_1^p} & \frac{1}{n_1^{p+1}} & \cdots & \frac{1}{n_1^{p+m-1}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \frac{1}{n_m^p} & \frac{1}{n_m^{p+1}} & \cdots & \frac{1}{n_m^{p+m-1}} \end{pmatrix} \begin{pmatrix} d_0 - y(t) \\ (d_p - c_p(t))h^p \\ \vdots \\ (d_{p+m-1} - c_{p+m-1}(t))h^{p+m-1} \end{pmatrix} = \omega(h) \quad (15)$$

Durch Multiplikation von (15) mit A^{-1} auf beiden Seiten und Betrachtung der ersten Zeile erhält man schließlich die Behauptung. Im höherdimensionalen Fall ersetze man die Einträge $\frac{1}{n_k{}^{p+j}}$ in A durch $\frac{1}{n_k{}^{p+j}}\mathbb{E}_N$ und die Einträge in den Vektoren durch ihre höherdimensionalen Gegenstücke (man beachte die geringfügige Änderung der Notation in (14).

Insgesamt haben wir unter Verwendung von (11) also eine Approximation von y erhalten, deren Konvergenzordnung p+m beträgt.

Beispiel 6 *Wir betrachten den Fall* $u_h(t) = y(t) + c_p(t)h^p + c_{p+1}(t)h^{p+1} + \mathcal{O}(h^{p+2})$ und $\mathcal{P}(h) = d_0 + d_p h^p$. *Mit* (12) *erhält man für* $1 \leq j \leq N$*:*

$$d_0^j + d_p^j h^p = u_h^j(t) \text{ und } d_0^j + d_p^j (h/n_1)^p = u_{(h/n_1)}^j(t)$$

Eine einfache Rechnung ergibt

$$d_0 = u_{h/n_1}(t) + \frac{u_{h/n_1}(t) - u_h(t)}{n_1^p - 1}$$

Leicht verifiziert man für die Matrix A, welche im besprochenen Fall die Form

$$\begin{pmatrix} \mathbb{E}_N & \mathbb{E}_N \\ \mathbb{E}_N & \frac{1}{n_1^p}\mathbb{E}_N \end{pmatrix}$$

hat, folgendes:

$$A^{-1} = \begin{pmatrix} \frac{1}{1-n_1^p}\mathbb{E}_N & \frac{n_1^p}{n_1^p-1}\mathbb{E}_N \\ \frac{n_1^p}{n_1^p-1}\mathbb{E}_N & \frac{-n_1^p}{n_1^p-1}\mathbb{E}_N \end{pmatrix}$$

Aufgrund

$$\begin{pmatrix} \frac{1}{1-n_1^p}\mathbb{E}_N & \frac{n_1^p}{n_1^p-1}\mathbb{E}_N \end{pmatrix} \begin{pmatrix} c_{p+1}(t) \\ c_{p+1}(t)\frac{1}{n_1^{p+1}} \end{pmatrix} = c_{p+1}(t)\frac{n_1-1}{(1-n_1^p)n_1}$$

gilt:

$$d_0 = \mathcal{P}(0) = y(t) + \frac{n_1-1}{(1-n_1^p)n_1}c_{p+1}(t)h^{p+1} + \mathcal{O}(h^{p+2})$$

Die einfache stetige Differenzierbarkeit von c_{p+1} ergibt für $t = \ell h; \ell \in \mathbb{N}$ die folgende Taylorentwicklung (unter Berücksichtigung, dass $c_{p+1}(a) = 0$):

$$c_{p+1}(a + \ell h) = \mathcal{O}(h).$$

Somit folgt:

$$\mathcal{P}(0) = y(t) + \mathcal{O}(h^{p+2})$$

1.6 Schrittweitensteuerung

1.6.1 Einführung

Es sei wieder das AWP $y' = f(t,y) \quad y(a) = y_0$ mit gegebenener Verfahrensfunktion $\varphi : D_{a,b,N} \longrightarrow \mathbb{R}^n$ der Konsitenzordnung $p \geq 1$ zu lösen. Dazu ziehen wir diesmal die folgende Vorschrift heran:

$$\begin{aligned} &w = u_\ell + \tfrac{h_\ell}{2}\varphi\left(t_\ell, u_\ell; \tfrac{h_\ell}{2}\right) \\ &u_{2\times h/2} := u_{\ell+1} = w + \tfrac{h_\ell}{2}\varphi\left(t_\ell + \tfrac{h_\ell}{2}, w; \tfrac{h_\ell}{2}\right), \quad t_{\ell+1} = t_\ell + h_\ell, \quad \ell = 0, 1 \ldots \end{aligned} \tag{16}$$

Das Ziel ist bei jedem Schritt die Schrittweite h_ℓ so zu wählen, dass der Fehler der Approximation durch die u_ℓ kontrollierbar klein bleibt!

Bemerkung 4 *Der Schritt $(t_\ell, u_\ell) \to (t_{\ell+1}, u_{\ell+1})$ der hier angewandten Verfahrensvorschrift entspricht einfach zwei Schritten $(t_\ell, u_\ell) \to (t_{\ell+1/2}, u_{\ell+1/2}) \to (t_{\ell+1}, u_{\ell+1})$ des ganz zu Beginn behandelten Einschrittsverfahrens mit halber Schrittweite $h_\ell/2$.*

1.6.2 Problemstellung

Wir definieren zunächst $z : [t_\ell, b] \to \mathbb{R}$ als analytische Lösung des AWP

$$z' = f(t, z), \qquad t \in [t_\ell, b]; \qquad z(t_\ell) = u_\ell \tag{17}$$

Man beachte hier, dass $z(t)$ die **nicht** notwendig Lösung des AWP mit dem Anfangswert $z(a) = y_0$ ist.
Es soll nun an der Stelle $t_\ell \in [a, b]$ der Fehler der Approximation $u_\ell \approx y(t_\ell) \in \mathbb{R}$ in der Größenordnung eines bestimmter Wert $\varepsilon > 0$ liegen, also:

$$\|u_{\ell+1} - z(t_\ell + h_\ell)\| \approx \varepsilon \tag{18}$$

mit $u_{\ell+1} \in \mathbb{R}^N$ aus dem hier betrachteten Verfahren.

Bemerkung 5 *Die Forderung* (18) *für die Schrittweitensteurung beruht auf dem loaklen Verfahrensfehler. Man erhofft sich damit naheliegender weise einen möglichst geringen globalen Verfahrensfehler.*

Einerseits soll die $\|u_{\ell+1} - z(t_\ell + h_\ell)\|$ die vorgegebene Schranke $\epsilon >$ nicht überschreiten was dadurch erreicht wird, dass man die Schrittweite h_ℓ klein macht. Andererseits ist es aber auch nicht wünschenswert dass h_ℓ sehr klein gemacht wird sodass $\|u_{\ell+1} - z(t_\ell + h_\ell)\| << \epsilon$ da sich dann die Rundungsfehler häufen würden.

Bemerkung 6 *Die Näherung an z soll uns befähigen, von der Integralkurve von f, auf der (t_ℓ, u_ℓ) liegt, so wenig wie möglich abzuweichen. Es ist zu beachten dass $z(t)$* ***auch noch nicht bekannt*** *ist. Wir werden aber im nächsten Abschnitt eine Näherung besprechen. Diese kann aufgrund der Defition von z und der speziellen Form unseres Verfahrens mit geringem Rechenaufwand aus 1.10 bezogen werden.*

Es folgt eine schematische Beschreibung des Ablaufs der Schrittweitensteuerung. Um die Schrittweite h_ℓ zu bestimmen, für die unsere Forderung (18) erfüllt ist , nehmen wir eine nicht zu kleine Startschrittweite h^0 und gehen für $k = 0, 1 \ldots$ wie folgt vor:

- Zunächst berechnet man für $h^{(k)}$ den Vektor $u_{2\times h^{(k)}/2} \in \mathbb{R}^N$.
- Anschließend wird der Fehler $\|u_{2\times h^{(k)}/2} - z(t_\ell + h^{(k)})\|$ geschätzt. Man bricht den Iterationsprozess ab falls der Fehler kleiner als ε ist und nennt $k_\varepsilon := k$.
- Falls die Schätzung hingegen größer als ε ausfällt, ermittelt man eine neue Testschrittweite $h^{(k+1)} \leq h^{(k)}$ und fängt wieder von vorne an.

Wenn wir unser passendes $h^{(k_\varepsilon)}$ ermittelt haben fährt man damit, $h_\ell = h^{(k_\varepsilon)}$ und $t_{\ell+1} = t_\ell + h^{(k_\varepsilon)}$, in der Verfahrensvorschrift fort.

1.6.3 Näherung von z(t)

Zur Abschätzung des Fehlers $\|u_{2\times h^{(k)}/2} - z(t_\ell + h^{(k)})\|$ muss $z(t_\ell + h^{(k)})$ genähert werden. Wir schätzen $z(t_\ell + h^{(k)})$ durch $z_{h^{(k)}} \in \mathbb{R}^N$ wie folgt ab:

$$z_h := u_{2\times h/2} - \frac{v_h - u_{2\times h/2}}{2^p - 1} \quad \text{mit} \quad v_h := u_\ell + h\varphi(t_\ell, u_\ell; h), \quad h > 0 \tag{19}$$

Dabei erhält man z_h mittels lokaler Extrapolation entsprechend Beispiel 6 mit $n_1 = 2$. Damit erhält man durch Umformen und Bilden der Norm für den genähernden Fehler:

$$\delta^{(k)} := \|u_{2\times h^{(k)}/2} - z_{h^{(k)}}\| = \frac{\|v_h - u_{2\times h/2}\|}{2^p - 1} \tag{20}$$

Ist also die Abschätzung $\delta^{(k)} \leq \varepsilon$ begnügt man sich mit $h_\ell = h^{(k)}$ und fährt dann mit ℓ um eins erhöht fort.

1.6.4 Bestimmung einer neuen Testschrittweite

Gilt jedoch für ein $h^{(k)}$, dass $\delta^{(k)} > \varepsilon$, wiederhohlt man die Berechnung mit k um eins erhöht , also mit einer neuen Testschrittweite $h^{(k+1)} < h^{(k)}$. Es wäre prinzipiell möglich irgendein $h^{(k+1)}$ zu wählen mit der einzigen Anforderung, dass es kleiner als $h^{(k)}$ sein soll. Dies ist aber einerseits nicht optimal da es sein könnte dass wir unser $h^{(k+1)}$ unnötig klein machen und somit die Rundungsfehler erhöhen. Andererseits ist unser neues $h^{(k+1)}$ vielleicht immer noch zu groß und wir müssen einen neuen Versuch mit einem $h^{(k+2)}$, $h^{(k+3)}$, ... starten was in der praktischen Anwendung Rechenzeit kosten würde.
Darum bedient man sich bei der Festlegung einer neuen passenden Testschrittweite $h^{(k+1)}$ der Darstellung des genäherten Fehlers $u_{2\times h/2} - z_{t_\ell + h}$!

Lemma 1.4 *Unter den Bedingungen von Satz 1.8 über die Asymptotik des globalen Verfahrensfehlers (dort für $r = 2$), kann man den Fehler wie folgt abschätzen:*

$$\|u_{2\times h/2} - z_{t_\ell+h}\| = \left(\frac{h}{h^{(k)}}\right)^{p+1} \delta^{(k)} \quad + \quad \mathcal{O}\left(\left(h^{(k)}\right)^{p+2}\right), \quad 0 < h \leq h^{(k)} \tag{21}$$

Dieser genäherte Fehler soll also ungefähr gleich ε sein:

$$\varepsilon = \left(\frac{h}{h^{(k)}}\right)^{p+1} \delta^{(k)} + \mathcal{O}\left(\left(h^{(k)}\right)^{p+2}\right)$$

Unter Vernachlässigung des Fehlers der Ordung $\left(h^{(k)}\right)^{p+2}$ erhält man durch Umformen:

$$h^{(k+1)} \quad := \quad h \quad = \quad \left(\frac{\varepsilon}{\delta^{(k)}}\right)^{1/(p+1)} h^{(k)} \tag{22}$$

Nun haben wir unser passendes $h^{(k+1)}$ mit dem unserem Fehler $\delta^{(k)}$, der bis auf eine Abweichung der Größenordnung $\mathcal{O}\left(\left(h^{(k)}\right)^{p+2}\right)$, in Nähe unseres ε liegt! Mit diesem $h^{(k+1)}$ fahren wir jetz in unserer Schema fort.

Beweis von Abschätzung (21):

Gemäß Satz 1.10 exsistiert ein von h unabhängiger Vektor $b_{p+1} \in \mathbb{R}^N$ derart, dass:

$$u_{2\times h/2} - z_{t_\ell+h} \quad = \quad b_{p+1}h^{p+1} \quad + \quad \mathcal{O}(h^{p+2}), \qquad h > 0, \tag{23}$$

gilt. Wir werden hier jetzt eine Abschätzung von b_{p+1} liefern um die Richtigkeit der Aussage (21) zu zeigen .

Aus Beispiel 6 über die Asymptotik des globalen Verfahrensfehlers geht hervor dass der Fehler

$$z_h \quad - \quad z(t_\ell + h) \quad = \quad \mathcal{O}(h^{p+2})$$

die Ordnung h^{p+2} hat. Wenn man dies nun in (23) einsetzt erhält man:

$$u_{2\times h/2} - z_h \quad = \quad b_{p+1}h^{p+1} \quad + \quad \mathcal{O}(h^{p+2}) \tag{24}$$

Wenn man nun die Norm bildet und man statt dem allgemeinen h, $h^{(k)}$ schreibt, erhält man:

$$\|u_{2\times h^{(k)}/2} - z_{h^{(k)}}\| \quad = \quad \|b_{p+1}\|(h^{(k)})^{p+1} \quad + \quad \mathcal{O}\left((h^{(k)})^{p+2}\right)$$

Mit der Beziehung $\|u_{2\times h^{(k)}/2} - z_{h^{(k)}}\| = \delta^{(k)}$ und durch Umformen ergibt sich folgender Ausdruck:

$$\|b_{p+1}\| = \frac{\delta^{(k)}}{(h^{(k)})^{p+1}} \quad + \quad \mathcal{O}((h^{(k)}) \tag{25}$$

Nun bildet man von (23) die Norm und setzt den gerade erhaltenen Ausdruck für $\|b_{p+1}\|$ ein.

$$\begin{aligned}\|u_{2\times h/2} - z_h\| &= \left(\frac{\delta^{(k)}}{(h^{(k)})^{p+1}} \quad + \quad \mathcal{O}((h^{(k)})\right) * h^{(p+1)} \quad + \quad \mathcal{O}(h^{p+2}) \\ &= \quad \left(\frac{h}{h^{(k)}}\right)^{p+1} \delta^{(k)} \quad + \quad \mathcal{O}(h^{(k)}) * h^{p+1} \quad + \quad \mathcal{O}(h^{p+2})\end{aligned}$$

mit $\quad 0 < h \leq h^{(k)} \quad$ folgt:

$$\|u_{2\times h/2} - z_{t_\ell+h}\| = \left(\frac{h}{h^{(k)}}\right)^{p+1} \delta^{(k)} \quad + \quad \mathcal{O}\left(\left(h^{(k)}\right)^{p+2}\right),$$

also die Behauptung.

Bemerkung 7 *Ein paar Ergänzungen:*

- *Für den Startschritt empfiehlt sich die Wahl* $h^{(0)} = \epsilon^q$ *mit* $1/(p+2) < q < 1$.
- *Zu der hier vorgestellten Schrittweitenstrategie gibt es natürlich auch Alternativen. Ebenfalls sinnvoll wäre einen Epsilonbalken anzugeben indem das Abbruchskritrium als erfüllt gilt, wenn:* $c_1\varepsilon < \delta(k_\varepsilon) < c_2\varepsilon$ *gilt. Falls* h^k *nicht passt kann* δ^k *oberhalb oder unterhalb des Epsilonbereichs liegen, also kan hier offentsichlich auch eine Schrittweitenvergrößerung von* $h^{k+1} > h^k$ *stattfinden.*
- *Es könnte natürlich auch sein, dass die in gewissen Fällen das erfüllen der Bedingung, dass der Fehler in der Größenordnung von* ε *liegt, nicht erreicht werden kann. In diesen Fällen müsste man nach einer bestimmten Anzahl von Versuchschritten* $k^\varepsilon < \infty$ *abbrechen.*

Pseudocode zur Schrittweitensteuerung

Der folgende Pseudocode soll den schematischen Aufbau eines Computerprogramms zur Schrittweitensteuerung darstellen. Die Verfahrensvorschrift sei die selbe wie in der Einführung dieses Kapitels.

Zu Beginn seien: $t_0 = a, \quad u_0 = y_0, \quad l = 0, \quad$ und $h^{(0)}, \varepsilon > 0$ *repeat* $\quad k = 0;$

repeat

if $\quad k = 0 \quad$ *then* $\quad h = h^{(0)} \quad$ *else* $\quad h = \left(\frac{\varepsilon}{\delta}\right)^{p+1} * h \quad$ *end*;

$w = u_\ell + \frac{h}{2}\varphi\left(t_\ell, u_\ell; \frac{h}{2}\right); \qquad u_{\ell+1} = w + \frac{h}{2}\varphi\left(t_\ell + \frac{h}{2}, w; \frac{h}{2}\right);$

$v = u_\ell + h\varphi\left(t_\ell, u_\ell; h\right); \qquad \delta = \frac{\|v - u_{l+1}\|}{2^p - 1}; \qquad k = k + 1;$

until $\qquad \delta \leq \varepsilon;$

$t_{\ell+1} = t_\ell + h; \quad l = l + 1;$

until $\qquad t_\ell \geq b;$

2 Allgemeine Theorie der Mehrschrittverfahren

2.1 Grundlagen

Definition 6 *Ein m-Schrittverfahren zur näherungsweisen Bestimmung einer Lösung des Anfangswertproblems* $y' = f(t,y), \quad y(a) = y_0$ *hat die Form:*

$$\sum_{j=0}^{m} \alpha_j u_{\ell+j} = h\varphi(t_\ell, u_\ell, \ldots, u_{\ell+m}; h), \qquad \ell = 0, \ldots, n-m, \tag{26}$$

mit

- *Koeffizienten* $\alpha_j \in \mathbb{R}$ *mit* $\alpha_m \neq 0$ *und einer Funktion*

$$\varphi : [a,b] \times (\mathbb{R}^N)^{m+1} \times \mathbb{R}_+ \longrightarrow \mathbb{R}^N \tag{27}$$

- *Schrittweiten*

$$t_\ell = a + \ell h \qquad \text{für } \ell = 0, 1, \ldots, n \qquad \text{mit } h = \tfrac{b-a}{n}, \tag{28}$$

- *und Startwerten* $u_0, \ldots, u_{m-1} \in \mathbb{R}^N$.

Bemerkung 8 *Ein paar Erläuterungen zu Definition 6:*

- *In den meisten Fällen setzt man* $u_0 := y_0$ *und die weiteren Startwerte* $u_1, u_2, \ldots, u_{m-1} \in \mathbb{R}^N$ *werden in einer Anlaufrechnung (z.B. mittels eines simplen ESV)ermittelt.*
- *Nach der Anlaufrechnung wird für jedes* $\ell \in \{0, 1, \ldots, n-m\}$ *mit den bereits bekannten Näherungen* $u_\ell, \ldots, u_{\ell+m-1} \in \mathbb{R}^N$ *und der Verfahrensvorschrift* (26) *die Näherung* $u_{\ell+m} \in \mathbb{R}^N$ *berechntet, mit dem Ziel, dass*

$$u_{l+m} \approx y(t_{l+m}).$$

- *Zur Vereinfachung der Notation wird im Folgenden der Definitionsbereich der Funktion* φ *immer wie in* (27) *angegeben. Bei den meisten m-Schrittverfahren ist der Ausdruck* $\varphi(t, v_0, \ldots, v_{m-1}; h)$ *aber nur für* $h \leq \frac{b-t}{m}$ *wohldefiniert.*
- *Hängt in der Verfahrensvorschrift* (26) *die rechte Seite von der Unbekannten* u_{l+m} *ab, so spricht man von einem* ***impliziten*** *m-Schrittverfahren, andernfalls von einem* ***expliziten*** *m-Schrittverfahren.*
- *Auf variablen Gittern sind m-Schrittverfahren von der Form*

$$\sum_{j=0}^{m} \alpha_j u_{\ell+j} = h_{\ell+m}\varphi(t_\ell, \ldots, t_{\ell+m}, u_\ell, \ldots, u_{\ell+m}; h_{\ell+m}), \qquad \ell = 0, \ldots, n-m.$$

- *Hat die Funktion φ in der Verfahrensvorschrift (26) die Form*

$$\varphi(t, v_0, \ldots, v_m; h) = \sum_{j=0}^{m} \beta_j f(t + jh, v_j),$$

*so wird (26) als **lineares** m-Schrittverfahren bezeichnet.*

Beispiel 7 *Ein spezielles lineares 2-Schrittverfahren ist die **Mittelpunktregel**,*

$$u_{\ell+2} = u_l + 2hf(t_{\ell+1}, u_{\ell+1}), \qquad \ell = 0, 1, \ldots, n-2.$$

2.1.1 Konvergenz- und Konsistenzordnung

Die Approximationseigenschaften eines Mehrschrittverfahrens werden durch seine Konvergenzordnung beschrieben.

Definition 7 *Ein Mehrschrittverfahren (26) zur Lösung des Anfangswertproblems $y' = f(t, y), \quad y(a) = y_0$ besitzt die **Konvergenzordnung** $p \geq 1$ falls sich zu jeder Konstanten $c \geq 0$ und beliebigen Startwerten $u_0, \ldots, u_{m-1} \in \mathbb{R}$ mit $\|u_k - y(t_k)\| \leq ch^p$ für $k = 0, 1, \ldots, m-1$ der **globale Verfahrensfehler** in der Form*

$$\max_{l=m,\ldots,n} \|u_\ell - y(t_\ell)\| \leq Kh^p$$

abschätzen lässt mit einer von der Schrittweite h unabhängigen Konstante $K \geq 0$

.

Definition 8 *Für ein Mehrschrittverfahren (26) zur Lösung des Anfangswertproblems $y' = f(t, y), \quad y(a) = y_0$ bezeichnet*

$$\eta(h, t) := \left[\sum_{j=0}^{m} \alpha_j y(t + jh)\right] - h\varphi\left(t, y(t), y(t+h), \ldots, y(t+mh); h\right) \qquad 0 < h \leq \tfrac{b-t}{m} \tag{29}$$

*den **lokalen Verfahrensfehler** im Punkt $(t, y(t))$ (bezüglich der Schrittweite h).*

Definition 9 *Ein Mehrschrittverfahren (26) zur Lösung des Anfangswertproblems $y' = f(t, y), \quad y(a) = y_0$ besitzt die **Konsistenzordnung** $p \geq 1$, falls gilt:*

$$\exists_{H>0} \exists_{C \geq 0} \quad \|\eta(t, h)\| \leq Ch^{p+1}, \quad a \leq t \leq b, \quad 0 \leq h \leq H \tag{30}$$

Die **Konsistenzordnung** wird oft nur kurz als **Ordnung** eines Mehrschrittverfahrens bezeichnet.

2.1.2 Nullstabilität, Lipschitzbedingung

Bei der Behandlung der Konvergenzordnung eines Mehrschrittverfahrens wird auch die folgende Lipschitzbedingung an die Funktion $\varphi : [a,b] \times (\mathbb{R}^N)^{m+1} \times \mathbb{R}_+ \longrightarrow \mathbb{R}^N$ aus der Verfahrensvorschrift (26) eine Rolle spielen:

$$\|\varphi(t, v_0, \ldots, v_m; h) - \varphi(t, w_0, \ldots, w_m; h)\| \leq L_\varphi \sum_{j=0}^{m} \|v_j - w_j\| \quad (v_j, w_j \in \mathbb{R}^N) \tag{31}$$

Bemerkung 9 *Falls $f : [a,b] \times \mathbb{R}^N \longrightarrow \mathbb{R}^N$ eine stetige Funktion ist, die die Lipschitzbedingung* (2) *erfüllt, so ist für lineare Mehrschrittverfahren die Lipschitzbedingung* (31) *erfüllt mit der speziellen Lipschitzkonstanten $L_\varphi = L \max_{j=0,\ldots,m} |\beta_j|$.*

Bezüglich der Existenz und Eindeutigkeit der Approximationen, welche ja nicht a priori gewährleistet sind, läßt sich folgender Hilfssatz formulieren:

Lemma 2.1 *Aus der Lipschitz-Stetigkeit der Verfahrensfunktion eines m-Schrittverfahrens im Sinne von* (31) *folgt die Existenz und Eindeutigkeit der Näherungen u_k für $m \leq\ k \leq n$*

Beweis:
Es sei zu gegebener Verfahrensfunktion $\varphi : [a,b] \times (\mathbb{R}^{\mathbb{N}})^m \times \mathbb{R}_{\geq 0} \to \mathbb{R}^N$, Schrittweite $h \geq 0$ und erzeugendem Polynom $(\alpha_j)_{j=0,\ldots,m}$ (siehe (32)) mit $\alpha_m \neq 0$ die Funktion Ξ_ℓ für $0 \leq \ell \leq n-m$ wie folgt definiert:

$$\Xi_\ell : \mathbb{R}^N \to \mathbb{R}^N; \quad x \mapsto \frac{1}{\alpha_m} \left(h\varphi(u_\ell, u_{\ell+1}, \ldots, u_{\ell+m-1}, x) - \sum_{j=0}^{m-1} \alpha_j u_{\ell+j} \right)$$

Mit der Lipschitzstetigkeit von φ gilt mit $0 \leq h < H < \frac{|\alpha_m|}{L_\varphi}$:

$$\begin{aligned} \forall_{h \leq H} \forall_{x_1, x_2 \in \mathbb{R}^N} & \|\Xi_\ell(x_1) - \Xi_\ell(x_2)\| \\ &= \frac{1}{|\alpha_m|} \|h \left(\varphi(u_\ell, u_{\ell+1}, \ldots, u_{\ell+m-1}, x_1) - \varphi(u_\ell, u_{\ell+1}, \ldots, u_{\ell+m-1}, x_2)\right) \| \\ &\leq \frac{L_\varphi h}{|\alpha_m|} \|x_1 - x_2\| \leq \frac{L_\varphi H}{|\alpha_m|} \|x_1 - x_2\| \end{aligned}$$

Ξ_ℓ erfüllt also eine Lipschitzbedingung mit der Konstanten $\kappa := \frac{L_\varphi H}{|\alpha_m|} < 1$, ist somit eine Kontraktion und besitzt als solche einen Fixpunkt $u_{m+\ell}$.

Definition 10 *Ein m-Schrittverfahren* (26) *zur Lösung des Anfangswertproblems $y' = f(t,y), \quad y(a) = y_0$ heißt* ***nullstabil****, falls das* ***erzeugende Polynom***

$$\rho(t) := \sum_{j=0}^{m} \alpha_j t^j \tag{32}$$

die folgende ***Dahlquistsche Wurzelbedingung*** *erfüllt:*

$$\begin{aligned} \rho(t_0) = 0 \quad &\Longrightarrow \quad |t_0| \leq 1; \\ \rho(t_0) = 0, \quad |t_0| = 1 \quad &\Longrightarrow \quad t_0 \textit{ ist einfache Nullstelle von } \rho. \end{aligned}$$

2.2 Ein Konvergenzkriterium

Ein paar vorbereitende Hilfssätze werden der Hauptaussage dieses Abschnittes vorausgeschickt.

Lemma 2.2 (Charakterisierung der Stabilität einer Matrix) *Für $A \in \mathbb{C}^{n\times n}$ sind folgende Aussagen äquivalent:*

1. *$(A^k)_{k\in\mathbb{N}}$ ist beschränkt ($\Leftrightarrow_{Df} A$ ist stabil)*
2. *Alle Nullstellen des Minimalpolynoms von A sind betragsmäßig kleiner oder gleich 1 und all seine Wurzeln mit Vielfachheit > 1 liegen in $K_1(0) := \{\xi \in \mathbb{C} \,|\, |\xi| < 1\}$*
3. *Alle Diagonalelemente der Jordan'schen Normalform von A sind betragsmäßig kleiner oder gleich 1 und alle Jordanblöcke der Normalform mit Diagonalelementen, die auf $\partial K_1(0)$ liegen, haben Dimension 1.*
4. *Es exisitiert eine invertierbare $n \times n$-Matrix S mit $\| S^{-1}AS \|_\infty \leq 1$.*

Beweis: Wir zeigen (1.) $\Rightarrow$ (2.) $\Rightarrow$ (3.) $\Rightarrow$ (4.) $\Rightarrow$ (1.). Angenommen es gelte $\neg$(2.). Im ersten Fall können wir annehmen, dass es ein $v \in \mathbb{R}^n$ gibt, sodass $Av = \lambda v$ mit $|\lambda| > 1$. Somit gilt

$$\|A^k\|\|v\| \geq \|A^k v\| = |\lambda|^k \|v\|,$$

woraus

$$\|A^k\| \geq |\lambda|^k \to \infty; \; k \to \infty$$

im Widerspruch zu (1.) folgt. Im zweiten Fall (alle Eigenwerte liegen innerhalb der abgeschlossenen Kreisscheibe mit Radius 1 um 0 in der komplexen Ebene) sei $j \geq 2$ die Vielfachheit einer Wurzel λ des Minimalpolynoms mit $|\lambda| = 1$. Wir wählen $0 \neq u \in \ker(A - \lambda\mathbb{E}_n)^j$ und setzen $w := (A - \lambda\mathbb{E}_n)^{j-2}u$ und $v := (A - \lambda\mathbb{E}_n)w$. Dann gilt

$$Aw = \lambda w + v \text{ mit} \quad Av = \lambda v.$$

Aus $A^k w = \lambda^k w + k\lambda^{k-1}v \quad k \in \mathbb{N}$ folgt schliesslich

$$\|kv + \lambda w\| = \|\lambda^k w + k\lambda^{k-1}v\| = \|A^k w\| \leq \|A^k\|\|w\|$$

und somit

$$\|A^k\| \geq \frac{k\|v\|}{\|w\|} - 1 \to \infty; \quad k \to \infty$$

Da die Diagonalelemente der Jordan-Normalform den Eigenwerten von A entsprechen und die größten Jordanblocks bezüglich einfacher Nullstellen des Minimalpolynoms von Dimension 1 sind, folgt (3.) aus (2.). Für (3.) $\Rightarrow$ (4.) wähle man für S die reguläre Matrix, die A in ihre Normalform überführt.

$$\|A^k\| \leq \|S\|\|S^{-1}AS\|\|S^{-1}\| \leq \|S\|M\|S^{-1}AS\|_\infty\|S^{-1}\| \leq \|S\|M\|S^{-1}\|$$

(Normäquivalenz in endlichdimensionalen Räumen) liefert die letzte Implikation.

Lemma 2.3 (Gronwall) *Sei eine stetige Funktion $g : I \to \mathbb{R}_{\geq 0}$ gegeben ($I \in \mathbb{R}$ ein nichtleeres Intervall) mit der Eigenschaft*

$$\exists_{A,B\in\mathbb{R}_{\geq 0}}\forall_{t_0,t\in I} \quad g(t) \leq A + B\left|\int_{t_0}^{t} g(\tau)d\tau\right|$$

Dann gilt:

$$\forall_{t,t_0\in I} \quad g(t) \leq A\exp(B|t-t_0|)$$

Der **Beweis** ist in einer Fülle von Lehrbüchern, so z.B. in [3], nachzulesen.
Nicht ohne Beweis bleiben soll der nachfolgende Hilfssatz, der eine Variante des Gronwall-Lemmas darstellt:

Lemma 2.4 *Seien $s \in \mathbb{N}_0$, $h_j > 0$, $0 \leq j \leq s-1$ und $v_j \in \mathbb{R}$ für $0 \leq j \leq s$ sowie $A, B \in \mathbb{R}\geq 0$ gegeben, und es gelte für alle $0 \leq j \leq s$ die Abschätzung*

$$|v_j| \leq A + B\sum_{\kappa=0}^{j-1} h_\kappa|v_\kappa|.$$

Dann gilt:

$$\forall_{0\leq j\leq s} \quad |v_j| \leq A\exp(B\sum_{\kappa=0}^{j-1} h_\kappa)$$

Beweis:
Zunächst die kleine

Definition 11 (charakteristische Funktion) *Sei $M \in \mathbb{R}^N$. Dann heißt*

$$\mathbf{1}_M : \mathbb{R}^N \to \mathbb{R}; \quad x \mapsto \begin{cases} 1 & \text{für } x \in M \\ 0 & \text{sonst} \end{cases}$$

die **charakteristische Funktion** *oder* **Indikatorfunktion** *von M.*

Wir definieren nun $t_0 := 0$ und für $1 \leq j \leq s$: $t_j := \sum_{\kappa=0}^{j-1} h_\kappa$ $I_j := [t_{j-1}, t_j)$. Weiters sei die Funktion

$$\vartheta := \sum_{\kappa=1}^{s} |v_{\kappa-1}| \mathbf{1}_{I_\kappa} + |v_s| \mathbf{1}_{\{t_s\}}$$

gegeben. Sei nun $t \in I := \left(\bigcup_{j=1}^{s} I_j\right) \cup \{t_s\}$. Dann gilt (da $\exists_{0 \leq j \leq s-1} \quad t \in I_{j+1}$ oder $t = t_s$):

$$\vartheta(t) = |v_j| \leq A + B \sum_{\kappa=0}^{j-1} h_\kappa |v_\kappa| = A + B \sum_{\kappa=0}^{j-1} \int_{t_\kappa}^{t_{\kappa+1}} \vartheta(\tau) d\tau$$

$$= A + B \int_0^{t_j} \vartheta(\tau) d\tau \leq A + B \int_0^{t} \vartheta(\tau) d\tau,$$

und nach Lemma 2.4 gilt:

$$|v_j| = \vartheta(t_j) \leq A \exp(B t_j) = A \exp(B \sum_{\kappa=0}^{j-1} h_\kappa),$$

was zu beweisen war.

Ein weiteres Resultat, diesmal aus der linearen Algebra, welches im Beweis des Hauptsatzes der Theorie allgemeiner m-Schrittverfahren eine maßgebliche Rolle spielt, betrifft die Form des charakteristischen Polynoms einer bestimmten Matrix.

Lemma 2.5 *Sei für ein $1 < m \in \mathbb{N}$ $(a_j)_{j=0,\ldots,m-1} \in \mathbb{R}^m$ gegeben, und die Matrix $A \in \mathbb{R}^{m \times m}$ sei definiert durch*

$$A := \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \ldots & & 0 & 1 \\ -a_0 & -a_1 & -a_2 & \ldots & -a_{m-1} \end{pmatrix}$$

Dann gilt:

$$\chi_A(t) := \det(t\mathbb{E}_m - A) = \pi_m(t) := t^m + \sum_{j=0}^{m-1} a_j t^j$$

Der **Beweis** ist einfach und erfolgt z.B. induktiv (nach m). Er sei hiermit dem interessierten Leser überlassen.

Nun kommen wir zu einem der wichtigsten Ergebnisse aus der Theorie der Mehrschrittverfahren, das uns erlaubt, aus der Konsistenz mit Ordnung p und der Nullstabilität des Verfahrens auf die Konvergenz mit ebenderselben Ordnung zu schließen. Diese Aussage stützt sich auf das nun angeführte

Lemma 2.6 *Für ein nullstabiles m-Schrittverfahren mit im Sinne von* (31) *lipschitzstetiger Verfahrensfunktion φ gilt:*

$$\exists_{K>0} \quad \max_{0\le j\le n} \|u_j - y(t_j)\| \le K \left(\max_{0\le j\le m-1} \|u_j - y(t_j)\| + \frac{\max_{a\le t\le b-hm} \|\eta(t,h)\|}{h|\alpha_m|} \right)$$

Beweis: Wir behandeln den Fall N=1. Die Maximumsnorm ermöglicht dann, durch Betrachtung der Komponentenfunktionen den allgemeinen Fall auf den speziellen zurückzuführen. Im Folgenden sei $\|\cdot\| = \|\cdot\|_\infty$.
Vorab ein paar Definitionen ($h = \frac{b-a}{n}$):

$$\forall_{0\le j\le n} \quad y_j := y(t_j) = y(a+jh) \quad e_j := u_j - y_j$$

$$\forall_{0\le j\le n-m} \quad \eta_j := \eta(t_j,h) = \left(\sum_{k=0}^{m} \alpha_k y(t_j+kh) \right) - h\varphi(t_j, y(t_j), y(t_j+h), \ldots, y(t_j+mh), h)$$

Damit gilt:

$$\sum_{k=0}^{m} \alpha_k e_{k+j} = \overbrace{h\left(\varphi(t_j,u_j,u_{j+1},\ldots,u_{j+m},h) - \varphi(t_j,y(t_j),y(t_j+h),\ldots,y(t_j+mh),h)\right)}^{\omega_j} - \eta_j \tag{33}$$

Dies läßt sich in folgende Form bringen:

$$\underbrace{\begin{pmatrix} e_{j+1} \\ e_{j+2} \\ \vdots \\ \vdots \\ e_{j+m} \end{pmatrix}}_{\mathcal{E}_{j+1}} = \underbrace{\begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \ldots & & 0 & 1 \\ -\frac{\alpha_0}{\alpha_m} & -\frac{\alpha_1}{\alpha_m} & -\frac{\alpha_2}{\alpha_m} & \ldots & -\frac{\alpha_{m-1}}{\alpha_m} \end{pmatrix}}_{=:A} \underbrace{\begin{pmatrix} e_j \\ e_{j+1} \\ \vdots \\ \vdots \\ e_{j+m-1} \end{pmatrix}}_{\mathcal{E}_j} + \underbrace{\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \frac{\omega_j - \eta_j}{\alpha_m} \end{pmatrix}}_{\mathcal{F}_j}$$

Nun gilt aber nach Lemma 2.5:

$$\chi_A(t) = \pi_m(t) = t^m + \frac{1}{\alpha_m} \sum_{k=0}^{m-1} \alpha_k t^k,$$

was zur Folge hat, dass die Eigenwerte von A den Nullstellen des erzeugenden Polynoms ρ entsprechen. Laut Voraussetzung genügt dieses jedoch der Dahlquistschen Bedingung, womit sich auf A Lemma 2.2 anwenden läßt. Wir können also $M > 0$ als obere Schranke für $\{\|A^j\| | j \in \mathbb{N}_0\}$ annehmen. Weiters gilt:

$$\mathcal{E}_j = A^j \mathcal{E}_0 + \sum_{k=0}^{j-1} A^{j-(k+1)} \mathcal{F}_k, \quad 0 \leq j \leq n - m + 1 \tag{34}$$

welches durch Induktion nach j bewiesen wird: Für $j = 0$ ist nichts zu beweisen. Sei also (34) für ein beliebiges $0 \leq j < n - m$ bewiesen. Dann gilt:

$$\mathcal{E}_{j+1} = A\mathcal{E}_j + \mathcal{F}_j =_{\text{IV}} A\left(A^j \mathcal{E}_0 + \sum_{k=0}^{j-1} A^{j-(k+1)} \mathcal{F}_k\right) + \mathcal{F}_j = A^{j+1}\mathcal{E}_0 + \sum_{k=0}^{j} A^{j-k} \mathcal{F}_k$$

Was wiederum die Abschätzung

$$\|\mathcal{E}_j\| \leq M\left(\|\mathcal{E}_0\| + \sum_{k=0}^{j-1} \|\mathcal{F}_k\|\right) \tag{35}$$

bedingt. Bleibt noch der zweite Summand der rechten Seite von (35) abzuschätzen:

$$\begin{aligned} |\alpha_m| \|\mathcal{F}_k\| = |\eta_k - \omega_k| &\leq \max_{0 \leq k \leq m-n} |\eta_k| + hL_\varphi \sum_{\kappa=0}^{m} |e_{k+\kappa}| \\ &\leq \max_{0 \leq k \leq m-n} |\eta_k| + mhL_\varphi \|\mathcal{E}_k\| + hL_\varphi \|\mathcal{E}_{k+1}\|, \end{aligned}$$

was durch Summation auf

$$\begin{aligned} |\alpha_m| \sum_{k=0}^{j-1} \|\mathcal{F}_k\| &\leq j \max_{0 \leq k \leq m-n} |\eta_k| + mhL_\varphi \sum_{k=0}^{j-1} \|\mathcal{E}_k\| + hL_\varphi \left(\sum_{k=0}^{j-1} \|\mathcal{E}_k\| + \|\mathcal{E}_j\|\right) \\ &\leq n \max_{0 \leq k \leq m-n} |\eta_k| + h \underbrace{(m+1)L_\varphi}_{=:\gamma} \sum_{k=0}^{j-1} \|\mathcal{E}_k\| + hL_\varphi \|\mathcal{E}_j\| \end{aligned}$$

führt. Nun gilt mit (35) und $h < H \in \mathbb{R}_{>0}$, wobei $H < \frac{|\alpha_m|}{ML_\varphi}$ (o.B.d.A $L_\varphi > 0$):

$$\begin{aligned} \|\mathcal{E}_j\| &\leq M\left(\|\mathcal{E}_0\| + \sum_{k=0}^{j-1} \|\mathcal{F}_k\|\right) \\ &\leq M\left(\|\mathcal{E}_0\| + \frac{1}{|\alpha_m|}\left[n \max_{0 \leq k \leq m-n} |\eta_k| + h\gamma \sum_{k=0}^{j-1} \|\mathcal{E}_k\| + hL_\varphi \|\mathcal{E}_j\|\right]\right) \\ \Leftrightarrow (1 - \frac{hML_\varphi}{|\alpha_m|}) \|\mathcal{E}_j\| &\leq M\left(\|\mathcal{E}_0\| + \frac{1}{|\alpha_m|}\left[n \max_{0 \leq k \leq m-n} |\eta_k| + h\gamma \sum_{k=0}^{j-1} \|\mathcal{E}_k\|\right]\right) \\ \text{und } 0 < \underbrace{(1 - \frac{HML_\varphi}{|\alpha_m|})}_{=:\zeta} &\leq (1 - \frac{hML_\varphi}{|\alpha_m|}) \end{aligned}$$

Folglich gilt:

$$\|\mathcal{E}_j\| \leq \frac{M}{\zeta}\left(\|\mathcal{E}_0\| + \frac{n}{|\alpha_m|}\max_{0\leq k\leq m-n}|\eta_k|\right) + \frac{Mh}{\zeta}\gamma\sum_{k=0}^{j-1}\|\mathcal{E}_k\|,$$

woraus mit Lemma 2.4 folgt:

$$\begin{aligned}\|\mathcal{E}_j\| &\leq \frac{M}{\zeta}\left(\|\mathcal{E}_0\| + \frac{n}{|\alpha_m|}\max_{0\leq k\leq m-n}|\eta_k|\right)\exp\left(\frac{M(b-a)}{\zeta}\gamma\right)\\ &= \frac{M}{\zeta}\left(\|\mathcal{E}_0\| + \frac{(b-a)}{h|\alpha_m|}\max_{0\leq k\leq m-n}|\eta_k|\right)\exp\left(\frac{M(b-a)}{\zeta}\gamma\right)\\ &\leq \frac{M}{\zeta}\left(\|\mathcal{E}_0\| + \frac{1}{h|\alpha_m|}\max_{a\leq t\leq b-hm}|\eta(t,h)|\right)\exp\left((b-a)\left[1+\frac{M\gamma}{\zeta}\right]\right)\end{aligned}$$

Unter Beachtung der Definition von $\mathcal{E}_j$, $\quad 0 \leq j \leq n-m+1$ ist die Behauptung des Lemmas damit vollständig bewiesen.

Satz 2.1 (Korollar zu Lemma 2.6) *Ein nullstabiles m-Schrittverfahren mit im Sinne von* (31) *lipschitzstetiger Verfahrensfunktion und Konsistenzordnung $p \in \mathbb{R}_{\geq 1}$ ist konvergent von derselben Ordnung.*

2.3 Konsistenz linearer Mehrschrittverfahren

in diesem Abschnitt wollen wir ein einfaches Konsistenzkriterium für lineare m-Schrittverfahren angeben, das sich einerseits auf Differenzierbarkeitseigenschaften von f, andererseits auf bestimmte Relationen zwischen den Koeffizienten der erzegenden Polynome $\sum_{j=0}^{m}\alpha_j x^j$ und $\sum_{j=0}^{m}\beta_j x^j$ stützt.

Lemma 2.7 *Falls zu einem gegebenem linearen m-Schrittverfahren*

$$\sum_{k=0}^{m}\alpha_k u_{l+k} = \sum_{k=0}^{m}\beta_k f(t_l + kh, u_{l+k}), \quad 0 \leq l \leq n-m$$

und $p \in \mathbb{N}$ *die* $p+1$ *Gleichungen*

$$\sum_{k=0}^{m}(k^\nu\alpha_k - \nu k^{\nu-1}\beta_k) = 0, \quad 0 \leq \nu \leq p$$

erfüllt sind, so ist das Verfahren konsistent mit Ordnung p.

Der Beweis verwendet, wenig überraschend, das essentielle Hilfsmittel der numerischen Approximation gew. Differentialgleichungen: die Taylorentwicklung der exakten Lösung y. Da y als Lösung von $y'(t) = f(t, y(t)); y(a) = y_0 \quad p+1$ mal differenzierbar ist, haben folgende Talorentwicklungen in $t \in [a, b]$ Gültigkeit:

$$y(t+kh) = \sum_{l=0}^{p} \frac{y^{(\nu)}(t)}{\nu!} k^\nu h^\nu + \mathcal{O}(h^{p+1}),$$

$$y'(t+kh) = \sum_{\nu=0}^{p-1} \frac{y^{(\nu+1)}(t)}{\nu!} k^\nu h^\nu + \mathcal{O}(h^p) = \sum_{\nu=0}^{p} \nu \frac{y^{(\nu)}(t)}{\nu!} k^{(\nu-1)} h^{(\nu-1)} + \mathcal{O}(h^p)$$

Nun gilt für den lokalen Verfahrensfehler:

$$\eta(t,h) = \sum_{k=0}^{m} \alpha_k y(t+kh) - \sum_{k=0}^{m} \beta_k f(t+kh, y(t+kh)) = \sum_{k=0}^{m} \alpha_k y(t+kh) - h \sum_{k=0}^{m} \beta_k y'(t+kh) =$$

$$\sum_{\nu=0}^{p} h^\nu \frac{y^{(\nu)}(t)}{\nu!} \underbrace{\sum_{k=0}^{m} (\alpha_k k^\nu - \nu \beta_k k^{(\nu-1)})}_{=0} + \mathcal{O}(h^{p+1}),$$

was den Beweis vervollständigt.

2.4 Ausblick auf spezielle Mehrschrittverfahren

In diesem kurz gehaltenen Abschnitt wird versucht, einen kleinen Einblick in eine Idee zu verschaffen, der eine Vielzahl spezieller m-Schrittverfahren zugrundeliegt. Man bedient sich bei der Konstruktion der erwähnten Verfahren der folgenden Eigenschaft der exakten Lösung y des Anfangswertproblemes (1):

$$\forall_{t,t' \in [a,b]} \quad y(t) - y(t') = \int_{t'}^{t} y'(\tau) d\tau = \int_{t'}^{t} f(\tau, y(\tau)) d\tau \tag{36}$$

Man versucht nun, im j-ten Schritt des Verfahrens durch geeignete Polynome $\mathcal{P}_j$ den Verlauf von $f(t, y(t))$ anzunähern und definiert schliesslich:

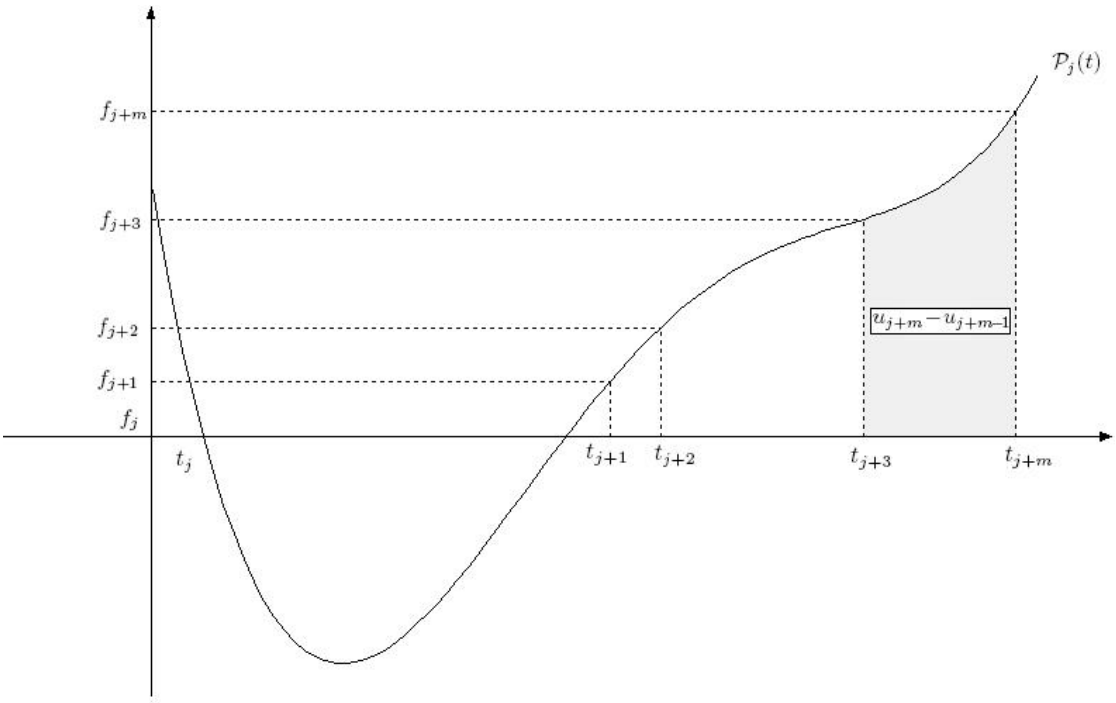
$$u_{j+m} - u_{j+m-1} := \int_{t_{j+m-1}}^{t_{j+m}} \mathcal{P}_j \tag{37}$$

Es bezeichne $\Pi_m^N := \left\{ f : \mathbb{R} \to \mathbb{R}^N | \exists_{a_0, a_1, \ldots, a_m \in \mathbb{R}^N} \forall_{t \in \mathbb{R}} \quad f(t) = \sum_{\nu=0}^{m} a_\nu t^\nu \right\}$.

Beispiel 8 (Adams-Moulton-Verfahren) *Für $0 \leq j \leq n-m$, $0 \leq k \leq m$ setze man $f_{j+k} := f(t_{j+k}, u_{j+k})$. $\mathcal{P}_j \in \Pi_m^N$ sei gegeben durch*

$$\forall_{0 \leq k \leq m} \quad \mathcal{P}_j(t_{j+k}) = f_{j+k}.$$

Anwendung von (37) *liefert dann das* ***Adams-Moulton-Verfahren*** *m-ter Klasse. Hierzu eine Illustration:*



Das Adams-Moulton Verfahren

Mit ein wenig Interpolationstheorie und sonstigem technischen Aufwand läßt sich folgender Satz beweisen:

Satz 2.2 *Das Adams-Moulton-Verfahren m-ter Klasse ist linear, nullstabil und im Falle der (m+1)-fachen stetigen Differenzierbarkeit von f konsistent mit Ordnung $m+1$.*

Der Beweis findet sich z.B. in [1], wo auch ähnlich strukturierte Verfahren unterschiedlicher Konsistenzordung vorgestellt werden.

Literatur

[1] Plato, R. : Numerische Mathematik kompakt, Vieweg 2004

[2] Deuflhard, P./ Bornemann, F. : Numerische Mathematik, Band 2, Teubner 1994

[3] Königsberger, K. : Analysis 2, Springer 2004

[4] Butcher, J.C. : Numerical Methods for Ordinary Differential Equations, Wiley 2003