Numerische Analysis

Vorlesung vom WS 2010/11

Mario Ohlberger

Institut für Numerische und Angewandte Mathematik
Fachbereich Mathematik und Informatik
Westfälische Wilhelms-Universität Münster

Dieses Skript beruht auf meinen Vorlesungen *Einführung in die Numerische Mathematik* und *Höhere Numerische Mathematik* vom Wintersemester 2007/2008 und Sommersemester 2008 an der Westfälische Wilhelms-Universität Münster.
Gegenüber den vorherigen Vorlesungen hat sich insbesondere die Zusammenstellung der Lehrinhalte geändert. Die Teile zu Interpolation und numerischer Quadratur bauen auf der Vorlesung vom WS 2007/2008 auf, während das Kapitel zur Numerik gewöhnlicher Differentialgleichungen auf der Vorlesung vom SS 2008 basiert.

Es besteht keine Garantie auf Richtigkeit und/oder Vollständigkeit des Manuskripts.

Mario Ohlberger

Inhaltsverzeichnis

Abbildungsverzeichnis

Kapitel 0

Einleitung

Das Ziel dieser Vorlesung ist eine Einführung in die numerische Analysis anhand von Differentialgleichungsproblemen. Zentrale Themen werden sein, die Interpolation von Funktionen, die numerische Integration und die numerische Differentiation im Zusammenhang mit der numerischen Behandlung von Rand- und Anfangswertproblemen.

Um das Zusammenspiel dieser Themen zu veranschaulichen, wollen wir anhand von Variationsproblemen motivieren, wie diese Bereiche der numerischen Mathematik zur Lösung konkreter Fragestellungen ineinandergreifen.

Variationsgleichungen und Galerkinapproximation

In der Physik, aber auch in anderen Anwendungsbereichen, spielt das Prinzip der *Energieminimierung* eine wichtige Rolle. Dieses Prinzip ermöglicht uns z.B. die Beschreibung des Verhaltens elastischer Körper. Dazu sei $u(x,t) \in \mathbb{R}^d$ die Auslenkung/der Verschiebevektor eines elastischen Körpers im Punkt x zum Zeitpunkt t und $\varepsilon(u) := \frac{1}{2}(\nabla u + \nabla u^\top)$ der Verzerrungstensor. Dann ist die potentielle Gesamtenergie eines belasteten, elastischen Körpers gegeben durch

$$E(u) = \frac{1}{2}\int_\Omega \sigma : \varepsilon(u)\mathrm{d}x - \int_\Omega f u \mathrm{d}x.$$

Dabei ist σ der symmetrische Spannungstensor und f eine äußere Volumenkraft. Für idealisierte Materialien ist der Spannungstensor proportional zum Verzerrungstensor, d.h. es gilt das lineare Materialgesetz

$$\sigma(u) = A\varepsilon(u)$$

mit dem symmetrischen und positiv definiten Elastizitätstensor A. Zur Modellierung elastischer Körper mit partiellen Differentialgleichungen wenden wir auf die Energie $E(u)$ das folgende Minimierungsprinzip an.

Definition 0.1 (Energieminimierung/Variationsprinzip)
a) Physikalisches Prinzip *Ein physikalisches System strebt immer in den Zustand minimaler Energie.*
b) Mathematische Äquivalenz: *Sei $\bar{u}(x,t)$ eine Zustandsvariable und $E(\bar{u})$ die Energie eines Systems in Abhängigkeit von $\bar{u}$. Dann strebt $\bar{u}$ gegen einen optimalen Zustand $u = u(x)$, der die Energie minimiert, d.h. falls E genügend glatt ist gilt*

$$\frac{d}{d\varepsilon} E(u + \varepsilon\varphi)\Big|_{\varepsilon=0} = 0$$

für beliebige zulässige Variationen φ.

Berechnen wir

$$\frac{d}{d\varepsilon} E(u + \varepsilon\varphi)|_{\varepsilon=0} = 0$$

für die potentielle Energie elastischer Körper, so erhalten wir die Gleichung

$$\int_\Omega A\varepsilon(u) : \varepsilon(\varphi)\mathrm{d}x = \int_\Omega f\varphi\mathrm{d}x$$

für alle zulässigen Variationen φ. Nehmen wir an, dass u genügend oft differenzierbar ist, so folgt mit partieller Integration für Testfunktionen (Variationen) φ, die auf dem Rand von Ω verschwinden:

$$\int_\Omega \Big(- \nabla\cdot(A\varepsilon(u)) - f\Big)\varphi dx = 0.$$

Mit dem Hauptsatz der Variationsrechnung folgt somit die Differentialgleichung

$$-\nabla\cdot(A\varepsilon(u)) = f.$$

oder in ausführlicher Notation

$$-\sum_{i=1}^{d}\sum_{k,l=1}^{d} \partial_{x_i} A_{ijkl}\varepsilon(u)_{kl} = f_i, \qquad \forall j = 1,\dots,d.$$

Im einfachsten eindimensionalen Fall, der Auslenkung eines dünnen Drahtes, reduziert sich diese Differentialgleichung zu

$$-\partial_x(A\partial_x u) = f, \quad \text{b.z.w.} - \partial_x^2 u = f,$$

für $A = Id$. Im einfachsten mehrdimensionalen Fall, der Auslenkung einer dünnen Membran, erhalten wir

$$-\nabla\cdot(A\nabla u) = f, \quad \text{b.z.w.} - \Delta u = f,$$

für $A = Id$. Wir erhalten also zur Beschreibung der Auslenkung elastischer Körper in Spezialfällen die Poisson-Gleichung.

Galerkin Verfahren

Idee: Energieminimierung in endlich dimensionalen Teilräumen

Die Idee der Galerkin Verfahren beruht auf dem Prinzip der Energieminimierung. Sei also V ein Funktionenraum und $E : V \to \mathbb{R}$ ein Energiefunktional. In Abschnitt 1.1 hatten wir gesehen, dass die Lösung u vieler physikalischer Fragestellungen formal wie folgt definiert werden können:

$$u = \arg\min_{v \in V} E(v).$$

Aus diesem kontinuierlichen Minimierungsproblem erhält man ein diskretes Minimierungsproblem, wenn man den Funktionenraum V durch einen endlichdimensionalen Teilraum $V_h \subset V$ ersetzt. Somit lässt sich die Lösung $u_h \in V_h$ eines allgemeinen Galerkin Verfahrens wie folgt schreiben

$$u_h = \arg\min_{v_h \in V_h} E(v_h).$$

Analog zum kontinuierlichen Variationsprinzip, ist die diskrete Lösung u_h dann charakterisiert durch

$$\frac{d}{d\varepsilon} E(u_h + \varepsilon v_h)\Big|_{\varepsilon=0} = 0, \qquad \text{für alle } v_h \in V_h,$$

falls E genügend regulär ist. Da V_h endlichdimensional ist können wir uns bei dieser Variation auf eine Basis von V_h einschränken. Sei also $\Phi := \{\varphi_i | i = 1, \cdot, N$ eine Basis von V_h, d.h. $N = \dim(V_h)$ und $u_h = \sum_{i=1}^N u_i \varphi_i$ die zugehörige Basisdarstellung von u_h, so folgt

$$\frac{d}{d\varepsilon} E\Big(\sum_{i=1}^N u_i \varphi_i + \varepsilon \varphi_j\Big)\Big|_{\varepsilon=0} = 0, \qquad \text{für alle } j = 1, \dots, N.$$

Dies ist ein System von N linearen oder nichtlinearen Gleichungen für die Unbekannten $u_i, i = 1, \dots, N$, das mit numerischen Methoden gelöst werden kann.
Galerkinapproximationen erhält man aus diesem allgemeinen Prinzip durch spezielle Wahl des endlichdimensionalen Teilraums V_h und zugehöriger Basis Φ mit der Eigenschaft, dass

$$\frac{d}{d\varepsilon} E\Big(\varphi_i + \varepsilon \varphi_j\Big)\Big|_{\varepsilon=0} \neq 0$$

nur für eine kleine Anzahl von Paaren (i, j) gilt.

Beispiel in einer Raumdimension

Wir betrachten das Energiefunktional eines elastischen Drahtes in einer Raumdimensuion, dass in der einfachsten Form gegeben ist durch:

$$E(u) := \int_0^1 \frac{1}{2} \partial_x u^2 + f u.$$

Ein Minimierer dieses Funktionals erfüllt

$$\frac{d}{d\varepsilon} E\Big(\partial_x u + \varepsilon \partial_x \varphi\Big)\Big|_{\varepsilon=0} = 0.$$

Wir berechnen

$$\begin{aligned} \frac{d}{d\varepsilon} E\Big(\partial_x u + \varepsilon \varphi\Big) &= \frac{d}{d\varepsilon} \int_0^1 \frac{1}{2} (\partial_x u + \varepsilon \partial_x \varphi)^2 + f(u + \varepsilon \varphi) \\ &= \frac{d}{d\varepsilon} \int_0^1 (\partial_x u + \varepsilon \partial_x \varphi) \partial_x \varphi + f \varphi. \end{aligned}$$

Also folgt

$$0 = \frac{d}{d\varepsilon} E\Big(u + \varepsilon \varphi\Big)\Big|_{\varepsilon=0} = \frac{d}{d\varepsilon} \int_0^1 \partial_x u \partial_x \varphi + f \varphi.$$

Bezeichnen wir mit $(\cdot,\cdot)$ das L^2-Skalarprodukt auf dem Intervall $(0,1)$, so lautet unser Problem also: Finde eine Funktion $u \in V$, so dass für alle Testfunktionen $\varphi \in V$ gilt

$$(\partial_x u, \partial_x \varphi) = (f, \varphi). \tag{1}$$

Dabei sei V ein geeigneter Funktionenraum. Sei nun $V_h \subset V$ ein endlichdimensionaler Teilraum von V der Dimension $\dim(V) = N$, dann ist eine endlichdimensionale Approximation u_h von u gegeben durch :

Finde eine Funktion $u_h \in V_h$, so dass für alle Testfunktionen $\varphi_h \in_h$ gilt

$$(\partial_x u_h, \partial_x \varphi_h) = (f, \varphi_h). \tag{2}$$

Ist $(\varphi_1, \ldots, \varphi_N)$ eine Basis von V_h, so können wir die Approximation u_h in dieser Basis darstellen, d.h.

$$u_h = \sum_{i=1}^{N} u_i \varphi_i.$$

Dabei sind u_i die Koeffizienten der Basisdarstellung von u_h. Da es in der Variationsformulierung (2) ausreicht mit den Basisfunktionen von V_h zu testen, folgt mit der Basisdarstellung von u_h für alle $j = 1, \ldots, N$:

$$\begin{aligned} & (\partial_x u_h, \partial_x \varphi_j) = (f, \varphi_j), \\ \Longrightarrow \quad & \textstyle\sum_{i=1}^{N} u_i (\partial_x \varphi_i, \partial_x \varphi_j) = (f, \varphi_j). \end{aligned}$$

Definieren wir die Matrix S und Vektoren U, F durch

$$\begin{aligned} S_{ji} &:= (\partial_x \varphi_i, \partial_x \varphi_j), \\ U_i &:= u_i, \\ F_j &:= (f, \varphi_j), \end{aligned}$$

so ist das discrete Problem (2) also äquivalent zur Lösung des linearen Gleichungssystems

$$SU = F.$$

Wir sehen an diesem Beispiel, dass wir zur numerischen Lösung des Variationsproblem insbesondere numerische Methoden zur Berechnung der Integrale $(\partial_x u_h, \partial_x \varphi_j)$, b.z.w. (f, φ_j) benötigen.
Integrale beliebiger Funktionen können in der Regel nicht in geschlossener Form angegeben werden. Andererseits ist es sehr einfach Polynome analytisch zu integrieren. Eine Möglichkeit beliebige Integrale zu approximieren besteht daher darin, diese Funktionen zunächst durch Polynome zu approximieren und diese approximierenden Polynome dann zu integrieren.
Dies motiviert, dass wir uns in Kapitel 1 zunächst mit der Polynominterpolation von Funktionen beschäftigen, um anschließend in Kapitel 2 die darauf aufbauenden Methoden zur numerischen Integration zu studieren. Mit diesen Vorbereitungen werden wir uns dann in Kapitel 3 mit der numerischen Lösung von Anfangswertproblemen und Randwertproblemen beschäftigen. Zur Approximation von Randwertproblemen, werden wir die gerade vorgestellte Methode wieder aufgreifen.

Kapitel 1

Interpolation

Sei $\{\Phi(x, a_0, \ldots, a_n) \mid a_0, \ldots, a_n \in \mathbb{R}\}$ eine Familie von Funktionen mit $x \in \mathbb{R}$, deren Elemente durch $(n+1)$ Parameter $a_0, \ldots, a_n \in \mathbb{R}$ gegeben sind.

Aufgabe: Zu $(x_k, f_k) \in \mathbb{R}^2,\ k = 0, \ldots, n$ mit $x_i \neq x_k$ für $i \neq k$, finde Parameter $a_0, \ldots, a_n$, so dass

$$\Phi(x_k, a_0, \ldots, a_n) = f_k \text{ für } k = 0, \ldots, n.$$

Falls Φ linear von seinen Parametern abhängt, spricht man von einem **linearen Interpolationsproblem**. (x_k, f_k) sind zum Beispiel Messdaten oder diskrete Werte aus einem anderen numerischen Verfahren, oder $f_k = f(x_k)$ für eine Funktion $f \in C^0$ und $x_0, \ldots, x_n$ sind die **Stützstellen**, an denen f interpoliert werden soll.

Beispiel:(Polynominterpolation)
$V \subset C^0(\mathbb{R})$ Vektorraum und $\dim(V) = n+1,\ \varphi_0, \ldots, \varphi_n$ Basis von V.
Setze $\Phi(x, a_0, \ldots, a_n) := \sum_{k=0}^{n} a_k \varphi_k(x)$.
Sei etwa $V = \mathbb{P}_n,\ \varphi_n(x) = x^k$, d.h. $\Phi(x, a_0, \ldots, a_n) = \sum_{k=0}^{n} a_k x^k$.
Problem: Finde ein Polynom höchstens n-ten Grades, so dass $p(x_k) = f_k$.

Weitere wichtige Beispiele:

Trigonomerische Interpolation

$$\Phi(x, a_0, \ldots, a_n) = a_0 + a_1 e^{ix} + a_2 e^{2ix} + \cdots + a_n e^{nix} = a_0 + \sum_{k=1}^{k} a_k(\cos(kx) + i \cdot \sin(kx))$$

Exponentielle Interpolation (nicht linear)

$$\Phi(x, a_0, \ldots, a_n, \lambda_0, \ldots, \lambda_n) = a_0 e^{\lambda_0 x} + \cdots + a_n e^{\lambda_n x}$$

Rationale Interpolation (nicht linear)

$$\Phi(x, a_0, \ldots, a_n, b_0, \ldots, b_m) = \frac{a_0 + a_1 x + \cdots + a_n x^n}{b_0 + b_1 x + \cdots + b_m x^m}$$

Erweitertes Problem: Hermite-Interpolation

Es werden nicht nur Funktionswerte f_k an den Stützstellen x_k vorgeschrieben, sondern auch Werte für die Ableitung von Φ.

Beispiel: $p(x) = \sum_{k=0}^{N} a_k x^k$, $N = 2(n+1)-1$ mit $p(x_k) = f_k,\ p'(x_k) = d_k$ für gegebene $(x_k, f_k, d_k) \in \mathbb{R}^3$ $(k = 0, \ldots, n)$.

Spline-Interpolation

Sei $q \in \mathbb{N}$ fest.

Gesucht: $\Phi \in C^q$ mit $\Phi(x_k) = f_k$ und $\Phi_{|_{[x_k, x_{k+1}]}} \in \mathbb{P}_r$, d.h. Φ ist stücksweise polynomial.

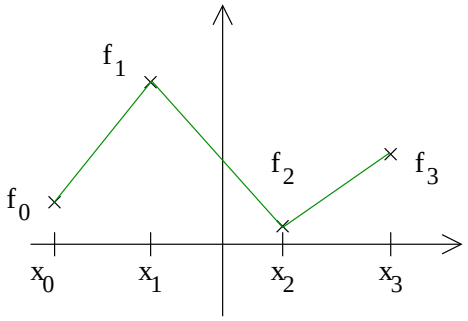


Abbildung 1.1: Spline-Interpolation

Abbildung 1.1 zeigt das Problem für $q = 0$ und $r = 1$, d.h. $\Phi \notin C^1$.

1.1 Polynominterpolation

Gegeben: $(x_0, f_0), \ldots, (x_n, f_n) \in \mathbb{R}^2$ mit $x_k \neq x_i$ $(k \neq i)$.

Gesucht: $p \in \mathbb{P}_N$ mit $p(x_i) = f_i$ $(i = 0, \ldots, n)$ mit N minimal.

Beispiel: $(x_0, f_0) = (0, 0),\ (x_1, f_1) = (1, 1)$

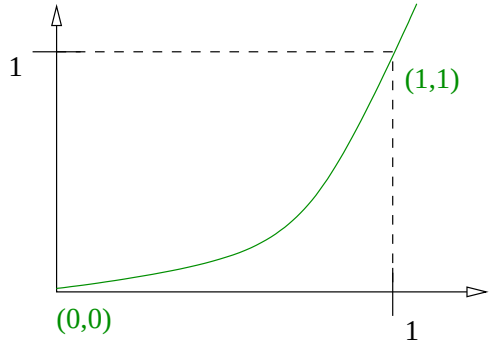


Abbildung 1.2: Polynominterpolation, Beispiel 1

Abbildung 1.2 verdeutlich das Beispiel, denn $p(x) = x^N$ erfüllt das Interpolationsproblem für alle $N \geq 1$. Gesuchtes Polynom ist dann: $p(x) = x$.

Satz 1.1
Es existiert genau ein $p \in \mathbb{P}_n$ mit $p(x_i) = f_i$ $(i = 0, \ldots, n)$.

Beweis: Sei $\varphi_0, \ldots, \varphi_n$ eine Basis von $\mathbb{P}_n$. Dann ist das Interpolationsproblem äquivalent dazu, einen Vektor $a = (a_0, \ldots, a_n)^\top$ zu finden, welcher das LGS $Aa = f$ löst mit $A = (\alpha_{ik}) \in \mathbb{R}^{(n+1)\times(n+1)}$, $\alpha_{ik} = \varphi_k(x_i)$ und $f = (f_0, \ldots, f_n)^\top \in \mathbb{R}^{n+1}$

$$\begin{aligned} \text{Es gilt : } Aa = f \quad &\iff \quad \sum_{k=0}^{n} \alpha_{ik} a_k = f \\ &\iff \quad \sum_{k=0}^{n} a_k \varphi_k(x_i) = f_i \\ &\iff \quad p(x_i) = f_i \text{ mit } p(x) = \sum_{k=0}^{n} a_k \varphi_k(x) \end{aligned}$$

Existenz und Eindeutigkeit: Es reicht zu zeigen, dass A regulär ist.

Sei also $a = (a_0, \ldots, a_n)^\top$ Lösung von $\sum_{k=0}^{n} a_k \varphi_k(x_i) = 0$ $(i = 0, \ldots, n)$. Dann hat $p(x) = \sum_{k=0}^{n} a_k \varphi_k(x) \in P_n$ die $(n+1)$-Nullstellen $x_0, \ldots, x_n$. Folglich muß $p \equiv 0$ sein, und somit $a_0 = \cdots = a_n = 0$.

Also gilt $Aa = 0 \implies a = 0 \implies A$ injektiv $\implies A$ regulär.
Also ist das Interpolationsproblem eindeutig lösbar. □

Bemerkung: Der Beweis von Satz 1.1 erlaubt es, Verfahren zur Lösung des Interpolationsproblems zu konstruieren. Dazu muss man eine Basis $\varphi_0, \ldots, \varphi_n$ von $\mathbb{P}_n$ wählen und das $(n+1) \times (n+1)$ LGS $Aa = f$ lösen.
Wählt man die Monombasis $\varphi_0(x) = 1, \varphi_1(x) = x, \varphi_2(x) = x^2, \ldots, \varphi_n(x) = x^n$ (also $\varphi_i(x) = x^i$), so gilt $p(x) = \sum_{i=0}^{n} \alpha_i \varphi_i(x)$ und es entsteht folgende Matrix:

$$A = \begin{pmatrix} \varphi_0(x_0) & \cdots & \varphi_n(x_0) \\ \vdots & \ddots & \vdots \\ \varphi_0(x_n) & \cdots & \varphi_n(x_n) \end{pmatrix} = \begin{pmatrix} 1 & x_0 & \cdots & x_0^n \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^n \end{pmatrix} \in \mathbb{R}^{(n+1)\times(n+1)}$$

A heißt die **Vandermondsche Matrix**; sie ist sehr schlecht konditioniert und voll besetzt. Daher ist das LGS $Aa = f$ sehr aufwändig zu lösen.

Die Darstellung des Interpolationspolynoms $\sum_{k=0}^{n} a_k x^k$ heißt **Normalform**. Diese Darstellung wird in der Praxis nicht verwendet. Andere Darstellungen sind numerisch effizienter, auch wenn dadurch das Interpolationspolynom nicht verändert wird!

a) **Lagrange-Form des Interpolationsproblems**

Am einfachsten ist $Aa = f$ zu lösen, falls $A = \mathbb{I}$, d.h. $\varphi_k(x_i) = \delta_{ki}$ $(0 \le k, i \le n)$. Wir erhalten mit $\varphi_k(x_i) = 0$ für $k \neq i$ den Ansatz

$$\varphi_k(x) = c \prod_{\substack{i=0 \\ i\neq k}}^{n} (x - x_i).$$

Aus $\varphi_k(x_k) = 1$ folgt dann

$$c = \left(\prod_{\substack{i=0 \\ i\neq k}}^{n} (x_k - x_i) \right)^{-1}$$

und somit erhalten wir

$$\varphi_k(x) = \prod_{\substack{i=0 \\ i\neq k}}^{n} \frac{(x - x_i)}{x_k - x_i}, \ (k = 0, \ldots, n).$$

Definition 1.2 (Lagrange-Polynome)
Die Polynome

$$l_k^n(x) := \prod_{\substack{i=0 \\ i\neq k}}^{n} \frac{(x - x_i)}{x_k - x_i}$$

heißen **Lagrange-Polynome**. *$(l_0^n, \ldots, l_n^n)$ bilden eine Basis von $\mathbb{P}_n$ und $p(x) = \sum_{k=0}^{n} f_k l_k^n(x)$ ist das Interpolationspolynom zu $(x_0, f_0), \ldots, (x_n, f_n)$.*

Für die Lagrange-Polynome gilt $l_i^n(x_j) = \delta_{ij}$.

Bemerkung: Diese Darstellung ist für die Theorie sehr brauchbar. Mit dieser Konstruktion zu arbeiten ist angenehm, weil für $p(x) = \sum_{i=1}^{n} f_i l_i^n(x)$ gilt

$$p(x_j) = \sum_{i=1}^{n} f_i l_j^n(x_j) = f_j.$$

Nachteil: Die Basispolynome ändern sich bei Hinzunahme von weiteren Stützstellen.

b) **Newton-Form des Interpolationsproblems**

Wähle eine Basis von $\mathbb{P}_n$, so dass A eine untere $\triangle$-Matrix wird

$$\varphi_k(x) := \prod_{j=0}^{k-1}(x - x_j),\ (k = 0, \ldots, n) \implies \varphi_k \in \mathbb{P}_k.$$

$$\begin{aligned} \text{etwa} : \varphi_0(x) &= 1 \left(\text{verwende Konvention } \prod_{j=j_0}^{j_n} a_j = 1 \text{ falls } j_n < j_0 \right) \\ \varphi_1(x) &= (x - x_0) \\ \varphi_2(x) &= (x - x_0)(x - x_1) \\ &\vdots \end{aligned}$$

Damit ist A untere $\triangle$-Matrix, da

$$\varphi_k(x_i) = 0 \text{ für } i < k.$$

Definition 1.3 (Newton-Polynome)
Die Polynome

$$N_k^n := \prod_{j=0}^{k-1}(x - x_j)$$

heißen **Newton-Polynome** *und das Interpolationspolynom* $p(x) = \sum_{k=0}^{n} a_k N_k^n(x)$ *heißt in* **Newton-Form**.

$$\begin{aligned} \text{Es gilt} : a_0 &= \frac{f_0}{\varphi_0(x_0)} = f_0, \\ a_1 &= \frac{(f_1 - \varphi_0(x_1)a_0)}{\varphi_1(x_1)} = \frac{f_1 - f_0}{x_1 - x_0} =: f[x_0, x_1], \\ a_2 &= \frac{(f_2 - \varphi_0(x_2)a_0 - \varphi_1(x_2)a_1)}{\varphi_2(x_2)} \\ &= \frac{\frac{f_2 - f_1}{x_2 - x_1} - \frac{f_1 - f_0}{x_1 - x_0}}{x_2 - x_0} = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0} =: f[x_0, x_1, x_2] \end{aligned}$$

Diese Koeffizienten werden berechnet über die **dividierten Differenzen** $f[x_0, \ldots, x_n]$ *(siehe Abschnitt 1.3).*

1.2 Funktionsinterpolation durch Polynome

Gegeben: Stützstellen $x_0, \ldots, x_n$ und f stetig.

Gesucht: Interpolationspolynom zu $(x_0, f(x_0)), \ldots, (x_n, f(x_n))$.

Satz 1.4 (Fehlerdarstellung)
Sei $f \in C^{n+1}(a,b)$ und $p \in \mathbb{P}_n$ das Interpolationspolynom zu den Stützstellen $x_0, \ldots, x_n$ paarweise verschieden. Dann existiert zu jedem $x \in (a,b)$ ein $\xi_x \in (a,b)$ mit

$$(*) \quad f(x) - p(x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi_x) \cdot \prod_{k=0}^{n} (x - x_k)$$

Beweis: Für $x = x_i$ $(i = 0, \ldots, n)$ ist nichts zu zeigen, da $f(x_i) = p(x_i)$ und $\prod_{k=0}^{n} (x - x_k) = 0$. Sei also $x \neq x_i$: Setze

$$\omega(t) := \prod_{i=0}^{n} (t - x_i)$$

und betrachte

$$\Phi(t) := f(t) - p(t) - \lambda \omega(t)$$

mit $\lambda = \frac{f(x)-p(x)}{\omega(x)} \in \mathbb{R}$ (beachte x fest). Es folgt $\Phi(x) = 0$ und $\Phi(x_i) = 0$ $(i = 0, \ldots, n)$ und somit hat Φ $n+2$ Nullstellen. Nach dem Satz von Rolle folgt weiter: Φ' hat $n+1$ Nullstellen und folglich $\Phi^{(n+1)}$ eine Nullstelle $\xi_x \in (a,b)$ mit

$$0 = \Phi^{(n+1)}(\xi_x) = f^{(n+1)}(\xi_x) - (n+1)! \frac{f(x) - p(x)}{\omega(x)}$$

Also haben wir $(*)$ bewiesen. □

Folgerung 1.5
Seien die Vorraussetzungen von Satz 1.4 erfüllt.
Dann gilt $\|f - p\|_\infty \leq \frac{1}{(n+1)!} \left\|f^{(n+1)}\right\|_\infty \|\omega\|_\infty$ mit dem **Knotenpolynom** $\omega(x) = \prod_{j=0}^{n} (x - x_j)$.

Beweis: Folgt direkt aus Satz 1.4. □
Bemerkung Folgerung 1.5 zeigt, dass die Approximation durch Polynominterpolation durch eine geeignete Wahl der Stützstellen optimiert werden kann. **Frage:** Wird der Interpolationsfehler bei wachsender Stützstellenzahl immer kleiner? Das folgende Beispiel zeigt, dass dies i.A. nicht der Fall ist.

Beispiel 1.6 (Runge)
Betrachten wir $f(x) = \frac{1}{1+x^2}$, $-5 \leq x \leq 5$ und $x_n^{(n)} := -5 + kh_n$, $0 \leq k \leq n$ mit $h_n := \frac{10}{n}$ (gleichmäßige Stützstellenverteilung). Sei $p_n(x_k^{(n)}) = f(x_k^{(n)})$.
Man kann zeigen (siehe Abb. 1.3), dass es ein $\tilde{x} \approx 3,6$ gibt, so dass für $|x| < \tilde{x}$ Konvergenz vorliegt, während der Interpolationsfehler für $|x| > \tilde{x}$ gegen unendlich geht.

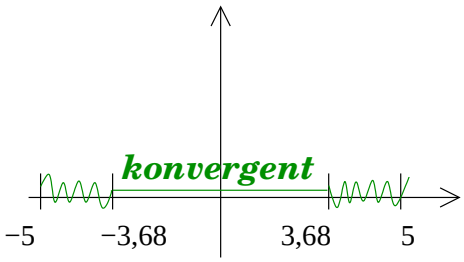


Abbildung 1.3: Interpolation von Funktionen, Beispiel 1.6

Allgemein gilt:

1. Für jede stetige Funktion f existiert eine Folge von Stützstellen mit $p_n \longrightarrow f$ gleichmäßig.
2. Zu jeder Folge von Stützstellen gibt es eine stetige Funktion f, so dass $p_n \not\longrightarrow f$ gleichmäßig.

Optimale Wahl von Stützstellen für das Referenzintervall $[-1,1]$

Idee: Wähle Stützstellen $x_0,\ldots,x_n \in [-1,1]$, so dass $\|w\|_{L^\infty([-1,1])}$ minimiert wird.

Bemerkung: Das Knotenpolynom $\omega(t) = \prod_{k=0}^{n}(t-x_k)$ ist ein normiertes Polynom $(n+1)$-tes Grades (d.h. Koeffizient 1 vor x^{n+1}) und die Stützstellen $x_0,\ldots,x_n$ sind die Nullstellen von ω. Wir erreichen also dann eine optimale Wahl der Stützstellen, wenn wir ein normiertes Polynom $(n+1)$-tes Grades finden, dass unter allen normierten Polynomen die ∞-Norm minimiert.

Definition 1.7 (Tschebyschev-Polynome)
Wir definieren die Tschebyschev-Polynome auf [-1,1] durch
$T_0(x) = 1,\ T_1(x) = x,\ T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x),\ \hat{T}_n(x) = 2^{1-n}T_n(x).$

D.h. $T_2(x) = 2x^2 - 1$, $T_3(x) = 4x^3 - 3x$, $\hat{T}_2(x) = x^2 - \frac{1}{2}$, $\hat{T}_3(x) = x^3 - \frac{3}{4}x$.

Satz 1.8
Für $x \in [-1,1]$ gilt:

$$T_n(x) = \cos\left(n\cos^{-1}(x)\right). \qquad (*)$$

Weiterhin gilt:

(i) $|T_n(x)| \le 1$,

(ii) $T_n\left(\cos\left(\frac{j\pi}{n}\right)\right) = (-1)^j \quad (0 \le j \le n)$,

(iii) $T_n\left(\cos\left(\pi\frac{2j-1}{2n}\right)\right) = 0 \quad (1 \le j \le n)$,

(iv) $T_n \in P_n(-1,1)$,

(v) $\hat{T}_n \in P_n(-1,1)$ mit Koeffizient 1 vor x^n (normiertes Polynom).

Beweis: Nach Additionstheorem gilt

$$\cos(A+B) = \cos(A)\cos(B) - \sin(A)\sin(B).$$

Also folgt

$$\cos((n+1)\Theta) = \cos(n\Theta)\cos(\Theta) - \sin(n\Theta)\sin(\Theta)$$

und

$$\cos((n-1)\Theta) = \cos(n\Theta)\cos(\Theta) + \sin(n\Theta)\sin(\Theta).$$

Zusammen erhalten wir

$$\cos((n+1)\Theta) + \cos((n-1)\Theta) = 2\cos(n\Theta)\cos(\Theta).$$

Setze $\Theta := \cos^{-1}(x)$, bzw. $x := \cos(\Theta)$, so folgt weiter

$$\cos\left((n+1)\cos^{-1}(x)\right) = 2\cos(n\cos^{-1}(x))x - \cos((n-1)\cos^{-1}(x))$$

d.h. $F_n := \cos\left(n\cos^{-1}(x)\right)$ genügt der Rekursionsformel von Definiton 1.7. Weiter ist $F_0(x) = 1$, $F_1(x) = \cos(\cos^{-1}(x)) = x$ und somit $F_n = T_n$, was $(*)$ beweist.

Die Eigenschaften (i), (ii), (iii) folgen direkt aus $(*)$; (iv) und (v) folgen aus der Rekursionsformel für T_n in Definition 1.7. □

Lemma 1.9
Sei $p \in \mathbb{P}_n$ ein normiertes Polynom auf $[-1,1]$. Dann gilt:

$$\|p\|_\infty = \max_{-1\le x\le 1} |p(x)| \ge 2^{1-n} \quad \text{und} \quad \left\|\hat{T}_n\right\|_\infty = 2^{1-n}.$$

Beweis: Annahme: Es gibt ein normiertes Polynom p mit $|p(x)| < 2^{1-n} \quad \forall\, x \in [-1,1]$.

Sei $x_i = \cos\left(\frac{i\pi}{n}\right)$. Nach Satz 1.8 (ii) folgt dann

$$(-1)^i p(x_i) \le |p(x_i)| < 2^{1-n} = (-1)^i \hat{T}_n(x_i)$$

und hieraus

$$(-1)^i \left(\hat{T}(x_i) - p(x_i)\right) > 0 \quad \text{für} \quad 0 \le i \le n.$$

D.h. das Polynom $\hat{T}_n - p$ wechselt $(n+1)$-Mal das Vorzeichen im Intervall $[-1,1]$. Da $\hat{T}_n$ und p normierte Polynome sind, ist dies ein Widerspruch dazu, dass $\hat{T}_n - p$ vom Grad $n-1$ ist.
Weiter folgt aus $|T_n(x)| \le 1$, dass $|\tilde{T}_n(x)| \le 2^{1-n}$ und mit $\tilde{T}_n(x_i) = 2^{1-n}(-1)^i$ folgt die Behauptung. □

Folgerung 1.10 (Optimale Wahl der Stützstellen)
Mit den Stützstellen $x_k = \cos\left(\pi\frac{2k-1}{2(n+1)}\right), \quad k = 1,\dots,n+1$ als die Nullstellen von T_{n+1} gilt, dass das Knotenpolynom gerade $\hat{T}_{n+1}$ ist, d.h. die Maximumsnorm des Knotenpolynoms ist $2^{1-(n+1)}$.

1.3 Dividierte Differenzen

Wiederholung: Die Newton-Form des Interpolationspolynom ist gegeben durch $p(x) = \sum_{k=0}^{n} a_k N_k(x)$ mit

$$N_k(x) = \begin{cases} 1 & : \quad k = 0 \\ \prod_{j=0}^{k-1}(x - x_j) & : \quad k \geq 1 \end{cases}.$$

Gesucht: Algorithmus zur effizienten Berechnung von $a_0, \ldots, a_n$.

Bemerkung: Setze $p_m(x) = \sum_{k=0}^{m} a_k N_k(x)$ für $m \leq n$, dann gilt $p_m(x_j) = f_j \quad (0 \leq j \leq m)$ und $p_m \in \mathbb{P}_m$, da $N_k \in \mathbb{P}_m$ für $0 \leq k \leq m$. D.h. p_m ist **das** Interpolationspolynom in $\mathbb{P}_m$ zu den Daten $(x_0, f_0), \ldots, (x_m, f_m)$. Insbesondere hängt a_k nur ab von $(x_0, f_0), \ldots, (x_k, f_k)$ für $0 \leq k \leq m$. Es wird daher die Schreibweise $f[x_0, \ldots, x_k]$ für a_k benutzt.
Beachte: a_m ist der Koeffizient vor dem x^m im Polynom p_m.

Definition 1.11 (Dividierte Differenzen)
Seien $i_o, \ldots, i_k \in \{0, \ldots, n\}$ paarweise verschieden und sei $p_{i_0,\ldots,i_k}$ das Interpolationspolynom zu den Daten $(x_{i_0}, f_{i_0}), \ldots, (x_{i_k}, f_{i_k})$. Mit $f[x_{i_0}, \ldots, x_{i_k}]$ bezeichnen wir den Koeffizienten vor x^k im Polynom $p_{i_0,\ldots,i_k}$.

$f[i_0, \ldots, i_k]$ wird als **dividierte Differenz** *der Ordnung k bezeichnet.*

Satz 1.12

(i) Die Polynome $p_{i_0,\ldots,i_k}$ genügen der Rekursionsformel

(1) $\quad p_{i_0,\ldots,i_k}(x) = \frac{(x - x_{i_0}) p_{i_1,\ldots,i_k}(x) - (x - x_{i_k}) p_{i_0,\ldots,i_{k-1}}}{x_{i_k} - x_{i_0}}$.

(ii) Die dividierten Differenzen genügen der Rekursionsformel

(2) $\quad f[x_{i_0}, \ldots, x_{i_k}] = \frac{f[x_{i_1},\ldots,x_{i_k}] - f[x_{i_0},\ldots,x_{i_{k-1}}]}{x_{i_k} - x_{i_0}}$, $f[x_{i_l}] = f_{i_l}$.

(iii) Die dividierten Differenzen sind unabhängig von der Reihenfolge ihrer Koeffizienten, d.h. ist $x_{j_0}, \ldots, x_{j_n}$ eine Permutation von $x_{i_0}, \ldots, x_{i_n}$, so gilt $f[x_{j_0}, \ldots, x_{j_n}] = f[x_{i_0}, \ldots, x_{i_n}]$.

Bemerkung: Die dividierten Differenzen können in der Form eines Tableaus geschrieben werden.

x_0	$a_0 = f_0$	$a_1 = f[x_0, x_1]$	$a_2 = f[x_0, x_1, x_2]$	$a_3 = f[x_0, x_1, x_2, x_3]$
x_1	f_1	$f[x_1, x_2]$	$f[x_1, x_2, x_3]$	
x_2	f_2	$f[x_2, x_3]$		
x_3	f_3			

Dabei ist z.B. $f[x_1, x_2, x_3] = \frac{f[x_2,x_3] - f[x_1,x_2]}{x_3 - x_1}$. Beachte, $f_k = f[x_k]$ und $p(x) = \sum_{k=0}^{n} f[x_0, \ldots, x_k] N_k(x)$ ist das gesuchte Interpolationspolynom.

Beispiel 1.13
Wir betrachten die Daten

x	3	1	5
f	1	-3	2

Die dividierten Differenzen liefern:

$$\begin{array}{cc|l} 3 & 1 & 2 = \frac{-3-1}{1-3} \quad -\frac{3}{8} = \frac{5/4-2}{5-3} \\ 1 & -3 & \frac{5}{4} = \frac{2-(-3)}{5-1} \\ 5 & 2 & \end{array}$$

Das Interpolationspolynom ist also $p(x) = 1 + 2(x-3) - \frac{3}{8}(x-3)(x-1)$.

Fügen wir eine Stützstelle hinzu, so betrachten wir die Daten:

x	3	1	5	6
f	1	-3	2	4

Die dividierten Differenzen liefern durch hinzufuegen einer "Diagonalen":

$$\begin{array}{cc|lll} 3 & 1 & 2 & -\frac{3}{8} & \frac{7}{40} \\ 1 & -3 & \frac{5}{4} & \frac{3}{20} & \\ 5 & 2 & 2 & & \\ 6 & 4 & & & \end{array}$$

Das Interpolationspolynom lautet:
$p(x) = 1 + 2(x-3) - \frac{3}{8}(x-3)(x-1) + \frac{7}{40}(x-3)(x-1)(x-5)$

Beweis: (von Satz 1.12)

(i) Setze $R(x)$ als rechte Seite von (1). Zu zeigen: $p_{i_0,\dots,i_n} = R(x)$.

Notation:
$p_k = p_{i_0,\dots,i_k}$, $p_{k-1} = p_{i_0,\dots,i_{k-1}}$ $q_k = p_{i_1,\dots,i_k}$

Dann ist

$$R(x) = \frac{(x - x_{i_0})q_k(x) - (x - x_{i_k})p_{k-1}(x)}{x_{i_k} - x_{i_0}}$$

$$\begin{aligned} \Longrightarrow \quad R(x_{i_0}) &= \tfrac{0-(x_{i_0}-x_{i_k})f_{i_0}}{x_{i_k}-x_{i_0}} = f_{i_0}, \\ R(x_{i_k}) &= \tfrac{(x_{i_k}-x_{i_0})f_{i_k}-0}{x_{i_k}-x_{i_0}} = f_{i_k}, \\ R(x_{i_l}) &= \tfrac{(x_{i_l}-x_{i_k})f_{i_l}-(x_{i_l}-x_{i_k})f_{i_l}}{x_{i_k}-x_{i_0}} = f_{i_l} \quad \forall\ 0 < l < k. \end{aligned}$$

Also ist R ist das Interpolationspolynom zu $(x_{i_0}, f_{i_0}), \dots (x_{i_n}, f_{i_n})$. Aufgrund der Eindeutigkeit folgt dann $p_{i_0,\dots,i_k} = R$.

(ii) Aus (i) folgt, dass der Koeffizient vor x^k in $R(x)$ durch $\frac{f[x_{i_1},\dots,x_{i_k}]-f[x_{i_0},\dots,x_{i_{k-1}}]}{x_{i_k}-x_{i_0}}$ gegeben ist. Nach Definition ist dieser Koeffizient gleich $f[x_{i_0}, \dots, x_{i_k}]$, also folgt (ii).

□

Satz 1.14 (Weitere Eigenschaften der dividierten Differenz)
Sei $f \in C^0(a,b)$, $x_0, \dots, x_n \in (a,b)$ paarweise verschieden und t fest gewählt mit $t \neq x_k\ \forall\ k = 0, \dots, n$.

(i) Wenn p das Interpolationspolynom von f an den Stützstellen $x_0, \ldots, x_n$ ist, so gilt:

$$f(t) - p(t) = f[x_0, \ldots, x_n, t] \prod_{j=0}^{n} (t - x_j)$$

(ii) Ist $f \in C^n(a,b)$, so existiert ein $\xi \in (a,b)$ mit $f[x_0, \ldots, x_n] = \frac{1}{n!} f^{(n)}(\xi)$.

Beweis: (Siehe Übungsaufgaben) □

Algorithmus 1.15 (Dividierte Differenzen)
Ziel: Das ganze Tableau soll berechnet und in eine Matrix gespeichert werden. Wenn eine weitere Stützstelle hinzugefügt wird, dann reicht es die Diagonale der dividierten Differenzen auszurechnen.

$c_{i0} := f_i$

Für $j = 1, \ldots, n$
  Für $i = 0, n-j$
    $c_{ij} := \frac{c_{i+1,j-1} - c_{i,j-1}}{x_{i+j} - x_i}$

$\Longrightarrow \quad c_{ij} = f[x_i, \ldots, x_{i+j}]$.

Nach Hinzunahme einer weiteren Stützstelle (x_{n+1}, f_{n+1}):

$c_{n+1,0} := f_{n+1}$
Für $j = 1, \ldots, n+1$
  $c_{n+1-j,j} := \frac{c_{n+1-j+1,j-1} - c_{n+1-j,j-1}}{x_{n+1} - x_{n+1-j}}$

Satz 1.16 (Auswertung des Interpolationspolynoms)
1. Fall: Die Koeffizienten $a_0, \ldots, a_n$ seien bekannt. Dann kann folgendes Schema zur Auswertung benutzt werden (**Horner-Schema**):

$$\begin{aligned} p(x) &= \sum_{k=0}^{n} a_k \prod_{j=0}^{k-1} \underbrace{(x - x_j)}_{:=\chi_j} \\ &= (((\cdots((a_n \chi_{n-1} + a_{n-1})\,\chi_{n-2} + a_{n-2})\,\chi_{n-3} \cdots)\,\chi_1 + a_1)\,\chi_0 + a_0) \end{aligned}$$

Algorithmus:
$p := a_n$
Für $k = n-1, \ldots, 0$
  $p := p(x - x_k) + a_k$

2. Fall: Das Interpolationspolynom p soll nur an einer Stelle ausgerechnet werden ohne vorher die Koeffizienten zu berechnen (**Neville-Schema**):

Sei $p_{i_0,\ldots,i_k} \in P_n$ das Interpolationspolynom zu $(x_{i_0}, f_{i_0}), \ldots, (x_{i_k}, f_{i_k})$. Das Neville-Schema verwendet die Rekursion aus 1.12 (i):

$$\begin{array}{cc|ccc} x_0 & f_0 = p_0(x) & p_{0,1}(x) & \cdots & p_{0,1,\ldots,n}(x) \\ x_1 & f_1 = p_1(x) & p_{1,2}(x) & \cdots & \\ \vdots & \vdots & \vdots & & \\ x_n & f_n = p_n(x) & p_{n,n+1}(x) & & \end{array}$$

$p_{0,1,\ldots,n}(x)$ ist gesucht, also der letzte Eintrag in der Tabelle.

Beispiel 1.17
Gegeben:

x_i	3	1	5	6
f_i	1	-3	2	4

Gesucht: $p(0)$

(i) **Mit dividierten Differenzen** erhält man

$$p(x) = 1 + 2(x-3) - \frac{3}{8}(x-3)(x-1) + \frac{7}{40}(x-3)(x-1)(x-5)$$

Mit dem **Horner-Schema** folgt anschließend:

$$\begin{array}{rcl} p(0) & = & \left(\left(\left(\frac{7}{40}(0-5) - \frac{3}{8}\right)(-1) + 2\right)(-3) + 1\right) \\ & = & \left(\left(\frac{5}{4} + 2\right)(-3) + 1\right) \\ & = & \left(-\frac{39}{4} + 1\right) = -\frac{35}{4} \end{array}$$

(ii) Mit dem **Neville-Schema** erhält man

$$\begin{array}{cc|ccc} x_0 & f_0 & & & \\ 3 & 1 & \frac{(0-3)(-3)-(0-1)1}{1-3} = -5 & -\frac{79}{8} & -\frac{35}{4} \\ 1 & -3 & \frac{(0-1)2-(0-5)(-3)}{5-1} = -\frac{17}{4} & -\frac{7}{2} & \\ 5 & 2 & \frac{(0-5)4-(0-6)2}{6-5} = -8 & & \\ 6 & 4 & & & \end{array}$$

1.4 Hermite Interpolation

Gegeben: $x_0, \ldots, x_m$ paarweise verschieden und für jede Stützstelle x_i Werte $c_{ij} \in \mathbb{R}$ für $0 \leq j \leq m_i - 1$.

Gesucht: Ein Polynom p mit $p^{(j)}(x_i) = c_{ij}$, $\forall i = 0, \ldots, m, j = 0, \ldots, m_i - 1$.

Die Anzahl der Bedingungen ist $n+1 := m_0 + m_1 + \cdots + m_m$, d.h. es macht Sinn $p \in \mathbb{P}_n$ zu suchen.

Satz 1.18
Es existiert genau ein $p_n \in \mathbb{P}_n$, welches die Bedingungen des Hermite Interpolationspolynoms erfüllt.

Beweis: (analog zu Satz 1.1) □

Satz 1.19 (Fehlerdarstellung für Hermite Interpolation)
Seien $f \in C^{n+1}(a,b)$ und $a \leq x_0 < \ldots < x_m \leq b$. Mit $m_0, \ldots, m_m \in \{1, \ldots, n+1\}$ und $n+1 = \sum_{j=0}^{n} m_j$

Sei $p_n \in \mathbb{P}_n$ das Hermite Interpolationspolynom zu den Daten

$$\begin{array}{cccc} (x_0, f(x_0)), & \ldots & , (x_0, f^{(m_0-1)}(x_0)) \\ (x_1, f(x_1)), & \ldots & , (x_1, f^{(m_1-1)}(x_1)) \\ \vdots & & \vdots \\ (x_m, f(x_m)), & \ldots & , (x_m, f^{(m_m-1)}(x_m)) \end{array}$$

Dann existiert für alle $x \in [a,b]$ ein $\xi_x \in [a,b]$ mit

$$f(x) - p_n(x) = \frac{f^{(n+1)}(\xi_x)}{(n+1)!}\Omega(x)$$

wobei $\Omega(x) := \prod_{k=0}^{n} (x - x_k)^{m_k}$.

Beweis: (analog zu Satz 1.4) □

Beispiel 1.20 (Newton-Form und dividierte Differenzen)
Gesucht: $p \in \mathbb{P}_2$ mit $p(x_0) = c_{00},\ p'(x_0) = c_{01},\ p(x_1) = c_{10}$

Durch dividierte Differenzen:

$$\begin{array}{cc|cc} x_i & f_i & & \\ x_0 & c_{00} & f[x_0,x_0] & f_[x_0,x_0,x_1] \\ x_0 & c_{00} & f[x_0,x_1] & \\ x_1 & c_{10} & & \end{array}$$

Nach Satz 1.14 gilt für $t \in (a,b):\ \exists\, \xi \in (x_0,t)$ mit $f[x_0,t] = f'(\xi)$.

Ist $f' \in C^0(a,b)$ so gilt:

$$\lim_{t \longrightarrow x_0} f[x_0,t] = f'(x_0)$$

Daher macht es Sinn $f[x_0,x_0] = f'(x_0)$ zu setzen. Im Beispiel folgt dann:

$$\begin{array}{cc|cc} x_0 & c_{00} & c_{01} & \frac{f[x_0,x_1]-f[x_0,x_0]}{x_1-x_0} \\ x_0 & c_{00} & \frac{c_{10}-c_{00}}{(x_1-x_0)} & \\ x_1 & c_{10} & & \end{array}$$

Wir erhalten also im Beispiel $f[x_0, x_0, x_1] = \frac{c_{10}-c_{00}}{(x_1-x_0)^2} - \frac{c_{01}}{x_1-x_0}$ und setzten das Interpolationspolynom in der Newtonform an:

$$p(x) = f[x_0] + f[x_0, x_0](x - x_0) + f[x_0, x_0, x_1](x - x_0)^2.$$

Dieser Ansatz läßt sich verallgemeinern zu:

$$p_n(x) = \sum_{k=0}^{n} f[z_0, \ldots, z_k] \prod_{j=0}^{k-1} (x - z_j)$$

mit

$$\begin{aligned} z_0 = \quad & \cdots = z_{m_0-1} && = x_0, \\ z_{k_0} = \quad & \cdots = z_{m_0-m_1-1} && = x_1, \\ & \vdots \\ & \text{usw.} \end{aligned}$$

Satz 1.21 (Rekursionsformel für dividierte Differenzen)
Sei $i_0, \ldots, i_n \in \{0, \ldots, n\}$ und o.B.d.A $z_{i_0} \leq z_{i_1} \leq \cdots \leq z_{i_k}$. Dann gilt

$$f[z_{i_0}, \ldots, z_{i_k}] = \begin{cases} \frac{f[z_{i_1}, \ldots, z_{i_k}] - f[z_{i_0}, \ldots, z_{i_{k-1}}]}{z_{i_k} - z_{i_0}} & : \quad z_{i_k} \neq z_{i_0} \\ \\ \frac{1}{k!} f^{(k)}(z_{i_0}) & : \quad z_{i_k} = z_{i_0} \end{cases}$$

Bemerkung 1.22

(i) Bei der Hermite Interpolation werden gerade die Werte vorgeschrieben, die bei den Dividierten Differenzen Tableau nicht durch die Rekursion gegeben sind.

(ii) Interpolationsprobleme, bei denen nicht für alle $j = 0, \ldots, m_i - 1$ die Werte $p^{(j)}(x_i)$ vorgeschrieben werden, sind nicht so einfach zu lösen (vergleiche Birkoff-Interpolation in den Übungsaufgaben).

1.5 Richardson Extrapolation

Gegeben: Eine Funktion $a : (0, \infty) \longrightarrow \mathbb{R}$.

Gesucht: $a(0) = \lim\limits_{h \searrow 0} a(h)$.

Idee: Wähle $h_0, \ldots, h_n$, setze $a_k = a(h_k)$ und bestimme das Interpolationspolynom zu $(h_0, a_0), \ldots, (h_n, a_n)$ und approximiere $a(0)$ durch $p(0)$.

Beispiel 1.23

(i) **Regel von L'Hospital**

Berechne $\lim\limits_{x \longrightarrow 0} \frac{\cos(x)-1}{\sin(x)}$, d.h. $a(h) = \frac{\cos(h)-1}{\sin(h)}$.

$$\begin{array}{rlll} \text{Setze}: & h_0 = \frac{1}{8} & , & a_0 = -6.258151 \cdot 10^{-2} \\ & h_1 = \frac{1}{16} & , & a_1 = -3.126018 \cdot 10^{-2} \\ & h_2 = \frac{1}{32} & , & a_2 = -1.562627 \cdot 10^{-2} \end{array}$$

$$\Longrightarrow \quad p(0) = -1.02\ldots \cdot 10^{-2}$$

Es ist $a(0) = \lim\limits_{h \searrow 0} \frac{\cos(h)-1}{\sin(h)} = \lim\limits_{h \searrow 0} \frac{-\sin(h)}{\cos(h)} = 0$.

(ii) **Numerische Verfahren** (etwa Differentiation von $f \in C^1$)

Wähle $a(h) = \frac{f(h)-f(-h)}{2h}$.

Ist f analytisch, so gilt die **asymptotische Entwicklung**

$$a(h) = a(0) + \sum_{i=1}^{\infty} \alpha_{2i} h^{2i} \text{ mit } a(0) = f'(0)$$

und

$$\begin{aligned} f(h) &= f(0) + \sum_{i=1}^{\infty} f^{(i)}(0) h^i, \\ f(-h) &= f(0) + \sum_{i=1}^{\infty} f^{(i)}(0)(-h)^i = \sum_{i=1}^{\infty} f^{(i)}(0)(-1)^i h^i. \end{aligned}$$

Das heißt, $a(h)$ ist eine gerade Funktion ($a(h) = a(-h)$) und das Interpolationspolynom solle nur h^{2k}-Terme enthalten.

Sei $f(x) = \sin(x) \quad \Longrightarrow \quad a(h) = \frac{\sin(h)-\sin(-h)}{2h} = \frac{\sin(h)}{h}$, so folgt für $p(x) = q(x^2), q \in \mathbb{P}_1$:

$$\begin{array}{rlll} h_0 = \frac{1}{8} & , & a_0 = 0.9973 \\ h_0 = \frac{1}{16} & , & a_0 = 0.99934 \\ h_0 = \frac{1}{32} & , & a_0 = 0.99983 \end{array}$$

$$\Longrightarrow \quad p(0) = 0.999999926.$$

Satz 1.24 (Extrapolationfehler)
Gelte für $a : (0, \infty) \to \mathbb{R}$ die asymptotische Entwicklung

$$a(h) = a(0) + \sum_{j=1}^{n} \alpha_j h^{qj} + a_{n+1}(h) h^{q(n+1)}$$

mit $q > 0$ und $a_{n+1}(h) = \alpha_{n+1} + o(1)$. Dabei seien $\alpha_1, \ldots, \alpha_{n+1} \in \mathbb{R}$ unabhängig von h. Sei $(h_k)_{k\in\mathbb{N}}$ eine monoton fallende Folge, $h_k > 0$ und $\frac{h_{k+1}}{h_k} \leq \rho < 1$ für $\rho > 0$ unabhängig von k. Mit $p_n^{(k)} \in \mathbb{P}_n$ bezeichnen wir das Interpolationspolynom in h zu den Daten $(h_k^q, a(h_k)), \ldots, (h_{k+n}^q, a(h_{k+n}))$. Dann gilt:

$$\left| a(0) - p_n^{(k)}(0) \right| = O(h_k^{q(n+1)}) \text{ für } k \longrightarrow \infty.$$

Beweis: Setze $z = h^q$, $z_k = h_k^q$. In der Lagrange Darstellung ist das Interpolationspolynom gegeben durch $p_n^{(k)}(z) = \sum_{i=0}^{n} a(h_{k+i}) L_{k,i}^n(z)$ mit $L_{k,i}^n(z) = \prod_{\substack{l=0 \\ l\neq i}}^{n} \frac{z - z_{k+l}}{z_{k+i} - z_{k+l}}$. Mit der asymptotischen Entwicklung von a folgt

$$\begin{aligned} p_n^{(k)}(0) &= \sum_{i=0}^{n} \left(a(0) + \sum_{j=1}^{n} \alpha_j z_{k+i}^j + \alpha_{n+1} z_{k+i}^{n+1} + o(1) z_{k+i}^{n+1} \right) L_{k,i}^n(0) \\ &= a(0) \sum_{i=0}^{n} L_{k,i}^n(0) + \sum_{j=1}^{n+1} \alpha_j \sum_{i=0}^{n} z_{k+i}^j L_{k,i}^n(0) + o(1) \sum_{i=0}^{n} z_{k+i}^{n+1} L_{k,i}^n(0). \end{aligned}$$

Um die Summanden $z^r L_{k,i}^n(0)$ zu berechnen, verwenden wir die Fehlerdarstellung aus Satz 1.4 mit $f(z) = z^r$, $r = 0, \ldots, n+1$ und dem Interpolationspolynom $q_n^{(k)}$ zu den Daten $(z_k, f(z_k)), \ldots, (z_{k+n}, f(z_{k+n}))$, d.h. $q_n^{(k)}(z) = \sum_{i=0}^{n} f(z_{k+i}) L_{k,i}^n(z)$, bzw. $q_n^{(k)}(0) = \sum_{i=0}^{n} z_{k+i}^r L_{k,i}^n(0)$.

Es gilt $f(0) - q_n^{(k)}(0) = \frac{1}{(n+1)!} f^{(n+1)}(\xi_0) \prod_{i=0}^{n} (0 - z_{k+i})$ und somit folgt

$$\begin{aligned} -\sum_{i=0}^{n} z_{k+i}^r L_{k,i}^n(0) &= \frac{1}{(n+1)!} f^{(n+1)}(\xi_0) (-1)^{n+1} \prod_{i=0}^{n} z_{k+i} - f(0) \\ &= \begin{cases} -1 & : \quad r = 0 \\ 0 & : \quad r = 1, \ldots, n \\ (-1)^{n+1} \prod_{i=0}^{n} z_{k+i} & : \quad r = n+1 \end{cases} \end{aligned}$$

Damit erhalten wir

$$p_n^{(k)}(0) = a(0) + \alpha_{n+1} (-1)^n \prod_{i=0}^{n} z_{k+i} + \sum_{i=0}^{n} o(1) z_{k+i}^{n+1} L_{k,i}^n(0).$$

Es gilt:

$$\begin{aligned} \left| \alpha_{n+1} (-1)^n \prod_{i=0}^{n} z_{k+i} \right| &\leq |\alpha_{n+1}| \prod_{i=0}^{n} z_k = |\alpha_{n+1}| \, z_k^{(n+1)} \\ &= |\alpha_{n+1}| \, h_k^{q(n+1)} = O(h_k^{q(n+1)}). \end{aligned}$$

Außerdem gilt $\left| L_{k,i}^n(0) \right| = \prod_{\substack{l=0 \\ l\neq i}}^{n} \frac{1}{\left| \frac{z_{k+i}}{z_{k+l}} - 1 \right|} \leq C(\rho, n, q)$, unabhängig von k:

$$\Longrightarrow \left|\sum_{i=0}^{n} o(1)L_{i,k}^{n}(0)z_{k+i}^{n+1}\right| \leq C(\rho,n,q)o(1)z_k^{n+1}$$
$$\leq C(\rho,n,q)o(1)h_k^{q(n+1)} = O(h_k^{q(n+1)}),$$

$$\Longrightarrow \left|p_n^{(k)}(0) - a(0)\right| = O(h_k^{q(n+1)}).$$ □

Algorithmus: (Richardson Extrapolation)
Zur Berechnung von $p_n^{(k)}(0)$ eignet sich das Neville Schema:

$$p_n^{(k)}(0) = p_{n-1}^{(k+1)}(0) + \frac{p_{n-1}^{(k+1)}(0) - p_{n-1}^{(k)}(0)}{\frac{z_k}{z_{k+n}} - 1}$$

mit $p_n^{(k)}(z) = p_{k,k+1,\ldots,k+n}(z)$.
Mit $a_{k,n} := p_n^{(k-n)}(0)$ erhält man als Rekursion für $a_{k,n}$:

$$a_{k,n} = a_{k,n-1} + \frac{a_{k,n-1} - a_{k-1,n-1}}{\left(\frac{h_{k-n}}{h_k}\right)^q - 1}.$$

Als Tableau mit Startwert $a_{k,0} = a(h_k)$ ergibt sich dann

$$\begin{array}{llllll} h_0 & a_{0,0} & & & & \\ h_1 & a_{1,0} & a_{1,1} & & & \\ h_2 & a_{2,0} & a_{2,1} & a_{2,2} & & \\ \vdots & \vdots & \vdots & \vdots & \ddots & \\ h_k & a_{k,0} & a_{k,1} & a_{k,2} & \cdots & a_{k,k} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \end{array}$$

Beispiel 1.25
Berechnung von $e = \lim_{n\to\infty}(1+\frac{1}{n})^n = \lim_{h\to 0}(1+h)^{\frac{1}{h}}$, d.h. $a(h) = (1+h)^{\frac{1}{h}}$.

Wähle $h_k = 2^{-k} \Longrightarrow a_{k,0} = a(h_k) = (1+2^{-k})^{2^k}$.

$\Longrightarrow a_{0,0} = 2,\ a_{1,0} = \frac{9}{4},\ a_{2,0} = \frac{625}{256} \approx 2.44.$

Als Tableau:

		$n=0$	$n=1$	$n=2$	
$k=0:$	$h_0=1$	2			
$k=1:$	$h_1=\frac{1}{2}$	$\frac{9}{4}$	$\frac{5}{2}$		
$k=2:$	$h_2=\frac{1}{4}$	$\frac{625}{256}$	$\frac{337}{128}$	$\frac{257}{96} \approx 2.67708$	

Es folgt also z.B. $a_{22} \approx 2.67708$ als Approximation von $e \approx 2.718281828$. Bereits $a_{5,5}$ liefert $a_{5,5} \approx 2.71827$, während $a_{5,0} \approx 2.6769 \approx a_{22}$.

Nebenrechnung:

$$a_{1,1} = a_{1,0} + \frac{a_{1,0}-a_{0,0}}{\left(\frac{h_0}{h_1}\right)-1} = \frac{9}{4} + \frac{\frac{9}{4}-2}{2-1} = \frac{5}{2}$$
$$a_{2,1} = a_{2,0} + \frac{a_{2,0}-a_{1,0}}{2-1} = \frac{337}{128}$$

Allgemein: $\frac{h_{k-n}}{h_k} = 2^{-k+n+k} = 2^n$

$$a_{2,2} = a_{2,1} + \frac{a_{2,1}-a_{1,1}}{2^2-1} = \frac{257}{96}$$

Aufwand: Die Richardson Extrapolation eignet sich vor allem, falls $a(h)$ sehr teuer zu berechnen ist, etwa falls für den Aufwand $A(h)$ gilt $A(h) = O(1/h)$. In unserem Beispiel folgt dann für $a(1/32)$ der Aufwand $A(h) = 32$, während der Aufwand zur Berechnung von a_{22} gegeben ist durch $A(1) + A(1/2) + A(1/4) = 7$.

1.6 Trigonometrische Interpolation

Gegeben: $(x_0, y_0), \ldots, (x_n, y_n)$, x_k paarweise verschieden, $x_k \in [0, \omega)$, $\omega > 0$.

Gesucht: Periodische Funktion $t_n : \mathbb{R} \longrightarrow \mathbb{R}$ mit Periode ω, welche die Daten interpoliert, d.h. $\forall\, x \in \mathbb{R} :\ t_n(x + \omega) = t_n(x)$ und $t_n(x_k) = y_k,\ k = 0, \ldots, n$.

Annahme (O.B.d.A): $\omega = 2\pi$.

Die Fourier Analysis legt nahe, die gesuchte Funktion t_n aus Funktionen der Form

$$1, \cos(x), \cos(2x), \ldots$$
$$\sin(x), \sin(2x), \ldots$$

zusammensetzen.

Ansatz: Suche Koeffizienten (a_k, b_k) mit

$$t_n(x) = \frac{a_0}{2} + \sum_{k=1}^{m} (a_k \cos(kx) + b_k \sin(kx)) + \frac{\Theta}{2} a_{m+1} \cos((m+1)x),$$

wobei

$$\Theta := \begin{cases} 0 & : \ n \text{ gerade} \\ 1 & : \ n \text{ ungerade} \end{cases}, \quad m := \begin{cases} \frac{n}{2} & : \ n \text{ gerade} \\ \frac{n-1}{2} & : \ n \text{ ungerade} \end{cases}.$$

Viele Aussagen der diskreten Fourier Analysis lassen sich kompakter über $\mathbb{C}$ formulieren, wobei die **Eulersche Formel**

$$e^{iz} = \cos(z) + i \cdot \sin(z)$$

benutzt wird.

Definition 1.26 (Trigonometrische Polynome)
Wir definieren den Raum der Trigonometrischen Polynome vom Grad n durch

$$T_n := \left\{ t^* : \ \mathbb{C} \longrightarrow \mathbb{C} \ \middle|\ t^*(z) = \sum_{k=0}^{n} c_k e^{ikz} \right\}$$

Mit $w := e^{iz}$ *gilt* $t^*(z) = \sum_{k=0}^{n} c_k w^k$.

Lemma 1.27

(i) Seien $(a_k)_{k=0}^{\infty}$, $(b_k)_{k=0}^{\infty}$ reelle Folgen. Setze $b_0 = 0$, $a_{-k} = a_k$, $b_{-k} = -b_k$ und $c_k = \frac{1}{2}(a_k - i \cdot b_k)$ für $k \in \mathbb{Z}$. Dann gilt:

$$\frac{a_0}{2} + \sum_{k=1}^{m} (a_k \cos(kx) + b_k \sin(kx)) = \sum_{k=-m}^{m} c_k e^{ikx}.$$

(ii) Sei $(c_k)_{k=-m}^m,\ c_k \in \mathbb{C}$. Setze $a_k = c_k + c_{-k},\ b_k = i \cdot (c_k - c_{-k}),\ k = 0, \dots, m$. Dann gilt:

$$\frac{1}{2}a_0 + \sum_{k=1}^{m} (a_k \cos(kx) + b_k \sin(kx)) = \sum_{k=-m}^{m} c_k e^{ikx}.$$

Beweis: (Ohne Beweis) □

Voraussetzungen und Notationen für diesen Abschnitt

- Äquidistante Stützstellen, d.h. $x_k = \frac{2\pi}{n+1}k,\ k = 0, \dots, n$.
- $w(x) := e^{ix},\ E_k(x) := e^{ikx} = w^k(x)\ (k \in \mathbb{Z})$.
- $\hat{w} := e^{i \cdot \frac{2\pi}{n+1}} \in \mathbb{C},\ w_k := w(x_k) = e^{ik\frac{2\pi}{n+1}} = \hat{w}^k\ (k \in \mathbb{Z})$.

Lemma 1.28

(i) $(E_k)_{k\in\mathbb{Z}}$ bilden ein **Orthonormalsystem**, d.h. $\langle E_k, E_l \rangle = \delta_{kl}$.

(ii) $w_k^{n+1} = 1$, d.h. $w_0, \dots, w_n$ sind die $(n+1)$ Einheitswurzeln und $w_0, \dots, w_n$ sind paarweise verschieden.

(iii) $w_k^l = w_l^k,\ w_{n+1-k}^l = w_{-k}^l,\ w_k^{-l} = \overline{w_k^l}$.

(iv) $\frac{1}{n+1} \sum_{j=0}^{n} w_j^{k-l} = \delta_{kl},\ 0 \le k, l \le n$.

(v) Für festes $j \in \mathbb{N}$: $\sum_{k=0}^{n} \sin(jx_k) = 0,\ \sum_{k=0}^{n} \cos(jx_k) = \begin{cases} n+1 & : \ (n+1) \mid j \\ 0 & : \ \text{sonst} \end{cases}$.

Beweis: (Siehe Übungsaufgaben) □

Satz 1.29 (Trigonometrische Interpolation in $\mathbb{C}$)
Zu gegebenen Daten $y_0, \dots, y_n \in \mathbb{C}$ existiert genau ein $t_n^* \in T_n$ mit $t_n^*(x_k) = y_k$ für $k = 0, \dots, n$.

Die Koeffizienten c_k sind gegeben durch:

$$c_k = \frac{1}{n+1} \sum_{j=0}^{n} y_j e^{-ijx_k} = \frac{1}{n+1} \sum_{j=0}^{n} y_j w_k^{-j}.$$

Beweis: Um die Existenz und Eindeutigkeit zu zeigen, verwenden wir Satz 1.1, der auch im Komplexen gezeigt werden kann. Es existiert daher genau ein $p \in \mathbb{P}_n$ mit $p(x) = \sum_{k=0}^{n} c_k x^k$ mit $c_k \in \mathbb{C}$ und $p(w_k) = y_k\ (k = 0, \dots, n)$ (Interpolationspolynom zu $(w_0, y_0), \dots, (w_n, y_n)$).

Mit $t_n^*(x) = \sum_{k=0}^{n} c_k e^{ikx}$ gilt: $t_n^*(x_l) = \sum_{k=0}^{n} c_k e^{ikx_l} = \sum_{k=0}^{n} c_k w_l^k = p(w_l) = y_l$.

Um die explizite Darstellung der Koeffizienten zu zeigen, verwenden wir Lemma 1.28:

$$\begin{aligned}\sum_{j=0}^{n} y_j w_k^{-j} &= \sum_{j=0}^{n} p(w_j) w_k^{-j} = \sum_{j=0}^{n} \left(\sum_{l=0}^{n} c_l w_j^l \right) w_j^{-k} \\ &= \sum_{l=0}^{n} c_l \left(\sum_{j=0}^{n} w_j^{l-k} \right) = \sum_{l=0}^{n} c_l (n+1)\delta_{lk} = (n+1)c_k.\end{aligned}$$

Also folgt $c_k = \frac{1}{n+1} \sum_{j=0}^{n} y_j w_k^{-j}$. □

Satz 1.30 (Trigonometrische Interpolation in $\mathbb{R}$)

Für $n \in \mathbb{N}$ gegeben, setze $m = \begin{cases} \frac{n}{2} & : \ n \text{ gerade} \\ \frac{n-1}{2} & : \ n \text{ ungerade} \end{cases}$, und $\Theta = \begin{cases} 0 & : \ n \text{ gerade} \\ 1 & : \ n \text{ ungerade} \end{cases}$.

Zu gegebenen Daten $y_0, \ldots, y_n \in \mathbb{R}$ existiert genau eine Funktion

$$t_n(x) = \frac{a_0}{2} + \sum_{k=1}^{m} (a_k \cos(kx) + b_k \sin(kx)) + \frac{\Theta}{2} a_{m+1} \cos((m+1)x)$$

mit $t_n(x_k) = y_k, \ k = 0, \ldots, n$.

Für die Koeffizienten a_k, b_k gilt:

$$a_k = \frac{2}{n+1} \sum_{j=0}^{n} y_j \cos(jx_k),$$

$$b_k = \frac{2}{n+1} \sum_{j=0}^{n} y_j \sin(jx_k).$$

Beweis: 1. Sei t_n^* das komplexe Interpolationspolynom zu $(x_0, y_0), \ldots, (x_n, y_n)$. Nach Satz 2.29 gilt:

$$t_n^*(x) = \sum_{k=0}^{n} c_k e^{ikx} \text{ mit } c_k = \frac{1}{n+1} \sum_{j=0}^{n} y_j w_k^{-j}.$$

Setze $c_{-k} = c_{n+1-k}, \ k = 1, \ldots, m$, d.h. $c_{-1} = c_n, \ c_{-2} = c_{n-1}, \ldots, c_{-m} = \begin{cases} c_{m+1} & : \ n \text{ gerade} \\ c_{m+2} & : \ n \text{ ungerade} \end{cases}$.

Setze $a_k = c_k + c_{-k}, \ b_k = i \cdot (c_k - c_{-k}), \ k = 0, \ldots, m$ und $a_{m+1} = \begin{cases} 0 & : \ n \text{ gerade} \\ 2c_{m+1} & : \ n \text{ ungerade} \end{cases}$.

Nach Lemma 1.27 gilt dann:

$$(*) \ \frac{a_0}{2} + \sum_{k=1}^{m} (a_k \cos(kx) + b_k \sin(kx)) = \sum_{k=-m}^{m} c_k e^{ikx}.$$

Es folgt

$$\begin{aligned} y_l = \sum_{k=0}^{n} c_k w_l^k &= \sum_{k=0}^{m} c_k w_k^l + \sum_{k=1}^{m} c_{-k} w_{n+1-k}^l + \Theta c_{m+1} w_{m+1}^l \\ &= \sum_{k=-m}^{m} c_k w_k^l + \Theta c_{m+1} w_{m+1}^l. \end{aligned}$$

Für n ungerade gilt: $m+1 = \frac{n+1}{2}$ und daher

$$\begin{aligned} w_{m+1}^l &= \cos((m+1)x_l) + i \cdot \sin((m+1)x_l) \\ &= \cos((m+1)x_l) + i \cdot 0, \text{ da } (m+1)x_l = \tfrac{n+1}{2} l \tfrac{2\pi}{n+1} = l\pi. \end{aligned}$$

$$\begin{aligned} \Longrightarrow \quad y_l &= \sum_{k=-m}^{m} c_k w_k^l + \Theta c_{m+1} w_{m+1}^l \\ &\overset{(*)}{=} \frac{a_0}{2} + \sum_{k=1}^{m} (a_k \cos(kx_l) + b_k \sin(kx_l)) + \frac{\Theta}{2} a_{m+1} \cos((m+1)x_l) \\ &= t_n(x_l). \end{aligned}$$

2. Eindeutigkeit folgt, da das LGS, welches die Koeffiziente a_k, b_k bestimmt, für jede rechte Seite $y_0, \ldots, y_n$ lösbar ist. Daher ist die Matrix regulär.

3. Die explizite Darstellung der Koeffizienten folgt aus:

$$c_{-k} = c_{n+1-k} = \frac{1}{n+1} \sum_{j=0}^{n} y_j w_{n+1-k}^{-j} = \frac{1}{n+1} \sum_{j=0}^{n} y_j w_k^j.$$

Wir erhalten:

$$a_k = c_k + c_{-k} = \frac{1}{n+1} \left(\sum_{j=0}^{n} y_j \left(e^{-ijx_k} + e^{ijx_k} \right) \right) = \frac{2}{n+1} \sum_{j=0}^{n} y_j \cos(jx_k),$$

$$b_k = i \cdot (c_k - c_{-k}) = \frac{i}{n+1} \left(\sum_{j=0}^{n} y_j \left(e^{-ijx_k} - e^{ijx_k} \right) \right) = \frac{2}{n+1} \sum_{j=0}^{n} y_j \sin(jx_k).$$

Bemerkung: $t_n(x_l) = t_n^*(x_l)$, aber im Allgemeinen ist $t_n(x) \neq t_n^*(x)$ für $x \neq x_l$ $(l = 0, \ldots, n)$. Es gilt sogar $t_n(x) \neq Re(t_n^*(x))$. □

Beispiel 1.31
Gegeben: $n = 2,\ x_0 = 0,\ x_1 = \frac{2}{3}\pi,\ x_2 = \frac{4}{3}\pi.$

Es gilt: $\cos(x_1) = \cos(x_2) = -\frac{1}{2},\ \sin(x_1) = -\sin(x_2) =: \xi,\ 2x_1 = x_2$ und $2x_2 \equiv x_1 \mod 2\pi.$

$c_0 = \frac{1}{3} \left(y_0 e^{-i0} + y_1 e^{-i0} + y_2 e^{-i0} \right) = \frac{1}{3}(y_0 + y_1 + y_2),$

$c_1 = \frac{1}{3} \left(y_0 e^{-i0} + y_1 e^{-i\frac{2}{3}\pi} + y_2 e^{-i2\frac{2}{3}\pi} \right) = \frac{1}{3} y_0 - \frac{1}{6}(y_1 + y_2) + i \cdot \frac{\xi}{3}(y_1 - y_2),$

$c_2 = \frac{1}{3} \left(y_0 e^{-i0} + y_1 e^{-i\frac{4}{3}\pi} + y_2 e^{-i2\frac{4}{3}\pi} \right) = \frac{1}{3} y_0 - \frac{1}{6}(y_1 + y_2) + i \cdot \frac{\xi}{3}(y_2 - y_1).$

Im Reellen: $(m = 1, \Theta = 0)$

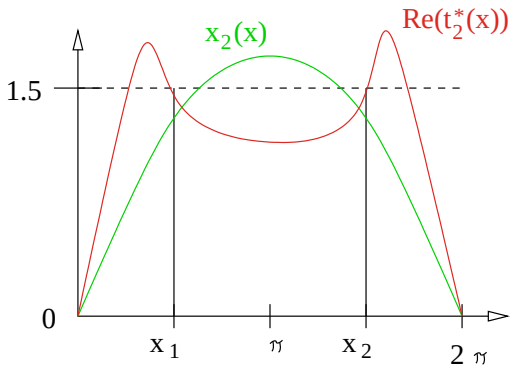


Abbildung 1.4: Beispiel 1.31

$a_0 = \frac{2}{3}(y_0 + y_1 + y_2)\,,\ a_1 = \frac{2}{3}\left(y_0 - \frac{1}{2}y_1 - \frac{1}{2}y_2\right)$

$b_1 = \frac{2\xi}{3}(y_1 - y_2)$

Seien $y_0 = 0,\ y_1 = y_2 = \frac{3}{2}$, so erhalten wir

$$t_2(x) = 1 - \cos(x).$$

Dahingegen erhalten wir als Realteil des komplexen Interpolationspolynoms (siehe auch Abb. 1.4):

$$Re(t_2^*(x)) = 1 - \frac{1}{2}(\cos(x) + \cos(2x))\,.$$

1.6.1 Schnelle Fourier Transformation (FFT)

Die Schnelle Fourier Transformation wird auch **FFT** (Fast Fourier Transformation) genannt.

Ziel: Effiziente Berechnung von $c_0, \ldots, c_n$. (a_k, b_k) können dann im zweiten Schritt schnell bestimmt werden.

Idee: *Divide and Conquer*-Verfahren: Das Problem der Größe n wird in 2 äquivalente Probleme der Größe $\frac{n}{2}$ aufgeteilt und separat gelöst, dann werden die beiden Lösungen wieder zu einer gesammten Lösung zusammengefügt. Am einfachsten ist die FFT darstellbar, falls $n = 2^Q - 1$, d.h. für 2^Q Daten $y_0, \ldots, y_n$.

Sei n ungerade und seien $m = \frac{n-1}{2}$, $l \in \{0, \ldots, n\}$ fest. Dann folgt

$$\begin{aligned} c_l &= \frac{1}{n+1}\sum_{j=0}^{n} y_j w_j^{-l} = \frac{1}{n+1}\left(\sum_{j=0}^{m} y_{2j} w_{2j}^{-l} + \sum_{j=0}^{m} y_{2j+1} w_{2j+1}^{-l}\right) \\ &= \frac{1}{n+1}\left(\sum_{j=0}^{m} y_{2j} w_{2j}^{-l} + \hat{w}^{-l}\left(\sum_{j=0}^{m} y_{2j+1} w_{2j}^{-l}\right)\right)\,, \text{ mit } \hat{w} = e^{i\frac{2\pi}{n+1}}. \end{aligned}$$

Da $n+1 = 2(m+1)$ folgt:

$$c_l = \frac{1}{2}\left(\frac{1}{m+1}\sum_{j=0}^{m} y_{2j} w_{2j}^{-l} + \hat{w}^{-l}\frac{1}{m+1}\sum_{j=0}^{m} y_{2j+1} w_{2j}^{-l}\right).$$

Sei $l_1 \equiv l \bmod (m+1)$, d.h. $l_1 \in \{0,\ldots,m\}$ und $l = \lambda(m+1)+l_1\ \lambda \in \mathbb{N}$. Dann folgt $l = \frac{1}{2}\lambda(n+1)+l$ und somit

$$w_{2j}^{-l} = e^{-il2j\frac{2\pi}{n+1}} = e^{-i\lambda j 2\pi - i\cdot l_1 2j\frac{2\pi}{n+1}} = e^{-i\lambda j 2\pi} w_{2j}^{-l_1} = w_{2j}^{-l_1}$$

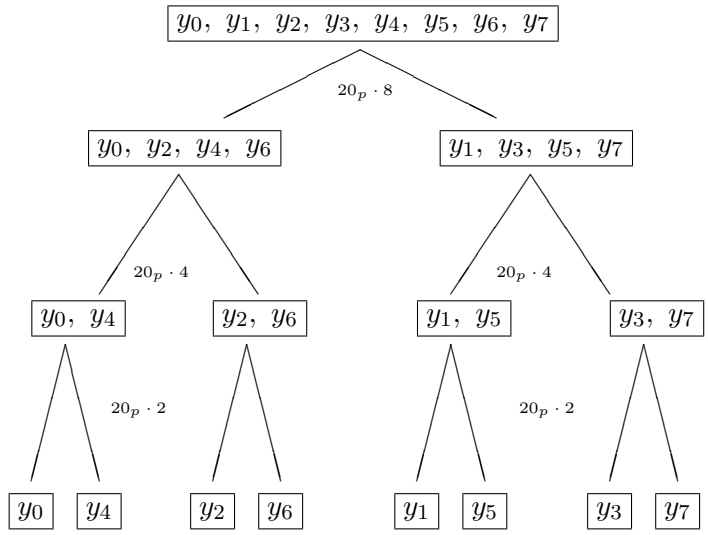
$$\Longrightarrow \quad c_l = \frac{1}{2}\left(c_{l_1}^{even} + c_{l_1}^{odd}\hat{w}^{-l}\right)$$

mit

$$\begin{aligned} c_{l_1}^{even} &= \frac{1}{m+1}\sum_{j=0}^{m} y_{2j} w_{2j}^{-l_1}, \\ c_{l_1}^{odd} &= \frac{1}{m+1}\sum_{j=0}^{m} y_{2j+1} w_{2j}^{-l_1},\ l_1 \in \{0,\ldots,\frac{n+1}{2}\}. \end{aligned}$$

Dabei sind $c_{l_1}^{even}$, $c_{l_1}^{odd}$ geraade die Koeffizienten des komplexen trigonometrischen Polynoms zu den Daten $(x_0,y_0),(x_2,y_2),\ldots,(x_{n-1},y_{n-1})$, bzw. zu $(x_0,y_1),(x_2,y_3),\ldots,(x_{n-1},y_n)$.

Idee des Algorithmus:

Stützstellen $(Q = 3)$	**Daten**
$x_0,\ x_1,\ x_2,\ x_3,\ x_4,\ x_5,\ x_6,\ x_7$	
$8 \cdot 2$	
$x_0,\ x_2,\ x_4,\ x_6$	
$4 \cdot 4$	
$x_0,\ x_4$	
$2 \cdot 8$	
Rechenaufwand: 0	
$48 = (3 \cdot 2 \cdot (n+1))$	

Allgemein: Pro Level $2(n+1)$ Operationen bei $\log_2(n)$ Levels $\implies$ Anzahl der Operationen zur Berechnung von $c_0, \ldots, c_n$ beträgt $2(n+1)\log_2(n) = O(n \log_2 n)$.

Satz 1.32

Sei $n = 2m+1, m \in \mathbb{N}$ und $y_0, \ldots, y_n$ gegeben. $t_n^*(x) = \sum_{j=0}^{n} c_j e^{ijx}$ sei das komplexe trigonometrische Interpolationspolynom zu $(x_0, y_0), \ldots, (x_n, y_n)$.

Sei $t_n^{even}(x) = \sum_{j=0}^{m} c_j^{even} e^{ijx}$ das Interpolationspolynom zu $(x_0, y_0), (x_2, y_2) \ldots, (x_{2m}, y_{2m})$ und $t_n^{odd}(x) = \sum_{j=0}^{m} c_j^{odd} e^{ijx}$ zu $(x_0, y_1), , (x_2, y_3) \ldots, (x_{2m}, y_{2m+1})$. Dann gilt

$$(*) \quad t_n^*(x) = \frac{1}{2}\left(1 + e^{i\cdot(m+1)x}\right) t_n^{even}(x) + \frac{1}{2}\left(1 - e^{i\cdot(m+1)x}\right) t_n^{odd}\left(x - \frac{\pi}{m+1}\right)$$

und es ist $c_l = \frac{1}{2}\left(c_l^{even} + \hat{w}^{-l} c_l^{odd}\right)$, $c_{l+m+1} = \frac{1}{2}\left(c_l^{even} - \hat{w}^{-l} c_l^{odd}\right)$ für $l = 0, \ldots, m$ und $\hat{w} = e^{i\frac{\pi}{m+1}}$.

Beweis: Sei r_n die rechte Seite von $(*)$, d.h.

$$\begin{aligned} r_n(x) &= \tfrac{1}{2}\sum_{j=0}^{m}\left[\left(1+e^{i(m+1)x}\right)c_j^{even}e^{ijx}+\left(1-e^{i(m+1)x}\right)c_j^{odd}e^{ij\left(x-\frac{\pi}{m+1}\right)}\right] \\ &= \tfrac{1}{2}\sum_{j=0}^{m}\left[c_j^{even}\left(e^{ijx}+e^{i(j+m+1)x}\right)+c_j^{odd}\left(e^{ijx}-e^{i(j+m+1)x}\right)e^{-ij\frac{\pi}{m+1}}\right] \\ &= \tfrac{1}{2}\sum_{j=0}^{m}\left(c_j^{even}+e^{-ij\frac{\pi}{m+1}}c_j^{odd}\right)e^{ijx}+\tfrac{1}{2}\sum_{j=m+1}^{2m+1}\left(c_{j-(m+1)-l}^{even}-e^{-ij\frac{\pi}{m+1}}c_{j-(m+1)}^{odd}e^{ijx}\right) \\ &= \sum_{j=0}^{n}\hat{c}_j e^{ijx}\in T_n. \end{aligned}$$

Wegen der Eindeutigkeit des Interpolationspolynom folgt $t_n^* = r_n$, falls r_n die Interpolationsbedingung erfüllt. Für $x_l = \frac{2\pi}{n+1}l$ gilt:

$$e^{i(m+1)x_l} = e^{i\frac{2\pi}{2n+1}(n+1)l} = e^{il\pi} = \begin{cases} 1 & : \ l \text{ gerade} \\ -1 & : \ l \text{ ungerade} \end{cases}$$

$$\Longrightarrow \quad r_n(x_l) \overset{(*)}{=} \begin{cases} t_n^{even}(x_l) & : \ l \text{ gerade} \\ t_n^{odd}\left(x_l-\frac{\pi}{m+1}\right) & : \ l \text{ ungerade} \end{cases}$$

$$= \begin{cases} t_n^{even}(x_l) & : \ l \text{ gerade} \\ t_n^{odd}(x_{l-1}) & : \ l \text{ ungerade} \end{cases}$$

Also ist $r_n(x_l) = y_l$ und damit $t_n^* \equiv r_n \implies \hat{c}_j = c_j$.
Da $e^{-ij\frac{\pi}{m+1}} = \hat{w}^{-j}$, folgt die Formel für c_l aus der Definition von $\hat{c}_l$. □

Algorithmus:

Für $q = 0, \ldots, Q$ sei $t_k^q(x) = \sum_{j=0}^{2^q-1} c_{k,j}^q e^{ijx}, \ k = 0, \ldots, 2^{Q-q} - 1$
das Interpolationspolynom zu $\left(x_{j2^{Q-q}}, y_{j2^{Q-q}+k}\right)_{j=0}^{2^q-1}$.
Sei $q \geq 1$:
Nach Satz 1.32 mit $m = 2^q - 1$, bzw. $n = 2^{q+1} - 1$ und den Daten
$(x_{j2^{Q-q}}, y_{j2^{Q-q}+k})_{j=0}^{2^q-1}$ gilt für $k = 0, \ldots, 2^{Q-(q+1)} - 1$:

$$\begin{aligned} c_{k,l}^{q+1} &= \frac{1}{2}\left(c_{k,l}^q + e^{-i\frac{2\pi}{2^{q+1}}l} c_{k+2^{Q-q-1},l}^q\right) \ l = 0, \ldots, 2^q - 1, \\ c_{k,l+2^q}^{q+1} &= \frac{1}{2}\left(c_{k,l}^q - e^{-i\frac{2\pi}{2^{q+1}}l} c_{k+2^{Q-q-1},l}^q\right). \end{aligned}$$

Start der Iteration (q=0): $\boxed{c_{k,0}^0 = y_k}$ für $k = 0, \ldots, 2^Q - 1$.

Speicherbedarf: Für q und $q+1$ müssen Matrizen berechnet werden:

$$C^q = (c_{k,l}^q), \ C^{q+1} = (c_{k,l}^{q+1})$$

$C^q \in \mathbb{C}^{2^{Q-q} \times 2^q}$ bzw. $C^{q+1} \in \mathbb{C}^{2^{Q-q-1} \times 2^{q+1}}$

Beide Matrizen sind von der selben Dimension: $2^{Q-q}2^q = 2^Q = n+1$ und $2^{Q-q-1}2^{q+1} = 2^Q = n+1$

Daher sollen die Koeffizienten $c_{k,l}^q, \ c_{k,l}^{q+1}$ in Vektoren der Dimension $n+1$ gespeichert werden

$$C[2^q k + l] := c_{k,l}^q,$$

$$D[2^{q+1} k + l] := c_{k,l}^{q+1}.$$

Es gilt: $e^{-i\frac{2\pi}{2^{q+1}}l} = e^{-il\frac{2\pi}{2^Q}2^{Q-q-1}l} =: W[2^{Q-q-1}l]$, wobei der Vektor $W[l] := e^{-i\frac{2\pi}{2^Q}l}, \ l = 0, \ldots, 2^Q - 1$ vorab nur einmal berechnet werden muss.

Mit $\hat{w} := e^{-i\frac{2\pi}{2^Q}}$ erhalten wir dann folgenden Algorithmus:

Algorithmus 1.33 (FFT)

Für $l = 0, \dots, 2^Q - 1$:

$q = 0$ | $C[l] = y_l$; $W[l] = \hat{w}^l$

Für $q = 0, \dots, Q-1$

$q \longrightarrow q+1$ | Für $k = 0, \dots, 2^{Q-(q+1)} - 1$

Für $l = 0, \dots, 2^q - 1$

$(*)$ | $u = C[2^q k + l]$
$v = W[2^{Q-q-1} l] C[2^q (k + 2^{Q-q-1}) + l]$
$D[2^{q+1} k + l] = \frac{1}{2}(u+v)$
$D[2^{q+1} k + l + 2^q] = \frac{1}{2}(u-v)$

Für $l = 0, \dots, 2^Q - 1$

$C[l] = D[l]$

Aufwand: $(*)$ benötigt 3 Operationen. Anzahl der Durchläufe von $(*)$

$$Q 2^{Q-q-1} 2^q = Q 2^{Q-1} = \log_2(n+1) \frac{n+1}{2}.$$

Daher ist der gesamte Aufwand gleich

$$3 \log_2(n+1) \frac{n+1}{2} = O(n \log_2 n).$$

Bemerkung: Der Algorithmus kann so umgeschrieben werden, dass der Vektor D nicht gebraucht wird. Es existieren auch Varianten für den Fall $n \neq 2^Q - 1$.

1.7 Spline-Interpolation

Motivation: Bei großen Werten von n führt die Polynominterpolation zu stark oszillierenden Interpolationspolynomen, da $p_n \in C^\infty(I)$. Das Problem tritt besonders dann auf, wenn die Stützstellen vorgegeben sind. Daher verwendet man häufig stückweise polynomielle Funktionen, d.h.

$$P|_{[x_{i-1}, x_i]} \in \mathbb{P}_r$$

mit $r \ll n$. Die Interpolationsbedingung $p(x_i) = y_i$ führt zu $p \in C^0(I)$, aber p ist i.a. nicht in $C^\infty(I)$, sondern $p \in C^q(I)$. Die Parameter (r, q) sind geeignet zu wählen:

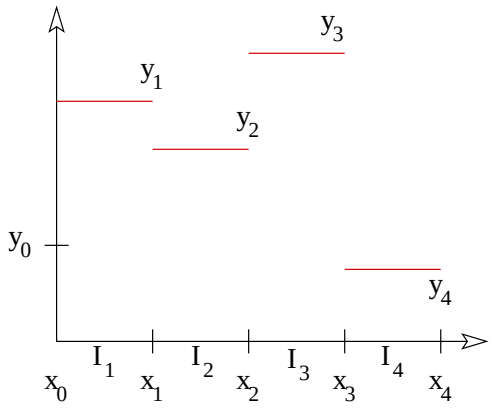


Abbildung 1.5: Beispiel 1.34: Treppenfunktionen

$$P|_{[x_{i-1},x_i]} \in \mathbb{P}_r, \quad p(x) = \begin{cases} p_1(x) & : \quad x \in (x_0, x_1] \\ p_2(x) & : \quad x \in (x_1, x_2] \\ \vdots & \\ p_n(x) & : \quad x \in (x_{n-1}, x_n] \end{cases}$$

$p_i \in \mathbb{P}_r$ hat die Interpolationsbedingungen $p_i(x_{i-1}) = y_{i-1},\ p_i(x_i) = y_i \iff p(x_k) = y_k,\ k = 0, \ldots, n$.

Notation: $\Delta = (x_0, \ldots, x_n)$ ist eine Zerlegung von $I = [a, b]$ mit $x_0 = a,\ x_n = b,\ x_{i-1} < x_i$ $(1 \leq i \leq n)$.
Mit $h_i := x_i - x_{i-1} > 0$ bezeichnen wir die Länge des Teilintervals $I_i := (x_{i-1}, x_i),\ I_0 := \{a\},\ i = 1, \ldots, n$. Die Feinheit der Zerlegung ist gegeben durch:

$$h = \max_{1 \leq i \leq n} h_i.$$

Für $r, q \in \mathbb{N}$ definieren wir den Raum der Splines durch

$$S_\Delta^{r,q} := \left\{ P \in C^q(I) \;\middle|\; P|_{I_i} =: p_i \in \mathbb{P}_r \text{ für } 1 \leq i \leq n \right\}$$

Gegeben: Zerlegung Δ, Daten $y_0, \ldots, y_n$ und $r, q \in \mathbb{N}$.

Gesucht: $P_\Delta \in S_\Delta^{r,q}$ mit $P_\Delta(x_k) = y_k,\ k = 0, \ldots, n$.

Beispiel 1.34
$\underline{r = 0}$: Die einzig mögliche Interpolation durch stückweise konstante Funktionen ist gegeben durch $P_\Delta(x) = y_i$ für $x \in I_i$ bzw. $p_i(x) = y_i$. Für $q \geq 0$ ist das Problem nicht lösbar.
Abbildung 1.5 zeigt die entstandene Treppenfunktion. In diesem Fall ist P_Δ nicht stetig!

$\underline{r = 1}$: Es soll gelten: $p_i \in \mathbb{P}_1$ und $p_i(x_{i-1}) = y_{i-1},\ p(x_i) = y_i$.
Abbildung 1.6 zeigt die eindeutig bestimmten Funktionen p_i definiert durch $p_i(x) = y_i + \frac{y_i - y_{i-1}}{h_i}(x - x_i)$. Damit gilt: $P_\Delta \in S_\Delta^{1,0}$.

$\underline{r=3}$: Annahme: $y_k = f(x_k)$ mit $f \in C^4(I)$

(i) **Fall**: Wähle für $i = 1, \ldots, n$ Werte $x_{ij} \in I_i$ für $j = 1, 2$ und definiere p_i als Interpolationspolynom zu $(x_{i-1}, y_{i-1}), (x_{i1}, f(x_{i1})), (x_{i2}, f(x_{i2})), (x_{i+1}, f(x_{i+1})) \implies p_i \in \mathbb{P}_3$ und $P_\Delta \in S_\Delta^{3,0}$. Nach Satz 1.4 gilt:

$$\begin{aligned} |f(x) - P_\Delta(x)| &= |f(x) - p_i(x)| = f^{(4)}(\xi_x)\tfrac{1}{4!}h_i^4 \text{ für } x \in I_i \\ &\leq \left\|f^{(4)}\right\|_\infty \tfrac{1}{4!}h^4. \end{aligned}$$

(ii) **Fall**: Wähle $p_i \in \mathbb{P}_3$ durch Hermiteinterpolation zu (x_{i-1}, y_{i-1}), $(x_{i-1}, f'(x_{i-1}))$, (x_i, y_i), $(x_i, f'(x_i)) \implies P_\Delta \in S_\Delta^{3,1}$ und $\|f - P_\Delta\|_\infty \leq h^4 \frac{1}{4!} \left\|f^{(4)}\right\|_\infty$.

Frage: Existiert ein $P_\Delta \in S_\Delta^{3,2}$?

Bemerkung: Für $n > r$ ist das Interpolationsproblem in $S_\Delta^{r,q}$ für $q \geq r$ i.a. schlecht gestellt (d.h. nicht lösbar):

Freiheitsgrade: $p_i \in \mathbb{P}_r$ führt auf $(r+1)$ Koeffizienten, also: $n(r+1)$ Freiheitsgrade.

Anzahl der Bedingungen:

Auf I_1 : 2 Interpolationsbedingungen
I_2 : 2 Interpolationsbedingungen + q Stetigkeitsbedingungen in x_1
$\vdots$
I_n : 2 Interpolationsbedingungen + q Stetigkeitsbedingungen in x_{n-1}

$\implies$ $2n + q(n-1) = n(q+2) - q$ Bedingungen.

Ist $q \geq r$ so folgt: $2n + q(n-1) \geq 2n + r(n-1) = n(r+1) + n - r > n(r+1)$. Für $n - r > 0$ existieren also mehr Bedingungen als Freiheitsgrade und das Problem ist i.A. nicht lösbar.

Spezialfall: $q = r - 1$ (Eigentliche "Spline-Interpolation")

Bedingungen: $n(q+2) - q = n(r+1) - q$, d.h. es müssen noch $q = r - 1$ Freiheitsgrade zusätzlich festgelegt werden.

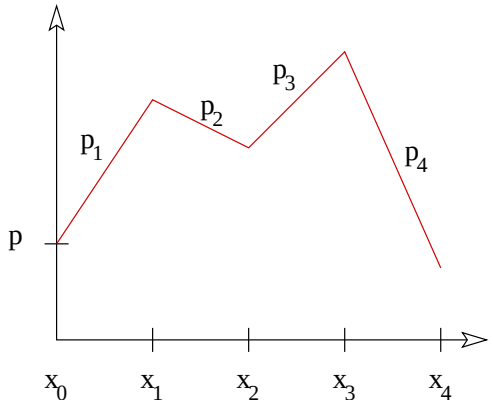


Abbildung 1.6: Beispiel 1.34: Gerade

1.7.1 Kubische Spline-Interpolation

Gegeben: $\Delta = (x_0, \ldots, x_n)$ Zerlegung des Intervalls $I = [a, b]$ und Daten $y_0, \ldots, y_n \in \mathbb{R}$.

Gesucht: $P_\Delta \in S^{3,2}_\Delta$ mit $P_\Delta(x_i) = y_i$ $(0 \le i \le n)$ und eine der Bedingungen a) bis d):

a) $P''_\Delta(a) = M_a,\ P''_\Delta(b) = M_b$ für $M_a, M_b \in \mathbb{R}$ gegeben.
Im Fall $M_a = M_b = 0$ spricht man von **natürlichen kubischen Splines**.

b) $p'(a) = g_a,\ p'(b) = g_b$ für $g_a, g_b \in \mathbb{R}$ gegeben.

c) P_Δ sei **periodisch fortsetzbar** in $\mathbb{C}^2(\mathbb{R})$, d.h. $y_0 = y_n$ und $p'(a) = p'(b),\ p''(a) = p''(b)$.

d) **not-a-knot**-Bedingung: $P_{\Delta|_{[I_1 \cup I_2]}} \in \mathbb{P}_3,\ P_{\Delta|_{[I_{n-1} \cup I_n]}} \in \mathbb{P}_3$, d.h. die Zusatzbedingungen werden verwendet, um die Sprünge in P'''_Δ für $x = x_1,\ x = x_{n-1}$ zu eliminieren.

Satz 1.35 (Existenz und Eindeutigkeit)
Zu gegebener Zerlegung Δ und Daten $y_0, \ldots, y_n$ existiert genau ein $P_\Delta \in S^{3,2}_\Delta$ mit $p(x_k) = y_k$, welches eine der Bedingungen a), b), c), oder d) erfüllt. Im Fall c) muss gelten: $y_0 = y_n$.

Beweis: **Idee:** Stelle LGS für die Momente $M_j := P''_\Delta(x_j)$ auf.

Da p''_j linear auf $I_j = (x_{j-1}, x_j]$ ist, muß gelten: $p''_j(x) = \frac{1}{h_j}\left(M_j(x - x_{j-1}) + M_{j-1}(x_j - x)\right)$.

Durch zweimalige Integration folgt für geeignete Integrationskonstanten $a_j, b_j \in \mathbb{R}$:

$$p_j(x) = \frac{1}{6h_j}\left(M_j(x - x_{j-1})^3 + M_{j-1}(x_j - x)^3\right) + b_j\left(x - \frac{x_j + x_{j-1}}{2}\right) + a_j. \quad (*)$$

Aus den Interpolationsbedingungen $p_j(x_{j-1}) = y_{j-1},\ p_j(x_j) = y_j$ folgt:

$$\begin{aligned} y_{j-1} &= \tfrac{1}{6h_j} M_{j-1} h_j^3 - b_j \tfrac{1}{2} h_j + a_j, \\ y_j &= \tfrac{1}{6h_j} M_j h_j^3 + b_j \tfrac{1}{2} h_j + a_j. \end{aligned}$$

Dies ist ein 2×2 LGS für a_j, b_j mit der Lösung

$$\begin{aligned} (**)\ a_j &= \tfrac{1}{2}\left(y_j + y_{j-1}\right) - \tfrac{1}{12} h_j^2 \left(M_j + M_{j-1}\right), \\ b_j &= \tfrac{1}{h_j}\left(y_j - y_{j-1}\right) - \tfrac{1}{6} h_j \left(M_j - M_{j-1}\right). \end{aligned}$$

Damit hängen die p_j nur von den Momenten $M_0, \ldots, M_n$ ab.

Es bleiben noch die $n - 1$ Bedingungen $p'(x_j) = p'_{j+1}(x_j)$ für $j = 1, \ldots, n-1$:
Aus $(*)$ und $(**)$ folgt: $p'_j(x) = \frac{1}{2h_j}\left(M_j\left(x - x_{j-1}\right)^2 - M_{j-1}\left(x_j - x\right)^2\right) + \frac{1}{h_j}\left(y_j - y_{j-1}\right) - \frac{1}{6} h_j \left(M_j - M_{j-1}\right)$.

Daher ist $p'_j(x_j) = p'_{j+1}(x_j)$ äquivalent zu
$\frac{1}{2} M_j \left(h_{j+1} + h_j\right) + \frac{1}{6} h_{j+1}\left(M_{j+1} - M_j\right) - \frac{1}{6} h_j \left(M_j - M_{j-1}\right) = \frac{1}{h_{j+1}}\left(y_{j+1} - y_j\right) - \frac{1}{h_j}\left(y_j - y_{j-1}\right)$ für
$j = 1, \ldots, n-1$,
bzw.
$\frac{1}{6} h_j M_{j-1} + \frac{1}{3}\left(h_j + h_{j+1}\right) M_j + \frac{1}{6} h_j M_{j+1} = y[x_j, x_{j+1}] - y[x_{j-1}, x_j]$.

Mit der zweiten dividierten Differenz $y[x_{j-1}, x_j, x_{j+1}] = \frac{y[x_j, x_{j+1}] - y[x_{i-1} - x_j]}{x_{j+1} - x_{j-1}}$ und $x_{j+1} - x_{j-1} = h_j + h_{j+1}$ folgt:

$$\mu_j M_{j-1} + M_j + \lambda_j M_{j+1} = 3y[x_{j-1}, x_j, x_{x+1}]$$

mit $\mu_j = \frac{h_j}{2(h_j + h_{j+1})}$, $\lambda_j = \frac{h_{j+1}}{2(h_j + h_{j+1})}$.

Wir erhalten ein $(n-1) \times (n-1)$ LGS für die (n+1) Momente $M_0, \ldots, M_n$.

Fall a) $M_0 = M_a$, $M_n = M_b$.

Dies führt auf das $(n-1) \times (n-1)$ LGS für $M_1, \ldots, M_{n-1}$ der Form

$$A \begin{pmatrix} M_1 \\ \vdots \\ M_{n-1} \end{pmatrix} = \begin{pmatrix} 3y[x_0, x_1, x_2] - \mu_1 M_a \\ 3y[x_1, x_2, x_3] \\ \vdots \\ 3y[x_{n-2}, x_{n-1}, x_n] - \lambda_{n-1} M_b \end{pmatrix}$$

mit

$$A = \begin{pmatrix} 1 & \lambda_1 & & 0 \\ \mu_2 & \ddots & \ddots & \\ & \ddots & \ddots & \lambda_{n-2} \\ 0 & & \mu_{n-1} & 1 \end{pmatrix}.$$

A ist regulär nach dem folgenden Lemma 1.36, da $\mu_j + \lambda_j = 1/2 < 1$ und $\lambda_1 < 1$, $\mu_{n-1} < 1$.

Die Fälle b),c),d) führen analog auf einfach strukturierte LGS mit regulären Matrizen, d.h. $p_1, \ldots, p_n$ eindeutig duch $(*), (**)$ festgelegt. □

Lemma 1.36
Sei $A \in \mathbb{R}^{n \times n}$ eine **tridiagonale Matrix**, d.h.

$$A = \text{tridiag}(b_i, a_i, c_i) = \begin{pmatrix} a_1 & c_1 & & 0 \\ b_2 & \ddots & \ddots & \\ & \ddots & \ddots & c_{n-1} \\ 0 & & b_n & a_n \end{pmatrix}$$

Es gelte: $|a_1| > |c_1| > 0$ und $|a_n| > |b_n| > 0$ und $|a_i| \geq |b_i| + |c_i|$, $b_i \neq 0$, $c_i \neq 0$, $2 \leq i \leq n-1$.

Dann gilt:

(i) A ist regulär.

(ii) $A = LR$ mit $L = \text{tridiag}(b_i, \alpha_i, 0)$ und $R = \text{tridiag}(0, 1, \gamma_i)$ mit $\alpha_1 = a_1$, $\gamma_1 = c_1 \alpha_1^{-1}$ und für $2 \leq i \leq n$: $\alpha_i = a_i - b_i \gamma_{i-1}$, $\gamma_i = c_i \alpha_i^{-1}$.

Daher kann $Ax = b$ in $O(n)$ Operationen gelöst werden.

Beweis: Siehe Übungsaufgaben □

Lemma 1.37
Die Spline-Interpolation mit kubischen Splines und einer der Zusatzbedingungen a), b), c) oder d) kann mit $O(n)$ Operationen gelöst werden.

Beweis: a) folgt aus 1.35, 1.36.
b), c), d): (siehe z.B. Schaback,Werner: Numerische Mathematik, Berlin, Springer, 1992.) □

Historisch: Interpolation durch biegsamen Stab (engl: spline) und Brett mit Nägeln bei (x_k, y_k). Der Stab hat minimale Krümmung, d.h. die Funktion minimiert $\int_I \frac{(y''(t))^2}{1+(y'(t))^2} \mathrm{d}t$ über alle glatten Funktionen y mit $y(x_k) = y_k$. Für den Fall kleiner erster Ableitungen entspricht dies näherungsweise $\int_I y''(t)^2 \mathrm{d}t$.

Satz 1.38 (Minimierungseigenschaft kubischer Splines)
Sei $\Delta = (x_0, \ldots, x_n)$ eine Zerlegung von $I = [a, b]$ und $y_0, \ldots, y_n \in \mathbb{R}$ gegeben. Sei $P_\Delta \in S_\Delta^{3,2}$ ein kubischer Spline mit $P_\Delta(x_k) = y_k$ und einer der Bedingungen a), b), oder c):

a) $P''_\Delta(a) = 0,\ P''_\Delta(b) = 0,$

b) $P'_\Delta(a) = g_a,\ P'_\Delta(b) = g_b,$

c) P_Δ periodisch fortsetzbar in $C^2(\mathbb{R})$.

Dann gilt für alle $f \in C^2(a, b)$ mit denselben Interpolationsbedingungen, d.h. mit $f(x_k) = y_k$ und a), b) oder c) und $\int_a^b |f''|^2 \leq \infty$:

$$\int_a^b \left|f''(x)\right|^2 \mathrm{d}x \geq \int_a^b \left|P''_\Delta(x)\right|^2 \mathrm{d}x.$$

Beweis: Zum Beweis dieser Aussage benötigen wir das folgende Lemma. □

Lemma 1.39 (Holladay Identität)
Sei $f \in C^2(a, b)$ mit $\int_a^b |f'''|^2 < \infty$ und $P_\Delta \in S_\Delta^{3,2}$, dann gilt:

$$\int_a^b \left|f'' - P''_\Delta\right|^2 = \int_a^b \left|f''\right|^2 - \int_a^b \left|P''_\Delta\right|^2$$
$$-2\left(\left[(f'(x) - P'_\Delta(x))P''_\Delta(x)\right]_{x=a}^{b} - \sum_{i=1}^{n}\left[(f(x) - P_\Delta(x))P'''_\Delta(x)\right]_{x=x_{i-1}^+}^{x_i^-}\right).$$

Dabei wurden die folgenden Abkürzungen benutzt:

$$[g(x)]_{x=a}^{b} = g(b) - g(a),$$

$$[g(x)]_{x=x_{i-1}^+}^{x_i^-} = \lim_{x \nearrow x_i} g(x) - \lim_{x \searrow x_{i-1}} g(x). \text{ Beachte : } P'''_\Delta \text{ ist unstetig!}$$

Beweis: Es ist

$$\begin{aligned}\int_a^b |f''-P''_\Delta|^2 &= \int_a^b |f''|^2 - 2\int_a^b f''P''_\Delta + \int_a^b |P''_\Delta|^2\\ &= \int_a^b |f''|^2 - \int_a^b |P''_\Delta|^2 - 2\int_a^b (f''-P''_\Delta)P''_\Delta\\ &= \int_a^b |f''|^2 - \int_a^b |P''_\Delta|^2 - 2\sum_{i=1}^n \underbrace{\int_{I_i}(f''-p_i'')p_i''}_{=:A_i}.\end{aligned}$$

Mit partieller Integration folgt für A_i:

$$\begin{aligned}A_i &= \int_{x_{i-1}}^{x_i}(f''-p_i'')p_i'' = [(f'-p_i')p_i'']_{x=x_{i-1}}^{x_i} - \int_{x_{i-1}}^{x_i}(f'-p_i')p_i'''\\ &= [(f'-p_i')p_i'']_{x=x_{i-1}}^{x_i} - [(f-p_i)p_i''']_{x_{i-1}^+}^{x_i^-} + \int_{x_{i-1}}^{x_i}(f-p_i)p_i^{(4)}.\end{aligned}$$

Es ist $p_i^{(4)} \equiv 0$, da $p_i \in P_3$ und

$$\begin{aligned}\sum_{i=1}^n [(f'-p_i')p_i'']_{x=x_{i-1}}^{x_i} &\overset{p_i''\in C^0}{=} \sum_{i=1}^n [(f'-P'_\Delta]P''_\Delta]_{x=x_{i-1}}^{x_i}\\ &= \sum_{i=1}^n [(f'(x_i)-p'_\Delta(x_i))P''_\Delta(x_i) - (f'(x_{i-1}-P'_\Delta(x_{i-1}))P''_\Delta(x_{i-1})]\\ &= (f'(x_n)-P'_\Delta(x_n))P''_\Delta(x_n) - f'(x_0)-P'_\Delta(x_0))P''_\Delta(x_0)\\ &= [(f'(x)-P'_\Delta(x))P''_\Delta(x)]_{x=a}^b.\end{aligned}$$

$$\implies \sum_{i=1}^n A_i = [(f'(x)-P'_\Delta(x))P''_\Delta(x)]_{x=a}^b - \sum_{i=1}^n [(f(x)-p_i(x))p_i'''(x)]_{x=x_{i-1}}^{x_i}.$$

Also folgt die Holladay Identität. □

Beweis: (Fortsetzung des Beweises von Satz 1.38)
In den 3 Fällen a), b), c) verschwindet der Term $2(\cdots)$ in der Holladay Identität $\implies$ $0 \le \int_a^b |f''-P''_\Delta|^2 = \int_a^b |f''|^2 - \int_a^b |P''_\Delta|^2$. □

Satz 1.40 (Fehlerabschätzung)
Sei Δ eine Zerlegung von I mit $h \le Kh_i$ $(1 \le i \le n)$ für ein $K > 0$. Sei $f \in C^4(a,b)$ mit $\left|f^{(4)}\right| < L$ für $x \in (a,b)$.
Sei $P_\Delta \in S_\Delta^{3,2}$ mit $P_\Delta(x_k) = f(x_k)$ und $P'_\Delta(a) = f'(a),\ P'_\Delta(b) = f'(b)$.

Dann gilt für $l = 0,1,2,3:\ \left|f^{(l)}(x) - P_\Delta^{(l)}(x)\right| \le 2LKh^{4-l}$.
Also insbesondere

$$|f(x) - P_\Delta(x)| \le 2LKh^4.$$

Beweis: (Ohne Beweis. Siehe z.B. Stoer, Bulirsch. Numerische Mathematik 1. Berlin, Springer 2007.) □

Basiswahl für den Splineraum $\mathbf{S_\Delta^{r,r-1}}$: B-Splines

Ziel: Konstruktion einer einfachen Basis von $S_\Delta^{r,r-1}$ mit

1. positiven Basisfunktionen für numerische Stabilität,
2. möglichst kleinem Träger.

Definition 1.41 (B-Splines)
Sei $(t_i)_{i\in\mathbb{Z}}$ *eine monoton nicht-fallende Folge mit* $\lim_{i \longrightarrow \pm\infty} t_i = \pm\infty$. *Dann sind die* **B-Splines** $B_{i,k} : \mathbb{R} \to \mathbb{R}$ vom Grad $k \in \mathbb{N}$ rekursiv definiert durch

$$B_{i,0}(x) = \begin{cases} 1 & : \quad t_i < x \leq t_{i+1} \\ 0 & : \quad \text{sonst} \end{cases}$$

und

$$B_{i,k} = \omega_{i,k}(x) B_{i,k-1}(x) + (1 - \omega_{i+1,k}(x)) B_{i+1,k-1}(x)$$

mit

$$\omega_{i,k}(x) = \begin{cases} \frac{x-t_i}{t_{i+k}-t_i} & : \quad t_i < t_{i+k} \\ 0 & : \quad \text{sonst} \end{cases} .$$

Beispiel:

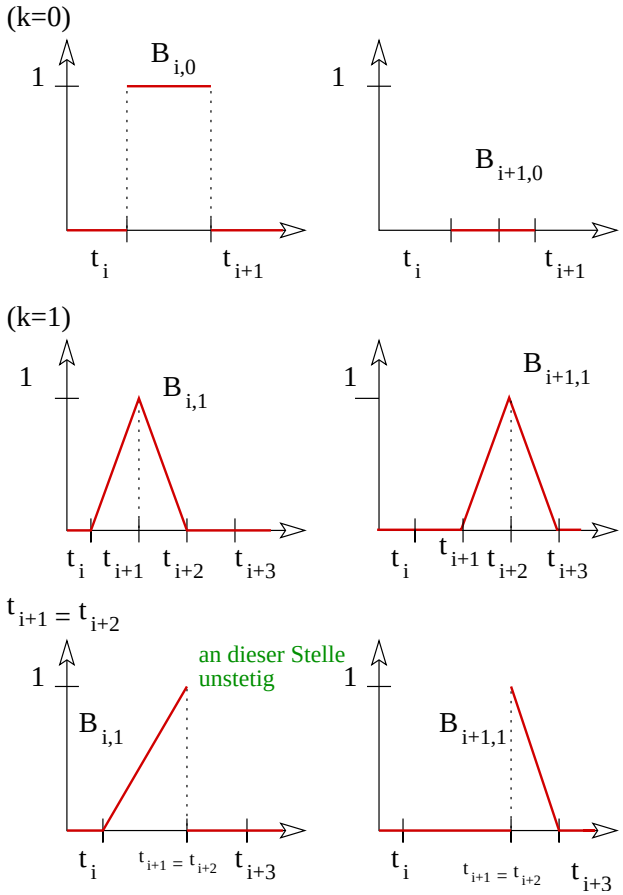


Abbildung 1.7: B-Splines

Die Abb. 1.7 zeigt 6 verschiedene Beispiele, die bei den B-Splines auftreten können.

Satz 1.42 (Eigenschaften der B-Splines)
Sei $(t_i)_{i\in\mathbb{Z}}$ eine monoton nicht-fallende Knotenfolge, wie in Definition 1.41. Dann gilt:

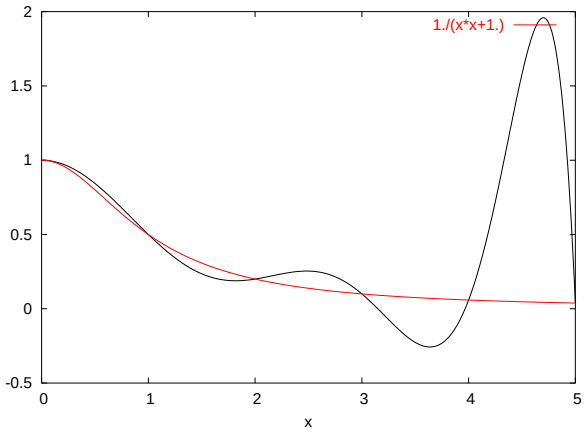
(i) $B_{i,k}|_{[t_j,t_{j+1}]} \in \mathbb{P}_k \ \forall\ i,j \in \mathbb{Z}, k \in \mathbb{N}$,

(ii) $supp(B_{i,k}) \subset [t_i, t_{i+k+1}]$, falls $t_i < t_{i+k+1}$ und $B_{i,k} \equiv 0$, falls $t_i = t_{i+k+1}$,

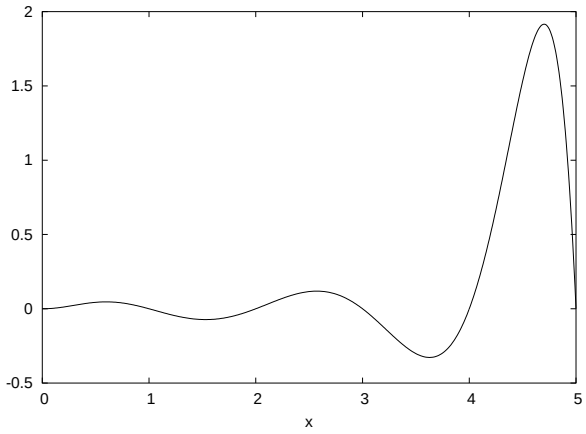
(iii) $B_{i,k} \geq 0,\ \sum\limits_{i\in\mathbb{Z}} B_{i,k}(x) = 1,\ \forall\ x \in \mathbb{R}$ (Zerlegung der 1).

(iv) Falls $\forall\ i \in \mathbb{Z}:\ t_i < t_{i+1}$, dann ist $B_{i,k} \in C^{k-1}$ und $(B_{i,k})_{i\in\mathbb{Z}}$ bildet eine Basis von $S_\Delta^{k,k-1}$.

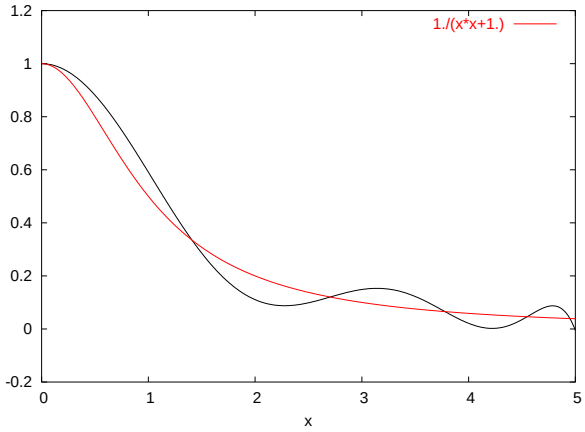
Beweis: (Ohne Beweis) □
Beispiel Ein Vergleich der angesprochenen Interpolationen für $f(x) = \frac{1}{x^2+1}$ ist in Abb. 1.8 dargestellt. Die Spline-Interpolation ergibt hier das beste Ergebnis.



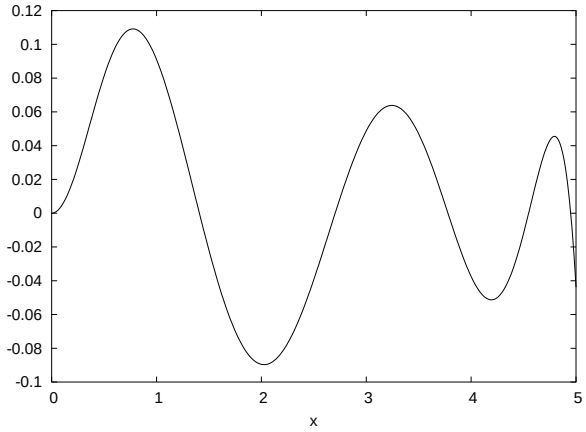
Polynomielle mit gleichmäßig verteilten Stützstellen



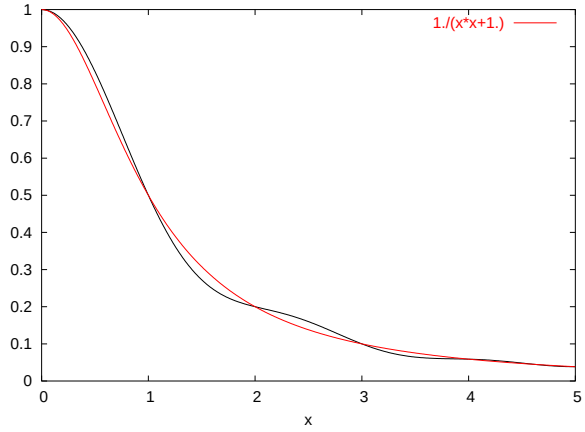
Fehler der Interpolation



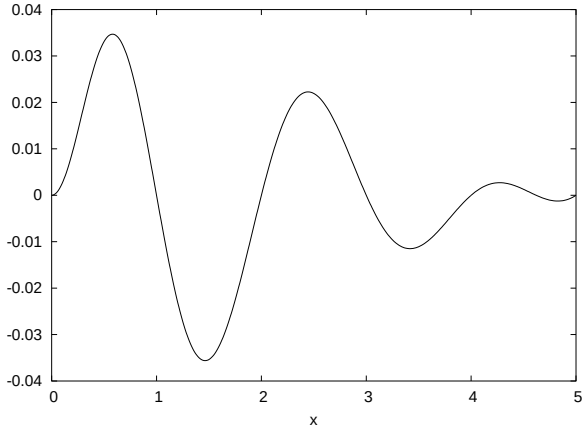
Tschebyschev-Interpolation



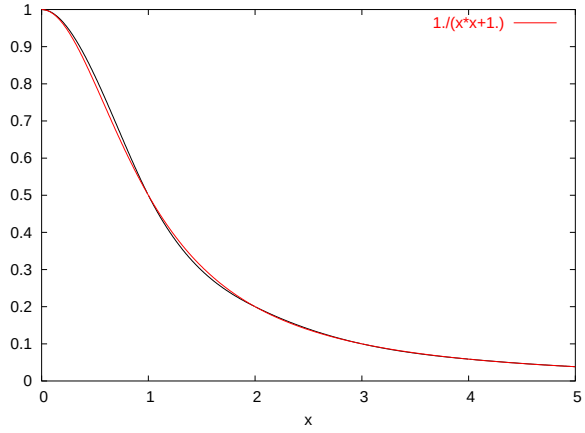
Fehler der Interpolation



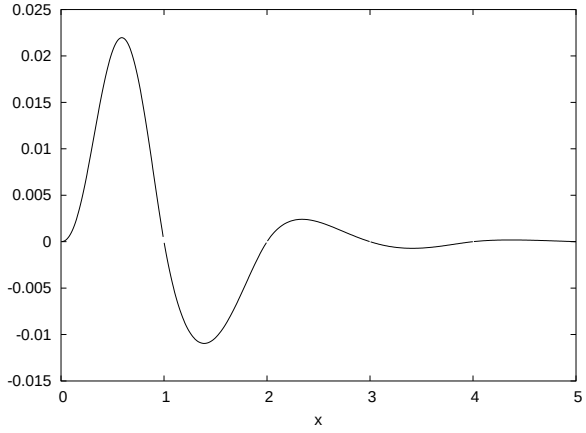
Trigonometrische Interpolation



Fehler der Interpolation



Spline-Interpolation



Fehler der Interpolation

Abbildung 1.8: Unterschiede einiger Interpolationen

Kapitel 2

Numerische Integration

Ziel: Approximation von

$$I(f) := \int_a^b \omega(x) f(x) \mathrm{d}x$$

für $f \in C^k(a,b)$ und für eine gegebene Gewichtsfunktion $\omega \in L^1(a,b)$.

Ansatz: Approximiere $I(f)$ durch eine Summe

$$I_n(f) := \sum_{j=0}^{m} \sum_{l=0}^{m_j - 1} f^{(l)}(x_j) \omega_j^l$$

Definition 2.1 (Quadratur)
Eine Funktional $I_n : C^k(a,b) \longrightarrow \mathbb{R}$ *der Form*

$$I_n(f) := \sum_{j=0}^{m} \sum_{l=0}^{m_j - 1} f^{(l)}(x_j) \omega_j^l$$

heißt **Quadraturformel** *mit den Stützstellen* $x_j \in [a,b]$ *und den* **Gewichten** $\omega_j^l \in \mathbb{R}$. *Dabei ist* $m \in \mathbb{N}$ *und* $m_j \in \{1, \ldots, k+1\}$ *und* $n+1 = \sum_{j=0}^{m} m_j$.

Die Quadratur heißt **exakt** *für* $\mathbb{P}_n$ *(bezüglich* ω*), g.d.w.*

$$I_n(p) = I(p) \ \forall \ p \in \mathbb{P}_n$$

.

$$R(f) = I(f) - I_n(f)$$

ist das zu I_n *gehörende* **Fehlerfunktional**.

Bemerkung: Für die allgemeine Definition der Quadratur vergleiche mit der Definition der Hermite Interpolation. Im folgenden betrachten wir meistens Quadraturen der Form

$$I_n(f) = \sum_{l=0}^{n} \omega_l f(x_l), \text{ d.h. } m_j = 1.$$

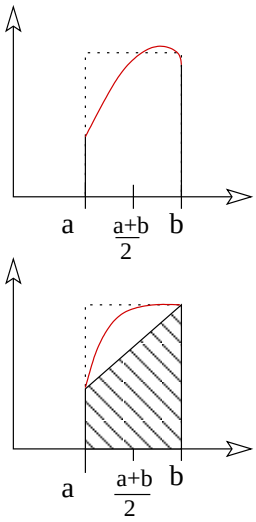


Abbildung 2.1: Beispiel 2.1

Beispiel 2.2 ($\omega \equiv 1$)
Die Abbildung 2.1 verdeutlicht diese Beispiele:

(i) Mittelpunktregel: $I_0(f) = (b-a) f\left(\frac{a+b}{2}\right)$.

(ii) Trapezregel: $I_1(f) = \frac{b-a}{2}\left(f(a) + f(b)\right)$.

(iii) Simpsonregel: $I_2(f) = \frac{b-a}{6}\left(f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right)$.

Satz 2.3
Gegeben seien $\omega \in L^1(a,b)$ und paarweise disjunkte Stützstellen $x_0, \ldots, x_n$. Dann existiert genau eine Quadraturformel der Form

$$I_n(f) = \sum_{j=0}^{n} \omega_j f(x_j),$$

welche exakt ist auf $\mathbb{P}_n$. Dabei sind die Gewichte gegeben durch

$$\omega_j := \int_a^b \omega(x) L_j^n(x) \mathrm{d}x,$$

wobei $L_j^n(x) = \prod_{\substack{l=0 \\ l \neq j}}^{n} \frac{(x-x_l)}{(x_j - x_l)}$ die Lagrange Polynome sind.

Beweis: I_n exakt auf $\mathbb{P}_n$
$\iff$ $I_n(p) = I(p)\ \forall p \in \mathbb{P}_n$
$\iff$ $I_n(L_l^n) = I(L_l^n)$ für $l = 0, \ldots, n$; da I_n, I linear und L_l^n Basis von $\mathbb{P}_n$
$\iff$ $\int_a^b \omega(x) L_l^n(x) = \sum_{j=0}^{n} \omega_j L_l^n(x_j) = \omega_l$, da $L_l^n(x_j) = \delta_{lj}$. □

Bemerkung: Es ist $I_n(f) = I(p_n)$, wobei $p_n \in \mathbb{P}_n$ das eindeutig bestimmte Interpolationspolynom zu $(x_0, f(x_0)), \ldots, (x_n, f(x_n))$ ist:

$$I_n(f) = \sum_{l=0}^{n} \omega_l f(x_l) = \sum_{l=0}^{n} \int_a^b \omega(x) L_l^n(x) f(x_l) \mathrm{d}x$$
$$= \int_a^b \omega(x) \underbrace{\sum_{l=0}^{n} L_l^n(x) f(x_l)}_{p_n(x)} \mathrm{d}x = \int_a^b \omega(x) p_n(x) \mathrm{d}x = I(p_n).$$

Definition 2.4
Eine Quadraturformel $I_n(f) = \sum_{l=0}^{n} \omega_l f(x_l)$ *zu gegebenen Stützstellen* $a \leq x_0 < x_1 < \ldots < x_n \leq b$ *und Gewichtsfunktion* $\omega \in L^1(a,b)$ *heißt* **Interpolationsquadratur**, *wenn sie auf* $\mathbb{P}_n$ *exakt ist. Nach Satz 2.3 ist sie eindeutig.*

Satz 2.5
Seien $x_0, \ldots, x_n \in [a,b]$ und $\omega \in L^1(a,b)$ gegeben mit den Symmetrieeigenschaften

(i) $x_j - a = b - x_{n-j}$ $(0 \leq j \leq n)$ (Symmetrie bzgl. $\frac{a+b}{2}$)

(ii) $\omega(x) = \omega(a+b-x)$ $(x \in [a,b])$ (gerade Funktion bzgl. $\frac{a+b}{2}$)

Dann gilt $\omega_{n-j} = \omega_j$ $(0 \leq j \leq n)$, d.h. die Interpolationsquadratur ist symmetrisch. Falls n gerade ist, so ist I_n exakt auf $\mathbb{P}_{n+1}$.

Beweis: Sei $\tilde{I}_n(f) := \sum_{j=0}^{n} \omega_{n-j} f(x_j)$. Dann gilt $\tilde{I}_n(p) = I_n(p) \; \forall \; p \in P_n$. Damit ist aber $\tilde{I}_n$ exakt auf P_n und nach Satz 2.3 gilt $\tilde{I}_n = I_n$ und folglich $\omega_{n-j} = \omega_j$.

Sei nun $n = 2m$ und damit $x_m = \frac{a+b}{2}$ wegen i). Sei $p_n \in \mathbb{P}_n$ das Interpolationspolynom zu $(x_0, f(x_0)), \ldots, (x_n, f(x_n))$ und sei $q_{n+1} \in \mathbb{P}_{n+1}$ das Hermite Interpolationspolynom zu $(x_0, f(x_0))$, $\ldots$, $(x_{m-1}, f(x_{m-1}))$, $(x_m, f(x_m)), (x_m, f'(x_m))$, $(x_{m+1}, f(x_{m+1})), \ldots, (x_n, f(x_n))$. Mit

$$c := \frac{f'(x_m) - p_n'(x_m)}{\prod\limits_{\substack{l=0 \\ l \neq m}}^{n} (x_m - x_l)}$$

und $N(x) = \prod_{l=0}^{n} (x - x_l) \in \mathbb{P}_{n+1}$ definiere

$$\tilde{q}_{n+1}(x) := p_n(x) + cN(x).$$

Dann ist $\tilde{q}_{n+1} \in \mathbb{P}_{n+1}$ und $\tilde{q}_{n+1}(x_l) = p_n(x_l) + cN(x_l) = f(x_l) + 0$. Weiter folgt

$$\tilde{q}_{n+1}'(x_m) = p_n'(x_m) + c \prod_{\substack{l=0 \\ l \neq m}}^{n} (x_m - x_l) = f'(x_m).$$

Wegen der Eindeutigkeit der Hermite Interpolation gilt daher $q_{n+1} = \tilde{q}_{n+1}$.

Es gilt wegen i): $N(x) = \prod_{l=0}^{m-1} (x - x_l) \left(x - \frac{a+b}{2}\right) \prod_{l=0}^{m-1} (x - (a+b-x_l))$ und somit folgt $N(a+b-x) = (-1)^{n+1} N(x) = -N(x)$.

Wegen ii) gilt damit:

$$\begin{aligned}\int_a^b \omega(x)N(x)\mathrm{d}x &= \int_a^{x_m} \omega(x)N(x)\mathrm{d}x + \int_{x_m}^b \omega(x)N(x)\mathrm{d}x \\ \overset{t=a+b-x}{=} &\int_a^{x_m} \omega(x)N(x)\mathrm{d}x - \int_{x_m}^a \omega(a+b-t)N(a+b-t)\mathrm{d}t \\ = &\int_a^{x_m} \omega(x)N(x)\mathrm{d}x + \int_a^{x_m} \omega(t)(-N(t))\mathrm{d}t = 0.\end{aligned}$$

Wir erhalten: $\int_a^b q_{n+1}(x)\omega(x)\mathrm{d}x = \int_a^b p_n(x)\omega(x)\mathrm{d}x = I(p_n) = I_n(f)$, da p_n Interpolationspolynom zu f.

Sei nun $f \in \mathbb{P}_{n+1} \implies f = q_{n+1}$ und daher $I_n(f) = I(q_{n+1}) = I(f)$. □

Satz 2.6 (Fehlerabschätzung)
Sei I_n eine Interpolationsquadratur (I.Q.) auf $\mathbb{P}_n(a,b)$ mit Gewichtsfunktion $\omega \equiv 1$. $R_n(f) := I_n(f) - I(f)$ sei das zugehörige Fehlerfunktional. Dann gilt:

(i) $|R_n(f)| \leq \frac{\|f^{(n+1)}\|_\infty}{(n+1)!}(b-a)^{n+2}$ für alle $f \in C^{n+1}(a,b)$, falls n ungerade ist,

(ii) $|R_n(f)| \leq \frac{\|f^{(n+2)}\|_\infty}{(n+2)!}(b-a)^{n+3}$ für alle $f \in C^{n+2}(a,b)$, falls n gerade ist und die Bedingung i) aus Satz 2.5 erfüllt ist.

Beweis:

(i) Es ist $I_n(f) = I(p_n)$, wobei $p_n \in \mathbb{P}_n$ das Interpolationspolynom zu den Daten $(x_i, f(x_i))$, $i = 0, \ldots, n$ ist.

$$\begin{aligned}\implies |R_n(f)| &= \left|\int_a^b (f - p_n)\right| \leq \int_a^b |f(x) - p_n(x)|\,\mathrm{d}x \\ &\overset{\text{Satz 4.4}}{=} \int_a^b \left|\frac{f^{(n+1)}(\xi_x)}{(n+1)!}\prod_{k=0}^n (x - x_k)\right|\mathrm{d}x \\ &\leq \frac{\|f^{(n+1)}\|_\infty}{(n+1)!}(b-a)^{n+2}\end{aligned}$$

(ii) Aus dem Beweis vom Satz 2.5 folgt: $I_n(f) = I(q_{n+1})$. Dann folgt die Behauptung mit Satz 4.19.

□

Bemerkung:

1. Die Abschätzungen lassen sich leicht verallgemeinern auf den Fall $\omega \in L^1(a,b)$.
2. Die Abschätzung $\int_a^b \left|\prod_{k=0}^n (x - x_k)\right| \leq (b-a)^{n+2}$ kann für gegebene $x_0, \ldots, x_n$ deutlich verbessert werden zu $\int_a^b \left|\prod_{k=0}^n (x - x_k)\right| \leq K(b-a)^{n+2}$ mit $K \ll 1$.

Satz 2.7 (Koordinatentransformation)
Sei $\hat{I}_n(\hat{f}) = \sum_{k=0}^{n} \hat{\omega}_k \hat{f}(t_k)$ mit $t_k \in [-1,1]$ eine I.Q. auf dem „Einheitsintervall“ $[-1,1]$. Dann wird durch

$$I_n(f) := \sum_{k=0}^{n} \omega_k f(x_k)$$

mit

$$\omega_k = \frac{b-a}{2}\hat{\omega}_k, \;\; x_k = \frac{b-a}{2}t_k + \frac{b+a}{2}$$

eine I.Q. auf dem Intervall $[a,b]$ definiert.

Gilt für das Fehlerfunktional $\hat{R}_n$ zu $\hat{I}_n$ die Abschätzung $|R_n(f)| \le K \left\| \hat{f}^{(m)} \right\|_\infty 2^{m+1}$, so gilt für R_n zu I_n: $|R_n(f)| = K \left\| f^{(m)} \right\|_\infty (b-a)^{m+1}$.

Beweis: Sei $p \in \mathbb{P}_n$ und $\hat{p}(t) = p(x(t))$ mit $x(t) := \frac{b-a}{2}t + \frac{b+a}{2}$.

Da $x(t)$ linear ist, gilt $\hat{p} \in \mathbb{P}_n$ und

$$\begin{aligned} I(p) = \int_a^b p(x)\mathrm{d}x &= \int_{-1}^{1} p(x(t))x'(t)\mathrm{d}t \\ &= \tfrac{(b-a)^2}{2} \int_{-1}^{1} \hat{p}(t)\mathrm{d}t = \tfrac{b-a}{2}\hat{I}_n(\hat{p}) \\ &= \sum_{k=0}^{n} \tfrac{b-a}{2}\hat{\omega}_k \hat{p}(t_k) \\ &= \sum_{k=0}^{n} \omega_k p(x_k) = I_n(p). \end{aligned}$$

Daher ist I_n exakt auf $\mathbb{P}_n$ und $I_n(f) = \frac{b-a}{2}\hat{I}_n(\hat{f})$ mit $\hat{f}(t) = f(x(t))$.

Es ist dann $\hat{f}'(t) = x'(t)f'(x(t)) = \frac{b-a}{2}f'(x(t))$ und induktiv $\hat{f}^{(m)}(t) = \left(\frac{b-a}{2}\right)^m f^{(m)}(x(t))$.
Also folgt

$$\left\| \hat{f}^{(m)} \right\|_\infty = 2^{-m}(b-a)^m \left\| f^{(m)}(x(t)) \right\|_\infty .$$

$$\begin{aligned} \Longrightarrow \quad |R_n(f)| &= \left| \int_a^b f(x)\mathrm{d}x - I_n(f) \right| = \left| \tfrac{b-a}{2} \left[\int_{-1}^{1} \hat{f}(t)\mathrm{d}t - \hat{I}_n(\hat{f}) \right] \right| \\ &= \tfrac{b-a}{2} \left| \hat{R}_n(\hat{f}) \right| \le \tfrac{b-a}{2} \left\| \hat{f}^{(m)} \right\|_\infty K 2^{m+1} \\ &= (b-a)^{m+1} \left\| f^{(m)} \right\|_\infty K. \end{aligned}$$

□

Bemerkung:

1. Es reicht also aus I.Q.en auf $[-1,1]$ zu konstruieren. Zu $-1 \le t_0 < \ldots < t_n \le 1$ wird durch

$$\hat{\omega}_j := \int_{-1}^{1} \prod_{\substack{k=0 \\ k \ne j}}^{n} \frac{t - t_k}{t_j - t_k}\mathrm{d}t$$

die I.Q. für $\omega = 1$ auf $[-1,1]$ zu $\mathbb{P}_n$ definiert. Mit $(\omega_j, x_j)_{j=0}^n$ wie im Satz 2.7 wird dann die I.Q. auf $[a,b]$ definiert.

2. Da für jede I.Q. $I_n(1) = (b-a)$ gilt, muss $\sum_{k=0}^{n} \omega_k = (b-a)$ gelten.

3. Wie bei der Polynominterpolation treten Probleme für große Werte von n auf, wie z.B. negative Gewichte. Daher geht man dazu über, Quadraturen auf Teilintervallen aufzusummieren:

$$\int_a^b f(x)\mathrm{d}x = \sum_{i=1}^{N} \int_{a_{i-1}}^{a_i} f(x)\mathrm{d}x \quad a = a_0 < \ldots < a_N = b$$

Satz 2.8 (Zusammengesetze Quadraturen)
Sei $\hat{I}_n(\hat{f}) = \sum_{k=0}^{n} \hat{\omega}_k \hat{f}(t_n)$ eine I.Q. auf $[-1,1]$ mit $\left|\hat{R}_n(\hat{f})\right| \leq K \left\|\hat{f}^{(m)}\right\|_\infty 2^{m+1}$. Zu $a < b,\ N \in \mathbb{N}$ setze $a_l := a + lH,\ l = 0, \ldots, N$ mit $H := \frac{b-a}{N}$.
Dann ist

$$I_h(f) := \frac{H}{2} \sum_{l=1}^{N} \sum_{k=0}^{n} \hat{\omega}_k f\left(\frac{H}{2}(t_k - 1) + a + lH\right)$$

eine Quadraturformel mit der Abschätzung
$|R_h(f)| := |I(f) - I_h(f)| \leq K \left\|f^{(m)}\right\|_\infty (b-a)H^m$.

Beweis: Wir wenden Satz 2.7 auf $[a_{l-1}, a_l]$ an:

$$\begin{aligned} \Longrightarrow \quad I_n^l(f) &:= \sum_{k=0}^{n} \tfrac{a_l - a_{l-1}}{2} \hat{\omega}_k f\left(\tfrac{a_l - a_{l-1}}{2} t_k + \tfrac{a_l + a_{l-1}}{2}\right) \\ &= \tfrac{H}{2} \sum_{k=0}^{n} \hat{w}_k f\left(\tfrac{H}{2}(t_k - 1) + a + lH\right). \end{aligned}$$

Also gilt $I_h(f) = \sum_{l=1}^{N} I_n^l(f)$ und es folgt:

$$\begin{aligned} |R_h(f)| &\leq \sum_{l=1}^{N} \left|R_n^l(f)\right| \overset{\text{Satz 2.7}}{\leq} K \left\|f^{(m)}\right\|_\infty \sum_{l=1}^{N} (a_l - a_{l-1})^{m+1} \\ &= K \left\|f^{(m)}\right\|_\infty \underbrace{NH}_{=(b-a)} H^m. \end{aligned}$$

□

2.1 Newton-Cotes Formeln

- Die Newton-Cotes Formeln sind I.Q.en mit äquidistanten Stützstellen $x_k = a + kh,\ h = \frac{b-a}{n}$.
- Als offene Newton-Cotes Formeln bezeichnet man I.Q.en zu äquidistanten Stützstellen $x_k = a + (k+1)h,\ h = \frac{b-a}{n+2}$, d.h. die Randpunkte a, b sind keine Stützstellen.

1. $n = 1$ (Trapezregel)

$x_0 = a,\ x_1 = b,\ \omega_0 = \int_a^b \frac{x-b}{a-b}\mathrm{d}x = \frac{b-a}{2},\ \omega_1 = \frac{b-a}{2}$,
$T(f) = I_1(f) = \frac{b-a}{2}(f(a) + f(b))$ (vgl. Abb. 2.1).

$$\begin{aligned} |R_T(f)| &\leq \tfrac{\|f''\|_\infty}{2} \int_a^b |x-a|\,|x-b| \\ &= \tfrac{\|f''\|_\infty}{2} \tfrac{(b-a)^3}{6} = \tfrac{\|f''\|_\infty}{12}(b-a)^3. \end{aligned}$$

2. $n = 2$ (Simpson-Regel)

$$x_0 = a,\ x_1 = \tfrac{a+b}{2},\ x_2 = b,\ \omega_0 = \omega_2 = \tfrac{b-a}{6},\ \omega_1 = \tfrac{2(b-a)}{3},$$
$$S(f) = I_2(f) = \tfrac{b-a}{6}\left(f(a) + 4f\left(\tfrac{a+b}{2}\right) + f(b)\right),$$
$$|R_S(f)| \leq \tfrac{\|f^{(4)}\|_\infty}{2880}(b-a)^5$$

Zusammengesetze Newton-Cotes Formeln

1. **Zusammengesetze Trapezregel** (Satz 2.8, $n = 1,\ h = H$)

$$T_h(f) = \tfrac{h}{2}\sum_{l=1}^{N}[f(a+lh-h) + f(a+lh)] = \tfrac{h}{2}\left(f(a) + 2\sum_{l=1}^{N-1} f(a+lh) + f(b)\right),$$
$$|R_h(f)| \leq \tfrac{\|f''\|_\infty}{12}(b-a)h^2.$$

2. **Zusammengesetze Simpson-Regel** (Satz 2.8, $n = 2,\ h = \frac{H}{2}, x_i := a + ih$)

$$S_h(f) = \tfrac{h}{3}\left(f(a) + 2\sum_{l=1}^{N-1} f(x_{2l}) + 4\sum_{l=1}^{N} f(x_{2l-1}) + f(b)\right),$$
$$|R_n(f)| \leq \tfrac{\|f^{(4)}\|_\infty}{180}(b-a)h^4.$$

Bemerkung: Bei den Newton-Cotes Formeln bleiben die Gewichte bis $n = 6$ positiv. Dei den offenen Newton-Cotes Formeln nur bis $n = 2$.

2.2 Gauß-Quadraturen

Idee: Wir suchen eine Quadratur Q_n, welche für $\mathbb{P}_m$ mit möglichst großem m exakt ist. Dies ist **nicht** möglich für $m = 2n + 2$ (Gegenbeispiel konstruierbar). Aber für $m = 2n + 1$ wird dies mit der Gauß-Quadrature (G.Q.) erreicht.

Definition 2.9 (Gauß-Quadraturen)
Sei $\omega \in L^1(a,b)$ gegeben. Eine Quadraturformel $Q_n : C([a,b]) \longrightarrow \mathbb{R},\ Q_n(f) := \sum_{k=0}^{n} \omega_k f(x_k)$ heißt **Gauß-Quadratur**, *falls Q_n exakt ist auf $\mathbb{P}_{2n+1}$.*

Satz 2.10
Sei $\omega \in L^1(a,b)$ und eine Quadratur $Q_n(f) := \sum_{k=0}^{n} \omega_k f(x_k)$ gegeben. Setze $p_{n+1}(x) := \prod_{k=0}^{n}(x - x_k)$. Dann sind äquivalent:

(i) Q_n ist Gauß-Quadratur.

(ii) Q_n ist Interpolationsquadratur und $\int_a^b \omega(x)p_{n+1}(x)q(x)\mathrm{d}x = 0\ \forall\ q \in \mathbb{P}_n$.

Beweis: „(i) $\Longrightarrow$ (ii)“: Sei $q \in \mathbb{P}_n$. Dann ist

$$\int_a^b \omega(x) p_{n+1}(x) q(x) \mathrm{d}x = Q_n(p_{n+1}q) = \sum_{k=0}^{n} \omega_k \underbrace{p_{n+1}(x_k)}_{=0 \text{ nach Def.}} q(x_k) = 0.$$

„(ii) $\Longrightarrow$ (i)“: Sei $p \in \mathbb{P}_{2n+1}$. Mit Polynomdivision gilt: $p = qp_{n+1} + r$ mit $q, r \in P_n$. Damit folgt

$$\begin{aligned} \int_a^b \omega(x) p(x) \mathrm{d}x &= \int_a^b \omega(x) \Big(\underbrace{q(x) p_{n+1}(x)}_{=0} + r(x) \Big) \mathrm{d}x \\ &\overset{\text{Vor. (ii)}}{=} 0 + \int_a^b \omega(x) r(x) \mathrm{d}x \\ &= 0 + Q_n(r) \\ &= Q_n(p_{n+1}q) + Q_n(r) \\ &= Q_n(p). \end{aligned}$$

□

Definition 2.11

(i) Eine Funktion $\omega \in L^1(a,b)$ heißt **zulässige Gewichtsfunktion**, *falls gilt $\omega \geq 0$ und* $\int_a^b w(x)\mathrm{d}x > 0$.

(ii) Ist ω eine zulässige Gewichtsfunktion, so wird durch

$$\langle p, q \rangle_\omega := \int_a^b \omega(x) p(x) q(x) \mathrm{d}x$$

ein Skalarprodukt auf $\mathbb{P}_n$ definiert.

Satz 2.12
Sei ω eine zulässige Gewichtsfunktion. Dann liefert die durch das Gram-Schmidtsche Orthogonalisierungsverfahren definierte Folge $(p_n)_{n\in\mathbb{N}}$

$$p_{n+1}(x) = x^{n+1} - \sum_{i=0}^{n} \frac{\langle x^{n+1}, p_i \rangle_\omega}{\langle p_i, p_i \rangle_\omega} p_i(x), p_0 = 1$$

das eindeutig bestimmte normierte Polynom $p \in \mathbb{P}_{n+1}$ der Form

$$(*)\ p(x) = \prod_{k=0}^{n} (x - x_k) \quad x_k \in \mathbb{C},\ 0 \leq k \leq n$$

mit

$$(**)\ \langle p, q \rangle_\omega = 0 \quad \forall\ q \in \mathbb{P}_n.$$

Außerdem ist $\{p_0, p_1, \ldots, p_{n+1}\}$ eine **Orthogonalbasis** von $\mathbb{P}_{n+1}$ bezüglich $\langle \cdot, \cdot \rangle_\omega$.

Beweis: (Induktion über n)

$n = 0$: klar.

$n-1 \longrightarrow n$: Sei $\{p_0, \ldots, p_n\}$ eine Orthogonalbasis von $\mathbb{P}_n$. Setze

$$\mathbb{P}_n^{\perp} := \left\{ p \in \mathbb{P}_{n+1} \;\middle|\; \langle p, q \rangle_\omega = 0 \quad \forall\, q \in \mathbb{P}_n \right\} \implies \dim(\mathbb{P}_n^{\perp}) = 1.$$

Da $(*)$ verlangt, dass der Koeffizient vor x^{n+1} gleich 1 ist, gibt es genau ein $p \in \mathbb{P}_{n+1}$, welches $(*)$ und $(**)$ erfüllt. Nach Konstruktion ist $p = p_{n+1}$, da $\langle p_{n+1}, p_k \rangle_\omega = 0 \quad \forall\, 0 \le k \le n$.
Also folgt $\langle p_{n+1}, q \rangle_\omega = 0 \quad \forall\, q \in \mathbb{P}_n$. □

Satz 2.13
Sei ω eine zulässige Gewichtsfunktion. Dann gilt: die Nullstellen $x_0, \ldots, x_n$ von p_{n+1} aus Satz 2.12 sind reell, einfach und liegen im Intervall (a, b).

Beweis: Wir setzen:
$q(x) = 1$, $k = -1$, falls es keine reelle Nullstelle ungerader Vielfachheit von p_{n+1} in (a, b) gibt,

$q(x) = \prod_{j=0}^{k} (x - x_j)$ andernfalls, wobei x_j, $0 \le j \le k$ alle solche Nullstellen sind.

Zu zeigen: $k = n$ und somit $q = p_{n+1}$.

Annahme: $k < n$: Nach Definition hat $p := p_{n+1} q$ kein Vorzeichenwechsel in (a, b). Da $k < n$, folgt $q \in \mathbb{P}_n$ und somit $\langle p_{n+1}, q \rangle_\omega = 0 \implies \omega p_{n+1} q = 0$ (fast überall) und somit $\omega = 0$ (fast überall). Dies ist ein Widerspruch zur Definition von ω. □

Satz 2.14
Sei ω eine zulässige Gewichtsfunktion. Dann gibt es genau eine G.Q. Q_n für ω, nämlich die, deren Stützstellen $x_0, \ldots, x_n$ die Nullstellen von p_{n+1} aus Satz 2.12 sind und deren Gewichte definiert sind durch

$$\omega_j := \int_a^b \omega(x) L_j(x) \mathrm{d}x$$

mit

$$L_j(x) := \prod_{\substack{k=0 \\ k \neq j}}^{n} \frac{(x - x_k)}{(x_j - x_k)}$$

Es gilt $\omega_j > 0 \quad \forall\, j$.

Beweis: Folgt aus den Sätzen 2.10, 2.12, 2.13 und aus dem Satz 2.3.

Noch zu zeigen: $\omega_j > 0 \quad \forall\, j$: Da $L_j^2 \in \mathbb{P}_{2n}$ ist, folgt:

$$0 < \int_a^b \omega(x) L_j^2(x) \mathrm{d}x = Q_n(L_j^2) = \sum_{k=0}^{n} \omega_k L_j^2(x_k) = \omega_j.$$

□

Satz 2.15 (Deutung der Gauß-Quadratur als Interpolationsquadratur)
Seien p das eindeutige bestimmte Polynom in $\mathbb{P}_{2n+1}$ mit den Eigenschaften $p(x_i) = f(x_i),\ p'(x_i) = f'(x_i)$ für $i = 0, \ldots, n$ und x_i die Nullstellen von p_{n+1}. Dann gilt:

$$Q_n(f) = Q_n(p) = I(p).$$

Beweis: (Übungsaufgabe) □

Folgerung 2.16
Für $f \in C^{2n+2}(a,b)$ gibt es ein $\xi \in (a,b)$ mit $I(f) - Q_n(f) = \frac{f^{(2n+2)}(\xi)}{(2n+2)!} \langle p_{n+1}, p_{n+1} \rangle_\omega$

Beweis: (Übungsaufgabe) □
Bemerkung: Die G.Q.en sind für stetige Funktionen auf kompakten Intervallen konvergent bei Graderhöhung, d.h. $|I(f) - Q_n(f)| \overset{n}{\longrightarrow} \overset{\infty}{} 0$.

Beispiel 2.17

1. **Gauß-Legendre-Quadratur**
   $\omega(x) = 1$, $[-1, 1]$.
   Es gilt $p_n(x) = \frac{(2n)!}{2^n (n!)^2} P_n(x)$, wobei $P_n(x)$ die **Legendre-Polynome** sind mit $P_0(x) = 1,\ P_1(x) = x$, und
   $$P_{n+1}(x) = \frac{2n+1}{n+1} x P_n(x) - \frac{n}{n+1} P_{n-1}(x).$$
   Es gilt: $I(f) - Q_n(f) = 2^{2n} \frac{n+1}{2n+2} \frac{(n!)^4}{((2n+1)!)^3} f^{(2n+2)}(\xi)$.

   Für $n = 1$: $Q_1(f) = f\left(-\frac{1}{\sqrt{3}}\right) + f\left(\frac{1}{\sqrt{3}}\right)$ („2-Punkt-Gauß-Quadratur“).

   $n = 2:\ Q_2(f) = \frac{1}{9}\left(5f\left(-\sqrt{\frac{3}{5}}\right) + 8f(0) + 5f\left(\sqrt{\frac{3}{5}}\right)\right)$.

   Die G.Q. auf $[a,b]$ erhält man durch Koordinatentransformation (vgl. 2.7).

2. **Gauß-Tschebyscheff-Quadratur**
   $\omega(x) = \sqrt{1 - x^2}^{-1}$, $[-1, 1]$.
   $p_n(x) = \frac{1}{2^{n-1}} T_n(x)$ und T_n die Tschebyscheff-Polyome 1. Art mit

   $T_0(x) = 1,\ T_1(x) = x$ und
   $$T_{n+1}(x) = 2x T_n(x) - T_{n-1}(x),$$
   $\Longrightarrow\ T_n(x) = \cos(n \arccos(x))$.
   Nullstellen von $p_{n+1}:\ x_j^{(n)} = \cos\left(\frac{2j+1}{2n+1}\pi\right) \quad j = 0, \ldots, n.$
   Gewichte: $\omega_j^{(n)} = \frac{\pi}{n+1}$.
   Fehler: $I(f) - Q_n(f) = \frac{\pi}{2^{2n+1}(2n+2)!} f^{(2n+2)}(\xi)$.

3. **Gauß-Laguerre-Quadratur**

   $\omega(x) = e^{-x},\ [0, \infty)$.
   $p_n(x) = (-1)^n L_n(x)$ und L_n Laguerre-Polynome mit

   $L_0(x) = 1,\ L_1(x) = 1 - x$ und

   $$L_{n+1}(x) = (1 + 2n - x)L_n(x) - n^2 L_{n-1}(x).$$

   Fehler: $I(f) - Q_n(f) = \frac{n+1}{2} \frac{(n!)^2}{(2n+1)!} f^{(2n+2)}(\xi)$.

4. **Gauß-Hermite-Quadratur**

   $\omega(x) = e^{-x^2},\ (-\infty, \infty)$.
   $p_n(x) = 2^n H_n(x)$ und H_n die Hermite Polynome mit
   $H_0(x) = 1,\ H_1(x) = 2x$ und

   $$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x).$$

   Fehler: $I(f) - Q_n(f) = \frac{\sqrt{\pi} n!}{2^{n+1}(2n+1)!} f^{(2n+2)}(\xi)$.

5. **Gauß-Jacobi-Quadratur**

   $\alpha, \beta > -1,\ \omega(x) = (1 - x)^\alpha (1 + x)^\beta, [-1, 1]$.
   $p_n(x) = J_n(x, \alpha, \beta)$ und $J_n(x, \alpha, \beta)$ sind die Jacobi-Polynome, definiert durch

   $$J_n(x, \alpha, \beta) := \frac{1}{2^n n! \omega(x)} \frac{d^k}{dx^n} \left((x^2 - 1)^n \omega(x) \right).$$

Definition 2.18 (Zusammengesetze Gauß-Quadraturen)
Zu $a < b,\ N \in \mathbb{N}$ setze $a_l := a + lH,\ l = 0, \ldots, N$ mit $H := \frac{b-a}{N}$. Sei $Q_n^l(f), n \in \mathbb{N}$ eine Gauß-Quadratur auf $[a_{l-1}, a_l]$, dann ist durch

$$Q_h(f) := \sum_{l=1}^{N} Q_n^l(f)$$

eine zusammengesetzte Gauß-Quadratur definiert.

Beispiel: (Zusammengesetze 2-Punkt G.Q. mit $n = 1$, $\omega = 1$)
Setze $h = \frac{b-a}{N},\ a_l = a + lh$ für $l = 0, \ldots, N$. Dann ist die zusammengesetze 2-Punkt G.Q. gegeben durch

$$Q_h(f) = \frac{h}{2} \sum_{j=0}^{N-1} \left(f(a_j + h') + f(a_{j+1} - h') \right)$$

mit $h' = \frac{h}{2} \left(1 - \frac{1}{\sqrt{3}} \right)$.

2.3 Romberg Verfahren

Idee: Anwendung der Richardson Extrapolation auf eine zusammengesetze Quadraturformel, d.h.

$$a(h) = T_h(f)$$

wobei $h_k = h_0 2^{-k}$ gewählt wird (Romberg Folge). Besonders geeignet ist die zusammengesetze Trappezregel $T_h(f)$, da sie eine asympotische Entwicklung in h^2 erlaubt, d.h. $q = 2$ in Satz 4.23. Um dies zu beweisen, führen wir zunächst die Bernoulli Polynome ein.

Definition 2.19 (Bernoulli Polynome/Zahlen)
Die durch $B_0(t) = 1$ *und* $\frac{\partial}{\partial x} B_k(t) = B_{k-1}(t)$, $\int_0^1 B_k(t)\mathrm{d}t = 0, k \geq 1$, *definierten Polynome heißen* **Bernoulli Polynome**. *Es ist also*

$$B_0(t) = 1, \quad B_1(t) = t - \frac{1}{2}, \quad B_2(t) = \frac{1}{2}t^2 - \frac{1}{2}t + \frac{1}{12}, \quad \dots$$

Die **Bernoulli Zahlen** *sind gegeben durch*

$$B_k := k! \cdot B_k(0).$$

Lemma 2.20 (Eigenschaften der Bernoulli Polynome)
Für die Bernoulli Polynome gilt:

(i) $B_k(0) = B_k(1)$ für $k \geq 2$,

(ii) $B_k(t) = (-1)^k B_k(1-t)$ für $k \geq 0$,

(iii) $B_{2k+1}(0) = B_{2k+1}\left(\frac{1}{2}\right) = B_{2k+1}(1) = 0$ für $k \geq 1$.

Beweis: (ohne Beweis) □

Satz 2.21 (Euler-MacLaurin'sche Summenformel)
Sei $f \in C^{2m}(a,b)$, $m \in \mathbb{N}$ und $h := \frac{b-a}{n}$, $n \in \mathbb{N}$. Dann gilt:

$$T_h(f) = \int_a^b f(x)\mathrm{d}x + \sum_{k=1}^{m-1} h^{2k} \frac{B_{2k}}{(2k)!} \left(f^{(2k-1)}(b) - f^{(2k-1)}(a) \right) + O(h^{2m}).$$

Beweis: Sei $\varphi \in C^{2m}(0,1)$ beliebig. Dann gilt mit $B_1' = B_0,\ B_1(1) = \frac{1}{2},\ B_1(0) = -\frac{1}{2}$:

$$\begin{aligned}
\int_0^1 \varphi(t)\mathrm{d}t &= \int_0^1 B_0(t)\varphi(t)\mathrm{d}t \\
&= [B_1(t)\varphi(t)]_{t=0}^1 - \int_0^1 B_1(t)\varphi'(t)\mathrm{d}t \\
&= \tfrac{1}{2}(\varphi(1)+\varphi(0)) - [B_2(t)\varphi'(t)]_{t=0}^1 + \int_0^1 B_2(t)\varphi''(t)\mathrm{d}t \\
&\overset{2.20.i}{=} \tfrac{1}{2}(\varphi(1)+\varphi(0)) - B_2(0)(\varphi'(1)-\varphi'(0)) + \underbrace{\left[B_3(t)\varphi''(t)\right]_{t=0}^1}_{=0\ \ 2.20.iii} - \int_0^1 B_3(t)\varphi'''(t)\mathrm{d}t \\
&= \cdots \\
&= \tfrac{1}{2}(\varphi(1)-\varphi(0)) - \sum_{k=1}^{m-1} B_{2k}(0)\left(\varphi^{(2k-1)}(1) - \varphi^{(2k-1)}(0)\right) + \int_0^1 B_{2m}(t)\varphi^{(2m)}(t)\mathrm{d}t
\end{aligned}$$

Setze $\varphi_j(t) := hf(x_{j-1}+th), \quad 1 \le j \le n$, dann gilt:

- $\int_0^1 \varphi_j(t)\mathrm{d}t = \int_{x_{j-1}}^{x_j} f(x)\mathrm{d}x$,
- $\varphi_j^{(k-1)}(t) = h^k f^{(k-1)}(x_{j-1}+th)$,
- $\varphi_j(1) = hf(x_j) = \varphi_{j+1}(0)$,
- $\varphi_j^{(2k-1)}(1) = \varphi_{j+1}^{(2k-1)}(0)$.

Daher gilt:

$$\begin{aligned}
\int_a^b f(x)\mathrm{d}x &= \sum_{j=1}^n \int_{x_{j-1}}^{x_j} f(x)\mathrm{d}x = \sum_{j=1}^n \int_0^1 \varphi_j(t)\mathrm{d}t \\
&= \sum_{j=1}^n \tfrac{1}{2}(\varphi_j(0)+\varphi_j(1)) - \sum_{j=1}^n \sum_{k=1}^{m-1} B_{2k}(0)\left(\varphi_j^{(2k-1)}(1) - \varphi_j^{(2k-1)}(0)\right) \\
&\quad + \sum_{j=1}^n \int_0^1 B_{2m}(t)\varphi_j^{(2m)}(t)\mathrm{d}t \\
&= \sum_{j=1}^n \tfrac{h}{2}(f(x_j)+f(x_{j-1})) - \sum_{k=1}^{m-1} B_{2k}(0)\left(\varphi_k^{(2k-1)}(1) - \varphi_k^{(2k-1)}(0)\right) \\
&\quad + \sum_{j=1}^n \int_0^1 B_{2m}(t)h^{2m+1}f^{(2m)}(x_{j-1}+th)\mathrm{d}t \\
&= T_h(f) - \sum_{k=0}^{m-1} B_{2k}(0)\left(f^{(2k-1)}(b) - f^{(2k-1)}(a)\right)h^{2k} \\
&\quad + h^{2m}\left[h\sum_{j=1}^m \int_0^1 B_{2m}(t)f^{(2m)}(x_{j-1}+(h))\mathrm{d}t\right].
\end{aligned}$$

Der letzte Term ist $O(h^{2m})$, falls $[\cdot]$ durch eine Konstante unabhängig von h abgeschätzt werden kann. Wir erhalten

$$\begin{aligned}
\left|h\sum_{j=1}^n \int_0^1 B_{2m}(t)f^{(2m)}(x_{j-1}(h))\mathrm{d}t\right| &\le h\sum_{j=1}^n \|B_{2m}\|_\infty \|f(2m)\|_\infty \\
&= n\cdot h\cdot \|B_{2m}\|_\infty \cdot \left\|f^{(2m)}\right\|_\infty \\
&= (b-a)\|B_{2m}\|_\infty \cdot \left\|f^{(2m)}\right\|_\infty = \text{konstant}.
\end{aligned}$$

Bemerkung: Die Summenformel zeigt die assymptotische Entwicklung und dass die Trapezregel auch ohne Extrapolation sehr gut für die Integration periodischer Funktionen geeignet ist.

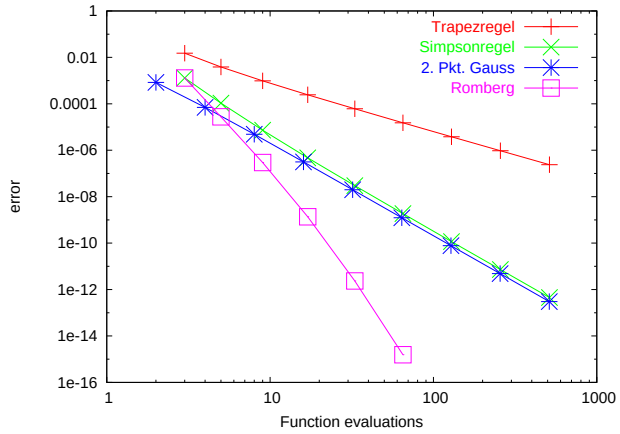
Definition 2.22 (Experimentelle Konvergenzordnung (*EOC*))
Sei $f \in C^k(a,b)$ und $I : C^k(a,b) \longrightarrow \mathbb{R}$ ein Funktional, I_h eine Quadraturformel, die I auf einer Zerlegung der Feinheit h approximiert. Gelte $h_1 > h_2$.

Die experimentelle Konvergenz *$EOC\,(e_{h_1 \longrightarrow h_2})$ (engl. **e**xperimental **o**rder of **c**onvergence) für den Fehler $e_h := |I(f) - I_h(f)|$ ist definiert durch*

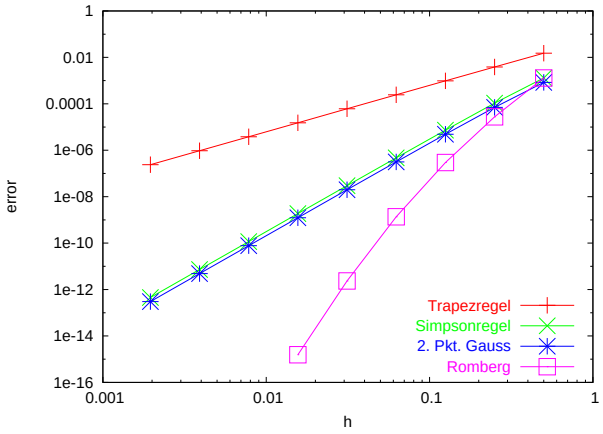
$$EOC\,(e_{h_1 \longrightarrow h_2}) := \frac{\log\left(\frac{e_{h_1}}{e_{h_2}}\right)}{\log\left(\frac{h_1}{h_2}\right)}.$$

Bemerkung: Für $h \longrightarrow 0$ verhält sich der Fehler wie h^p, wobei p vom angewandten Verfahren abhängt. Mit der EOC hat man die Möglichkeit, p numerisch zu bestimmen.

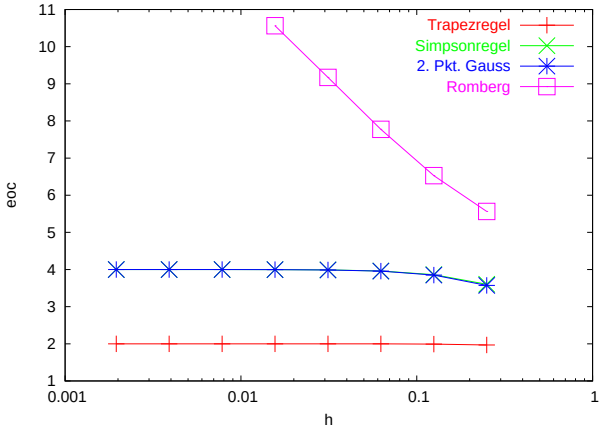
Beispiel 2.23 (Fehler der Approximierung der Integration)
Gegeben seien $I = [0,1]$ und $f(x) := \frac{1}{x+1}$, $g(x) := \frac{3}{2}\sqrt{x}$. Es gilt $\int_0^1 \frac{1}{x+1}\mathrm{d}x = \ln(2)$, $\int_0^1 \frac{3}{2}\sqrt{x}\mathrm{d}x = 1$. Die Abbildung 2.2 ziegt das Verhalten des Approximationsfehlers von 4 Verfahren: Trapezregel (rot), Simposon-Regel (grün), zwei-Punkt Quadratur (blau) und Romberg Verfahren (lila). ***Typ 1*** ist der Fehler im Vergleich zu der Anzahl der Funktionauswertungen, im Prinzip ein Maß für den Berechnungsaufwand. ***Typ 2*** ist der Fehler im Vergleich zu h, d.h. zu der Unterteilung bei den zusammengesetzen Quadraturen. ***Typ 3*** ist die EOC im Verhältnis zu h.



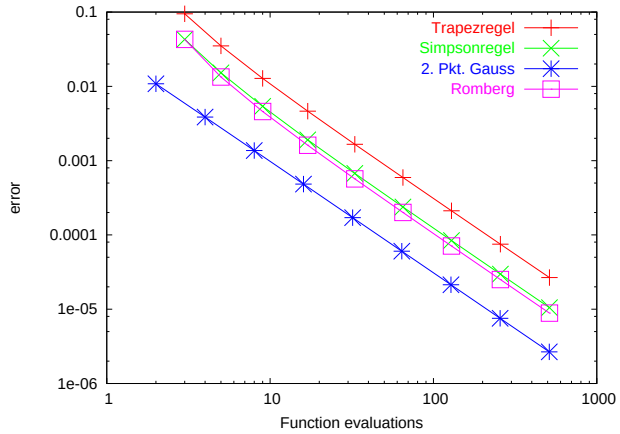
f Typ 1



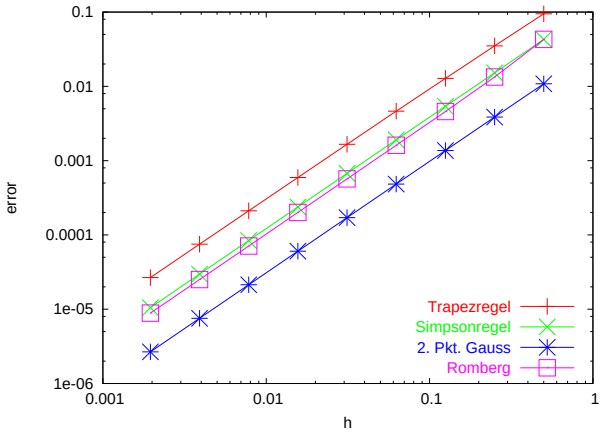
f Typ 2



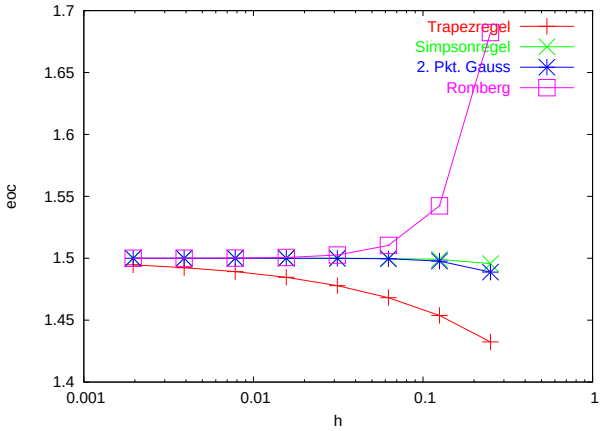
f Typ 3



g Typ 1



g Typ 2



g Typ 3

Abbildung 2.2: Fehler der Quadraturen

Kapitel 3

Numerik Gewöhnlicher Differentialgleichungen

3.1 Einleitung

Viele Fragestellungen aus den Naturwissenschaften, der Ökonomie und Medizin führen auf mathematische Probleme, die numerisch gelöst werden müssen. In dieser Vorlesung wird die Theorie und Praxis grundlegender numerischer Algorithmen zur Behandlung gewöhnlicher Differentialgleichungen behandelt. Dazu gehören Einschritt- und Mehrschrittverfahren zur Approximation von Anfangswertproblemen, sowie Finite Differenzen und Finite Elemente Verfahren zur Diskretisierung von Randwertproblemen. Weitere Themen sind Gradientenverfahren, Eigenwertprobleme und Grundzüge der Approximation.

Beginnen wir mit ein paar Grundbegriffen und Beispielen zu gewöhnlichen Differentialgleichungen:

Definition 3.1 (Gewöhnliche Differentialgleichung)
Seien $n \in \mathbb{N}$ und $I \subset \mathbb{R}$ ein Intervall und $F : I \times \mathbb{R}^{n+1} \to \mathbb{R}$ gegeben.
Unter einer skalaren gewöhnlichen Differentialgleichung (DGL) n-ter Ordnung für eine Funktion $y \in C^n(I)$ versteht man eine Gleichung der Form

$$F(x, y(x), y'(x), y''(x), \ldots, y^{(n)}(x)) = 0, \quad (x \in I). \tag{3.1}$$

Falls eine Funktion $f : I \times \mathbb{R}^n \to \mathbb{R}$ existiert, so dass (3.1) in der Form

$$y^{(n)}(x) = f(x, y(x), y'(x), y''(x), \ldots, y^{(n-1)}(x)), \quad (x \in I) \tag{3.2}$$

geschrieben werden kann, so heißt (3.1) explizit (sonst implizit). Ein Funktion $y \in C^n(I)$, die (3.1) erfüllt, heißt Lösung der gewöhnlichen Differentialgleichung.

Beispiel 3.2 (Freier Fall/Gravitation)
Sei $t \in I := [0, \infty]$ die Zeit, $x(t) \in \mathbb{R}$ der Ort eines Massepunktes zur Zeit t, $v(t) = x'(t)$ die Geschwindigkeit des Massepunktes und $a(t) = v'(t) = x''(t)$ die Beschleunigung des Massepunktes, sowie $K = K(t, x(t), v(t))$ eine äußere Kraft, die auf den Massepunkt wirkt.

Newtonsches Kraftgesetz:

$$\begin{aligned} & mx''(t) = K(t, x(t), x'(t)), \\ \Longrightarrow \quad & x''(t) = \tfrac{1}{m} K(t, x(t), x'(t)). \end{aligned} \tag{3.3}$$

Für gegebene Kraft K ist (3.3) eine explizite DGL zweiter Ordnung.

A) Freier Fall ohne Luftwiderstand:

Erbeschleunigung: $g = 9.81 \frac{m}{s^2} \Longrightarrow K(t, x(t), x'(t)) = -mg$.
Einsetzen in (3.3) ergibt:

$$\begin{aligned} x''(t) = & \quad -\frac{1}{m} mg = -g, \\ \Longrightarrow \quad x'(t) = & \quad -gt + c_1, \\ \Longrightarrow \quad x(t) = & \quad -\frac{1}{2} gt^2 + c_1 t + c_2, \quad \text{mit } c_1, c_2 \in \mathbb{R}. \end{aligned}$$

Die Lösung dieser DGL ist also nicht eindeutig bestimmt.

B) Freier Fall mit Luftwiderstand:
Hier ist die Kraft gegeben durch:

$$K(t, x(t), x'(t)) = -\sigma x'(t) - mg, \quad \text{mit } \sigma > 0.$$

Wir erhalten die DGL:

$$x''(t) = -\frac{\sigma}{m} x'(t) - g.$$

Mit $x'(t) = v(t)$ ergibt sich eine DGL erster Ordnung für v:

$$v'(t) = -\frac{\sigma}{m} v(t) - g. \tag{3.4}$$

Die Funktionen $v(t) = c_1 \exp(-\frac{\sigma}{m} t) - \frac{mg}{\sigma}$ sind für alle $c_1 \in R$ Lösungen von (3.4), denn

$$\begin{aligned} & v'(t) = -c_1 \tfrac{\sigma}{m} \exp(-\tfrac{\sigma}{m} t), \\ \text{und} \quad & -\tfrac{\sigma}{m} v(t) - g = -c_1 \tfrac{\sigma}{m} \exp(-\tfrac{\sigma}{m} t) + g - g. \end{aligned}$$

Bemerkung: Man kann zeigen, dass alle Lösungen von (3.4) von dieser Form sein müssen.

Für die Lösung von (3.3) gilt nach Integration:

$$x(t) = -c_1 \frac{m}{\sigma} \exp(-\frac{\sigma}{m} t) - \frac{mg}{\sigma} t + c_2.$$

Auch hier ist die Lösung nicht eindeutig.

C) Freier Fall aus großer Höhe:
Da sich die Gravitation mit der Höhe ändert, müssen wir in diesem Fall das Gravitationsgesetz ansetzen als:

$$K(t, x(t), x'(t)) = -mg \frac{R^2}{x^2(t)}$$

Dabei bezeichnet R den Erddurchmesser. Wir erhalten die explizite DGL zweiter Ordnung:

$$x''(t) = -g \frac{R^2}{x^2(t)}. \tag{3.5}$$

Ansat zur Lösung: $x(t) = at^b$. Damit erhlaten wir: $x'(t) = abt^{b-1}$ und $x''(t) = ab(b-1)t^{b-2}$. Einsetzen in (3.5) ergibt:

$$ab(b-1)t^{b-2} = -gR^2\frac{1}{a^2t^{2b}} = -\frac{gR^2}{a^2}t^{-2b}.$$

Also muß gelten:

1.) $b - 2 = -2b \implies b = \frac{2}{3}$,

2.) $ab(b-1) = -\frac{gR^2}{a^2} \implies a = \left(\frac{9gR^2}{2}\right)^{1/3}$.

Also ist

$$x(t) = \left(\frac{9gR^2}{2}\right)^{1/3} t^{2/3}$$

eine Lösung von (3.5).

Beispiel 3.3
Für die explizite DGL erster Ordnung

$$y'(x) = x^2 + (y(x))^2$$

kann keine lösung in geschlossener Form angegeben werden. Es existieren jedoch Lösungen!

Bemerkung: Wir halten fest:

1. Die Bestimmung einer Lösung (in Formelgestalt) ist sehr oft nicht möglich.
2. Falls eine lösung existiert, ist sie i.A. nicht eindeutig.
3. Die *freien Konstanten* $(c_1, c_2, \dots)$ können durch zusätzliche Bedingungen festgelegt werden: Anfangswerte oder Randwerte.

In diesem Kapitel werden wir uns mit numerischen Verfahren zur Lösung von gewöhnlichen Differentialgleichungen beschäftigen. Dabei werden wir uns zunächst mit Anfangswertproblemen auseinandersetzen.

3.2 Exkurs zur Theorie gewöhnlicher Differentialgleichungen

Definition 3.4 (Anfangswertproblem (AWP))
Sei folgende Voraussetzung erfüllt:

(V) Seien $I := [a, b]$, $b < a$, $G \subset \mathbb{R}^n$ *zusammenhängende und offene Teilmenge (Gebiet),* $S := I \times G$ *und* $f : S \longrightarrow \mathbb{R}^n$ *stetig und* $y_0 \in G$.

Dann heißt $y : I \longrightarrow \mathbb{R}^n$ *Lösung des* **AWP**, *g.d.w.*

(i) $y \in C^1(I, \mathbb{R}^n)$ *und* $y(I) \subseteq G$

(ii) $y'(x) = f(y, y(x)) \ \forall x \in I$

(iii) $y(a) = y_0$

Satz 3.5
Unter der Voraussetzung (V) sind folgende Aussagen äquivalent:

a) $y : I \longrightarrow \mathbb{R}^n$ löst (AWP).

b) $y \in C^0(I, G)$ und $y(x) = y_0 + \int_a^x f(s, y(s))ds \ \ \forall x \in I$.

Beweis: siehe Ü.A. □

Definition 3.6 (Picard-Lindelöf Iteration)
Definiere Operator $T : C^0(I, \mathbb{R}^n) \longrightarrow C^0(I, \mathbb{R}^n)$ *durch*

$$(Ty)(x) := y_0 + \int_a^x f(s, y(s))ds.$$

Dann ist 3.2 b) äquivalent zu $Ty = y$.

Fixpunktiteration: $y^{(0)} \in C^0(I, \mathbb{R}^n)$ *gegeben,* $y^{(n+1)} = Ty^{(n)}$.

Frage: Wann konvergiert diese Iteration?

Satz 3.7 (Picard-Lindelöf, lokale Version)
Es gelten die Voraussetzungen (V). Erfüllt f auf S die Lipschitzbedingung

$$(L) \qquad \|f(x,y) - f(x,z)\|_\infty \leq L \|y - z\|_\infty \quad \forall (x,y),(x,z) \in S,$$

so hat das AWP lokal eine eindeutige Lösung $\tilde{y}$, d.h. $\exists \varepsilon > 0, \ \exists \tilde{y} \in C^1(I_\varepsilon, \mathbb{R}^n)$ mit $\tilde{y}$ löst AWP auf $I_\varepsilon := [a, a + \varepsilon]$.

Beweis: Es ist

$$\begin{aligned} \|(Ty)(x) - (Tz)(x)\|_\infty &\overset{(L)}{\leq} L \int_a^x \|y(s) - z(s)\|_\infty \, \mathrm{d}s \\ &\leq L \|y - z\|_\infty (x - a) \\ &\leq L\varepsilon \|y - z\|_\infty, \text{ für } x \in I_\varepsilon. \end{aligned}$$

Wähle $\varepsilon > 0$, so dass $L\varepsilon < 1$. Dann folgt, dass T Kontraktion ist und die Aussage folgt mit dem Banachschen Fixpunktsatz. □

Satz 3.8 (Picard-Lindelöf, globale Version)
Gelte (V) und (L). Erfüllt f auf S die Bedingung

$$(M) \qquad \|f(x,y)\|_\infty \leq M \qquad \forall (x,y) \in S$$

und G erfülle zusätzlich $G \supseteq \overline{B_\sigma(y_0)} := \{x \in \mathbb{R}^n \mid \|x - y_0\|_\infty \leq \sigma\}$, wobei $\sigma \geq (b-a)M$ sei.
Dann gilt:

a) Das AWP hat auf I genau eine Lösung $\tilde{y}$.

b) $\forall x \in I$ gilt $(x, \tilde{y}(x)) \in K_M = K_M(a, y_0) := \{(x,y) \in \mathbb{R} \times \mathbb{R}^n | \, \|y - y_0\|_\infty \leq M|x - a|\} \cap S$ (siehe Abb. 3.1)).

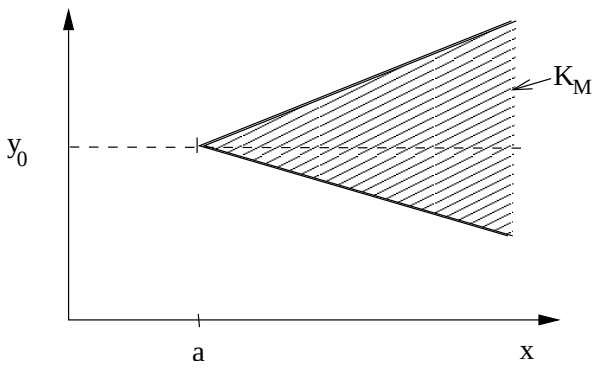


Abbildung 3.1: Picard-Lindelöf: Grafische Darstellung von K_M.

c) Die Fixpunktiteration von Picard-Lindelöf konvergiert gleichmäßig auf I gegen $\tilde{y}$. Für den Fehler gilt:

$$\left\| \tilde{y}(x) - y^{(k)}(x) \right\|_\infty \leq \frac{L^k (x-a)^k}{k!} e^{L(x-a)} \qquad \forall x \in I.$$

Beweis: siehe Ü.A. □

Satz 3.9 (Satz von Peano)
Sei S ein Rechteckgebiet, das $K_M(a, y_0)$ enthält und es gelten (V) und (M). Dann ex. eine Lösung $\tilde{y}$ des AWP auf I.

Beweis: (siehe z.B. Walter [9]) □

Lemma 3.10 (Lemma von Gronwall)
Sei $p, q \in C^0([a,b])$ mit $p, q \geq 0$. Erfüllt die Funktion $e : [a,b] \longrightarrow \mathbb{R}$ die Integralbedingung

$$0 \leq e(x) \leq p(x) + \int_a^x q(s)e(s)ds \qquad \forall x \in [a,b],$$

so gilt:

$$0 \leq e(x) \leq p(x) + \int_a^x q(s)p(s) \exp\left(\int_s^x q(t)dt \right) ds.$$

Beweis: siehe Ü.A. □

Satz 3.11 (Stetigkeitssatz für AWP)
Es gelten die Vor. (V) und (L). Sei $\tilde{y}$ Lösung des AWP und $\tilde{z}$ sei Lösung des gestörten AWP:

$$z'(x) = f(x, z(x)) + \varepsilon(x) \,, \quad z(a) = z_0 \in G.$$

Gelte $z_0 - y_0 \leq \tilde{\varepsilon}$ und sei $\varepsilon(x) \in C^0(I, \mathbb{R}^n)$ mit $\|\varepsilon(x)\|_\infty \leq \varepsilon \qquad \forall x \in I.$
Dann gilt:

$$\|\tilde{z}(x) - \tilde{y}(x)\|_\infty \leq (\tilde{\varepsilon} + \varepsilon(x-a))e^{L(x-a)}.$$

Beweis: Folgt direkt aus 3.7 mit $e(x) := \|\tilde{z}(x) - \tilde{y}(x)\|_\infty$.

$$\begin{aligned} \Longrightarrow \quad e(x) &= \left\| z_0 + \int_a^x f(\tau, \tilde{z}(\tau)) - \varepsilon(\tau)\mathrm{d}\tau - y_0 - \int_a^x f(\tau, \tilde{y}(\tau))\mathrm{d}\tau \right\| \\ &\leq \|z_0 - y_0\| + \varepsilon(x-a) + \int_a^x L\ e(\tau)\mathrm{d}\tau \\ &= p(t) + \int_a^t q(\tau)e(\tau)\mathrm{d}\tau \text{ mit} \\ &\quad p(\tau) = \tilde{\varepsilon} + \varepsilon(x-a),\ q(\tau) = L \end{aligned}$$

Aus dem Lemma von Gronwall folgt:

$$e(x) \leq \tilde{\varepsilon} + \varepsilon(x-a) + L\int_a^x (\tilde{\varepsilon} + \varepsilon(\tau - a))e^{L(\tau-a)}\mathrm{d}\tau.$$

Durch Berechnung des Integrals folgt die Behauptung. □

Satz 3.12 Methode der Trennung der Variablen
Läßt sich die Differentialgleichnung $y' = f(x,y)$ als $y' = \frac{p(x)}{q(y)}$ schreiben und ist $\frac{dq}{dx} = \frac{dp}{dy}$ erfüllt, so gilt: $y(x)$ ist gegeben durch

$$F(x, y(x)) = konst$$

mit

$$F(x,y) = \int_a^x p(t)dt + \int_{y_0}^y q(s)ds.$$

Beweis: Nachrechnen. □

Definition 3.13 (Lineare AWP)
Sei $I = [a,b]$. Gesucht $y \in C^1(I, \mathbb{R})$:

(L-AWP) $\quad \left| \begin{array}{ll} y'(x) = \alpha y(x) + \beta, & \alpha, \beta \in \mathbb{R} \\ y(a) = y_0 \end{array} \right.$

Die Lösung des homogenen AWPs $z' = \alpha z$, $z(a) = z_0$ ist dann gegeben durch $z(x) = z_0 e^{\alpha(x-a)}$.
Durch "Variation der Konstanten" folgt:

$$\underline{\text{Ansatz:}} \quad y(x) = z(x)v(x) \quad \Longrightarrow \quad y(x) = (\tfrac{\beta}{\alpha} + y_0)e^{\alpha(x-a)} - \tfrac{\beta}{\alpha}.$$

Anwendung: Bakterienwachstum
$y_0 \hat{=}$ # Bakterien am Anfang, $-\beta > 0$ Sterberate, $\alpha > 0$ Wachstumsfaktor. Das qualitative Verhalten der Lösung ist in Abb. 3.2 dargestellt.

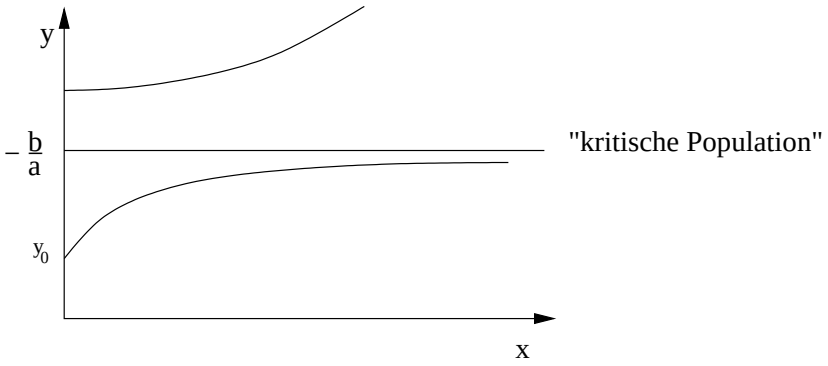


Abbildung 3.2: Lineare (AWP):Anwendung, Bakterienwachstum.

Definition 3.14 (Systeme linearer AWP)
Sei $I = [a, b]$ und $A \in \mathbb{R}^{n\times n}$ diagonalisierbar, $y_0, B \in \mathbb{R}^n$. Gesucht $y \in C^1(I, \mathbb{R}^n)$:

$$\textit{(SL-AWP)} \qquad \left| \begin{array}{l} y'(x) = Ay(x) + B \\ y(a) = y_0 \end{array} \right.$$

Wir definieren die "Matrixexponentielle" $exp(A)$ durch

$$exp(A) := \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

Dann löst auch hier $z(x) = z_0 exp(A(x-a))$ das homogene System $z' = Az\ ,\ \ z(a) = z_0$. Eine Lösung von SL-AWP erhält man durch Variation der Konstanten mit dem Ansatz:

$$y(x) = exp(A(x-a))v(x).$$

Definition 3.15 (Differentialgleichungen höherer Ordnung)
Geg: $I = [a, b]\ ;\ G \subset \mathbb{R}^{m-1}$ *Gebiet;* $S = I \times G\ ,\ f : S \longrightarrow \mathbb{R}$.
Ges: $y \in C^m(I, \mathbb{R}^n)\ ;\ \ y^{(k)}(I) \subseteq P_k(G)\ ;\ \ P_k$ *Projektion*

$$y^{(m)}(x) = f(x, y'(x), \ldots, y^{(m-1)}(x)) \ \ \forall x \in I,$$
$$y^{(k)}(a) = y_0^{(k)} \subset P_k(G)\ , k = 0, \ldots, m-1.$$

Reduktion auf System 1. Ordnung:
Setze $y_k := y^{(k)}\ ;\ k = 0, \ldots, m-1$
und $F(x, y_0, \ldots, y_{m-1}) = (y_0, \ldots, y_{m-1}, f(x, y_0, \ldots, y_{m-1}))^T$.
Dann folgt mit $y = (y_0, \ldots, y_{m-1})^T$:

$$y' = F(x, y),$$
$$y(a) = (y_0^{(0)}, \ldots, y_0^{(m-1)})^T.$$

Generalvereinbarung:

1) Alle Sätze werden im Folgenden für skalare AWP 1. Ordnung formuliert.

2) Sollte ein Satz für Systeme nicht gelten, so wird dies explizit angegeben.

3.3 Einschrittverfahren

Wir betrachten das skalare AWP:

$$\text{(AWP)} \qquad \left| \begin{array}{l} y'(x) = f(x,y), \\ y(a) = y_0. \end{array} \right.$$

Es seien stets die Bedingungen (V), (M), (L) erfüllt und es sei stets $S \subseteq K_M(a,y_0)$ ein Rechteckgebiet, wobei

$$K_M = K_M(a,y_0) := \{(x,y) \in I \times \mathbb{R} |\ |y-y_0| \leq M(x-a) + \varepsilon_0\}, \quad \varepsilon_0 > 0.$$

Wir bezeichnen mit $\tilde{y}$ die eindeutig bestimmte Lösung von (AWP).

Definition 3.16 (Diskretes Verfahren zur Lösung von (AWP))
Definiere Gitter $I_h = \{x_0, \ldots, x_n\} \subseteq I, \quad x_{j+1} = x_j + h_j, \quad h_j > 0, \ j = 0, \ldots, n-1$. *Setze* $x_0 = a, x_n = b$.
Schrittweitenvektor: $\overline{h} = (h_0, \ldots, h_{n-1})^T$.
I_h *heisst zulässiges Gitter auf* I. *Definiere* $h := max_{j=0,\ldots,n-1} h_j$ *als Feinheit von* I_h.
Ein numerisches Verfahren Φ

- *findet ein Gitter* I_h
- *ordnet* I_h *eine Gitterfunktion* $u_h : I_h \longrightarrow \mathbb{R}$ *zu.*

u_h *ist auf* I_h *diskrete Näherungslösung von* $\tilde{y}$.

Definition 3.17 (Globaler Fehler von Verfahren Φ)
Für $x \in I_h$ *setze* $e_h(x) := \tilde{y}(x) - u_h(x)$ *("Fehlerfunktion"). Definiere*

1) $e_i := e_h(x_i), \quad x_i \in I_h$,

2) $\|e_h\|_h := max_{x \in I_h} |e_h(x)|$.

$\|e_h\|_h$ *heist Diskretisierungsfehler von* Φ *auf* I_h.

Definition 3.18 (Konvergenz/Konvergenzordnung)

1) Φ *heisst konvergent auf* I, *falls* $\|e_h\|_h \longrightarrow 0$ *für* $h \longrightarrow 0$.

2) Φ *hat auf* I *die Konvergenzordnung* $p > 0$, *falls für genügend glattes* f *gilt:*

$$\|e_h\|_h = O(h^p) \quad \text{für } h \longrightarrow 0.$$

Definition 3.19 (Explizite Einschrittverfahren (ESV))
Ein Verfahren Φ heisst explizites Einschrittverfahren, falls es eine Verfahrensfunktion $\varphi = \varphi(x,u,h)$, $(x,u) \in K_M(a,y_o)$, $h \in (0, b-a]$, $\varphi(x,u,h) \in [-M,M]$ gibt, so dass u_h auf zulässigem Gitter I_h definiert ist durch:

$$\left|\begin{array}{ll} u_0 & := y_0 + \varepsilon_0, \\ u_{j+1} & := u_j + h_j \varphi(x_j, u_j, h_j), \quad j = 0, \dots n-1, \\ u_h(x_j) & := u_j, \qquad \forall x_j \in I_h. \end{array}\right.$$

Beispiel 3.20
1) Eulerverfahren (explizit): $\varphi(x,u,h) = f(x,u)$.

2) Verbessertes Eulerverfahren: $\varphi(x,u,h) = f(x + \frac{h}{2}, u + \frac{h}{2} f(x,u))$.

Definition 3.21 (Abschneidefehler/Konsistenz)

1) Für $h \in (0, b-a]$, $x \in [a, b-h]$, $y : [x, x+h] \longrightarrow \mathbb{R}$ mit $y'(\xi) = f(\xi, y(\xi))$ für $\xi \in [x, x+h]$ sowie $(x, y(x)) \in K_M$, heißt

$$\tau_h(x,y) := \frac{y(x+h) - y(x)}{h} - \varphi(x, y(x), h)$$

lokaler Abschneidefehler oder Diskretisierungsfehler von φ.

2) φ heißt konsistent mit (AWP), falls gilt

$$|\tau_h(x, \tilde{y})| \overset{h \to 0}{\longrightarrow} 0 \quad \forall x \in I \setminus \{b\}.$$

φ hat Konsistenzordnung $p \in \mathbb{N}$, falls für genügend glattes f gilt: $|\tau_h(x,\tilde{y})| = O(h^p)$, $h \longrightarrow 0$ $\forall x \in I \setminus \{0\}$

3) Das ESV Φ heißt konsistent mit (AWP), falls $e_0 \overset{h \to 0}{\longrightarrow} 0$ und φ konsistent mit (AWP) ist.
Φ hat Konsistenzordnung $p \in \mathbb{N}$, falls

$$|e_0| = O(h^p) \ , \quad h \longrightarrow 0$$

und φ die Konsistenzordnung p hat.

Lemma 3.22 (Diskretes Lemma von Gronwall)
Seien $(p_n)_{n\in\mathbb{N}}, (q_n)_{n\in\mathbb{N}}, (e_n)_{n\in\mathbb{N}}$ positive Folgen mit $e_{n+1} \le (1+q_n)e_n + p_n$ für $n < N$. Dann gilt:

$$e_n \le \Big(e_0 + \sum_{j=1}^{n-1} p_j\Big) \exp\Big(\sum_{j=0}^{n-1} q_j\Big) \qquad \text{für } n < N.$$

Beweis: Durch vollständige Induktion. □

Satz 3.23 (Konvergenzsatz für ESV)
Φ sei ESV mit Verfahrensfunktion $\varphi = \varphi(x,u,h)$. φ sein bzgl. u auf K_M global Lipschitz stetig. Dann gilt:
Ist Φ mit (AWP) konsistent (mit Ordnung p), so ist Φ auf I konvergent (mit Ordnung p).

Beweis: Sei $a \le x < x+h \le b$. Es ist

$$\frac{\tilde{y}(x+h)-\tilde{y}(x)}{h} = \varphi(x,\tilde{y}(x),h) + \tau_h(x,\tilde{y}).$$

Für $\tilde{y}_j = \tilde{y}(x_j)$; $x_j \in I_h$; $\tau_j := \tau_h(x_j,\tilde{y})$ gilt also

$$\tilde{y}_{j+1} - \tilde{y}_j = h_j\varphi(x_j,\tilde{y}_j,h_j) + h_j\tau_j \;\; ; \;\; j = 0,\dots,n-1$$

Ist $\tilde{L}$ die Lipschitzkonstante von φ bzgl. u, so gilt:

$$|e_{j+1}| \le |e_j|(1+h_j\tilde{L}) + |\tau_j|h_j$$

$$\overset{3.22}{\Longrightarrow} \quad \|e_h\|_h \le (|e_0| + (b-a)\max_{j=0,\dots,n-1}|\tau_j|)e^{(b-a)\tilde{L}}$$

Nach Voraussetzung ist Φ konsistent (Ordnung p), also folgt

1) $\max_{j=0,\dots,n-1}|\tau_j| \longrightarrow 0 \quad (= O(h^p))$

2) $|e_0| \longrightarrow 0 \quad (= O(h^p))$

$\Longrightarrow$ Behauptung. □

Beispiel 3.24

1) Das Eulerverfahren ist konvergent mit der Ordnung 1. ($\varphi(x,u,h) = f(x,u)$)

2) Das verbesserte Eulerverfahren ist konvergent mit Ordnung 2.

Definition 3.25 (Implizites ESV)
Ein Verfahren Φ zur Lösung von (AWP) heiß implizites ESV, falls es eine Verfahrensfunktion $\tilde{\varphi} = \tilde{\varphi}(x,u,v,h)$, (x,u), $(x+h,v) \in K_M$, $h \in (0, b-a]$, $\tilde{\varphi}(x,u,v,h) \in [-M,M]$ gibt, so dass für zulässige Gitter I_h mit genügend kleiner Feinheit h u_h definiert ist durch:

$$\left|\begin{array}{ll} u_0 & = y_0 + \varepsilon_0, \\ u_{j+1} & = u_j + h_j\tilde{\varphi}(x_j,u_j,u_{j+1},h_j), \quad j = 0,\dots,n-1, \\ u_h(x_j) & := u_j, \qquad \forall x_j \in I_h. \end{array}\right.$$

Beispiel 3.26

1) Implizites Eulerverfahren:

$$\tilde{\varphi}(x,u,v,h) = f(x+h,v) \quad \Longrightarrow \quad u_{j+1} = u_j + h_j f(x_{j+1},u_{j+1}).$$

Dies entspricht der Rechteckregel:

$$\frac{1}{h}\int_x^{x+h} f(t,y(t))dt \approx f(x+h,y(x+h)).$$

2) Crank-Nicholson-Verfahren oder Trapezverfahren:

$$\tilde{\varphi}(x,u,v,h) = \frac{1}{2}[f(x,u) + f(x+h,v)].$$

Dies entsricht einer Integration mit der Trapezregel.

Satz 3.27
Sei Φ ein implizites ESV mit $\tilde{\varphi} = \tilde{\varphi}(x,u,v,h)$ Lipschitz-stetig bzgl. u und v auf dem Definitionsbereich von $\tilde{\varphi}$. Dann existiert für genügend kleines h eine bzgl. u global Lipschitz-stetige Funktion $v(x,u,h)$, so das gilt:

$$u_{j+1} = u_j + h_j\tilde{\varphi}(x_j,u_j,v(x_j,u_j,h_j),h_j) \;\; j = 0,\ldots,n-1.$$

D.h. Φ ist mit $\varphi(x,u,h) := \tilde{\varphi}(x,u,v(x,u,h),h)$ lokal ein explizites ESV im Sinne von 3.16.

Beweis:

1) Für feste x,u,h sei T_u definiert als
$T_u : \mathbb{R} \longrightarrow \mathbb{R}\,, \;\; v \longmapsto u + h\tilde{\varphi}(x,u,v,h)$
Sei L_1 die Lipschitzkonstante von $\tilde{\varphi}$ bzgl v.

$$|T_u v_1 - T_u v_2| \leq \underbrace{h \cdot L_1}_{<1,\ falls\ h\ klein\ genug} |v_1 - v_2|$$

$\Longrightarrow$ T_u ist kontrahierende Selbstabb. auf $\mathbb{R}$.
Nach Banachschen Fixpunktsatz existiert genau ein $\tilde{v}$, so dass $T_u\tilde{v} = \tilde{v}$
$\Longrightarrow$ $\tilde{v} = \tilde{v}(x,u,h)$ ist wohldefiniert und $(x+h,\tilde{v}) \in K_M$.

2) Seien v_1, v_2 die eindeutigen Fixpunkte von T_{u_1}, T_{u_2}. L_2 sei die Lipschitzkonstante von $\tilde{\varphi}$ bzgl. u. Dann gilt:

$$\begin{aligned} |v_1 - v_2| &= |T_{u_1}v_1 - T_{u_2}v_2| \leq |T_{u_1}v_1 - T_{u_1}v_2| + |T_{u_1}v_2 - T_{u_2}v_2| \\ &\leq hL_1|v_1 - v_2| + |u_1 - u_2| + h \cdot L_2|u_1 - u_2| \\ \Longrightarrow \; |v_1 - v_2| &= \underbrace{\frac{1+hL_2}{1-hL_1}}_{:=\tilde{L}_H} |u_1 - u_2|. \end{aligned}$$

Also folgt für $v_1 = v(x,u_1,h), v_2 = v(x,u_2,h)$:

$$|v(x,u_1,h) - v(x,u_2,h)| \leq \tilde{L}_H|u_1 - u_2|, \quad h \leq H < \frac{1}{L_1}.$$

□

Definition 3.28 Taylorverfahren
Das Taylorverfahren der Ordnung p ist gegeben durch:

$$\varphi_p(x,u(x),h) := f(x,u(x)) + \frac{h}{2}\frac{d}{dx}f(x,u(x)) + \ldots + \frac{h^{p-1}}{p!}\frac{d^{p-1}}{dx^{p-1}}f(x,u(x)).$$

Satz 3.29

Sei $f \in C^p(S)$ mit $p \in \mathbb{N}$. Dann ist das Taylorverfahren Φ_p mit Verfahrensfunktion φ_p konvergent mit Ordnung p, falls $|e_0| = O(h^p)$, $h \to 0$.

Beweis:

1) Für x, y, h folgt mit der Taylorentwicklung mit Integralrestglied:

$$\begin{aligned}\tau_h(x,y) &= \frac{y(x+h)-y(x)}{h} - \varphi_p(x, y(x), h) = \frac{1}{h}\frac{1}{p!}\int_x^{x+h}(x-\xi)^p y^{(p+1)}(\xi)d\xi \\ &= \frac{h^p}{p!}\int_0^1 (1-t)^p y^{(p+1)}(x+th)dt \\ &= O(h^p)\end{aligned}$$

2) $\frac{d^k}{dx^k}f(x,u,h)$ habe Lipschitz-Konstante L_k , $k = 0, \ldots, p-1$.
Dann gilt:

$$|\varphi_p(x, u_1(x), h) - \varphi_p(x, u_2(x)h)| \leq \underbrace{\sum_{j=0}^{p-1} \frac{(b-a)^j}{j!} L_j}_{=:L} |u_1(x) - u_2(x)|.$$

Aus 1) und 2) folgt die Behauptung mit Satz 3.23. □

Bemerkung:

1) Satz 3.26 zeigt, dass es ESV mit beliebig hoher Konvergenzordnung gibt.

2) Taylorverfahren sind in der Praxis unbedeutend, da

- φ_p nicht unabhängig vom Richtungsfeld f ist,

- höhere Ableitungen von f "teuer" auszurechnen sind.

Definition 3.30 (Explizite Runge-Kutta Verfahren)
Das ESV Φ mit Verfahrensfunktion φ heißt explizites Runge-Kutta Verfahren, falls gilt:

$$
\begin{array}{ll}
u_0 & := y_0 + \varepsilon_0, \\
u_{j+1} & := u_j + h_j \varphi(x_j, u_j, h_j), \quad j = 0, \dots, n-1, \\
 & \\
\varphi(x,u,h) & := \sum_{i=1}^{m} \gamma_i k_i(x,u,h) \text{ mit } \gamma_i \in \mathbb{R}\,, \ \sum_{i=1}^{m} \gamma_i = 1, \\
k_1(x,u,h) & := f(x,u), \\
k_2(x,u,h) & := f(x + \alpha_2 h, u + \beta_{2,1} k_1(x,u,h)), \\
\vdots & \\
k_m(x,u,h) & := f\left(x + \alpha_m h, u + h \sum_{l=1}^{m-1} \beta_{m,l} k_l(x,u,h)\right), \\
\text{mit } \alpha_i \in [0,1]. &
\end{array}
$$

Die Koeffizienten $\alpha_i, \gamma_i, \beta_{m,l}$ sind von x, u, h unabhängig und so gewählt, dass die Konsistenzordnung von φ maximal wird. m heißt Stufenzahl des Runge-Kutta-Verfahrens und k_i heißt i-te Stufenfunktion des Verfahrens Φ.

Notation: *(Butcher-Tableau)*

$$
\begin{array}{c|cccc}
\alpha_1 & 0 & & & 0 \\
\alpha_2 & \beta_{2,1} & \ddots & & \\
\vdots & \vdots & \ddots & \ddots & \\
\alpha_m & \beta_{m,1} & \dots & \beta_{m,m-1} & 0 \\
\hline
 & \gamma_1 & & \gamma_{m-1} & \gamma_m
\end{array}
\qquad \textit{bzw.} \qquad
\begin{array}{c|c}
\alpha & \beta \\
\hline
 & \gamma^t
\end{array}
$$

Bemerkung 3.31

1) $\sum_{i=1}^{m} \gamma_i = 1 \iff$ Konsistenz des entsprechenden R.-K. Verfahrens.

2) Oft wird zusätzlich gefordert, dass $\sum_{i=0}^{l-1} \beta_{l,i} = \alpha_l \ \ \forall l = 2, \dots, m$ gilt, damit $k_l(x_j, u_j, h_j)$ bei genügend glattem f die Ableitung $\tilde{y}'(x_j + \alpha_l h)$ bis auf Größen der Ordnung $O(h^2)$ annähert.

Beispiel 3.32

1) Eulerverfahren: $m = p = 1$ $\quad \begin{array}{c|c} 0 & 0 \\ \hline & 1 \end{array}$
$\implies k_1(x,u,h) = f(x,u)\,; \ \ \varphi(x,u,h) = k_1(x,u,h) = f(x,u)$

2) Verbessertes Eulerverfahren: $m = p = 2$ $\quad \begin{array}{c|cc} 0 & & \\ \frac{1}{2} & \frac{1}{2} & \\ \hline & 0 & 1 \end{array}$

$k_1(x,u,h) = f(x,u)$
$k_2(x,u,h) = f(x + \frac{1}{2}h, u + \frac{1}{2}hk_1(x,u,h))$
$\varphi(x,u,h) = 1 \cdot k_2(x,u,h) = f(x + \frac{h}{2}, u + \frac{h}{2} f(x,u))$

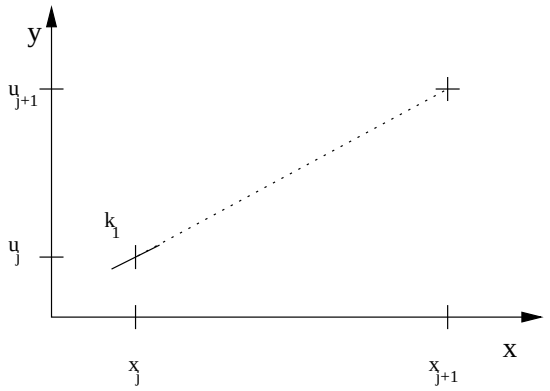


Abbildung 3.3: Runge-Kutta: Eulerverfahren

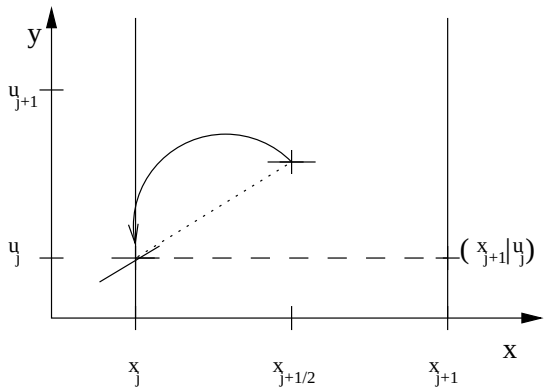


Abbildung 3.4: Runge-Kutta: verbessertes Eulerverfahren

3) Verfahren von Heun: $m = p = 3$

$$\begin{array}{c|ccc} 0 & & & \\ \frac{1}{3} & \frac{1}{3} & & \\ \frac{2}{3} & 0 & \frac{2}{3} & \\ \hline & \frac{1}{4} & 0 & \frac{3}{4} \end{array}$$

$k_1(x,u,h) = f(x,u)$
$k_2(x,u,h) = f(x+\frac{1}{3}h, u+\frac{1}{3}hk_1)$
$k_3(x,u,h) = f(x+\frac{2}{3}h, u+\frac{2}{3}hk_2)$
$\varphi(x,u,h) = \frac{1}{4}k_1 + \frac{3}{4}k_3$

4) Klassisches R.-K. Verfahren: $m = p = 4$

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

Bemerkung 3.33

1) Im Spezialfall $f(x,y) = f(x)$ werden expl. R.-K. Verfahren zu zusammengesetzten Quadraturen auf I: z.B.

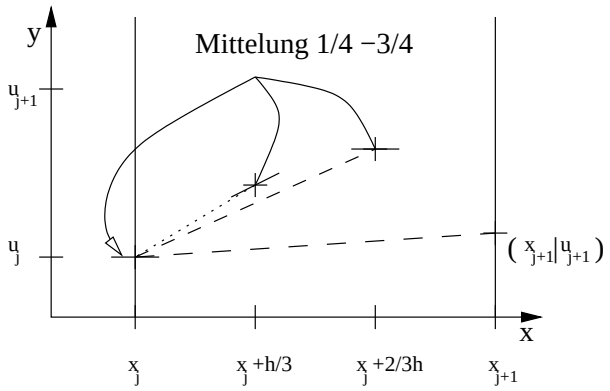


Abbildung 3.5: Runge-Kutta: Verfahren von Heun

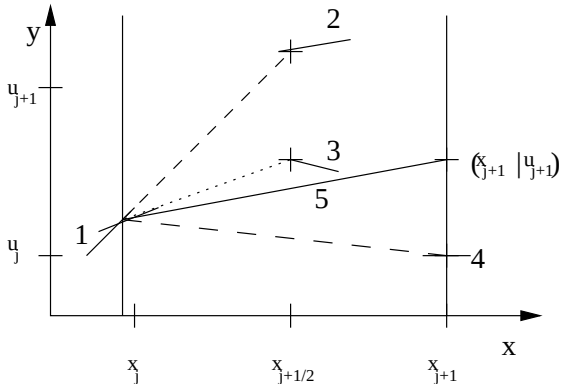


Abbildung 3.6: Runge-Kutta: klassisch

Eulerverfahren $\longleftrightarrow$ Rechteckregel
verb. Eulerverfahren $\longleftrightarrow$ Tangententrapezregel
klass. R-K. Verfahren $\longleftrightarrow$ Simpsonregel

Idee: $\frac{1}{h_j}\int_{x_j}^{x_{j+1}} f(t)\,dt$ wird ersetzt durch $\frac{1}{6}f(x_j) + \frac{2}{3}f(x_j + \frac{h_j}{2}) + \frac{1}{6}f(x_{j+1})$.

2) Zur Konstruktion von R.-K. Verfahren genügt es Spezialfälle $f(x,y) = f(y)$ zu betrachten (autonome Differentialgleichungen). Denn

(i) Koeff. von R.-K. Verfahren bleiben für Systeme von Differentialgleichungen 1. Ordnung unverändert.

(ii) Jede skalare Differentialgleichung 1. Ordnung lässt sich als autonomes System schreiben:

$$y'(x) = f(x, y(x)) \;,\quad y(a) = y_0$$

$$\Longleftrightarrow \quad \left| \begin{array}{ll} y_1'(x) = f(y_2, y_1) & , y_1(a) = y_0 \\ y_2'(x) = 1 & , y_2(a) = a \end{array} \right.$$

Hinweis: i) gilt nur wenn $\sum_{i=0}^{l-1} \beta_{e,i} = \alpha_l$.

3) Stufenzahl m und Konsistenzordnung $p = p(m)$ sind im Allgemeinen nicht korreliert:

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

Für $m \geq 10$ gilt immer $p(m) \leq m-3$. z.B. $m = 17, p = 10$.

3.34 Konstruktion von R.-K. Verfahren mit zwei Stufen
Es gilt für ESV: $\tau_h(x,y) := \frac{y(x+h)-y(x)}{h} - \varphi(x, y(x), h)$.

Ansatz für $m = 2$ und $f(x,y) = f(y)$:

$$\begin{aligned} \varphi(x, y(x), h) &= \gamma_1 k_1(x, y(x), h) + \gamma_2 k_2(x, y(x), h) \\ &= \gamma_1 f(y(x)) + \gamma_2 f\left(y(x) + h\beta_{2,1} f(y(x))\right) \end{aligned}$$

$$\overset{Taylor}{\Longrightarrow} \quad \varphi(x,y,h) = (\gamma_1+\gamma_2) f(y) + h\beta_{2,1}\gamma_2 (ff')(y) + \tfrac{h^2}{2}\gamma_2\beta_{2,1}^2 (f^2 f'')(y) + O(h^3)$$

wobei $y = y(x)$, $(f \cdot g)(y) = f(y)g(y)$; $f^{(k)} = \frac{d^k}{dy^k} f(y)$.

Andererseits folgt auch mit Taylorentwicklung

$$\frac{y(x+h)-y(x)}{h} = f(y) + \frac{h}{2}(ff')(y) + \frac{h^2}{6}(f''f^2 + f'^2 f)(y) + O(h^3)$$

Also folgt:

$$\begin{aligned} \tau_h(x,y) &= \Big(1 - (\gamma_1+\gamma_2)\Big) f(y) + h\Big(\tfrac{1}{2} - \beta_{2,1}\gamma_2\Big)(f'f)(y) \\ &\quad + \tfrac{h^2}{2}\Big(\tfrac{1}{3}(f''f^2 + f'^2 f) - \gamma_2\beta_{2,1}(f^2 f'')\Big)(y) + O(h^3) \end{aligned}$$

Bedingungen:
Für $p = 1$: $\gamma_1 + \gamma_2 = 1$
Für $p = 2$: $\frac{1}{2} - \beta_{2,1}\gamma_2 = 0$
Für $p = 3$: nicht möglich.

Wähle $\alpha_2 = \beta_{2,1} \in [0,1]$ frei!
Spezialfälle:

1) $\gamma_1 = 0, \gamma_2 = 1 \implies \beta_{2,1} = \frac{1}{2}$,

2) $\gamma_1 = \gamma_2 = \frac{1}{2} \implies \beta_{2,1} = 1$.

Bemerkung 3.35
Die Vorgehensweise für $m > 2$ ist analog, wobei für den Aufwand gilt:

Ordnung p	1	2	3	4	5	6	7	8	9	10
# Bedingungen	1	2	4	8	16	37	85	200	486	1205

Diese Bedingungen sind nichtlineare Gleichungen!

$\rightsquigarrow$ Systematisch Behandlung von Butcher 1964: "grafische Methode".

Satz 3.36

1) Hat ein Runge-Kutta Verfahren die Konsistenzordnung p, so gilt für hinreichend glattes f:

$$\tau_h(x,y) = h^p \overline{\tau}(x,y) + O(h^{p+1}) \ , \ h \longrightarrow 0$$

wobei $\overline{\tau}(x,y) = \sum_k \varepsilon_k D_k f(x,y)$ mit $D_k f \triangleq$ Produkt partieller Ableitungen, $\varepsilon_k \triangleq$ Fehlerkoeffizient.

2) Explizite R.-K. Verfahren mit Konistenzordnung p sind konvergent mit Ordnung p.

Beweis:

1) Klar nach Konstruktion in 3.31.

2) Es genügt zu zeigen, dass die k_l, $l = 1, \dots, n$ Lipschitz-stetig sind (dann folgt die Beh. mit Satz 3.23).
Beweis durch Induktion über l:
I.A. $l = 1: \quad k_1(x,u,h) = f(x,y) \implies L_1 = L$
I.S. $l \to l+1$: Seien $k_1, \dots, k_l$ Lipschitz-stetig mit Konstanten $L_1, \dots, L_l$. Dann folgt:

$$\begin{aligned} & |k_{l+1}(x,u_1,h) - k_{l+1}(x,u_2,h)| \\ = \quad & |f(x+\alpha_{l+1}h, u_1 + h\textstyle\sum_{j=1}^{l} k_j(x,u_1,h)\beta_{l+1,j}) - \\ & f(x+\alpha_{l+1}h, u_2 + h\textstyle\sum_{j=1}^{l} k_j(x,u_2,h)\beta_{l+1,j})| \\ \text{I.V.} & \\ \leq \quad & L|u_1-u_2| + Lh\textstyle\sum_{j=1}^{l} |\beta_{l+1,j}|L_j|u_1-u_2| \\ \leq \quad & \underbrace{L(1+(b-a)\sum_{j=1}^{l}|\beta_{l+1,j}|L_j)}_{=:L_{l+1}}\,|u_1-u_2|. \end{aligned}$$

Damit folgt die Behauptung.

□

Definition 3.37 (Implizite Runge-Kutta Verfahren)
Φ sei ESV mit Verfahrensfunktion φ gegeben durch

$$\varphi(x,u,h) := \sum_{i=1}^{m} \gamma_i k_i(x,u,h) \text{ mit } \sum_{i=1}^{m} \gamma_i = 1.$$

Φ heißt implizites R.-K. Verfahren mit m Stufen, falls

$$k_j(x,u,h) = f\left(x+\alpha_j h, u + h\sum_{l=1}^{m}\beta_{j,l}k_l(x,u,h)\right), \quad j = 1,\dots,m$$

wobei $\gamma_j, \alpha_j, \beta_{j,l}$ für optimale Runge-Kutta Verfahren so gewählt sind, dass φ maximale Konsistenzordnung hat.

Satz 3.38

a) Das implizite R.-K. Verfahren mit m Stufen ist wohldefiniert, falls für die Schrittweite h gilt:

$$hL \max_{k=1,\dots,m} \sum_{j=1}^{m} |\beta_{k,j}| < 1,$$

wobei L die Lipschitz Konstante von f bzgl. y ist.

b) Sei $f(x,y) = f(x)$. Dann stimmt das optimale implizite R.-K. Verfahren mit m Stufen überein mit zusammengesetzten Gauß'schen Quadraturformeln auf I mit jeweils m Knoten in den Teilintervallen $[x_j, x_{j+1}]$, $j = 0, \ldots, n-1$.

c) Optimale m-stufige implizite R.-K. Verfahren haben Konsistenz- und Konvergenzordnung $2m$, $m \geq 1$.

Beweis:

a) folgt aus der Kontraktionseigenschaft im Banach'schem Fixpunktsatz.

b) & c) siehe Deuflhard/Bornemann §6.2 und §6.3.

□

Bemerkung 3.39 (Vorteil impliziter R.-K. Verfahren)
Qualitative Eigenschaften der Lösung von (AWP), wie etwa das Monotonieverhalten, werden bei gröberen Gittern bereits wiedergegeben und die Konvergenzordnung ist sehr hoch. (Nachteil: Es ist ein System von m gekoppelten nicht-linearen Gleichungen zu lösen).

Ein Kompromiss zwischen Vor- und Nachteilen sind die sogenannten diagonal impliziten Rung-Kutta Verfahren (DIRK-Verfahren). Sie zeichnen sich dadurch aus, dass gilt $\beta_{j,l} = 0$ für $l > j$, d.h. es muessen nur m skalare nicht-lineare Gleichungen gelöst werden.

Beispiel 3.40

1) $m = 1$, $p = 2$:

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

2) $m = 2$, $p = 4$:

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

Nächstes Ziel: Schrittweitensteuerung bei ESV.
Dazu ist es notwendig eine genauere Abschätzung des globalen Fehlers zu erhalten, als diejenige, die durch das Lemma von Gronwall entsteht (vgl. Beweis von Satz 3.23).
Wir benötigen eine asymptotische Entwicklung von e_h, um

1.) den tatsächlichen Fehler zu schätzen,

2.) Extrapolation anwenden zu können,

3.) Schrittweiten steuern zu können.

Satz 3.41 (Hauptterm der asymptotischen Fehlerentwicklung bei ESV)
Φ Sei ein ESV auf äquidistantem Gitter I_h. Φ habe Konsistenzordnung p mit

$$e_0 = \rho_0 h^p + \mathcal{O}(h^{p+1}) \quad (h \longrightarrow 0),$$

$$\tau_h(x,y) = \overline{\tau}(x)h^p + \mathcal{O}(h^{p+1}).$$

φ und f seien auf ihrem Definitionsbereich in C^2, $\overline{y}$ sei die eindeutig bestimmte Lösung der linearen Störgleichung

$$(\text{SG}) \left| \begin{array}{ll} \overline{y}'(x) & = \rho(x)\overline{y}(x) - \overline{\tau}(x), \\ \overline{y}(a) & = \rho_0, \end{array} \right|$$

mit $\rho(x) := f_y(x, \tilde{y}(x)) \ \ \forall x \in I$

Dann gilt für $x_j \in I_h$:

$$u_j = \tilde{y}(x_j) + \overline{y}(x_j)h^p + \mathcal{O}(h^{p+1}) \quad (h \longrightarrow 0).$$

Beweis: Für $e_j := u_j - \tilde{y}(x_j)$, $\overline{\tau}_j = \overline{\tau}(x_j)$ gilt (vgl 3.23)

$$e_{j+1} = e_j + h(\varphi(x_j, u_j, h) - \varphi(x_j, \tilde{y}_j, h)) - h^{p+1}\overline{\tau}_j + O(h^{p+2}).$$

Mit Taylorentwicklung folgt:

$$\varphi(x_j, u_j, h) = \varphi(x_j, \tilde{y}_j + e_j, h) = \varphi(x_j, \tilde{y}_j, h) + e_j\varphi_y(x_j, \tilde{y}_j, h) + \frac{1}{2}e_j^2\varphi_{yy}(x_j, \eta_j, h)$$

für ein $\eta_j \in I(\tilde{y}_j, \tilde{y}_j + e_j)$.
Wegen der Konsistenz gilt weiter: $\lim\limits_{h \longrightarrow 0} \varphi(x, y, h) = f(x, y)$

$$\begin{array}{lrl} \Longrightarrow & \varphi_y(x, y, 0) & = \quad f_y(x, y) \\ \Longrightarrow & \varphi_y(x_j, \tilde{y}_j, h) & \overset{Taylor}{=} \quad \varphi_y(x_j, \tilde{y}_j, 0) + O(h) \\ & & = \quad f_y(x_j, \tilde{y}_j) + O(h) \\ & & \overset{Vor.}{=} \quad \rho(x_j) + O(h). \end{array}$$

Einsetzen in die Gleichung für e_{j+1} ergibt:

$$e_{j+1} = e_j(1 + h\rho(x_j)) - h^{p+1}\overline{\tau}_j + O(h^2 e_j) + O(he_j^2) + O(h^{p+2}.)$$

Setze $\overline{e}_j = h^{-p}e_j$, so folgt

$$\underbrace{\overline{e}_{j+1} = \overline{e}_j(1 + h\rho(x_j)) - h\overline{\tau}_j}_{\triangleq \, Euler-Verfahren \, angewendet \, auf \, (SG)} \quad + O(h^2\overline{e}_j) + O(h^{p+1}\overline{e}_j^2) + O(h^2)$$

Konsistenz des Euler-Verfahrens

$$\Longrightarrow \quad \overline{e}_j = \overline{y}(x_j) + O(h).$$

Also folgt

$$e_j = u_j - \tilde{y}(x_j) = h^p\overline{e}_j = h^p\overline{y}(x_j) + O(h^{p+1}).$$

□

Beispiel 3.42
Sei Φ das Eulerverfahren und $f(x, y) = \lambda y$, $\lambda \in \mathbb{R}$ y_0 gegeben. Sei $I = [0, b]$, $b > 0$ und $u_0 = y_0$.

$$\Longrightarrow \quad \tilde{y}(x) = y_0 e^{\lambda x}. \quad (*)$$

Für den Abschneidefehler gilt:

$$\begin{aligned}\tau_h(x,\tilde{y}) &= \tfrac{h}{2}\tilde{y}''(x)+O(h^2)\\ &\overset{(*)}{=} \\ &= h\underbrace{\frac{1}{2}\lambda^2 y_0 e^{\lambda x}}_{=:\bar{\tau}(x)}+O(h^2)\end{aligned}$$

sowie $\rho(x) = f_y(x,\tilde{y}) = \lambda$ und $\rho_0 = 0$.
Also lautet (SG):

$$\left|\begin{array}{l}\bar{y}'(x) = \lambda\bar{y}(x) - y_0\frac{\lambda^2}{2}e^{\lambda x}\\ y(0) = 0\end{array}\right.$$

Lösung von (SG) ist $\bar{y}(x) = -y_0\frac{\lambda^2}{2}xe^{\lambda x}$

Mit Satz 3.41 folgt:

$$\begin{aligned}u_j &= \tilde{y}_j + \bar{y}(x_j)h + O(h^2)\\ &= \tilde{y}_j - y_0\tfrac{\lambda^2}{2}x_j e^{\lambda x_j}h + O(h^2)\end{aligned}$$

Direkte Verifikation: Nach dem Euler-Verfahren ist

$$u_{j+1} = (1+\lambda h)u_j.$$

Induktion $\Longrightarrow$ $u_{j+1} = (1+\lambda h)^j y_0$ für $j = 1,\ldots,n-1$.
Weiter ist

$$\begin{aligned}&(1+h\lambda) \overset{\text{Taylor für } e^{\lambda h}}{=} e^{\lambda h} - \tfrac{h^2}{2}\lambda^2 + O(h^3)\\ \Longrightarrow\ &(1+h\lambda)^j \overset{\text{Binomi}}{=} e^{\lambda x_j} - x_j\tfrac{\lambda^2}{2}e^{\lambda x_j}h + O(h^2)\\ \Longrightarrow\ &\underbrace{(1+h\lambda)^j y_0}_{=u_j} = \underbrace{e^{\lambda x_j}y_0}_{=\tilde{y}(x_j)}\underbrace{-y_0 x_j\frac{\lambda^2}{2}e^{\lambda x_j}}_{=\bar{y}(x_j)}\cdot h + O(h^2).\end{aligned}$$

Folgerung 3.43
Gilt spezieller als in Satz 3.41

$$e_0 = \rho_0 h^p + \rho_1 h^{p+1} + \ldots + \rho_k h^{p+k} + O(h^{p+k+1})$$

und

$$\tau_n(x,y) = \bar{\tau}_o(x)h^p + \bar{\tau}_1(x)h^{p+1} + \ldots + \bar{\tau}_k h^{p+k} + O(h^{p+k+1}),$$

so gibt es Funktionen $\bar{y}_0,\ldots,\bar{y}_k$, so dass

$$u_j = \tilde{y}(x_j) + \bar{y}_0(x_j)h^p + \ldots + \bar{y}_k(x_j)h^{p+k} + O(h^{p+k+1}) \quad .$$

Die Funktionen $\bar{y}_0,\ldots,\bar{y}_k$ genügen den Störgleichungen

$$(SG_i)\left|\begin{array}{ll}\bar{y}_i(x) = \rho(x)\bar{y}_i(x) - \bar{\tau}_i(x), & \forall i = 0,\ldots,k.\\ \bar{y}_i(a) = \rho_i,\end{array}\right.$$

Beweis: Die Behauptung folgt induktiv unter Verwendung von 3.41. □

3.44 Extrapolation
Seien $I_h, I_{h'}$ zwei äquidistante Gitter auf I mit $h' = qh$, $0 < q < 1$. Sei weiter $x \in I_h \cap I_{h'}$ und es gelten die Voraussetzungen von Satz 3.41.
Seien $u_h, u_{h'}$ die Nährungslösungen des gleichen Verfahrens Φ auf I_h bzw. $I_{h'}$. Dann gelten:

1) $u_h(x) = \tilde{y}(x) + \bar{y}(x)h^p + O(h^{p+1})$,

2) $u_{h'}(x) = \tilde{y}(x) + \bar{y}(x)q^p h^p + O(h^{p+1})$.

Setzt man nun $u_h^{(1)} := \alpha u_h(x) + \beta u_{h'}(x)$ mit $\alpha = -\frac{q^p}{1-q^p}$ und $\beta = 1 - \alpha$, so folgt

$$u_h^{(1)}(x) = \tilde{y}(x) + O(h^{p+1}).$$

(Dies ist die Idee der Richardson-Extrapolation!)

3.45 Schrittweitensteuerung beim ESV
Φ sei ESV mit Verfahrensfunktion φ, I_h Gitter auf I und $u_h : I_h \longrightarrow \mathbb{R}$ Näherungslösung.
Nach Beweis von Satz 3.23 gilt:

$$\|e_h\|_h \leq \frac{1}{L}\tau e^{L(b-a)} + |e_0|e^{L(b-a)},$$

wobei τ obere Schranke für den Abschneidefehler ist.
Sei $|e_0| \leq \varepsilon$. Ist dann $\tau \leq \frac{1}{2}\frac{L\tilde{\varepsilon}}{e^{L(b-a)}} =: \eta$ und $e_0 \leq \frac{1}{2}\frac{\tilde{\varepsilon}}{e^{L(b-a)}}$, so gilt:

$$\|e_h\|_h \leq \tilde{\varepsilon}.$$

Idee: Wähle in jedem Schritt die Gitterweite $h_j > 0$ so, dass h_j maximal ist und $\tau \leq \eta$ erfüllt ist!

Darstellung des Verfahrensfehlers:
$\hat{y}$ genüge dem lokalen AWP

$$\left| \begin{array}{l} y' = f(x,y), \\ y(x^*) = z^*, \end{array} \right.$$

wobei f die Voraussetzungen aus Satz 3.41 erfülle.
Dann gilt für $x \in \{x^*, x^* + h\}$

$$u_h = \hat{y}(x) + \bar{y}_0(x)h^p + \bar{y}_1(x)h^{p+1} + O(h^{p+2}).$$

Es folgt:

$$\begin{array}{rcl} \tau_h(x^*, y) & = & -\bar{y}_0(x^* + h)h^{p-1} - \bar{y}_1(x^* + h)h^p + O(h^{p+1}) \\ & \overset{\text{Taylor}}{=} & -h^p\ \bar{y}_0'(x^*) + O(h^{p+1}). \end{array}$$

Ansatz: Berechne h näherungsweise durch

$$h^p\ |\bar{y}_o'(x^*)| = \eta \qquad (A)$$

1.Variante: Schätzung von h mittels Extrapolation.

1.) Bestimme Näherungslösung v_1 in $x^* + \tilde{h}$ ausgehend von (x^*, z^*) bei Schrittweite $\tilde{h}$.

2.) Bestimme Näherungslösung v_2 in $x^* + \tilde{h}$ ausgehend von (x^*, z^*) bei Schrittweite $\frac{\tilde{h}}{2}$.

Mit 3.41 folgt

(1) $\hat{y}(x^* + \tilde{h}) - v_1 = -\bar{y}_o'(x^*)\tilde{h}^{p+1} + O(\tilde{h}^{p+2})$,

(2) $\hat{y}(x^* + \tilde{h}) - v_2 = -\bar{y}_o'(x^*)\left(\frac{\tilde{h}}{2}\right)^{p+1} + O(\tilde{h}^{p+2})$.

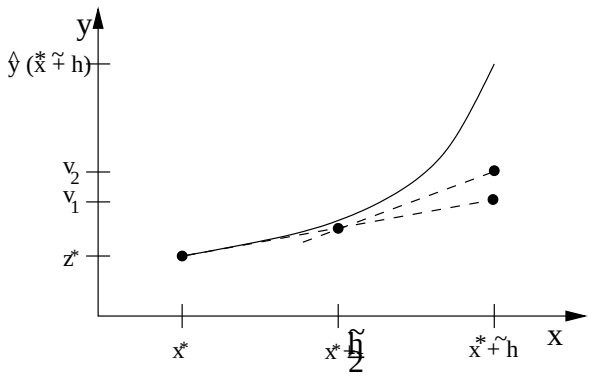


Abbildung 3.7: Schätzung mittels Extrapolation

$$\overset{(2)-(1)}{\Longrightarrow} \quad \tilde{h}^{p+1}\bar{y}_0'(x^*) = \tfrac{v_1-v_2}{1-2^{-(p+1)}} + O(\tilde{h}^{p+2}),$$
$$\Longrightarrow \quad \bar{y}_0'(x^*) = \tilde{h}^{-(p+1)}\tfrac{v_1-v_2}{1-2^{-(p+1)}} + O(\tilde{h}).$$

Einsetzen in (A) ergibt für h die Schätzung

$$h = \tilde{h}\sqrt[p]{\frac{2^{p+1}-1}{2^{p+1}}\frac{\eta\tilde{h}}{|v_1-v_2|}}.$$

Empfehlung:

Ist $h \in \left[\frac{\tilde{h}}{2}, 2\tilde{h}\right]$, so arbeite weiter mit $h_j := h$,

andernfalls führe eine neue Schätzung mit $\tilde{h} := h$ durch.

Grund: Die Schrittweite soll nicht zu stark schwanken.

Sinnvoll: Abbruchbedingung: Wenn $h < 10^{-d}$, $d > 0$, dann Abbruch.

2. Variante: Schätzung von h mit Verfahren höherer Konsistenzordnung.

Seien $\Phi, \tilde{\Phi}$ Verfahren mit Konsistenzordnung p, q, $q > p$ und seien für $\Phi, \tilde{\Phi}$ die Voraussetzungen von Satz 3.41 erfüllt.

1.) Berechne v_1 in $x^* + \tilde{h}$ ausgehend von (x^*, z^*) mit Φ und Schrittweite $\tilde{h}$.
2.) Berechne v_2 in $x^* + \tilde{h}$ ausgehend von (x^*, z^*) mit $\tilde{\Phi}$ und Schrittweite $\tilde{h}$.

Mit Satz 3.41 folgt:

$$(1) \quad \hat{y}(x^* + \tilde{h}) - v_1 = -\bar{y}_0'(x^*)\tilde{h}^{p+1} + O(h^{p+2}),$$
$$(2) \quad \hat{y}(x^* + \tilde{h}) - v_2 = -\bar{y}_0'(x^*)\tilde{h}^{q+1} + O(h^{q+2}).$$

$$\overset{(2)-(1)}{\Longrightarrow} \quad \bar{y}_0'(x^*) = \tilde{h}^{-(p+1)}\tfrac{v_1-v_2}{1-\tilde{h}^{q-p}} + O(\tilde{h}.$$

Einsetzen in (A) ergibt:

$$h = \tilde{h}\sqrt[p]{\frac{\tilde{h}\eta(1-\tilde{h}^{q-p})}{|v_1-v_2|}}.$$

Aufwandsvergleich:

Variante 1: 100% Mehraufwand (Zwischenpunkt)

Variante 2: Bei verschiedenen Verfahren i.A. Summe aus Aufwand von Φ und $\tilde{\Phi}$ (für die Praxis besser!).

Realisierung von Variante 2 mit eingebetteten R.-K. Verfahren

Idee: m-Stufen

$$\begin{array}{c|cccc} 0 & & & & \\ \alpha_2 & \beta_{2,1} & & & \\ \vdots & \vdots & \ddots & & \\ \alpha_m & \beta_{m,1} & \cdots & \beta_{m,m-1} & \\ \hline & \gamma_1 & \cdots & \gamma_{m-1} & \gamma_m \\ & \tilde{\gamma}_1 & \cdots & \tilde{\gamma}_{m-1} & \tilde{\gamma}_m \end{array}$$

wobei $\tilde{\Phi}$ definiert durch $\tilde{\gamma}_i, i = 1, \ldots, m$ die optimale Konsistenzordnung
und Φ definiert durch $\gamma_i, i = 1, \ldots, m$ eine um 1 kleinere Konsistenzordnung liefert.

Vorteil: Die Stufenfunktionen k_i müssen für Φ und $\tilde{\Phi}$ nur einmal berechnet werden.

Beispiel: Das Verfahren von Dormand-Prince ($m = 7$)

$$\begin{array}{c|ccccccc} 0 & & & & & & & \\ \frac{1}{5} & \frac{1}{5} & & & & & & \\ \frac{3}{10} & \frac{3}{40} & \frac{9}{40} & & & & & \\ \frac{4}{5} & \frac{44}{45} & -\frac{56}{15} & \frac{32}{9} & & & & \\ \frac{8}{9} & \frac{19372}{6561} & -\frac{25360}{2187} & \frac{64448}{6561} & -\frac{212}{729} & & & \\ 1 & \frac{9017}{3168} & -\frac{355}{33} & \frac{46732}{5247} & \frac{49}{176} & -\frac{5130}{18656} & & \\ 1 & \frac{35}{348} & 0 & \frac{500}{1113} & \frac{125}{192} & -\frac{2187}{6784} & \frac{11}{84} & \\ \hline \gamma & \frac{5179}{57900} & 0 & \frac{7571}{16695} & \frac{393}{640} & -\frac{92097}{339200} & \frac{187}{2100} & \frac{1}{40} \\ \tilde{\gamma} & \frac{35}{348} & 0 & \frac{500}{1113} & \frac{125}{192} & -\frac{2187}{6784} & \frac{11}{84} & 0 \end{array}$$

$\Longrightarrow \quad p(\gamma) = 4 \ , \ \ p(\tilde{\gamma}) = 5.$

3.4 Mehrschrittverfahren

Um Mehrschrittverfahren zu untersuchen benötigen wir einige Sätze aus der Theorie linearer Differenzengleichungen, die wir im folgenden Abschnitt formulieren.

3.4.1 Theorie der linearen Differenzengleichungen

Definition 3.46 (Differenzengleichung)
Sei $U := \{u : \mathbb{N}_0 \longrightarrow \mathbb{C}\}$ *die Menge der komplexen Zahlenfolgen mit* $u_i = u(i)$. *Seien* $k \in \mathbb{N}_0$, $f \in U$, $\alpha_i \in \mathbb{C}$, $i = 0, \ldots, k-1$ *gegeben. Dann heißt:*

$$(DG_k) \qquad u_{n+k} + \alpha_{k-1} u_{n+k-1} + \ldots + \alpha_0 u_n = f_n \;\; ; \;\; n \in \mathbb{N}_0$$

Differenzengleichung der Ordnung k.
(DG_k) *heißt homogen, falls* $f = 0$, *sonst inhomogen.*

Lemma 3.47
Sei eine homogene Differenzengleichung mit Ordnung k gegeben.

1) Seien $u_0, \ldots, u_{k-1} \in \mathbb{C}$ gegebene Anfangswerte einer Folge $u \in U$. Dann gilt: Es gibt genau eine Folge $u \in U$ mit diesen Anfangsdaten, welche der Differenzengleichung genügt.

2) Die Menge aller Lösungen der Differenzengleichungen bildet einen k-dimensionalen Unterraum von U. Basislösungen sind definiert durch:

$$\left|\begin{array}{l} u^{(j)} = (u_n^{(j)})_{n \in \mathbb{N}_0}, \\ u_n^{(j)} = \delta_{nj} \quad 0 \le n, j \le k-1, \\ u^{(j)} \text{ erfüllt die Differenzengleichung für } 0 \le j \le k-1, \end{array}\right.$$

Beweis: 1) Durch vollständige Induktion, 2) nachrechnen. □

Definition und Satz 3.48 (Charakterisisches Polynom)
Für (DG_k) homogen heißt

$$\rho(t) = t^k + \alpha_{k-1} t^{k-1} + \ldots + \alpha_0$$

charakterisisches Polynom. Seien $\lambda_1, \ldots, \lambda_m$ Nullstellen des Poloynoms ρ mit Vielfachheit $m_1, \ldots, m_m$ mit $\sum_{i=1}^m m_i = k$. Dann sind äquivalent:

1) u ist eine Lösung von (DG_k),

2) $u = (u_n)_{n \in \mathbb{N}_0}$, $u_n = \sum_{i=1}^m p_i(n) \lambda_i^n$, $n \in \mathbb{N}_0$, wobei $p_i \in \mathbb{P}_{m_i - 1}$ sind für $i = 1, \ldots, m$.

Beweis:

1) $\Rightarrow$ 2) $u_0, \dots, u_{k-1}$ seien Anfangswerte einer gegebenen Lösung von (DG_k). Setze

$$A:= \begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ 0 & & 0 & 1 \\ -\alpha_0 & \dots & \dots & -\alpha_{k-1} \end{pmatrix} \in \mathbb{C}^{k\times k} \text{ "Begleitmatrix von } \rho\text{"}$$

und $\vec{u}_n := \begin{pmatrix} u_n \\ \vdots \\ u_{n+k-1} \end{pmatrix} \in \mathbb{C}^k$ "Folgenabschnitt n".

Dann gilt:

$$\vec{u}_n = A\vec{u}_{n-1} \overset{\text{Induktion}}{\Longrightarrow} \vec{u}_n = A^n\vec{u}_0.$$

Nach linearer Algebra ex. eine invertierbare Matrix S, so dass $A = SJS^{-1}$, wobei J die Jordansche Normalform von A ist.

Also $J = \begin{pmatrix} J_1 & & 0 \\ & \ddots & \\ 0 & & J_n \end{pmatrix}$ mit $J_i = \begin{pmatrix} \lambda_i & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{pmatrix} \in \mathbb{C}^{m_i\times m_i}$.

Dann folgt induktiv: $A^n = SJ^nS^{-1}$, wobei

$$J_i^n = \begin{pmatrix} \lambda_i^n & \binom{n}{1}\lambda_i^{n-1} & \binom{n}{2}\lambda_i^{n-2} & \dots & \binom{n}{m_i-1}\lambda_i^{n-m_i+1} \\ 0 & \ddots & \binom{n}{1}\lambda_i^{n-1} & \dots & \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & 0 & \lambda_i^n \end{pmatrix}$$

mit $\binom{n}{j} = 0$ für $j > n$.
$\binom{n}{j}$ ist für festes j und $n \in \mathbb{N}_0$ ein Polynom vom Grad j in n.
$\Longrightarrow$ J_i^n enthält Koeffizienten nur vom Typ $p_i(n)\lambda_i^n$ mit Grad von $p_i \le m_i - 1$.
Also hat u_n die angegebene Form.

2) $\Rightarrow$ 1) Ü.A.

□

Beispiel 3.49
$u_{n+3} - 4u_{n+2} + 5u_{n+1} - 2u_n = 0$ mit Startwerten $u_0 = 1,\ u_1 = 2,\ u_2 = -1$.
Das charakteristische Polynom ist:

$$\begin{aligned} \rho(t) &= t^3 - 4t^2 + 5t - 2 \\ &= (t-1)^2(t-2), \end{aligned}$$

$\Longrightarrow$ $\lambda_1 = 1$ mit $m_1 = 2$, $\lambda_2 = 2$ mit $m_2 = 1$.

Mit Satz 3.45 folgt:

$$u_n = p_1(n)1^n + p_2(n)2^n.$$

Mit $p_1(n) = \alpha + \beta n$ und $p_2(n) = \gamma$ folgen die Bedingungen:

$$\left.\begin{aligned} 1 &= u_0 = \alpha + \gamma \\ 2 &= u_1 = \alpha + \beta + 2\gamma \\ -1 &= u_2 = \alpha + 2\beta + 4\gamma \end{aligned}\right\} \Longrightarrow \begin{aligned} \alpha &= 5 \\ \beta &= 5 \\ \gamma &= -4 \end{aligned}$$

Also ist die eindeutige Lösung der Differenzengleichunggegeben durch:

$$u_n = 5(1+n) - 2^{n+2}.$$

Bemerkung 3.50
Für die Matrix A gilt stets $Rg(A - \lambda I) \geq k - 1$ für $\lambda \in \mathbb{C}$.
Ist λ Eigenwert von A, so gilt $\dim \mathrm{Eig}(\lambda) = \dim \ker(A - \lambda I) \leq 1$.
$\Longrightarrow$ Für jeden Eigenwert λ gibt es höchstens einen Jordanblock.

Satz 3.51 (Wurzelbedingung von Dahlquist)
Sei (DG_k) eine homogene, lineare Differenzengleichung mit charakteristischem Polynom ρ. Dann sind äquivalent:

1) Jede Lösung von (DG_k) ist beschränkt.

2) Die Nullstellen λ_i von ρ erfüllen

   (i) $|\lambda_i| \leq 1$,
   (ii) $|\lambda_i| = 1 \implies \lambda_i$ ist einfache Nullstelle.

Beweis:

2) $\Rightarrow$ 1) Die Behauptung folgt direkt aus Satz 3.48, da

$$\lim_{n\to\infty} |u_n| = \lim_{n\to\infty} \sum_{i=1}^{m} |p_i(n)||\lambda_i|^n \leq C.$$

1) $\Rightarrow$ 2) Sei $u_n = \lambda_i^n$. Dann folgt aus $\lim\limits_{n\to\infty} |u_n| \leq C$ auch $\lim\limits_{n\to\infty} |\lambda_i|^n \leq C$ und somit $|\lambda_i| \leq 1$.
Sei weiter λ_i NST mit Vielfachheit $m_i \geq 2$. Dann ist $u_n = n\lambda_i^n$ eine Lösung und mit Satz 3.48 folgt:
$\lim\limits_{n\to\infty} |u_n| \leq C \implies \lim\limits_{n\to\infty} n|\lambda_i^n| \leq C \implies |\lambda_i| < 1.$

□

Definition 3.52 (Verschiebeoperatoren)
$E : U \longrightarrow U\ ,\ u \longmapsto Eu$ *mit* $(Eu)_n = u_{n+1}$

$$\Longrightarrow\ (u_0, u_1, u_2, \ldots) \longmapsto (u_1, u_2, u_3, \ldots)$$

$$E^{-1} : U \longrightarrow U\ ,\ u \longmapsto E^{-1}u \text{ mit } (E^{-1}u)_n = \begin{cases} u_{n-1}\ , & n \geq 1 \\ 0\ , & n = 0 \end{cases}$$

$$\Longrightarrow\ (u_0, u_1, u_2, \ldots) \longmapsto (0, u_0, u_1, u_2, \ldots)$$

Bemerkung 3.53

1) $EE^{-1}u = u$, aber $E^{-1}Eu = u - u_0 e^{(0)}$ mit $e_n^{(j)} := \delta_{jn}$ für $j, n \in \mathbb{N}_0$.

2) Es ist $E^j := \underbrace{E \cdot \ldots \cdot E}_{j-mal}$
$\implies \rho(E)u = f$ "Differenzengleichung" (DG_k).

Lemma 3.54
Sei (DG_k) gegeben. Definiere $v^{(j)} := E^{-j-1}u^{(k-1)}$ mit $u^{(k-1)}$ Basislösung von $\rho(E)u = 0$. Dann gilt:

$$\rho(E)v^{(j)} = e^{(j)}, \quad \text{für } j \in \mathbb{N}_0.$$

Beweis: Ü.A. □

Satz 3.55
Sei $\rho(E)u = f$ inhomogene Differenzengleichung der Ordnung k. Dann gilt:

(i) $\bar{u} := \sum_{n=0}^{\infty} f_n v^{(n)}$ ist eine Lösung.

(ii) Alle Lösungen haben die Gestalt:
$u = \tilde{u} + \bar{u}$ mit $\tilde{u}$ ist Lösung von $\rho(E)u = 0$.
($\implies$ Lösungsmenge ist k-dimensionale Untermannigfaltigkeit von U)

(iii) Für gegebenes AWP $\rho(E)u = f$ mit $u_i = \beta_i$ für $i = 0, \ldots, k-1$ gilt:

$$u = \tilde{u} + \bar{u},$$

wobei $\tilde{u}$ Lösung von $\rho(E)u = 0$ ist mit $\tilde{u}_i = \beta_i$, $i = 0, \ldots, k-1$.

Beweis: Nachrechnen. □

Bemerkung 3.56
Es gilt $v_n^{(j)} = 0$ für $n \leq k + j - 1$, also insbesondere $\bar{u}_n := \sum_{j=0}^{n-k} f_j u_{n-j-1}^{(k-1)}$.

Beispiel 3.57

1) Sei $\rho(t) = t^2 - 1$; $f = (f_n)_{n\in\mathbb{N}}$ mit $f_n := n$
$\implies$ NST von $\rho(t) : \lambda_{1/2} = \pm 1$

a) Homogene Lösung: $\tilde{u}_n = \alpha 1^n + \beta(-1)^n$ (nach Satz 3.48).

b) Spezielle Lösung des inhomogenen Problems:

1. Ansatz: $\bar{u}_n = An + B$
$$\implies \quad n = f_n = (\rho(E)\bar{u})_n = \bar{u}_{n+2} - \bar{u}_n = A(n+2) + B - An - B = 2A$$
Ansatz schlägt fehl, da A unabhängig von n gewählt werden muss!

2. Ansatz: $\bar{u}_n = An^2 + Bn$
$\Longrightarrow \quad n = f_n = \bar{u}_{n+2} - \bar{u}_n = A \cdot (n+2)^2 + B(n+2) - An^2 - Bn$
Sortiere Terme mit Faktor $n^2, n, 1$. Dann folgt durch Koeffizientenvergleich:

$$\left.\begin{array}{ll} \text{Terme mit } n^2: & \diagup \\ \text{Terme mit } n: & 4A + B - B = 1 \\ \text{Terme mit } 1: & 4A + 2B = 0 \end{array}\right\} \Longrightarrow \begin{array}{l} A = \frac{1}{4} \\ B = -\frac{1}{2} \end{array}$$

Eine spezielle Lösung ist also:

$$\bar{u}_n = \frac{1}{4}n^2 - \frac{1}{2}n$$

c) Wir erhalten eine allgemeinere Lösung: $u = \tilde{u} + \bar{u}$

$$u_n = \frac{1}{4}n^2 - \frac{1}{2}n + \alpha + \beta(-1)^n$$

2) Sei $\rho(t) = t^2 - 1$ und $f_n = n + 2^n$

a) Homogene Lösung wie in 1).

b) Inhomogene Lösung:
Ansatz: $\bar{u}_n = An^2 + Bn + C2^n$
Koeffizientenvergleich: $c = \frac{1}{3}\ ,\ A = \frac{1}{4}\ ,\ B = -\frac{1}{2}$

c) Allgemeine Lösung:

$$u_n = \frac{1}{4}n^2 - \frac{1}{2}n + \frac{1}{3}2^n + \alpha + \beta(-1)^n$$

Lösung von Differenzengleichungen im Hauptfall:

Seien $f_n := \sum_{j=1}^{i} q_j(n)\mu_j^n$ mit $\mu_j \in \mathbb{C}\, q_j \in \mathbb{P}$ mit $\text{Grad}(q_j) =: g_j$. μ_j sei m_j-fache NST von ρ mit $0 \leq m_j \leq k$ ($m_j = 0 \implies \mu_j$ keine NST von ρ).

Ansatz: $\bar{u}_n = \sum_{j=1}^{i} n^{m_j}\tilde{q}_j(n)\mu_j^n$ mit $\tilde{q}_j$ allgemeines Polynom vom Grad g_j.

In Beispiel 1): $i = 1\ ;\ \mu_1 = 1\ ;\ q_1(n) = n.$

In Beispiel 2): $i = 2\ ;\ \mu_1 = 1\ ;\ \mu_2 = 2\ ;\ q_1(n) = n\ ;\ q_2(n) = 1.$

3) Weiteres Beispiel: $f_n = n^2(-1)^n + n^4 + n^5 2^n$.
$\Longrightarrow i = 3; \mu_1 = -1; \mu_2 = 1, \mu_3 = 2, q_1(n) = n^2, q_2(n) = n^4, q_3(n) = n^5.$

3.4.2 Lineare k-Schrittverfahren

Es gelten die Voraussetzungen aus Abschnitt 3.3 für (AWP) und zusätzlich sei I_h äquidistant, d.h. $h_j = h \ \ \forall j$.

Definition 3.58 (k-Schrittverfahren)

a) Ein Verfahren Φ zur Lösung des (AWP) auf I heißt k-Schrittverfahren mit $k \in \mathbb{N}$, falls es eine Verfahrensfunktion $\varphi = \varphi(x, v_0, v_1, \ldots, v_k, h)$ mit $(x + hi, v_i) \in K_M \ \forall i = 0, \ldots, k$, $h \in (0, \frac{b-a}{k}]$, $\varphi(x, v_0, v_1, \ldots, v_k, h) \in [-M, M]$ gibt, so dass für (äquidistante) zulässige Gitter I_h mit genügend kleiner Feinheit die Näherungslösung $u_h(x_j) =: u_j$ durch die Differenzengleichung k-ter Ordnung

$$\sum_{i=0}^{k} a_i u_{j+i} = h\varphi(x_j, u_j, u_{j+1}, \ldots, u_{j+k}, h)$$

wohldefiniert ist für geeignete Startwerte, $u_0, \ldots, u_{k-1}$.

b) Ein k-Schrittverfahren heißt linear, falls φ die Form

$$\varphi(x, v_0, v_1, \ldots, v_k, h) = \sum_{i=0}^{k} b_i f(x + ih, v_i)$$

hat.

Definition und Bemerkung 3.59

1) Ist $\rho(t) = \sum_{i=0}^{k} a_i t^i$ das sogenannte 1. charakteristische Polynom von Φ, so schreiben wir

$$\Phi = (\rho, \varphi).$$

Ist das Verfahren linear und $\sigma(t) := \sum_{i=0}^{k} b_i t^i$ das so genannte 2. charakteristische Polynom von Φ, so schreiben wir

$$\Phi = (\rho, \sigma).$$

2) Implizitheit ist grundsätzlich zulässig.

3) Für $k = 1$ erhalten wir dei Definition von Einschrittverfahren.

Definition 3.60 (Abschndeidefehler und Konsistenz)

1) Für $h \in (0, \frac{b-a}{k}]$, $x \in [a, b - kh]$ und $y \in C^1(I) : y'\xi) = f(\xi, y(\xi))$ auf $[x, x + kh]$ mit $(x + hi, y(x + hi)) \in K_M, j = 0, \ldots, k$ heißt

$$\tau_h(x, y) := \frac{1}{h} \sum_{i=0}^{k} a_i y(x + hi) - \varphi(x, y(x), y(x + h), \ldots, y(x + kh), h)$$

der lokale Abschneidefehler.

b) Ist $\tau_h(x, \tilde{y}) = o(1)$ für $h \longrightarrow 0$, so heißt Φ konsistent mit (AWP), falls die Anfangswerte $u_0, \ldots, u_{k-1}$ konsistent sind, d.h. $u_i \xrightarrow{h \to 0} y_0$ für $i = 0, \ldots, k - 1$.
Ist $\tau_h(x, \tilde{y}) = O(h^p)$, $p \in \mathbb{N}$ für $h \longrightarrow 0$, so hat Φ die Konsistenzordnung p, falls die Anfangswerte die Konsistenzordnung p haben, d.h. $|u_i - y_0| = O(h^p)$ für $i = 0, \ldots, k-1$.

3.61 Mehrschrittverfahren resultierend aus Quadraturen

Ansatz: Interpolationsquadratur des Richtungsfeldes.
Aus $y' = f(x, y)$ folgt für $x_j, x_{j+k} \in [a, b]$

$$y(x_{j+k}) = y(x_{j+q}) + \int_{x_{j+q}}^{x_{j+k}} f(t, y(t))dt$$

für $0 \leq q \leq k-1$.

Idee: Approximation mit geeigneter Quadraturformel (siehe Abb. 3.8).
Approximiere $\int_{x_{j+q}}^{x_{j+k}} f(t, y(t))dt \approx h \sum_{i=0}^{s} b_i \underbrace{f(x_{j+i}, u_{j+i})}_{=:f_{j+i}}$ mit $0 \leq s \leq k$.

Erhalte MSV:

$$u_{j+k} = u_{j+q} + h \sum_{i=0}^{s} b_i f_{j+i}.$$

Ist $s < k$, oder $b_k = 0 \implies$ Verfahren ist explizit.
Ist $s = k$ und $b_k \neq 0 \implies$ Verfahren ist implizit.

Ansatz zu Bestimmung der b_i: Ersetze $f(t, y(t)$ durch Polynom p_j vom Grad s, $0 \leq s \leq k$, so dass p_j die Daten (x_{j+i}, f_{j+i}) $i = 0, \ldots, s$ interpoliert.
Berechne die Koeffizienten b_i durch Lagrange-Ansatz:
$p_j(t) = \sum_{i=0}^{s} L_{ij}(t) f_{j+i}$ mit $L_{ij}(t) = \prod_{l=0, l\neq i}^{s} \frac{t - x_{j+l}}{x_{j+i} - x_{j+l}} \implies L_{ij}(x_{j+l}) = \delta_{il}$

Dann ist:

$$b_i = \frac{1}{h} \int_{x_{j+q}}^{x_{j+k}} L_{ij}(t)dt.$$

Bemerke: Auf äquidistanten Gittern hängen die b_i nicht von h und j ab.
Im Sinne von 3.58 und 3.59 folgt:

$$u_{j+k} - u_{j+q} = h\varphi(x_j, u_j, \ldots, u_{j+k}, h)$$

mit $\varphi(x, v_0, v_i, \ldots, v_k, h) = \sum_{i=0}^{s} b_i f(x + ih, v_i)$.

$\implies$ $\rho(t) = t^k - t^q$ 1. charakteristisches Polynom,
$\implies$ $\sigma(t) = \sum_{i=0}^{s} b_i t^i$ 2. charakteristisches Polynom,
$\implies$ $\Phi = (\rho, \sigma)$.

Satz 3.62
Seien $f \in C^{s+1}(S)$ und $\Phi = (\rho, \sigma)$ ein k-Schrittverfahren der Klasse 3.61. Dann hat Φ die Konsistenzordnung $s + 1$, falls die Anfangswerte diese Konsistenzordnung haben.
(Für k gerade, $q = 0$ und $s = k$ gilt sogar "s+2"!)

Beweis: Setze $t_i = x + ih$, $i = 0, \ldots, s$. Sei $p(t_i) = y'(t_i) = f(t_i, y(t_i))$

$$\implies p(t) - f(t, y(t)) \overset{\text{Satz 1.4}}{=} \frac{1}{(s+1)!} \prod_{i=0}^{s} (t - t_i)\ y^{(s+2)}(\xi_t) \qquad (*)$$

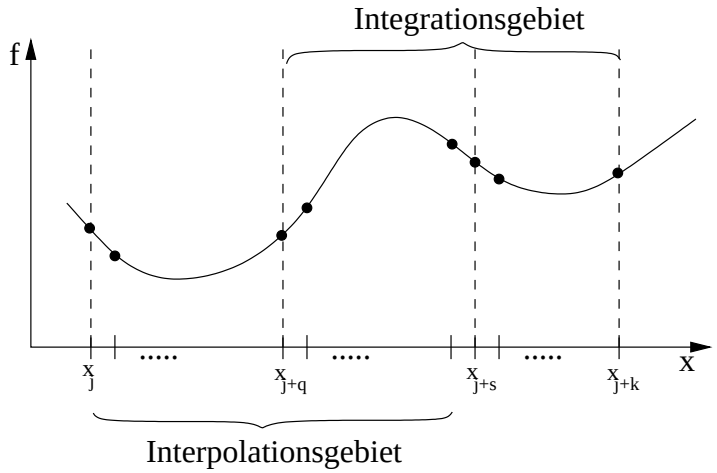


Abbildung 3.8: spezielle lineare Mehrschrittverfahren

für ein $\xi_t \in (t_0, t_s)$.

$$\begin{aligned}
\Longrightarrow \quad \tau_h(x,\tilde{y}(x)) &= \tfrac{1}{h}\left[\tilde{y}(x+kh)-\tilde{y}(x+qh)\right] - \tfrac{1}{h}\sum_{i=0}^{s} b_i f(x+ih,\tilde{y}(x+ih)) \\
&= \tfrac{1}{h}\int_{x+qh}^{x+kh}\tilde{y}'(t)dt - \tfrac{1}{h}\int_{x+qh}^{x+kh} p(t)dt \\
&= \tfrac{1}{h}\int_{x+qh}^{x+kh}\Big(f(t,\tilde{y}(t)) - p(t)\Big)dt \\
&\overset{(*)}{=} \tfrac{1}{h}\tfrac{1}{(s+1)!}\int_{x+qh}^{x+kh}\prod_{i=0}^{s}(t-t_i)\ \tilde{y}^{(s+2)}(\xi_t)dt,
\end{aligned}$$

$$\begin{aligned}
\Longrightarrow \quad |\tau_h(x,\tilde{y}(x))| &\le \tfrac{1}{h}\tfrac{1}{(s+1)!}\left\|\tilde{y}^{(s+2)}\right\|_{\infty,I}(sh)^{s+1}(k-q)h \\
&\le \underbrace{C(\tilde{y},s)}_{konstant} h^{s+1}.
\end{aligned}$$

□

3.63 Berechnung der Koeffizienten b_i

Es ist $b_{s-r} = (-1)^r \sum_{i=r}^{s}\binom{i}{r} b_i^*$, $0 \le r \le s$ mit $b_i^* = \int_{q-k}^{k-s}\binom{t+i-1}{i}dt$.
Mit $a := q-k$, $b := k-s$ entstehen die $b_i^* = b_i^*(a,b)$ als Taylorkoeffizienten der Funktion $g(z,a,b)$ bei der Entwicklung um $z=0$, wobei

$$g(z,a,b) := \frac{(1-z)^{-a}}{\ln(1-z)} - \frac{(1-z)^{-b}}{\ln(1-z)}$$

$\Longrightarrow$ $g(z,a,b) = \sum_{i=0}^{\infty} b_i^*(a,b)z^i$ mit $a < b \le 1$, $|z| \le R < 1$. g heißt erzeugende Funktion für b_i^*.

Definition 3.64

1) $q = k - 1$ *"Adams-Verfahren"*

2) $q = k - 1,\ s = k - 1$ *"Adams-Bashforth-Verfahren", explizit*

3) $q = k - 1,\ s = k$ *"Adams-Moulton-Verfahren", implizit*

4) $q = k - 2,\ s = k - 1$ *"Nyström-Verfahren", explizit*

5) $q = k - 2,\ s = k$ *"Milne-Simpson-Verfahren", implizit*

3.65 MSV resultierend aus Differentiationsformeln (BDF-Verfahren)

Ansatz: Interpoliere die Daten $(x_j, u_j), \ldots, (x_{j+k}, u_{j+k})$ mit $p(x)$ und verwende $p'(x_{j+q})$ als Approximation für $y'(x_{j+q}) = f(x_{j+q}, y(x_{j+q}))$.

$$\begin{array}{ll} q = k & \Longrightarrow \text{ implizites Verfahren} \\ 0 \le q \le k-1 & \Longrightarrow \text{ explizites Verfahren} \end{array}$$

Lagrangeansatz für $p(x)$:

$$h \cdot p'(x_{j+q}) = \sum_{i=0}^{k} \underbrace{L'_{ij}(x_{j+q})h}_{=:a_i} \cdot u_{j+i} = h \underbrace{f(x_{j+q}, u_{j+q})}_{=:\varphi},$$

wobei a_i unabhängig von j und von h ist.

$\Longrightarrow \quad \rho(t) = \sum_{i=0}^{k} a_i t^i\ ,\ \ \sigma(t) = t^q, \quad \Longrightarrow \quad \Phi = (\rho, \sigma)$ "BDF-k Verfahren".

Satz 3.66
Sei $f \in C^k(S)$. Dann gilt für das BDF-k Verfahren aus 3.65:
Die Konsistenzordnung ist $p = k$ bei hinreichend konsistenten Startwerten.

Beweis: Ü.A. □

Satz 3.67 (Charakterisierung der Konsistenz von MSV)
Sei $\Phi(\rho, \varphi)$ k-Schrittverfahren mit $k \ge 1$. Seien $u_0, \ldots u_{k-1}$ Startwerte. Dann sind äquivalent:

1) Φ ist konsistent mit (AWP).

2) a) $u_i \longrightarrow y_0$ für $h \longrightarrow 0$; $i = 0, \ldots, k-1$,
   b) $\rho(1)\tilde{y}(x) = 0$,
   c) $\varphi(x, \tilde{y}(x), \ldots, \tilde{y}(x + kh), h) - \rho'(1) f(x, \tilde{y}(x)) \longrightarrow 0$ für $h \longrightarrow 0$.

Beweis:

"1) $\Longrightarrow$ 2)"

a) folgt direkt aus der Definition von Konsistenz.

b) & c):

$$\begin{aligned}\sum_{i=0}^{k} a_i \tilde{y}(x+ih) &= \sum_{i=0}^{k} a_i \left(\tilde{y}(x) + ih\tilde{y}'(x) + o(h)\right)\\ &= \rho(1)\tilde{y}(x) + \rho'(1)h\tilde{y}'(x) + o(h)\end{aligned}$$

Es folgt:

$$\tau_h(x,\tilde{y}(x)) = \frac{1}{h}\left[\rho(1)\tilde{y}(x) + \rho'(1)h\tilde{y}(x)\right] - \varphi(x,\tilde{y}(x),\ldots,\tilde{y}(x+kh),h) + o(1) \quad (*)$$

$$\begin{aligned}\left|\tfrac{\rho(1)\tilde{y}(x)}{h}\right| &\leq |\tau_h(x,\tilde{y}(x))| + |\rho'(1)f(x,\tilde{y}(x)) - \varphi(\ldots)| + o(1)\\ &\leq C \text{ unabhängig von h}\end{aligned}$$

Also folgt $\rho(1)\tilde{y}(x) = 0$.
Einsetzen in (*) ergibt dann:
$\rho'(1)f(x,\tilde{y}(x)) - \varphi(x,\tilde{y}(x),\ldots,\tilde{y}(x+kh),h) = o(1)$ $(h \to 0)$. $\implies$ 2c).

"2) $\implies$ 1)" Folgt direkt aus (*).

□

Folgerung 3.68
Ist $\Phi = (\rho,\sigma)$ ein lineares k-Schrittverfahren, so gilt:
Φ ist konsistent mit (AWP), genau dann wenn 3.67 2a) und $\rho(1) = 0,\ \rho'(1) = \sigma(1)$ erfüllt ist.

Beweis: Zu zeigen: $\rho'(1) = \sigma(1)$ $\iff$ 3.67 2c) gilt (2a) und 2b) gelten sofort).

Da $\varphi(x,\tilde{y}(x),\ldots,\tilde{y}(x+kh),h) = \sum_{i=0}^{k} b_i f(x+ih,\tilde{y}(x+ih)) \longrightarrow \underbrace{\sum_{i=0}^{k} b_i}_{=\sigma(1)} f(x,\tilde{y}(x))$

für $h \longrightarrow 0$, da $f,\tilde{y}$ stetig.
$\implies$ 2c) ist erfüllt $\iff$ $\rho'(1) = \sigma(1)$.

□

Bemerkung 3.69
Konsistenz und Lipschitzstetigkeit von φ bzgl. $v_0,\ldots,v_k$ ist *nur* für $k = 1$ hinreichend für die Konvergenz! (vgl. Übungsaufgabe).

Definition 3.70 (Asymptotische Stabilität)
Sei $\Phi = (\rho,\varphi)$ k-Schrittverfahren mit $k \geq 1$. Sei $C_h := \{u : I_h \longrightarrow \mathbb{R}\}$ der Funktionenraum der Gitterfunktionen. Sei $F_h : C_h \longrightarrow C_h$ für $v \in C_h$ definiert durch

$$\begin{aligned}(F_h(v))_i &:= v_i - u_i\ ,\quad i = 0\ldots k-1\ ,\quad u_i \text{ Startwerte,}\\ (F_h(v))_{j+k} &:= \tfrac{1}{h}\sum_{i=0}^{k} a_i v_{i+j} - \varphi(x_j,v_j,\ldots,v_{j+k},h) \text{ für } j = 0,\ldots,n-k.\end{aligned}$$

F_h heißt Defektfunktion.

Das Verfahren Φ heißt asymptotisch stabil, genau dann wenn
$\exists K,H > 0$: $\forall h \in (0,H)$ und $\forall v,w \in C_h$:

$$\|v-w\|_h \leq K\,\|F_h(v) - F_h(w)\|_h\,.$$

Bemerkung 3.71

1) Durch $F_h(u) = 0$ wird die Verfahrenslösung des k-Schrittverfahrens Φ für (AWP) implizit beschrieben, falls $u_0, \dots, u_{k-1}$ Startwerte sind.

2) Ist $F_h(v_h) = \varepsilon_h$ mit ε_h Störtertm (etwa durch Rundungsfehler), so folgt aus der Stabilität:

$$\|u_h - v_h\|_h \leq K \| \underbrace{F(u_h)}_{=0} - \underbrace{F(v_h)}_{=\varepsilon_h} \|_h = K \, \|\varepsilon_h\|_h$$

3) $F_h(\tilde{y}_{j+k}) = \tau_h(x_{j+k}, \tilde{y})$

Satz 3.72 (Charakterisierung stabiler k-Schrittverfahren)
Sei $\Phi = (\rho, \varphi)$ k-Schrittverfahren mit $k \geq 1$. Gelte:

(i) $f = 0 \implies \varphi = 0$ (bei linearen Verfahren stets erfüllt),

(ii) φ sei Lipschitzstetig bzgl. $v_o, \dots, v_k$, d.h. es ex. ein $\tilde{L} > 0$ mit

$$|\varphi(x, v_o, \dots, v_k, h) - \varphi(x, w_o, \dots, w_k, h)| \leq \tilde{L} \max_{0 \leq i \leq k} |v_i - w_i|$$

(bei linearen Verfahren erfüllt, falls L-Bedingung gilt.)

Dann sind äquivalent:

a) Φ ist asymptotisch stabil.

b) ρ erfüllt die Wurzelbedingung von Dahlquist (vgl. Satz 3.51).

Beweis: Gelte stets (i), (ii).

a) $\implies$ b) Sei Φ asymptotisch stabil. Dann folgt für $f = 0$ $\varphi = 0$ nach (i) und es ist

$$(F_h(v))_i = v_i - u_i \,, \quad i = 0, \dots, k-1,$$

$$(F_h(v))_{j+k} = \tfrac{1}{h} \sum_{i=0}^{k} a_i v_{j+i} \,, \quad j = 0, \dots, n-k.$$

Nach a) $\exists K, H > 0$:

$$\|u - v\|_h \leq K \, \|F_h(u) - F_h(v)\|_h = K \, \|F_h(u - v)\|_h$$

$\forall h \in (0, H)$ und $\forall u, v \in C_h$.
Also gilt für alle $w \in C_h$: $\|w\|_h \leq K \, \|F_h(w)\|_h$.
Für die Lösung $\tilde{u}$ der Gleichung $F_h(\tilde{u})_{j+k} = 0$, $j = 0, \dots, n-k$ (homogene Gleichung) folgt:

$$\|\tilde{u}\|_h \leq K \max_{0 \leq i \leq k-1} |\tilde{u}_i| \,.$$

$\implies$ Für beliebige Startwerte $\tilde{u}_0, \dots, \tilde{u}_{k-1}$ ist die Lösung der homogenen k-stufigen Differenzengleichung $\sum_{i=0}^{k} a_i v_{j+i} = 0$ und $v_i = \tilde{u}_i$ für $i = 0, \dots, k-1$ beschränkt.
Also folgt mit Satz 3.51, dass ρ die Wurzelbedingung von Dahlquist erfüllt.

b) $\Longrightarrow$ a) Seien $v, w \in C_h$, dann gilt:

$(F_h(v) - F_h(w))_i = v_i - w_i\,, \quad i = 0, \ldots, k-1$

$(F_h(v) - F_h(w))_{j+k} = \frac{1}{h}\sum_{i=0}^{k} a_i(v_{j+i} - w_{j+i})$

$\quad -\varphi(x_j, v_j, \ldots, v_{j+k}, h) + \varphi(x_j, w_j, \ldots, w_{j+k}, h)$ für $j = 0, \ldots, n-k$.

Setze $\delta_\mu := v_\mu - w_\mu\,, \quad \mu = 0, \ldots, n$ und $d_\mu := (F_h(v) - F_h(w))_\mu,\ \mu = 0, \ldots, n$ sowie

$$(*)\quad \eta_\mu := d_{\mu+k} + \varphi(x_\mu, v_\mu, \ldots v_{\mu+k}, h) - \varphi(x_\mu, w_\mu, \ldots w_{\mu+k}, h)$$

Dann gilt $\sum\limits_{i=0}^{k} a_i\delta_{i+j} = h\eta_j$ für $j = 0, \ldots, n-k$.

Sei δ_μ vorgegeben für $\mu = 0, \ldots, k-1$. Mit Satz 3.55 folgt

$$(**)\quad \delta_{j+k} = \sum_{i=0}^{k-1} \delta_i u_{j+k}^{(i)} + \sum_{\mu=0}^{j} h\eta_\mu u_{j+k-\mu-1}^{(k-1)},$$

wobei $u^{(\nu)}$ für $\nu = 0, \ldots, k-1$ Basislösung der homogenen Differenzengleichung ist, d.h. $\sum_{i=0}^{k} a_i u_{j+i}^{(\nu)} = 0$ und $u_n^{(\nu)} = \delta_{n,\mu}$ für $0 \le n, \mu \le k-1$.

Nach Voraussetzung a) und Satz 3.51 folgt $\left|u_n^{(\nu)}\right| \le M, \quad n \in \mathbb{N}_0, \quad 0 \le \nu \le k-1$. Mit (*) gilt

$$|\eta_\mu| \overset{(ii)}{\le} |d_{\mu+k}| + \tilde{L} \underbrace{\max_{0\le i\le k} |\delta_{\mu+i}|}_{=:\varepsilon_\mu}.$$

Also folgt aus (**) und $\delta_i = d_i$ für $i = 0, \ldots, k-1$:

$$|\delta_{j+k}| \le Mdk + Mh\sum_{\mu=0}^{j}(d + \tilde{L}\varepsilon_\mu) = Md(k + (j+1)h) + Mh\tilde{L}\sum_{\mu=0}^{j}\varepsilon_\mu,$$

mit $d := \|F_h(v) - F_h(w)\|_h = \max\limits_{0\le\mu\le n} |d_\mu|$.

Beachte: Die rechte Seite ist monoton wachsend in j. Also folgt

$$\varepsilon_j \le Md(k + (j+1)h) + hM\tilde{L}\sum_{\mu=0}^{j}\varepsilon_\mu,$$

$$\Longrightarrow\quad \varepsilon_j(1 - hM\tilde{L}) \le Md(k + (j+1)h) + hM\tilde{L}\textstyle\sum_{\mu=0}^{j-1}\varepsilon_\mu.$$

Für $h \le H := \frac{1}{2M\tilde{L}}$ ist $hM\tilde{L} \le \frac{1}{2}$ und wir erhalten

$$\varepsilon_j \le \underbrace{2M(k + (b-a))}_{=:c} d + \underbrace{2M\tilde{L}}_{=:D}\sum_{\mu=0}^{j-1} h\varepsilon_\mu.$$

Mit diskretem Lemma von Gronwall folgt:

$$\varepsilon_j \le ce^{D(b-a)}d, \quad \text{für } j = 0, \ldots, n-k.$$

Mit $K := ce^{D(b-a)}$ folgt hieraus schließlich

$$\|v - w\|_h \le K\,\|F_h(v) - F_h(w)\|_h\,.$$

□

Folgerung 3.73

(i) Für L-stetige φ sind ESV stabil.

(ii) $\Phi(\rho, \varphi)$ und φ L-stetig, dann gilt:
Φ stabil $\iff$ $\Phi_0 = (\rho, 0)$ stabil.

Satz 3.74 (Konvergenzsatz von Dahlquist)
Sei $\Phi = (\rho, \varphi)$ k-Schrittverfahren mit $k \geq 1$. Sei Φ konsistent mit (AWP) von der Ordnung $p \geq 1$ und seien die Voraussetzungen aus Satz 3.72 erfüllt. Dann sind äquivalent:

(i) Φ ist konvergent mit Ordnung p.

(ii) Φ ist asymptotisch stabil.

Beweis:

(i) $\Longrightarrow$ (ii) Indirekter Beweis. Annahme: Φ ist nicht stabil.
Dann folgt mit den Sätzen 3.51 und 3.72, dass ρ die Wurzelbedingung nicht erfüllt. Folglich hat $\rho(E)u = 0$ eine unbeschränkte Lösung.
Wähle $f = 0, \varphi = 0$ und konsistente Startwerte, so dass u_n divergiert für $h \longrightarrow 0$.Dies ist ein Widerspruch zu (i).

(ii) $\Longrightarrow$ (i) Sei $\tilde{y}_h := \tilde{y}|_{I_h}$ Dann gilt:
$(F_h(\tilde{y}_h))_i = \tilde{y}(x_i) - u_i\ ,\ \ 0 \leq i \leq k-1,$
$(F_h(\tilde{y}_h))_{j+k} = \tau_h(x_j, \tilde{y})\ ,\ \ j = 0, \ldots, n-k.$
Also folgt:

$$F_h(\tilde{y}_h) - F_h(u_h) = F_h(\tilde{y}_h) - 0$$

und somit

$$\|\tilde{y} - u_h\|_h \overset{\Phi\text{stabil}}{\leq} K \|F_h(\tilde{y}_h) - F_h(u_h)\|_h = K \|F_h(\tilde{y}_h)\|_h = (O(h^p)),$$

da Φ konsistent von der Ordnung p ist.

□

Bemerke: Aus dem Beweis des Konvergenzsatzes von Dahlquist folgt für ein asymptotisch stabiles k-Schrittverfahren $\Phi = (\rho, \varphi)$ die Fehlerabschätzung

$$\|\tilde{y} - u_h\|_h \leq K \max\Big(\max_{i=0}^{k-1} |\tilde{y}_i - u_i|, \max_{i=0}^{n-k} \tau_h(x_i, \tilde{y}) \Big).$$

Folgerung 3.75

1) Für MSV der Klasse 3.61 gilt:
Ist $0 \leq q < k$ und $k - 1 \leq s \leq k$, so ist das Verfahren konvergent mit Ordnung $s + 1$.

2) Für BDF-Verfahren (3.65) gilt:
Die Verfahren sind konvergent, falls

a) $q = k \leq 6$ (implizit),
b) $q = k - 1 \leq 1$ (explizit).

Beweis:

1) Nach Satz 3.62 sind die MSV mit Ordnung $s+1$ konsistent für $f \in C^{s+1}(S)$ und es ist $\rho(t) = t^k - t^q = t^q(t^{k-q} - 1)$.
$\Longrightarrow$ NST von ρ sind: $\lambda_1 = 0$ (q-fache NST) und $\lambda_{j+2} = e^{i\frac{2\pi}{k-q}j}$ für $j = 0, \ldots, k-q-1$ (einfache NSTen).
$\Longrightarrow$ ρ erfüllt die Wuzelbedingung.

$\overset{3.72}{\Longrightarrow}$ Verfahren ist asymptotisch stabil.

$\overset{3.74}{\Longrightarrow}$ Konvergenz mit Ordnung $s+1$.

2) Ü.A.

□

Beispiel 3.76 (Milne-Simpson für $k = 2$)
Verfahren:

$$u_{j+2} - u_j = \frac{h}{3}[f(x_{j+2}, u_{j+2}) + 4f(x_{j+1}, u_{j+1}) + f(x_j, u_j)]$$

Betrachten wir dieses Verfahren für $y' = \lambda y, \lambda < 0, y(0) = 1$:
$\Longrightarrow$ $\tilde{y}(x) = e^{\lambda x}$.
Für $t := \lambda h$ folgt für das Verfahren:

$$u_{j+2}(1 - \frac{t}{3}) - \frac{4}{3}tu_{j+1} - (1 + \frac{t}{3})u_j = 0,$$

$\Longrightarrow$ $\rho_t(\mu) = \mu^2(1 - \frac{t}{3}) - \mu\frac{4}{3}t - (1 + \frac{t}{3})$.

Nullstellen:

$$\mu_1 = 1 + t + O(t^2) = e^t(1 + O(t^2)),$$
$$\mu_2 = -1 + \tfrac{t}{3} + O(t^2) = -e^{-\frac{t}{3}}(1 + O(t^2))$$

Mit den Startwerten $u_0 = 1, u_1 = 1 + t$ folgt:

$$u_j = [1 + \frac{\lambda h}{2} + O(\lambda^2 h^2)] \underbrace{e^{\lambda x_j}}_{guter\ Term\ (konvergent)} (1 + O(\lambda h))$$
$$+[-\frac{\lambda h}{2} + O(\lambda^2 h^2)] \underbrace{e^{-\lambda x_j}(-1)^j}_{parzitärer\ Term\ (divergent)} (1 + O(\lambda h)).$$

Der parazitäre Term wird für $\lambda < 0$ exponentiell groß und dominiert das Verhalten.

Satz 3.77 (Konsistenzordnungskriterien für lineare MSV)
Sei $\Phi = (\rho, \sigma)$ lineares MSV mit $k \geq 1$, d.h.

$$\sum_{i=0}^{k} a_i u_{j+i} = h \sum_{i=0}^{k} b_i f_{j+i} \quad \text{mit} \quad f_{j+i} := f(x_{j+1}, u_{j+i}).$$

Für $g \in C^1(\tilde{I}), \tilde{I} = [0, b-a]$ definiere Operator

$$K(g,h) := \frac{1}{h} \sum_{i=0}^{k} (a_i g(ih) - h b_i g'(ih)).$$

Sei $f \in C^p(S)$. Dann sind äquivalent:

(i) $\tau_h(x, \tilde{y}) = O(h^p)$,

(ii) $K(t^s, 1) = 0$ für $s = 0, \ldots, p$,

(iii) $\lambda = 0$ ist $(p+1)$-fache NST von $\rho(e^\lambda) - \lambda\sigma(e^\lambda)$,

(iv) $\lambda = 1$ ist p-fache NST von $\frac{\rho(\lambda)}{\ln(\lambda)} - \sigma(\lambda)$

(v) Die Konsistenzordnung für $y' = y, y(0) = 1$ ist p.

(vi) Die Konsistenzordnung für $y' = sx^{s-1}, y(0) = 1$ ist p für alle $s = 0, \ldots, p$.

Beweis: Wir zeigen die Äquivalenz von (i), (ii) und (iii). Für die Äquivalenz zu (iv), (v) und (vi) siehe Ü.A.

(i) $\iff$ (ii): Ist $f \in C^p(S)$, so folgt $\tilde{y} \in C^{p+1}(I)$
Taylorentwicklung von $\tilde{y}$ und $\tilde{y}'$ liefert:

(*) $\left| \begin{array}{l} \tilde{y}(x+ih) = \sum_{s=0}^{p} \frac{\tilde{y}^{(s)}(x)}{s!} i^s h^s + O(h^{p+1}), \\ \tilde{y}'(x+ih) = \sum_{s=1}^{p} \frac{\tilde{y}^{(s)}(x)}{s!} \cdot s i^{s-1} h^{s-1} + O(h^p). \end{array} \right.$

Mit $x_i = x + ih$ folgt für den Abschneidefehler:

$$\begin{aligned} h\tau_h(x, \tilde{y}) &= \sum_{i=0}^{k} a_i \tilde{y}(x_i) - h b_i \tilde{y}'(x_i) \\ &\overset{(*)}{=} \rho(1)\tilde{y}(x) + \sum_{s=1}^{p} \left(\frac{y^{(s)}(x)}{s!} h^s \left(\sum_{i=0}^{k} (a_i i^s - b_i s i^{s-1}) \right) \right) + O(h^{p+1}) \\ &\overset{\text{Def. } K(g,h)}{=} \sum_{s=0}^{p} h^s \tilde{y}^{(s)}(x) \tfrac{1}{s!} K(t^s, 1) + O(h^{p+1}). \end{aligned}$$

Also folgt die Behauptung.

(ii) $\iff$ (iii): Nach Teil 1 gilt:

$$h\tau_h(x, e^x) = e^x \sum_{s=0}^{p} \frac{h^s}{s!} K(t^s, 1) + O(h^{p+1}).$$

Andererseits folgt aus der Definition:

$$\begin{aligned} h\tau_h(x, e^x) = hK(e^{x+t}, h) &= e^x \sum_{i=0}^{k} (a_i e^{ih} - h b_i e^{ih}) \\ &= e^x(\rho(e^h) - h\sigma(e^h)). \end{aligned}$$

Also folgt

$$\underbrace{\rho(e^h) - h\sigma(e^h)}_{=:I(h)} = \sum_{s=0}^{p} \frac{h^s}{s!} K(t^s, 1) + O(h^{p+1}).$$

Entwicklung von I in $h = 0$ ergibt die Behauptung. □

Folgerung 3.78

1) Ist $g(t) = y(x + t)$, so folgt $K(g, h) = \tau_h(x, y)$.

2) Ist $f \in C^q(S)$ und $q \geq p$, p Konsistenzordnung, so gilt:

$$\tau_h(x, \tilde{y}) = \sum_{s=p+1}^{q} h^{s-1} \frac{\tilde{y}^{(s)}(x)}{s!} K(t^s, 1) + O(h^q).$$

$K(t^{p+1}, 1)$ heißt "Hauptfehlerkoeffizient".

Bemerkung 3.79

1) Für *spezielle Klassen* von linearen MSV kann analog zu Satz 3.41 eine asymptotische Entwicklung des globalen Fehlers angegeben werden.

2) Insbesondere gilt bei MSV: Die Existenz einer asymptotischen Entwicklung hängt von der Wahl der Startwerte ab.
Siehe hierzu auch Stoer/Bulirsch: Numerische Mathematik II [8].

3.4.3 Das Extrapolationsverfahren von Gragg

Idee: Anwendung der Richardsonextrapolation auf Mehrschrittverfahren.

Definition 3.80 (Mittelpunktverfahren)
Sei $u_h : I_h \longrightarrow \mathbb{R}$ *definiert durch* $u_h(x_i) = u_i$, $i = 0, \ldots, n$ *mit*

$$\left| \begin{array}{l} u_0 = y_0, \\ u_1 = u_0 + hf(x_0, u_0), \\ u_{i+1} - u_{i-1} = 2hf(x_i, u_i), \ i = 1, \ldots, n-1. \end{array} \right.$$

Satz 3.81 (Satz von Gragg)

Für genügend glattes f gilt für den globalen Fehler des Mittelpunktverfahrens:

$$u_j = \tilde{y}(x_j) + \sum_{i=1}^{N} h^{2i}(v_{2i}(x_j) + \underbrace{(-1)^j w_{2i}(x_j)}_{''oszillierender\ Term''}) + \mathcal{O}(h^{2N+2})$$

mit $n \in \mathbb{N}$ und v_{2i}, w_{2i} unabhängig von h.

(ohne Beweis)

Problem: Der oszillierende Term führt zu numerischer Auslöschung.

Definition 3.82 (Graggsche Funktion)
Setze

$$s_j := S(x_j, h) := \frac{1}{2}[u_j + u_{j-1} + hf(x_j, u_j)],$$

wobei u_j Lösung des Mittelpunktverfahrens ist.
Dann folgt aus Satz 3.81:

$$s_j = \tilde{y}(x_j) + h^2[v_{21}(x_j) + \frac{1}{4}\tilde{y}''(x_j)] + \mathcal{O}(h^4)$$

D.h. die Graggsche Funktion "glättet" den oszillierenden Term.

Algorithmus 3.83 (Algorithmus von Gragg)

Seien $f(x,y), x_0, y_0, H > 0$ und $n \in \mathbb{N}$ gegeben.
Setze $\bar{x} = x_0 + H, h := \frac{H}{n}, x_i := x_0 + ih$, also $x_n = \bar{x}$.
Gesucht ist eine Näherung für $\tilde{y}(\bar{x})$.
Definiere dazu:

$$\begin{aligned} u_0 &= y_0, \\ u_1 &= u_0 + hf(x_0, u_0), \\ u_{i+1} &= u_{i-1} + 2hf(x_i, u_i). \end{aligned}$$

Dann ist

$$S(\bar{x}, h) = \frac{1}{2}[u_n - u_{n-1} + hf(x_n, u_n)]$$

eine gute Näherung von $\tilde{y}(\bar{x})$.

Extrapolationsschema:

Sei $(n_i)_{i\in\mathbb{N}}$ eine Folge mit $0 < n_0 < n_1 < n_2 < \ldots$, z.B. Rombergfolge $n_i := 2^i$.
Setze $h_i = \frac{H}{n_i}$ für $i \in \mathbb{N}_0$

Berechne: $S(\bar{x}, h_0), S(\bar{x}, h_1), S(\bar{x}, h_2), \ldots$

Mit dem Neville-Aitken Schema berechnet man dann

$$\begin{array}{lllll} S(\bar{x}, h_0) = & T_{00} & \searrow & & \\ S(\bar{x}, h_1) = & T_{10} & \to & T_{11} & \\ \vdots & \vdots & & & \ddots \\ S(\bar{x}, h_i) = & T_{i0} & \to & T_{1i} & \ldots \quad T_{ii} \end{array}$$

$$T_{ij} = \frac{-h_{i-j} T_{i,j-1} + h_i T_{i-1,j-1}}{h_i - h_{i-j}}$$

T_{ii} ist Extrapolation in $h = 0 \implies T_{ii} \approx \tilde{y}(\bar{x})$.

3.4.4 Prädiktor-Korrektor-Verfahren

Definition 3.84
Sei $\Phi^* = (\rho^*, \sigma^*)$ *ein explizites lineares k-Schrittverfahren*

$$\sum_{i=0}^{k} a_i^* u_{j+i} = h \sum_{i=0}^{k-1} b_i^* f_{j+i}$$

mit $a_k^* = 1$. Φ^* *heißt Prädiktorformel.*

Sei $\Phi = (\rho, \sigma)$ *implizites lineares k-Schrittverfahren*

$$\sum_{i=0}^{k} a_i u_{j+i} = h \sum_{i=0}^{k} b_i f_{j+i}$$

mit $a_k = 1, b_k \neq 0$. Φ *heißt Korrektorformel*

Algorithmus:

(P) $u_{j+k}^{(0)} + \sum_{i=0}^{k-1} a_i^* u_{j+i} = h \sum_{i=0}^{k-1} b_i^* f_{j+i}$
`Für` $\mu = 1, \ldots, l$:
  (E) $f_{j+k}^{(\mu-1)} := f(x_{j+k}, u_{j+k}^{(\mu-1)})$
  (C) $u_{j+k}^{(\mu)} + \sum_{i=0}^{k-1} a_i u_{j+i} = h b_k f_{j+k}^{(\mu-1)} + h \sum_{i=0}^{k-1} b_i f_{j+i}$
(E) `Setze` $u_{j+k} := u_{j+k}^{(l)}$; $f_{j+k} := f(x_{j+k}, u_{j+k}^{(l)})$

Das Verfahren heißt $P(EC)^l E$*-Verfahren.* $(P) \triangleq$ *explizite Näherung,* $(EC)^l \triangleq$ *Fixpunktiteration für das implizite Verfahren.*

Bemerkung 3.85
1) Die $P(EC)^l E$-Verfahren sind explizite, i. A. nichtlineare Mehrschrittverfahren.

2) Alternativ zum letzten E-Schritt kann man setzen:

$$u_{j+k} := u_{j+k}^{(l)}, \quad f_{j+k} := f(x_{j+k}, u_{j+k}^{(l-1)}) = f_{j+k}^{(l-1)}.$$

Dieses Verfahren heißt $P(EC)^l$-Verfahren und ist für die Praxis wichtiger.

Satz 3.86
Sei Φ^* explizites k-Schrittverfahren mit Konsistenzordnung $p^* \geq 1$ und Φ sei implizites k-Schrittverfahren mit Konsistenzordnung $p \geq 1$. Dann gilt für das zugehörige $P(EC)^l E$-Verfahren:

1) $P(EC)^l E$ ist konsistent mit Ordnung $\bar{p} := \min\{p^* + l, p\}$

2) Erfüllt ρ die Wurzelbedingung, so ist das $P(EC)^l E$- Verfahren konvergent mit Ordnung $\bar{p}$, falls die Startwerte konvergent mit Ordnung $\bar{p}$ sind.

3) Ist $p < p^* + l$, so hat das $P(EC)^l E$-Verfahren den Hauptfehlerkoeffizienten des Korrektorverfahrens.

(ohne Beweis)

Bemerkung 3.87

1) Eine analoge Aussage zu Satz 3.86 gilt auch für die $P(EC)^l$-Verfahren.

2) In der Praxis wählt man häufig Adams-Bashforth (P) und Adams-Moulton (C) der gleichen Konsistenzordnung und führt nur $l = 1$ Iterationen durch. Man profitiert von dem deutlich kleineren Hauptfehlerkoeffizienten.

3) Kennt man die Hauptfehlerkoeffizienten zu (P) und (C), so kann man anlog zum Vorgehen bei ESV eine Schrittweitenkontrolle für die $P(EC)^l E$ erhalten.

3.5 Steife Differentialgleichungen und Stabilitätsbegriffe

Beispiel 3.88
Betrachte AWP

$$\left| \begin{array}{l} y'(x) = q(y - g(x)) + g'(x) \\ y(a = 0) = y_0 \; ; \;\; g_0 := g(a) \end{array} \right.$$

Sei $q < 0$, $|q| \gg 0$, $|g'(x)| \ll 1$.
Dann ist die exakte Lösung gegeben durch:

$$\tilde{y}(x) = g(x) + e^{qx}(y_0 - g_0).$$

Für $\tilde{y}_j := \tilde{y}(x_j)$; $g_j := g(x_j)$ folgt:

$$\tilde{y}_{j+1} - g_{j+1} = \underbrace{e^{qh}}_{<1}(\tilde{y}_j - g_j).$$

$\Longrightarrow$ Für die exakte Lösung wird der Abstand von $\tilde{y}$ und g monoton kleiner.

Ziel: Das numerische Verfahren soll dieses Verhalten reproduzieren.

a) Euler explizit: Startwert $u_0 = y_0$ und $h > 0$:

$$\Longrightarrow \quad u_{j+1} - g_{j+1} = (1 + hq)(u_j - g_j) + O(h^2)$$

Bei der Näherung erhält man nur dann eine Dämpfung, wenn $|1 + hq| < 1$ $\Longrightarrow$ Bedingung an h $(h < -1/q)$.

b) Euler implizit: Startwert $u_0 = y_0$ und $h > 0$:

$$\Longrightarrow \quad u_{j+1} - g_{j+1} = \underbrace{\frac{1}{1 - hq}}_{<1,\ \forall h>0}(u_j - g_j) + O(h^2)$$

$\Longrightarrow$ Immer Dämpfung, unabhängig von h!

Allgemeine Situation 3.89
Seien u und v Lösungen des Systems

$$y' = f(x, y) \;\; \text{im} \;\; \mathbb{R}^n$$

mit $u(a) = u_0$, $v(a) = v_0$. Dann ist:

$$u'(x) - v'(x) = J(x)(u(x) - v(x))$$

mit

$$J(x) = \int_0^1 f_y(x, tu(x) + (1 - t)v(x))dt.$$

[Für $n = 1$, z.B. $f(x, y) = qy$, $v(x) = g(x)$ (siehe Beispiel 3.85)]

Definition 3.90 (Steife Differentialgleichung)
Das AWP $y' = f(x, y)$ heißt steife Differentialgleichung *rechts von a, falls gilt:*

1) λ EW von $J(x)$ $\implies$ Re $\lambda < 0$.

2) Für $n \geq 2$ gilt:
$\exists$ EW λ_1 mit $\lambda_1 < 0$ und $|\lambda_1| \gg 0$ und $\exists$ EW λ_2 mit $|\lambda_2| \ll |\lambda_1|$.

Beispiel 3.91

y_i $i = 1, 2, 3$ seien Konzentrationen von Substanzen zur Zeit t, k_i seien Reaktionsraten:

$$\begin{aligned} y_1' &= -k_1 y_1 + k_2 y_2 y_3, \\ y_2' &= k_1 y_1 - k_2 y_2 y_3 - k_3 y_2^2, \\ y_3' &= k_3 y_2^2. \end{aligned}$$

Seien $y_1(0) = 1$, $y_2(0) = y_3(0) = 0$.

Dann folgt für die EW von $J(x)$

x	λ_1	λ_2	λ_3
0	0	0	$-0,04$
10^{-2}	0	$-0,36$	-2180
100	0	$-0,0048$	-4240
∞	0	0	-10^4

$\implies$ "sehr steifes System"

Definition 3.92 ((Absolute Stabilität))
Sei $y' = qy$ Testgleichung. *Sei* $\Phi = (\rho, \sigma)$ *lineares k-Schrittverfahren, d.h.*

$$(*) \quad \sum_{i=0}^{k} (a_i - hqb_i) u_{j+i} = 0 \; ; \quad j = k, \ldots, n-k$$

und $u_0, \ldots, u_{k-1}$ *gegeben.*

Definiere das Stabilitätspolynom

$$\rho_t(\lambda) = \rho(\lambda) - t \cdot \sigma(\lambda)$$

D.h. $\rho_t(\lambda)$ *ist charakteristisches Polynom von* $(*)$ *mit* $t = hq$.

Φ *heißt* absolut stabil *für* $t \in \mathbb{C}$ *falls gilt:*

$$t \in D_s := \{t \in \mathbb{C} \mid \rho_t(\lambda) = 0 \; für \; ein \; \lambda \in \mathbb{C} \implies |\lambda| < 1\}.$$

D_s *heißt Stabilitätsgebiet von* Φ.
Für ESV $\Phi = \varphi$ *der Form* $u_{j+1} = g(t)u_j$ *ist das Stabilitätspolynom gegeben durch*

$$\rho_t(\lambda) = \lambda - g(t).$$

Beispiel 3.93 (Stabilitätsgebiete für explizite Runge-Kutta-Verfahren)
Siehe Folie aus Deuflhard, Bornemann [4, Seite 230].

Beispiel 3.94
Das Verfahren von Milne-Simpson ist asymptotisch stabil, aber nicht absolut stabil. Dies wurde in Beispiel 3.73 gezeigt.

Weitere Forderung an " gute" numerische Verfahren:

A) Ein fallender Exponentialterm der Lösung von $y' = qy$ soll stets (für alle h) fallend genähert werden

Dies führt auf den Begriff der A-Satabilität

Definition 3.95 ((A-Stabilität))
Φ *heißt A-stabil, gdw.* $D_s \supset H_- = \{t \in \mathbb{C} \mid \; Re \; t < 0\}$.

Abschwächungen dazu sind:
$A(\alpha)$*-Stabilität:* $\iff \; D_s \supset \{t \in \mathbb{C} \mid \; |arg(-t)| < a\}$
$A(0)$*-Stabilität:* $\iff \; D_s \supset \mathbb{R}^-$

Satz 3.96
Sei $\Phi = (\rho, \sigma)$ lineares MSV mit $k \geq 2$ und Φ sei A-stabil. Dann gilt

1) Φ ist implizit,

2) Φ hat Konvergenzordnung 2.

Das Trapezverfahren hat unter allen A-stabilen MSV die kleinste Fehlerschranke.

(ohne Beweis)

Satz 3.97

1) Die BDF-k Verfahren sind A-stabil für $1 \leq k = s \leq 2$.

2) Implizite Runge-Kutta Verfahren sind A-stabil.

(ohne Beweis)

3.6 Numerische Lösung von Randwertproblemen

Beispiel 3.98 (Wärmeleitender Stab)

Gegeben: Stab $\triangleq [a,b]$
Wärmequellendichte $f(x),\ x \in [a,b]$
Randwerte: $y(a) = y_l, y(b) = y_r$
Wärmleitfähigkeitskoefffizient $k(x),\ x \in [a,b]$

Gesucht: Temperatur $y(x)$ mit

$$(*) \qquad \left| \begin{array}{l} -(k(x)y'(x))' = f(x) \ \ \forall x \in (a,b), \\ y \in C^2((a,b)) \cap C^0([a,b]), \\ y(a) = u_l,\ y(b) = u_r. \end{array} \right.$$

Formulierung als System 1. Ordnung: Setze $y_1 = y(x),\ y_2(x) = y'(x)$

$$(*) \implies \left| \begin{array}{l} y_1'(x) = y_2(x), \\ y_2'(x) = -\frac{k'(x)}{k(x)} y_2(x) - \frac{f(x)}{k(x)} \end{array} \right.$$

mit $y_1(a) = u_l,\ y_1(b) = u_r$.

Definition 3.99 (Lineare Randwertprobleme)
Seien $B_a, B_b \in \mathbb{R}^{n\times n},\ A \in C^0(I, \mathbb{R}^{n\times n}),\ I := [a,b],\ f \in C^0(I,\mathbb{R}^n),\ g \in \mathbb{R}^n$.
Dann heißt $y : I \longrightarrow \mathbb{R}^n$ *Lösung des* linearen Randwertproblems (RWP), *falls gilt:*

1) $y \in C^1(I, \mathbb{R}^n)$,

2) $y'(x) = Ay + f$ *in* I,

3) $B_a y(a) + B_b y(b) = g$.

Beispiel 3.100
Die Differentialgleichung

$$y''(x) + y(x) = 0,\ \ x \in [0,\pi] \iff \left| \begin{array}{l} y_1'(x) - y_2(x) = 0 \\ y_2'(x) + y_1(x) = 0 \end{array} \right.$$

hat die allgemeine Lösung $y(x) = c_1 \sin(x) + c_2 \cos(x)$.
Für verschiedene Randbedingungen ergibt sich unterschiedliches Vehalten:

1) $y(0) = y(\pi);\ y'(0) = y'(\pi) \implies$ ergibt die eindeutige Lösung $y \equiv 0$.

2) $y(0) = y(\pi) = 0 \implies$ ergibt unedlich viele Lösungen $y(x) = c_1 \sin(x)$.

3) $y(0) = 0,\ y(\pi) = 1 \implies$ ergibt keine Lösung.

Ziel: Bedingung für eindeutige Lösbarkeit!

Definition 3.101 (Fundamentalsystem)
Für

$$(*) \qquad y' = Ay + f \quad in\ I$$

definiere das Fundametalsystem $\{y_1, \ldots, y_n\}$, $y_i : I \longrightarrow \mathbb{R}^n$, *durch*

1) $y_i \in C^1(I, \mathbb{R}^n)$,

2) $y_i' = Ay_i$ *in* I,

3) $y_i(a) = e_i$, *mit* e_i *i-ter Einheitsvektor im* $\mathbb{R}^n$.

Die Matrix

$$Y(x) := \begin{pmatrix} y_{11}(x) & \ldots & y_{n1}(x) \\ \vdots & \ddots & \vdots \\ y_{1n}(x) & \ldots & y_{nn}(x) \end{pmatrix}$$

heißt Fundamentalmatrix.
Ist $y_0 \in C^1(I, \mathbb{R}^n)$, $y_0(a) = 0$ *eine spezielle Lösung von* $(*)$, *so läßt sich jede Lösung von* $(*)$ *darstellen als*

$$y(x) = y_0(x) + Y(x) \cdot s$$

mit einem Vektors $\in \mathbb{R}^n$.

Satz 3.102 (Existenz und Eindeutigkeit linearer RWPe)
Die folgenden Aussagen sind äquivalent:

1) (RWP) besitzt eine eindeutige Lösung.

2) Das zugehörige homogene RWP mit $f = 0$, $g = 0$ besitzt nur die triviale Lösung.

3) Die Matrix $B_a + B_b Y(b) \in \mathbb{R}^{n \times n}$ ist regulär.

Beweis: Seien $y_0(x)$ und $Y(x)$ wie in Definition 3.98 definiert.
Dann gilt $y_0(a) = 0$ und $Y(a) = I$.
Also ist $y(x) = y_0(x) + Y(x)s$ genau dann Lösung von (RWP), wenn gilt

$$[B_a + B_b Y(b)]s = g - B_b y_0(b)$$

Also ist $s \in \mathbb{R}^n$ genau dann eindeutig bestimmt, wenn 2) oder 3) gelten. □

3.6.1 Sturm-Liouville Probleme

Definition 3.103 (Sturm-Liouville Probleme)
Seien $p \in C^1(I),\ q, f \in C^0(i),\ I = [a,b]$ *gegeben mit*

$$q(x) \geq 0\ \forall x \in I,\ p(x) \geq p_0 > 0\ \forall x \in I$$

Dann heißt $y \in C^2((a,b)) \cap C^0([a,b]) \cap C^1((a,b])$ *Lösung des* Sturm-Liouville Problems (SLP) *mit homogenen Dirichlet und Neumann Randwerten, falls gilt:*

$$(SLP) \left| \begin{array}{l} -(p(x)y'(x))' + q(x)y(x) = f(x)\ \ \forall x \in I \\ y(a) = 0;\ y'(b) = 0 \end{array} \right.$$

Bemerkung: Allgemein Randbedingungen sind: $\left| \begin{array}{l} \alpha_1 y'(a) + \alpha_0 y(a) = g_a \\ \beta_1 y'(b) + \beta_0 y(b) = g_b \end{array} \right.$

Definition 3.104 (Variationsproblem)
Für $v \in X := \{u \in C^1([a,b]) \mid u(a) = 0\}$ *definiere das sogenannte* Energiefunktional $I : X \to \mathbb{R}$ *durch*

$$I(v) = \frac{1}{2}\int_a^b p(x)(v'(x))^2 dx + \frac{1}{2}\int_a^b q(x)(v(x))^2 dx - \int_a^b f(x)v(x)dx.$$

$y \in X$ *heißt Lösung des mit Variationsproblems, falls gilt*

$$I(y) = \inf_{v \in X} I(v).$$

Lemma 3.105 (Eulergleichung und natürliche Randbedingung)
Ist $y \in X$, so dass

$$I(y) \leq I(v)\ \ \forall v \in X,$$

so gilt $\forall \varphi \in X$:

$$(*)\quad \int_a^b py'\varphi' + qy\varphi = \int_a^b f\varphi$$

$(*)$ heißt schwache Formulierung der Eulergleichung.

Ist zusätzlich $y \in C^2((a,b))$, so folgt weiter:

$$(**) \left| \begin{array}{l} -(py')' + qy = f \quad \text{in } I, \\ y(a) = 0;\ y'(b) = 0. \end{array} \right.$$

Beweis:
Teil 1: Sei $y \in X$ mit $I(y) \leq I(v)\ \ \forall v \in X$.
$\implies I(y) \leq I(y + \varepsilon\varphi)\ \ \forall \varepsilon \in \mathbb{R}\ ;\ \ \forall \varphi \in X.$
$\implies G(0) \leq G(\varepsilon)\ \ \forall \varepsilon \in \mathbb{R}$, wobei $G(\varepsilon) := I(y + \varepsilon\varphi)$.
Da G differezierbar in ε is, folgt $G'(0) = 0$.
Es ist

$$\begin{aligned} G(\varepsilon) &= \tfrac{1}{2}\int_a^b p(y'+\varepsilon\varphi')^2 + q(y+\varepsilon\varphi)^2 - 2f(y+\varepsilon\varphi) \\ &= I(y) + \varepsilon\int_a^b py'\varphi' + qy\varphi - f\varphi + \varepsilon^2(I(\varphi) + \int_a^b f\varphi) \\ \Longrightarrow\quad G'(0) &= \int_a^b py'\varphi' + qy\varphi - f\varphi \overset{!}{=} 0, \quad \forall\varphi \in X. \end{aligned}$$

Teil 2: Ist $y \in C^2((a,b))$, so folgt mit partieller Integration

$$\int_a^b py'\varphi' = -\int_a^b (py')'\varphi + \Big[py'\varphi\Big]_a^b.$$

Einsetzen in $(*)$ ergibt

$$(\dagger) \quad -\int_a^b (py')'\varphi + \int_a^b qy\varphi - f\varphi + [py'\varphi]_a^b = 0.$$

Also folgt für alle $\varphi \in C_0^1((a,b))$:

$$\int_a^b [-(py')' + qy - f]\varphi = 0.$$

$$\Longrightarrow \quad -(py')' + qy = f.$$

Durch Einsetzen in $(\dagger)$ erhält man für alle $\varphi \in X$:

$$0 = [py'\varphi]_a^b = p(b)y'(b)\varphi(b) - p(a)y'(a)\underbrace{\varphi(a)}_{=0} = p(b)y'(b)\varphi(b)$$

Wähle $\varphi(b) \neq 0 \implies$, so folgt $y'(b) = 0$, da $p(b) > 0$. $y'(b) = 0$ heißt natürliche Randbedingung.

Nächstes Ziel: Existenz einer schwachen Lösung, d.h. □

$$\exists y \in X \text{ mit } I(y) = \inf_{v\in X} I(v).$$

Dazu benötigen wir folgenden Hilfssatz.

Satz 3.106 (Poincaré Ungleichung)
Für alle $v \in X$ gilt mit einer Konstanten $c_p \leq \frac{1}{2}(b-a)^2$:

$$\int_a^b (v(x))^2 dx \leq c_p \int_a^b (v'(x))^2 dx$$

Beweis: Es ist $|v(x)| = |v(x) - \underbrace{v(a)}_{=0}| \leq \int_a^x |v'(s)|\, ds$.

$$\Longrightarrow \quad (v(x))^2 \leq \left(\int_a^x |v'(s)|\, ds\right)^2 \leq (x-a)\int_a^x |v'(s)|^2\, ds.$$

$$\Longrightarrow \quad \int_a^b (v(x))^2 dx \leq \int_a^b (x-a)\int_a^b |v'(s)|^2\, ds = \tfrac{1}{2}(b-a)^2 \int_a^b |v'(s)|^2\, ds.$$ □

Lemma und Definition 3.107
Definiere auf X die Norm

$$\|v\|_X := \left(\int_0^1 (v(x))^2 + (v'(x))^2 dx\right)^{\frac{1}{2}}.$$

Sei $(v_n)_{n\in\mathbb{N}}$, $v_n \in X$ eine Minimalfolge, d.h.

$$\lim_{n\longrightarrow\infty} I(v_n) = \inf_{v\in X} I(v).$$

Dann ist $(v_n)_{n\in\mathbb{N}}$ eine Cauchyfolge in X, d.h.

$$\|v_n - v_m\|_X \longrightarrow 0 \quad (n,m \longrightarrow \infty).$$

Beweis: Setze $d := \inf_{v\in X} I(v)$.

1. Schritt: Zeige $d > -\infty$: Es ist

$$\begin{aligned} I(v) &\overset{\text{Vor. in (SLP)}}{\geq} \frac{p_0}{2}\int_a^b (v'(x))^2dx - \left(\int_a^b (f(x))^2\right)^{\frac12}\left(\int_a^b (v(x))^2\right)^{\frac12} \\ &\geq \frac{p_0}{2}\int_a^b (v'(x))^2dx - \underbrace{\sqrt{c_p}\left(\int_a^b (f(x))^2\right)^{\frac12}}_{=:a}\underbrace{\left(\int_a^b (v'(x))^2\right)^{\frac12}}_{=:b}. \end{aligned}$$

Mit der Ungleichung $|ab| \leq \frac{\delta}{2}a^2 + \frac{1}{2\delta}b^2 \ \ \forall a,b\in\mathbb{R}$ folgt

$$I(v) \geq \left(\frac{p_0}{2} - \frac{\delta}{2}\right)\int_a^b (v'(x))^2dx - \frac{\sqrt{c_p}}{2\delta}\|f\|^2_{L^2(a,b)}.$$

Für $\delta = p_0$ folgt: $I(v) \geq -\frac{\sqrt{c_p}}{2p_0}\|f\|^2_{L^2(a,b)}$. Also folgt $d > -\infty$, da $f \in C^0([a,b])$.

2. Schritt: Sei $(v_n)_{n\in\mathbb{N}}$ Folge in X mit $\lim_{n\to\infty} I(v_n) = d > -\infty$.

Zeige: (v_n) ist Cauchyfolge in X.

Notation: Definiere die Bilinearform $B : X\times X \longrightarrow \mathbb{R}$ durch

$$B(v,w) := \int_a^b pv'w' + \int_a^b qvw.$$

Dann gilt $B(v,v) = \int_a^b p\,(v')^2 + \int_a^b qv^2 \geq p_0\int_a^b (v'(x))^2dx$.

Aufgrund der Poincaré Ungleichung gilt weiter:

$$\begin{aligned} B(v,v) &\geq p_0\left(\int_a^b (v'(x))^2dx + c_p\int_a^b (v'(x))^2dx\right)\frac{1}{1+c_p} \\ &\geq \frac{p_0}{1+c_p}\|v\|^2_X. \end{aligned}$$

Damit folgt:

$$\begin{aligned} \frac{p_0}{1+c_p}\|v_n - v_m\|_X &\leq B(v_n - v_m, v_n - v_m) \\ &\overset{\text{Parallelogrammidentität}}{=} 2B(v_n,v_n) + 2B(v_m,v_m) - 4B\left(\tfrac{v_n+v_m}{2}, \tfrac{v_n+v_m}{2}\right) \\ &\overset{I(v)=\frac12 B(v,v)-\int_a^b fv}{=} 4[I(v_n) + I(v_m) - 2I(\tfrac{v_n+v_m}{2})] \\ &\leq 4[I(v_n) + I(v_m) - 2d] \\ &\longrightarrow 4[d + d - 2d] = 0 \quad \text{für } n,m\to\infty. \end{aligned}$$

Also ist $(v_n)_{n\in\mathbb{N}}$ Cauchyfolge.

□

Problem: Der Raum $(X, \|\cdot\|_X)$ ist nicht vollständig!

Lösung: Vervollständige X bzgl. der Norm $\|\cdot\|_X$ und erhalte vollständigen Raum $\bar{X}^{\|\cdot\|_X}$.

Beispiel aus der Analysis III: Ist $Y = C^0([a,b])$, so ist $\bar{Y} = L^2((a,b))$ die Vervollständigung von Y bzgl. $\|v\|_Y = \left(\int_a^b v^2\right)^{\frac{1}{2}}$.

Um $\bar{X}$ zu charakterisieren, benötigen wir den Begriff der schwachen Ableitung.

Definition 3.108 (Schwache Ableitung)
$v \in L^1(a,b)$ besitzt eine schwache Ableitung k-ter Ordnung $D^k v \in L^1(a,b)$, falls für alle $\varphi \in C_0^\infty(a,b)$ gilt:

$$\int_a^b v(x)\varphi^{(k)}(x)dx = (-1)^k \int_a^b (D^k v)(x)\varphi(x)dx.$$

Beispiel 3.109
Sei $v(x) = |x|$ mit $x \in [-1,1]$.
Dann gilt: $D^1(v)(x) = \text{sign}(x)$, denn für $\varphi \in C_0^\infty(-1,1)$ gilt:

$$\begin{aligned}
\int_{-1}^1 v(x)\varphi'(x)dx &= \int_{-1}^0 (-x)\varphi'(x)dx + \int_0^1 x\varphi(x)dx \\
&= [(-x)\varphi(x)]_{-1}^0 + \int_{-1}^0 \varphi(x)dx + [x\varphi(x)]_0^1 - \int_0^1 \varphi(x)dx \\
&= -\int_{-1}^0 (-1)\varphi(x)dx - \int_0^1 \varphi(x)dx \\
&= -\int_{-1}^1 \text{sign}(x)\varphi(x)dx.
\end{aligned}$$

$\rightsquigarrow$ Allgemein gilt in einer Raumdimension: Stückweise differenzierbare Funktionen, die global stetig sind, sind schwach differenzierbar. Die schwache Ableitung ist durch die stückweise definierte Ableitung gegeben.

Satz 3.110 (Eindimensionale Sobolevräume)

1) Der Sobolevraum

$$\begin{aligned}
H^m(a,b) := \{&v \in L^2(a,b) \mid v \text{ hat schwache Ableitungen} \\
&D^k v \in L^2(a,b)\ \forall k = 0,\dots,m\}
\end{aligned}$$

ist mit dem Skalarprodukt

$$(u,v)_{H^m(a,b)} := \sum_{k=0}^m \int_a^b D^k u D^k v$$

ein Hilbertraum. Durch $\|u\|_{H^m(a,b)} := \sqrt{(u,u)_{H^m(a,b)}}$ erhalten wir eine Norm auf $H^m(a,b)$.

2) $u \in H^m(a,b)$ ist fast überall gleich einer Funktion $\tilde{u} \in C^{m-1}([a,b])$ und es ist

$$\|\tilde{u}\|_{C^{m-1}([a,b])} \le c\,\|u\|_{H^m(a,b)}.$$

3) Zu $u \in H^m(a,b)$ gibt es eine Folge $(u_j)_{j\in\mathbb{N}}$, $u_j \in C^m([a,b])$, so dass

$$\|u - u_j\|_{H^m(a,b)} \longrightarrow 0 \qquad (j \to \infty).$$

Ist $\tilde{u}(a) = 0$, so kann man u_j so wählen, dass $u_j(a) = 0$ ist.

Beweis: z.B [Alt. Lineare Funktionalanalysis, Springer, 1992]. □

Folgerung 3.111
Die Vervollständigung von X bzgl. $\|\cdot\|_X$ ist gegeben durch

$$\bar{X} = \{v \in H^1(a,b) \mid v(a) = 0\}.$$

Satz 3.112 (Existenz und Eindeutigkeit einer schwachen Lösung von (SLP))
Seien die Voraussetzungen aus Definition 3.100 erfüllt. Dann gibt es genau ein $y \in \bar{X}$, so dass $\forall \varphi \in \bar{X}$ gilt:

$$\int_a^b py'\varphi' + \int_a^b qy\varphi = \int_a^b f\varphi.$$

Beweis:

Existenz: Sei $(v_n)_{n\in\mathbb{N}}$ Minimalfolge in $\bar{X}$ (anstelle von X).
Analog zu Lemma 3.104 folgern wir, dass $(v_n)_{n\in\mathbb{N}}$ Cauchyfolge in $\bar{X}$ ist.
Da $\bar{X}$ vollständig ist (Satz 3.107) $\exists y \in \bar{X}$ mit $\|v_n - y\|_{\bar{X}} \to 0 \ \ (n \to \infty)$.

Zeige: $I(y) = d \ (= \inf_{v\in\bar{X}} I(v))$.

Es ist

$$\begin{aligned}
&|I(\ y\ ) - I(v_n)| = \\
&\quad = \left|\tfrac{1}{2}\int_a^b p((y')^2 - (v_n')^2) + q(y^2 - v_n^2) - \int_a^b f(y - v_n)\right| \\
&\quad \le \tfrac{1}{2}\|p\|_\infty \int_a^b |y'^2 - v_n'^2| + \tfrac{1}{2}\|q\|_\infty \int_a^b |y^2 - v_n^2| + \int_a^b |f|\,|y - v_n| \\
&\quad \le \tfrac{1}{2}\|p\|_\infty \int_a^b |y' - v_n'|\,(|y'| + |v_n'|) + \tfrac{1}{2}\|q\|_\infty \int_a^b |y - v_n|\,(|y| + |v_n|) \\
&\qquad + \int_a^b |f|\,|y - v_n| \\
&\overset{\text{C.S.}}{\le} \underbrace{\left[\left(\frac{1}{2}\|p\|_\infty + \frac{1}{2}\|q\|_\infty\right)(\|y\|_{\bar{X}} + \|v_n\|_{\bar{X}}) + \|f\|_{L^2(a,b)}\right]}_{\le C < \infty} \underbrace{\|y - v_n\|_{\bar{X}}}_{\to 0 \ (n\to\infty)}
\end{aligned}$$

Also folgt $I(y) = \lim_{n\to\infty} I(v_n) = d$ und damit die Existenz der Lösung.

Eindeutigkeit: Seien $y_1,\ y_2$ Lösungen, so folgt für $v = y_1 - y_2$:

$$\int_a^b pv'\varphi' + \int_a^b qv\varphi \ \ \forall \varphi \in \bar{X}.$$

Wähle $\varphi = v$:

$$\Longrightarrow^{3.106} \ 0 = B(v,v) \ge \frac{p_0}{1 + c_p}\|v\|_X \ \Longrightarrow \ \|v\|_X = 0$$
$$\Longrightarrow \ v = 0 \ \Longrightarrow \ y_1 = y_2.$$

□

Satz 3.113 (A priori Abschätzung)
Für jede schwache Lösung $y \in \bar{X}$ von (SLP) gilt:

1) $\|y'\|_{L^2(a,b)} \leq C_1 \|f\|_{L^2(a,b)}$.

2) Ist $y \in H^2(a,b)$, so gilt
$\|y''\|_{L^2(a,b)} \leq C_2 \|f\|_{L^2(a,b)}$.

Dabei sind $C_1 = \frac{\sqrt{c_p}}{p_0}$, $C_2 = \frac{\sqrt{3}}{p_0}\left(\frac{\sqrt{c_p}}{p_0}\|p'\|_\infty + \frac{c_p}{p_0}\|q\|_\infty + 1\right)$

Beweis: Ü.A. □

3.6.2 Das Ritz-Galerkin Vefahren

Definition 3.114 (Ritz-Galerkin Verfahren für (SLP))
Sei $X := \{v \in H^1(a,b) \mid v(a) = 0\}$ *und* $I : X \longrightarrow \mathbb{R}$ *das Energiefunktional aus Definition 3.101. Sei* $X_h \in X$ *ein endlichdimensionaler Teilraum. Dann heißt* $u_h \in X_h$ Ritz-Galerkin Approximation *der schwachen Lösung* $y \in X$ *von (SLP), g.d.w*

$$I(u_h) = \inf_{v_h \in X_h} I(v_h).$$

Idee: Minimierung von I *auf endlichen Teilräumen.*

Folgerung 3.115 (Schwache diskrete Differentialgleichung)

1) Da $X_h \in X$ ist, folgt $I(u) \leq I(u_h)$.

2) Ist $u_h \in X_h$ Ritz-Galerkin Approximation von (SLP), so gilt:

$$(D - SLP) \qquad B(u_h, \varphi_h) = \int_a^b f\varphi_h, \qquad \forall \varphi_h \in X_h$$

mit $B(u,v) := \int_a^b pu'v' + \int_a^b quv$.

Beweis: Analog zum kontinuierlichen Fall in Lemma 3.102. □

Bemerkung 3.116 (Formulierung von (D-SLP) als lineares Gleichungssystem)
Seien $N := \dim(X_h)$ und $\{\varphi_1, \ldots, \varphi_N\}$ eine Basis von X_h. Dann läßt sich u_h darstellen als

$$u_h(x) = \sum_{j=1}^{N} u_j \varphi_j(x)$$

mit Koeffizienten $u_j \in \mathbb{R}$, $j = 1, \ldots, N$.
Damit ist (D-SLP) äquivalent zu

$$\sum_{j=1}^{N} B(\varphi_j, \varphi_k) u_j = \int_a^b f\varphi_k \;\; \forall k = 1, \ldots, N.$$

Mit den Definitionen $u := (u_1, \ldots, u_N)$; $S_{kj} := B(\varphi_j, \varphi_k)$, $k, j = 1, \ldots, N$; $b_k := \int_a^b f\varphi_k$; $b := (b_1, \ldots, b_N)$ ist somit (D-SLP) äquivalent zu dem linearen Gleichungssystem

$$\mathbf{Su} = \mathbf{b}$$

Die Matrix S wird "Steifigkeitsmatrix" genannt.

Satz 3.117 (Abstrakte Fehlerabschätzung)
Seien X ein normierter Raum, $X_h \subset X$ ein Teilraum. Weiter sei $f \in X' = L(X; \mathbb{R})$ ein lineares Funktional auf X.
Gilt dann

$$u \in X : B(u, \varphi) = f(\varphi), \qquad \forall \varphi \in X,$$
$$u_h \in X_h : B(u_h, \varphi_h) = f(\varphi_h) \qquad \forall \varphi_h \in X_h$$

mit einer Bilinearform $B : X \times X \longrightarrow \mathbb{R}$ die koerziv ist, d.h.

$$\exists C_0 > 0, \text{ so dass } \forall \varphi \in X : B(\varphi, \varphi) \geq C_0 \|\varphi\|_X^2$$

und stetig, d.h.

$$\exists C_1 \geq 0, \text{ so dass } \forall \varphi, \psi \in X : |B(\varphi, \psi)| \leq C_1 \|\varphi\|_X \cdot \|\psi\|_X ,$$

so gilt die Fehlerabschätzung

$$\|u - u_h\|_X \leq \frac{C_1}{C_0} \inf_{v_h \in X_h} \|u - v_h\|_X$$

und es ist

$$B(u - u_h, \varphi_h) = 0 \qquad \forall \varphi_h \in X_h. \quad \text{"Galerkin-Orthogonalität"}$$

Beweis: Wegen $X_h \subset X$, folgt $\forall \varphi_h \in X_h$
$B(u, \varphi_h) = f(\varphi_h)$ und $B(u_h, \varphi_h) = f(\varphi_h)$
$\Longrightarrow \quad B(u - u_h, \varphi_k) = 0 \ \forall \varphi_h \in X_h$ (Galerkin-Orthogonalität $\surd$).

Mit Koerzivität und Stetigkeit von B folgt nun:

$$\begin{aligned} C_0 \|u - u_h\|_X^2 &\leq B(u - u_h, u - u_h) \\ &= B(u - u_h, u) - \underbrace{B(u - u_h, u_h)}_{=0} \\ &\overset{v_h \in X_h}{=} B(u - u_h, u) - \underbrace{B(u - u_h, v_h)}_{=0} \\ &= B(u - u_h, u - v_h) \\ &\overset{\text{stetig}}{\leq} C_1 \|u - u_h\|_X \|u - v_h\|_X \end{aligned}$$

Also folgt: $\|u - u_h\|_X \leq \frac{C_1}{C_0} \|u - v_h\|_X \ \ \forall v_h \in X_h$. □

Bemerkung 3.118
Für (SLP) ist

$$B(u, v) = \int_a^b pu'v' + \int_a^b quv$$

und

$$f(v) := \int_a^b fv.$$

Ist $X := \{v \in H^1(a,b) \,|\, v(a) = 0\}$, so ist $f \in X'$, falls $f \in L^2(a,b)$, da $\forall v \in X$ gilt

$$|f(v)| \leq \|f\|_{L^2(a,b)} \|v\|_{L^2(a,b)} \leq \|f\|_{L^2(a,b)} \|v\|_X .$$

Beispiel 3.119 (Wahl von X_h)

1) Polynomapproximation: $X_h = \mathbb{P}_k$ ($\hat{=}$ Polynome von Grad $\leq k$)
   $\Longrightarrow$ $N = \dim(X_h) = k + 1$.

2) Eigenräume: $X_h := \text{span}\{\varphi_1, \ldots, \varphi_N\}$, wobei $\varphi_1, \ldots, \varphi_N$ die ersten N normierten Eigenfunktionen des Operators
   $$Lu := -(pu')' + qu\ , \quad u(a) = 0,\ u'(b) = 0$$
   sind, d.h.
   $$L\varphi_j = \lambda_j \varphi_j\ ;\ \ 0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_N.$$

3) Stückweise Polynome: $X_h = \{\varphi_h \in C^0([a,b]) \,|\, \varphi_h(a) = 0, \varphi_h \restriction_{[x_j, x_{j+1}]} \in \mathbb{P}_k\}$ wobei $a = x_0 < x_1 < \ldots < x_n = b$ eine Zerlegung von $[a,b]$ ist.
   $\Longrightarrow$ Finite Elemente!

3.6.3 Finite Elemente Verfahren

Definition 3.120 (Finite Elemente Verfahren)

1) Sei $I := [a,b]$ *und* $I_h := \{x_0, \dots, x_n\} \subset I$ *ein Gitter mit* $x_0 = a$, $x_n = b$, $x_{j+1} = x_j + h_j$, $h_j > 0$. *Definiere* $I_j := (x_j, x_{j+1})$ *und* $h := \max_{j=0,\dots,n-1} h_j$. *Für festes* $k \in \mathbb{N}$ *wählen wir*

$$X_h := \{\varphi_h \in C^0([a,b]) \,|\, \varphi_h(a) = 0, \varphi_h \restriction_{I_j} \in \mathbb{P}_k \; \forall j = 0, \dots, n-1\}.$$

Ein $u_h \in X$ *heißt Finite Elemente Approximation von (SLP), falls für alle* $\varphi_h \in X_h$ *gilt:*

$$B(u_h, \varphi_h) = f(\varphi_h) \quad \forall \varphi_h \in X_h.$$

Dabei sind B, f *wie in Bemerkung 3.115 definiert.*

2) (Finite Elemente Basis für $k = 1$*)*
Sei $k = 1$. *Für* $j = 1, \dots, n$ *sei* $\varphi_j \in X_h$ *definiert durch*

$$\varphi_j(x_i) = \delta_{ij} \quad \forall x_i \in I_h.$$

D.h. für $j = 1, \dots, n$ *ist*

$$\varphi_j(x) = \begin{cases} \frac{x - x_{j-1}}{h_{j-1}} & x \in I_{j-1} \\ \frac{x_{j+1} - x}{h_j} & x \in I_j \\ 0 & sonst \end{cases}.$$

Es ist $supp(\varphi_j) = I_j \cup I_{j-1}$ *und somit folgt für die Steifigkeitsmatrix* $S_{kj} := B(\varphi_j, \varphi_k) = 0$, *falls* $|j - k| \geq 2$.

3) (Fehlerabschätzung für $k = 1$*)*
Sei $k = 1$. *Definiere Interpolierende* $\tilde{u}_h \in X_h$ *zu* $y \in X$ *durch*

$$\tilde{u}_h(x_j) := y(x_j) \quad \forall j = 0, \dots, n \implies \tilde{u}_h(x) = \sum_{j=1}^{n} y(x_j)\varphi_j(x).$$

Dann gilt die Fehlerabschätzung (nach Satz 3.114)

$$\|y - u_h\|_X \leq \frac{C_1}{C_0} \inf_{v_h \in X_h} \|y - v_h\|_X \leq \frac{C_1}{C_0} \|y - \tilde{u}_h\|_X .$$

D.h. der Fehler der Finite Elemente Approximation ist durch den Interpolationsfehler abgeschätzt.

Satz 3.121 (Interpolationsfehler auf dem Einheitselement)
Für $v \in H^1(0,1)$ sei $\tilde{v}_h \in \mathbb{P}_1$ die lineare Interpolierende zu v, d.h. $\tilde{v}_h \in \mathbb{P}_1$, $\tilde{v}_h(0) = v(0)$, $\tilde{v}_h(1) = v(1)$. Dann gilt:

1) $\|v - \tilde{v}_h\|_{L^2(0,1)} \leq \sqrt{c_p} \|v'\|_{L^2(0,1)}$,

2) $\|\tilde{v}_h'\|_{L^2(0,1)} \leq \|v'\|_{L^2(0,1)}$.

Ist zusätzlich $v \in H^2(0,1)$, so ist

3) $\|v - \tilde{v}_h\|_{L^2(0,1)} \leq c_p \|v''\|_{L^2(0,1)}$,

4) $\|(v' - \tilde{v}_h')\|_{L^2(0,1)} \leq \sqrt{c_p} \|v''\|_{L^2(0,1)}$.

Beweis: Analog zu Hilfssatz 3.103 (Poincaré Ungleichung) zeigt man , dass für alle $w \in H^1(0,1)$ mit $w(0) = w(1) = 0$ gilt

$$(*) \quad \|w\|^2_{L^2(0,1)} \leq c_p \left\|w'\right\|^2_{L^2(0,1)},$$

wobei $c_p \leq \frac{1}{2}$ ist. Setze $w = v - \tilde{v}_h$. Dann folgt aus $(*)$ unter Verwendung von

$$(**) \quad \tilde{v}_h'(x) = v(1) - v(0) = \int_0^1 v'(s)ds :$$

$$\begin{array}{rcl} \|v - \tilde{v}_h\|^2_{L^2(0,1)} & \leq & c_p \|v - \tilde{v}_h)'\|^2_{L^2(0,1)} = c_p \int_0^1 (v'(x) - \tilde{v}_h'(x))^2 dx \\ & = & c_p \int_0^1 (v'(x))^2 - 2\tilde{v}_h'(x)v'(x) + (\tilde{v}_h'(x))^2 dx \\ & \overset{(**)}{=} & c_p \int_0^1 (v'(x))^2 dx - 2(v(1) - v(0))^2 + (v(1) - v(0))^2 \\ & = & c_p \int_0^1 (v'(x))^2 dx - (v(1) - v(0))^2 \\ & \leq & c_p \int_0^1 (v'(x))^2 dx \end{array}$$

Also haben wir 1) gezeigt.
2) folgt direkt aus $(**)$, da

$$\left\|\tilde{v}_h'\right\|_{L^2(0,1)} = |v(1) - v(0)| \leq \left\|v'\right\|_{L^2(0,1)}$$

3) folgt aus 1) und $(*)$, da $(*)$ auch für $w \in H^1(0,1)$ mit $\int_0^1 w(x)dx = 0$ gilt.

4) folgt analog zur Herleitung von Gleichung 2), wenn man beachtet, dass $\tilde{v}_h'' = 0$ gilt. □

Folgerung 3.122 (Interpolationsabschätzung)
Sei $y \in X$ und $\tilde{u}_h \in X_h$ die Interpolierende von y für $k = 1$ aus Definition 3.117, 3). Dann gilt, falls $y \in H^2(a,b)$:

1) $\|y - \tilde{u}_h\|_{L^2(a,b)} \leq c_p h^2 \|y''\|_{L^2(a,b)}$,

2) $\|(y - \tilde{u}_h)'\|_{L^2(a,b)} \leq \sqrt{c_p} h \|y''\|_{L^2(a,b)}$.

Beweis: (Folgt aus 3.118 mit Skalierungsargument)

zu 1): Es ist $\|y - \tilde{u}_h\|^2_{L^2(a,b)} = \sum_{j=0}^{n-1} \|y - \tilde{u}_h\|^2_{L^2(I_j)}$.

Durch die Transformation $x = F(\bar{x}) = h_j\bar{x} + x_j$ und $\bar{y}(\bar{x}) = y(F(\bar{x}))$ folgt

$$I_j = F((0,1))$$

und

$$\begin{array}{rcl} \tilde{u}_h|_{I_j} & = & y(x_j) + \frac{x - x_j}{h_j}(y(x_{j+1}) - y(x_j)) \\ & = & \bar{y}(0) + \bar{x}(\bar{y}(1) - \bar{y}(0)) =: \tilde{\bar{y}}_h(\bar{x}) \end{array}$$

Also gilt wegen $F'(\bar{x}) = h_j$ und $\bar{y}''(\bar{x}) = y''(F(\bar{x})) \cdot (F'(\bar{x}))^2$

$$\begin{aligned} \|y - \tilde{u}_h\|^2_{L^2(I_j)} &= h_j \|\bar{y} - \tilde{\bar{y}}_h\|^2_{L^2(0,1)} \\ &\overset{3.118\ 3)}{\leq} c_p^2 h_j \|\bar{y}''\|^2_{L^2(0,1)} \\ &\leq c_p^2 h_j^4 \|y''\|^2_{L^2(I_j)} . \end{aligned}$$

$$\begin{aligned} \Longrightarrow \quad \|y - \tilde{u}_h\|^2_{L^2(a,b)} &\leq c_p^2 \sum_{j=0}^{n-1} h_j^4 \|\bar{y}''\|^2_{L^2(I_j)} \\ &\leq c_p^2 h^4 \|y''\|^2_{L^2(a,b)} . \end{aligned}$$

Also ist 1) gezeigt.

zu 2): (Analoge Vorgehensweise!)

$$\|(y - \tilde{u}_h)'\|^2_{L^2(a,b)} = \sum_{j=0}^{n-1} \|(y - \tilde{u}_h)'\|^2_{L^2(I_j)} .$$

$$\begin{aligned} \|(y - \tilde{u}_h)'\|^2_{L^2(I_j)} &= \tfrac{1}{h_j} \|(\bar{y} - \tilde{u}_h)\|^2_{L^2(0,1)} \\ &\overset{3.118\ 4)}{\leq} c_p \tfrac{1}{h_j} \|\bar{y}''\|^2_{L^2(0,1)} \\ &= c_p h_j^2 \|y''\|^2_{L^2(I_j)} . \end{aligned}$$

$$\begin{aligned} \Longrightarrow \quad \|(y - \tilde{u}_h)'\|^2_{L^2(a,b)} &\leq c_p \sum_{j=0}^{n-1} h_j^2 \|y''\|^2_{L^2(I_j)} \\ &\leq c_p h^2 \|y''\|^2_{L^2(a,b)} . \end{aligned}$$

□

Satz 3.123 (Fehlerabschätzung für lineare Finite Elemente)
Sei $y \in X$ die schwache Lösung von (SLP) und es gelten die Voraussetzungen aus Definition 3.100. Sei zusätzlich $f \in L^2(a,b)$, $p, p', q \in L^\infty(a,b)$ und es gelte $y \in H^2(a,b)$.
Dann existiert eine Konstante $c > 0$, so dass für die stückweise lineare Finite Elemente Approximation $u_h \in X_h$ (mit $k = 1$) gilt:

$$\|y - u_h\|_X \leq c\, h \|f\|_{L^2(a,b)}$$

Beweis: Nach 3.117, 3) gilt $\|y - u_h\|_X \leq \frac{c_1}{c_0} \|y - \tilde{u}_h\|_X$, wobei $\tilde{u}_h$ die Interpolierende von y in X_h ist und

$$c_1 := \max\{\|p\|_\infty, \|q\|_\infty\},\ c_0 := \frac{p_0}{1 + c_p}.$$

Dann folgt weiter mit Interpolationsfehlerabschätzung 3.119:

$$\begin{aligned} \|y - u_h\|_X &\leq \tfrac{c_1}{c_0} \left(\|y - \tilde{u}_h\|^2_{L^2(a,b)} + \|(y - \tilde{u}_h)'\|^2_{L^2(a,b)} \right)^{\frac{1}{2}} \\ &\overset{3.119}{\leq} \tfrac{c_1}{c_0} (c_p^2 h^4 + c_p h^2)^{1/2} \|y''\|_{L^2(a,b)} \\ &\overset{3.110}{\leq} \underbrace{\frac{c_1 c_2}{c_0} \sqrt{c_p} (c_p h^2 + 1)^{1/2}}_{\leq C \text{ (z.B. für } h \leq h_{\max})} h \|f\|_{L^2(a,b)} . \end{aligned}$$

□

Bemerkung 3.124

1) Man kann zeigen, dass mit den Voraussetzungen $p \in C^1(a,b),\ q, f \in C^0(a,b);\ p, p', q \in L^\infty(a,b)$ und $f \in L^2(a,b)$ folgt, dass $y \in H^2(a,b)$ ist.

2) Für $k > 1$ glit die Fehlerabschätzung:

$$\|y - u_h\|_X \leq c \cdot h^k \left\| y^{(k+1)} \right\|_{L^2(a,b)}$$

falls p, q, f und y regulär genug sind.

3) Weiteres zu Finiten Elementen findet man z.B. in den Büchern von Ciarlet [3] oder Braess [2].

Bemerkung 3.125 (Weitere Diskretisierungsverfahren für (SLP))

1) Finite Differenzenverfahren:
Idee: Ersetze Ableitungen durch Differenzenquotienten, z.B.

$$y' \approx \frac{y_j - y_{j-1}}{h},\ \ y'' \approx \frac{y_{j+1} - 2y_j + y_{j-1}}{h^2}.$$

2) Schießverfahren:
Berechne mit ESV oder MSV Approximationen der Anfangswertprobleme für $y_0,\ Y$ aus Definition 3.98 und erhalte $y_{0,h},\ Y_h$.
Wie im kontinuierlichen Fall ist dann

$$y_h := y_{0,h} + Y_h s_h$$

wobei s_h Lösung des Gleichnungssystems

$$[B_a + B_b Y_h(b)] s_h = g - B_b y_{0,h}(b)$$

ist (vgl. Satz 3.99 mit Beweis).

Kapitel 4

Ausblick: Partielle Differentialgleichungen

Partielle Differentialgleichungen sind Gleichungen, die eine Funktion mit mehreren Variablen mit ihren partiellen Ableitungen in Beziehung setzt.

Beispiel: (Wärmeleitungsgleichung in drei Raumdimensionen)
Gesucht ist die zeitabhängige Temperaturverteilung $T = T(x,t)$ in einem Gebiet $\Omega \subset \mathbb{R}^3$ und Zeitinterval $[0, T_{\max}]$.

Im folgenden sei $\Omega \subset \mathbb{R}^n$ ein Gebiet. Wir beschränken und auf skalare partielle Differentialgleichungen zweiter Ordnung vom linearen Typ für $u \in C^2(\Omega)$.

Definition 4.1 (Partielle Differentialgleichung zweiter Ordnung)
$u \in C^2(\Omega)$ heißt klassiche Lösung einer linearen partielle Differentialgleichung zweiter Ordnung, falls gilt:

$$\sum_{i,j=1}^{n} a_{ij}(x) u_{x_i,x_j}(x) + \sum_{i=1}^{n} b_i(x) u_{x_i}(x) + c(x)u(x) = f(x), \quad \forall x \in \Omega. \tag{4.1}$$

Dabei sind $a_{ij}, b_i, c, f \in C^0(\Omega)$ für $i,j = 1,\dots,n$, $x = (x_1,\dots,x_n) \in \Omega$ und $u_{x_i} := \frac{\partial u}{\partial x_i}(x)$. Da $u \in C^2(\Omega)$ist, gilt $u_{x_i,x_j}(x) = u_{x_j,x_i}(x)$ und wir können ohne Beschränkung der Allgemeinheit annehmen, dass $a_{ij}(x) = a_{ji}(x)$ gilt. Dann ist

$$A(x) = \Big(a_{ij}(x)\Big)_{i,j=1,\dots,n}$$

symmetrisch und hat n reelle Eigenwerte.
Wir setzen $\mathbf{b}(x) := (b_1(x),\dots,b_n(x))$.

Definition 4.2 (Klassifizierung für Partielle Differentialgleichungen zweiter Ordnung)

(i) Die Partielle Differentialgleichung (4.1) heißt hyperbolisch in $x \in \Omega$, falls $n-1$ Eigenwerte von $A(x)$ gleiches Vorzeichen besitzen und ein Eigenwert das entgegengesetzte Vorzeichen hat.

(ii) Die Partielle Differentialgleichung (4.1) heißt elliptisch in $x \in \Omega$, falls alle Eigenwerte von $A(x)$ das gleiche Vorzeichen besitzen.

(iii) Die Partielle Differentialgleichung (4.1) heißt parabolisch in $x \in \Omega$, falls $n-1$ Eigenwerte von $A(x)$ gleiches Vorzeichen haben, ein Eigenwert verschwindet und zusätzlich gilt:

$$Rg(A(x), \boldsymbol{b}(x)) = n.$$

Bemerkung 4.3
Definition 4.2 deckt nicht alle denkbaren Typen ab. Allerdings sind die definierten Typen die wichtigsten in den Anwendungen.

4.1 Die Wellengleichung

Definition 4.4 (Wellengleichung)
Seien $u_0, u_1, a \in \mathbb{R}$ und $g, h : [0,\pi] \to \mathbb{R}$ mit

$$\begin{aligned} g(0) &= u_0, \quad g(\pi) = u_1, \\ h(0) &= 0, \quad h(\pi) = 0. \end{aligned} \tag{4.2}$$

$u : [0,\pi] \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ heißt Lösung der Wellengeichung, falls gilt:

$$u_{tt}(x,t) = a^2 u_{xx}(x,t), \qquad \forall (x,t) \in (0,\pi) \times (0,\infty) \tag{4.3}$$

und

$$u(0,t) = u_0, u(\pi,t) = u_1, \quad \forall t \geq 0, \tag{4.4}$$

$$u(x,0) = g(x), \qquad \forall x \in [0,\pi], \tag{4.5}$$

$$u_t(x,0) = h(x), \qquad \forall x \in [0,\pi]. \tag{4.6}$$

Die Wellengleichung beschreibt für $u_0, u_1 = 0$ die Auslenkung einer in $x = 0$ und $x = \pi$ fest eingespannten Saite:

Bemerkung 4.5

- Die Wellengleichung ist hyperbolisch. Es ist

$$A(x,t) = \begin{pmatrix} -a^2 & 0 \\ 0 & 1 \end{pmatrix}.$$

- Die Bedingungen (4.2) heißen Verträglichkeitsbedingungen.

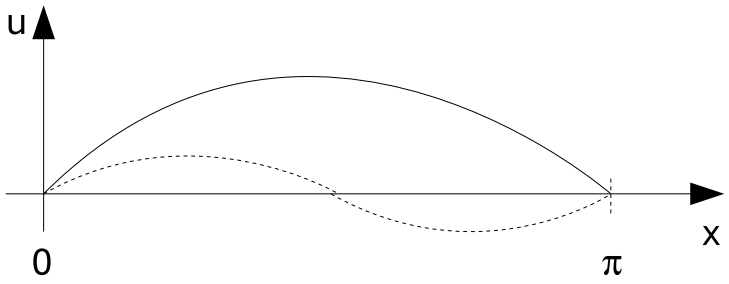


Abbildung 4.1: Erste und zweite Mode einer schwingenden Saite.

4.2 Die Poisson Gleichung

Sei $\Omega \subset R^2$ ein Gebiet mit glattem Rand $\partial\Omega$ und $\bar{\Omega} = \Omega \cup \partial\Omega$.

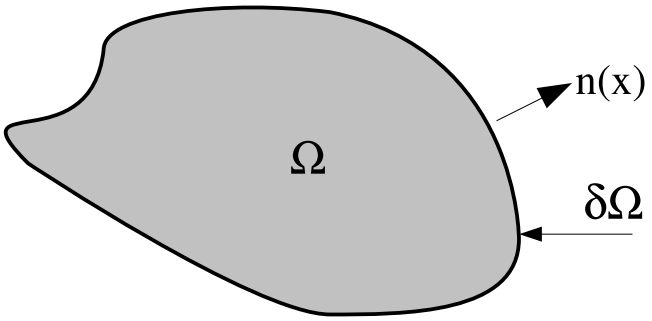


Abbildung 4.2: Skizze eines Gebietes Ω mit glattem Rand.

Definition 4.6 (Poisson Gleichung)
Seien $f : \Omega \to \mathbb{R}$ und $g : \partial\Omega \to \mathbb{R}$ gegebene Funktionen. Dann heißt eine Funktion $u \in C^2(\Omega) \cap C(\bar{\Omega})$ Lösung des Dirichletproblems für die Poisson Gleichung, falls gilt

$$\Delta u(x) := \sum_{i=1,2} u_{x_i,x_i} = f(x), \quad \forall x \in \Omega, \tag{4.7}$$

und

$$u(x) = g(x), \qquad \forall x \in \partial\Omega. \tag{4.8}$$

Falls $f \equiv 0$ gilt, so heißt Gleichung (4.7) auch Laplace Gleichung.

Bemerkung 4.7

- Die Poisson Gleichung ist elliptisch. Es ist

$$A(x) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

- Neben dem Dirichlet-Problem betrachtet man auch das Neumann-Problem, d.h. (4.8) wird ersetzt durch
$$\nabla u(x) \cdot \mathbf{n}(x) = g(x), \quad \forall x \in \partial\Omega.$$
Dabei ist $\mathbf{n}(x)$ die äußere Normale an den Rand von Ω.

- Die Poisson Gleichung modelliert z. B. die stationäre Wärmeverteilung, oder die Verteilung eines elektrischen Potentials.

4.3 Die Wärmeleitungsgleichung

Sei $\Omega \subset R^2$ ein Gebiet mit glattem Rand $\partial\Omega$.

Definition 4.8 (Wärmeleitungsgleichung)
Seien $u_0 : \Omega \to \mathbb{R}$ und $g : \partial\Omega \times [0,\infty) \to \mathbb{R}$ mit folgender Verträglichkeitsbedingung gegeben:

$$u_0(x) = g(x,0), \qquad \forall x \in \partial\Omega. \tag{4.9}$$

Weiter seien $a > 0$ und $f : \Omega \times [0,\infty) \to \mathbb{R}$ gegeben.
Dann heißt eine Funktion $u \in C^2(\Omega \times (0,\infty)) \cap C(\bar{\Omega} \times (0,\infty))$ Lösung der Wärmeleitungsgleichung, falls gilt

$$u_t(x) = a\Delta u(x) + f(x), \qquad \forall x \in \Omega \times (0,\infty) \tag{4.10}$$

und

$$u(x,0) = u_0(x), \qquad \forall x \in \Omega, \tag{4.11}$$
$$u(x,t) = g(x,t), \qquad \forall (x,t) \in \partial\Omega \times [0,\infty). \tag{4.12}$$

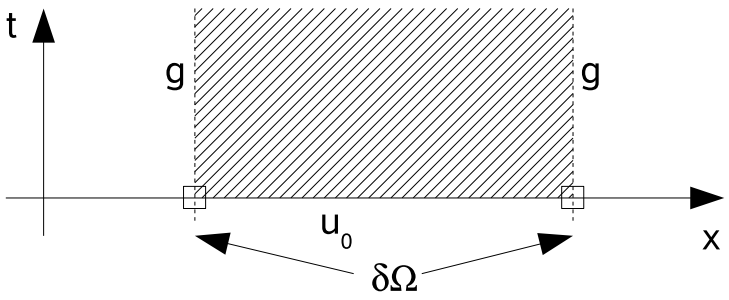


Abbildung 4.3: Skizze zu der Verträglichkeitsbedingung (4.9).

Bemerkung 4.9

- Die Wärmeleitungsgleichung ist parabolisch. Es ist
$$A(x) = \begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

und

$$\mathrm{Rg}(A, (0,0,1)^t) = 3.$$

Index

Literaturverzeichnis

[1] M. Bollhöfer und V. Mehrmann.*Numerische Mathematik*, Vieweg Verlag, Wiesbaden 2004.

[2] P. Braess. *Finite Elemente*, Springer, Berlin 1992.

[3] P.G. Ciarlet.*The finite element methods for elliptic problems*, North-Holland, Amsterdam 1987.

[4] P. Deuflhard und F. Bornemann. *Numerische Mathematik II*, 3. Auflage, Walter de Gruyter, Berlin 2002.

[5] R. D. Grigorieff. *Numerik gewöhnlicher Differentialgleichungen*, 2 Auflage, Teubner, Stuttgart 1977.

[6] W. Hackbusch. *Iterative Lösung großer schwachbesetzter Gleichungssysteme. Leitfäden der Angewandten Mathematik und Mechanik, 69*, Teubner Studienbücher Mathematik, Teubner, Stuttgart 1991.

[7] G. Hämmerlin und K.-H. Hoffmann.*Numerische Mathematik*, Springer, Berlin 1989.

[8] J. Stoer und R. Bulirsch. *Einführung in die Theorie der Numerischen Mathematik I & II*, Heidelberger Taschenbücher, Springer, Berlin 2000.

[9] W. Walter. *Gewöhnliche Differentialgleichungen*, 5. Auflage, Springer, Berlin, 1993.