universität wien

DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

„Numerische Integration – Verfahren zur effizienten Berechnung von Integralen“

verfasst von / submitted by

Alexander Thomaso

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magister der Naturwissenschaften (Mag. rer. nat.)

Wien, 2018 / Vienna, 2018

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 190 406 412
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Lehramtsstudium: UF Mathematik UF Physik
Betreut von / Supervisor:	ao. Univ.-Prof. Mag. Dr. Peter Raith

Vorwort und Danksagung

Sehr geehrte Leserinnen und Leser!

In meiner Diplomarbeit möchte ich Ihnen die numerische Integration näher bringen, indem ich eine Reihe verschiedener Verfahren vorstellen werde, deren Vor- und Nachteile erläutere und diverse Einsatzmöglichkeiten beschreibe.

Ich habe mich für dieses Thema entschieden, da ich mich allgemein sehr für numerische Mathematik interessiere und ich es spannend finde, Algorithmen zu entwickeln bzw. zu beschreiben, welche man an einen Computer übergibt, um innerhalb kurzer Zeit an Ergebnisse zu gelangen, die vor einigen Jahrhunderten mit einem wesentlichen Rechenaufwand verbunden waren (z. B. die Berechnung von Logarithmen oder die Ermittlung von Nullstellen).

Unter den verschiedenen Teilbereichen der numerischen Mathematik entschied ich mich für die numerische Integration, da es hier eine sehr umfangreiche Auswahl an Verfahren mit teilweise komplett unterschiedlichen Ansätzen und Zugangsweisen gibt, welche ich kennenlernen wollte und in gewisser Weise handhabbar machen möchte. Außerdem ist die numerische Integration – im Vergleich zu zahlreichen anderen numerischen Verfahren – für den Schulalltag durchaus geeignet, weshalb sie auch für meine berufliche Zukunft relevant sein wird.

Neben diesem kurzen Einblick in die persönlichen Beweggründe für die Wahl dieses Themas möchte ich an dieser Stelle meiner Diplomarbeit auch die Gelegenheit nutzen, um mich bei all jenen zu bedanken, die mich während des Studiums und beim Verfassen dieser Arbeit unterstützt haben.

Mein Dank geht insbesondere an Herrn Professor Raith, der mich durch meine gesamte Studienzeit im Unterrichtsfach Mathematik begleitet hat, mich bei der Auswahl des Themas beraten hat und mir schließlich bei der Ausarbeitung hilfreich zur Seite gestanden ist.

Außerdem möchte ich mich bei der Universität Wien für die Ermöglichung eines reibungslosen und verzögerungsfreien Studiums bedanken. Dasselbe gilt für meinen derzeitigen Arbeitgeber, das Schulzentrum Ungargasse, der mir bereits einen sehr frühen Eintritt in das Berufsleben ermöglichte und mir die zum Abschluss meines Studiums notwendigen Freistellungen bereitwillig gewährte.

Mein besonderer Dank gilt außerdem meiner Familie, die mir mein Studium ermöglicht hat und mir stets beratend zur Seite stand. Im Speziellen möchte ich mich für die pädagogischen Ratschläge meiner Mutter bedanken, welche mir bei meinen ersten Praxiserfahrungen im Klassenzimmer sehr hilfreich waren. Abschließend bedanke ich mich auch bei meinen Freunden für eine schöne Studienzeit.

Wien, im Jänner 2018

Inhaltsverzeichnis

1 Einleitung

Diese Diplomarbeit beschäftigt sich mit der numerischen Integration, also der Approximation von bestimmten Integralen. Eine derartige Näherung kann aus ganz unterschiedlichen Gründen notwendig bzw. sinnvoll sein. Bei vielen Funktionen, wie zum Beispiel $x \mapsto \frac{\sin x}{x}$ oder $x \mapsto \exp(-x^{-2})$, ist die Stammfunktion nicht geschlossen darstellbar. Es kann aber auch vorkommen, dass die Funktion nur an endlich vielen Stellen bekannt ist (z. B. bei der Durchführung eines physikalischen Experiments, bei welchem nur für einzelne Zeitpunkte Messwerte vorliegen). In vielen Fällen (insbesondere wenn kein Computeralgebrasystem zur Verfügung steht) ist es aber auch sonst wesentlich einfacher, ein numerisches Verfahren anzuwenden, anstatt die Stammfunktion zu ermitteln (selbst wenn diese geschlossen darstellbar ist).

Häufig wird der Integrand bei numerischen Integrationsverfahren durch eine Polynomfunktion approximiert, da deren Stammfunktionen sehr einfach zu bestimmen sind. Auch in Hinblick auf die Fehlerabschätzung weisen Polynome einige nützliche Eigenschaften auf. Aus diesem Grund widmet sich der erste Teil dieser Diplomarbeit den Polynomen. Es werden zunächst einige allgemeine Eigenschaften von stetigen Funktionen (und damit insbesondere von Polynomen) behandelt, bevor anschließend ausführlich auf die Polynominterpolation eingegangen wird. In den beiden darauffolgenden Kapiteln werden zwei spezielle Klassen von Polynomen vorgestellt, nämlich die Bernoulli-Polynome und die orthogonalen Polynome.

Nachdem die im weiteren Verlauf der Arbeit benötigten Begriffe und Eigenschaften beschrieben wurden, wird die numerische Integration zunächst allgemein behandelt. In den drei darauffolgenden Kapiteln (dem Hauptteil dieser Arbeit) werden drei große Typen von numerischen Integrationsverfahren – die Newton-Cotes-Formeln, das Romberg-Verfahren und das Gauß-Verfahren – dargestellt und mit Beispielen veranschaulicht.

Ziel dieser Arbeit ist es, die Vielfalt der numerischen Integrationsverfahren durch eine strukturierte Beschreibung und durch zahlreiche Vergleiche und Gegenüberstellungen handhabbar zu machen. Es wird dabei insbesondere auf die Fehlerordnung und die Abschätzung des Fehlers sowie auf die Anzahl der für die näherungsweise Berechnung des Integrals notwendigen Stützwerte eingegangen. Letztere wird als Maß für die Effizienz eines Verfahrens betrachtet, da bei vergleichbarer Genauigkeit der Ergebnisse der geringere Rechenaufwand ausschlaggebend ist, und dieser besteht zu einem Großteil aus der Berechnung der einzelnen Funktionswerte.

Da die numerische Mathematik eng mit dem Einsatz von Computern und der damit einhergehenden Steigerung der Rechengeschwindigkeit verbunden ist, werden in dieser Diplomarbeit auch regelmäßig Codes des Computeralgebrasystems MATHEMATICA vorgestellt, mit denen verschiedene Resultate berechnet werden können. Außerdem werden am Ende der Arbeit Codes zur Durchführung der behandelten Integrationsverfahren zur Verfügung gestellt und deren Benutzung ausführlich erklärt.

2 Allgemeine Grundlagen

In diesem Kapitel werden einige grundlegende Eigenschaften behandelt, auf welche im weiteren Verlauf dieser Arbeit zurückgegriffen wird.

Definition 2.1. Sei $n \in \mathbb{N}$ und seien $a_0, ..., a_n \in \mathbb{R}$ mit $a_n \neq 0$. Dann bezeichnet man $p(x) := a_0 + a_1x + a_2x^2 + ... + a_nx^n$ als reelles Polynom vom Grad n. Ist $a_n = 1$, dann nennt man das Polynom normiert. Die Menge aller Polynome mit Höchstgrad n bezeichnet man mit $\mathcal{P}_n$.

2.1 Stetigkeit und Differenzierbarkeit

Satz 2.2. Sei p ein Polynom, dann ist p stetig.

Satz 2.3. Sei p ein Polynom, dann ist p differenzierbar.

Lemma 2.4. Sei $p(x) = a_nx^n + ... + a_1x + a_0$. Dann gilt $p^{(n)}(x) = a \cdot n!$, $\forall x \in \mathbb{R}$.

Korollar 2.5. Sei p ein normiertes Polynom mit Grad n. Dann gilt $p^{(n)}(x) = n!$, $\forall x \in \mathbb{R}$.

Korollar 2.6. Sei p ein Polynom mit Grad n. Dann gilt $p^{(n+1)}(x) = 0$, $\forall x \in \mathbb{R}$.

Satz 2.7. Sei p ein Polynom, dann gilt $p \in C^\infty$. Polynome sind also glatt.

2.2 Sätze über stetige Funktionen

Die folgenden Sätze gelten allgemein für Funktionen, welche die entsprechenden Stetigkeits- und Differenzierbarkeitsbedingungen erfüllen und somit nach Satz 2.7 insbesondere auch für Polynome.

Satz 2.8. Sei $f : [a, b] \to \mathbb{R}$ eine Funktion. Dann gilt f differenzierbar $\Rightarrow$ f stetig $\Rightarrow$ f Riemann-integrierbar.

Satz 2.9. (Satz von Rolle) Sei $f : [a, b] \to \mathbb{R}$ stetig auf $[a, b]$ und differenzierbar auf (a, b) und sei $f(a) = 0$ und $f(b) = 0$. Dann gibt es ein $\xi \in (a, b)$ mit $f'(\xi) = 0$.

Beweis: Dieser Beweis wurde anhand von [20] durchgeführt. Nach dem Satz vom Minimum und Maximum (vgl. Weierstraß), welcher hier nicht bewiesen wird, besitzt die Funktion f ein Minimum m_1 und ein Maximum m_2. Nun werden zwei Fälle unterschieden:

1. Beide Extremstellen befinden sich am Rand. Somit ist das Maximum und das Minimum der Funktion 0, weshalb die Funktion der Nullfunktion entsprechen muss, also $f(x) = 0$, $\forall x \in [a, b]$ und damit auch $f'(x) = 0$, $\forall x \in (a, b)$. Insbesondere existiert damit ein $\xi \in (a, b)$ mit $f'(\xi) = 0$.

2. Zumindest eine Extremstelle befindet sich im Inneren des Intervalls, also $m_1 \in (a,b)$ oder $m_2 \in (a,b)$. Sei ohne Beschränkung der Allgemeinheit $m_2 \in (a,b)$. Dann ist m_2 eine lokale Extremstelle, und damit gilt $f'(m_2) = 0$. Setze $\xi := m_2 \Rightarrow f'(\xi) = 0$. □

Die Aussage dieses Satzes ist also, dass zwischen zwei Nullstellen einer stetig differenzierbaren Funktion f die Ableitung f' ebenfalls eine Nullstelle besitzt.

Satz 2.10. (Mittelwertsatz der Differentialrechnung) Seien $a, b \in \mathbb{R}$ mit $a < b$, und sei $f : [a,b] \to \mathbb{R}$ stetig auf $[a,b]$ und differenzierbar auf (a,b). Dann existiert ein $\xi \in (a,b)$ mit $f(b) - f(a) = f'(\xi) \cdot (b-a)$.

Beweis: Als Grundlage für den Beweis wurde [20] herangezogen. Man definiere

$$g(x) := f(x) - f(a) - \frac{f(b) - f(a)}{b-a} \cdot (x-a).$$

Diese Funktion ist stetig auf $[a,b]$, da f stetig auf $[a,b]$ ist und Polynome stetig sind. Außerdem gilt $g'(x) = f'(x) - \frac{f(b)-f(a)}{b-a}$, weshalb g differenzierbar auf (a,b) ist.

Wegen

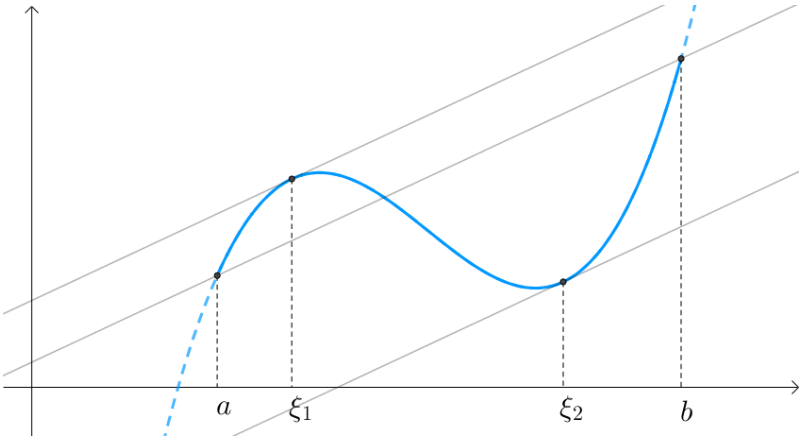
$$g(a) = \underbrace{f(a) - f(a)}_{=0} - \frac{f(b) - f(a)}{b-a} \cdot \underbrace{(a-a)}_{=0} = 0$$

und

$$g(b) = f(b) - f(a) - \frac{f(b) - f(a)}{b-a} \cdot (b-a) = f(a) - f(b) - (f(a) - f(b)) = 0$$

existiert aufgrund des Satzes von Rolle ein $\xi \in (a,b)$ mit $0 = g'(\xi) = f'(\xi) - \frac{f(b)-f(a)}{b-a}$. Daraus ergibt sich $f(b) - f(a) = f'(\xi) \cdot (b-a)$. □

Bemerkung: Folgende Skizze veranschaulicht die Aussage des Mittelwertsatzes der Differentialrechnung geometrisch. Im Intervall (a,b) gibt es mindestens eine Stelle, an welcher die Steigung der Tangente an den Funktionsgraphen der Sekantensteigung im Intervall $[a,b]$ entspricht. In diesem Beispiel gibt es sogar zwei derartige Stellen.



Satz 2.11. (Nullstellensatz von Bolzano) Sei $f : [a,b] \to \mathbb{R}$ stetig auf $[a,b]$ mit $f(a) \leq 0$ und $f(b) \geq 0$ bzw. $f(a) \geq 0$ und $f(b) \leq 0$. Dann existiert ein $x \in [a,b]$ mit $f(x) = 0$.

Beweis: Grundlage für diesen Beweis ist [20]. Es sei ohne Beschränkung der Allgemeinheit $f(a) \leq 0$ und $f(b) \geq 0$. Man definiere die Menge $A := \{x \in [a,b] \mid f(x) \leq 0\}$. Diese Menge

ist nichtleer, weil $a \in A$ und nach oben beschränkt, da b eine obere Schranke darstellt. Somit existiert ein $x_0 \in [a, b]$ mit $x_0 := \sup(A)$ (kleinste obere Schranke). Daraus ergeben sich zwei zu unterscheidende Fälle:

1. Fall: Aus $x_0 = b$ folgt, dass $f(x_0) = f(b) \geq 0$.
2. Fall: Für $x_0 < b$ definiert man $x_n := x_0 + \frac{1}{n} \overset{n\to\infty}{\longrightarrow} x_0$. Es gilt $\exists N : \forall n \geq N : x_n \in [a, b]$. Außerdem ist $x_n \notin A$, da $x_0 = \sup(A)$ und $x_n > x_0$. Weil f stetig ist, gilt

$$f(x_0) = \lim_{n\to\infty} \underbrace{f(x_n)}_{>0, \text{ da } x_n \notin A} \geq 0.$$

In beiden Fällen erhält man somit $f(x_0) \geq 0$. Im weiteren Verlauf wird ein Widerspruchsbeweis geführt. Angenommen, es ist $f(x_0) \neq 0$, also $f(x_0) > 0$. Da f stetig in x_0 ist, existiert für alle $\varepsilon > 0$ ein $\delta > 0$, sodass für alle $x \in [a, b]$ mit $|x - x_0| < \delta$ gilt: $|f(x) - f(x_0)| < \varepsilon$. Außerdem gilt $f(x) - f(x_0) \leq |f(x) - f(x_0)|$. Man definiere die frei wählbare Variable $\varepsilon := f(x_0)$. Dies ist möglich, da laut Annahme $f(x_0) > 0$ gilt. Insgesamt erhält man

$$f(x) - f(x_0) \leq |f(x) - f(x_0)| < f(x_0).$$

Daraus folgt, dass $f(x) > 0$ gelten muss, da sonst $|f(x) - f(x_0)| < f(x_0)$ nicht erfüllt wäre.

Weil $x_0 = \sup(A)$ ist, gibt es ein $x \in A$ mit $x_0 - \delta < x \leq x_0$, und somit gilt $|x - x_0| < \delta$. Wie im vorigen Absatz beschrieben folgt daher $f(x) > 0$ für dieses $x \in A$. Andererseits gilt für alle $x \in A$, dass $f(x) \leq 0$ ist. Man erhält mit $0 < f(x) \leq 0$ einen Widerspruch zur Annahme, dass $f(x_0) \neq 0$ gilt, weshalb $f(x_0) = 0$ gelten muss. □

Satz 2.12. (Zwischenwertsatz) Sei I ein Intervall, $f : I \to \mathbb{R}$ stetig und $\alpha \in \mathbb{R}$. Falls es $x_1, x_2 \in I$ gibt mit $f(x_1) \leq \alpha$ und $f(x_2) \geq \alpha$, dann gibt es ein $x \in I$ mit $f(x) = \alpha$.

Beweis: Dieser Beweis wurde anhand von [20] durchgeführt. Es sind hier zwei Fälle zu unterscheiden:

1. Fall: Für $x_1 \neq x_2$ definiert man $g(x) := f(x) - \alpha$. Es gilt dann $g(x_1) = f(x_1) - \alpha \leq 0$, da $f(x_1) \leq \alpha$ und $g(x_2) = f(x_2) - \alpha \geq 0$, da $f(x_2) \geq \alpha$. Nach dem Nullstellensatz von Bolzano (Satz 2.11) existiert nun ein $x \in I$ mit $0 = g(x) = f(x) - \alpha \Rightarrow f(x) = \alpha$.
2. Fall: Für $x_1 = x_2$ folgt wegen

$$f(x_1) \leq \alpha \leq \underbrace{f(x_2)}_{=f(x_1)},$$

dass für $x := x_1$ somit $f(x) = \alpha$ gelten muss. □

Satz 2.13. (Mittelwertsatz der Integralrechnung) Sei $f : [a, b] \to \mathbb{R}$ stetig und sei $g : [a, b] \to \mathbb{R}$ integrierbar mit $g(x) \leq 0,\ \forall x \in [a, b]$ oder $g(x) \geq 0,\ \forall x \in [a, b]$ (also g hat keinen Vorzeichenwechsel). Dann gibt es ein $\xi \in [a, b]$, sodass folgende Eigenschaft erfüllt ist:

$$\int_a^b f(x)g(x)\,dx = f(\xi) \int_a^b g(x)\,dx$$

Beweis: Dieser Beweis wurde angelehnt an [21, Seite 85]. Sei γ_1 das Minimum von f auf $[a, b]$ und sei γ_2 das Maximum von f auf $[a, b]$. Dann gilt $\gamma_1 \leq f(x) \leq \gamma_2,\ \forall x \in [a, b]$. Sei nun ohne Beschränkung der Allgemeinheit $g(x) \geq 0, \forall x \in [a, b]$ (der andere Fall ändert in

den nachfolgenden Überlegungen nur die Richtung der Vergleichszeichen). Dann gilt

$$\gamma_1 g(x) \leq f(x)g(x) \leq \gamma_2 g(x), \ \forall x \in [a,b].$$

Aus der Monotonie[1] und der Linearität[2] des Riemann-Integrals folgt

$$\gamma_1 \int_a^b g(x)\,dx \leq \int_a^b f(x)g(x)\,dx \leq \gamma_2 \int_a^b g(x)\,dx$$

Man definiere nun $I := \int_a^b g(x)\,dx$. Es sind zwei Fälle zu unterscheiden:

1. Fall: Sei $I = 0$. Dann erhält man aus dem sogenannten Sandwich-Satz folgendes Resultat:

$$0 \leq \int_a^b f(x)g(x)\,dx \leq 0 \Rightarrow \int_a^b f(x)g(x)\,dx = 0$$

Sei nun $\xi \in [a,b]$ beliebig. Dann gilt

$$\underbrace{\int_a^b f(x)g(x)\,dx}_{=0} = f(\xi) \underbrace{\int_a^b g(x)\,dx}_{=0},$$

womit die Aussage in diesem Fall bewiesen ist.

2. Fall: Sei $I \neq 0$. Dann gilt

$$\gamma_1 \leq \frac{1}{I} \int_a^b f(x)g(x)\,dx \leq \gamma_2.$$

Setze $\eta := \frac{1}{I}\int_a^b f(x)g(x)\,dx$. Es gilt dann $\gamma_1 \leq \eta \leq \gamma_2$. Aus dem Zwischenwertsatz (Satz 2.12) folgt, dass es ein $\xi \in [a,b]$ gibt, mit $f(\xi) = \eta$ (weil f stetig ist). Aus

$$f(\xi) = \eta = \frac{1}{I} \int_a^b f(x)g(x)\,dx$$

folgt wegen der Definition $I := \int_a^b g(x)\,dx$ schließlich

$$\int_a^b f(x)g(x)\,dx = f(\xi) \int_a^b g(x)\,dx.$$

□

[1] Für Riemann-integrierbare Funktionen f, g mit $f(x) \leq g(x), \ \forall x$ gilt $\int_a^b f(x)\,dx \leq \int_a^b g(x)\,dx$.

[2] Für Riemann-integrierbare Funktionen f, g und für $c_1, c_2 \in \mathbb{R}$ gilt folgende Eigenschaft: $\int_a^b (c_1 f(x) + c_2 g(x))\ dx = c_1 \int_a^b f(x)\,dx + c_2 \int_a^b g(x)\,dx$

Satz 2.14. (Satz von de l'Hospital) Sei $a, b \in \mathbb{R}$ und seien $f, g : (a, b) \to \mathbb{R}$ differenzierbar. Sei außerdem $g(x) \neq 0,\ \forall x \in (a, b)$ und $g'(x) \neq 0,\ \forall x \in (a, b)$.

Falls $\lim\limits_{x \to a^+} \frac{f'(x)}{g'(x)}$ existiert, dann gelten:

(i) Falls $\lim\limits_{x \to a^+} f(x) = \lim\limits_{x \to a^+} g(x) = 0$, dann existiert $\lim\limits_{x \to a^+} \frac{f(x)}{g(x)}$ und es gilt $\lim\limits_{x \to a^+} \frac{f(x)}{g(x)} = \lim\limits_{x \to a^+} \frac{f'(x)}{g'(x)}$.

(ii) Falls $\lim\limits_{x \to a^+} f(x) = \infty$, dann gilt $\lim\limits_{x \to a^+} \frac{f(x)}{g(x)} = \lim\limits_{x \to a^+} \frac{f'(x)}{g'(x)}$.

Falls $\lim\limits_{x \to b^-} \frac{f'(x)}{g'(x)}$ existiert, dann gelten:

(iii) Falls $\lim\limits_{x \to b^-} f(x) = \lim\limits_{x \to b^-} g(x) = 0$, dann existiert $\lim\limits_{x \to b^-} \frac{f(x)}{g(x)}$ und es gilt $\lim\limits_{x \to b^-} \frac{f(x)}{g(x)} = \lim\limits_{x \to b^-} \frac{f'(x)}{g'(x)}$.

(iv) Falls $\lim\limits_{x \to b^-} f(x) = \infty$, dann gilt $\lim\limits_{x \to b^-} \frac{f(x)}{g(x)} = \lim\limits_{x \to b^-} \frac{f'(x)}{g'(x)}$.

Der obige Satz wurde übernommen von [20].

Satz 2.15. (Substitutionsregel) Seien $[a, b]$ und $[A, B]$ zwei Intervalle. Sei $f : [A, B] \to \mathbb{R}$ eine stetige Funktion, und sei $\varphi : [a, b] \to [A, B]$ stetig differenzierbar. Dann gilt

$$\int_{\varphi(a)}^{\varphi(b)} f(x)\, dx = \int_a^b f(\varphi(t)) \cdot \varphi'(t)\, dt.$$

Beweis: Der Beweis wurde an [29, Folie 18, Satz 8.8] angelehnt. Sei F eine Stammfunktion von f (nach dem Hauptsatz der Differential- und Integralrechnung existiert eine solche). Dann gilt

$$\int_{\varphi(a)}^{\varphi(b)} f(x)\, dx = F(\varphi(b)) - F(\varphi(a)) = (F \circ \varphi)(t)\Big|_a^b = \int_a^b (F \circ \varphi)'(t)\, dt.$$

Aufgrund der Kettenregel folgt schließlich

$$\int_a^b (F \circ \varphi)'(t)\, dt = \int_a^b F'(\varphi(t)) \cdot \varphi'(t)\, dt = \int_a^b f(\varphi(t)) \cdot \varphi'(t)\, dt.$$

□

Beispiel: Es soll das bestimmte Integral $\int_0^1 \cos(3x + 5)\, dx$ berechnet werden. Dazu wird die Funktion $t = \varphi(x) = 3x + 5$ definiert, wodurch man $\varphi'(x) = 3$ erhält. Unter Verwendung von Satz 2.15 (beginnend auf der rechten Seite) ergibt sich:

$$\int_0^1 \cos(3x + 5)\, dx = \frac{1}{3} \int_0^1 \cos(\varphi(x)) \cdot \underbrace{\varphi'(x)}_{=3}\, dx \overset{2.15}{=} \frac{1}{3} \int_{\varphi(0)}^{\varphi(1)} \cos(t)\, dt = \frac{1}{3} \sin(t)\Big|_5^8 \approx 0{,}649.$$

Definition 2.16. Sei M ein kompakter Raum (z. B. die Menge $\mathbb{R}$ der reellen Zahlen oder ein abgeschlossenes Intervall $[a, b]$). Sei $f : M \to \mathbb{R}$ eine stetige Funktion. Dann definiert man die Maximumsnorm bzw. Unendlichnorm von f auf M durch

$$\|f\|_\infty := \max_{x \in M} |f(x)|.$$

Satz 2.17. (Approximationssatz von Weierstraß) Sei $f : [a, b] \to \mathbb{R}$ stetig und sei $\varepsilon > 0$ beliebig. Dann gilt auf $[a, b]$ Folgendes:

$$\exists n \in \mathbb{N} : \exists p \in \mathcal{P}_n : \|f - p\|_\infty < \varepsilon$$

Der Beweis kann unter [16] nachgelesen werden.

Satz 2.18. (Formel von Faà di Bruno) Seien $x \mapsto g(x) =: t$ und $t \mapsto f(t)$ zwei n-mal differenzierbare Funktionen, die von jeweils einer Variable abhängen. Dann gilt für die Komposition $(f \circ g)(x) := f(g(x))$ Folgendes, wobei $k_1, ..., k_n \in \mathbb{N}_0$ und $k_1 + ... + k_n = n$ ist:

$$\frac{d^n}{dx^n}(f \circ g)(x) = \sum_{1k_1+2k_2+...+nk_n=n} \frac{n!}{k_1! \cdots k_n!} \cdot \frac{d^{k_1+...+k_n}}{dt^{k_1+...+k_n}} f(t) \cdot \prod_{\substack{m=1 \\ k_m \geq 1}}^{n} \left(\frac{d^m g}{dx^m} \cdot \frac{1}{m!} \right)^{k_m}$$

Ein Beweis kann unter [3, Seite 101, Satz 7.1] aufgefunden werden. Als weitere Quelle (hauptsächlich in Bezug auf die Formulierung des Satzes) diente [25].

Korollar 2.19. Sei f eine n-mal differenzierbare Funktion und sei g eine lineare Funktion, dann gilt

$$\frac{d^n}{dx^n}(f \circ g)(x) = \frac{d^n}{dt^n} f(t) \cdot \left(\frac{dg}{dx} \right)^n.$$

Beweis: Da g eine lineare Funktion ist, gilt $\frac{d^m g}{dx^m} = 0$ für alle $m > 1$. Daher ist $(k_1, ..., k_n) = (n, 0, ..., 0)$ die einzige Partition von n, bei welcher kein Faktor des Produkts 0 ergibt, denn außer k_1 sind alle $k_m = 0$.

Somit erhält man

$$\frac{d^n}{dx^n}(f \circ g)(x) = \frac{n!}{n! \cdot 0! \cdots 0!} \cdot \frac{d^n}{dt^n} f(t) \cdot \left(\frac{d^1 g}{dx^1} \cdot \frac{1}{1!} \right)^n = \frac{d^n}{dt^n} f(t) \cdot \left(\frac{dg}{dx} \right)^n.$$

□

2.3 Nullstellen

Satz 2.20. (Fundamentalsatz der Algebra) Sei $p = \sum_{k=0}^{n} a_k x^k$ ein Polynom mit $\operatorname{grad} p \geq 1$. Dann besitzt p genau n (nicht unbedingt verschiedene) Nullstellen in $\mathbb{C}$.

Satz 2.21. Sei $N \in \mathbb{N}$, sei p ein Polynom mit Grad N, und seien $x_1 < x_2 < ... < x_n$ paarweise verschiedene reelle Nullstellen von p mit den Vielfachheiten $k_1, ..., k_n$, sodass gilt $N = \sum_{i=1}^{n} k_i$. Es hat also p insgesamt N Nullstellen. Dann besitzt $p^{(m)}$ für $1 \leq m \leq N$ mindestens $N - m$ reelle Nullstellen.

Beweis: Es reicht zu zeigen, dass die Anzahl der reellen Nullstellen pro Ableitung um genau 1 abnimmt, da die ursprüngliche Funktion N Nullstellen hat. Es existiert ein $a \in \mathbb{R}$, sodass

man p folgendermaßen darstellen kann:

$$p(x) = a \cdot \prod_{i=1}^{n} (x - x_i)^{k_i}$$

Aus der Produktregel folgt für p' folgendes Ergebnis:

$$p'(x) = a \cdot \sum_{j=1}^{n} k_j (x - x_j)^{k_j - 1} \prod_{i=1, i \neq j}^{n} (x - x_i)^{k_i}$$

Bemerkung: Man hat nun also insgesamt N Summanden, denn gemäß der Produktregel wird jeder der N Linearfaktoren einzeln abgeleitet, weshalb man zu jeder Nullstelle x_j insgesamt k_j-mal den Faktor $(x - x_j)^{k_j - 1}$ erhält. Das Produkt dahinter enthält alle übrigen Linearfaktoren der anderen Nullstellen.

Nach dem Satz von Rolle (Satz 2.9) gibt es für alle $i \in \{1, ..., n-1\}$ ein $\xi_i \in (x_i, x_{i+1})$ mit $p'(\xi_i) = 0$. Das bedeutet, p' hat nun insgesamt mindestens $n - 1$ „neue“ Nullstellen. Andererseits wird von jeder ursprünglichen Nullstelle die Vielfachheit um 1 herabgesetzt. Somit verliert p' gegenüber p insgesamt n Nullstellen. Für die Anzahl $\tilde{N}$ der Nullstellen von p' gilt also $\tilde{N} \geq N + (n-1) - n = N - 1$. □

Lemma 2.22. Sei $f : [a, b] \to \mathbb{R}$ eine glatte Funktion mit $f(a) = 0$ und $f(b) = 0$. Sei $x_0 \in (a, b)$ mit $f(x_0) = 0$. Dann gibt es ein $\xi \in (a, b)$ mit $f''(\xi) = 0$.

Beweis: Es sind zwei Fälle zu unterscheiden:

1. Fall: Es gibt $\alpha_1, \alpha_2 \in (a, b)$ mit $\alpha_1 < \alpha_2$ und $f(x) = 0,\ \forall x \in (\alpha_1, \alpha_2)$. Dann gilt $f''(x) = 0,\ \forall x \in (\alpha_1, \alpha_2)$, womit die Aussage gezeigt ist.
2. Fall: Die Funktion f hat nur endlich viele Nullstellen und es sei x_0 eine davon. Dann muss es $\alpha_1, \alpha_2 \in (a, b)$ geben mit $\alpha_1 < x < \alpha_2$ sodass ohne Beschränkung der Allgemeinheit gilt $f(\alpha_1) > 0$ und $f(\alpha_2) < 0$. Wegen $f(a) = 0$ und $f(\alpha_1) > 0$ sowie wegen $f(\alpha_2) < 0$ und $f(b) = 0$ muss es $\beta_1 \in (a, \alpha_1)$ und $\beta_3 \in (\alpha_2, b)$ geben, sodass $f'(\beta_1) > 0$ und $f'(\beta_3) > 0$ gilt. Wegen $f(\alpha_1) > 0$ und $f(\alpha_2) < 0$ muss es außerdem ein $\beta_2 \in (\alpha_1, \alpha_2)$ geben, mit $f'(\beta_2) < 0$. Daher muss es $\gamma_1 \in (\beta_1, \beta_2)$ und $\gamma_2 \in (\beta_2, \beta_3)$ geben mit $f'(\gamma_1) = f'(\gamma_2) = 0$.

   Hier sind erneut zwei Fälle zu unterscheiden:

   - Es gibt kein $x \in (\gamma_1, \gamma_2)$ mit $f'(x) \neq 0$. Dann ist $f''(x) = 0,\ \forall x \in (\gamma_1, \gamma_2)$ und die Aussage ist bestätigt.
   - Es gibt ohne Beschränkung der Allgemeinheit ein $\delta \in (\gamma_1, \gamma_2)$ mit $f'(\delta) > 0$. Dann muss es ein $\varepsilon_1 \in (\gamma_1, \delta)$ geben mit $f''(\varepsilon_1) > 0$ und es muss ein $\varepsilon_2 \in (\delta, \gamma_2)$ geben mit $f''(\varepsilon_2) < 0$. Aus dem Nullstellensatz von Bolzano (Satz 2.11) folgt somit, dass es ein $\xi \in (\varepsilon_1, \varepsilon_2) \subseteq (a, b)$ gibt mit $f''(\xi) = 0$. □

2.4 Eigenschaften gerader und ungerader Funktionen

Lemma 2.23. Sei p ein Polynom. Dann gilt:

(i) Ist p gerade, dann ist $\int_0^x p(t)\,dt$ ein ungerades Polynom.

(ii) Ist p ungerade, dann ist $\int_0^x p(t)\,dt$ ein gerades Polynom.

Beweis: (i) Sei $n \in \mathbb{N}$ so, dass $2n$ der Höchstgrad von p ist. Da p gerade ist, kann man das Polynom als Summe gerader Potenzen schreiben, also

$$p(x) = \sum_{k=0}^{n} a_{2k} x^{2k}.$$

Es gilt außerdem folgende Nebenrechnung:

$$\int_0^x t^{2k}\, dt = \left.\frac{t^{2k+1}}{2k+1}\right|_0^x = \frac{x^{2k+1}}{2k+1}$$

Damit erhält man schließlich

$$\int_0^x p(t)\, dt = \int_0^x \left(\sum_{k=0}^{n} a_{2k} t^{2k}\right) dt = \sum_{k=0}^{n} a_{2k} \int_0^x t^{2k}\, dt = \sum_{k=0}^{n} \frac{a_{2k}}{2k+1}\, x^{2k+1},$$

wobei es sich um ein ungerades Polynom handelt.

(ii) Sei nun $n \in \mathbb{N}_0$ so, dass $2n+1$ der Höchstgrad von p ist. Dieser Beweis verläuft analog zu (i) mit dem Ansatz

$$p(x) = \sum_{k=0}^{n} a_{2k+1} x^{2k+1}.$$

□

Lemma 2.24. Sei $a \in \mathbb{R}$ und sei $f : [-a, a] \to \mathbb{R}$ Riemann-integrierbar. Dann gelten folgende Eigenschaften:

(i) $\int_{-a}^{a} f(x)\, dx = 0$, falls f ungerade ist.

(ii) $\int_{-a}^{a} f(x)\, dx = 2\int_0^a f(x)\, dx$, falls f gerade ist.

Beweis: Durch Aufteilen des Integrals erhält man

$$\int_{-a}^{a} f(x)\, dx = \int_{-a}^{0} f(x)\, dx + \int_0^a f(x)\, dx.$$

Aus Satz 2.15 erhält man mit $x = \varphi(u) = -u$ und $\varphi'(u) = -1$ das folgende Resultat:

$$\int_a^0 f(-u)\, du = (-1) \cdot \int_a^0 f(\varphi(u)) \cdot \underbrace{\varphi'(u)}_{=-1}\, du = -\int_{\varphi(a)}^{\varphi(0)} f(x)\, dx = -\int_{-a}^{0} f(x)\, dx$$

Unter Verwendung der Eigenschaft $\int_a^b f(t)\, dt = -\int_b^a f(t)\, dt$ erhält man somit

$$\int_{-a}^{0} f(x)\, dx + \int_0^a f(x)\, dx = -\int_a^0 f(-u)\, du + \int_0^a f(x)\, dx = \int_0^a f(-u)\, du + \int_0^a f(x)\, dx.$$

(i) Für ungerade Funktionen folgt nun wegen $f(-u) = -f(u)$

$$\int_{-a}^{a} f(x)\, dx = \int_0^a \underbrace{f(-u)}_{=-f(u)}\, du + \int_0^a f(x)\, dx = 0.$$

(ii) Für gerade Funktionen erhält man wegen $f(-u) = f(u)$

$$\int_{-a}^{a} f(x)\,dx = \int_{0}^{a} \underbrace{f(-u)}_{=f(u)}\,du + \int_{0}^{a} f(x)\,dx = 2\int_{0}^{a} f(x)\,dx.$$

□

Lemma 2.25. Sei $f : \mathbb{R} \to \mathbb{R}$ eine ungerade Funktion, also $f(-x) = -f(x)$, $\forall x \in \mathbb{R}$. Dann gilt $f(0) = 0$.

Beweis: Wegen $f(-x) = -f(x)$ gilt $f(0) = f(-0) = -f(0)$, und daher muss $f(0) = 0$ sein. □

2.5 Taylorpolynome

Die nachfolgend vorgestellte Taylor-Formel kann verwendet werden, um Funktionen in der Nähe einer bestimmten Stelle (der Entwicklungsstelle) durch sogenannte Taylorpolynome anzunähern.

Definition 2.26. Sei I ein Intervall, sei $n \in \mathbb{N}_0$, und sei $f : I \to \mathbb{R}$ eine $(n+1)$-mal stetig differenzierbare Funktion. Sei außerdem $\tilde{x} \in I$ die sogenannte Entwicklungsstelle. Dann bezeichnet man das durch

$$T_n f(x) := \sum_{k=0}^{n} \frac{f^{(k)}(\tilde{x})}{k!} \cdot (x - \tilde{x})^k$$

definierte Polynom als n-tes Taylorpolynom der Funktion f an der Entwicklungsstelle $\tilde{x}$.

Um nun ermitteln zu können, wie groß für ein $x \in I$ die Abweichung des Funktionswertes des Taylorpolynoms vom tatsächlichen Funktionswert ist, wird nachfolgend eine Formel für die Abschätzung dieses sogenannten Restglieds hergeleitet.

Definition 2.27. Sei $x \in I$ und $T_n f(x)$ das n-te Taylorpolynom einer $(n+1)$-mal stetig differenzierbaren Funktion. Dann definiert man das Restglied R_{n+1} dieses Taylorpolynoms folgendermaßen:

$$R_{n+1}(x) := f(x) - T_n f(x)$$

Satz 2.28. Sei $n \in \mathbb{N}_0$, $f : I \to \mathbb{R}$ eine $(n+1)$-mal stetig differenzierbare Funktion und $\tilde{x} \in I$ die Entwicklungsstelle des Taylorpolynoms. Dann gilt für ein $x \in I$ die folgende Formel zur Berechnung des Restglieds $R_{n+1}(x)$:

$$R_{n+1}(x) = \frac{1}{n!} \int_{\tilde{x}}^{x} (x - t)^n f^{(n+1)}(t)\,dt$$

Beweis: Der Beweis stützt sich auf [7, Seite 3 und 4]. Es wird ein Induktionsbeweis geführt. Für $n = 0$ erhält man aus dem Hauptsatz der Differential- und Integralrechnung wegen $\int_{\tilde{x}}^{x} f'(t)\,dt = f(x) - f(\tilde{x})$ Folgendes:

$$R_1(x) := f(x) - f(\tilde{x}) = \int_{\tilde{x}}^{x} f'(t)\,dt = \frac{1}{0!} \cdot \int_{\tilde{x}}^{x} f'(t)\,dt$$

Es ist also $f(\tilde{x})$ das 0-te Glied des Taylorpolynoms (also jenes, das man durch Einsetzen von $k = 0$ in die Summe aus Definition 2.26 erhält) und $R_1(x)$ des entsprechende Restglied, womit die Behauptung für $n = 0$ gezeigt ist.

Für $n = 1$ wird das Restglied $R_1(x)$ partiell integriert, wobei für den Faktor 1 die Stammfunktion $t - x = -(x - t)$ verwendet wird. Zu beachten ist hierbei, dass x in diesem Fall als Konstante zu behandeln ist. Es gilt:

$$R_1(x) = \int_{\tilde{x}}^{x} f'(t)\,dt = \int_{\tilde{x}}^{x} 1 \cdot f'(t)\,dt = -(x-t) \cdot f'(t)\Big|_{\tilde{x}}^{x} + \int_{\tilde{x}}^{x} (x-t) f''(t)\,dt =$$

$$= (x - \tilde{x}) \cdot f'(x) + \int_{\tilde{x}}^{x} (x-t) f''(t)\,dt = \underbrace{\frac{f'(\tilde{x})}{1!} \cdot (x-\tilde{x})^1}_{=:\gamma} + \underbrace{\frac{1}{1!} \int_{\tilde{x}}^{x} (x-t)^1 f''(t)\,dt}_{=R_2(x)}$$

Hier ist γ das 1. Glied des Taylorpolynoms (also jenes, das man durch Einsetzen von $k = 1$ in die Summe aus Definition 2.26 erhält) und $R_2(x)$ das entsprechende Restglied, womit für $n = 1$ die Behauptung ebenfalls gezeigt wäre. Durch Abspalten des 1. Gliedes des Taylorpolynoms aus dem Restglied R_1 erhält man somit einen Ausdruck für das Restglied R_2. Dieses Prinzip der partiellen Integration wird nun fortgesetzt. Als Induktionsvoraussetzung gelte für ein beliebiges aber festes n

$$R_n(x) = \frac{1}{(n-1)!} \int_{\tilde{x}}^{x} (x-t)^{n-1} f^{(n)}(t)\,dt$$

Für den Induktionsschritt $n \mapsto n + 1$ erhält man folgendes Resultat:

$$R_n(x) = \frac{1}{(n-1)!} \int_{\tilde{x}}^{x} (x-t)^{n-1} f^{(n)}(t)\,dt =$$

$$= \frac{1}{(n-1)!} \cdot \left(\frac{(x-t)^n}{n} \cdot (-1) \cdot f^{(n)}(t)\Big|_{\tilde{x}}^{x} + \frac{1}{n} \int_{\tilde{x}}^{x} (x-t)^n f^{(n+1)}(t)\,dt \right) =$$

$$= \frac{(x-\tilde{x})^n}{n!} \cdot f^{(n)}(t) + \underbrace{\frac{1}{n!} \int_{\tilde{x}}^{x} (x-t)^n f^{(n+1)}(t)\,dt}_{=R_{n+1}(x)}$$

Der erste Summand entspricht dem n-ten Glied des Taylorpolynoms (also jenes, das man durch Einsetzen von $k = n$ in die Summe aus Definition 2.26 erhält) und $R_{n+1}(x)$ ist das entsprechende Restglied, womit die Behauptung für alle $n \in \mathbb{N}_0$ bewiesen wurde. □

Definition 2.29. Man bezeichnet das Restglied des Taylorpolynoms in der Form

$$R_{n+1}(x) = \frac{1}{n!} \int_{\tilde{x}}^{x} (x-t)^n f^{(n+1)}(t)\,dt$$

als Integralrestglied bzw. Restglied in Integralform.

Nachfolgend wird eine weitere Darstellungsform des Restglieds behandelt, die sogenannte Lagrange-Darstellung. Zunächst wird aber eine Eigenschaft gezeigt, die für den späteren Beweis benötigt wird.

Lemma 2.30. Seien $x, \tilde{x} \in \mathbb{R}$ und $n \in \mathbb{N}_0$. Dann hat $(x-t)^n$ für alle $t \in [\min(x, \tilde{x}), \max(x, \tilde{x})]$ ein einheitliches Vorzeichen.

Beweis: Es werden zwei Fälle unterschieden:

1. Fall: $x \geq \tilde{x} \Rightarrow t \leq x \Rightarrow x - t \geq 0, \forall t \in [\tilde{x}, x]$.
2. Fall: $x \leq \tilde{x} \Rightarrow t \geq x \Rightarrow x - t \leq 0, \forall t \in [x, \tilde{x}]$.

Insgesamt hat daher $x - t$ und somit auch $(x - t)^n$ für alle $t \in [\min(x, \tilde{x}), \max(x, \tilde{x})]$ ein konstantes Vorzeichen. □

Satz 2.31. Es existiert ein $\xi \in [\min(x, \tilde{x}), \max(x, \tilde{x})]$, sodass Folgendes gilt:

$$R_{n+1}(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - \tilde{x})^{n+1}$$

Beweis: Der Beweis wird in Anlehnung an [7, Seite 4] geführt. Ausgangsbasis für diesen Beweis bildet die in Satz 2.28 gewonnene Formel

$$R_{n+1}(x) = \frac{1}{n!} \int_{\tilde{x}}^{x} (x - t)^n f^{(n+1)}(t)\, dt.$$

Weil $(x-t)^n$ gemäß Lemma 2.30 auf dem Intervall $[\min(x, \tilde{x}), \max(x, \tilde{x})]$ keinen Vorzeichenwechsel aufweist, gibt es aufgrund des Mittelwertsatzes der Integralrechnung (Satz 2.13) ein $\xi \in [\min(x, \tilde{x}), \max(x, \tilde{x})]$, sodass gilt:

$$\frac{1}{n!} \int_{\tilde{x}}^{x} (x - t)^n f^{(n+1)}(t)\, dt = \frac{f^{(n+1)}(\xi)}{n!} \int_{\tilde{x}}^{x} (x - t)^n\, dt = \frac{f^{(n+1)}(\xi)}{n!} \cdot \frac{(x-t)^{n+1}}{n+1} \cdot (-1)\Big|_{\tilde{x}}^{x} =$$

$$= \frac{f^{(n+1)}(\xi)}{(n+1)!} \cdot (x - \tilde{x})^{n+1}$$

□

Definition 2.32. Man bezeichnet das Restglied des Taylorpolynoms in der Form

$$R_{n+1}(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - \tilde{x})^{n+1}$$

als Lagrange-Restglied bzw. Restglied in Lagrange-Form.

3 Polynominterpolation

Unter Polynominterpolation versteht man das Ermitteln einer Polynomfunktion, deren Graph durch vorgegebene Punkte (die sogenannten Stützpunkte) verläuft. Ein Stützpunkt ist dabei eine Kombination von Stützstelle und Stützwert. Anlass dazu kann beispielsweise sein, dass man nur endlich viele Messpunkte eines physikalischen Experiments kennt und den Zusammenhang durch eine Polynomfunktion modellieren möchte, deren Graph durch diese Messpunkte verläuft. Aufgrund der leichten Handhabbarkeit von Polynomen kann es jedoch auch nützlich sein, eine bereits gegebene Funktion an bestimmten Stellen auszuwerten und durch eine Polynomfunktion zu approximieren, deren Graph durch diese Punkte verläuft.

Definition 3.1. Seien $x_0, ..., x_n \in \mathbb{R}$ vorgegebene Stützstellen. Dann bezeichnet man

$$\phi(x) := \prod_{k=0}^{n} (x - x_k)$$

als das normierte Stützstellenpolynom zu den Stützstellen $x_0, ..., x_n$.

Es handelt sich dabei um jenes normierte Polynom, dessen Nullstellen sich genau an den Stützstellen befinden[1].

Beispiel: Es seien die fünf Stützstellen $(x_0, x_1, x_2, x_3, x_4) := (-3, -1, 0, 2, 5)$ gegeben. Das entsprechende normierte Stützstellenpolynom lautet dann

$$\phi(x) = (x+3) \cdot (x+1) \cdot x \cdot (x-2) \cdot (x-5) = x^5 - 3x^4 - 15x^3 + 19x^2 + 30x.$$

Definition 3.2. Sei $n \in \mathbb{N}$, seien $x_0, ..., x_n \in \mathbb{R}$ paarweise verschiedene Stützstellen, und seien $y_0, ..., y_n \in \mathbb{R}$ vorgegebene Stützwerte. Dann bezeichnet man ein Polynom p mit $p(x_i) = y_i$ für $i \in \{0, ..., n\}$ als Interpolationspolynom durch die gegebenen Stützpunkte.

Nachfolgend werden allgemeine Eigenschaften der Polynominterpolation erläutert sowie drei Verfahren vorgestellt, mit welchen Interpolationspolynome zu vorgegebenen Stützpunkten bestimmt werden können. Außerdem werden die Vor- und Nachteile dieser Methoden und deren Einsatzmöglichkeiten beschrieben.

3.1 Eigenschaften

Zu Beginn soll gezeigt werden, dass es für $n + 1$ Stützpunkte ein eindeutig bestimmtes Interpolationspolynom mit Höchstgrad n gibt. Dafür muss jedoch zunächst die Vandermonde-

[1] Da die Nullstellen eines Polynoms durch Multiplikation mit $a \in \mathbb{R} \setminus \{0\}$ erhalten bleiben, gibt es unendlich viele nichtnormierte Polynome mit dieser Eigenschaft.

Matrix, benannt nach Alexandre-Théophile Vandermonde (Frankreich, 1735 – 1796), behandelt werden.

Definition 3.3. Seien $x_0, ..., x_n \in \mathbb{R}$. Dann ist die Vandermonde-Matrix $\mathcal{V}(x_0, ..., x_n)$ folgendermaßen definiert:

$$\mathcal{V}(x_0, ..., x_n) := \begin{pmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{pmatrix}$$

Es handelt sich dabei um eine quadratische Matrix. Aus der obigen Definition geht auch hervor, dass für $x_i = 0$ die Definition $0^0 := 1$ gewählt wurde. Die Determinante einer Vandermonde-Matrix wird als Vandermonde-Determinante bezeichnet. Für sie gilt die folgende Berechnungsformel.

Lemma 3.4. Sei $\mathcal{V}(x_0, ..., x_n)$ eine Vandermonde-Matrix. Dann gilt:

$$\det \mathcal{V}(x_0, ..., x_n) = \prod_{0 \le k < l \le n} (x_l - x_k)$$

Die Bedingung $0 \le k < l \le n$ bedeutet, dass alle Faktoren $(x_l - x_k)$ vorkommen, bei denen $k, l \in \{0, ..., n\}$ ist und $k < l$ erfüllt wird.

Beweis: Grundlage des Beweises ist [13, Seite 2, Z3]. Es wird ein Induktionsbeweis geführt.

Für den Induktionsanfang $n = 0$ ist die linke Seite $\det(\mathcal{V}(x_0)) = \det(1) = 1$. Die rechte Seite entspricht dem leeren Produkt, welches 1 ergibt, wodurch die Aussage für $n = 0$ erfüllt ist.

Angenommen, die zu zeigende Behauptung gilt für ein beliebiges aber festes $n \in \mathbb{N}$ (Induktionsvoraussetzung).

Für $n + 1$ sieht die linke Seite folgendermaßen aus:

$$\det(\mathcal{V}(x_0, ..., x_{n+1})) = \det \begin{pmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^{n+1} \\ 1 & x_1 & x_1^2 & \cdots & x_1^{n+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n+1} \\ 1 & x_{n+1} & x_{n+1}^2 & \cdots & x_{n+1}^{n+1} \end{pmatrix}$$

Nun wird von jeder Spalte der Matrix das x_{n+1}-fache der unmittelbar links davon stehenden Spalte subtrahiert. Dies ist möglich, da der Wert der Determinante unverändert bleibt, wenn man zu einer Spalte bzw. Zeile ein Vielfaches einer anderen Spalte bzw. Zeile addiert. Die erste Spalte bleibt als einzige unverändert, da hier 0 subtrahiert wird. Insgesamt erhält man dadurch:

$$\det \begin{pmatrix} 1 & x_0 - x_{n+1} & x_0^2 - x_{n+1}x_0 & \cdots & x_0^{n+1} - x_{n+1}x_0^n \\ 1 & x_1 - x_{n+1} & x_1^2 - x_{n+1}x_1 & \cdots & x_1^{n+1} - x_{n+1}x_1^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n - x_{n+1} & x_n^2 - x_{n+1}x_n & \cdots & x_n^{n+1} - x_{n+1}x_n^n \\ 1 & x_{n+1} - x_{n+1} & x_{n+1}^2 - x_{n+1}x_{n+1} & \cdots & x_{n+1}^{n+1} - x_{n+1}x_{n+1}^n \end{pmatrix} =$$

$$= \det \begin{pmatrix} 1 & x_0 - x_{n+1} & (x_0 - x_{n+1})x_0 & \cdots & (x_0 - x_{n+1})x_0^n \\ 1 & x_1 - x_{n+1} & (x_1 - x_{n+1})x_1 & \cdots & (x_1 - x_{n+1})x_1^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n - x_{n+1} & (x_n - x_{n+1})x_n & \cdots & (x_n - x_{n+1})x_n^n \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

Entwickelt man nun die Determinante gemäß des Laplaceschen Entwicklungssatzes in Zeilenform[2], also durch Anwenden der Formel

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} \cdot a_{ij} \cdot \det A_{ij},$$

nach der letzten Zeile, so erhält man

$$(-1)^{(n+2)+1} \cdot 1 \cdot \det \begin{pmatrix} x_0 - x_{n+1} & (x_0 - x_{n+1})x_0 & \cdots & (x_0 - x_{n+1})x_0^n \\ x_1 - x_{n+1} & (x_1 - x_{n+1})x_1 & \cdots & (x_1 - x_{n+1})x_1^n \\ \vdots & \vdots & \ddots & \vdots \\ x_n - x_{n+1} & (x_n - x_{n+1})x_n & \cdots & (x_n - x_{n+1})x_n^n \end{pmatrix}$$

da alle anderen Summenden wegen dem Faktor $a_{(n+2),j} = 0,\ \forall j > 1$ wegfallen. Es bleibt also nur der erste Summand der obigen Formel erhalten. Für den Exponent ist hier zu beachten, dass die ursprüngliche Matrix $n + 2$ Zeilen hatte.

Bei der somit entstandenen Matrix wird für $i \in \{1, ..., n + 1\}$ aus der i-ten Zeile der Matrix der Faktor $x_{i-1} - x_{n+1}$ herausgehoben, was an der Determinante der Matrix nichts ändert. Man erhält damit folgenden Term:

$$(-1)^{n+3} \cdot (x_0 - x_{n+1}) \cdots (x_n - x_{n+1}) \cdot \det \begin{pmatrix} 1 & x_0 & \cdots & x_0^n \\ 1 & x_1 & \cdots & x_1^n \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^n \end{pmatrix}$$

Nun werden alle $n + 1$ herausgehobenen Faktoren mit -1 multipliziert. Der übrig bleibende Faktor $(-1)^2 = 1$ fällt weg. Jetzt kann die Determinante durch die Induktionsvoraussetzung ersetzt werden. Man erhält schließlich

$$(x_{n+1} - x_0) \cdots (x_{n+1} - x_n) \cdot \prod_{0 \le k < l \le n} (x_l - x_k) = \prod_{0 \le k < l \le n+1} (x_l - x_k).$$

□

[2]Hier ist A_{ij} jene Matrix, die aus A entsteht, wenn die i-te Zeile und die j-te Spalte gestrichen werden.

Definition 3.5. Eine $n \times n$-Matrix A heißt regulär bzw. invertierbar, wenn eine inverse Matrix A^{-1} existiert, sodass $A \cdot A^{-1} = A^{-1} \cdot A = I$ gilt, wobei I die Einheitsmatrix mit Format $n \times n$ ist.

Lemma 3.6. Eine $n \times n$-Matrix A ist genau dann regulär, wenn $\det A \neq 0$ gilt.

Lemma 3.7. Die Vandermonde-Matrix $\mathcal{V}(x_0, ..., x_n)$ ist genau dann regulär, wenn $x_0, ..., x_n$ paarweise verschieden sind.

Beweis: Dies ist sofort ersichtlich, da für $x_k \neq x_l$ der Faktor $(x_l - x_k)$ ungleich 0 ist und somit das Produkt aus Lemma 3.4 nicht 0 ergeben kann. □

Der nachfolgende Satz wird zeigen, dass es immer ein Interpolationspolynom $p \in \mathcal{P}_n$ (also mit Höchstgrad n) gibt, und dass dieses auch eindeutig bestimmt ist.

Satz 3.8. Sei $n \in \mathbb{N}$, seien $x_0, ..., x_n \in \mathbb{R}$ paarweise verschiedene Stützstellen, und seien $y_0, ..., y_n \in \mathbb{R}$ die zugehörigen Stützwerte. Dann existiert genau ein Polynom p, welches höchstens Grad n besitzt und für welches $p(x_k) = y_k$, $\forall k \in \{0, ..., n\}$ gilt.

Beweis: Die Beweisführung erfolgte in Anlehnung an [8, Seite 195]. Setze $p(x) := a_0 + a_1 x + ... + a_n x^n$. Es handelt sich dabei um ein Polynom mit Höchstgrad n. Durch Einsetzen der Stützstellen x_k und der Stützwerte y_k mit $k \in \{0, ..., n\}$ erhält man folgendes lineares Gleichungssystem:

$$a_0 + a_1 x_k + ... + a_n x_k^n = y_k, \quad k \in \{0, ..., n\}$$

Die Koeffizientenmatrix dieses linearen Gleichungssystems mit den Unbekannten $a_0, ..., a_n$ lautet dann

$$\begin{pmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{pmatrix}.$$

Dabei handelt es sich um die Vandermonde-Matrix $\mathcal{V}(x_0, x_1, ..., x_n)$, welche nach Lemma 3.7 regulär ist, weil die Stützstellen $x_0, ..., x_n$ gemäß Voraussetzung paarweise verschieden sind. Somit ist das lineare Gleichungssystem eindeutig lösbar und das Polynom p eindeutig bestimmt. □

Definition 3.9. Sei $n \in \mathbb{N}$, seien $x_0, ..., x_n \in \mathbb{R}$ paarweise verschiedene Stützstellen, und seien $y_0, ..., y_n \in \mathbb{R}$ die zugehörigen Stützwerte. Dann wird jenes nach Satz 3.8 eindeutig bestimmte Interpolationspolynom mit Höchstgrad n als minimales Interpolationspolynom bezeichnet und nachfolgend mit $\psi(x)$ abgekürzt.

Bemerkung 1: Möchte man hervorheben, dass dadurch beispielsweise die Funktion f interpoliert wird, so wird auch $\psi_f(x)$ geschrieben.

Bemerkung 2: Wie bereits aus der Formulierung von Satz 3.8 hervorgeht, muss das Interpolationspolynom nicht unbedingt Grad n besitzen. Folgendes Beispiel wird zeigen, dass es auch einen niedrigeren Grad aufweisen kann.

Beispiel: Gegeben seien die Stützpunkte $(x_0, y_0) = (1, 1), (x_1, y_1) = (2, 2)$ und $(x_2, y_2) = (3, 3)$. Man erhält daraus das lineare Gleichungssystem

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Die Lösung lautet $(a_0, a_1, a_2) = (0, 1, 0)$. Somit ist das Interpolationspolynom $p(x) = x$, also ein Polynom vom Grad 1, obwohl $n = 2$ gilt.

Durch den obigen Satz wurde gezeigt, dass es für n Stützpunkte genau ein Interpolationspolynom gibt, dessen Grad höchstens n ist. Jedoch gibt es unendlich viele Interpolationspolynome höheren Grades.

Satz 3.10. Seien $n, m \in \mathbb{N}$ mit $m > n$, seien $x_0, ..., x_n \in \mathbb{R}$ paarweise verschiedene Stützstellen, und seien $y_0, ..., y_n \in \mathbb{R}$ die zugehörigen Stützwerte. Dann besitzen genau diejenigen Polynome p mit Grad m die Interpolationseigenschaft $p(x_i) = y_i$ für $i \in \{0, ..., n\}$, welche in der Form $p(x) = \psi(x) + \phi(x)q(x)$ dargestellt werden können. Dabei sind $\phi(x)$ das Stützstellenpolynom aus Definition 3.1 und $\psi(x)$ das minimale Interpolationspolynom aus Definition 3.9. Das Polynom q besitzt Grad $m - n - 1$.

Beweis: Grundlage für diesen Beweis ist [8, Seite 196]. Zunächst wird gezeigt, dass p ein Interpolationspolynom ist. Es gilt $\phi(x_i) = 0$ für $i \in \{0, 1, ..., n\}$. Daraus folgt $p(x_i) = \psi(x_i) = y_i$ für $i \in \{0, 1, ..., n\}$, weshalb $p(x)$ ein Interpolationspolynom ist.

Es muss jetzt noch gezeigt werden, dass jedes Interpolationspolynom die obige Darstellung besitzt. Der Grad von p ist nach Voraussetzung m. Der Grad von ϕ beträgt $n + 1$. Ist p ein Interpolationspolynom, so gilt $p(x_i) = y_i = \psi(x_i),\ \forall i \in \{0, ..., n\}$. Daraus folgt $p(x_i) - \psi(x_i) = 0,\ \forall i \in \{0, ..., n\}$. Außerdem gilt $\phi(x_i) = 0,\ \forall i \in \{0, ..., n\}$. Somit muss es ein Polynom q mit Grad $m-(n+1) = m-n-1$ geben, sodass $p(x)-\psi(x) = \phi(x) \cdot q(x)$. □

Bemerkung: Man erhält das Polynom q des obigen Satzes durch die Division $(p - \psi) : \phi$.

Nachdem nun allgemeine Eigenschaften bezüglich Interpolationspolynomen behandelt wurden, werden in den nächsten beiden Kapiteln zwei Methoden vorgestellt, durch welche Interpolationspolynome konkret berechnet werden können.

Durch Satz 3.8 und das anschließende Beispiel wurde gezeigt, dass das minimale Interpolationspolynom ψ immer durch Lösen eines linearen Gleichungssystems bestimmt werden kann. Es gibt jedoch auch andere Methoden, um ψ zu ermitteln, welche in den nachfolgenden Abschnitten erläutert werden.

3.2 Lagrange-Interpolation

Die Formel für die Berechnung des minimalen Interpolationspolynoms $\psi \in \mathcal{P}_n$ lautet bei dieser Methode

$$\psi(x) = \sum_{i=0}^{n} y_i \ell_i(x).$$

Dabei sind $\ell_i \in \mathcal{P}_n$ die sogenannten Lagrange-Polynome und es gilt $\ell_i(x_j) = \delta_{ij}$ für $i, j \in \{0, ..., n\}$, wobei

$$\delta_{ij} := \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

das sogenannte Kronecker-Delta ist. Durch diese Forderungen ergibt sich die Interpolationsbedingung $\psi(x_i) = y_i$.

Für die einzelnen Lagrange-Polynome ℓ_i sind durch $\ell_i(x_j) = \delta_{ij}$ die $n+1$ Stützpunkte festgelegt. Bei $x_0, ..., x_{i-1}, x_{i+1}, ..., x_n$ besitzt das Polynom ℓ_i jeweils eine Nullstelle (also Funktionswert 0) und an der Stelle x_i hat es den Funktionswert 1. Nach Satz 3.8 sind die Lagrange-Polynome ℓ_i somit eindeutig bestimmt. Sie besitzen jeweils Höchstgrad n. Für die Berechnung der Lagrange-Polynome ℓ_i kann die nachfolgende Formel verwendet werden.

Lemma 3.11. Die Lagrange-Polynome $\ell_i \in \mathcal{P}_n$ des Lagrangeschen Ansatzes, welche die Bedingung $\ell_i(x_j) = \delta_{ij}$ erfüllen sollen, können durch die Formel

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

berechnet werden.

Beweis: Für $j = i$ gilt

$$\ell_i(x_i) = \prod_{\substack{k=0 \\ k \neq i}}^{n} \underbrace{\frac{x_i - x_k}{x_i - x_k}}_{=1} = \prod_{\substack{k=0 \\ k \neq i}}^{n} 1 = 1^n = 1.$$

Sei nun $j \in \{0, 1, ..., n\} \setminus \{i\}$. Dann gilt

$$\ell_i(x_j) = \prod_{\substack{k=0 \\ k \neq i}}^{n} \frac{x_j - x_k}{x_i - x_k} = \frac{x_j - x_0}{x_i - x_0} \cdots \frac{x_j - x_{j-1}}{x_i - x_{j-1}} \cdot \frac{\overbrace{x_j - x_j}^{=0}}{x_i - x_j} \cdot \frac{x_j - x_{j+1}}{x_i - x_{j+1}} \cdots \frac{x_j - x_n}{x_i - x_n} = 0.$$

Somit erfüllen die Lagrange-Polynome die Bedingung $\ell_i(x_j) = \delta_{ij}$.

□

Vor- und Nachteile: Ein Vorteil dieses Verfahrens ist der Verzicht auf das Lösen eines Gleichungssystems (wie dies auf Seite 22 durchgeführt wurde). Vorteil dieser Interpolationsmethode ist aber insbesondere, dass die Lagrange-Polynome ℓ_i unabhängig von den Stützwerten y_i sind. Hat man somit unveränderliche Stützstellen x_i, so können für beliebige Datensätze $y_0, ..., y_n$ stets dieselben Lagrange-Polynome verwendet werden. Eine mögliche Anwendung wäre die Modellierung physikalischer Zusammenhänge, wenn zu jeweils fixen Zeitpunkten ein Messwert bestimmt wird. Hier sind die von Experiment zu Experiment unverändert bleibenden Stützstellen $x_0, ..., x_n$ vorgegeben und die Datensätze für die Messwerte ändern sich stets.

Ein gravierender Nachteil dieses Verfahrens ist es, dass keine weiteren Stützstellen hinzugenommen werden können, ohne dabei alle Lagrange-Polynome neu berechnen zu müssen.

Dieser Nachteil wird durch die im folgenden Kapitel vorgestellte Methode wegfallen.

Beispiel: Gegeben sind die Stützpunkte $(x_0, y_0) = (0, 2)$, $(x_1, y_1) = (1, 5)$ und $(x_2, y_2) = (3, 7)$. Zunächst werden die drei Lagrange-Polynome berechnet:

$$\begin{aligned}
\ell_0(x) &:= \frac{x-x_1}{x_0-x_1} \cdot \frac{x-x_2}{x_0-x_2} = \frac{x-1}{0-1} \cdot \frac{x-3}{0-3} = \tfrac{1}{3}\,x^2 - \tfrac{4}{3}\,x + 1 \\
\ell_1(x) &:= \frac{x-x_0}{x_1-x_0} \cdot \frac{x-x_2}{x_1-x_2} = \frac{x-0}{1-0} \cdot \frac{x-3}{1-3} = -\tfrac{1}{2}\,x^2 + \tfrac{3}{2}\,x \\
\ell_2(x) &:= \frac{x-x_0}{x_2-x_0} \cdot \frac{x-x_1}{x_2-x_1} = \frac{x-0}{3-0} \cdot \frac{x-1}{3-1} = \tfrac{1}{6}\,x^2 - \tfrac{1}{6}\,x
\end{aligned}$$

Das minimale Interpolationspolynom zu den gegebenen Stützpunkten erhält man nun durch:

$$\psi(x) = \sum_{i=0}^{n} y_i \ell_i(x) = 2 \cdot \left(\tfrac{1}{3}\,x^2 - \tfrac{4}{3}\,x + 1\right) + 5 \cdot \left(-\tfrac{1}{2}\,x^2 + \tfrac{3}{2}\,x\right) + 7 \cdot \left(\tfrac{1}{6}\,x^2 - \tfrac{1}{6}\,x\right)$$

$$\psi(x) = \tfrac{1}{3} \cdot \left(-2x^2 + 11x + 6\right)$$

Durch Einsetzen von x_0, x_1 und x_2 erkennt man, dass dieses Polynom die gegebenen Stützpunkte interpoliert. Sollten sich nun die Stützwerte ändern, so müssen lediglich die entsprechenden Faktoren in der Summe geändert werden. Die Lagrange-Polynome ℓ_0, ℓ_1 und ℓ_2 können jedoch übernommen werden, wodurch der größte Teil des Rechenaufwands wegfällt.

3.3 Newton-Interpolation

Hier wird für das minimale Interpolationspolynom ψ folgender Ansatz verwendet:

$$\psi(x) = \gamma_0 + \gamma_1(x - x_0) + \gamma_2(x - x_0)(x - x_1) + ... + \gamma_n(x - x_0) \cdots (x - x_{n-1})$$

Es bleibt nun zu klären, wie die Koeffizienten $\gamma_0, ..., \gamma_n$ berechnet werden. Dazu muss zunächst der Begriff der dividierten Differenz definiert werden:

Definition 3.12. Seien $x_0, ..., x_n \in \mathbb{R}$ paarweise verschieden und seien $y_0, ..., y_n \in \mathbb{R}$. Dann definiert man laut [2, Seite 2] für $k \in \{0, ..., n\}$ Folgendes:

- $f[x_0] := y_0$
- $f[x_0, x_1] := \frac{f[x_1] - f[x_0]}{x_1 - x_0}$
- $f[x_0, ..., x_k] := \frac{f[x_1, ..., x_k] - f[x_0, ..., x_{k-1}]}{x_k - x_0}$

Man bezeichnet $f[x_0, ..., x_k]$ als die k-te dividierte Differenz des gegebenen Datensatzes.

Das folgende Lemma beschreibt, welche Gestalt die Koeffizienten $\gamma_0, ..., \gamma_n$ haben müssen, damit der Newtonsche Ansatz die Interpolationsbedingungen $\psi(x_i) = y_i$ erfüllt.

Lemma 3.13. Die Koeffizienten $\gamma_0, ..., \gamma_n$ für die Newtonsche Interpolationsformel entsprechen den dividierten Differenzen aus Definition 3.12, d. h. es gilt für alle $k \in \{0, ..., n\}$

$$\gamma_k = f[x_0, ..., x_k].$$

Vor- und Nachteile: Im Vergleich zur Lagrangeschen Interpolationsmethode müssen hier bei Hinzunahme eines zusätzlichen Stützpunktes nicht alle bisherigen Faktoren neu berechnet werden, sondern lediglich ein weiterer Term zum bereits existierenden Polynom addiert werden.

Der große Nachteil dieses Verfahrens ist es jedoch, dass die Faktoren $\gamma_0, ..., \gamma_n$ von den Stützwerten $y_0, ..., y_n$ abhängig sind. Somit müssen diese bei einer Änderung der Stützwerte stets neu berechnet werden.

Beispiel: Gegeben sind wie bereits beim Beispiel auf Seite 25 die Stützpunkte $(x_0, y_0) = (0, 2)$, $(x_1, y_1) = (1, 5)$ und $(x_2, y_2) = (3, 7)$. Im ersten Schritt werden die dividierten Differenzen ermittelt. Dies kann anhand folgenden Schemas durchgeführt werden, wobei die obersten Werte jeder Spalte den γ-Werten entsprechen:

$$f[x_0] := y_0 = 2$$

$$f[x_0, x_1] := \frac{f[x_1] - f[x_0]}{x_1 - x_0} = 3$$

$$f[x_1] := y_1 = 5 \qquad\qquad f[x_0, ..., x_2] := \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0} = -\frac{2}{3}$$

$$f[x_1, x_2] := \frac{f[x_2] - f[x_1]}{x_2 - x_1} = 1$$

$$f[x_2] := y_2 = 7$$

Es ist also $\gamma_0 = 2, \gamma_1 = 3$ und $\gamma_2 = -\frac{2}{3}$. Das Interpolationspolynom erhält man nun folgendermaßen:

$$\psi(x) = \gamma_0 + \gamma_1 \cdot (x - x_0) + \gamma_2 \cdot (x - x_0) \cdot (x - x_1) = 2 + 3 \cdot (x - 0) - \tfrac{2}{3} \cdot (x - 0) \cdot (x - 1)$$

$$\psi(x) = \tfrac{1}{3} \cdot \left(-2x^2 + 11x + 6\right)$$

Man erhält also dasselbe Ergebnis wie bei der Lagrange-Interpolation (nachfolgender Satz wird dies allgemein beweisen). Würde sich hier jedoch ein Stützwert ändern, so würde dies die linke Spalte des oben vorgestellten Schemas beeinflussen und somit in weiterer Folge alle Spalten.

Satz 3.14. Seien $x_0, ..., x_n \in \mathbb{R}$ paarweise verschiedene Stützstellen und $y_0, ..., y_n \in \mathbb{R}$ die zugehörigen Stützwerte. Dann sind die durch das Lagrange-Verfahren und durch das Newton-Verfahren gewonnenen Interpolationspolynome identisch.

Beweis: Aus dem Ansatz der Newton-Interpolation ist leicht ersichtlich, dass bei $n + 1$ Stützpunkten ein Polynom mit Höchstgrad n entsteht. Auch die Lagrange-Polynome ℓ_i haben bei $n+1$ Stützpunkten allesamt den Höchstgrad n, an welchem auch die Multiplikation mit den Stützwerten und die anschließende Summenbildung nichts ändert. Da also beide Interpolationspolynome den minimalen Höchstgrad n besitzen, folgt unmittelbar aus Satz

3.8, also der Eindeutigkeit des minimalen Interpolationspolynoms, dass die Polynome gleich sind. □

Fazit: Hat man einen vorgegebenen bzw. unveränderlichen Satz an Stützstellen $x_0, ..., x_n$ für welchen immer neue Stützwerte gemessen bzw. ermittelt werden, so ist es sinnvoller, das Lagrange-Verfahren zu wählen, da hier die Lagrange-Polynome ℓ_i nur einmal berechnet werden müssen und anschließend übernommen werden können, während die Faktoren γ_i des Newton-Verfahrens stets neu berechnet werden müssten.

Wird jedoch eine bestehende Menge an Stützpunkten $(x_0, y_0), ..., (x_n, y_n)$ laufend erweitert, so ist es sinnvoll, das Newton-Verfahren zu benutzen, da man in diesem Fall wie bereits erwähnt, das Interpolationspolynom nur um einen weiteren Term ergänzen muss, während man beim Lagrange-Verfahren alle Lagrange-Polynome ℓ_i neu berechnen müsste.

3.4 Fehlerabschätzung

Dieser Abschnitt widmet sich der Abschätzung der Abweichung des minimalen Interpolationspolynoms ψ_f von der dadurch interpolierten Funktion f. Bedeutung findet diese Abschätzung u. a. bei der numerischen Integration, also dem Hauptthema dieser Arbeit. Die Abweichung wird allgemein als Restglied bezeichnet.

Definition 3.15. Sei $f : [a, b] \to \mathbb{R}$ eine stetige Funktion und seien $x_0, ..., x_n \in [a, b]$ paarweise verschiedene Stützstellen. Dann definiert man $R_n f := f - \psi_f$ als das Restglied der Polynominterpolation.

Für $(n + 1)$-mal stetig differenzierbare Funktionen ergibt sich die nachfolgend behandelte Berechnungsformel für das Restglied $R_n f$.

Satz 3.16. Sei $f : [a, b] \to \mathbb{R}$ eine $(n + 1)$-mal stetig differenzierbare Funktion, und seien $x_1, ..., x_r \in [a, b]$ paarweise verschiedene Stützstellen mit den Vielfachheiten $k_1, ..., k_r \in \mathbb{N}$, sodass $\sum_{j=1}^{r} k_j = n + 1$ gilt. Sei $\varphi \in \mathcal{P}_n$ ein Polynom mit Höchstgrad n, welches für alle $j \in \{1, ..., r\}$ die Bedingung $\varphi(x_j) = f(x_j)$ erfüllt. Dann gilt für das Restglied $R_n f(x) := f(x) - \varphi(x)$, dass für alle $x \in [a, b]$ ein $\xi \in [a, b]$ existiert, sodass gilt:

$$R_n f(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \prod_{j=1}^{r} (x - x_j)^{k_j}$$

Beweis: Als Grundlage für diesen Beweis wurde [17, Satz 9.10] verwendet. Sei $x \in [a, b]$. Es sind nun zwei Fälle zu unterscheiden.

1. Fall: Die Stelle x ist selbst eine Stützstelle, d. h. es existiert ein $i \in \{1, ..., r\}$ mit $x = x_i$. Setzt man in die zu beweisende Formel ein, so erhält man

$$R_n f(x_i) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \underbrace{\prod_{j=1}^{r} (x_i - x_k)^{k_j}}_{=0} = 0.$$

Da für $j \in \{1, ..., r\}$ gilt $\varphi(x_j) = f(x_j)$, stimmt dies mit der Definition $R_n f(x_j) :=$

$f(x_j) - \varphi(x_j) = 0$ überein, weshalb in diesem Fall die Aussage für jedes beliebige $\xi \in [a, b]$ bestätigt wurde.

2. Fall: Sei nun x keine Stützstelle. Dann definiert man α als

$$\alpha(x) := \prod_{j=1}^{r} (x - x_j)^{k_j}.$$

Es handelt sich dabei um ein normiertes Polynom mit Grad $n + 1$. Außerdem wird für ein beliebiges aber festes $x \in [a, b]$ mit $x \neq x_i$, $\forall i \in \{1, ..., r\}$ die Hilfsfunktion $g : [a, b] \to \mathbb{R}$ folgendermaßen definiert:

$$g(y) := f(y) - \varphi(y) - \alpha(y) \cdot \frac{f(x) - \varphi(x)}{\alpha(x)}$$

Nach Voraussetzung ist $f(y) \in C^{n+1}[a, b]$ und wegen Satz 2.7 sind auch die Polynome $\varphi(y), \alpha(y) \in C^{n+1}[a, b]$. Außerdem ist der Bruch eine Konstante (weil x fest gewählt wurde), und der Nenner des Bruches ist ungleich null, da x keine Stützstelle sein darf. Somit ist $g \in C^{n+1}[a, b]$. Man erhält

$$g^{(n+1)}(y) = f^{(n+1)}(y) - \underbrace{0}_{\text{Kor. 2.6}} - \underbrace{(n+1)!}_{\text{Lemma 2.5}} \cdot \overbrace{\frac{f(x) - \varphi(x)}{\alpha(x)}}^{=R_n f(x)} = f^{(n+1)}(y) - (n+1)! \cdot \frac{R_n f(x)}{\alpha(x)}.$$

Des Weiteren besitzt g die $n + 2$ Nullstellen $x, x_1^{(k_1)}, ..., x_r^{(k_r)}$, wobei die Vielfachheiten in hochgestellter Klammer angeführt werden, denn für $y = x$ gilt

$$g(y) = \underbrace{f(y)}_{=f(x)} - \underbrace{\varphi(y)}_{=\varphi(x)} - \underbrace{\alpha(y)}_{=\alpha(x)} \cdot \frac{f(x) - \varphi(x)}{\alpha(x)} = 0,$$

und für $y \in \{x_1, ..., x_r\}$ erhält man wegen $\alpha(x_i) = 0$ und der Interpolationsbedingung $f(x_i) = \varphi(x_i)$

$$g(y) = \underbrace{f(y) - \varphi(y)}_{=0} - \underbrace{\alpha(y)}_{=0} \cdot \frac{f(x) - \varphi(x)}{\alpha(x)} = 0.$$

Gemäß Satz 2.21 besitzt $g^{(n+1)}$ höchstens $n + 1$ Nullstellen weniger als g. Da g laut den vorherigen Überlegungen $n + 2$ Nullstellen aufweist, hat $g^{(n+1)}$ daher mindestens eine Nullstelle. Das heißt, es gibt ein $\xi \in [a, b]$ mit $g^{(n+1)}(\xi) = 0$.

Somit erhält man schließlich

$$0 = g^{(n+1)}(\xi) = f^{(n+1)}(\xi) - (n+1)! \cdot \frac{R_n f(x)}{\alpha(x)} \Rightarrow R_n f(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \cdot \prod_{j=1}^{r} (x - x_j)^{k_j}.$$

□

Bemerkung: Die Stelle $\xi \in [a, b]$ ist von der Stelle x, an welcher die Abweichung bestimmt werden soll, abhängig, also $\xi(x)$.

Im Allgemeinen ist die Stelle ξ mühsam zu bestimmen, weshalb man oft folgende Abschätzung für den Höchstfehler verwendet.

Korollar 3.17. Für das Restglied der Polynominterpolation gilt die Abschätzung

$$|R_n f(x)| \leq \frac{1}{(n+1)!} \cdot \max_{x \in [a,b]} \left(|f^{(n+1)}(x)| \cdot \prod_{j=1}^{r} |(x - x_j)^{k_j}| \right).$$

Beweis: Dies folgt sofort aus Satz 3.16. □

3.5 Algorithmus von Aitken und Neville

In manchen Fällen möchte man das Interpolationspolynom nicht explizit ermitteln sondern nur an einer bestimmten Stelle auswerten. Eine Möglichkeit, den Funktionswert rekursiv zu bestimmen, stellt der Algorithmus von Aitken und Neville dar – benannt nach dem neuseeländischen Mathematiker Alexander Aitken (1895 – 1967) und dem englischen Mathematiker Eric Harold Neville (1889 – 1961).

Satz 3.18. (Lemma von Aitken): Seien $x_0, ..., x_n \in \mathbb{R}$ paarweise verschiedene Stützstellen mit den zugehörigen Werten $y_0, ..., y_n \in \mathbb{R}$. Sei nun $0 \leq j \leq k \leq n$, und sei $p_{k,j}$ das eindeutig bestimmte Interpolationspolynom mit Höchstgrad j zu den insgesamt $j+1$ Stützpunkten $(x_{k-j}, y_{k-j}), ..., (x_k, y_k)$. Dann kann für ein beliebiges aber festes $x \in \mathbb{R}$ der Funktionswert $p_{k,j}(x)$ durch folgende Rekursion berechnet werden:

(i) $p_{k,0}(x) := y_k$

(ii) $p_{k,j}(x) := \frac{(x_{k-j} - x) \cdot p_{k,j-1}(x) - (x_k - x) \cdot p_{k-1,j-1}(x)}{x_{k-j} - x_k}$

Insbesondere gilt für das Interpolationspolynom $p_{n,n}$ aller $n+1$ Stützstellen:

$$p_{n,n}(x) = \frac{(x_0 - x) \cdot p_{n,n-1}(x) - (x_n - x) \cdot p_{n-1,n-1}(x)}{x_0 - x_n}$$

Beweis: Der Beweis wird mit Hilfe von [11, Seite 54 und 55] geführt. Es handelt sich um einen Induktionsbeweis. Für den Induktionsanfang $j = 0$ sind $p_{k,j} = p_{k,0}$ konstante Funktionen, welche wegen $p_{k,0}(x_k) = y_k$ jeweils den Punkt (x_k, y_k) interpolieren und somit die geforderte Interpolationseigenschaft erfüllen.

Es wird nun der Induktionsschritt $j - 1 \mapsto j$ betrachtet: Das Polynom $p_{k,j-1}$ interpoliert nach Induktionsvoraussetzung die Punkte

$$(x_{k-(j-1)}, y_{k-(j-1)}), ..., (x_k, y_k) = (x_{k-j+1}, y_{k-j+1}), ..., (x_k, y_k)$$

und das Polynom $p_{k-1,j-1}$ interpoliert nach Induktionsvoraussetzung die Punkte

$$(x_{(k-1)-(j-1)}, y_{(k-1)-(j-1)}), ..., (x_{k-1}, y_{k-1}) = (x_{k-j}, y_{k-j}), ..., (x_{k-1}, y_{k-1}).$$

Somit sind die drei Fälle $x = x_{k-j}$, $x \in \{x_{k-j+1}, ..., x_{k-1}\}$ und $x = x_k$ zu untersuchen:

1. Fall: Es sei $x = x_{k-j}$. Dann gilt Folgendes:

$$p_{k,j}(x_{k-j}) = \frac{\overbrace{(x_{k-j} - x_{k-j})}^{=0} \cdot p_{k,j-1}(x_{k-j}) - (x_k - x_{k-j}) \cdot \overbrace{p_{k-1,j-1}(x_{k-j})}^{=y_{k-j}}}{x_{k-j} - x_k} = y_{k-j}$$

2. Fall: Es sei $x = x_i \in \{x_{k-j+1}, ..., x_{k-1}\}$. Dann gilt Folgendes:

$$p_{k,j}(x_i) = \frac{(x_{k-j} - x_i) \cdot \overbrace{p_{k,j-1}(x_i)}^{=y_i} - (x_k - x_i) \cdot \overbrace{p_{k-1,j-1}(x_i)}^{=y_i}}{x_{k-j} - x_k} =$$

$$= \frac{(x_{k-j} - x_i) \cdot y_i - (x_k - x_i) \cdot y_i}{x_{k-j} - x_k} = \frac{(x_{k-j} - x_k) \cdot y_i}{x_{k-j} - x_k} = y_i$$

3. Fall: Es sei $x = x_k$. Dann gilt Folgendes:

$$p_{k,j}(x_k) = \frac{(x_{k-j} - x_k) \cdot \overbrace{p_{k,j-1}(x_k)}^{=y_k} - \overbrace{(x_k - x_k)}^{=0} \cdot p_{k-1,j-1}(x_k)}{x_{k-j} - x_k} = y_k$$

Somit erfüllt $p_{k,j}$ die Interpolationseigenschaft $p_{k,j}(x_i) = y_i$, $\forall i \in \{k-j, k\}$. □

Beispiel: Gegeben seien erneut die drei Stützpunkte $(x_0, y_0) = (0, 2)$, $(x_1, y_1) = (1, 5)$ und $(x_2, y_2) = (3, 7)$, welche bereits auf Seite 25 und auf Seite 26 verwendet wurden, und von denen bereits bekannt ist, dass das minimale Interpolationspolynom $\psi(x) = \frac{1}{3} \cdot \left(-2x^2 + 11x + 6\right)$ lautet.

Es soll mithilfe des Algorithmus von Aitken und Neville der Wert dieses Interpolationspolynoms an der Stelle $x = 2$ ermittelt werden, ohne das Polynom dabei explizit zu bestimmen. Dazu werden schrittweise die Polynome $p_{k,j}$ aus Satz 3.18 an der Stelle $x = 2$ ausgewertet. Man erhält:

$$p_{0,0}(2) = y_0 = 2$$
$$p_{1,0}(2) = y_1 = 5$$
$$p_{2,0}(2) = y_2 = 7$$

$$p_{1,1}(2) = \frac{(x_0 - 2) \cdot p_{1,0}(2) - (x_1 - 2) \cdot p_{0,0}(x)}{x_0 - x_1} = \frac{-2 \cdot 5 - (-1) \cdot 2}{-1} = 8$$
$$p_{2,1}(2) = \frac{(x_1 - 2) \cdot p_{2,0}(2) - (x_2 - 2) \cdot p_{1,0}(x)}{x_1 - x_2} = \frac{-1 \cdot 7 - 1 \cdot 5}{-2} = 6$$

$$p_{2,2}(2) = \frac{(x_0 - 2) \cdot p_{2,1}(2) - (x_2 - 2) \cdot p_{1,1}(x)}{x_0 - x_2} = \frac{-2 \cdot 6 - 1 \cdot 8}{-3} = \frac{20}{3}$$

Setzt man in den aus den vorherigen Beispielen bereits bekannten Funktionsterm von ψ ein, so erhält man ebenfalls den Wert $\psi(2) = \frac{20}{3}$.

4 Bernoulli-Polynome

Dieses Kapitel behandelt die nach dem Schweizer Mathematiker Jakob I. Bernoulli (1655–1705)[1] benannten Bernoulli-Polynome. Diese speziellen Polynome spielen im Rahmen dieser Arbeit beim Romberg-Verfahren (Kapitel 8) eine große Rolle. Hauptquelle für dieses Kapitel ist [30], weshalb diese Quelle bei den einzelnen Beweisen nicht jedes Mal angeführt wird.

Definition 4.1. Sei $n \in \mathbb{N}_0$. Dann wird das Bernoulli-Polynom $B_n : [0,1] \to \mathbb{R}$ nach [22, Seite 4] und [30] durch die folgenden Eigenschaften rekursiv definiert:

(i) $B_0(x) := 1$

(ii) $B_n'(x) = n \cdot B_{n-1}(x), \ \forall n \geq 1$

(iii) $\int_0^1 B_n(x)\,dx = 0, \ \forall n \geq 1$

Definition 4.2. Für $n \in \mathbb{N}_0$ werden die Zahlen $B_n := B_n(0)$ als Bernoulli-Zahlen bezeichnet.

Lemma 4.3. Sei $n \geq 1$. Dann gilt

$$B_n(x) = B_n + n \cdot \int_0^x B_{n-1}(t)\,dt.$$

Beweis: Nach Definition 4.1 (ii) gilt $B_n'(t) = n \cdot B_{n-1}(t)$ für $n \geq 1$. Integriert man beide Seiten im Intervall $[0, x]$, so erhält man

$$\int_0^x B_n'(t)\,dt = n \cdot \int_0^x B_{n-1}(t)\,dt.$$

Aus dem Hauptsatz der Differential- und Integralrechnung erhält man auf der linken Seite

$$B_n(t)\big|_0^x = B_n(x) - B_n(0) = n \cdot \int_0^x B_{n-1}(t)\,dt.$$

Daraus folgt schließlich unter Verwendung von Definition 4.2

$$B_n(x) = B_n(0) + n \cdot \int_0^x B_{n-1}(t)\,dt = B_n + n \cdot \int_0^x B_{n-1}(t)\,dt.$$

□

Lemma 4.4. Für $n \neq 1$ gilt $B_n(0) = B_n(1)$.

Beweis: Für $n = 0$ folgt diese Aussage sofort aus Definition 4.1 (i), denn $B_0(0) = 1 = B_0(1)$.

[1] nach gregorianischem Kalender

Sei nun $n \geq 2$. Dann gilt wegen Lemma 4.3 und Definition 4.1 (iii)

$$B_n(1) = \underbrace{B_n}_{=B_n(0)} + n \cdot \underbrace{\int_0^1 B_{n-1}(t)\,dt}_{=0} = B_n(0).$$

□

Satz 4.5. Für $n \in \mathbb{N}_0$ gilt

$$B_n(x) = \sum_{k=0}^{n} \binom{n}{k} \cdot B_k \cdot x^{n-k}.$$

Beweis: Es wird ein Induktionsbeweis geführt.

Für den Induktionsanfang $n = 0$ gilt nach Definition 4.1 (i) $B_0(x) = 1$. Andererseits ist

$$\sum_{k=0}^{0} \binom{0}{k} \cdot B_k \cdot x^{0-k} = \underbrace{\binom{0}{0}}_{=1} \cdot \underbrace{B_0}_{=1} \cdot \underbrace{x^0}_{=1} = 1,$$

wodurch die Gleichung erfüllt ist.

Als Induktionsvoraussetzung (IV) sei für ein beliebiges aber festes $n \in \mathbb{N}$ die Eigenschaft

$$B_{n-1}(x) = \sum_{k=0}^{n-1} \binom{n-1}{k} \cdot B_k \cdot x^{n-1-k}$$

erfüllt. Für den Induktionsschritt $n - 1 \mapsto n$ erhält man somit nach Lemma 4.3:

$$B_n(x) = B_n + n \cdot \int_0^x B_{n-1}(t)\,dt \overset{\text{IV}}{=} B_n + n \cdot \int_0^x \sum_{k=0}^{n-1} \binom{n-1}{k} \cdot B_k \cdot t^{n-1-k}\,dt$$

Da endliche Summen und Integrale vertauscht werden können, erhält man durch die Nebenrechnung

$$\int_0^x t^{n-1-k}\,dt = \left.\frac{t^{n-k}}{n-k}\right|_0^x = \frac{x^{n-k}}{n-k}$$

folgendes Resultat:

$$B_n + n \cdot \sum_{k=0}^{n-1} \left[\binom{n-1}{k} \cdot B_k \cdot \int_0^x t^{n-1-k}\,dt \right] = B_n + \sum_{k=0}^{n-1} n \cdot \binom{n-1}{k} \cdot \frac{1}{n-k} \cdot B_k \cdot x^{n-k}$$

Aus der Definition $\binom{n}{k} := \frac{n!}{k! \cdot (n-k)!}$ des Binomialkoeffizienten folgt

$$n \cdot \binom{n-1}{k} \frac{1}{n-k} = \frac{\overbrace{n \cdot (n-1)!}^{n!}}{k! \cdot \underbrace{(n-1-k)! \cdot (n-k)}_{=(n-k-1)! \cdot (n-k) = (n-k)!}} = \frac{n!}{k! \cdot (n-k)!} = \binom{n}{k}.$$

Mit $B_n = \binom{n}{n} \cdot B_n \cdot x^{n-n}$ erhält man schließlich

$$B_n + \sum_{k=0}^{n-1} \binom{n}{k} \cdot B_k \cdot x^{n-k} = \sum_{k=0}^{n} \binom{n}{k} \cdot B_k \cdot x^{n-k}.$$

Damit ist die zu zeigende Eigenschaft für alle $n \in \mathbb{N}_0$ beweisen. □

Korollar 4.6. Die Bernoulli-Polynome sind normiert, d. h. für $n \in \mathbb{N}_0$ erfüllt das n-te Bernoulli-Polynom in der Darstellung $B_n(x) = a_n x^n + ... + a_1 x + a_0$ die Eigenschaft $a_n = 1$.

Beweis: Aus Satz 4.5 erhält man

$$B_n(x) = \sum_{k=0}^{n} \binom{n}{k} \cdot B_k \cdot x^{n-k} = \underbrace{\binom{n}{0}}_{=1} \cdot \underbrace{B_0}_{=1} \cdot x^{n-0} + \sum_{k=1}^{n} \binom{n}{k} \cdot B_k \cdot x^{n-k} =$$

$$= 1 \cdot x^n + \sum_{k=1}^{n} \binom{n}{k} \cdot B_k \cdot x^{n-k}.$$

Da die Exponenten der übrigen Summanden wegen x^{n-k} kleiner als n sind, ist die Aussage dadurch bewiesen. □

Nun wird eine rekursive Formel zur Berechnung der n-ten Bernoulli-Zahl B_n hergeleitet.

Lemma 4.7. Für $n \in \mathbb{N}, n \geq 2$ gilt

$$B_n = \sum_{k=0}^{n} \binom{n}{k} B_k.$$

Beweis: Dies folgt sofort aus Satz 4.5, Lemma 4.4 und der Definition der Bernoulli-Zahlen:

$$B_n := B_n(0) \overset{4.4}{=} B_n(1) \overset{4.5}{=} \sum_{k=0}^{n} \binom{n}{k} \cdot B_k \cdot \underbrace{1^{n-k}}_{=1} = \sum_{k=0}^{n} \binom{n}{k} \cdot B_k.$$

□

Lemma 4.8. Für $n \geq 1$ gilt

$$B_n = -\frac{1}{n+1} \sum_{k=0}^{n-1} \binom{n+1}{k} B_k.$$

Beweis: Aus Lemma 4.7 folgt für $n \geq 1$

$$B_{n+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} B_k = \sum_{k=0}^{n-1} \binom{n+1}{k} B_k + \underbrace{(n+1)}_{=\binom{n+1}{n}} \cdot B_n + \underbrace{1}_{=\binom{n+1}{n+1}} \cdot B_{n+1}.$$

Durch Umformen dieser Gleichung nach B_n folgt daraus die zu zeigende rekursive Formel für B_n. □

Die Bernoulli-Zahl B_0 ist durch Definition 4.1 (i) festgelegt, denn es gilt $B_0 = B_0(0) = 1$. Die Bernoulli-Zahl B_1 erhält man durch nachfolgendes Lemma.

Lemma 4.9. Die Bernoulli-Zahl B_1 lautet $-\frac{1}{2}$.

Beweis: Nach Definition 4.1 (ii) gilt

$$B_1'(x) = 1 \cdot \underbrace{B_0(x)}_{=1} = 1,$$

und somit folgt für ein $c \in \mathbb{R}$

$$B_1(x) = \int_0^x 1 \, dt = x + c.$$

Aus Definition 4.1 (iii) erhält man nun

$$\int_0^1 B_1(x) \, dx = 0 \Rightarrow \int_0^1 (x + c) \, dx = 0 \Rightarrow \int_0^1 x \, dx + \int_0^1 c \, dx = 0$$

$$\Rightarrow \frac{x^2}{2}\Big|_0^1 + c \cdot 1 = 0 \Rightarrow \frac{1}{2} - \frac{0}{2} + c = 0 \Rightarrow c = -\frac{1}{2}.$$

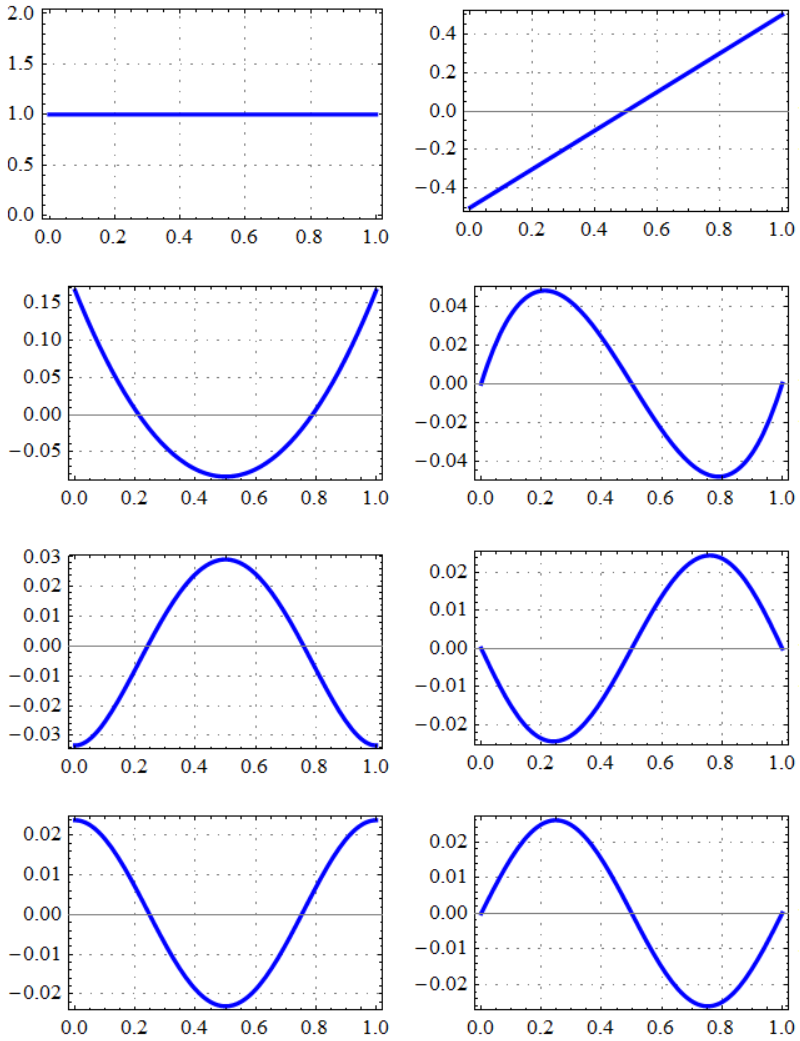
□

Damit sind durch Definition 4.1 (i), Lemma 4.9 und Lemma 4.8 die Bernoulli-Zahlen B_n für alle $n \in \mathbb{N}_0$ eindeutig bestimmt. Wegen Satz 4.5 sind daher auch alle Bernoulli-Polynome $B_n(x)$ eindeutig festgelegt. Wie man aus Definition 4.1 (i), Lemma 4.9 und Lemma 4.8 außerdem sehen kann, gilt für alle Bernoulli-Zahlen $B_n \in \mathbb{Q}$.

Nachfolgende Tabelle gibt einen Überblick über die Bernoulli-Polynome $B_n(x)$ und die Bernoulli-Zahlen B_n für $n \leq 5$.

n	B_n	$B_n(x)$
0	1	1
1	$-\frac{1}{2}$	$x - \frac{1}{2}$
2	$\frac{1}{6}$	$x^2 - x + \frac{1}{6}$
3	0	$x^3 - \frac{3}{2}x^2 + \frac{x}{2}$
4	$-\frac{1}{30}$	$x^4 - 2x^3 + x^2 - \frac{1}{30}$
5	0	$x^5 - \frac{5}{2}x^4 + \frac{5}{3}x^3 - \frac{x}{6}$

In nachfolgender Abbildung sind die ersten acht Bernoulli-Polynome zu sehen, wobei in der linken Spalte die geraden (also von oben nach unten B_0, B_2, B_4 und B_6) und in der rechten Spalte die ungeraden (also von oben nach unten B_1, B_3, B_5 und B_7) abgebildet sind.



Man erkennt anhand der Funktionsgraphen, dass die Bernoulli-Polynome an der Stelle $x = \frac{1}{2}$ Symmetrieeigenschaften aufweisen: Für gerade n sind sie gerade, für ungerade n sind sie ungerade. Nachfolgender Satz (Satz 4.10) wird dies allgemein zeigen. Außerdem erkennt man, dass für ungerade n bei $x = \frac{1}{2}$ eine Nullstelle vorliegt (siehe Lemma 4.12) und ab $n = 3$ für ungerade n zusätzlich bei $x = 0$ und $x = 1$ zwei weitere Nullstellen existieren (siehe Lemma 4.11).

Satz 4.10. Sei $n \in \mathbb{N}_0$ und sei $P_n(x) := B_n(x + \frac{1}{2})$. Dann gilt:

(i) Falls n gerade ist, ist $P_n(x)$ eine gerade Funktion.

(ii) Falls n ungerade ist, ist $P_n(x)$ eine ungerade Funktion.

Beweis: Als Grundlage für diesen Beweis wurde [22, Seite 4 und 5] herangezogen. Es wird ein Induktionsbeweis erbracht. Für $n = 0$ erhält man mit $B_0(x) = 1$

$$P_0(x) = B_0\left(x + \frac{1}{2}\right) = 1,$$

was wegen $P_0(-x) = 1 = P_0(x)$ eine gerade Funktion darstellt. Für $P_1(x)$ erhält man mit

$B_1(x) = x - \frac{1}{2}$

$$P_1(x) = B_1\left(x + \frac{1}{2}\right) = \left(x + \frac{1}{2}\right) - \frac{1}{2} = x,$$

was wegen $P_1(-x) = -x = -P_1(x)$ eine ungerade Funktion ist.

Sei nun als Induktionsvoraussetzung $P_{2n}(x)$ für ein beliebiges aber festes $n \geq 0$ eine gerade Funktion. Als Induktionsschritt wird nun $2n \mapsto 2n+1$ betrachtet. Wegen Definition 4.1 (ii) gilt

$$P'_{2n+1}(x) = B'_{2n+1}\left(x + \frac{1}{2}\right) \overset{\text{(ii)}}{=} (2n+1) \cdot B_{2n}\left(x + \frac{1}{2}\right) = (2n+1) \cdot P_{2n}(x). \tag{E1}$$

Aus Definition 4.1 (iii) folgt außerdem

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} P_{2n+1}(x)\,dx = \int_0^1 P_{2n+1}\left(x - \frac{1}{2}\right)\,dx = \int_0^1 B_{2n+1}(x)\,dx \overset{\text{(iii)}}{=} 0. \tag{E2}$$

Für die Funktion P_{2n+1} müssen also wegen Definition 4.1 (ii) und (iii) die Eigenschaften E1 und E2 erfüllt sein. Es wird nun $q(x) := (2n+1) \cdot \int_0^x P_{2n}(t)\,dt$ definiert. Die Ableitung ist $q'(x) = (2n+1) \cdot P_{2n}(x)$. Nach Induktionsvoraussetzung ist P_{2n} gerade, weshalb nach Lemma 2.23 (i) die Funktion $\int_0^x P_{2n}(x)\,dx$ ungerade ist, und damit ist auch q ungerade. Deshalb gilt wiederum nach Lemma 2.24 (i)

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} q(x)\,dx = (2n+1) \cdot \underbrace{\int_{-\frac{1}{2}}^{\frac{1}{2}} P_{2n}(x)\,dx}_{=0} = 0.$$

Somit erfüllt q die zuvor für P_{2n+1} geforderten Eigenschaften E1 und E2. Man kann also $P_{2n+1}(x) = q(x) = (2n+1) \cdot \int_0^x P_{2n}(t)\,dt$ setzen. Weil bereits gezeigt wurde, dass q ungerade ist, ist auch P_{2n+1} ungerade.

Es gilt wegen Lemma 4.3 $P_{2n+2}(x) = c + (2n+2) \cdot \int_0^x P_{2n+1}(t)\,dt$ für ein $c \in \mathbb{R}$. Weil P_{2n+1} ungerade ist, folgt aus Lemma 2.23 (ii), dass $\int_0^x P_{2n+1}(t)\,dt$ ein gerades Polynom ist. Somit ist P_{2n+2} gerade. □

Lemma 4.11. Für ungerade $n \geq 3$ gilt $B_n(0) = 0 = B_n(1)$ und $B_n = 0$.

Beweis: Nach Lemma 4.4 gilt für $n \geq 2$ die Eigenschaft $B_n(0) = B_n(1)$. Außerdem gilt für alle ungeraden n wegen Satz 4.10

$$B_n(0) = P_n\left(-\frac{1}{2}\right) = -P_n\left(\frac{1}{2}\right) = -B_n(1).$$

Aus $B_n(0) = B_n(1)$ und $B_n(0) = -B_n(1)$ folgt für ungerade $n \geq 3$ schließlich $B_n(0) = B_n(1) = 0$. Aus der Definition der Bernoulli-Zahlen folgt, dass auch $B_n = 0$ ist. □

Lemma 4.12. Für ungerade n gilt $B_n(\frac{1}{2}) = 0$.

Beweis: Nach Satz 4.10 ist $P_n(x) := B_n(x + \frac{1}{2})$ für ungerade n eine ungerade Funktion. Wegen Lemma 2.25 gilt somit für $P_n(0) = B_n(0 + \frac{1}{2}) = B_n(\frac{1}{2}) = 0$. □

Satz 4.13. Sei $n \geq 1$. Dann gelten folgende Eigenschaften:

(i) $(-1)^n B_{2n-1}(x) > 0$ für $x \in (0, \frac{1}{2})$

(ii) $(-1)^n (B_{2n}(x) - B_{2n}) \geq 0$ für $x \in (0, 1)$

(iii) $(-1)^{n+1} B_{2n} > 0$

Beweis: Für diesen Beweis wurde [1, Seite 463] verwendet. Es wird ein Induktionsbeweis für Aussage (i) geführt. Zusätzlich wird gezeigt, dass die Aussagen (ii) und (iii) für ein beliebiges aber festes $n \geq 1$ richtig sind, sofern Aussage (i) für dieses n richtig ist.

Der Induktionsanfang ist erfüllt, da man für $n = 1$ Folgendes erhält:

$$(-1)^1 \cdot B_1(x) = -(x - \tfrac{1}{2}) = \tfrac{1}{2} - x > 0, \ \forall x \in (0, \tfrac{1}{2})$$

Sei nun als Induktionsvoraussetzung (IV) Aussage (i) für ein beliebiges aber festes $n \geq 1$ richtig. Unter Verwendung von Lemma 4.3 erhält man

$$(-1)^n (B_{2n}(x) - B_{2n}) = (-1)^n \left(B_{2n} + 2n \int_0^x B_{2n-1}(t)\, dt - B_{2n} \right) =$$

$$= 2n \int_0^x \underbrace{(-1)^n \cdot B_{2n-1}(t)}_{\overset{\text{IV}}{>} 0, \ \forall t \in \left(0, \frac{1}{2}\right)} dt > 0, \ \forall x \in (0, \tfrac{1}{2})$$

Da nach Satz 4.10 (i) $B_{2n}(x)$ symmetrisch um $x = \frac{1}{2}$ ist, gilt die Ungleichung für $x \in (\frac{1}{2}, 1)$. Aufgrund der Stetigkeit von Polynomen kann dieser Term somit auch an der Stelle $x = \frac{1}{2}$ nicht negativ sein. Daher ist Aussage (ii) gezeigt.

Wegen Aussage (ii) ist $(-1)^n (B_{2n}(x) - B_{2n}) \geq 0$ für $x \in [0, 1]$, wobei die Gleichheit nur an den Stellen $0, \frac{1}{2}$ und 1 möglich ist. Somit gilt

$$0 < \int_0^1 (-1)^n (B_{2n}(x) - B_{2n})\, dx = (-1)^n \underbrace{\int_0^1 B_{2n}(x)\, dx}_{\overset{4.1\,(iii)}{=} 0} - (-1)^n \underbrace{\int_0^1 B_{2n}\, dx}_{= B_{2n}} =$$

$$= (-1) \cdot (-1)^n B_{2m} = (-1)^{n+1} B_{2n}.$$

Damit wäre auch Aussage (iii) gezeigt.

Für den Induktionsschritt $n \mapsto n + 1$ verwendet man die aus Lemma 4.11 und Lemma 4.12 folgende Eigenschaft $B_{2n+1}(0) = B_{2n+1}(\frac{1}{2}) = B_{2n+1}(1) = 0$.

Angenommen, es hätte $B_{2n+1}(x)$ im Intervall $(0, \frac{1}{2})$ eine weitere Nullstelle. Weil $B_{2n+1}(x)$ glatt ist, müsste wegen Lemma 2.22 auch $B''_{2n+1}(x)$ eine Nullstelle $x_0 \in (0, \frac{1}{2})$ besitzen.

Nach Definition 4.1 (ii) gilt

$$B'_{2n+1}(x) = (2n + 1) B_{2n}(x)$$

und damit

$$B''_{2n+1}(x) = (2n + 1) B'_{2n}(x) = (2n + 1) \cdot 2n \cdot B_{2n-1}(x).$$

Aus $B''_{2n+1}(x_0) = 0$ folgt somit $B_{2n-1}(x_0) = 0$, was einen Widerspruch zur Induktionsvor-

aussetzung darstellt. Daher besitzt $B_{2n+1}(x)$ auf $(0, \frac{1}{2})$ keinen Vorzeichenwechsel.

Das Vorzeichen von $B_{2n+1}(x)$ auf $(0, \frac{1}{2})$ entspricht wegen $B_{2n+1}(0) = 0$ dem Vorzeichen von $B'_{2n+1}(0)$, was nach Definition 4.1 (ii) gleich ist mit $(2n+1)B_{2n}(0)$ und nach Definition 4.2 gleich ist mit $(2n+1)B_{2n}$. Weil Aussage (iii) gleichbedeutend damit ist, dass $(-1)^{n+1}$ und B_{2n} dasselbe Vorzeichen besitzen, entspricht das Vorzeichen von $B_{2n+1}(x)$ auf $(0, \frac{1}{2})$ dem Vorzeichen von $(-1)^{n+1}$. Daraus folgt schließlich:

$$(-1)^{n+1} B_{2n+1}(x) > 0 \Leftrightarrow \underbrace{(-1)^{n+1} \cdot (-1)^{n+1}}_{=(-1)^{2(n+1)}=1} = 1 > 0$$

□

5 Orthogonale Polynome

Dieses Kapitel dient als Vorbereitung für das in Kapitel 9 behandelte Gauß-Verfahren. Es werden zunächst einige allgemeine Definitionen und grundlegende Eigenschaften behandelt. Anschließend wird eine Rekursionsformel zur Generierung orthogonaler Polynome vorgestellt und bewiesen. Den Abschluss des Kapitels bildet ein Überblick über historisch bedeutsame Folgen von orthogonalen Polynomen, die auch im späteren Verlauf dieser Arbeit zum Einsatz kommen werden.

5.1 Vektorräume, Basen und Skalarprodukte

Definition 5.1. Es sei V eine Menge, $(K, +, \cdot)$ ein Körper, $\oplus : V \times V \to V$ die sogenannte Vektoraddition und $\odot : K \times V \to V$ die sogenannte Skalarmultiplikation. Dann nennt man $(V, \oplus, \odot)$ Vektorraum bezüglich K, falls folgende Eigenschaften erfüllt sind:

(i) $(x \oplus y) \oplus z = x \oplus (y \oplus z),\ \forall x, y, z \in V$ (Assoziativität der Vektoraddition)

(ii) $\exists 0_V \in V : 0_V \oplus x = x \oplus 0_V = x,\ \forall x \in V$ (neutrales Element)

(iii) $\forall x \in V : \exists - x \in V : x \oplus (-x) = (-x) \oplus x = 0_V$ (Existenz inverser Elemente)

(iv) $x \oplus y = y \oplus x,\ \forall x, y \in V$ (Kommutativität der Vektoraddition)

(v) $a \odot (x \oplus y) = (a \odot x) \oplus (a \odot y),\ \forall x, y \in V,\ \forall a \in K$ (Distributivgesetz 1)

(vi) $(a + b) \odot x = (a \odot x) \oplus (b \odot x),\ \forall x \in V,\ \forall a, b \in K$ (Distributivgesetz 2)

(vii) $(a \cdot b) \odot x = a \odot (b \odot x),\ \forall x \in V,\ \forall a, b \in K$ (Assoziativität der Skalarmultiplikation)

(viii) $1_K \odot x = x,\ \forall x \in V$ (Neutralität des Einselements)

Satz 5.2. Sei $[a, b]$ ein Intervall und $\mathcal{C}([a, b]) := \{f : [a, b] \to \mathbb{R} \mid f \text{ ist stetig}\}$ die Menge aller stetigen Funktionen auf $[a, b]$. Dann ist $\mathcal{C}([a, b])$ mit der „gewöhnlichen“ Addition von Funktionen und der skalaren Multiplikation von Funktionen ein Vektorraum bezüglich $\mathbb{R}$.

Satz 5.3. Sei $n \in \mathbb{N}_0$ und sei $\mathcal{P}_n$ die Menge aller Polynome mit Höchstgrad n. Dann ist $\mathcal{P}_n$ mit der „gewöhnlichen“ Addition von Funktionen und der skalaren Multiplikation von Funktionen ein Vektorraum bezüglich $\mathbb{R}$.

Definition 5.4. Sei V ein Vektorraum bezüglich K. Dann bezeichnet man eine Teilmenge $B \subseteq V$ als Basis von V, wenn für alle $x \in V$ gilt, dass es für alle $b \in B$ ein eindeutig bestimmtes $k_b \in K$ gibt (wobei höchstens endlich viele $k_b \neq 0$ sind), sodass $x = \sum_{b \in B} k_b \cdot b$ gilt.

Das heißt, jedes Element des Vektorraums lässt sich eindeutig als Linearkombination von Basiselementen darstellen.

Definition 5.5. Sei V ein reeller Vektorraum. Dann bezeichnet man eine Abbildung $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ als Skalarprodukt, falls folgende Eigenschaften erfüllt sind:

(i) $\langle x, x \rangle \geq 0, \ \forall x \in V$

(ii) $\langle x, x \rangle = 0 \Leftrightarrow x = 0$

(iii) $\langle x, y \rangle = \langle y, x \rangle, \ \forall x, y \in V$

(iv) $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle, \ \forall x, y, z \in V, \ \forall \alpha, \beta \in \mathbb{R}$

Definition 5.6. Sei V ein Vektorraum, $\langle \cdot, \cdot \rangle$ ein Skalarprodukt auf V und B eine Basis von V. Dann nennt man $B = \{b_1, b_2, ...\}$ Orthogonalbasis, falls $\langle b_i, b_j \rangle = 0$ gilt für alle $i \neq j$.

Satz 5.7. Sei $n \in \mathbb{N}_0$ und sei $\mathcal{P}_n$ der Vektorraum der Polynome mit Höchstgrad n. Seien $p_0, ..., p_n \in \mathcal{P}_n$ so, dass $\operatorname{grad} p_k = k, \ \forall k \in \{0, ..., n\}$. Setze $B := \{p_0, ..., p_n\}$. Dann ist B eine Basis von $\mathcal{P}_n$.

Das bedeutet, um eine Basis für $\mathcal{P}_n$ zu erhalten, muss lediglich für jedes $k \in \{0, ..., n\}$ genau ein beliebiges Polynom mit Grad k ausgewählt werden.

Definition 5.8. Eine Funktion $\alpha : [a, b] \to \mathbb{R}$ nennt man Gewichtsfunktion, falls folgende Eigenschaften erfüllt sind:

(i) α ist stetig.

(ii) $\alpha(x) > 0, \ \forall x \in (a, b)$

(iii) $\alpha(x) \geq 0$ für $x \in \{a, b\}$.

Satz 5.9. Sei α eine Gewichtsfunktion und seien f, g stetige Funktionen auf $[a, b]$. Dann ist

$$\langle f, g \rangle_\alpha := \int_a^b \alpha(x) f(x) g(x) \, dx$$

ein Skalarprodukt.

Beweis: Es müssen die vier Eigenschaften aus Definition 5.5 überprüft werden.

(i) Es ist gemäß der Definition der Gewichtsfunktion $\alpha(x) \geq 0$ für $x \in [a, b]$. Für die Funktion f gilt außerdem $f(x) \cdot f(x) = f(x)^2 \geq 0$, $\forall x \in [a, b]$. Somit erhält man

$$\langle f, f \rangle_\alpha = \int_a^b \underbrace{\alpha(x)}_{\geq 0} \underbrace{f(x)^2}_{\geq 0} \, dx \geq 0.$$

(ii) $[\Leftarrow]$: Sei $f = 0$. Dann gilt

$$\langle f, f \rangle_\alpha = \langle 0, 0 \rangle_\alpha = \int_a^b \alpha(x) \cdot 0^2 \, dx = 0.$$

$[\Rightarrow]$: Angenommen, es gibt ein $x_0 \in [a, b]$, sodass $f(x_0) \neq 0$. Weil f stetig ist gibt es daher ein $\delta > 0$, sodass $f(x) \neq 0, \ \forall x \in (x_0 - \delta, x_0 + \delta)$. Somit erhält man

$$\langle f,f\rangle_\alpha = \int_a^b \alpha(x)f(x)^2\,dx = \underbrace{\int_a^{x_0-\delta} \alpha(x)f(x)^2\,dx}_{\geq 0} + \underbrace{\int_{x_0-\delta}^{x_0+\delta} \alpha(x)f(x)^2\,dx}_{>0} + \underbrace{\int_{x_0+\delta}^{b} \alpha(x)f(x)^2\,dx}_{\geq 0} > 0.$$

Daher war die Annahme falsch und es ist $f(x) = 0,\ \forall x \in [a,b]$.

(iii) Seien f und g zwei stetige Funktionen. Dann gilt

$$\langle f,g\rangle_\alpha = \int_a^b \alpha(x)\underbrace{f(x)g(x)}_{=g(x)f(x)}\,dx = \int_a^b \alpha(x)g(x)f(x)\,dx = \langle g,f\rangle_\alpha.$$

(iv) Seien f,g und h stetige Funktionen und seien $\lambda,\mu \in \mathbb{R}$. Dann gilt:

$$\langle \lambda f + \mu g, h\rangle_\alpha = \int_a^b \alpha(x)\left(\lambda f(x) + \mu g(x)\right)h(x)\,dx =$$

$$\lambda\int_a^b \alpha(x)f(x)h(x)\,dx + \mu\int_a^b \alpha(x)g(x)h(x)\,dx = \lambda\langle f,h\rangle_\alpha + \mu\langle g,h\rangle_\alpha.$$

□

Definition 5.10. Sei α eine Gewichtsfunktion. Dann bezeichnet man das in Satz 5.9 definierte Skalarprodukt $\langle\cdot,\cdot\rangle_\alpha$ als gewichtetes Skalarprodukt bezüglich α.

Definition 5.11. Sei α eine Gewichtsfunktion und seien f,g stetige Funktionen auf $[a,b]$. Dann nennt man f und g orthogonal bezüglich α (α-orthogonal), falls $\langle f,g\rangle_\alpha = 0$ gilt.

5.2 Rekursionsformel

Definition 5.12. Sei I ein Intervall und sei α eine Gewichtsfunktion auf I. Dann nennt man $(p_n)_{n\in\mathbb{N}_0}$ ein orthogonales Polynomsystem, falls für alle $i,j \in \mathbb{N}_0$ mit $i \neq j$ gilt, dass $\langle p_i,p_j\rangle_\alpha = 0$ ist.

Satz 5.13. Sei α eine Gewichtsfunktion. Die Polynomfolge $(p_k)_{k\in\mathbb{N}_0}$ wird definiert durch

(i) $p_0(x) := 1$

(ii) $p_1(x) := x - a_1$

(iii) $p_k(x) := xp_{k-1}(x) - a_kp_{k-1}(x) - b_kp_{k-2}(x)$ für $k \geq 2$,

wobei die Koeffizienten a_k und b_k folgendermaßen definiert sind:

$$a_k := \frac{\langle xp_{k-1},p_{k-1}\rangle_\alpha}{\langle p_{k-1},p_{k-1}\rangle_\alpha} \qquad \text{und} \qquad b_k := \frac{\langle xp_{k-1},p_{k-2}\rangle_\alpha}{\langle p_{k-2},p_{k-2}\rangle_\alpha}$$

Diese Polynomfolge $(p_k)_{k\in\mathbb{N}_0}$ bildet ein orthogonales Polynomsystem.

Beweis: Der Beweis wird in Anlehnung an [6, Seite 183, Satz 6.3.4] geführt. Es ist aus der Definition sofort ersichtlich, dass p_k ein Polynom mit Grad k ist. Insbesondere ist $p_k \neq 0$, weshalb der Nenner von a_k und b_k nicht 0 ist.

Es wird mittels Induktion gezeigt, dass für $k \in \mathbb{N}_0$ gilt, dass $\langle p_k, p_i \rangle_\alpha = 0$ ist für alle $i < k$.

Als Induktionsanfang gilt für $k = 1$ wegen $p_0 := 1$, dass $x = x \cdot 1 = xp_0$ und $a_1 = a_1 \cdot 1 = a_1 p_0$ erfüllt ist. Setzt man nun die Definition von p_1 ein, so erhält man unter Verwendung der soeben beschriebenen Eigenschaften und der Definition des Skalarprodukts das gewünschte Resultat:

$$\langle p_1, p_0 \rangle_\alpha = \langle x - a_1, p_0 \rangle_\alpha = \langle x, p_0 \rangle_\alpha - \langle a_1, p_0 \rangle_\alpha = \langle xp_0, p_0 \rangle_\alpha - \langle a_1 p_0, p_0 \rangle_\alpha =$$

$$= \langle xp_0, p_0 \rangle_\alpha - a_1 \langle p_0, p_0 \rangle_\alpha = \langle xp_0, p_0 \rangle_\alpha - \frac{\langle xp_0, p_0 \rangle_\alpha}{\langle p_0, p_0 \rangle_\alpha} \langle p_0, p_0 \rangle_\alpha = 0.$$

Sei als Induktionsvoraussetzung (IV) die Aussage für ein beliebiges aber festes $k \geq 1$ bewiesen. Es wird nun der Induktionsschritt $k \mapsto k + 1$ durchgeführt.

Zunächst erhält man für $\langle p_{k+1}, p_k \rangle_\alpha$ durch Einsetzen von $p_{k+1} = xp_k - a_{k+1}p_k - b_{k+1}p_{k-1}$

$$\langle p_{k+1}, p_k \rangle_\alpha = \langle xp_k, p_k \rangle_\alpha - a_{k+1} \langle p_k, p_k \rangle_\alpha - b_{k+1} \underbrace{\langle p_{k-1}, p_k \rangle_\alpha}_{\overset{\text{IV}}{=} 0} =$$

$$= \langle xp_k, p_k \rangle_\alpha - \frac{\langle xp_k, p_k \rangle_\alpha}{\langle p_k, p_k \rangle_\alpha} \langle p_k, p_k \rangle_\alpha = 0.$$

Auf analoge Weise erhält man

$$\langle p_{k+1}, p_{k-1} \rangle_\alpha = \langle xp_k, p_{k-1} \rangle_\alpha - a_{k+1} \underbrace{\langle p_k, p_{k-1} \rangle_\alpha}_{\overset{\text{IV}}{=} 0} - b_{k+1} \langle p_{k-1}, p_{k-1} \rangle_\alpha =$$

$$= \langle xp_k, p_{k-1} \rangle_\alpha - \frac{\langle xp_k, p_{k-1} \rangle_\alpha}{\langle p_{k-1}, p_{k-1} \rangle_\alpha} \langle p_{k-1}, p_{k-1} \rangle_\alpha = 0.$$

Sei nun $i \in \{k - 2, ..., 1\}$. Dann gilt

$$\langle p_{k+1}, p_i \rangle_\alpha = \langle xp_k, p_i \rangle_\alpha - a_{k+1} \underbrace{\langle p_k, p_i \rangle_\alpha}_{\overset{\text{IV}}{=} 0} - b_{k+1} \underbrace{\langle p_{k-1}, p_i \rangle_\alpha}_{\overset{\text{IV}}{=} 0} = \langle xp_k, p_i \rangle_\alpha$$

Mit Definition 5.10 erhält man schließlich

$$\langle xp_k, p_i \rangle_\alpha = \int_a^b \alpha(x) \underbrace{xp_k(x)p_i(x)}_{= p_k(x) x p_i(x)} = \langle p_k, xp_i \rangle_\alpha.$$

Formt man nun die aus Rekursionsvorschrift (iii) gewonnene Gleichung $p_{i+1} = xp_i - a_{i+1}p_i - b_{i+1}p_{i-1}$ um und setzt für xp_i ein, so erhält man

$$\langle p_k, xp_i \rangle_\alpha = \langle p_k, p_{i+1} + a_{i+1}p_i + b_{i+1}p_{i-1} \rangle_\alpha =$$

$$= \underbrace{\langle p_k, p_{i+1} \rangle_\alpha}_{\overset{\text{IV}}{=} 0} + a_{i+1} \underbrace{\langle p_k, p_i \rangle_\alpha}_{\overset{\text{IV}}{=} 0} + b_{i+1} \underbrace{\langle p_k, p_{i-1} \rangle_\alpha}_{\overset{\text{IV}}{=} 0} = 0.$$

Weil man für $i = 0$ nicht wie oben die allgemeine Rekursionsvorschrift verwenden kann, da hier der Term $b_{i+1}p_{i-1} = b_1 p_{-1}$ nicht definiert wäre, muss dieser Fall gesondert betrachtet werden. Dazu formt man Rekursionsvorschrift (ii) mittels $p_0 := 1$ folgendermaßen um: $p_1 = x - a_1 = x \cdot 1 - a_1 \cdot 1 = xp_0 - a_1 p_0 \Leftrightarrow xp_0 = p_1 + a_1 p_0$. Damit erhält man

$$\langle p_{k+1}, p_0\rangle_\alpha = \langle xp_k, p_0\rangle_\alpha - a_{k+1}\underbrace{\langle p_k, p_0\rangle_\alpha}_{\overset{\text{IV}}{=}0} - b_{k+1}\underbrace{\langle p_{k-1}, p_0\rangle_\alpha}_{\overset{\text{IV}}{=}0} = \langle p_k, xp_0\rangle_\alpha =$$

$$= \langle p_k, p_1 + a_1 p_0\rangle_\alpha = \underbrace{\langle p_k, p_1\rangle_\alpha}_{\overset{\text{IV}}{=}0} + a_1\underbrace{\langle p_k, p_0\rangle_\alpha}_{\overset{\text{IV}}{=}0} = 0,$$

wobei die Umformungen aus der Linearität des Skalarprodukts in der zweiten Komponente folgen, die sich aus den Eigenschaften (iii) und (iv) von Definition 5.5 ergibt. Es gilt somit für alle $k \in \mathbb{N}_0$, dass $\langle p_k, p_i\rangle_\alpha = 0$, $\forall i < k$. □

Korollar 5.14. Die in Satz 5.13 definierte Polynomfolge $(p_k)_{k\in\{0,\dots,n\}}$ ist eine Orthogonalbasis von $\mathcal{P}_n$.

Beweis: Nach Definition von $(p_k)_{k\in\mathbb{N}_0}$ ist sofort ersichtlich, dass $\operatorname{grad} p_k = k$ ist. Außerdem gilt nach Satz 5.13, dass $\langle p_k, p_i\rangle_\alpha = 0$ ist für alle $i < k$. Somit folgt nach Satz 5.7, dass $(p_k)_{k\in\{0,\dots,n\}}$ eine Orthogonalbasis für $\mathcal{P}_n$ ist. □

Satz 5.15. Sei $(p_k)_{k\in\mathbb{N}}$ die in Satz 5.13 definierte Polynomfolge. Dann gilt für ein festes $n \in \mathbb{N}$ und das dadurch festgelegte Polynom p_{n+1} für alle $q \in \mathcal{P}_n$, dass $\langle p_{n+1}, q\rangle_\alpha = 0$ ist.

Beweis: Es ist $(p_k)_{k\in\{0,\dots,n\}}$ gemäß Korollar 5.14 eine Orthogonalbasis von $\mathcal{P}_n$. Damit kann man q als Linearkombination von Basiselementen darstellen:

$$q = \sum_{i=0}^{n} a_i p_i$$

Man erhält dadurch wegen der in Satz 5.13 gezeigten Eigenschaft $\langle p_k, p_i\rangle_\alpha = 0$ für alle $k \in \mathbb{N}$ und alle $i < k$

$$\langle p_{n+1}, q\rangle_\alpha = \langle p_{n+1}, a_0 p_0 + \dots + a_n p_n\rangle_\alpha = a_0\underbrace{\langle p_{n+1}, p_0\rangle_\alpha}_{=0} + \dots + a_n\underbrace{\langle p_{n+1}, p_n\rangle_\alpha}_{=0} = 0.$$

□

Bemerkung: Hat das Polynom q einen höheren Grad als n, so gilt diese Aussage nicht, da die Linearkombination von Basiselementen auch p_{n+1} enthalten könnte, und das Skalarprodukt in diesem Fall nicht verschwindet.

5.3 Spezielle orthogonale Polynomsysteme

In dieser Arbeit wird auf vier spezielle Folgen von orthogonalen Polynomen eingegangen, welche später beim Gauß-Verfahren (Kapitel 9) verwendet werden:

- Legendre-Polynome
- Tschebyschow-Polynome

- Laguerre-Polynome
- Hermite-Polynome

Bemerkung: Die durch die Rekursion von Satz 5.13 gewonnenen Polynome sind normiert. Somit unterscheiden sie sich in vielen Fällen durch einen Faktor von den klassischen Vertretern der genannten Polynomfolgen. Diesen Faktor kann man jedoch durch die hier vorgestellte Rekursion nicht erhalten, sondern nur durch Behandeln der klassischen Problemstellung, deren Lösungen diese Polynomfolgen sind. In den nachfolgenden Abbildungen sind jeweils die Funktionsgraphen der klassischen Polynome dargestellt.

Neben den hier vorgestellten orthogonalen Polynomsystemen sind auch die Jacobi-Polynome und die Bessel-Polynome erwähnenswert.

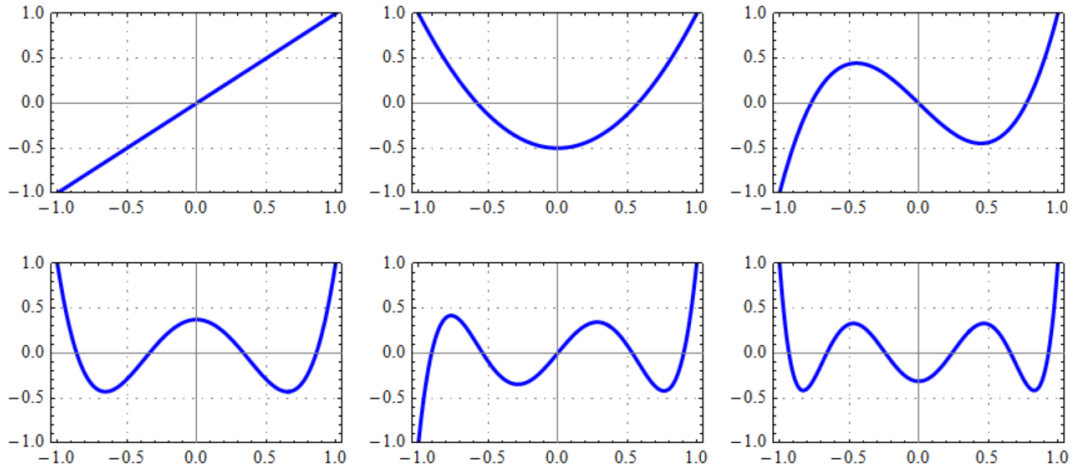
5.4 Legendre-Polynome

Die Legendre-Polynome, benannt nach dem französischen Mathematiker Adrien-Marie Legendre (1752 – 1833), erhält man durch die Rekursion von Satz 5.13 unter Verwendung der Gewichtsfunktion $\alpha(x) = 1$ und des Intervalls $[-1, 1]$.

Die ersten Vertreter dieser Polynomfolge sind:

$$\begin{aligned}
p_0(x) &= 1 \\
p_1(x) &= x \\
p_2(x) &= x^2 - \tfrac{1}{3} \\
p_3(x) &= x^3 - \tfrac{3}{5}x \\
p_4(x) &= x^4 - \tfrac{6}{7}x^2 + \tfrac{3}{35} \\
p_5(x) &= x^5 - \tfrac{10}{9}x^3 + \tfrac{5}{21}x \\
p_6(x) &= x^6 - \tfrac{15}{11}x^4 + \tfrac{5}{11}x^2 - \tfrac{5}{231}
\end{aligned}$$

Die Funktionsgraphen der klassischen Legendre-Polynome $p_1, ..., p_6$ sehen folgendermaßen aus:



Ihren Ursprung haben die Legendre-Polynome als Lösungen der Legendre-Differentialgleichungen

$$(1-x^2)f''(x) - 2xf'(x) + n(n+1)f(x) = 0.$$

Sie unterscheiden sich jedoch in den meisten Fällen um einen Faktor von den oben genannten Polynomen. Die explizite Formel zur Berechnung der klassischen Legendre-Polynome lautet gemäß [5, Seite 98, Beispiel 85] folgendermaßen:

$$p_n(x) = \frac{1}{2^n n!} \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^k \binom{n}{k} \frac{(2n-2k)!}{(n-2k)!} x^{n-2k}$$

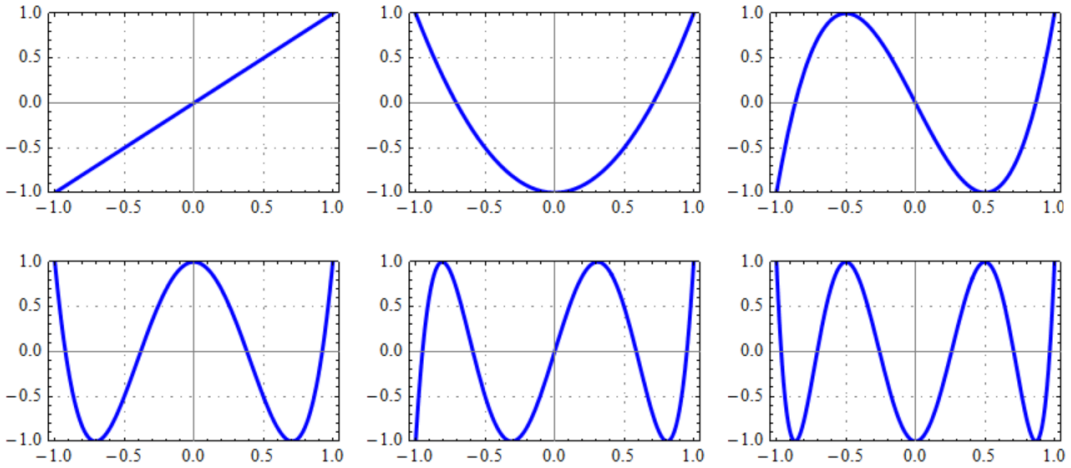
5.5 Tschebyschow-Polynome

Bei dieser Polynomfolge, welche nach dem russischen Mathematiker Pafnuti Lwowitsch Tschebyschow (1821 – 1894) benannt ist, verwendet man ebenfalls das Intervall $[-1, 1]$, jedoch in Kombination mit der Gewichtsfunktion $\alpha(x) = \frac{1}{\sqrt{1-x^2}}$.

Die ersten Vertreter dieser Polynomfolge sind folgende Polynome:

$$\begin{aligned}
p_0(x) &= 1 \\
p_1(x) &= x \\
p_2(x) &= x^2 - \tfrac{1}{2} \\
p_3(x) &= x^3 - \tfrac{3}{4}x \\
p_4(x) &= x^4 - x^2 + \tfrac{1}{8} \\
p_5(x) &= x^5 - \tfrac{5}{4}x^3 + \tfrac{5}{16}x \\
p_6(x) &= x^6 - \tfrac{3}{2}x^4 + \tfrac{9}{16}x^2 - \tfrac{1}{32}
\end{aligned}$$

Nachfolgend sind die Funktionsgraphen der klassischen Tschebyschow-Polynome $p_1, ..., p_6$ dargestellt:



Ihren Ursprung haben die Tschebyschow-Polynome als Lösungen der Tschebyschow-Differentialgleichungen

$$(1-x^2)f''(x) - xf'(x) + n^2 f(x) = 0.$$

Allgemein können die klassischen Tschebyschow-Polynome durch die Rekursionsformel $p_{n+1}(x) = 2xp_n(x) - p_{n-1}(x)$ berechnet werden, wobei $p_0(x) = 1$ und $p_1(x) = x$ ist. Im Intervall $[-1, 1]$ können die Funktionswerte außerdem durch die explizite Formel

$$p_n(x) = \cos(n \cdot \arccos(x))$$

ermittelt werden. Interessant für die spätere Gauß-Quadratur ist dieses orthogonale Polynomsystem insbesondere deshalb, weil die Nullstellen im Vergleich zu vielen anderen orthogonalen Polynomsystemen mit Hilfe einer einfachen Formel berechnet werden können. Die Nullstellen von p_n sind laut [14, Seite 19] für $k \in \{1, ..., n\}$ gegeben durch $\cos\left(\frac{2k-1}{2n} \cdot \pi\right)$.

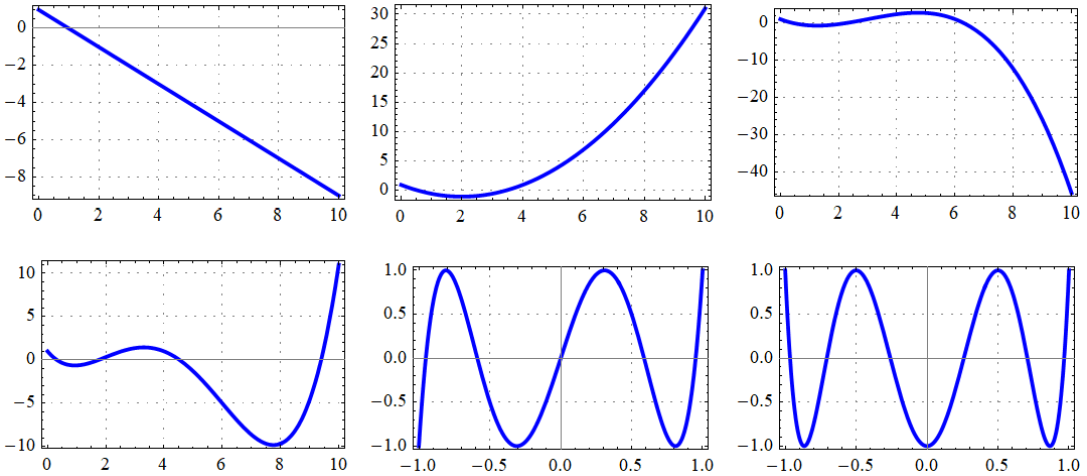
5.6 Laguerre-Polynome

Bei dieser nach Edmond Laguerre (Frankreich, 1834 – 1886) benannten Polynomfolge wird die Dichtefunktion $\alpha(x) = e^{-x}$ und das Intervall $[0, +\infty)$ verwendet.

Die ersten Vertreter dieser Polynomfolge sind:

$$\begin{aligned}
p_0(x) &= 1 \\
p_1(x) &= x - 1 \\
p_2(x) &= x^2 - 4x + 2 \\
p_3(x) &= x^3 - 9x^2 + 18x - 6 \\
p_4(x) &= x^4 - 16x^3 + 72x^2 - 96x + 24 \\
p_5(x) &= x^5 - 25x^4 + 200x^3 - 600x^2 + 600x - 120 \\
p_6(x) &= x^6 - 36x^5 + 450x^4 - 2400x^3 + 5400x^2 - 4320x + 720
\end{aligned}$$

Nachfolgende Abbildung zeigt die Funktionsgraphen der klassischen Laguerre-Polynome $p_1, ..., p_6$.



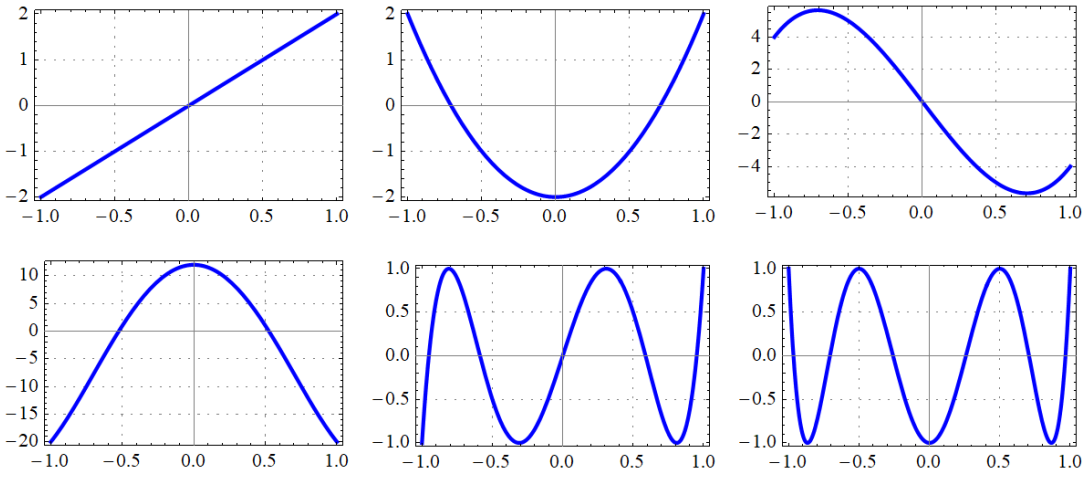
5.7 Hermite-Polynome

Diese nach Charles Hermite (Frankreich, 1822 – 1901) benannte Polynomfolge verwendet die Dichtefunktion $\alpha(x) = e^{-x^2}$ und das Intervall $\mathbb{R}$.

Die ersten Vertreter der Polynomfolge sind:

$$\begin{aligned} p_0(x) &= 1 \\ p_1(x) &= x \\ p_2(x) &= x^2 - \tfrac{1}{2} \\ p_3(x) &= x^3 - \tfrac{3}{2}x \\ p_4(x) &= x^4 - 3x^2 + \tfrac{3}{4} \\ p_5(x) &= x^5 - 5x^3 + \tfrac{15}{4}x \\ p_6(x) &= x^6 - \tfrac{15}{2}x^4 + \tfrac{45}{4}x^2 - \tfrac{15}{8} \end{aligned}$$

Nachfolgende Abbildung zeigt die Funktionsgraphen der klassischen Hermite-Polynome $p_1, ..., p_6$.



Gemäß [19, Seite 149, Formel 6.280] können die Hermite-Polynome durch folgende Formel explizit berechnet werden:

$$p_n(x) = \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^k \, (2x)^{n-2k} \, \frac{n!}{k!(n-2k)!}$$

6 Numerische Integrationsverfahren

Ziel der numerischen Integration – auch als numerische Quadratur bezeichnet – ist es, eine Näherung für das reellwertige Integral

$$\int_a^b f(x)\,dx$$

zu bestimmen. In folgenden beiden Situationen kommen numerische Integrationsverfahren typischerweise zum Einsatz:

- Das Integral ist nicht geschlossen lösbar, d. h. die Stammfunktion ist nicht durch elementare Funktionen darstellbar. Zwei bekannte Beispiele hierfür sind der Umfang der Ellipse und die Dichtefunktion der Normalverteilung.
- Der Integrand ist nur an endlich vielen Stellen bekannt. Dies ist beispielsweise der Fall, wenn die zu integrierende Funktion nur in Form von endlich vielen Messwerten vorliegt (siehe Beispiel auf Seite 62).

Da das Berechnen möglichst genauer Funktionswerte des Integranden in vielen Fällen mit hohem Rechenaufwand verbunden ist und die Ermittlung einer großen Anzahl von Messwerten ebenfalls aufwändig ist und oftmals mit mehr Zeitaufwand bei der Durchführung von Experimenten bzw. höheren Kosten verbunden ist, kann man die Qualität eines numerischen Integrationsverfahrens u. a. daran messen, wie viele Stützstellen für das Erreichen einer bestimmten Anzahl an korrekten Stellen benötigt werden.

6.1 Interpolationsquadratur

Bei dieser Methode handelt es sich um eine Familie von numerischen Integrationsverfahren, denen die Idee zugrunde liegt, den Integranden f (stückweise) durch ein Interpolationspolynom anzunähern, welches anschließend exakt integriert werden kann.

Nachfolgend werden zur besseren Übersichtlichkeit in vielen Abschnitten die Stützstellen mit einer Tilde versehen, also beispielsweise $\tilde{x}_i$, und die Teilintervallgrenzen ohne Tilde notiert, also z. B. $[x_i, x_{i+1}]$.

Definition 6.1. Seien $f : [a, b] \to \mathbb{R}$ die zu integrierende Funktion, $n \in \mathbb{N}$ und $\tilde{x}_0, ..., \tilde{x}_n \in [a, b]$ paarweise verschiedene Stützstellen. Dann bezeichnet man die Näherung des Integrals $\int_a^b f(x)\,dx$ durch das Integral $\int_a^b \psi_f(x)\,dx$ des zugehörigen Interpolationspolynoms $\psi_f(x)$ als Interpolationsquadratur.

Zu jeder Interpolationsquadratur existiert eine sogenannte Quadraturformel Q, welche lediglich vom Integrationsintervall $[a, b]$ und der Auswahl der Stützstellen $\tilde{x}_0, ..., \tilde{x}_n \in [a, b]$ abhängig ist, nicht jedoch von der zu integrierenden Funktion f.

Definition 6.2. Sei $[a,b]$ ein Intervall, $n \in \mathbb{N}$ und $\tilde{x}_0, ..., \tilde{x}_n \in [a,b]$ paarweise verschieden. Dann bezeichnet man

$$Q(f) := (b-a) \sum_{i=0}^{n} w_i f(\tilde{x}_i)$$

mit den Gewichten $w_0, ..., w_n \in \mathbb{R}$ als Quadraturformel, falls für alle $f : [a,b] \to \mathbb{R}$ gilt $Q(f) = \int_a^b \psi_f(x)\, dx$, wobei $\psi_f(x)$ das Interpolationspolynom von f zu den vorgegebenen Stützstellen ist.

Der folgende Satz soll zeigen, wie man bei gegebenen Stützstellen unter Verwendung der Lagrange-Interpolation die zugehörigen Gewichte erhält.

Satz 6.3. Seien $f : [a,b] \to \mathbb{R}$ die zu integrierende Funktion, $n \in \mathbb{N}$ und $\tilde{x}_0, ..., \tilde{x}_n \in [a,b]$ paarweise verschiedene Stützstellen. Für die zu diesen Stützstellen gehörende Quadraturformel $Q(f)$ sind die Gewichte $w_0, ..., w_n$ gegeben durch

$$w_i = \frac{1}{b-a} \int_a^b \prod_{\substack{k=0 \\ k \neq i}}^{n} \frac{x - \tilde{x}_k}{\tilde{x}_i - \tilde{x}_k}\, dx.$$

Beweis: Gemäß Definition 6.1 wird zu den $n+1$ paarweise verschiedenen Stützstellen das Interpolationspolynom $\psi_f(x)$ mit Höchstgrad n in Lagrange-Darstellung ermittelt. Es gilt:

$$Q(f) = \int_a^b \psi_f(x)\, dx = \int_a^b \left(\sum_{i=0}^{n} f(\tilde{x}_i) \ell_i(x) \right) dx$$

Da Integrale und endliche Summen vertauscht werden können, erhält man schließlich:

$$Q(f) = \sum_{i=0}^{n} \int_a^b f(\tilde{x}_i) \ell_i(x)\, dx = \sum_{i=0}^{n} f(\tilde{x}_i) \int_a^b \ell_i(x)\, dx = (b-a) \sum_{i=0}^{n} f(\tilde{x}_i) \underbrace{\frac{1}{b-a} \int_a^b \ell_i(x)\, dx}_{=:w_i}$$

Für die Gewichte w_i gilt also:

$$w_i = \frac{1}{b-a} \int_a^b \ell_i(x)\, dx = \frac{1}{b-a} \int_a^b \prod_{\substack{k=0 \\ k \neq i}}^{n} \frac{x - \tilde{x}_k}{\tilde{x}_i - \tilde{x}_k}\, dx.$$

□

Satz 6.4. Sei $[a,b]$ ein Intervall und seien $\tilde{x}_0, ..., \tilde{x}_n \in [a,b]$ paarweise verschiedene Stützstellen. Dann ist die zugehörige Quadraturformel eindeutig bestimmt.

Beweis: Dies ergibt sich sofort aus der Eindeutigkeit des Interpolationspolynoms ψ_f, also Satz 3.8. □

Somit ist klar, dass Quadraturformeln für ein gegebenes Intervall $[a,b]$ nur von der Wahl der Stützstellen abhängen. In den weiterführenden Kapiteln werden sich die verschiedenen Integrationsmethoden dementsprechend durch eine unterschiedliche Herangehensweise bei der Auswahl dieser Stützstellen auszeichnen.

Von besonderem Interesse ist bei den diversen numerischen Integrationsmethoden die Abschätzung des Fehlers $R(f)$. Für diesen werden zu den verschiedenen Integrationsmethoden sogenannte Fehlerschranken hergeleitet, mit welchen man den maximal auftretenden Fehler berechnen kann.

Definition 6.5. Sei $f \in C[a,b]$ und sei $Q(f)$ eine beliebige Quadraturformel. Dann definiert man den Quadraturfehler $R(f)$, welcher bei der Näherung des Integrals auftritt, folgendermaßen:

$$R(f) := \int_a^b f(x)\,dx - Q(f)$$

Bemerkung: Bei zusammengesetzten Quadraturformeln (also wenn das Intervall $[a,b]$ in mehrere Teilintervalle zerlegt wird, auf welche dann jeweils die Quadraturformel angewendet wird), wird der Quadraturfehler eines einzelnen Intervalls nachfolgend oftmals mit $r(f)$ bezeichnet.

Definition 6.6. Sei Q eine Quadraturformel, dann besitzt Q die Fehlerordnung m, wenn für den Quadraturfehler R Folgendes gilt:

(i) $R(p) = 0,\ \forall p \in \mathcal{P}_{m-1}$

(ii) $\exists p \in \mathcal{P}_m : R(p) \neq 0$

Dies bedeutet, dass durch eine Interpolationsquadratur mit Fehlerordnung m alle Polynome mit Höchstgrad $m-1$ exakt integriert werden.

Satz 6.7. Sei $m \in \mathbb{N}$ und sei $p \in \mathcal{P}_m$ mit $p(x) = \alpha_m x^m + ... + \alpha_1 x + \alpha_0$. Sei außerdem $Q(p) = (b-a)\sum_{i=0}^{n} w_i p(\tilde{x}_i)$ eine Quadraturformel. Dann gilt für den Quadraturfehler:

$$R(p) = \sum_{j=0}^{m} \alpha_j R(x^j)$$

Beweis: Aus der Definition des Quadraturfehlers erhält man durch Einsetzen in $p(x)$:

$$R(p) = \int_a^b p(x)\,dx - (b-a)\sum_{i=0}^{n} w_i p(\tilde{x}_i) = \sum_{j=0}^{m} \alpha_j \int_a^b x^j\,dx - (b-a)\underbrace{\sum_{i=0}^{n} w_i \sum_{j=0}^{m} \alpha_j \tilde{x}_i^j}_{=\sum_{i=0}^{n}\sum_{j=0}^{m} w_i \alpha_j \tilde{x}_i^j}$$

Da man endliche Summen vertauschen kann, erhält man:

$$\sum_{j=0}^{m} \alpha_j \int_a^b x^j\,dx - (b-a)\underbrace{\sum_{j=0}^{m}\sum_{i=0}^{n} w_i \alpha_j \tilde{x}_i^j}_{=\sum_{j=0}^{m}\alpha_j \sum_{i=0}^{n} w_i \tilde{x}_i^j} = \sum_{j=0}^{m} \alpha_j \underbrace{\left(\int_a^b x^j\,dx - (b-a)\sum_{i=0}^{n} w_i \tilde{x}_i^j \right)}_{=R(x^j)} = \sum_{j=0}^{m} \alpha_j R(x^j)$$

□

Wegen dieser Linearitätseigenschaft des Quadraturfehlers genügt es also, jenes minimale $m \in \mathbb{N}$ zu finden, für das gilt $R(x^m) \neq 0$ und $R(x^k) = 0$, $\forall k \in \{0, ..., m-1\}$, um die Fehlerordnung zu ermitteln.

Satz 6.8. Sei $f : [a, b] \to \mathbb{R}$ eine Funktion, und sei $Q(f) := (b-a) \sum_{i=0}^{n} w_i f(x_i)$ eine Quadraturformel mit Fehlerordnung $m \geq 1$. Dann gilt $\sum_{i=0}^{n} w_i = 1$.

Beweis: Da die Fehlerordnung von Q zumindest 1 ist, gilt nach Definition 6.6 die Eigenschaft $R(x^0) = R(1) = 0$. Somit folgt aus der Definition des Quadraturfehlers

$$\int_a^b 1\,dx = Q(f) \Rightarrow b - a = (b-a) \sum_{i=0}^{n} w_i f(\tilde{x}_i) \Rightarrow 1 = \sum_{i=0}^{n} w_i \underbrace{f(\tilde{x}_i)}_{=1,\, \forall x \in [a,b]} \Rightarrow \sum_{i=0}^{n} w_i = 1.$$

□

Nachfolgender Satz liefert eine allgemeine Formel zur Abschätzung des Quadraturfehlers.

Satz 6.9. Sei $f : [a, b] \to \mathbb{R}$ eine $(n+1)$-mal stetig differenzierbare Funktion, und seien $\tilde{x}_0, ..., \tilde{x}_n \in [a, b]$ paarweise verschiedene Stützstellen. Sei ψ_f das minimale Interpolationspolynom von f zu diesen Stützstellen, und sei $Q(f)$ die nach Satz 6.4 eindeutig bestimmte Quadraturformel zu den Stützstellen $\tilde{x}_0, ..., \tilde{x}_n$. Dann gilt

$$\left| \int_a^b f(x)\,dx - Q(f) \right| \leq \frac{1}{(n+1)!} \cdot \max_{x \in [a,b]} |f^{(n+1)}(x)| \cdot \int_a^b \prod_{k=0}^{n} |(x - \tilde{x}_k)|\,dx.$$

Beweis: Grundlage dieses Beweises ist [6, Seite 168, Satz 6.1.3]. Nach Definition 6.2 gilt für alle $f : [a, b] \to \mathbb{R}$ die Eigenschaft $Q(f) = \int_a^b \psi_f(x)\,dx$. Somit erhält man

$$\left| \int_a^b f(x)\,dx - Q(f) \right| = \left| \int_a^b f(x)\,dx - \int_a^b \psi_f(x)\,dx \right| = \left| \int_a^b \left(f(x) - \psi_f(x)\right)\,dx \right|.$$

Des Weiteren gilt

$$\left| \int_a^b (f(x) - \psi_f(x))\,dx \right| \leq \int_a^b |f(x) - \psi_f(x)|\,dx.$$

Durch die Abschätzung des Restglieds der Polynominterpolation (Korollar 3.17) erhält man schließlich folgendes Ergebnis, wodurch die Aussage bewiesen wäre:

$$\int_a^b |f(x) - \psi_f(x)|\,dx \leq \int_a^b \frac{1}{(n+1)!} \cdot \max_{x \in [a,b]} |f^{(n+1)}(x)| \cdot \prod_{k=0}^{n} |(x - \tilde{x}_k)|\,dx =$$

$$= \frac{1}{(n+1)!} \cdot \max_{x \in [a,b]} |f^{(n+1)}(x)| \cdot \int_a^b \prod_{k=0}^{n} |(x - \tilde{x}_k)|\,dx$$

□

Nun wird mit Hilfe des vorherigen Satzes gezeigt, dass eine Quadraturformel zu $n+1$ Stützstellen zumindest die Fehlerordnung $m = n+1$ besitzt, d. h. alle Polynome mit Höchstgrad n werden exakt integriert.

Korollar 6.10. Sei $[a,b]$ ein Intervall und sei Q die Quadraturformel zu den $n+1$ paarweise verschiedenen Stützstellen $\tilde{x}_0, ..., \tilde{x}_n \in [a,b]$. Dann gilt für alle Polynome p mit Höchstgrad n, dass $\int_a^b p(x)\,dx - Q(p) = 0$ ist.

Beweis: Nach Satz 6.9 gilt unter Verwendung von Korollar 2.6

$$\left| \int_a^b p(x)\,dx - Q(p) \right| \leq \frac{1}{(n+1)!} \cdot \max_{x\in[a,b]} |\underbrace{p^{(n+1)}(x)}_{=0}| \cdot \int_a^b \prod_{k=0}^{n} |(x-\tilde{x}_k)|\,dx = 0.$$

□

Die Fehlerordnung einer Quadraturformel zu $n+1$ Stützstellen ist also mindestens $n+1$, d. h. die Fehlerordnung ist nach unten beschränkt. Satz 9.1 wird später zeigen, dass die Fehlerordnung einer Quadraturformel zu $n+1$ Stützstellen auch eine obere Schranke besitzt.

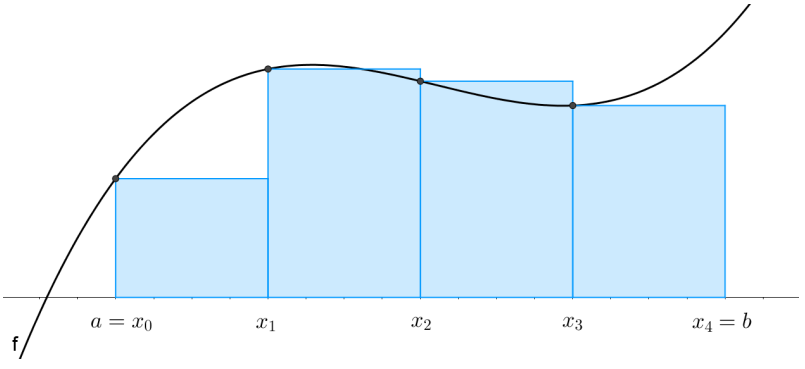
6.2 Rechteckregeln

Diese Klasse von Quadratuformeln bilden die wohl einfachste aber auch bei gleichem Rechenaufwand ungenaueste Herangehensweise. Hier wird die Funktion f in den n äquidistanten[1] Teilintervallen jeweils durch eine konstante Funktion mit Funktionswert $f(\tilde{x}_i)$ für ein $\tilde{x}_i \in [x_i, x_{i+1}]$ angenähert. Die Quadraturformel lautet

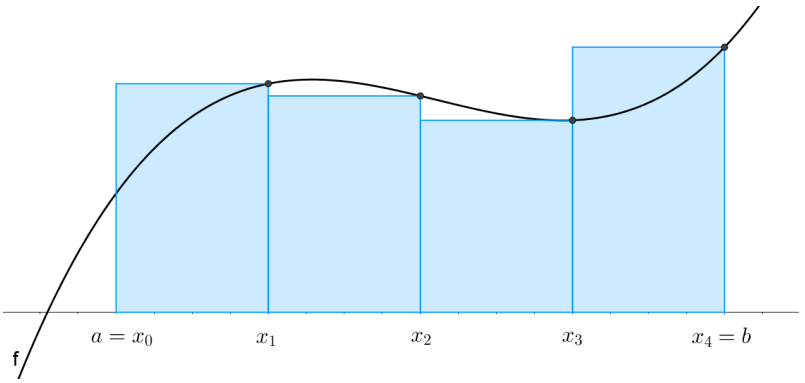
$$Q_R(f) = \sum_{i=1}^{n} h \cdot f(\tilde{x}_i)$$

mit $h = x_{i+1} - x_i$. Es handelt sich dabei also um eine Summe von Flächeninhalten von Rechtecken, weshalb diese Art von Quadraturformel auch als Rechteckregel bezeichnet wird.

Sinnvoll ist es, für $f(\tilde{x}_i)$ entweder den Funktionswert $f(x_i)$ oder den Funktionswert $f(x_{i+1})$ zu verwenden, also jenen am linken Rand des Intervalls (wie in der oberen der beiden nachfolgenden Skizzen) bzw. am rechten Rand des Intervalls (wie in der unteren der beiden nachfolgenden Skizzen).



[1] Grundsätzlich könnte man auch unterschiedlich breite Teilintervalle verwenden.



Satz 6.11. Sei $f : [a, b] \to \mathbb{R}$ stetig differenzierbar. Dann gilt für den Quadraturfehler $R(f)$ der beiden oben beschriebenen und illustrierten Quadraturformeln für äquidistante Teilintervalle der Breite h die Abschätzung

$$|R(f)| \leq \frac{h}{2} \max_{x \in [a,b]} |f'(x)| \cdot (b - a).$$

Beweis: Dieser Beweis beruht auf [10, Seite 8]. Es wird ohne Beschränkung der Allgemeinheit der Beweis anhand der Rechteckregel mit linkem Randwert geführt. Das bedeutet, dass sich die n Stützstellen bei $x_i = a + ih$ befinden, wobei $i \in \{0, ..., n - 1\}$ gilt.

Zunächst wird der Fehler $r(f)$ eines einzelnen Teilintervalls $[x_i, x_i + h] \subseteq [a, b]$ abgeschätzt. Aus dem Ansatz

$$r(f) = \int_{x_i}^{x_i+h} f(x)\, dx - h \cdot f(x_i)$$

erhält man durch Ableiten des Fehlers nach h unter Verwendung der Linearität der Ableitung

$$r'(f) = \frac{d}{dh} \int_{x_i}^{x_i+h} f(x)\, dx - f(x_i) = f(x_i + h) - f(x_i).$$

Man definiert nun

$$S_i := \max_{x \in [x_i, x_i+h]} |f'(x)|.$$

Durch nochmaliges Ableiten von $r'(f)$ nach h erhält man schließlich $r''(f) = f'(x_i + h) \leq S_i$. Aus $|r''(f)| \leq S_i$ folgt $|r'(f)| \leq S_i \cdot h$, woraus man wiederum $|r(f)| \leq S_i \cdot \frac{h^2}{2}$ erhält.

Für den zusammengesetzten Quadraturfehler $|R(f)|$ gilt nun Folgendes:

$$|R(f)| \leq \tfrac{h^2}{2} \sum_{i=0}^{n-1} S_i \leq \tfrac{h^2}{2} \sum_{i=0}^{n-1} \max_{x \in [a,b]} |f'(x)| \leq \tfrac{h \cdot (b-a)}{2n} \cdot n \cdot \max_{x \in [a,b]} |f'(x)| = \tfrac{h}{2} \cdot \max_{x \in [a,b]} |f'(x)| \cdot (b-a)$$

□

Beispiel: Gegeben ist die Funktion $f(x) := 0.03x^3 - 0.2x^2 + 0.2x + 3$ (welche auch bereits für die beiden Abbildungen oben verwendet wurde). Diese soll im Intervall $[-2, 6]$ mittels zusammengesetzter Rechtecksregel (einmal mit linkem Randwert und einmal mit rechtem Randwert) näherungsweise integriert werden, wobei als Schrittweite $h = 2$ verwendet wird.

Nachfolgende Tabelle zeigt die Funktionswerte bei den verwendeten Stützstellen:

x	-2	0	2	4	6
$f(x)$	1.56	3.00	2.84	2.52	3.48

Aus dieser Wertetabelle erhält man folgende Näherungswerte für das Integral:

$$Q_{\text{links}}(f) = 2 \cdot (1.56 + 3.00 + 2.84 + 2.52) = 19.84$$
$$Q_{\text{rechts}}(f) = 2 \cdot (3.00 + 2.84 + 2.52 + 3.84) = 24.40$$

Außerdem ist

$$\max_{x \in [-2,6]} |f'(x)| = 1.36,$$

was dem Wert an der Stelle $x = -2$ entspricht. Gemäß Satz 6.11 beträgt die maximale Abweichung somit $\frac{2}{2} \cdot 1.36 \cdot (6 - (-2)) = 10.88$. Der tatsächliche Wert des Integrals beträgt ungefähr 21.87. Die Näherungswerte sind somit deutlich näher am tatsächlichen Wert, als dies durch Satz 6.11 prognostiziert wurde.

7 Newton-Cotes-Formeln

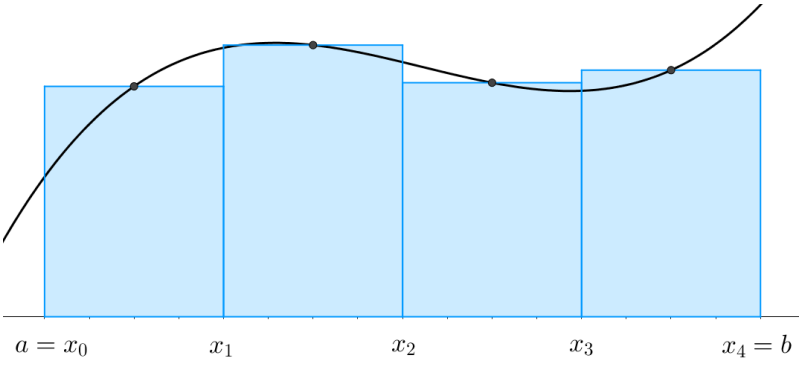
Grundsätzlich können die Stützstellen im Intervall $[a, b]$ beliebig gewählt werden. Bei den in diesem Kapitel behandelten Newton-Cotes-Formeln, benannt nach Isaac Newton (England, 1643 – 1727)[1] und Roger Cotes (England, 1682 – 1716), sind die Stützstellen jedoch äquidistant. Beim Gauß-Verfahren, welches in Kapitel 9 behandelt wird, werden die Stützstellen beispielsweise anders ermittelt.

Nachfolgend werden zunächst einige bekannte Vertreter der Familie der Newton-Cotes-Formeln vorgestellt, bevor diese in ein verallgemeinertes Modell eingegliedert werden. Es wird dabei stets von Beginn an die zusammengesetzte Version betrachtet. Im Beispiel der Mittelpunktregel bedeutet dies, dass nicht vom gesamten Intervall $[a, b]$ der Mittelpunkt als einzige Stützstelle herangezogen wird, sondern die Mittelpunktregel auf eine Reihe von Teilintervallen angewendet wird, deren Ergebnisse anschließend summiert werden.

Je nach behandelter Regel kann es vorkommen, dass für die Gesamtanzahl k an Stützstellen bestimmte Bedingungen existieren. Bei der Simpsonregel werden beispielsweise pro Teilintervall drei Stützstellen benötigt, wobei bis auf a und b alle Teilintervallgrenzen doppelt verwendet werden. Somit muss die Gesamtanzahl k an Stützstellen von der Form $k = 3\ell - (\ell - 1) = 2\ell + 1$ für $\ell \in \mathbb{N}$ sein[2].

7.1 Mittelpunktregel

Es handelt sich hier um eine spezielle Form der Rechteckregel welche ihren Namen deshalb erhielt, weil als Stützstellen die Mitten der äquidistanten Teilintervalle herangezogen werden. Gelegentlich wird sie auch als Tangententrapezregel bezeichnet, da das Tangententrapez flächengleich mit dem Rechteck ist.



[1] nach gregorianischem Kalender

[2] Prinzipiell werden pro Teilintervall drei Stützstellen benötigt, jedoch werden bei den insgesamt ℓ Teilintervallen die $\ell - 1$ inneren Teilintervallgrenzen doppelt verwendet.

Definition 7.1. Sei $f : [a, b] \to \mathbb{R}$ die zu integrierende Funktion. Dann lautet die Mittelpunktregel

$$\int_a^b f(x)\,dx \approx (b-a) \cdot f\left(\frac{a+b}{2}\right).$$

Aus den n Teilintervallen $[a + ih, a + (i + 1)h]$ mit $i \in \{0, ..., n - 1\}$ erhält man die n Stützstellen

$$\tilde{x}_i := \frac{(a + ih) + (a + (i + 1)h)}{2} = a + ih + \frac{h}{2}$$

für $i \in \{0, ..., n-1\}$ und dementsprechend die Stützwerte $f(\tilde{x}_i)$. Die Gewichte w_i entsprechen jeweils der Schrittweite h.

Die Quadraturformel für die zusammengesetzte Mittelpunktregel lautet somit

$$Q_M(f) = h \sum_{i=0}^{n-1} f\left(\tilde{x}_i\right).$$

Satz 7.2. Sei $f \in C^2[a, b]$. Dann gilt für den Quadraturfehler $R_M(f)$ der Mittelpunktregel

$$|R_M(f)| \leq \frac{h^2}{24} \max_{x\in[a,b]} |f''(x)| \cdot (b - a).$$

Beweis: Als Grundlage für diesen Beweis wurde [8, Seite 293] verwendet. Es wird zunächst der Fehler $r_M(f)$ für die einfache Mittelpunktregel, also für ein einzelnes Teilintervall $[x_i, x_{i+1}]$, berechnet. Dafür sei $h := x_{i+1} - x_i$. Die Stützstellen seien $\tilde{x}_i := x_i + \frac{h}{2}$. Nach Definition des Quadraturfehlers gilt

$$r_M(f) = \int_{x_i}^{x_{i+1}} f(x)\,dx - hf(\tilde{x}_i).$$

Da $f \in C_2[a, b]$ ist, kann man für ein $x \in [x_i, x_{i+1}]$ den Funktionswert $f(x)$ durch das 1. Taylorpolynom (siehe Definition 2.26) mit der Entwicklungsstelle $\tilde{x}_i$ annähern. Es gilt dann gemäß Satz 2.31 für ein von x abhängiges $\xi_i^* \in [x_i, x_{i+1}]$

$$f(x) = f(\tilde{x}_i) + f'(\tilde{x}_i) \cdot (x - \tilde{x}_i) + \frac{f''(\xi_i^*)}{2} \cdot (x - \tilde{x}_i)^2.$$

Somit folgt unter Verwendung von $x_{i+1} = x_i + h$ und $\tilde{x}_i = x_i + \frac{h}{2}$:

$$r_M(f) = \underbrace{\int_{x_i}^{x_{i+1}} f(\tilde{x}_i)\,dx}_{=hf(\tilde{x}_i)} + \underbrace{\int_{x_i}^{x_{i+1}} f'(\tilde{x}_i) \cdot (x - \tilde{x}_i)\,dx}_{=0} + \int_{x_i}^{x_{i+1}} \frac{f''(\xi_i^*)}{2} \cdot (x - \tilde{x}_i)^2\,dx - hf(\tilde{x}_i)$$

Damit reduziert sich die Berechnung auf

$$r_M(f) = \int_{x_i}^{x_{i+1}} \frac{f''(\xi_i^*)}{2} \cdot (x - \tilde{x}_i)^2 \, dx.$$

Sei $\xi_i \in [x_i, x_{i+1}]$ so, dass $f''(\xi_i) = \max_{x \in [x_i, x_{i+1}]} f''(x)$. Man beachte, dass ξ_i nun von x unabhängig ist. Es gilt

$$\left| \frac{f''(\xi_i^*)}{2} \cdot (x - \tilde{x}_i)^2 \right| \leq \left| \frac{f''(\xi_i)}{2} \cdot (x - \tilde{x}_i)^2 \right|.$$

Da nun im Integranden $f''(\xi_i^*)$ durch eine von x unabhängige Konstante ersetzt wurde, kann das Integral berechnet werden, und es folgt:

$$|r_M(f)| = \left| \int_{x_i}^{x_{i+1}} \frac{f''(\xi_i^*)}{2} (x - \tilde{x}_i)^2 \, dx \right| \leq \int_{x_i}^{x_{i+1}} \left| \frac{f''(\xi_i^*)}{2} (x - \tilde{x}_i)^2 \right| \, dx \leq$$

$$\leq \int_{x_i}^{x_{i+1}} \left| \frac{f''(\xi_i)}{2} (x - \tilde{x}_i)^2 \right| \, dx = \frac{f''(\xi_i)}{2} \int_{x_i}^{x_{i+1}} \left| (x - \tilde{x}_i)^2 \right| \, dx = \frac{f''(\xi_i)}{2} \int_{x_i}^{x_{i+1}} (x - \tilde{x}_i)^2 \, dx$$

Für die Berechnung des Integrals wird $x_{i+1} := x_i + h$ und $\tilde{x}_i := x_i + \frac{h}{2}$ gesetzt. Mittels MATHEMATICA[3] erhält man für das Integral den Wert $\frac{h^3}{12}$. Somit gilt

$$|r_M(f)| \leq \frac{f''(\xi_i)}{2} \cdot \frac{h^3}{12} = \frac{h^3}{24} \cdot f''(\xi_i).$$

Der gesamte Quadraturfehler ergibt sich nun durch

$$|R_M(f)| \leq \underbrace{\frac{h^3}{24}}_{= \frac{h^2}{24} \cdot \frac{b-a}{n}} \cdot \underbrace{\sum_{i=0}^{n-1} \left| f''(\xi_i) \right|}_{\leq n \cdot \max\limits_{x \in [a,b]} |f''(x)|} \leq \frac{h^2}{24} \max_{x \in [a,b]} |f''(x)| \cdot (b - a)$$

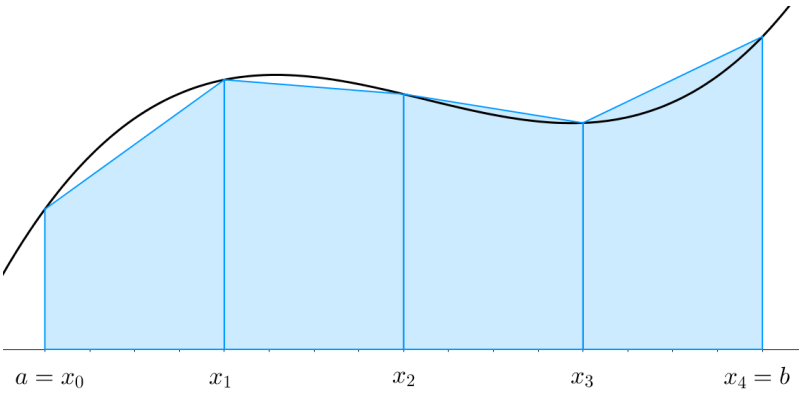
□

Bemerkung: Es ist hier anzumerken, dass das Fehlerglied quadratisch von der Intervalllänge h abhängt, obwohl die Funktion f durch eine konstante Funktion angenähert wird.

7.2 Sehnentrapezregel

Bei der Sehnentrapezregel (häufig unpräzise nur als Trapezregel bezeichnet) wird die zu integrierende Funktion in den n äquidistanten Teilintervallen stückweise linear interpoliert, wobei als Stützpunkte die Punkte des Funktionsgraphen an den Teilintervallgrenzen herangezogen werden. Die dadurch entstehende Quadraturformel kann daher als Summe von Flächeninhalten von Trapezen aufgefasst werden.

[3] Eingabe: `Integrate[(x - (xi + h/2))^2, x, xi, xi + h]`



Definition 7.3. Sei $f : [a, b] \to \mathbb{R}$ die zu integrierende Funktion. Dann lautet die Sehnentrapezregel

$$\int_a^b f(x)\,dx \approx (b-a) \cdot \frac{f(a)+f(b)}{2}.$$

Zerlegt man das Intervall $[a, b]$ in n äquidistante Teilintervalle $[x_i, x_{i+1}]$ mit $i \in \{0, ..., n-1\}$, $h = \frac{b-a}{n}$ und $x_i = a + ih$, so kann man aufgrund der Tatsache, dass in der Summe außer den beiden äußeren Stützstellen alle doppelt vorkommen, die zusammengesetzte Sehnentrapezregel folgendermaßen formulieren:

$$Q_T(f) = h \cdot \left(\frac{f(x_0)}{2} + \sum_{i=1}^{n-1} f(x_i) + \frac{f(x_n)}{2} \right)$$

Satz 7.4. Sei $f \in C^2[a, b]$. Für den Quadraturfehler $R_T(f)$ der Sehnentrapezregel gilt

$$|R_T(f)| \leq \frac{h^2}{12} \max_{x\in[a,b]} |f''(x)| \cdot (b-a).$$

Beweis: Als Grundlage für diesen Beweis diente [18, Satz 13.7]. Zunächst wird wieder der Fehler $r_T(f)$ für ein einzelnes Teilintervall $[x_i, x_{i+1}]$ abgeschätzt. Es gilt

$$r_T(f) = \int_{x_i}^{x_{i+1}} f(x)\,dx - \tfrac{h}{2} \cdot (f(x_i) + f(x_{i+1})) = \int_{x_i}^{x_{i+1}} (f(x) - \psi_f(x))\,dx.$$

Die letzte Umformung ist möglich, da das Integral $\int_{x_i}^{x_{i+1}} \psi_f(x)\,dx$ des linearen Interpolationspolynoms genau dem Flächeninhalt des Trapezes entspricht. Durch Erweitern erhält man

$$r_T(f) = \int_{x_i}^{x_{i+1}} (x-x_i)(x-x_{i+1}) \cdot \frac{f(x)-\psi_f(x)}{(x-x_i)(x-x_{i+1})}\,dx.$$

Für die Funktion $g(x) := (x-x_i)(x-x_{i+1}) = x^2 - (x_i+x_{i+1})x + x_i x_{i+1}$ gilt $g(x) \leq 0,\ \forall x \in [x_i, x_{i+1}]$, denn $x - x_i \geq 0,\ \forall x \in [x_i, x_{i+1}]$ und $x - x_{i+1} \leq 0,\ \forall x \in [x_i, x_{i+1}]$. Damit hat g im Intervall $[x_i, x_{i+1}]$ keinen Vorzeichenwechsel.

Als nächstes wird die Stetigkeit der Funktion

$$h(x) := \frac{f(x) - \psi_f(x)}{(x - x_i)(x - x_{i+1})}$$

untersucht. Die Funktion f ist nach Voraussetzung stetig, und ψ_f ist stetig, da Polynome laut Satz 2.7 stetig sind. Der Nenner, also die Funktion g, ist ebenfalls stetig, da es sich auch hier um ein Polynom handelt. Es müssen nur noch die beiden Stellen x_i und x_{i+1}, also die Nullstellen von g, näher untersucht werden.

Die Nullstelle von $g'(x) = 2x - (x_i + x_{i+1})$ liegt bei $x = \frac{x_i + x_{i+1}}{2}$. Weil somit $g'(x_i) \neq 0$ und $g'(x_{i+1}) \neq 0$ ist, existieren die beiden Grenzwerte

$$\lim_{x \to x_i^+} \frac{(f(x) - \psi_f(x))'}{g'(x)} \qquad \text{und} \qquad \lim_{x \to x_{i+1}^-} \frac{(f(x) - \psi_f(x))'}{g'(x)}.$$

Aufgrund der Interpolationseigenschaft gilt

$$\lim_{x \to x_i^+} (f(x) - \psi_f(x)) = \lim_{x \to x_{i+1}^-} (f(x) - \psi_f(x)) = 0.$$

Außerdem ist

$$\lim_{x \to x_i^+} g(x) = \lim_{x \to x_{i+1}^-} g(x) = 0.$$

Somit folgt aus Satz 2.14 (i) und (iii), dem Satz von de l'Hospital, dass die einseitigen Grenzwerte

$$\lim_{x \to x_i^+} h(x) = \lim_{x \to x_i^+} \frac{f(x) - \psi_f(x)}{g(x)} \qquad \text{und} \qquad \lim_{x \to x_{i+1}^-} h(x) = \lim_{x \to x_{i+1}^-} \frac{f(x) - \psi_f(x)}{g(x)}$$

existieren. Daher ist h auf (x_i, x_{i+1}) stetig, in x_i rechtsseitig stetig und in x_{i+1} linksseitig stetig. Daraus folgt, dass h stetig auf $[x_i, x_{i+1}]$ ist.

Aus dem Mittelwertsatz der Integralrechnung (Satz 2.13) folgt daher, dass es ein $\xi \in [x_i, x_{i+1}]$ gibt mit

$$r_T(f) = \int_{x_i}^{x_{i+1}} g(x)\, h(x)\, dx = h(\xi) \int_{x_i}^{x_{i+1}} g(x)\, dx.$$

In einer Nebenrechnung erhält man

$$\int_{x_i}^{x_{i+1}} (x - x_i)(x - x_{i+1})\, dx = \frac{(x_i - x_{i+1})^3}{6} = -\frac{(x_{i+1} - x_i)^3}{6}.$$

Außerdem erhält man unter Verwendung von Satz 3.16:

$$h(\xi) = \frac{f(\xi) - \psi_f(\xi)}{(\xi - x_i)(\xi - x_{i+1})} \overset{3.16}{=} \frac{\frac{f''(\xi)}{2!} \prod_{k=i}^{i+1} (\xi - x_k)}{(\xi - x_i)(\xi - x_{i+1})} = \frac{f''(\xi)}{2}$$

Somit ist

$$r_T(f) = -\frac{f''(\xi)}{12}(x_{i+1} - x_i)^3$$

für ein unbekanntes $\xi \in [x_i, x_{i+1}]$, weshalb man folgende Abschätzung durchführen kann:

$$|r_T(f)| \leq \frac{(x_{i+1} - x_i)^3}{12} \max_{x \in [x_i, x_{i+1}]} |f''(x)|.$$

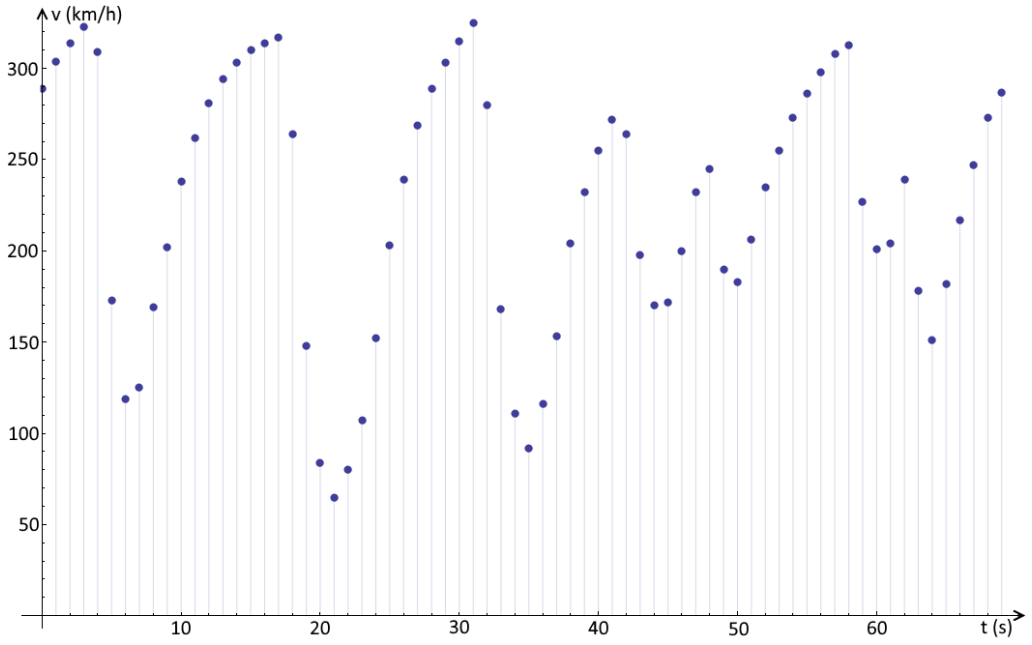
Zerlegt man nun $[a, b]$ in n äquidistante Teilintervalle mit den Grenzen $x_0, ..., x_n$, so erhält man:

$$|R_T(f)| \leq \underbrace{\frac{h^3}{12}}_{=\frac{h^2}{12} \cdot \frac{b-a}{n}} \cdot \underbrace{\sum_{i=0}^{n-1} \max_{x \in [x_i, x_{i+1}]} |f''(x)|}_{\leq n \cdot \max\limits_{x \in [a,b]} |f''(x)|} \leq \frac{h^2}{12} \max_{x \in [a,b]} |f''(x)| \cdot (b - a)$$

□

Bemerkung: Vergleicht man das Resultat mit Satz 7.2, so erkennt man, dass die Fehlerabschätzung der Sehnentrapezregel um den Faktor 2 größer ist, als jene der Mittelpunktregel.

Beispiel: An dieser Stelle wird ein anwendungsorientiertes Beispiel zur Sehnentrapezregel behandelt, welches die Einsatzmöglichkeiten der numerischen Integration demonstrieren soll. Es wird dabei anhand eines Videos[4] die Pole-Position-Runde von Felipe Massa beim Großen Preis von Österreich 2014 analysiert. Dieses Video wurde ausgewählt, da hier während der gesamten Runde einerseits die Geschwindigkeit des Fahrzeuges und andererseits die Rundenzeit eingeblendet wird (was bei vielen anderen Videos dieser Art nicht der Fall ist). Es war somit einfach, zu jeder vollen Sekunde die Geschwindigkeit zu bestimmen und in eine Liste einzutragen. Dafür wurde jeweils jenes Bild des Videos verwendet, bei welchem die Rundenzeit der jeweiligen vollen Sekunde am nächsten war. Man erhielt folgendes Diagramm, welches jeder Sekunde die Geschwindigkeit in km/h zuordnet:



[4]Quelle: `https://www.youtube.com/watch?v=vvpSiBUNDrc` (Zugriff: 19. Juli 2017)

Wendet man auf diesen Datensatz nun die zusammengesetzte Trapezregel mit n Teilintervallen an, so erhält man Näherungswerte für $\frac{1}{3600}\int_0^{69} f(x)\,dx$, was der Streckenlänge in Kilometern entspricht. Der Faktor $\frac{1}{3600}$ ist für die Umrechnung der Zeiteinheit notwendig. Da die Stützstellen im Allgemeinen nicht ganzzahlig sind, wurde jeweils auf die nächste ganze Zahl gerundet. Folgende Tabelle zeigt die ersten zehn Näherungswerte Q_n für jeweils n Teilintervalle.

n	1	2	3	4	5	6	7	8	9	10
Q_n	5.520	3.824	3.801	4.557	5.512	4.153	4.310	4.291	4.430	4.261

Insbesondere ist für $6 \leq n \leq 69$ das Maximum 4.47 km und das Minimum 4.06 km, was einer Spannweite von 0.41 km entspricht. Ab sieben Stützstellen, also für $n \geq 6$, bleibt der Näherungswert somit in einem relativ kleinen Intervall. Die tatsächliche Streckenlänge des Red Bull Rings beträgt 4.318 km und liegt somit in diesem Intervall. Es ist durchaus erstaunlich, wie exakt man diesen Wert mit lediglich sieben Stützstellen vorhersagen kann. Zusätzlich müssen hier folgende Unsicherheiten berücksichtigt werden:

- Die verwendeten Stützwerte für die Stützstellen stimmen oftmals nicht mit den tatsächlichen Werten überein, da die Stützstellen auf ganze Sekunden gerundet wurden und der zugehörige Geschwindigkeitswert genommen wurde.
- Die aus dem Video abgelesenen Messwerte (Geschwindigkeit und Zeit) weisen jeweils selbst eine gewisse Unsicherheit auf und müssen auch nicht exakt zusammenpassen. Beispielsweise könnte der Geschwindigkeitswert zeitverzögert sein.
- Die tatsächliche Rundenzeit betrug 68.759 s und nicht 69 s.

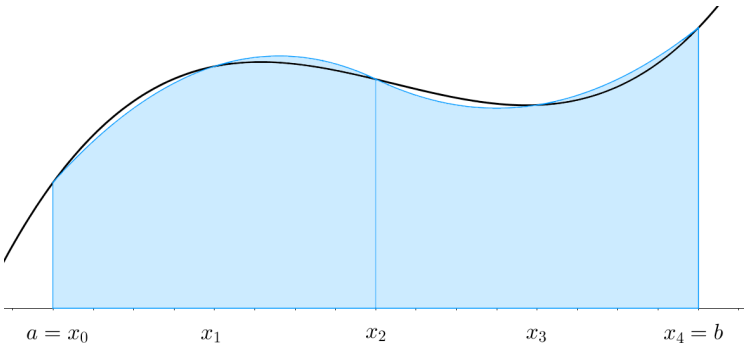
7.3 Simpsonregel

Mit der Simpsonregel (auch als Keplersche Fassregel bekannt) wird die Funktion über drei Stützstellen hinweg durch ein Polynom zweiten Grades interpoliert.

Definition 7.5. Sei $f : [a, b] \to \mathbb{R}$ die zu integrierende Funktion. Dann lautet die Simpsonregel

$$\int_a^b f(x)\,dx \approx (b-a) \cdot \left(\frac{1}{6} \cdot f(a) + \frac{2}{3} \cdot f\left(\frac{a+b}{2} \right) + \frac{1}{6} \cdot f(b) \right).$$

Nachfolgende Abbildung zeigt die Anwendung der zusammengesetzten Simpsonregel, wobei das Intervall in vier gleich große Teilintervalle zerlegt wurde, von denen jeweils zwei aufeinanderfolgende Teilintervalle für die Anwendung der Simpsonregel verwendet wurden.



Es ergibt sich somit, dass für die Anwendung der zusammengesetzten Simpsonregel das Intervall $[a, b]$ immer in eine gerade Anzahl an Teilintervallen zerlegt werden muss.

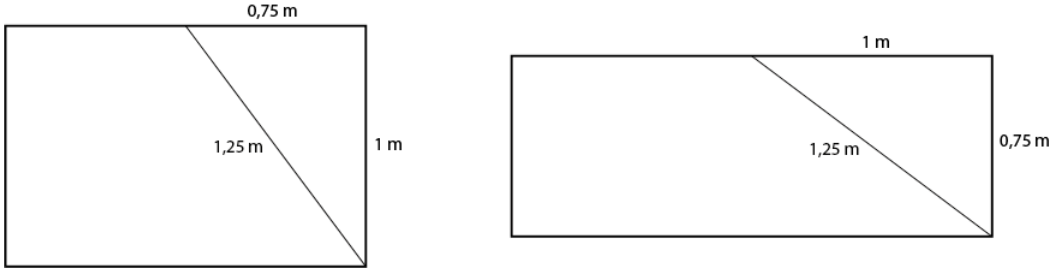
Historisches: Erstmals wurde die Simpsonregel von Evangelista Torricelli (Italien, 1608 – 1647) benutzt. Sie wurde jedoch nach Thomas Simpson (England, 1710 – 1761) benannt. Ihren Ursprung hat sie aber bereits etwas früher, als Johannes Kepler (Deutschland, 1571 – 1630) die sogenannte Keplersche Fassregel aufstellte.

Anstoß für die Untersuchung des Volumens von Weinfässern war für Kepler laut [24] die Bezahlung der Weinfässer für seine zweite Hochzeit im Jahr 1613. Er wunderte sich über die damals übliche Volumsbestimmung mittels der sogenannten Visiermethode. Dabei wurde eine Messrute durch das Spundloch bis zum gegenüberliegenden Ende der Fassböden gesteckt. Die dafür notwendige Länge galt als Maß für das Fassvolumen.

Nach [15, Seite 99 und 100] schrieb Kepler in der Widmung des Buches *Nova Stereometria doliorum vinariorum* (Neue Inhaltsberechnung von Weinfässern) Folgendes:

Als ich im November des letzten Jahres (1613) meine Wiedervermählung feierte, zu einer Zeit, da an den Donauufern bei Linz die aus Niederösterreich herbeigeführten Weinfässer nach einer reichlichen Lese aufgestapelt und zu einem annehmbaren Preise zu kaufen waren, da war es die Pflicht des neuen Gatten und sorglichen Familienvaters, für sein Haus den nötigen Trunk zu besorgen. Als einige Fässer eingekellert waren, kam am 4. Tage der Verkäufer mit der Meßrute, mit der er alle Fässer, ohne Rücksicht auf ihre Form, ohne jede weitere Überlegung oder Rechnung ihrem Inhalte nach bestimmte. Die Visierrute wurde mit ihrer metallenen Spitze durch das Spundloch quer bis zu den Rändern der beiden Böden eingeführt, und als die beiden Längen gleich gefunden worden waren, ergab die Marke am Spundloch die Zahl der Eimer im Fasse. Ich wunderte mich, daß die Querlinie durch die Faßhälfte ein Maß für den Inhalt abgeben könne, und bezweifelte die Richtigkeit der Methode, denn ein sehr niedriges Fass mit etwas breiteren Böden und daher sehr viel kleinerem Inhalt könne dieselbe Visierlänge besitzen. Es schien mir als Neuvermähltem nicht unzweckmäßig, ein neues Prinzip mathematischer Arbeiten, nämlich die Genauigkeit dieser bequemen und allgemein wichtigen Bestimmung nach geometrischen Grundsätzen zu erforschen und die etwa vorhandenen Gesetze ans Licht zu bringen.

Folgendes Beispiel zweier zylinderförmiger Fässer soll veranschaulichen, was Kepler mit dem vorletzten Satz des obigen Zitates andeutet:



Das linke Fass hat mit $\frac{3\pi}{8}\,\mathrm{m}^3$ ($\approx 1.18\,\mathrm{m}^3$) trotz gleicher Visierlänge ein wesentlich größeres Volumen als das rechte Fass mit $\frac{9\pi}{32}\,\mathrm{m}^3$ ($\approx 0.88\,\mathrm{m}^3$), nämlich etwa um ein Drittel größer. In gewisser Weise ist es erstaunlich, dass man zu dieser Zeit nicht einmal auf Unterschiede bei zylinderförmigen Fässern einging. Mit Keplers Formel war es später möglich das Volumen für Kreiszylinder, Kegel, Kegelstumpf, Kugel, Rotationsellipsoid, elliptisches Paraboloid und einschaliges Hyperboloid exakt zu berechnen.

Satz 7.6. Sei $f \in C^4[a,b]$. Für den Quadraturfehler $R_S(f)$ der Simpsonregel gilt

$$|R_S(f)| \leq \frac{h^4}{180} \max_{x\in[a,b]} |f^{(4)}(x)| \cdot (b-a).$$

Dabei ist $h := (b-a)/2n$ der Abstand zwischen zwei Stützstellen.

Beweis: Dieser Beweis wird unter Verwendung von [18, Satz 13.10] geführt. Es wird zunächst ein Teilintervall der Länge $2h$ mit den drei Stützstellen $x_i,\ x_{i+1} = x_i + h$ und $x_{i+2} = x_i + 2h$ untersucht, wobei i durch 2 teilbar ist. Es sei $\psi_f(x)$ das minimale Interpolationspolynom von f bezüglich dieser drei Stützstellen.

Im nächsten Schritt definiert man die kubische Interpolierende ρ von f folgendermaßen:

$$\rho(x) := \psi_f(x) + \frac{4}{(x_{i+2}-x_i)^2} \cdot \big(\psi_f'(x_{i+1}) - f'(x_{i+1})\big) \cdot (x-x_i) \cdot (x-x_{i+1}) \cdot (x-x_{i+2})$$

Für die Ableitung ρ' erhält man für $S(x) := (x-x_{i+1})\cdot(x-x_{i+2}) + (x-x_i)\cdot(x-x_{i+2}) + (x-x_i)\cdot(x-x_{i+1})$ folgende Funktion:

$$\rho'(x) = \psi_f'(x) + \frac{4}{(x_{i+2}-x_i)^2} \cdot \big(\psi_f'(x_{i+1}) - f'(x_{i+1})\big) \cdot S(x)$$

Für $k \in \{0,1,2\}$ gelten die Interpolationseigenschaften $\rho(x_{i+k}) = \psi_f(x_{i+k}) = f(x_{i+k})$, weil $(x_{i+k}-x_i)\cdot(x_{i+k}-x_{i+1})\cdot(x_{i+k}-x_{i+2}) = 0$ gilt und somit der zweite Summand der Definition von ρ wegfällt. Daher handelt es sich tatsächlich um ein kubisches Interpolationspolynom von f bezüglich der Stützstellen $x_i,\ x_{i+1}$ und x_{i+2}. Die Eigenschaft $\rho'(x_{i+1}) = f'(x_{i+1})$ erhält man aus folgender Umformung:

$$\rho'(x_{i+1}) = \psi_f'(x_{i+1}) + \underbrace{\frac{4}{(x_{i+2}-x_i)^2}}_{=\frac{4}{(2h)^2}} \cdot \big(\psi_f'(x_{i+1}) - f'(x_{i+1})\big) \cdot \underbrace{(x_{i+1}-x_i)}_{=h} \cdot \underbrace{(x_{i+1}-x_{i+2})}_{=-h} =$$

$$= \psi_f'(x_{i+1}) - \frac{4h^2}{4h^2}\big(\psi_f'(x_{i+1}) - f'(x_{i+1})\big) = \psi_f'(x_{i+1})$$

Eine Nebenrechnung mit MATHEMATICA[5] ergibt, dass $\int_{x_i}^{x_{i+2}} (x-x_i)\cdot(x-x_{i+1})\cdot(x-x_{i+2})\,dx = 0$ gilt. Daraus folgt $\int_{x_i}^{x_{i+2}} \rho(x)\,dx = \int_{x_i}^{x_{i+2}} \psi_f(x)\,dx$, da die Faktoren vor $(x-x_i)\cdot(x-x_{i+1})\cdot(x-x_{i+2})$ in $\rho(x)$ konstant ist.

Man definiert nun $g(x) := (x-x_i)\cdot(x-x_{i+1})^2\cdot(x-x_{i+2})$. Dann gilt:

$$g(x) = \underbrace{(x-x_i)}_{\geq 0}\cdot\underbrace{(x-x_{i+1})^2}_{\geq 0}\cdot\underbrace{(x-x_{i+2})}_{\leq 0} \leq 0,\ \forall x \in [x_i, x_{i+2}]$$

Somit erhält man:

$$r_S(f) := \int_{x_i}^{x_{i+2}} (f(x)-\psi_f(x))\,dx = \int_{x_i}^{x_{i+2}} (f(x)-\rho(x))\,dx = \int_{x_i}^{x_{i+2}} g(x)\,\frac{f(x)-\rho(x)}{g(x)}\,dx$$

Die Funktion $g(x)$ ist ein Polynom und daher nach Satz 2.2 stetig und deshalb nach Satz 2.8 Riemann-integrierbar. Außerdem gilt $g(x) \leq 0,\ \forall x \in [x_i, x_{i+2}]$, also es gibt keinen Vorzeichenwechsel.

Nun wird gezeigt, dass der Bruch

$$z(x) := \frac{f(x)-\rho(x)}{g(x)}$$

im Integranden stetig ist. Der Nenner besitzt offensichtlich die Nullstellen $x_i,\ x_{i+1}$ und x_{i+2}. Somit ist z auf (x_i, x_{i+1}) und auf (x_{i+1}, x_{i+2}) stetig, weshalb nur noch die Stellen $x_i,\ x_{i+1}$ und x_{i+2} überprüft werden müssen. Wegen $\rho(x_{i+k}) = f(x_{i+k})$ für $k \in \{0,1,2\}$ besitzt der Zähler die Nullstellen $x_i,\ x_{i+1}$ und x_{i+2}. Laut MATHEMATICA[6] sind die Nullstellen von g' die Stellen $x_{i+1} - \frac{h}{\sqrt{2}},\ x_{i+1}$ und $x_{i+1} + \frac{h}{\sqrt{2}}$.

Wegen $g'(x) \neq 0$ für $x \in \{x_i, x_{i+2}\}$ existieren die Grenzwerte

$$\lim_{x\to x_i^+} \frac{(f(x)-\rho(x))'}{g'(x)} \qquad \text{und} \qquad \lim_{x\to x_{i+2}^-} \frac{(f(x)-\rho(x))'}{g'(x)}.$$

Somit folgt aus Satz 2.14 (i) und (iii), dass die einseitigen Grenzwerte

$$\lim_{x\to x_i^+} z(x) = \lim_{x\to x_i^+} \frac{f(x)-\rho(x)}{g(x)} \qquad \text{und} \qquad \lim_{x\to x_{i+2}^-} z(x) = \lim_{x\to x_{i+2}^-} \frac{f(x)-\rho(x)}{g(x)}$$

existieren. Daher ist z in x_i rechtsseitig stetig und in x_{i+2} linksseitig stetig. Für x_{i+1} gilt

$$\lim_{x\to x_{i+1}^+} \frac{f(x)-\rho(x)}{g(x)} = \lim_{x\to x_{i+1}^+} \frac{f'(x)-\rho'(x)}{g'(x)} = \lim_{x\to x_{i+1}^-} \frac{f'(x)-\rho'(x)}{g'(x)} = \lim_{x\to x_{i+1}^-} \frac{f(x)-\rho(x)}{g(x)}.$$

Die Gleichheit der beiden Grenzwerte folgt daraus, dass sowohl der Zähler als auch der Nenner stetig sind. Außerdem gilt $f'(x_{i+1}) - \rho'(x_{i+1}) = 0$ und $g'(x_{i+1}) = 0$. Daher kann erneut die Regel von de l'Hospital (Satz 2.14) angewendet werden und man erhält

[5] Eingabe: `Integrate[(x - x0) (x - (x0 + h)) (x - (x0 + 2 h)), {x, x0, x0 + 2 h}]`
[6] Eingabe: `Simplify[Solve[D[(x - x0) (x - (x0 + h))^2 (x - (x0 + 2 h)), x] == 0, x]]`

$$\lim_{x\to x_{i+1}^+} \frac{f'(x)-\rho'(x)}{g'(x)} = \lim_{x\to x_{i+1}^+} \frac{f''(x)-\rho''(x)}{g''(x)} = \lim_{x\to x_{i+1}^-} \frac{f''(x)-\rho''(x)}{g''(x)} = \lim_{x\to x_{i+1}^-} \frac{f'(x)-\rho'(x)}{g'(x)}.$$

Die Gleichheit der Grenzwerte folgt erneut aus der Stetigkeit von Zähler und Nenner. Eine Nebenrechnung mit MATHEMATICA[7] ergibt außerdem, dass $g''(x_{i+1}) \neq 0$ ist, weshalb dieser Grenzwert existiert. Daher existiert nach der Regel von de l'Hospital (Satz 2.14) auch der Grenzwert $\lim_{x\to x_{i+1}} z(x)$.

Somit wurde gezeigt, dass z stetig in x_{i+1} und damit stetig auf $[x_i, x_{i+2}]$ ist. Aus dem Mittelwertsatz der Integralrechnung (Satz 2.13) folgt nun, dass es ein $\eta \in [x_i, x_{i+2}]$ gibt, sodass gilt:

$$\int_{x_i}^{x_{i+2}} g(x)\,\frac{f(x)-\rho(x)}{g(x)}\,dx = \frac{f(\eta)-\rho(\eta)}{g(\eta)} \int_{x_i}^{x_{i+2}} g(x)\,dx$$

Nach Satz 3.16 gibt es für alle $x \in [x_i, x_{i+2}]$, und damit auch für η, ein $\xi \in [x_i, x_{i+2}]$, sodass für die kubische Interpolierende ρ, welche inklusive Vielfachheiten vier Schnittstellen (Stützstellen) mit f hat[8], Folgendes gilt:

$$f(\eta)-\rho(\eta) = \frac{f^{(4)}(\xi)}{4!}\cdot g(\eta)$$

Eine Nebenrechnung mittels MATHEMATICA[9] liefert

$$\int_{x_i}^{x_{i+2}} g(x)\,dx = -\frac{(x_{i+2}-x_i)^5}{120}.$$

Daraus folgt letztendlich

$$r_S(f) = -\frac{\frac{f^{(4)}(\xi)}{4!}\cdot g(\eta)}{g(\eta)} \cdot \frac{\overbrace{(x_{i+2}-x_i)^5}^{=32h^5}}{120} = -\frac{f^{(4)}(\xi)\cdot h^5}{90}.$$

Es gilt also

$$|r_S(f)| \leq \frac{|f^{(4)}(\xi)|\cdot h^5}{90}.$$

Für den Gesamtfehler $R_S(f)$ aller Teilintervalle erhält man mit $h := (b-a)/2n$ folgende Abschätzung:

$$|R_S(f)| \leq \underbrace{\frac{h^5}{90}}_{=\frac{h^4}{90}\cdot\frac{b-a}{2n}} \cdot \underbrace{\sum_{i=0}^{n-1} |f^{(4)}(\xi_i)|}_{\leq n\cdot \max\limits_{x\in[a,b]}|f^{(4)}(x)|} \leq \frac{h^4}{180} \max_{x\in[a,b]} |f^{(4)}(x)|\cdot(b-a)$$

□

[7] Eingabe: `Simplify[Solve[D[(x - x0) (x - (x0 + h))^2 (x - (x0 + 2 h)), x, 2] == 0, x]]`
[8] nämlich x_i, x_{i+1} und x_{i+2}, wobei x_{i+1} doppelt vorkommt, da auch $\rho'(x_{i+1}) = f'(x_{i+1})$ gilt
[9] Eingabe: `Integrate[(x - x0) (x - (x0 + x2)/2)^2 (x - x2), {x, x0, x2}]`

7.4 Allgemeines Modell

Die in den vorherigen Kapiteln vorgestellten Quadraturformeln lassen sich in ein allgemeines Modell eingliedern, nämlich in die Klasse der sogenannten Newton-Cotes-Formeln. Hierbei unterscheidet man zwischen abgeschlossenen Newton-Cotes-Formeln, offenen Newton-Cotes-Formeln und den sogenannten Maclaurin-Formeln. Der Unterschied liegt in der Position der Stützstellen. Das gesamte Intervall $[a, b]$ wird dabei jeweils in Teilintervalle gleicher Länge zerlegt. Bei den abgeschlossenen Formeln sind die Intervallgrenzen zugleich die Stützstellen. Bei den offenen Formeln werden die beiden äußeren Grenzen, also a und b, nicht als Stützstellen verwendet. Bei den Maclaurin-Formeln wird die Mitte jedes Teilintervalls als Stützstelle verwendet.

Definition 7.7. Man nennt für ein $n \in \mathbb{N}$ die Interpolationsquadratur n-ter Ordnung mit den Stützstellen $\tilde{x}_k = a + kh$ für $k = 0, ..., n$ abgeschlossene Newton-Cotes-Formel n-ter Ordnung. Dabei ist $h = \frac{b-a}{n}$ die sogenannte Schrittweite.

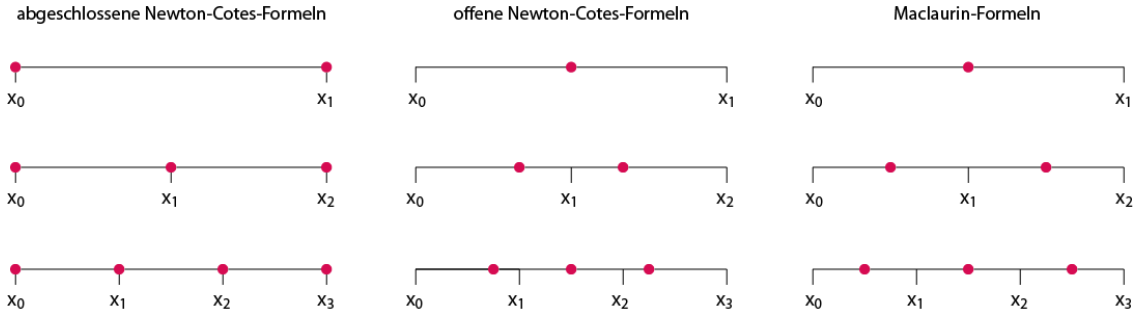
Definition 7.8. Man nennt für ein $n \in \mathbb{N}_0$ die Interpolationsquadratur n-ter Ordnung mit den Stützstellen $\tilde{x}_k = a + \frac{k+1}{n+2} \cdot (b - a)$ für $k = 0, ..., n$ offene Newton-Cotes-Formel n-ter Ordnung.

Definition 7.9. Man nennt für ein $n \in \mathbb{N}_0$ die Interpolationsquadratur n-ter Ordnung mit den Stützstellen $\tilde{x}_k = a + \frac{2k+1}{2n+2} \cdot (b - a)$ für $k = 0, ..., n$ Maclaurin-Quadraturformel n-ter Ordnung.

Bemerkung 1: Die offenen Newton-Cotes-Formeln werden nach [23] auch Steffensen-Formeln genannt.

Bemerkung 2: Die Maclaurin-Formeln wurden nach Colin Maclaurin (Schottland, 1698 – 1746) benannt.

Demnach ist die Mittelpunktregel die offene Newton-Cotes-Formel 0-ter Ordnung (und auch die Maclaurin-Formel 0-ter Ordnung), die Sehnentrapezregel die abgeschlossene Newton-Cotes-Formel erster Ordnung und die Simpsonregel die abgeschlossene Netwon-Cotes-Formel zweiter Ordnung.



Auf die Maclaurin-Formeln wird im weiteren Verlauf nicht mehr eingegangen. Nachfolgend werden Berechnungsmethoden hergeleitet, mit denen für Newton-Cotes-Formeln beliebiger

Ordnung die Gewichte ermittelt und die Fehler abgeschätzt werden können. Zunächst werden nun Formeln hergeleitet, mit welchen die Gewichte unabhängig vom Intervall $[a, b]$ berechnet werden können.

Satz 7.10. Sei $n \in \mathbb{N}$. Dann gilt für die Gewichte der abgeschlossenen Newton-Cotes-Formel n-ter Ordnung:

$$w_i = \frac{1}{n} \int_0^n \prod_{\substack{k=0 \\ k \neq i}}^{n} \frac{u-k}{i-k} \, du, \qquad \forall i \in \{0, ..., n\}$$

Beweis: Es ist $\tilde{x}_k = a + kh$ und $\tilde{x}_i = a + ih$. Außerdem folgt aus $h = \frac{b-a}{n}$, dass $b = a + nh$ und $\frac{h}{b-a} = \frac{1}{n}$ gelten. Somit erhält man nach Satz 6.3:

$$\frac{1}{b-a} \int_a^b \prod_{\substack{k=0 \\ k \neq i}}^{n} \frac{x - \tilde{x}_k}{\tilde{x}_i - \tilde{x}_k} \, dx = \frac{1}{b-a} \int_a^{a+nh} \prod_{\substack{k=0 \\ k \neq i}}^{n} \frac{\overbrace{x - (a + kh)}^{=x-a-kh}}{\underbrace{(a + ih) - (a + kh)}_{=h(i-k)}} \, dx =$$

$$= \frac{1}{b-a} \int_a^{a+nh} \prod_{\substack{k=0 \\ k \neq i}}^{n} \frac{x - a - kh}{h(i-k)} \, dx$$

Im nächsten Schritt wird $u := \frac{x-a}{h}$ substituiert. Somit gilt $x = uh + a$ und $dx = h \cdot du$. Die neuen Integralgrenzen sind nun $\frac{a-a}{h} = 0$ und $\frac{(a+nh)-a}{h} = n$. Man erhält schließlich:

$$\frac{1}{b-a} \int_0^n \prod_{\substack{k=0 \\ k \neq i}}^{n} \frac{\overbrace{uh + a - a - kh}^{h(u-k)}}{h(i-k)} \cdot h \, du = \frac{1}{n} \int_0^n \prod_{\substack{k=0 \\ k \neq i}}^{n} \frac{u-k}{i-k} \, du$$

□

Für die Berechnung der Gewichte der abgeschlossenen Newton-Cotes-Formel n-ter Ordnung kann folgender MATHEMATICA-Code verwendet werden:

```
p[i_, n_] := Product[If[k != i, (x - k)/(i - k), 1], {k, 0, n}]

w[i_, n_] := Simplify[1/n*Integrate[p[i, n], {x, 0, n}]]

t[n_] := Table[w[i, n], {i, 0, n}]
```

Der nächste Satz widmet sich der Tatsache, dass die Gewichte der abgeschlossenen Newton-Cotes-Formeln symmetrisch sind.

Satz 7.11. Sei $n \in \mathbb{N}$. Dann gilt für die Gewichte der abgeschlossenen Newton-Cotes-Formel n-ter Ordnung $w_{n-i} = w_i$ für alle $i \in \{0, ..., n\}$.

Beweis: Für den Beweis wurde [4, Seite 106, Lemma 5.10] verwendet. Sei $i \in \{0, ..., n\}$. Dann gilt laut Satz 7.10:

$$w_{n-i} = \frac{1}{n} \int_0^n \prod_{\substack{k=0 \\ k \neq n-i}}^{n} \frac{u-k}{(n-i)-k}\, du.$$

Im Produkt wird nun die Substitution $j := n - k$ durchgeführt. Somit gilt $k = n - j$, und für Start- und Endwert des Produktes erhält man $n - 0 = n$ und $n - n = 0$. Außerdem folgt aus $k \neq n - i \Rightarrow n - j \neq n - i \Rightarrow j \neq i$. Man erhält:

$$w_{n-i} = \frac{1}{n} \int_0^n \prod_{\substack{j=n \\ j \neq i}}^{0} \underbrace{\frac{u-(n-j)}{(n-i)-(n-j)}}_{= \frac{u-n+j}{-i+j} \cdot \overset{(-1)}{=} \frac{(n-u)-j}{i-j}} du = \frac{1}{n} \int_0^n \prod_{\substack{j=n \\ j \neq i}}^{0} \frac{(n-u)-j}{i-j}\, du.$$

Fasst man das Produkt als Funktion in Abhängigkeit von $n-u$ auf, so erhält man durch die Substitution $v = \varphi(u) = n - u$ und $\varphi'(u) = -1$ unter Verwendung von Satz 2.15 folgendes Resultat:

$$\int_0^n f(n-u)\, du = (-1) \cdot \int_0^n f(\varphi(u)) \cdot \underbrace{\varphi'(u)}_{=-1} du = -\int_{\varphi(0)}^{\varphi(n)} f(v)\, dv = -\int_n^0 f(v)\, dv$$

Mit der Eigenschaft $\int_a^b f(x)\, dx = -\int_b^a f(x)\, dx$ erhält man insgesamt

$$w_{n-i} = \frac{1}{n} \int_0^n \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{v-j}{i-j}\, dv = w_i.$$

□

In Anlehnung an die vorherigen Sätze (Satz 7.10 und Satz 7.11) wird nun eine Formel für die Berechnung der Gewichte der offenen Newton-Cotes-Formeln hergeleitet und gezeigt, dass diese ebenfalls symmetrisch sind.

Satz 7.12. Sei $n \in \mathbb{N}_0$. Dann gilt für die Gewichte der offenen Newton-Cotes-Formel n-ter Ordnung:

$$w_i = \frac{1}{n+2} \int_{-1}^{n+1} \prod_{\substack{k=0 \\ k \neq i}}^{n} \frac{u-k}{i-k}\, du, \qquad \forall i \in \{0, ..., n\}$$

Beweis: Für $i \in \{0, ..., n\}$ gilt nach Definition 7.8 für die Stützstellen der offenen Newton-Cotes-Formeln $\tilde{x}_i = a + \frac{i+1}{n+2}(b-a)$. Außerdem ist $b - a = (n+1) \cdot h \Leftrightarrow h = \frac{b-a}{n+1} \Leftrightarrow b = a + h \cdot (n+1)$. Damit erhält man ausgehend von Satz 6.3

$$w_i = \frac{1}{b-a} \int_a^b \prod_{\substack{k=0 \\ k \neq i}}^{n} \frac{x - \tilde{x}_k}{\tilde{x}_i - \tilde{x}_k}\, dx =$$

$$= \frac{1}{b-a} \int_a^{a+h\cdot(n+1)} \prod_{\substack{k=0 \\ k \neq i}}^{n} \frac{x - (a + \frac{k+1}{n+2}(b-a))}{(a + \frac{i+1}{n+2}(b-a)) - (a + \frac{k+1}{n+2}(b-a))}\, dx.$$

Nach Vereinfachung des Bruches im Produkt erhält man

$$\frac{1}{b-a}\int_a^{a+h\cdot(n+1)} \prod_{\substack{k=0\\k\neq i}}^{n} \frac{x-a-(k+1)\frac{b-a}{n+2}}{(i-k)\frac{b-a}{n+2}}\,dx.$$

Nun wird $u := \frac{x-a-\frac{b-a}{n+2}}{\frac{b-a}{n+2}}$ substituiert. Es ist dann $x = u\frac{b-a}{n+2} + a + \frac{b-a}{n+2}$ und $dx = \frac{b-a}{n+2}\,du$. Außerdem gilt für die Integralgrenzen

$$\frac{a-a-\frac{b-a}{n+2}}{\frac{b-a}{n+2}} = -1$$

und

$$\frac{a+(n+1)\cdot\overbrace{h}^{=\frac{b-a}{n+1}}-a-\frac{b-a}{n+2}}{\frac{b-a}{n+2}} = \frac{(b-a)-\frac{b-a}{n+2}}{\frac{b-a}{n+2}} = \frac{(b-a)\frac{n+2}{b-a}-1}{1} = n+2-1 = n+1.$$

Durch Einsetzen in obigen Ausdruck erhält man

$$\frac{1}{b-a}\int_{-1}^{n+1} \prod_{\substack{k=0\\k\neq i}}^{n} \frac{u\frac{b-a}{n+2}+a+\frac{b-a}{n+2}-a-(k+1)\frac{b-a}{n+2}}{(i-k)\frac{b-a}{n+2}}\,\frac{b-a}{n+2}\,du.$$

Durch Vereinfachen erhält man schließlich

$$\frac{1}{n+2}\int_{-1}^{n+1} \prod_{\substack{k=0\\k\neq i}}^{n} \frac{u-k}{i-k}\,du.$$

□

Satz 7.13. Sei $n \in \mathbb{N}_0$. Dann gilt für die Gewichte der offenen Newton-Cotes-Formel n-ter Ordnung $w_{n-i} = w_i$ für alle $i \in \{0, ..., n\}$.

Beweis: Der Beweis erfolgt in Analogie zu jenem von Satz 7.11. □

Satz 7.14. (Satz von Kusmin) Laut [4, Seite 113, Satz 5.16] gilt für die Gewichte $w_0, ..., w_n$ der abgeschlossenen Newton-Cotes-Formeln

$$\lim_{n\to\infty}\sum_{k=0}^{n}|w_k| = \infty.$$

Die Summe der Beträge der Gewichte wächst also im Vergleich zur Summe der Gewichte (welche laut Satz 6.8 immer 1 ist), wodurch es nach [4, Seite 113] bei ansteigendem n zu Instabilitäten kommt.

Die nächsten Seiten widmen sich einer allgemeinen Fehlerabschätzung der abgeschlossenen Newton-Cotes-Formeln.

Definition 7.15. Sei $k \in \mathbb{N}_0$. Dann definiere

$$(x)_+^k := \begin{cases} x^k, & \text{falls } x \geq 0, \\ 0, & \text{falls } x < 0. \end{cases}$$

Satz 7.16. Sei $m \in \mathbb{N}, m \geq 1$, und sei $f : [0,1] \to \mathbb{R}$ eine m-mal stetig differenzierbare Funktion. Sei $Q(f) = \sum_{i=0}^{n} w_i f(x_i)$ eine Quadraturformel mit Fehlerordnung m und $R(f) = \int_0^1 f(x)\,dx - Q(f)$ der zugehörige Quadraturfehler. Dann gibt es eine Funktion $K_m : [0,1] \to \mathbb{R}$, sodass gilt:

$$R(f) = \int_0^1 K_m(x) f^{(m)}(x)\,dx$$

Beweis: Für den Beweis wurden [27, Seite 52, Satz 3.10] und [28, Seite 61, Satz 30] verwendet. Die m-mal stetig differenzierbare Funktion f kann als Taylorpolynom (siehe Definitionen 2.26 und 2.29) mit Entwicklungsstelle $\tilde{x} = 0$ folgendermaßen dargestellt werden:

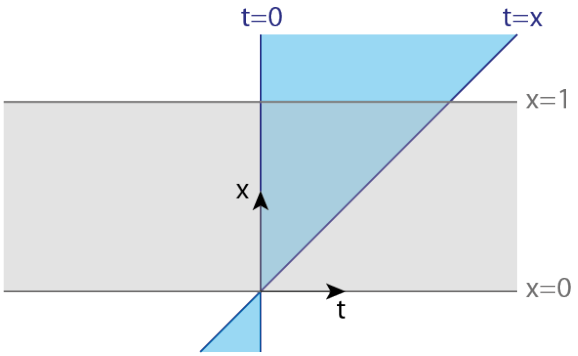
$$f(x) = \underbrace{\sum_{k=0}^{m-1} \frac{f^{(k)}(0)}{k!} x^k}_{=:p(x)} + \underbrace{\frac{1}{(m-1)!} \int_0^x (x-t)^{m-1} f^{(m)}(t)\,dt}_{=:q(x)}$$

Es ist $p \in \mathcal{P}_{m-1}$, und daher wird p durch die Quadraturformel Q exakt integriert, also $R(p) = 0$. Es gilt somit $R(f) = \underbrace{R(p)}_{=0} + R(q) = R(q)$. Durch Einsetzen erhält man:

$$R(f) = \int_0^1 \frac{1}{(m-1)!} \int_0^x f^{(m)}(t)(x-t)^{m-1}\,dt\,dx - \sum_{i=0}^{n} w_i \frac{1}{(m-1)!} \int_0^{x_i} f^{(m)}(t)(x_i-t)^{m-1}\,dt =$$

$$= \frac{1}{(m-1)!} \left(\int_0^1 \int_0^x f^{(m)}(t)(x-t)^{m-1}\,dt\,dx - \sum_{i=0}^{n} w_i \int_0^{x_i} f^{(m)}(t)(x_i-t)^{m-1}\,dt \right)$$

Nachfolgende Abbildung soll den Tausch der beiden Integrationsvariablen im Doppelintegral veranschaulichen. Insgesamt wird über den doppelt hinterlegten dreieckigen Bereich integriert. Aktuell werden horizontale Streifen summiert und nach der Vertauschung werden vertikale Streifen summiert.



Man erhält somit

$$\frac{1}{(m-1)!}\left(\int_0^1\int_t^1 f^{(m)}(t)(x-t)^{m-1}\,dx\,dt-\sum_{i=0}^{n}w_i\int_0^{x_i}f^{(m)}(t)(x_i-t)^{m-1}\,dt\right).$$

Unter Verwendung von Definition 7.15 kann man die Obergrenze des zweiten Integrals von x_i auf 1 setzen, wodurch die Integrationsgrenzen von i unabhängig sind. Außerdem ist im ersten Integral $f^{(m)}(t)$ unabhängig von x.

$$\frac{1}{(m-1)!}\left(\int_0^1 f^{(m)}(t)\int_t^1(x-t)^{m-1}\,dx\,dt-\sum_{i=0}^{n}w_i\int_0^1 f^{(m)}(t)(x_i-t)_+^{m-1}\,dt\right)$$

Da bei endlichen Summen das Integral- und das Summenzeichen vertauscht werden dürfen, erhält man

$$\begin{aligned}\frac{1}{(m-1)!}\left(\int_0^1 f^{(m)}(t)\int_t^1(x-t)^{m-1}\,dx\,dt-\int_0^1 f^{(m)}(t)\sum_{i=0}^{n}w_i(x_i-t)_+^{m-1}\,dt\right)&=\\ =\int_0^1\frac{1}{(m-1)!}\left(\int_t^1(x-t)^{m-1}\,dx-\sum_{i=0}^{n}w_i(x_i-t)_+^{m-1}\right)f^{(m)}(t)\,dt.&\end{aligned}$$

Mit MATHEMATICA[10] erhält man $\int_t^1(x-t)^{m-1}\,dx=\frac{(1-t)^m}{m}$. Somit folgt insgesamt:

$$R(f)=\int_0^1\underbrace{\frac{1}{(m-1)!}\left(\frac{(1-t)^m}{m}-\sum_{i=0}^{n}w_i(x_i-t)_+^{m-1}\right)}_{=:K_m(t)}f^{(m)}(t)\,dt.$$

Damit existiert eine Funktion K_m, welche die geforderte Bedingung erfüllt. □

Definition 7.17. Sei $m\in\mathbb{N}$. Die Funktion K_m aus Satz 7.16 bezeichnet man als Peano-Kern, benannt nach Giuseppe Peano (Italien, 1858 – 1932).

Korollar 7.18. Es gilt folgende Fehlerabschätzung:

$$|R(f)|\le\max_{x\in[0,1]}|f^{(m)}(x)|\cdot\int_0^1|K_m(t)|\,dt$$

Beweis: Aus Satz 7.16 erhält man

$$|R(f)|=\left|\int_0^1 K_m(t)\underbrace{f^{(m)}(t)}_{\le\max_{x\in[0,1]}f^{(m)}(x)}dt\right|\le\left|\max_{x\in[0,1]}f^{(m)}(x)\cdot\int_0^1K_m(t)\,dt\right|=$$

[10] Eingabe: `Integrate[(x-t)^(m-1),{x,t,1}, Assumptions -> m ≥ 1]`

$$= \max_{x\in[0,1]} |f^{(m)}(x)| \cdot \underbrace{\left| \int_0^1 K_m(t)\,dt \right|}_{\leq \int_0^1 |K_m(t)|\,dt} \leq \max_{x\in[0,1]} |f^{(m)}(x)| \cdot \int_0^1 |K_m(t)|\,dt.$$

□

Satz 7.19. Gilt $K_m(x) \geq 0$, $\forall x \in [0,1]$ oder $K_m(x) \leq 0$, $\forall x \in [0,1]$, dann gibt es ein $\xi \in [0,1]$, sodass man für den Fehler $R(f)$ folgende Formel erhält:

$$R(f) = f^{(m)}(\xi) \underbrace{\int_0^1 K_m(x)\,dx}_{=:c_m}$$

Beweis: Weil f nach Voraussetzung m-mal stetig differenzierbar ist, ist $f^{(m)}$ stetig. Außerdem besitzt die Funktion $K_m(x)$ nach Voraussetzung keinen Vorzeichenwechsel und ist gemäß ihrer Definition integrierbar. Somit gibt es laut Mittelwertsatz der Integralrechnung (Satz 2.13) ein $\xi \in [0,1]$, sodass die Behauptung stimmt. □

Definition 7.20. Hat der Peano-Kern keinen Vorzeichenwechsel, so definiert man

$$c_m := \int_0^1 K_m(x)\,dx$$

laut [27, Seite 53, Definition 3.14] als Fehlerkonstante der zugehörigen Quadraturformel mit Fehlerordnung m.

Folgender Satz wird eine sehr einfache Möglichkeit zur Berechnung der Fehlerkonstante c_m liefern, für welche man den Peano-Kern nicht explizit kennen muss. Man muss lediglich wissen, dass dieser auf $[0,1]$ keinen Vorzeichenwechsel hat.

Satz 7.21. Gilt $K(x) \geq 0$, $\forall x \in [0,1]$ oder $K(x) \leq 0$, $\forall x \in [0,1]$, dann kann die Fehlerkonstante c_m einer Quadraturformel mit Fehlerordnung m durch die Formel

$$c_m = \frac{R(x^m)}{m!}$$

berechnet werden.

Beweis: Als Grundlage für diesen Beweis wurde [27, Seite 54] verwendet. Man setze $f(x) := x^m$. Nach Satz 2.7 ist f m-mal stetig differenzierbar und erfüllt damit die Voraussetzung. Nach Korollar 2.5 ist $f^{(m)}(x) = m!$, $\forall x \in [0,1]$. Damit erhält man mit Satz 7.19 und Definition 7.20

$$R(x^m) = m! \cdot c_m.$$

Weil c_m nicht von f sondern nur von der gewählten Quadraturformel Q abhängig ist, gilt somit allgemein $c_m = \frac{R(x^m)}{m!}$. □

Nun fehlt bei der Fehlerabschätzung nur noch der Übergang von Funktionen mit Definitionsmenge $[0,1]$ zu Funktionen mit Definitionsmenge $[a,b]$.

Satz 7.22. Sei $f : [a,b] \to \mathbb{R}$ eine Funktion und seien $n, N \in \mathbb{N}$. Weiters sei $Q(f) = \sum_{i=0}^{n} w_i f(x_i)$ eine abgeschlossene Newton-Cotes-Formel mit Fehlerordnung m, für welche

Satz 7.19 verwendbar ist. Das Intervall $[a, b]$ wird in N gleich große Teilintervalle zerlegt, auf welche jeweils die Quadraturformel Q angewendet werden soll. Setze dazu $h := \frac{b-a}{N \cdot n}$. Dann gilt:

$$|R(f)| \leq (b-a) \cdot n^m h^m \cdot |c_m| \cdot \max_{x \in [a,b]} |f^{(m)}(x)|$$

Beweis: Für diesen Beweis wurde [27, Seite 54 und 55] herangezogen. Zunächst wird nur das erste der N Teilintervalle, also $[a, a + nh]$, betrachtet. Wegen Satz 2.15 gilt für $x = \varphi(t) = a + nht$ und $\dot{\varphi}(t) = nh$ der folgende Zusammenhang:

$$nh \int_0^1 f(a + nht)\, dt = \frac{nh}{nh} \int_0^1 f(\varphi(t)) \cdot \underbrace{\dot{\varphi}(t)}_{=nh}\, dt = \int_{\varphi(0)}^{\varphi(1)} f(x)\, dx = \int_a^{a+nh} f(x)\, dx$$

Somit erhält man

$$\int_a^{a+nh} f(x)\, dx = nh \int_0^1 \underbrace{f(a + nht)}_{=:g(t)}\, dt = nh \int_0^1 g(t)\, dt.$$

Die Intervallgrenzen sind nun also 0 und 1. Die Quadraturformel für das erste Teilintervall lautet unter Verwendung der Stützstellen $x_i = \frac{i}{n}$

$$Q(f) = nh \cdot Q(g) = nh \sum_{i=0}^{n} w_i g(x_i).$$

Den Fehler $r_1(f)$ des ersten Teilintervalls erhält man nun folgendermaßen:

$$r_1(f) = \int_a^{a+nh} f(x)\, dx - Q(f) = nh \left(\int_0^1 g(t)\, dt - Q(g) \right) = nh \cdot r_1(g)$$

Es ist $g(t) = f(x(t)) = f(a + nht)$. Daher gilt laut Korollar 2.19:

$$\frac{d^m g}{dt^m} = \frac{d^m f}{dt^m} = \frac{d^m f}{dx^m} \underbrace{\left(\frac{dx}{dt} \right)^m}_{=nh} = f^{(m)}(x) \cdot (nh)^m$$

Unter Verwendung von Satz 7.19 existiert somit ein $\xi_1 \in [a, a + nh]$, sodass gilt $r_1(g) = f^{(m)}(\xi_1) \cdot (nh)^m \cdot c_m$. Daher folgt $r_1(f) = f^{(m)}(\xi_1) \cdot (nh)^{m+1} \cdot c_m$. Es wird nun der Gesamtfehler $R(f)$ untersucht:

$$|R(f)| = \left| \sum_{i=1}^{N} \left(\int_{a+(i-1)nh}^{a+inh} f(x)\, dx - Q(f) \right) \right| \leq \sum_{i=1}^{N} |r_i(f)| \leq$$

$$\leq \sum_{i=1}^{N} (nh)^{m+1} \cdot |c_m| \cdot \underbrace{|f^{(m)}(\xi_i)|}_{\leq \max_{x \in [a,b]} |f^{(m)}(x)|} \leq$$

$$\leq \underbrace{N(nh)^{m+1}}_{=Nnh(nh)^m=(b-a)(nh)^m} \cdot |c_m| \cdot \max_{x\in[a,b]} |f^{(m)}(x)| = (b-a)\, n^m h^m \cdot |c_m| \cdot \max_{x\in[a,b]} |f^{(m)}(x)|$$

□

Korollar 7.23. Für $N = 1$, also die einmalige Anwendung der Quadraturformel auf das gesamte Intervall $[a, b]$, gilt

$$|R(f)| \leq n^{m+1} h^{m+1} \cdot |c_m| \cdot \max_{x\in[a,b]} |f^{(m)}(x)|.$$

Beweis: Dies folgt sofort aus $b - a = nh$. □

Bemerkung: Prinzipiell könnten Satz 7.16, Korollar 7.18 und Satz 7.19 auch verwendet werden, um den Fehler der offenen Newton-Cotes-Formeln oder anderer Quadraturformeln abzuschätzen, sofern die Voraussetzungen erfüllt sind. Beispielsweise muss die Funktion ausreichend oft differenzierbar sein und für Satz 7.19 darf der Peano-Kern K_m keinen Vorzeichenwechsel aufweisen. Bei der Beweisführung von Satz 7.22 müssten dann jedoch die Stützstellen und Teilintervallgrenzen angepasst werden, da hier explizit die Stützstellenverteilung der abgeschlossenen Newton-Cotes-Formeln verwendet wurde.

7.5 Überblick

Dieses Kapitel gibt einen Überblick über die abgeschlossenen und über die offenen Newton-Cotes-Formeln bis einschließlich Ordnung 6. Es werden in den folgenden Tabellen jeweils die Ordnung n der entsprechenden Newton-Cotes-Formel, die Gewichte w_i (nach Satz 7.10 und Satz 7.12) und die Fehlerordnung m (nach Satz 6.7) angegeben. Die Stützstellen $\tilde{x}_i$ werden aus Platzgründen wegen ihrer einfachen Berechnung (siehe Definitionen 7.7 und 7.8) nicht angeführt.

Für die abgeschlossenen Newton-Cotes-Formeln werden zusätzlich die Fehlerkonstante c_m (nach Satz 7.21)[11], der Faktor $\gamma := n^{m+1} \cdot |c_m|$ des Höchstfehlers der „einfachen“ Formel (nach Korollar 7.23) und der Faktor $\kappa := n^m \cdot |c_m|$ des Höchstfehlers der zusammengesetzten Formel (nach Satz 7.22) angegeben.

Abgeschlossene Newton-Cotes-Formeln: Die abgeschlossenen Newton-Cotes-Formeln erster und zweiter Ordnung (Sehnentrapezregel und Simpsonregel) wurden bereits ausführlich behandelt. Die Formel dritter Ordnung wird häufig als 3/8-Regel bzw. Pulcherrima bezeichnet. Für $n = 4$ spricht man von der Milne-Regel (selten auch Boole-Regel). Die Formel sechster Ordnung heißt Weddle-Regel (nach Thomas Weddle, 1817 – 1853). Nachfolgend werden die genannten Kennzahlen aufgelistet:

Eine Besonderheit tritt erstmals bei $n = 8$ auf: Hier kommen nämlich negative Gewichte vor. Für den Fall $\sum_{i=0}^{n} |w_i| > 1$ (was bei negativen Gewichten der Fall ist) kann es zu problematischen Situationen kommen. Als extremes Beispiel wird im nächsten Kapitel gezeigt, wie es für einen Integranden mit $f(x) > 0$, $\forall x \in [a, b]$ zu einem negativen Wert für $\int_a^b f(x)\, dx$ kommen kann (siehe Beispiel 3, Seite 81).

[11] Dass der Peano-Kern auf $[0, 1]$ keinen Vorzeichenwechsel aufweist, wurde mittels MATHEMATICA überprüft, indem der Graph von K_m geplottet wurde.

n	w_i	m	c_m	γ	κ
1	$\frac{1}{2}\ \frac{1}{2}$	2	$-\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$
2	$\frac{1}{6}\ \frac{2}{3}\ \frac{1}{6}$	4	$-\frac{1}{2880}$	$\frac{1}{90}$	$\frac{1}{180}$
3	$\frac{1}{8}\ \frac{3}{8}\ \frac{3}{8}\ \frac{1}{8}$	4	$-\frac{1}{6480}$	$\frac{3}{80}$	$\frac{1}{80}$
4	$\frac{7}{90}\ \frac{16}{45}\ \frac{2}{15}\ \frac{16}{45}\ \frac{7}{90}$	6	$-\frac{1}{12600}$	$\frac{8}{945}$	$\frac{2}{945}$
5	$\frac{19}{288}\ \frac{25}{96}\ \frac{25}{144}\ \frac{25}{144}\ \frac{25}{96}\ \frac{19}{288}$	6	$-\frac{1}{12600}$	$\frac{275}{12096}$	$\frac{55}{12096}$
6	$\frac{41}{840}\ \frac{9}{35}\ \frac{9}{280}\ \frac{34}{105}\ \frac{9}{280}\ \frac{9}{35}\ \frac{41}{840}$	8	$-\frac{1}{453600}$	$\frac{9}{1400}$	$\frac{3}{2800}$

Eine weitere Besonderheit macht sich ab $n = 12$ bemerkbar: Hier gilt nämlich für den Betrag der drei mittleren Gewichte $|w_5| = w_5 > 1$, $|w_6| = -w_6 > 1$ und $|w_7| = w_7 > 1$. Auch dies kann zu problematischen Situationen führen (siehe Beispiel 4, Seite 82).

Aus obiger Tabelle ist folgendes Muster für die Fehlerordnung erkennbar:

Vermutung 7.24. Sei $n \in \mathbb{N}$, und sei Q die abgeschlossene Newton-Cotes-Formel n-ter Ordnung. Dann gilt für die Fehlerordnung m von Q:

(i) Falls n gerade ist, so ist $m = n + 2$.

(ii) Falls n ungerade ist, so ist $m = n + 1$.

Dies bedeutet, dass abgeschlossene Newton-Cotes-Formeln mit gerader Ordnung auch Polynome exakt integrieren, deren Grad um 1 höher ist als jener des Interpolationspolynoms. Formeln mit ungerader Ordnung hingegen liefern nur bei Polynomen mit höchstens gleichem Grad wie jener des Interpolationspolynoms exakte Ergebnisse.

Offene Newton-Cotes-Formeln: Von den offenen Newton-Cotes-Formeln wurde bisher lediglich jene mit Ordnung $n = 0$ behandelt, also die Mittelpunktregel. Die Ausführungen am Beginn des nächsten Kapitels werden jedoch ergeben, dass alle anderen Formeln dieser Klasse in der Praxis nicht relevant sind, da es bessere Klassen von offenen Quadraturformeln gibt. Nachfolgend werden dennoch die Gewichte und die Fehlerordnungen der offenen Newton-Cotes-Formeln bis zur sechsten Ordnung aufgelistet:

n	w_i	m
0	1	2
1	$\frac{1}{2}\ \frac{1}{2}$	2
2	$\frac{2}{3}\ \frac{-1}{3}\ \frac{2}{3}$	4
3	$\frac{11}{24}\ \frac{1}{24}\ \frac{1}{24}\ \frac{11}{24}$	4
4	$\frac{11}{20}\ \frac{-7}{10}\ \frac{13}{10}\ \frac{-7}{10}\ \frac{11}{20}$	6
5	$\frac{611}{1440}\ \frac{-151}{480}\ \frac{281}{720}\ \frac{281}{720}\ \frac{-151}{480}\ \frac{611}{1440}$	6
6	$\frac{92}{189}\ \frac{-106}{105}\ \frac{244}{105}\ \frac{-2459}{945}\ \frac{244}{105}\ \frac{-106}{105}\ \frac{92}{189}$	8

Man erkennt, dass die Aussage von Vermutung 7.24 auch für die offenen Newton-Cotes-Formeln plausibel erscheint.

Vermutung 7.25. Sei $n \in \mathbb{N}$, und sei Q die offene Newton-Cotes-Formel n-ter Ordnung. Dann gilt für die Fehlerordnung m von Q:

(i) Falls n gerade ist, so ist $m = n + 2$.

(ii) Falls n ungerade ist, so ist $m = n + 1$.

In den Abschnitten 11.1 und 11.2 werden MATHEMATICA-Codes zur Anwendung der Newton-Cotes-Formeln beliebiger Ordnung und mit beliebiger Anzahl an Teilintervallen vorgestellt und erklärt. Außerdem können damit die in den beiden obigen Tabellen aufgelisteten Kennzahlen ermittelt werden.

7.6 Wahl einer geeigneten Newton-Cotes-Formel

In diesem Kapitel werden die Vor- und Nachteile der verschiedenen Newton-Cotes-Formeln erläutert. Konkret werden folgende Fragen behandelt:

1. In welchen Fällen sollte man sich für eine abgeschlossene Newton-Cotes-Formel entscheiden, und wann ist es sinnvoller, eine offene Newton-Cotes-Formel zu verwenden?
2. Ist es besser die Ordnung n, also den Grad des Interpolationspolynoms, oder die Anzahl N der Teilintervalle von $[a, b]$ zu erhöhen?

Abgeschlossene Formeln vs. offene Formeln

Im Zusammenhang mit Experimenten ist es in vielen Fällen notwendig, eine abgeschlossene Formel zu verwenden, da man die Funktionswerte (Messwerte) nur an den Stützstellen (Messzeitpunkte) und nicht im Inneren der Teilintervalle kennt. Wird jedoch bereits bei der Planung des Experiments darauf geachtet, die Stützstellen gemäß der offenen Newton-Cotes-Formeln über den Messzeitraum zu verteilen, so ist dieser Punkt hinfällig.

Aus mathematischer Sicht sind offene Formeln sinnvoll, wenn der Funktionswert des Integranden an einer der Integralgrenzen nicht definiert ist, z. B. bei $\int_0^1 1/\sqrt{x}\,dx$. Jedoch werden hierfür heute meist keine Newton-Cotes-Formeln verwendet, da diese laut [27, Seite 49, erster Absatz] eine schlechtere Fehlerordnung als die ebenfalls offenen Gauß-Formeln (siehe Kapitel 9) haben. Die einzige Ausnahme bildet laut [27, Seite 49, erster Absatz] aufgrund ihrer einfachen Umsetzbarkeit die Mittelpunktregel.

Höhere Ordnung vs. mehr Teilintervalle

Wie bereits oben erläutert, ist es aufgrund der dort vorkommenden negativen Gewichte nicht empfehlenswert, abgeschlossene Newton-Cotes-Formeln mit $n \geq 8$ zu verwenden. Ein Grund dafür ist das Phänomen der Auslöschung, welches bei der Subtraktion von annähernd gleich großen Gleitkommazahlen auftreten kann. Auf diesen Effekt der numerischen Mathematik wird in dieser Arbeit jedoch nicht eingegangen. Es wird stets davon ausgegangen, dass bei der Berechnung der Näherungswerte ausreichend genaue Zwischenergebnisse vorliegen (wie dies bei MATHEMATICA üblich ist). Wie einige der nachfolgenden Beispiele demonstrieren werden, kommt es aber durch die negativen Gewichte auch zu anderen, wesentlich gravierenderen Problemen, durch welche das Ergebnis auch bei der Verwendung exakter Zahlen unbrauchbar wird.

Ist N die Anzahl der Teilintervalle in welche $[a, b]$ zerlegt wird und n die Ordnung der Newton-Cotes-Formel, so musst bei einer zusammengesetzten Formel die Funktion f an insgesamt $N(n + 1) - (N - 1) = Nn + 1$ verschiedenen Stellen ausgewertet werden. Da

(bei bekannten Gewichten) das Bestimmen der Funktionswerte in den meisten Fällen den größten Teil des Rechenaufwandes beanspruchen wird, kann dies als Maß für die Effizienz der jeweiligen Formel betrachtet werden. Das bedeutet, es ist jene Kombination von N und n zu bevorzugen, welche bei ähnlicher Stützstellenanzahl $Nn + 1$ die höchste Genauigkeit erzielt. Dies ist allgemein schwierig zu beurteilen, da die auftretenden Effekte von der konkreten Funktion f abhängig sind. Aus diesem Grund werden nachfolgend für einige Kombinationen von N und n und für verschiedene Funktionen die Genauigkeiten berechnet und daraus Schlüsse gezogen.

Die grau dargestellten Stellen der Näherungswerte geben bei den folgenden Beispielen (und allgemein im weiteren Verlauf dieser Arbeit) jeweils an, dass diese Stelle vom tatsächlichen Wert abweicht.

Beispiel 1: $f : [0, 1] \to \mathbb{R}, \quad f(x) = e^x, \quad \int_0^1 e^x \, dx = e - 1 \approx 1.718281828$

Für dieses Beispiel werden die abgeschlossenen Newton-Cotes-Formeln der Ordnungen 1, 2, 3, 4, 5, 6, 7, 8, 10, 16 und 32 untersucht. Die Anzahl N an Teilintervallen wird so gewählt, dass insgesamt ungefähr 100 Funktionsauswertungen notwendig sind, also $Nn + 1 \approx 100$.

Folgende Tabelle zeigt die Werte für n, N und $Nn + 1$, den auf 20 Nachkommastellen gerundeten Näherungswert des Integrals und die Abweichung vom tatsächlichen Wert.

n	N	$Nn+1$	$Q(f)$	$\lvert R(f)\rvert$
1	99	100	1.71829643818344849603	1.46×10^{-5}
2	49	99	1.71828182856253852844	1.03×10^{-10}
3	33	100	1.71828182868263559871	2.24×10^{-10}
4	25	101	1.71828182845904887174	3.64×10^{-15}
5	20	101	1.71828182845905304769	7.81×10^{-15}
6	16	97	1.71828182845904523562	2.55×10^{-19}
7	14	99	1.71828182845904523582	4.55×10^{-19}
8	12	97	1.71828182845904523536	1.64×10^{-23}
10	10	101	1.71828182845904523536	7.07×10^{-28}
16	6	97	1.71828182845904523536	5.68×10^{-40}
32	3	97	1.71828182845904523536	–

Der Fehler der letzten Näherung konnte nicht berechnet werden, da hier bereits die maximale Rechengenauigkeit des verwendeten Computers überschritten wurde. Man kann durch dieses Beispiel zwei interessante Zusammenhänge erkennen:

- Das Verfahren scheint problemlos anwendbar zu sein: Je größer die Ordnung der Newton-Cotes-Formel wird, umso besser ist die Näherung bei etwa gleich großer Anzahl an Stützstellen tendenziell. Man könnte aus diesem Beispiel den Trugschluss ziehen, dass das Verfahren umso besser ist, je größer man n wählt.
- Bei diesem Beispiel kann man sehr schön die in Vermutung 7.24 beschriebene Tatsache erkennen: Die Formeln mit gerader Ordnung n sind wesentlich genauer, als die jeweils vorherige Formel mit ungerader Ordnung. Es ist in diesem Fall sogar so, dass sie geringfügig genauer sind, als die Formeln mit nächsthöherer ungerader Ordnung.

Beispiel 2: $f : [0,1] \to \mathbb{R}, \quad f(x) = \sin^2(100x), \quad \int_0^1 f(x)\,dx \approx 0.502183243$

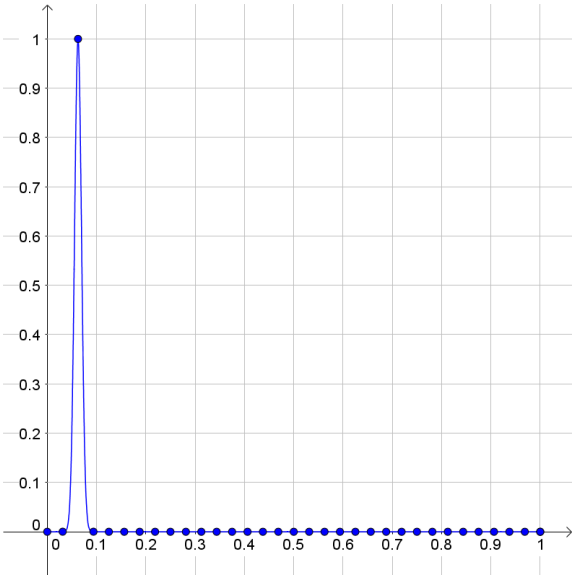
Auch in diesem Beispiel wurden die Näherungswerte für einige Kombinationen von N und n berechnet und verglichen:

n	N	$Nn+1$	$Q(f)$	$\lvert R(f)\rvert$
1	99	100	0.501384746460459654 25	7.98×10^{-4}
2	49	99	0.502577722194023034061	3.94×10^{-4}
3	33	100	0.508958072163332254 03	6.77×10^{-3}
4	25	101	0.503049505364541696 31	8.66×10^{-4}
5	20	101	0.503308005343760471 56	1.12×10^{-3}
6	16	97	0.449449143425939782 63	5.27×10^{-2}
7	14	99	0.504987482301735029 50	2.80×10^{-3}
8	12	97	0.497820297945089281 58	4.36×10^{-3}
10	10	101	0.490815073262764448 54	$1.14 \times 10{-}2$
16	6	97	0.664913291115311236 95	1.63×10^{-1}
32	3	97	-360.627070934939997 68	3.61×10^{2}
2	288	577	0.502183422108890348 73	1.79×10^{-7}
32	18	577	0.502183243243034986 45	1.33×10^{-22}

Bei dieser sehr stark oszillierenden Funktion kann man keine Regelmäßigkeit hinter der Entwicklung des Fehlers erkennen. Man sieht jedoch, dass der Fehler für jene Formeln mit negativen Gewichten (also ab $n = 8$) größer wird und teilweise absurde Näherungswerte entstehen.

Die letzte Zeile zeigt zwar, dass das Verfahren für genügend viele Funktionsauswertungen scheinbar gut funktioniert, jedoch ändert dies nichts an der Tatsache, dass es in manchen Fällen (die in den folgenden Beispielen künstlich herbeigeführt werden und in ähnlicher Form immer, also auch bei beliebig vielen Funktionsauswertungen, auftreten können) zu sehr problematischen Situationen kommt.

Beispiel 3: $f : [0,1] \to \mathbb{R}, \quad f(x) = e^{-\frac{(x-0.0625)^2}{0.0001}}, \quad \int_0^1 f(x)\,dx \approx 0.0177245$

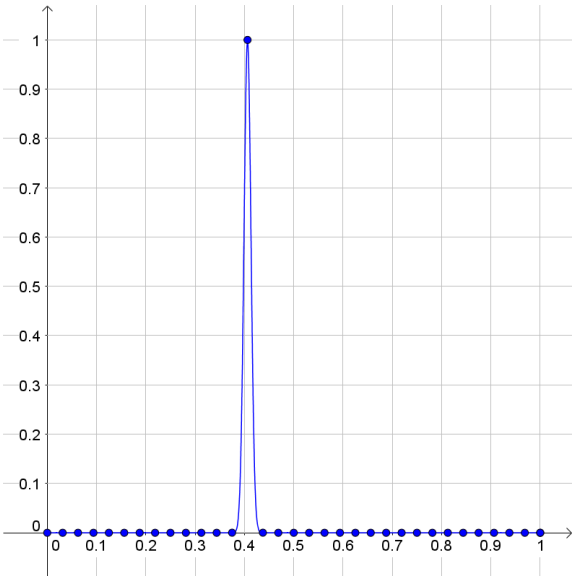


Diese Variation der Dichtefunktion der Normalverteilung wurde speziell so konstruiert, dass sie bei $\frac{2}{32}$, also an einer Stützstelle, an der bei $(n, N) \in \{(8,4), (16,2), (32,1)\}$ negative Gewichte auftreten, einen deutlich von 0 verschiedenen Funktionswert besitzt, nämlich den Funktionswert 1. Bei den anderen Stützstellen ist die Funktion fast 0. Insbesondere ist hervorzuheben, dass die Funktion auf ganz $\mathbb{R}$ positiv ist. Erneut wurden ausgewählte Kombinationen für n und N untersucht:

n	N	$Nn+1$	$Q(f)$	$\lvert R(f)\rvert$
1	32	33	0.03125	1.35×10^{-2}
2	16	33	0.02084	3.11×10^{-3}
4	8	33	0.01667	1.05×10^{-3}
8	4	33	-0.00818	2.59×10^{-2}
16	2	33	-0.09196	1.10×10^{-1}
32	1	33	-0.36306	3.81×10^{-1}

Bei $n \in \{8, 16, 32\}$ wird die Unbrauchbarkeit des Ergebnisses besonders deutlich sichtbar. Aufgrund der negativen Gewichte ist hier der Näherungswert des Integrals einer positiven Funktion negativ.

Beispiel 4: $f : [0,1] \to \mathbb{R}, \quad f(x) = e^{-\frac{(x-0.40625)^2}{0.0001}}, \quad \int_0^1 f(x)\,dx \approx 0.0177245$



In diesem Beispiel wird die „Nadel“ an die Stelle $x = \frac{13}{32}$ verschoben, da das zu dieser Stützstelle gehörende Gewicht für die Kombination $(n, N) = (32, 1)$ sehr viel größer als 1 ist.

n	N	$Nn+1$	$Q(f)$	$\lvert R(f)\rvert$
1	32	33	0.03125	1.34×10^{-2}
2	16	33	0.04167	2.39×10^{-2}
32	1	33	58963.2	5.90×10^{4}

Dadurch ergibt es sich, dass der Näherungswert überhaupt nichts mit dem tatsächlichen Wert des Integrals zu tun hat.

Beispiel 5: $f : [0,1] \to \mathbb{R}, \quad f(x) = \sqrt{x}, \quad f'(x) = \frac{1}{2\sqrt{x}}, \quad \int_0^1 f(x)\,dx = \frac{2}{3}$

Diese Funktion hat im Vergleich zu den Funktionen der vorherigen drei Beispiele keine besonderen Unregelmäßigkeiten, weshalb man annehmen könnte, dass das Integral relativ einfach approximiert werden kann. Die Funktion f ist jedoch an der Stelle $x = 0 \in [0,1]$ nicht differenzierbar. Es ist anhand dieses Beispiels gut ersichtlich, dass die Fehlerabschätzung von Satz 7.22 tatsächlich nur für Funktionen anwendbar ist, die genügend oft differenzierbar sind, denn für die Näherung mittels abgeschlossener Newton-Cotes-Formel der Ordnung $n = 32$ und der Unterteilung in drei Teilintervalle (also insgesamt 97 Stützstellen) beträgt der Fehler noch immer ungefähr 4×10^{-5}. Im Vergleich dazu wurde in Beispiel 1 bereits durch die Sehnentrapezregel ein gleichwertiger Fehler erzielt. Betrachtet man bei Beispiel 5 den Fehler für verschiedene Ordnungen n, so erkennt man, dass dieser stets ungefähr in der Größenordnung 10^{-4} bis 10^{-5} liegt (bereits bei $n = 1$), weshalb es tatsächlich plausibel erscheint, dass die Näherung bei etwa gleich großer Anzahl an Stützstellen unabhängig von der Ordnung nicht wesentlich besser wird.

Fazit: Abschließend lässt sich feststellen, dass die Newton-Cotes-Formeln mit größerem n bei gleicher Anzahl an Funktionsauswertungen zwar tendenziell exaktere Ergebnisse liefern, diese jedoch mit einem gewissen Risiko verbunden sind. Weiß man, dass die zu integrierende Funktion einigermaßen „normal“ verläuft, also keine großen Schwankungen aufweist, so spielen die vorgestellten Effekte möglicherweise keine so große Rolle. Im Allgemeinen sollte man sich jedoch trotzdem auf Formeln mit $n \leq 7$ beschränken. Der Effekt von Beispiel 4 wird außerdem weitestgehend verhindert, wenn die Gewichte möglichst gleich groß (und somit jedenfalls kleiner als 1) sind.

In allen Beispielen hat sich gezeigt, dass die Sehnentrapezregel ($n = 1$) zwar nicht die besten, aber durchaus akzeptable Näherungswerte liefert. Die Gewichte sind gleich groß, weshalb auch die soeben angesprochene Eigenschaft erfüllt ist. Außerdem ist die Berechnung äußerst einfach durchzuführen, da man (bei der zusammengesetzten Form) nur die inneren Stützwerte summieren und die Hälfte der beiden äußeren Stützwerte addieren muss. Im nachfolgenden Kapitel wird das sogenannte Romberg-Verfahren behandelt, welches auf die Sehnentrapezregel aufbaut.

8 Romberg-Verfahren

Ausgangspunkt für das nach dem deutschen Mathematiker Werner Romberg (1909 – 2003) benannte Verfahren stellt die Sehnentrapezregel

$$R_{0,0} := h_0 \cdot \frac{f(a) + f(b)}{2}$$

bzw. in zusammengesetzter Form

$$R_{k,0} := h_k \cdot \left(\frac{f(a)}{2} + \sum_{i=1}^{\frac{h_0}{h_k}-1} f(a + ih_k) + \frac{f(b)}{2} \right)$$

dar. Dabei ist $h_k := \frac{b-a}{2^k}$ und somit $R_{k,0}$ der Näherungswert des Integrals, den man durch die Sehnentrapezregel mit k Halbierungen des Intervalls $[a, b]$ erhält. In diesem Zusammenhang wird die Folge $\left(\frac{1}{2^k}\right)_{k\in\mathbb{N}_0} = 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, ...$ als Romberg-Folge bezeichnet.

Die Grundidee ist es, ausgehend von den Näherungswerten $R_{0,0}, ..., R_{k,0}$ mit Hilfe des Algorithmus von Aitken und Neville (siehe Kapitel 3.5) eine Extrapolation des Näherungswertes auf $h = 0$ durchzuführen. Aus diesem Grund spricht man auch häufig von der Romberg-Extrapolation. Dazu werden die Näherungswerte $R_{k,0}$ als Funktionswerte in Abhängigkeit von h^2 aufgefasst (der Grund dafür ist in Satz 8.5 ersichtlich) und das eindeutig bestimmte Interpolationspolynom mittels Aitken-Neville-Algorithmus an der Stelle $h = 0$ ausgewertet. Dieser Algorithmus erlaubt es einerseits, Polynomwerte zu berechnen, ohne das Polynom zuvor explizit ermittelt zu haben, und andererseits kann bei nicht ausreichender Genauigkeit jederzeit eine weitere Stützstelle (sinnvollerweise nahe bei 0) hinzugenommen werden. Die Tatsache, dass das Romberg-Verfahren solange durchgeführt werden kann, bis die gewünschte Genauigkeit erreicht wurde, stellt einen wesentlichen Vorteil dieser Methode dar.

Laut Lemma von Aitken (Satz 3.18) ergibt sich für $R_{k,j}$ an der Stelle $h = 0$ folgender Wert:

$$R_{k,j}(0) = \frac{h_{k-j}^2 \cdot R_{k,j-1}(0) - h_k^2 \cdot R_{k-1,j-1}(0)}{h_{k-j}^2 - h_k^2}$$

Bevor auf die Details dieses Verfahrens näher eingegangen wird, wird zum besseren Verständnis zunächst ein konkretes Rechenbeispiel vorgestellt. In Abschnitt 11.3 wird außerdem ein MATHEMATICA-Code zur Anwendung des Romberg-Verfahrens (und der später behandelten Variationen) dargestellt und erklärt.

8.1 Rechenbeispiel

In diesem Beispiel soll die Funktion $f(x) = e^x$ im Intervall $[0, 1]$ integriert werden. Das exakte Ergebnis lautet auf 40 Nachkommastellen genau

$$\int_0^1 e^x \, dx = e^x \,|_0^1 = e^1 - e^0 = e - 1 = 1.7182818284590452353602874713526624977572...$$

Es werden im ersten Schritt die Näherungswerte $R_{0,0}, ..., R_{4,0}$ berechnet. Die Anzahl der Funktionsauswertungen beträgt dabei jeweils $2^k + 1$, wobei die Funktionsauswertungen der vorherigen Näherung übernommen werden können. Insgesamt benötigt man für die Berechnung von $R_{4,4}$ also $2^4 + 1 = 17$ Funktionswerte. Die Ergebnisse sind in folgender Tabelle dargestellt:

k	$R_{k,0}$	$R_{k,1}$	$R_{k,2}$	$R_{k,3}$	$R_{k,4}$
0	1.85914				
1	1.75393	1.71886115			
2	1.72722	1.71831884	1.71828268792		
3	1.72051	1.71828415	1.71828184221	1.718281828794	
4	1.71884	1.71828197	1.71828182867	1.718281828460	1.718281828459078

Man erkennt hier, dass aus den ursprünglichen Näherungswerten $R_{k,0}$ mit einer Genauigkeit von höchstens drei richtigen Nachkommastellen ohne die Hinzunahme weiterer Stützstellen die Anzahl richtiger Nachkommastellen auf 13 erhöht werden konnte.

Die Simpsonregel und die Milne-Regel liefern bei 17 Stützstellen Näherungswerte, die wesentlich ungenauer sind:

- Simpsonregel: 1.7182819740518919044 (6 Nachkommastellen)
- Milne-Regel: 1.7182818286753582377 (9 Nachkommastellen)

Erhöht man die Anzahl der Funktionsauswertungen auf 257, so erhält man folgende Resultate:

- Simpsonregel: 1.718281828461 (10 Nachkommastellen)
- Milne-Regel: 1.71828182845904548 (16 Nachkommastellen)
- $R_{8,8} = 1.7182818284590452353602874713526625007$ (33 Nachkommastellen)

Dies legt die Vermutung nahe, dass das Romberg-Verfahren besser ist als die abgeschlossenen Newton-Cotes-Formeln niedriger Ordnung und im Vergleich zu den Formeln höherer Ordnung tatsächlich konvergiert. Diese Vermutungen werden in den folgenden Abschnitten näher untersucht.

8.2 Definition

Nachfolgend werden die im Romberg-Verfahren vorkommenden Größen exakt definiert.

Definition 8.1. Sei $f : [a, b] \to \mathbb{R}$ eine Funktion. Sei $h_i := \frac{b-a}{2^i}$ für $i \in \mathbb{N}_0$. Dann definiert man die Startwerte der Romberg-Extrapolation als

$$R_{k,0} := h_k \cdot \left(\frac{f(a)}{2} + \sum_{i=1}^{\frac{h_0}{h_k}-1} f(a + ih_k) + \frac{f(b)}{2} \right)$$

für $k \in \mathbb{N}_0$. Für ein festes $j \in \mathbb{N}$ und $k \geq j$ definiert man die aus den Startwerten erhaltenen Näherungswerte $R_{k,j}$ folgendermaßen:

$$R_{k,j} := \frac{h_{k-j}^2 \cdot R_{k,j-1} - h_k^2 \cdot R_{k-1,j-1}}{h_{k-j}^2 - h_k^2}$$

Die Näherungswerte $R_{j,j}$ werden nachfolgend als Romberg-Näherungen j-ter Stufe bezeichnet.

Lemma 8.2. Es gilt die folgende Gleichheit:

$$R_{k,j} := \frac{h_{k-j}^2 \cdot R_{k,j-1} - h_k^2 \cdot R_{k-1,j-1}}{h_{k-j}^2 - h_k^2} = R_{k,j-1} + \frac{R_{k,j-1} - R_{k-1,j-1}}{\left(\frac{h_{k-j}}{h_k}\right)^2 - 1}$$

Beweis: Zur besseren Übersichtlichkeit werden die Umbenennungen $R_1 := R_{k-1,j-1}$, $R_2 := R_{k,j-1}$, $h_1 := h_{k-j}$ und $h_2 := h_k$ vorgenommen. Es wird nun die rechte Seite der Gleichung umgeformt:

$$R_2 + \frac{R_2 - R_1}{\left(\frac{h_1}{h_2}\right)^2 - 1} = \frac{\frac{h_1^2}{h_2^2} R_2 - R_2 + R_2 - R_1}{\frac{h_1^2 - h_2^2}{h_2^2}} = \frac{h_1^2 R_2 - h_2^2 R_1}{h_1^2 - h_2^2}$$

□

8.3 Fehlerabschätzung

In diesem Abschnitt wird erläutert, warum es beim Romberg-Verfahren durch die Kombination der zusammengesetzten Trapezregel mit verschiedenen Schrittweiten zu einer Verbesserung der Genauigkeit kommt.

Voraussetzung dafür sind die in Kapitel 4 behandelten Bernoulli-Polynome, mit welchen nun zunächst die Euler-Maclaurinsche-Summenformel bewiesen wird. Diese wird später benutzt, um die asymptotische Fehlerdarstellung der zusammengesetzten Trapezregel zu begründen.

Satz 8.3. (Euler-Maclaurinsche-Summenformel) Sei $g : [0,1] \to \mathbb{R}$ eine $(2m+2)$-mal stetig differenzierbare Funktion. Dann gibt es ein $\xi \in [0,1]$, sodass Folgendes gilt:

$$\int_0^1 g(x)\,dx = \frac{g(0)+g(1)}{2} + \sum_{k=1}^{m} \frac{B_{2k}}{(2k)!} \cdot \left(g^{(2k-1)}(0) - g^{(2k-1)}(1)\right) - \frac{B_{2m+2}}{(2m+2)!} \cdot g^{(2m+2)}(\xi)$$

Beweis: Quelle dieses Beweises ist [6, Seite 172, Satz 6.1.8]. Durch Multiplikation mit dem Bernoulli-Polynom $B_0(x) = 1$ erhält man

$$\int_0^1 g(x)\,dx = \int_0^1 g(x) \cdot B_0(x)\,dx.$$

Wegen Definition 4.1 (ii), also $B_n'(x) = n \cdot B_{n-1}(x) \Leftrightarrow B_{n-1}(x) = \frac{1}{n} \cdot B_n'(x)$ für $n \geq 1$, ist $B_1(x)$ eine Stammfunktion von $B_0(x)$. Somit erhält man mittels partieller Integration

$$\int_0^1 g(x) \cdot B_0(x)\,dx = g(x) \cdot B_1(x)\Big|_0^1 - \int_0^1 g'(x) \cdot B_1(x)\,dx.$$

Wendet man dieses Prinzip der partiellen Integration nun fortlaufend insgesamt $2m$-mal jeweils auf das entstehende Integral an, so erhält man unter Verwendung der Tatsache, dass $\frac{1}{k} \cdot B_k(x)$ eine Stammfunktion von $B_{k-1}(x)$ ist, für den allgemeinen Fall folgende Formel:

$$\int_0^1 g^{(k-1)}(x) \cdot B_{k-1}(x)\,dx = g^{(k-1)}(x) \cdot \frac{1}{k} \cdot B_k(x)\Big|_0^1 - \frac{1}{k}\int_0^1 g^{(k)}(x) \cdot B_k(x)\,dx$$

Insgesamt kommt man auf diese Weise auf nachfolgende Darstellung. Man beachte dabei, dass aufgrund der Verschachtelung alle Terme mit ungeraden Bernoulli-Polynomen addiert werden und alle Terme mit geraden Bernoulli-Polynomen subtrahiert werden. Das übrig bleibende Integral ist negativ.

$$\int_0^1 g(x)\,dx = \sum_{k=1}^{m} \frac{1}{(2k-1)!} B_{2k-1}(x) g^{(2k-2)}(x)\Big|_0^1 - \sum_{k=1}^{m} \frac{1}{(2k)!} B_{2k}(x) g^{(2k-1)}(x)\Big|_0^1 +$$

$$-\frac{1}{(2m+1)!}\int_0^1 B_{2m+1}(x) g^{(2m+1)}(x)\,dx$$

Weil nach Lemma 4.4 für $k \geq 2$ gilt $B_k = B_k(0) = B_k(1)$, ist in diesem Fall

$$B_k(x) g^{(k-1)}(x)\Big|_0^1 = \underbrace{B_k(1)}_{=B_k} g^{(k-1)}(1) - \underbrace{B_k(0)}_{=B_k} g^{(k-1)}(0) = B_k\left(g^{(k-1)}(1) - g^{(k-1)}(0)\right).$$

Diese Umformung wird später für die zweite Summe eingesetzt.

Für die erste Summe bleibt wegen $B_k(0) = B_k(1) = 0$ für ungerade $n \geq 3$ (siehe Lemma 4.11) und $B_1(0) = -\frac{1}{2}$ und $B_1(1) = \frac{1}{2}$ nur folgender Term übrig:

$$\frac{1}{(2 \cdot 1 - 1)!} \cdot \left(\frac{1}{2}\, g(1) - \left(-\frac{1}{2}\right) g(0)\right) = \frac{g(0)+g(1)}{2}$$

Verwendet man als Stammfunktion von $B_{2m+1}(x)$ den Ausdruck

$$\frac{1}{2m+2}\left(B_{2m+2}(x) - B_{2m+2}\right),$$

also die Integrationskonstante $\frac{-B_{2m+2}}{2m+2}$, so erhält man mittels partieller Integration für das verbleibende Integral unter Verwendung der aus Lemma 4.4 resultierenden Tatsache, dass für $x = 0$ und $x = 1$ die Eigenschaft $B_{2m+2}(x) - B_{2m+2} = 0$ gilt, Folgendes:

$$\begin{aligned}\int_0^1 B_{2m+1}(x) g^{(2m+1)}(x)\,dx &= g^{(2m+1)}(x) \cdot \frac{1}{2m+2} \cdot \underbrace{\left(B_{2m+2}(x) - B_{2m+2}\right)}_{=0,\text{ für } x\in\{0,1\}}\Big|_0^1 + \\ &- \frac{1}{2m+2}\int_0^1 \left(B_{2m+2}(x) - B_{2m+2}\right) \cdot g^{(2m+2)}(x)\,dx = \\ &= -\frac{1}{2m+2}\int_0^1 \left(B_{2m+2}(x) - B_{2m+2}\right) \cdot g^{(2m+2)}(x)\,dx\end{aligned}$$

Insgesamt gilt somit also:

$$\begin{aligned}\int_0^1 g(x)\,dx &= \frac{g(0)+g(1)}{2} - \sum_{k=1}^{m} \frac{B_{2k}}{(2k)!}\left(g^{(2k-1)}(1) - g^{(2k-1)}(0)\right) + \\ &+ \frac{1}{(2m+2)!}\int_0^1 \left(B_{2m+2}(x) - B_{2m+2}\right) \cdot g^{(2m+2)}(x)\,dx\end{aligned}$$

Wegen Satz 4.13 (ii) hat $B_{2m+2}(x) - B_{2m+2}$ auf $[0,1]$ keinen Vorzeichenwechsel. Da die Funktion $g^{(2m+2)}$ nach Voraussetzung stetig ist, folgt aus dem Mittelwertsatz der Integralrechnung (Satz 2.13), dass es ein $\xi \in [0,1]$ gibt, sodass

$$\begin{aligned}\int_0^1 g(x)\,dx &= \frac{g(0)+g(1)}{2} - \sum_{k=1}^{m} \frac{B_{2k}}{(2k)!}\left(g^{(2k-1)}(1) - g^{(2k-1)}(0)\right) + \\ &+ \frac{g^{(2m+2)}(\xi)}{(2m+2)!}\int_0^1 \left(B_{2m+2}(x) - B_{2m+2}\right)\,dx\end{aligned}$$

gilt. Spaltet man das Integral auf, so folgt aus Definition 4.1 (iii) schließlich

$$\int_0^1 g(x)\,dx = \frac{g(0)+g(1)}{2} - \sum_{k=1}^{m} \frac{B_{2k}}{(2k)!}\left(g^{(2k-1)}(1) - g^{(2k-1)}(0)\right) - \frac{B_{2m+2}}{(2m+2)!} \cdot g^{(2m+2)}(\xi).$$

Durch Herausheben von (-1) aus der Summe erhält man die zu zeigende Formel. □

Definition 8.4. Sei $f : [a,b] \to \mathbb{R}$ eine Funktion. Dann definiert man die Trapezsumme T_n in Abhängigkeit von der Schrittweite $h_n := \frac{b-a}{n}$ nach [22, Seite 1] folgendermaßen:

$$T_n := h_n \cdot \left(\frac{f(a)}{2} + \sum_{i=1}^{n-1} f(a + ih_n) + \frac{f(b)}{2}\right).$$

Im Zusammenhang mit dem Romberg-Verfahren gilt $R_{k,0} = T_n$ für $n = 2^k$.

Satz 8.5. Sei $f : [a, b] \to \mathbb{R}$ eine $(2m+2)$-mal stetig differenzierbare Funktion. Dann besitzt die Trapezsumme T_n die asymptotische Fehlerdarstellung

$$\int_a^b f(x)\,dx - T_n = \tau_1 h_n^2 + \tau_2 h_n^4 + ... + \tau_m h_n^{2m} + r(h_n) \cdot h_n^{2m+2},$$

wobei $\tau_1, ...\tau_m \in \mathbb{R}$ Konstanten sind und $r(h_n)$ eine von h_n abhängige beschränkte Funktion ist, d. h. $|r(h_n)| \leq K$ für ein $K \in \mathbb{R}$.

Beweis: Grundlage dieser Beweisführung ist [6, Seite 173 und 174]. Zunächst wird die Funktion $g(t) := f(x_i + h_n t)$ für $t \in [0, 1]$ und $h_n := x_{i+1} - x_i$ definiert. Das heißt, es wird eines der n Teilintervalle der Trapezsumme T_n auf das Intervall $[0, 1]$ transformiert. Durch Substitution von $x = \varphi(t) = x_i + h_n t$ und $\dot{\varphi}(t) = h_n$ erhält man wegen Satz 2.15

$$\int_{x_i}^{x_{i+1}} f(x)\,dx = \int_0^1 f(x_i + h_n t) \cdot h_n\,dt = h_n \int_0^1 g(t)\,dt.$$

Weil f und damit auch g eine $(2m + 2)$-mal stetig differenzierbare Funktion ist, folgt unter Verwendung der Euler-Maclaurin-Summenformel (Satz 8.3) für ein $\xi_i \in (0, 1)$ Folgendes:

$$h_n \int_0^1 g(t)\,dt = \frac{h_n}{2}\,(g(0) + g(1)) + \sum_{k=1}^{m} \frac{B_{2k}}{(2k)!}\,h_n \left(g^{(2k-1)}(0) - g^{(2k-1)}(1)\right) +$$

$$-\frac{B_{2m+2}}{(2m+2)!}\,h_n\,g^{(2m+2)}(\xi_i)$$

Wegen Korollar 2.19 gilt für $g(t) = f(x(t)) = f(x_i + h_n t)$

$$\frac{d^j g}{dt^j} = \frac{d^j f}{dt^j} = \frac{d^j f}{dx^j} \underbrace{\left(\frac{dx}{dt}\right)^j}_{=h_n} = f^{(j)}(x) \cdot h_n^j.$$

Mit $g(0) = f(x_i)$ und $g(1) = f(x_i + h_n) = f(x_{i+1})$ erhält man somit

$$\int_{x_i}^{x_{i+1}} f(x)\,dx = \frac{h_n}{2}(f(x_i) + f(x_{i+1})) + \sum_{k=1}^{m} \frac{B_{2k}}{(2k)!} h_n^{2k} \left(f^{(2k-1)}(x_i) - f^{(2k-1)}(x_{i+1})\right) +$$

$$-\frac{B_{2m+2}}{(2m+2)!} h_n^{2m+3} f^{(2m+2)}(x_i + \xi_i h_n).$$

Für das gesamte Intervall $[a, b]$ erhält man unter Verwendung von Definition 8.4, der Vertauschbarkeit von endlichen Summen und durch Herausheben der vom jeweiligen Teilintervall unabhängigen Terme

$$\int_a^b f(x)\,dx = \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f(x)\,dx =$$

$$= T_n + \sum_{k=1}^{m} \frac{B_{2k}}{(2k)!}\; h_n^{2k} \sum_{i=0}^{n-1} \left(f^{(2k-1)}(x_i) - f^{(2k-1)}(x_{i+1})\right) +$$

$$-\frac{B_{2m+2}}{(2m+2)!}\, h_n^{2m+3} \sum_{i=0}^{n-1} f^{(2m+2)}(x_i + \xi_i h_n).$$

Dabei wurde folgender Zusammenhang verwendet:

$$\sum_{i=0}^{n-1} \frac{h_n}{2}(f(x_i) + f(x_{i+1})) = T_n$$

Es gilt

$$\min_{i\in\{0,\ldots,n-1\}} f^{(2m+2)}(x_i + \xi_i h_n) \leq \frac{1}{n} \sum_{i=0}^{n-1} f^{(2m+2)}(x_i + \xi_i h_n) \leq \max_{i\in\{0,\ldots,n-1\}} f^{(2m+2)}(x_i + \xi_i h_n).$$

Weil $f^{(2m+2)}$ stetig ist, gibt es aufgrund des Zwischenwertsatzes (Satz 2.12) ein $\xi \in (a, b)$ sodass

$$f^{(2m+2)}(\xi) = \frac{1}{n} \sum_{i=0}^{n-1} f^{(2m+2)}(x_i + \xi_i h_n) \Leftrightarrow n \cdot f^{(2m+2)}(\xi) = \sum_{i=0}^{n-1} f^{(2m+2)}(x_i + \xi_i h_n)$$

gilt. Es ist wegen $x_0 = a$ und $x_n = b$ außerdem

$$\sum_{i=0}^{n-1} \left(f^{(2k-1)}(x_i) - f^{(2k-1)}(x_{i+1}) \right) = f^{(2k-1)}(a) - f^{(2k-1)}(b).$$

Deshalb folgt unter Verwendung von $nh_n = b - a$

$$\int_a^b f(x)\, dx \;-\; T_n =$$

$$= \sum_{k=1}^{m} \frac{B_{2k}}{(2k)!}\, h_n^{2k} \left(f^{(2k-1)}(a) - f^{(2k-1)}(b) \right) - \frac{B_{2m+2}}{(2m+2)!}\, h_n^{2m+2}(b-a) f^{(2m+2)}(\xi).$$

Nun definiert man

$$\tau_k := \frac{B_{2k}}{(2k)!} \left(f^{(2k-1)}(a) - f^{(2k-1)}(b) \right)$$

und

$$r(h_n) := \frac{B_{2m+2}}{(2m+2)!}(b-a) f^{(2m+2)}(\xi).$$

Es ist r deshalb von h_n abhängig, weil die einzelnen ξ_i und damit auch ξ von der Wahl der Schrittweite h_n abhängig sind. Außerdem ist $r(h_n)$ beschränkt durch

$$|r(h_n)| \leq \frac{|B_{2m+2}|}{(2m+2)!}(b-a) \max_{x\in[a,b]} |f^{(2m+2)}(x)|.$$

Somit gilt die Fehlerabschätzung

$$\int_a^b f(x)\,dx - T_n = \tau_1 h_n^2 + \tau_2 h_n^4 + ... + \tau_m h_n^{2m} + r(h_n) \cdot h_n^{2m+2}.$$

□

Wie man bereits aus obiger Fehlerdarstellung sehen kann, verschwindet der Fehler für $n \to \infty$ bzw. für $h_n \to 0$ und somit konvergiert die zusammengesetzte Trapezregel gegen das tatsächliche Integral. Wäre die Funktion $r(h_n)$ nicht beschränkt, so wäre diese Konvergenz nicht sichergestellt.

Im vorherigen Satz wurde jedoch vorausgesetzt, dass die zu integrierende Funktion $(2m+2)$-mal stetig differenzierbar ist. Um zu zeigen, dass die zusammengesetzte Trapezregel konvergiert ist es laut [6, Seite 175] überhaupt nicht notwendig, dass die Funktion f diese Eigenschaft erfüllt. Die Konvergenz lässt sich auch zeigen, wenn f nur Riemann-integrierbar ist.

Satz 8.6. Sei $f : [a, b] \to \mathbb{R}$ Riemann-integrierbar. Dann gilt Folgendes:

$$\lim_{n\to\infty} \left| \int_a^b f(x)\,dx - T_n \right| = 0$$

Beweis: Die Beweisführung erfolgt auf Basis von [6, Seite 175, Satz 6.1.9]. Laut Definition 8.4 gilt

$$T_n := h_n \cdot \Bigg(\frac{f(a)}{2} + \underbrace{\sum_{i=1}^{n-1} f(a+ih_n)}_{=2\cdot\frac{1}{2}\cdot\sum_{i=1}^{n-1} f(a+ih_n)} + \frac{f(b)}{2} \Bigg).$$

Durch Umformen erhält man

$$\frac{h_n}{2}\left(f(a) + \sum_{i=1}^{n-1} f(a+ih_n) \right) + \frac{h_n}{2}\left(\sum_{i=1}^{n-1} f(a+ih_n) + f(b) \right) =$$

$$\frac{1}{2}\Bigg(\underbrace{h_n \cdot \sum_{i=0}^{n-1} f(a+ih_n)}_{=:\gamma_1(h_n)} + \underbrace{h_n \cdot \sum_{i=1}^{n} f(a+ih_n)}_{=:\gamma_2(h_n)} \Bigg).$$

Bei γ_1 und γ_2 handelt es sich um Riemann-Summen für f auf dem Intervall $[a, b]$. Weil f Riemann-integrierbar ist, gilt

$$\lim_{h_n\to 0} \gamma_1(h_n) = \lim_{h_n\to 0} \gamma_2(h_n) = \int_a^b f(x)\,dx.$$

Somit erhält man als Grenzwert der Trapezsumme

$$T_n \to \frac{1}{2}\left(\gamma_1(h_n) + \gamma_2(h_n)\right) = \frac{1}{2} \cdot 2 \cdot \int_a^b f(x)\,dx = \int_a^b f(x)\,dx.$$

□

Entscheidend am Resultat von Satz 8.5 ist jedoch, dass hier ausschließlich gerade Potenzen von h_n auftreten. Diese Tatsache kann man sich laut [6, Seite 175] beim Extrapolationsverfahren ausnutzen.

Um zu klären, was hinter der Funktionsweise des Romberg-Verfahrens steckt, betrachtet man die zwei verschiedenen Schrittweiten h_k und $h_{k+1} = \frac{h_k}{2}$. Nach Satz 8.5 folgt somit:

$$\int_a^b f(x)\,dx - R_{k,0} = \sum_{i=1}^{m} \tau_i h_k^{2i} + r(h_k) \cdot h_k^{2m+2}$$

$$\int_a^b f(x)\,dx - R_{k+1,0} = \sum_{i=1}^{m} \tau_i \left(\frac{h_k}{2}\right)^{2i} + r\left(\frac{h_k}{2}\right) \cdot \left(\frac{h_k}{2}\right)^{2m+2}$$

Durch Einsetzen dieser beiden Ausdrücke in Lemma 8.2 erhält man mit der Nebenrechnung

$$\left(\frac{h_k}{h_{k+1}}\right)^2 - 1 = \left(\frac{h_{k+1} \cdot 2}{h_{k+1}}\right)^2 - 1 = 2^2 - 1 = 3$$

folgenden Ausdruck für die Romberg-Näherung $R_{k+1,1}$:

$$R_{k+1,1} = R_{k+1,0} + \frac{R_{k+1,0} - R_{k,0}}{\left(\frac{h_k}{h_{k+1}}\right)^2 - 1} = \int_a^b f(x)\,dx - \sum_{i=1}^{m} \tau_i \left(\frac{h_k}{2}\right)^{2i} - r\left(\frac{h_k}{2}\right) \cdot \left(\frac{h_k}{2}\right)^{2m+2} +$$

$$+\frac{1}{3}\left(\int_a^b f(x)\,dx - \sum_{i=1}^{m} \tau_i \left(\frac{h_k}{2}\right)^{2i} - r\left(\frac{h_k}{2}\right) \cdot \left(\frac{h_k}{2}\right)^{2m+2} - \int_a^b f(x)\,dx + \sum_{i=1}^{m} \tau_i h_k^{2i} + r(h_k) \cdot h_k^{2m+2}\right)$$

Die beiden Integrale innerhalb der Klammer fallen weg. Durch Subtrahieren des ersten Integrals und Multiplizieren mit -1 erhält man:

$$\int_a^b f(x)\,dx - R_{k+1,1} = \frac{4}{3}\sum_{i=1}^{m} \tau_i \left(\frac{h_k}{2}\right)^{2i} + \frac{4}{3} r\left(\frac{h_k}{2}\right) \cdot \left(\frac{h_k}{2}\right)^{2m+2} - \frac{1}{3}\sum_{i=1}^{m} \tau_i h_k^{2i} - \frac{1}{3} r(h_k) \cdot h_k^{2m+2}$$

Für die einzelnen τ_i ergibt sich

$$\frac{4}{3}\tau_i \underbrace{\left(\frac{h_k}{2}\right)^{2i}}_{=h_k^{2i} \cdot \frac{1}{2^{2i}}} - \frac{1}{3}\tau_i h_k^{2i} = \left(\frac{4 \cdot 2^{-2i} - 1}{3}\right) \tau_i \cdot h_k^{2i} = \underbrace{\frac{4^{1-i} - 1}{3}\tau_i}_{=:\tilde{\tau}_i} \cdot h_k^{2i}.$$

Für das Restglied erhält man:

$$\frac{4}{3} \cdot r\left(\frac{h_k}{2}\right) \cdot \left(\frac{h_k}{2}\right)^{2m+2} - \frac{1}{3} \cdot r(h_k) \cdot h_k^{2m+2} = h_k^{2m+2} \cdot \underbrace{\left(\frac{4}{3} \cdot r\left(\frac{h_k}{2}\right) \cdot \frac{1}{2^{2m+2}} - \frac{1}{3} r(h_k)\right)}_{=:\tilde{r}(h_k)}$$

Man erkennt sofort, dass $\tilde{\tau}_1 = 0$ ergibt, weshalb der erste Term der Summe wegfällt. Somit bleibt für $R_{k+1,1}$ folgende asymptotische Fehlerdarstellung:

$$\int_a^b f(x)\,dx - R_{k+1,1} = \sum_{i=2}^{m} \tilde{\tau}_i h_k^{2i} + \tilde{r}(h_k) \cdot h_k^{2m+2}$$

Es ergibt sich also, dass durch Kombination von $R_{k,0}$ und $R_{k+1,0}$ der Fehler statt von h_k^2 nur noch von h_k^4 abhängt. Es ist $0 \leq 4^{1-i} \leq 1$ und somit $-1 \leq 4^{1-i} - 1 \leq 0$. Damit erhält man insgesamt

$$\left| \frac{4^{1-i} - 1}{3} \right| < 1, \ \forall i \in \mathbb{N}.$$

Somit ist $|\tilde{\tau}_i| < |\tau_i|$, und daher wird die obere Schranke der Summe der Fehlerdarstellung wegen

$$\sum_{i=2}^{m} |\tilde{\tau}_i| h_k^{2i} < \sum_{i=1}^{m} |\tau_i| h_k^{2i}$$

tatsächlich kleiner.

Dieses Prinzip lässt sich analog weiterführen, denn wenn man die Fehlerdarstellungen von $R_{k+1,1}$ und $R_{k+2,1}$ (welche h_k nur noch in mindestens vierter Potenz enthalten) in Lemma 8.2 einsetzt, um $R_{k+2,2}$ zu berechnen, erhält man unter Verwendung von $h_k = 2h_{k+1} = 4h_{k+2}$ in der Summe der Fehlerdarstellung von $R_{k+2,2}$ für jenen Term, der τ_2 enthält, Folgendes:

$$\tau_2 \left(\frac{h_k}{4} \right)^{2\cdot 2} + \frac{\tau_2 \left(\frac{h_k}{4} \right)^{2\cdot 2} - \tau_2 \left(\frac{h_k}{2} \right)^{2\cdot 2}}{\underbrace{\left(\frac{h_k}{h_{k+2}} \right)^2 - 1}_{=15}} = \tau_2 h_k^4 \cdot \left(\frac{1}{256} + \frac{\frac{1}{256} - \frac{1}{16}}{15} \right) = 0$$

Der Fehler enthält also h_k nur noch in sechster Potenz. Nachfolgender Satz stellt die hier angestellten Beobachtungen allgemein dar.

Satz 8.7. Sei $f : [a, b] \to \mathbb{R}$ eine $(2m+2)$-mal stetig differenzierbare Funktion. Dann besitzt $R_{k,j}$ mit $j \in \{0, ..., m-1\}$ und $k \geq j$ die asymptotische Fehlerdarstellung

$$\int_a^b f(x)\,dx - R_{k,j} = \sum_{i=j+1}^{m} \tau_i h_k^{2i} + r(h_k)\, h_k^{2m+2},$$

wobei $\tau_i \in \mathbb{R}$ Konstanten sind und $r(h_k)$ eine von h_k abhängige beschränkte Funktion ist.

Beweis: Die Idee für diesen Beweis wurde ansatzweise von [6, ab Seite 177] übernommen. Es wird ein Induktionsbeweis geführt. Für $j = 0$ erhält man die Aussage durch Satz 8.5 und die Tatsache, dass $R_{k,0} = T_n$ ist, für $n = 2^k$.

Als Induktionsvoraussetzung gelte nun für ein beliebiges aber festes $j - 1 \in \{0, ..., m-2\}$ und ein $k \geq j - 1$, dass $R_{k,j-1}$ die Fehlerdarstellung

$$\int_a^b f(x)\,dx - R_{k,j-1} = \sum_{i=j}^{m} \tau_i h_k^{2i} + r(h_k)\, h_k^{2m+2}$$

besitzt. Das heißt, die niedrigste Potenz von h_k hat den Exponenten $2j$.

Im Induktionsschritt $j - 1 \mapsto j$ muss also nur noch gezeigt werden, dass in der Fehlerdar-

stellung von $R_{k,j}$ das Summenglied mit h_k^{2j} wegfällt. Aus $h_{k-1} = 2h_k$ erhält man:

$$R_{k,j-1} = \int_a^b f(x)\,dx - \sum_{i=j}^{m} \tau_i h_k^{2i} - r(h_k) h_k^{2m+2}$$

$$R_{k-1,j-1} = \int_a^b f(x)\,dx - \sum_{i=j}^{m} \tau_i \cdot 2^{2i}\, h_k^{2i} - r(h_{k-1}) h_{k-1}^{2m+2}$$

Durch Einsetzen in Lemma 8.2 erhält man Folgendes:

$$R_{k,j} = \int_a^b f(x)\,dx - \sum_{i=j}^{m} \tau_i h_k^{2i} - r(h_k) h_k^{2m+2} +$$

$$+ \frac{\int_a^b f(x)\,dx - \sum_{i=j}^{m} \tau_i h_k^{2i} - r(h_k) h_k^{2m+2} - \int_a^b f(x)\,dx + \sum_{i=j}^{m} \tau_i \cdot 2^{2i}\, h_k^{2i} + r(h_{k-1}) h_{k-1}^{2m+2}}{\left(\frac{h_{k-j}}{h_k}\right)^2 - 1}$$

Im Zähler fällt das Integral weg und kurze Umformungen ergeben unter Verwendung von $h_{k-j} = 2^j \cdot h_k$ Folgendes:

$$\int_a^b f(x)\,dx - R_{k,j} = \sum_{i=j}^{m} \tau_i h_k^{2i} + r(h_k) h_k^{2m+2} +$$

$$+ \frac{\sum_{i=j}^{m} \tau_i h_k^{2i} + r(h_k) h_k^{2m+2} - \sum_{i=j}^{m} \tau_i \cdot 2^{2i}\, h_k^{2i} - r(h_{k-1}) h_{k-1}^{2m+2}}{2^{2j} - 1}$$

Für jenen Term dieser Fehlerdarstellung, welcher h_k^{2j} beinhaltet, erhält man:

$$\tau_j h_k^{2j} + \frac{\tau_j h_k^{2j} - \tau_j \cdot 2^{2j}\, h_k^{2j}}{2^{2j} - 1} = \tau_j h_k^{2j} \underbrace{\left(1 + \frac{1 - 2^{2j}}{2^{2j} - 1}\right)}_{=0} = 0.$$

Somit besitzt $R_{k,j}$ die Fehlerdarstellung

$$\int_a^b f(x)\,dx - R_{k,j} = \sum_{i=j+1}^{m} \tilde{\tau}_i h_k^{2i} + \tilde{r}(h_k)\, h_k^{2m+2}$$

für neue Konstanten $\tilde{\tau}_i \in \mathbb{R}$ und eine neue beschränkte Funktion $\tilde{r}(h_k)$, deren Gestalt den Überlegungen entnommen werden kann, welche im Vorfeld dieses Satzes angestellt wurden.

Analog folgt dies für alle anderen $R_{k,j}$ mit $k \geq j$, wodurch die Aussage bewiesen wurde. □

8.4 Zusammenhang mit Newton-Cotes-Formeln

In diesem Kapitel werden drei Beispiele vorgestellt, welche einen Zusammenhang zwischen dem Romberg-Verfahren und den abgeschlossenen Newton-Cotes-Formeln höherer Ordnung zeigen.

Beispiel 1: Es wird $R_{1,1}$ berechnet. Nach Definition 8.1 erhält man unter Verwendung von $h_0 = b - a$ und $h_1 = \frac{b-a}{2}$ folgende Näherungswerte:

$$R_{0,0} = (b-a) \cdot \left(\frac{f(a)}{2} + \frac{f(b)}{2} \right)$$

$$R_{1,0} = \frac{b-a}{2} \cdot \left(\frac{f(a)}{2} + f\left(\frac{a+b}{2} \right) + \frac{f(b)}{2} \right)$$

Durch Einsetzen in Lemma 8.2 erhält man nach kurzer Umformung ausgehend von

$$R_{1,1} = R_{1,0} + \frac{R_{1,0} - R_{0,0}}{\left(\frac{h_0}{h_1}\right)^2 - 1} = R_{1,0} + \frac{R_{1,0} - R_{0,0}}{3}$$

das Ergebnis

$$R_{1,1} = (b-a) \cdot \left(\frac{1}{6} \cdot f(a) + \frac{2}{3} \cdot f\left(\frac{b-a}{2} \right) + \frac{1}{6} \cdot f(b) \right).$$

Dies entspricht nach Definition 7.5 der Simpsonregel, also der abgeschlossenen Newton-Cotes-Formel 2. Ordnung.

Beispiel 2: Nun wird $R_{2,2}$ berechnet. Mit

$$R_{2,0} = \frac{b-a}{4} \cdot \left(\frac{f(a)}{2} + f\left(a + \frac{b-a}{4} \right) + f\left(\frac{a+b}{2} \right) + f\left(a + 3 \cdot \frac{b-a}{4} \right) + \frac{f(b)}{2} \right)$$

und $R_{1,0}$ aus dem vorigen Beispiel erhält man durch Einsetzen in Lemma 8.2 und kurzer Rechnung Folgendes:

$$R_{2,1} = (b-a) \cdot \left(\frac{1}{12} \cdot f(a) + \frac{1}{3} \cdot f\left(a + \frac{b-a}{4} \right) + \frac{1}{6} \cdot f\left(\frac{a+b}{2} \right) + \frac{1}{3} \cdot f\left(a + 3 \cdot \frac{b-a}{4} \right) + \frac{1}{12} \cdot f(b) \right)$$

Daraus folgt nun mit

$$R_{2,2} = R_{2,1} + \frac{R_{2,1} - R_{1,1}}{\left(\frac{h_0}{h_2}\right)^2 - 1} = R_{2,1} + \frac{R_{2,1} - R_{1,1}}{15}$$

nach kurzer Rechnung und durch Verwendung von $R_{1,1}$ aus dem ersten Beispiel

$$R_{2,2} = (b-a) \cdot \left(\frac{7}{90} \cdot f(a) + \frac{16}{45} \cdot f\left(a + \frac{b-a}{4} \right) + \frac{2}{15} \cdot f\left(\frac{a+b}{2} \right) + \right.$$

$$\left. + \frac{16}{45} \cdot f\left(a + 3 \cdot \frac{b-a}{4} \right) + \frac{7}{90} \cdot f(b) \right).$$

Durch Vergleich mit den Gewichten der ersten Tabelle in Kapitel 7.5 erkennt man, dass es sich hierbei um die Milne-Regel, also die abgeschlossene Newton-Cotes-Formel 4. Ordnung, handelt.

Beispiel 3: Für dieses Beispiel definiert man die Schrittweitenfolge abweichend von Definition 8.1 durch $h_i := \frac{b-a}{3^i}$ für $i \in \mathbb{N}_0$. Damit wird nun $R_{1,1}$ berechnet.

Mit den beiden Ausgangswerten

$$R_{0,0} = (b-a) \cdot \left(\frac{f(a)}{2} + \frac{f(b)}{2} \right) \quad \text{und}$$

$$R_{1,0} = \frac{b-a}{3} \cdot \left(\frac{f(a)}{2} + f\left(a + \frac{b-a}{3}\right) + f\left(a + 2 \cdot \frac{b-a}{3}\right) + \frac{f(b)}{2} \right)$$

erhält man unter Verwendung von Lemma 8.2 aus dem Ansatz

$$R_{1,1} = R_{1,0} + \frac{R_{1,0} - R_{0,0}}{\left(\frac{h_0}{h_1}\right)^2 - 1} = R_{1,0} + \frac{R_{1,0} - R_{0,0}}{8}$$

das Ergebnis

$$R_{1,1} = (b-a) \cdot \left(\frac{f(a)}{8} + \frac{3}{8} \cdot f\left(a + \frac{b-a}{3}\right) + \frac{3}{8} \cdot f\left(a + 2 \cdot \frac{b-a}{3}\right) + \frac{f(b)}{8} \right).$$

Durch erneuten Vergleich mit den Gewichten der ersten Tabelle in Kapitel 7.5 erkennt man, dass es sich hierbei um die sogenannte Pulcherrima handelt, also die abgeschlossene Newton-Cotes-Formel 3. Ordnung.

8.5 Verallgemeinerung

Dieser Abschnitt stellt abschließend eine allgemeinere Version des Romberg-Verfahrens vor, bei welcher die Schrittweitenfolge wie soeben in Beispiel 3 abweichend von $h_i = \frac{b-a}{2^i}$ definiert wird.

Definition 8.8. Sei $f : [a,b] \to \mathbb{R}$ eine Funktion und sei $(n_i)_{i \in \mathbb{N}_0}$ eine streng monoton wachsende Folge in $\mathbb{N}$. Dann definiert man die Schrittweitenfolge $(h_i)_{i \in \mathbb{N}_0}$ als $h_i := \frac{b-a}{n_i}$. Die Startwerte der Romberg-Extrapolation erhält man durch

$$R_{k,0} := h_k \cdot \left(\frac{f(a)}{2} + \sum_{i=1}^{\frac{h_0}{h_k}-1} f(a + i h_k) + \frac{f(b)}{2} \right)$$

für $k \in \mathbb{N}_0$. Für ein festes $j \in \mathbb{N}$ und $k \geq j$ definiert man die aus den Startwerten erhaltenen Näherungswerte $R_{k,j}$ folgendermaßen:

$$R_{k,j} := R_{k,j-1} + \frac{R_{k,j-1} - R_{k-1,j-1}}{\left(\frac{h_{k-j}}{h_k}\right)^2 - 1}$$

In der Praxis werden laut [9, Seite 81] meist die nachfolgenden drei Folgen zur Generierung der Schrittweite verwendet:

- Romberg-Folge: $n_i := 2^i$
- Bulirsch-Folge: $(n_0, n_1, n_2) := (1, 2, 3), n_i := 2n_{i-2}$, für $i \geq 3$
- Harmonische Folge: $n_i := i + 1$

Die Romberg-Folge $1, 2, 4, 8, 16, 32, 64, ...$ hat den Vorteil, dass für die Berechnung der nächsthöheren Schrittweite alle bisher berechneten Funktionsauswertungen übernommen werden können, da die Stützstellen nicht verworfen werden, sondern lediglich in der Mitte zweier aufeinanderfolgender Stützstellen eine neue hinzukommt. Ein Nachteil ist jedoch die rasch anwachsende Anzahl $n_i + 1$ an notwendigen Funktionsauswertungen.

Eine Verbesserung bringt die nach dem deutschen Mathematiker Roland Bulirsch (geboren 1932) benannte Bulirsch-Folge $1, 2, 3, 4, 6, 8, 12, ...$ Eine noch geringere Anzahl an Funktionsauswertungen erhält man durch die harmonische Folge $1, 2, 3, 4, 5, 6, 7, ...$ Der Nachteil an der harmonischen Folge ist jedoch, dass man im Falle von Primzahlen mit Ausnahme der Intervallgrenzen a und b auf keine bisherigen Stützstellen zurückgreifen kann.

Es wird nun nochmals das Eingangsbeispiel behandelt und die beiden neuen Variationen dieses Verfahrens ebenfalls untersucht. Die bereits oben angeführten Näherungen von $\int_0^1 e^x \, dx$ unter Verwendung von 257 Stützstellen sind:

- Simpsonregel: 1.71828182846**1** (10 Nachkommastellen)
- Milne-Regel: 1.7182818284590452**48** (16 Nachkommastellen)
- $R_{8,8} = 1.71828182845904523536028747135266250**07}$ (33 Nachkommastellen)

Unter Verwendung der Bulirsch-Folge erhält man für $R_{10,10}$ den folgenden Näherungswert:

- $R_{10,10} = 1.71828182845904523536028747135266249**89**$ (36 Nachkommastellen)

Da bei diesem Verfahren bei einer bestimmten Schrittweite nicht alle Stützstellen der direkt vorhergehenden Schrittweite übernommen werden, sondern die Schrittweite erst bei der übernächsten Näherung halbiert wird (die verwendeten Stützstellen alternieren also in gewisser Weise), müssen zur konkreten Bestimmung der Anzahl der benötigten Funktionsauswertungen die Bruchfolgen $\frac{k}{32}$ für $k \in \{0, ..., 32\}$ und $\frac{k}{48}$ für $k \in \{0, ..., 48\}$ verglichen werden, denn $h_9 = 32$ und $h_{10} = 48$. Es ergibt sich, dass hier insgesamt 65 verschiedene Stützstellen vorliegen. Man beachte, dass man mit 192 Stützstellen weniger drei richtige Nachkommastellen mehr erhält als mit der Romberg-Folge.

Mit der harmonischen Folge erhält man für $R_{11,11}$ den folgenden Näherungswert:

- $R_{11,11} = 1.7182818284590452353602874713526624**987**$ (35 Nachkommastellen)

Aufgrund der in der harmonischen Folge vorkommenden Primzahlen müssen hier die Bruchfolgen $\frac{k}{n}$ für $k \in \{1, ..., n\}$ und $n \in \{1, ..., 12\}$ verglichen werden und die Anzahl der unterschiedlichen Elemente gezählt werden. Man kommt zu dem Ergebnis, dass insgesamt 47 Stützstellen verwendet wurden. Damit konnte mit einer nochmals um 18 verringerten Anzahl an Funktionsauswertungen die Genauigkeit nahezu beibehalten werden.

Es erscheint Aufgrund der damit verbundenen Verringerung des Rechenaufwands sinnvoll zu sein, eine Folge zu wählen, welche bei gleichbleibender Genauigkeit die Anzahl der Stützstellen möglichst gering hält. Dazu werden nachfolgend drei neue Folgen untersucht:

1. Die erste Folge enthält 1 und alle geraden Zahlen, also $1, 2, 4, 6, 8, 10, ...$
2. Die zweite Folge beginnt mit 1. Bei den weiteren Folgengliedern handelt es sich um die durch 2 oder durch 3 teilbaren natürlichen Zahlen. Die ersten Glieder der Folge sind also $1, 2, 3, 4, 6, 8, 9, 10, 12, 14, 15, 16, 18, 20, ...$
3. Zuletzt wird jene Folge untersucht, welche neben der Zahl 1 alle Zahlen enthält, die ausschließlich aus den Primfaktoren 2 und 3 gebildet werden, also $1, 2, 3, 4, 6, 8, 9, 12, 16, 18, 24, 27, 32, 36, ...$

Die erste Folge erzeugt bei insgesamt 65 Funktionsauswertungen eine Näherung mit 35 korrekten Nachkommastellen:

- $R_{10,10} = 1.71828182845904523536028747135266249\color{gray}{84}$

Die zweite Folge erzeugt bei insgesamt 63 Funktionsauswertungen eine Näherung mit 37 korrekten Nachkommastellen:

- $R_{11,11} = 1.718281828459045235360287471352662497\color{gray}{770}$

Die letzte Folge erzeugt bei insgesamt 45 Funktionsauswertungen eine Näherung mit 33 korrekten Nachkommastellen:

- $R_{10,10} = 1.71828182845904523536028747135266\color{gray}{251}$

Für 39 Nachkommastellen werden hingegen bereits 63 Funktionsauswertungen benötigt:

- $R_{11,11} = 1.718281828459045235360287471352662497757\color{gray}{68}$

Man erkennt also, dass durch diese drei neu konstruierten Folgen zumindest für dieses konkrete Beispiel keine Verbesserung erzielt werden kann. Am effizientesten scheinen die harmonische Folge und die letzte der drei neuen Folgen zu sein, wobei weitere Berechnungen von $R_{n,n}$ mit $n > 11$ ergeben haben, dass die harmonische Folge zunehmend effizienter wird. Es muss jedoch berücksichtigt werden, dass alle hier durchgeführten Vergleiche anhand eines einzigen Integranden zustande gekommen sind. Somit kann man daraus bestenfalls eine Tendenz ableiten, jedoch keinen allgemeingültigen Zusammenhang aufstellen.

9 Gauß-Verfahren

Während bei den bisher vorgestellten numerischen Integrationsverfahren das Integrationsintervall $[a, b]$ in äquidistante Teilintervalle zerlegt wurde und die Stützstellen somit jeweils denselben Abstand zueinander hatten, sind sie bei den diversen Vertretern des Gauß-Verfahrens nicht mehr gleichmäßig verteilt.

Das Ziel ist es, auf diese Weise Quadraturformeln zu erhalten, welche Polynome mit möglichst hohem Grad exakt integrieren. Konkret ist es durch das Gauß-Verfahren möglich, mit $n+1$ Stützstellen Polynome vom Grad $2n+1$ exakt zu integrieren. Es wird außerdem zunächst gezeigt werden, dass dies zugleich die größtmögliche Ausbeute darstellt, die man mit einer Quadraturformel erreichen kann.

Zur Erinnerung: Mit den abgeschlossenen Newton-Cotes-Formeln n-ter Ordnung (also mit $n+1$ Stützstellen) konnte man Polynome vom Grad $n+1$ exakt integrieren, falls n gerade war (ansonsten nur Grad n).

Beim Gauß-Verfahren wird die zu integrierende Funktion f zerlegt in das Produkt einer Gewichtsfunktion α (siehe Kapitel 5.1) und einer weiteren Funktion g, also $f = \alpha \cdot g$. Für spezielle Polynome p gilt dann

$$\int_a^b f(x)\,dx = \int_a^b \alpha(x) \cdot g(x)\,dx \approx \int_a^b \alpha(x) \cdot p(x)\,dx = \sum_{i=0}^{n} w_i g(x_i).$$

9.1 Maximale Fehlerordnung

Der nächste Satz soll zeigen, dass keine Quadraturformel mit $n+1$ Stützstellen existiert, die alle Polynome mit Grad $2n+2$ exakt integriert.

Satz 9.1. Sei $[a, b]$ ein Intervall und seien $x_0, ..., x_n \in [a, b]$ beliebige Stützstellen. Sei $Q = \sum_{i=0}^{n} w_i f(x_i)$ eine Quadraturformel mit $n+1$ Stützstellen. Dann gibt es Polynome mit Grad $2n+2$, die nicht exakt integriert werden.

Beweis: Die Beweisführung erfolgt in Anlehnung an [6, Seite 180, erster Absatz]. Als Gewichtsfunktion wird $\alpha(x) = 1$ gewählt. Es gilt $\alpha(x) > 0$, $\forall x \in \mathbb{R}$. Man definiert nun $f(x) := g(x) = \prod_{i=0}^{n}(x - x_i)^2$. Dabei handelt es sich um ein Polynom vom Grad $2n+2$. Es ist außerdem $g(x) \geq 0$, $\forall x \in [a, b]$, wobei g nur endlich viele Nullstellen $x_0, ..., x_n$ besitzt. Somit ist $\alpha(x)g(x) \geq 0$ mit nur endlich vielen Nullstellen. Daraus folgt

$$\int_a^b f(x)\,dx = \int_a^b \alpha(x)g(x)\,dx > 0.$$

Andererseits erhält man durch die Quadraturformel

$$Q(f) = \sum_{i=0}^{n} w_i \underbrace{f(x_i)}_{=0} = 0.$$

Wegen $\int_a^b f(x)\,dx \neq Q(f)$ erkennt man, dass es Polynome mit Grad $2n+2$ gibt, die nicht exakt integriert werden. □

Nachfolgend wird eine hinreichende Bedingung hergeleitet, um eine Quadraturformel mit $n+1$ Stützstellen zu erhalten, welche alle Polynome mit Höchstgrad $2n+1$ exakt integriert.

Satz 9.2. Sei α eine Gewichtsfunktion und sei $p \neq 0$ ein Polynom mit Grad $n+1$, welches $n+1$ paarweise verschiedene Nullstellen besitzt und zu allen Polynomen mit Höchstgrad n orthogonal bezüglich α ist. Seien $x_0, ..., x_n$ die Nullstellen von p und seien die Gewichte definiert durch $w_i := \int_a^b \alpha(x)\ell_i(x)\,dx$ mit

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

Dann integriert die Quadraturformel $Q(f) = \sum_{i=0}^{n} w_i g(x_i)$ alle Funktionen f mit $f = g \cdot \alpha$ exakt, wobei g ein Polynom mit Höchstgrad $2n+1$ ist.

Beweis: Grundlage des Beweises ist [6, Seite 180, Satz 6.3.1]. Es sei g ein Polynom mit Grad $2n+1$. Durch Division mit Rest von g durch p erhält man $g = p \cdot q + r$, wobei q, r Polynome mit Höchstgrad n sind, denn nach Definition der Division mit Rest muss $\operatorname{grad}(r) < \operatorname{grad}(p) = n+1$ gelten und da $\operatorname{grad}(g) = 2n+1$ und $\operatorname{grad}(p) = n+1$ ist, kann der Grad von q höchstens n sein.

Da $x_0, ..., x_n$ die Nullstellen von p sind, gilt $p(x_i) = 0$. Daraus folgt $g(x_i) = p(x_i) \cdot q(x_i) + r(x_i) = r(x_i)$ für $i \in \{0, ..., n\}$.

Weil p nach Definition α-orthogonal zu allen Polynomen ψ mit Höchstgrad n ist, und somit die Eigenschaft $\langle p, \psi\rangle_\alpha = 0$ erfüllt ist, gilt insbesondere $\langle p, q\rangle_\alpha = 0$. Man erhält

$$\int_a^b f(x)\,dx = \int_a^b \alpha(x)g(x)\,dx = \underbrace{\int_a^b \alpha(x)p(x)q(x)\,dx}_{=:\langle p,q\rangle_\alpha = 0} + \int_a^b \alpha(x)r(x)\,dx = \int_a^b \alpha(x)r(x)\,dx.$$

Da Q eine Quadraturformel mit $n+1$ Stützstellen ist, werden dadurch nach Korollar 6.10 Polynome mit Höchstgrad n exakt integriert. Deshalb gilt $\int_a^b \alpha(x)r(x)\,dx = \sum_{i=0}^{n} w_i r(x_i)$ und somit insgesamt

$$\int_a^b f(x)\,dx = \int_a^b \alpha(x)g(x)\,dx = \int_a^b \alpha(x)r(x)\,dx = \sum_{i=0}^{n} w_i r(x_i) = \sum_{i=0}^{n} w_i g(x_i).$$

Somit wird die Funktion f exakt integriert. □

Findet man solche Polynome p mit Grad $n+1$, welche zu allen Polynomen mit Höchstgrad n orthogonal bezüglich einer Gewichtsfunktion α sind, so kann diese maximale Fehlerordnung erreicht werden.

Derartige Polynome können durch die in Satz 5.13 definierte Rekursion generiert werden. In Kapitel 5.3 wurden einige bedeutsame Folgen orthogonaler Polynome vorgestellt, welche im Anschluss näher behandelt werden.

Definition 9.3. Sei $\alpha : [a, b] \to \mathbb{R}$ eine Gewichtsfunktion. Sei p ein Polynom mit Grad $n+1$, das zu allen x^k mit $k \in \{0, ..., n\}$ orthogonal bezüglich α ist. Seien $x_0, ..., x_n$ die paarweise verschiedenen Nullstellen von p. Dann definiert man die Gewichte $w_0, ..., w_n$ durch

$$w_i := \int_a^b \alpha(x) \prod_{\substack{k=0 \\ k \neq i}}^{n} \frac{x - x_k}{x_i - x_k}.$$

Man bezeichnet die Quadraturformel

$$Q_n(f) := \sum_{i=0}^{n} w_i f(x_i)$$

als Gauß-Quadratur n-ter Ordnung, für eine stetige Funktion $f : [a, b] \to \mathbb{R}$.

Nachfolgend werden einige spezielle Gauß-Quadraturen aufgelistet.

Definition 9.4. Man unterscheidet folgende spezielle Klassen von Gauß-Quadraturen:

(i) Für die in Kapitel 5.4 behandelten Legendre-Polynome spricht man von der Gauß-Legendre-Quadratur, welche als Integrationsintervall $[-1, 1]$ verwendet.

(ii) Für die in Kapitel 5.5 behandelten Tschebyschow-Polynome spricht man von der Gauß-Tschebyschow-Quadratur, welche als Integrationsintervall $[-1, 1]$ verwendet.

(iii) Für die in Kapitel 5.6 behandelten Laguerre-Polynome spricht man von der Gauß-Laguerre-Quadratur, welche als Integrationsintervall $[0, \infty)$ verwendet.

(iv) Für die in Kapitel 5.7 behandelten Hermite-Polynome spricht man von der Gauß-Hermite-Quadratur, welche als Integrationsintervall $\mathbb{R}$ verwendet.

Die hier verwendeten Polynome sind stets nur für ein bestimmtes Intervall orthogonal. Um für ein vorgegebenes Integrationsintervall eine Quadraturformel zu erhalten, kann man nun entweder mit der in Satz 5.13 definierten Rekursion für das gewünschte Intervall unter Verwendung einer geeigneten Gewichtsfunktion selbst ein Polynom konstruieren, oder man transformiert mittels linearer Substitution die Integralgrenzen auf das Intervall $[-1, 1]$ und verwendet die Gauß-Legendre-Quadratur. Nachfolgend wird diese Transformation beschrieben.

Lemma 9.5. Seien $[a, b]$ und $[A, B]$ zwei Intervalle. Sei $f : [a, b] \to \mathbb{R}$ stetig. Definiere

$$\varphi : [A, B] \to [a, b], \quad x \mapsto \frac{a - b}{A - B} \cdot (x - A) + a.$$

Dann gilt folgende Transformation:

$$\int_a^b f(x)\, dx = \frac{a - b}{A - B} \int_A^B f\left(\frac{a - b}{A - B} \cdot (t - A) + a\right) dt$$

Beweis: Zunächst muss gezeigt werden, dass die Bildmenge der oben definierten Funktion tatsächlich $[a, b]$ ist. Für $x \in [A, B]$ gilt

$$\frac{a-b}{A-B} \cdot \underbrace{(x-A)}_{\in[0,B-A]} + a = \underbrace{\frac{b-a}{B-A} \cdot (x-A)}_{\in[0,b-a]} + a \in [a, b]$$

Da φ ein Polynom ist, ist φ stetig differenzierbar. Daher gilt aufgrund der Substitutionsregel der Integration (Satz 2.15)

$$\int_{\varphi(A)}^{\varphi(B)} f(x)\,dx = \int_A^B f(\varphi(t)) \cdot \varphi'(t)\,dt.$$

Es ist $\varphi(A) = a$ und $\varphi(B) = b$. Außerdem ist $\varphi'(x) = \frac{a-b}{A-B}$. Somit erhält man

$$\int_a^b f(x)\,dx = \int_A^B f\left(\frac{a-b}{A-B} \cdot (t-A) + a\right) \cdot \frac{a-b}{A-B}\,dt =$$

$$= \frac{a-b}{A-B} \int_A^B f\left(\frac{a-b}{A-B} \cdot (t-A) + a\right)\,dt.$$

□

Dieses Resultat wird nun verwendet, um eine Transformation auf das bei der Gauß-Legendre-Quadratur verwendete Integral $[-1, 1]$ herzuleiten.

Korollar 9.6. Es gilt

$$\int_a^b f(x)\,dx = \frac{b-a}{2} \int_{-1}^1 f\left(\frac{b-a}{2} \cdot t + \frac{a+b}{2}\right)\,dt.$$

Beweis: Der Beweis folgt sofort, indem man in obige Transformationsformel $A := -1$ und $B := 1$ einsetzt. □

In Kapitel 9.3 wird diese Transformation in Beispiel 1 (siehe Seite 107) verwendet.

Lemma 9.7. Die Gewichte w_i mit $i \in \{0, ..., n\}$ der Gauß-Tschebyschow-Quadratur n-ter Ordnung sind gemäß [12, Seite 200] gegeben durch

$$w_i = \frac{\pi}{n+1}.$$

Es ergibt sich also, dass die Gewichte der Gauß-Tschebyschow-Quadratur für ein festes Tschebyschow-Polynom identisch sind.

9.2 Konvergenz

In diesem Kapitel wird gezeigt, dass für eine Folge $(p_n)_{n\in\mathbb{N}_0}$ von orthogonalen Polynomen mit $n \to \infty$ die Abweichung des Näherungswertes $Q_n(f)$ vom tatsächlichen Wert des

Integrals für stetige Funktionen verschwindet, also

$$\lim_{n\to\infty}\left(\int_a^b f(x)\,dx - Q_n(f)\right) = 0.$$

Dazu werden zunächst einige grundlegende Eigenschaften von orthogonalen Polynomen im Zusammenhang mit den Gewichten der Gauß-Quadratur behandelt. Insbesondere wird hier gezeigt, dass die Nullstellen der orthogonalen Polynome paarweise verschieden sind (was bereits zuvor gelegentlich erwähnt wurde).

Satz 9.8. Sei $\alpha : [a,b] \to \mathbb{R}$ eine Gewichtsfunktion und sei p ein Polynom mit Grad $n+1$, welches zu allen x^k mit $k \in \{0,...,n\}$ orthogonal bezüglich α ist. Seien $x_0,...,x_n$ die Nullstellen von p. Dann definiere die Gewichte $w_0,...,w_n$ durch

$$w_i := \int_a^b \alpha(x) \prod_{\substack{k=0 \\ k\neq i}}^{n} \frac{x - x_k}{x_i - x_k}\,dx.$$

Es gelten folgende Eigenschaften:

(i) Es ist $x_i \in (a,b)$. Außerdem sind die Nullstellen reell und paarweise verschieden.

(ii) Alle Gewichte sind positiv, d. h. $w_i > 0$, $\forall i \in \{0,...,n\}$.

(iii) Es gilt $\sum_{i=0}^{n} w_i = \int_a^b \alpha(x)\,dx$.

Beweis: Die Beweisführung erfolgt anhand von [6, Seite 181, Satz 6.3.2].

(i) Da p zu x^k für $k \in \{0,...,n\}$ nach Definition α-orthogonal ist, gilt

$$\int_a^b x^k \alpha(x) p(x)\,dx = 0.$$

Für $k = 0$ erhält man

$$\int_a^b \alpha(x) p(x)\,dx = 0.$$

Aus $\alpha(x) > 0$ für $x \in (a,b)$ folgt aus der letzten Gleichung, dass p in (a,b) zumindest einen Vorzeichenwechsel hat (sonst könnte das Integral nicht Null ergeben). Sei m definiert als die Anzahl der Vorzeichenwechsel von p in (a,b). Setze $t_0 := a$ und $t_{m+1} := b$. Dann gibt es paarweise verschiedene $t_1,...,t_m \in (a,b)$, sodass p auf (t_i, t_{i+1}) konstantes Vorzeichen hat.

Definiere $q(x) := \prod_{i=1}^{m}(x - t_i)$. Es ist $\text{grad}(q) = m$. Außerdem hat dieses Polynom dieselben Vorzeichenwechsel wie p, weil alle Nullstellen paarweise verschieden sind. Somit gilt $p(x) \cdot q(x) \geq 0,\ \forall x \in (a,b)$ oder $p(x) \cdot q(x) \leq 0,\ \forall x \in (a,b)$, da die beiden Funktionen entweder überall gleiches oder entgegengesetztes Vorzeichen aufweisen.

Da es nur endliche viele Nullstellen gibt, gilt demnach

$$\int_a^b q(x)\alpha(x)p(x)\,dx \neq 0.$$

Da dies für $m < n+1$ ein Widerspruch wäre zur α-Orthogonalität von p bezüglich aller Polynome x^k mit $k \in 0,...,n$ und somit auch zu deren Linearkombinationen, muss $m \geq n+1$ gelten. Somit hat p mindestens $n+1$ Vorzeichenwechsel auf (a,b) und damit laut

Zwischenwertsatz (Satz 2.12) mindestens $n+1$ paarweise verschiedene reelle Nullstellen in (a,b).

Bemerkung: Aufgrund des Fundamentalsatzes der Algebra (siehe Satz 2.20) hat p höchstens $n+1$ reelle Nullstellen und somit hat p genau $n+1$ reelle Nullstellen.

(ii) Man definiere

$$p_k(x) := \frac{p(x)}{x - x_k}$$

für $k \in \{0, ..., n\}$. Es ist wegen (i) x_k eine einfache Nullstelle von p. Somit gilt $p_k(x_i) = 0$ für $i \neq k$. An der Stelle x_k hat p_k keine Polstelle, da in $p(x)$ ebenfalls der Linearfaktor $(x - x_k)$ vorkommt. Die Funktion p_k kann daher an der Stelle x_k stetig fortgesetzt werden mit $p_k(x_k) \neq 0$. Außerdem ist $p_k(x)^2 \geq 0, \forall x \in \mathbb{R}$ mit nur endlich vielen Nullstellen und es gilt $p_k^2 \in \mathcal{P}_{2n}$. Insbesondere wird p_k^2 von Q_n exakt integriert. Somit erhält man

$$0 < \int_a^b \alpha(x) p_k(x)^2 \, dx = \sum_{i=0}^{n} w_i p_k(x_i)^2 = w_k \underbrace{p_k(x_k)^2}_{>0} \Rightarrow w_k > 0.$$

(iii) Setze $f(x) := 1$. Da dieses Polynom mit Grad 0 ebenfalls exakt integriert wird, erhält man

$$\int_a^b \alpha(x) \underbrace{f(x)}_{=1} \, dx = \sum_{i=0}^{n} w_i \underbrace{f(x_i)}_{=1} \Rightarrow \int_a^b \alpha(x) \, dx = \sum_{i=0}^{n} w_i.$$

□

Satz 9.9. Sei f eine stetige Funktion und sei $(p_n)_{n \in \mathbb{N}_0}$ mit $\text{grad}(p_n) = n+1$ eine Polynomfolge, deren Glieder die Bedingungen von Satz 9.2 erfüllen. Seien Q_n die zugehörigen Quadraturformeln (ebenfalls gemäß Satz 9.2). Dann gilt

$$\lim_{n \to \infty} Q_n(f) = \int_a^b f(x) \, dx.$$

Beweis: Die Beweisführung erfolgt in Anlehnung an [6, Seite 186, Satz 6.3.7]. Sei $\varepsilon > 0$ eine beliebige reelle Zahl. Nach dem Approximationssatz von Weierstraß (siehe Satz 2.17) gibt es ein Polynom q mit

$$\|f - q\|_\infty < \varepsilon.$$

Sei $m := \text{grad}(q)$. Dann wähle ein $p_k \in (p_n)_{n \in \mathbb{N}_0}$ mit $2k + 1 \geq m$. Weil Q_k Polynome bis inklusive Grad $2k+1$ exakt integriert, gilt

$$Q_k(q) = \int_a^b \alpha(x) q(x) \, dx.$$

Damit erhält man unter Verwendung der Dreiecksungleichung $|x \pm y| \leq |x| + |y|$ für $x, y \in \mathbb{R}$ Folgendes:

$$\left| \int_a^b \alpha(x) f(x) \, dx - Q_k(f) \right| = \left| \int_a^b \alpha(x) f(x) \, dx - \int_a^b \alpha(x) q(x) + Q_k(q) - Q_k(f) \right| \leq$$

$$\leq \underbrace{\left|\int_a^b \alpha(x)f(x)\,dx - \int_a^b \alpha(x)q(x)\,dx\right|}_{=\left|\int_a^b \alpha(x)(f(x)-q(x))\,dx\right| \leq \int_a^b \alpha(x)|f(x)-q(x)|\,dx} + \underbrace{|(Q_k(q) - Q_k(f)|}_{=\sum_{i=0}^n w_i|q(x_i)-f(x_i)|}$$

Es gilt für alle $x \in \mathbb{R}$, dass $|q(x) - f(x)| = |f(x) - q(x)| < \varepsilon$. Daher ist obiger Ausdruck kleiner als

$$\varepsilon \int_a^b \alpha(x)\,dx + \varepsilon \sum_{i=0}^n w_i$$

Wegen Satz 9.8 (iii) gilt $\sum_{i=0}^n w_i = \int_a^b \alpha(x)\,dx$. Deshalb erhält man insgesamt

$$\left|\int_a^b \alpha(x)f(x)\,dx - Q_k(f)\right| < 2\varepsilon \int_a^b \alpha(x)\,dx.$$

Weil ε beliebig gewählt wurde, erhält man die zu zeigende Konvergenz. □

9.3 Konkrete Rechenbeispiele

Ein MATHEMATICA-Code zur Anwendung der verschiedenen Gauß-Verfahren wird in Kapitel 11.4 vorgestellt und ausführlich erklärt. Nachfolgend werden einige Rechenbeispiele exemplarisch behandelt und deren Resultate mit anderen Verfahren zur numerischen Integration verglichen.

Beispiel 1: Dieses Beispiel soll einerseits den Umgang mit der in Lemma 9.6 hergeleiteten Transformationsformel zeigen und andererseits die Gauß-Quadratur mit den bisher behandelten Verfahren vergleichen.

Es soll die Funktion $f(x) := \cos x$ im Intervall $[-\frac{\pi}{2}, \frac{\pi}{2}]$ integriert werden. Dazu wird die Gauß-Legendre-Quadratur 4. Ordnung verwendet. Das zugehörige Legendre-Polynom hat den Grad 5 und somit 5 Nullstellen. Es gilt dann laut Lemma 9.6

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos x\,dx = \frac{\pi}{2}\int_{-1}^{1} \cos\left(\tfrac{\pi}{2}\,t\right)\,dt \approx \frac{\pi}{2}\sum_{i=0}^{n} w_i \cos\left(\tfrac{\pi}{2}\,x_i\right).$$

Das Legendre-Polynom mit Grad 5 lautet

$$x^5 - \tfrac{10}{9}x^3 + \tfrac{5}{21}x.$$

Die Nullstellen wurden mittels MATHEMATICA ermittelt, und die Gewichte wurden gemäß Definition 9.3 berechnet. Man erhält dann folgenden Näherungswert:

$$Q_4(f) = 2.000000\,1102$$

Das exakte Ergebnis lautet 2. Somit erhält man mit fünf Funktionsauswertungen sechs korrekte Nachkommastellen. Unter Verwendung gleich vieler Stützstellen bzw. Funktionsauswertungen erhält man durch die Milne-Regel (welche zugleich dem Näherungswert $R_{2,2}$ des Romberg-Verfahrens entspricht), die Simpson-Regel auf zwei Teilintervallen und die Trapezregel auf vier Teilintervallen (was beim Romberg-Verfahren dem Näherungswert $R_{2,0}$

entspricht) folgende Ergebnisse:

- Milne-Regel: 1.9985707318
- Simpson-Regel: 2.0045597549
- Trapezregel: 1.8961188979

Man sieht also sofort, dass (zumindest in diesem Beispiel) die Gauß-Quadratur mit gleicher Anzahl an Stützstellen wesentlich exaktere Näherungswerte liefert.

Beispiel 2: Dieses Beispiel soll nun die in Satz 9.2 gezeigte Exaktheit der Gauß-Formeln veranschaulichen. Dazu wird das Tschebyschow-Polynom mit Grad 4 verwendet, also

$$p(x) = x^4 - x^2 + \tfrac{1}{8},$$

um die Funktion

$$f(x) = \alpha(x) \cdot g(x) := \tfrac{x^7 - x^6 + x^4}{\sqrt{1-x^2}}$$

im Integral $[-1, 1]$ zu integrieren. Das exakte Ergebnis ist $\frac{\pi}{16}$. Die vier Nullstellen von p sind

$$\pm\frac{\sqrt{2+\sqrt{2}}}{2} \qquad \text{und} \qquad \pm\frac{\sqrt{2-\sqrt{2}}}{2}.$$

Die Gewichte $w_0, ..., w_3$ sind laut Lemma 9.7 allesamt $\frac{\pi}{4}$. Tatsächlich erhält man durch die Quadraturformel $\sum_{i=0}^{3} w_i g(x_i)$ das exakte Ergebnis $\frac{\pi}{16}$. Es wurde also mit vier Stützstellen ein Polynom g mit Grad 7 exakt integriert.

Für die Funktion

$$f(x) = \alpha(x) \cdot g(x) := \tfrac{x^8}{\sqrt{1-x^2}},$$

deren Integral im Intervall $[-1, 1]$ den Wert $\frac{35\pi}{128} \approx 0.85903$ hat, erhält man mit derselben Quadraturformel die Näherung 0.834485. Man erkennt also, dass Polynome mit Grad 8 nicht exakt integriert werden.

Bemerkung: Durch die abgeschlossenen Newton-Cotes-Formeln könnte man die beiden hier behandelten Integrale überhaupt nicht näherungsweise berechnen, da die zu integrierenden Funktionen an den Stützstellen $x_0 = -1$ und $x_n = 1$ nicht definiert sind.

Beispiel 3: Durch dieses Beispiel soll einerseits die in Satz 9.9 gezeigte Konvergenz für p_n mit $n \to \infty$ plausibel gemacht werden. Dafür werden erneut die Tschebyschow-Polynome verwendet, da diese einfach mittels MATHEMATICA generierbar sind und weil deren Nullstellen durch Lemma 9.7 sehr bequem berechnet werden können. Somit kann auch für große n die Näherung relativ schnell berechnet werden. Andererseits soll dieses Beispiel aber auch zeigen, dass trotz des in Satz 9.2 gezeigten maximalen Exaktheitsgrades die Gauß-Formeln nicht zwangsweise besser als andere Quadraturformeln sind.

Konkret soll das Integral $\int_{-1}^{1} \cos(x)\, dx$ berechnet werden. Der tatsächliche Wert lautet auf 30 Nachkommastellen genau 1.68294196961579301330500464326 0.

Mit der Gauß-Tschebyschow-Quadratur erhält man folgende Näherungswerte in Abhängigkeit der Ordnung n der verwendeten Quadraturformel Q_n (also mit $n + 1$ Stützstellen):

Als Vergleich dazu erhält man mit der Trapezregel, der Simpsonregel und der Milne-Regel für 101 Stützstellen folgende Näherungswerte:

n	Q_n
10	1.687337547772424345543084720881
100	1.682986403000470632244385012976
1000	1.682942413996157917157917310315
10000	1.682941974059601312925322485040
100000	1.682941969660231096766264769691
1000000	1.682941969616237394139663750218

- Trapezregel: 1.682885871176148487361264906318
- Simpsonregel: 1.682941971111812669372209799409
- Milne-Regel: 1.682941969615565011381872056779

Mit dem Romberg-Verfahren (unter Verwendung der Romberg-Folge) erhält man durch die Näherung $R_{9,9}$ die vollständigen 30 Nachkommastellen (mit 513 Stützstellen).

Man sieht also, dass die Gauß-Quadratur durchaus Vorteile haben kann – z. B. wenn der Integrand an den Grenzen nicht definiert ist oder man eine Funktion integrieren möchte, welche dem Produkt einer Gewichtsfunktion und eines Polynoms entspricht (da man hier bei genügend hohem Grad ein exaktes Ergebnis erhält) – aber die einfachen Newton-Cotes-Formeln und das Romberg-Verfahren ebenfalls ihre Berechtigung haben.

Beispiel 4: Dieses Beispiel soll schließlich das uneigentliche Integral $\int_{-\infty}^{\infty} f(x)\,dx$ näherungsweise berechnen, was mit den bisher behandelten Verfahren nicht möglich wäre.

Konkret wurde hierfür folgende Funktion ausgewählt:

$$f(x) := \frac{1}{2^{(x^2)}}$$

Der exakte Wert des Integrals lautet

$$\int_{-\infty}^{\infty} \frac{1}{2^{(x^2)}}\,dx = \sqrt{\frac{\pi}{\log 2}} \approx 2.12893403886245235863.$$

In Abhängigkeit der Ordnung n der verwendeten Quadraturformel Q_n (also für $n+1$ Stützstellen) erhält man mittels Gauß-Hermite-Quadratur folgende Näherungswerte:

- $Q_5(f) = 2.128575544220681906398$
- $Q_{10}(f) = 2.128933970181098756 82$
- $Q_{15}(f) = 2.128934038849139361 83$

10 Ausblick

In diesem Kapitel werden Ansätze für weitere Verfahren der numerischen Integration vorgestellt, welche bisher in dieser Arbeit nicht behandelt wurden.

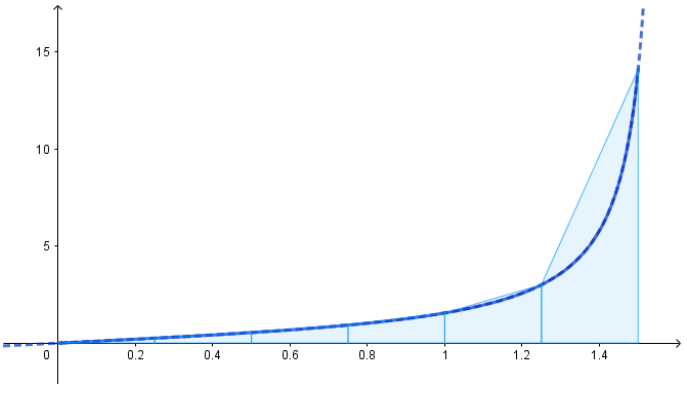
10.1 Schrittweitensteuerung

Bei den bisherigen Verfahren waren die Stützstellen entweder äquidistant (Newton-Cotes-Formeln und Romberg-Verfahren) oder entsprachen den Nullstellen bestimmter Polynome (Gauß-Quadratur). Einen weiteren Ansatz stellt die Schrittweitensteuerung dar. Hierbei ist die Wahl der Stützstellen von der konkret vorliegenden Funktion abhängig. Im Falle der zusammengesetzten Trapezregel ist es beispielsweise sinnvoll, in jenen Bereichen, in denen die Funktion eine betragsmäßig große Steigung aufweist, mehr Stützstellen zu wählen, während in annähernd konstanten Bereichen die Anzahl reduziert werden kann. Auf diese Weise kann für dieselbe Anzahl an Stützstellen die Genauigkeit verbessert werden. Folgendes Beispiel soll dies veranschaulichen.

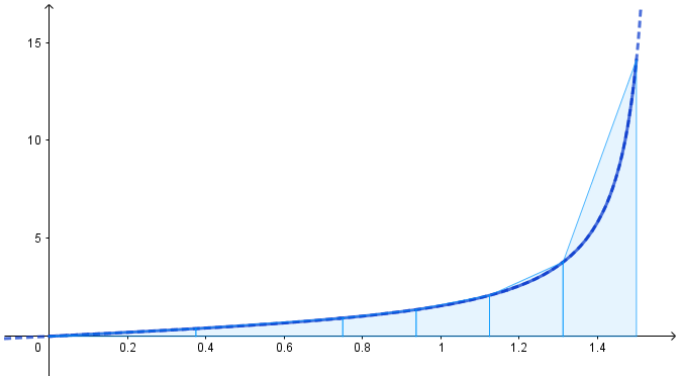
Beispiel: In diesem Beispiel soll das Integral

$$\int_0^{1.5} \tan x \, dx$$

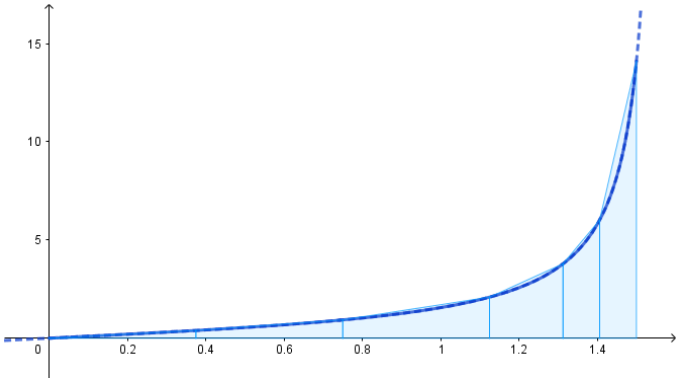
näherungsweise berechnet werden. Der tatsächliche Wert beträgt ungefähr 2.649. Verwendet man die zusammengesetzte Trapezregel mit $N = 6$, also sechs äquidistanten Teilintervallen und damit sieben Stützstellen (siehe Abbildung), so erhält man den Näherungswert 3.338.



Anhand des Funktionsgraphen erkennt man jedoch, dass dieser im linken Bereich eine geringe Steigung aufweist, während er im rechten Bereich stark ansteigt. Somit erscheint es sinnvoll, dort eine höhere Stützstellendichte zu verwenden. Benutzt man als Ausgangsbasis $N = 4$ (fünf Stützstellen) und fügt in die Mitte der beiden rechten Teilintervalle eine weitere Stützstelle hinzu (siehe Abbildung auf nächster Seite), so erhält man die Näherung 3.089, was bei gleicher Anzahl an Stützstellen den Fehler um ca. 36 % reduziert.



Eine weitere Verbesserung kann erreicht werden, wenn man erneut als Ausgangsbasis $N = 4$ verwendet, das letzte der vier Teilintervalle halbiert und von den beiden dadurch entstehenden Teilintervallen erneut das letzte halbiert. Hier beträgt der Näherungswert 2.843, womit der Fehler im Vergleich zum ersten Näherungswert um 72 % verringert wurde.



Eine konkrete Realisierung dieser Idee findet man unter [4, Seite 120 bis 122]. Hier wird eine Formel hergeleitet, mit welcher man die Verbesserung der Näherung bei Hinzunahme einer weiteren Stützstelle in der Mitte von zwei ursprünglichen Stützstellen abschätzen kann, ohne dabei die Ableitung der Funktion kennen zu müssen. Liegt diese Verbesserung betragsmäßig über einem vordefinierten Wert, so sollte man die neue Stützstelle hinzunehmen. Es bleibt jedoch fraglich, was genau an dieser Methode zu einer Erleichterung führt, da man zur Abschätzung der Verbesserung ohnehin den Funktionswert in der Mitte des Teilintervalls berechnen muss und man diesen somit eigentlich gleich hinzunehmen könnte.

10.2 Monte-Carlo-Methode

Die Monte-Carlo-Methode ist sinnvoll für Mehrfachintegrale oder falls eine geringe Genauigkeit ausreichend ist. Grob gesprochen werden n Zufallspunkte des Integrationsbereichs ausgewählt, für welche der Funktionswert ermittelt wird. Von diesen n Funktionswerten wird anschließend der arithmetische Mittelwert berechnet und mit dem Inhalt (Intervalllänge, Flächeninhalt, ...) multipliziert.

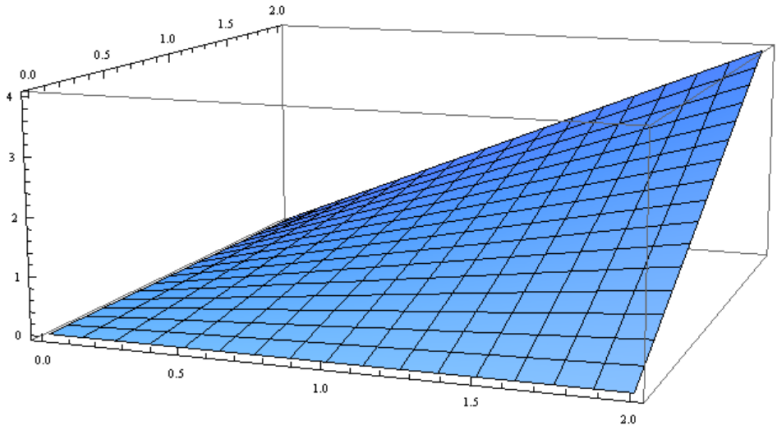
Für den zweidimensionalen Fall sieht dies für einen rechteckigen Integrationsbereich folgendermaßen aus: Sei $M := [a, b] \times [c, d] \subseteq \mathbb{R}^2$ und $f : M \to \mathbb{R}$ eine stetige Funktion. Dann gilt

für zufällig ausgewählte $x_1, ..., x_n \in [a, b]$ und $y_1, ..., y_n \in [c, d]$

$$\int_a^b \int_c^d f(x, y)\, dy\, dx \approx \frac{A}{n} \sum_{i=1}^{n} f(x_i, y_i),$$

wobei $A = (b - a) \cdot (d - c)$ der Flächeninhalt des Integrationsbereichs ist. In Abschnitt 11.5 wird ein MATHEMATICA-Code zur Anwendung dieses Verfahrens vorgestellt. Dieser wurde auch für das nachfolgende Beispiel verwendet.

Beispiel: Als Beispiel soll hier zur Funktion $f(x, y) := x \cdot y$ mit dem Funktionsgraphen



das Integral über $[0, 2] \times [0, 2]$ näherungsweise berechnet werden. Das exakte Ergebnis lautet 4. Für $n = 1\,000\,000$ erhielt man nach fünf Auswertungen folgende Ergebnisse:

- 4.00496
- 4.00394
- 4.00535
- 3.99592
- 3.99586

Man erkennt, dass die Näherungswerte zwar für viele praktische Anwendungsszenarien akzeptabel sein werden, jedoch in Anbetracht der hohen Anzahl an Stützstellen sehr ungenau sind.

11 Codes

In diesem Kapitel werden einige MATHEMATICA-Codes vorgestellt und erklärt, mit welchen die Berechnung der Näherungswerte durchgeführt werden kann.

11.1 Abgeschlossene Newton-Cotes-Formeln

Nachfolgend wird ein Code vorgestellt, mit dem abgeschlossene Newton-Cotes-Formeln beliebiger Ordnung und mit beliebiger Anzahl an Teilintervallen auf eine vorgegebene Funktion $f : [a, b] \to \mathbb{R}$ angewendet werden können. Außerdem können einige Kennzahlen (z. B. die Fehlerkonstante c_m) berechnet werden.

```
1  abgeschlossen[n_] := Table[Simplify[1/n*Integrate[Product[If[k != i, (x - k)
       /(i - k), 1], {k, 0, n}], {x, 0, n}]], {i, 0, n}]
2
3  f[x_] := Sqrt[x]
4  a := 0
5  b := 1
6
7  exakt := Integrate[f[x], {x, a, b}]
8
9  st[k_, n_] := a + k*(b - a)/n
10
11 ANC[n_] := (b - a)*Sum[Part[abgeschlossen[n], l + 1]*f[st[l, n]], {l, 0, n}]
12 ERG[n_] := {N[ANC[n], 20], Abs[N[ANC[n] - exakt, 20]], n}
13 zANC[int_, n_] := (b - a)/int*Sum[Sum[Part[abgeschlossen[n], l + 1]*f[st[l +
       q*n, n*int]], {l, 0, n}], {q, 0, int - 1}]
14 zERG[int_, n_] := {N[zANC[int, n], 20], Abs[N[zANC[int, n] - exakt, 20]], int
       , n, int*n + 1}
15
16 K[n_, m_, t_] := 1/(m - 1)!*( ((1 - t)^m)/m - Sum[Part[abgeschlossen[n], k +
       1]*If[(k/n - t)^(m - 1) > 0, (k/n - t)^(m - 1), 0], {k, 0, n}] )
17
18 c[m_, n_] := (Integrate[x^m, {x, 0, 1}] - Sum[Part[abgeschlossen[n], l + 1]*(
       l/n)^m, {l, 0, n}])/m!
19
20 R[m_, n_] := Abs[c[m, n]*n^(m + 1)]
21 zR[m_, n_] := Abs[c[m, n]*n^m]
```

Durch die in Zeile 1 definierte Funktion werden gemäß Satz 7.10 die Gewichte der abgeschlossenen Newton-Cotes-Formel n-ter Ordnung berechnet. In den Zeilen 3 bis 5 werden die zu integrierende Funktion und die Intervallgrenzen festgelegt. In Zeile 7 wird eine Variable mit dem exakten Wert des Integrals erstellt. Eine Funktion zur Berechnung der Stützstellen wird in Zeile 9 definiert, wobei $0 \leq k \leq n$ sein sollte.

Die in Zeile 11 definierte Funktion `ANC` berechnet den Näherungswert des Integrals $\int_a^b f(x)\,dx$ mittels abgeschlossener Newton-Cotes-Formel n-ter Ordnung (ohne Unterteilung in mehrere

Intervalle). Die Funktion `ERG` in der nächsten Zeile liefert den Näherungswert, die Abweichung vom tatsächlichen Wert und die Anzahl der verwendeten Stützstellen.

Die Funktionen in den Zeilen 13 und 14 führen analoge Berechnungen für die zusammengesetzten abgeschlossenen Newton-Cotes-Formeln durch, wobei `int` die Anzahl an Teilintervallen beschreibt.

In Zeile 16 wird eine Funktion zur Berechnung des in Satz 7.16 vorkommenden und in Definition 7.17 definierten Peano-Kerns K_m zur Verfügung gestellt. Die in Zeile 18 definierte Funktion kann zur Berechnung der Fehlerkonstante c_m (siehe Definition 7.20) verwendet werden, wobei hier die in Satz 7.21 vorgestellte Formel benutzt wird. In den Zeilen 20 und 21 werden schließlich Funktionen für die Berechnung der in Kapitel 7.5 definierten Faktoren γ und κ des Höchstfehlers der einfachen bzw. der zusammengesetzten Formeln definiert.

11.2 Offene Newton-Cotes-Formeln

Der hier behandelte MATHEMATICA-Code kann verwendet werden, um ein Integral mittels offener Newton-Cotes-Formel beliebiger Ordnung und mit beliebiger Anzahl an Teilintervallen zu approximieren.

```
1 offen[m_] := Table[Simplify[1/(m + 2)*Integrate[Product[If[k != i, (x - k)/(i
      - k), 1], {k, 0, m}], {x, -1, m + 1}]], {i, 0, m}]
2
3 f[x_] := Sqrt[x]
4 a := 0
5 b := 1
6
7 exakt := Integrate[f[x], {x, a, b}]
8
9 st[k_, n_] := a + (k + 1)*(b - a)/(n + 2)
10 st2[a1_, b1_, k_, n_] := a1 + (k + 1)*(b1 - a1)/(n + 2)
11
12 ONC[n_] := (b - a)*Sum[Part[offen[n], l + 1]*f[st[l, n]], {l, 0, n}]
13 ERG[n_] := {N[ONC[n], 20], Abs[N[ONC[n] - exakt, 20]], n + 1}
14 zONC[int_, n_] := (b - a)/int*Sum[Sum[Part[offen[n], l + 1]*f[st2[a + q*(b -
      a)/int, a + (q + 1)*(b - a)/int, l, n]], {l, 0, n}], {q, 0, int - 1}]
15 zERG[int_, n_] := {N[zONC[int, n], 20], Abs[N[zONC[int, n] - exakt, 20]], int
      *n + 1}
```

Die in den Zeilen 12 bis 15 definierten Funktionen sind sinngemäß identisch mit jenen der abgeschlossenen Formeln im vorherigen Kapitel. Für die Berechnung der Gewichte in Zeile 1 wurde jedoch hier Satz 7.12 verwendet. Außerdem benötigt man bei den zusammengesetzten Formeln die abgewandelte Funktion `st2` für die Berechnung der Stützstellen, da diese nicht gleichmäßig auf $[a, b]$ verteilt sind, sondern im Bereich der Teilintervallgrenzen ein größerer Abstand vorliegt.

11.3 Romberg-Verfahren

In diesem Abschnitt wird ein Code zur Approximation des Integrals $\int_a^b f(x)\,dx$ mittels Romberg-Verfahren vorgestellt, wobei die Romberg-Folge, die Bulirsch-Folge und die har-

monische Folge zur Verfügung stehen.

```
1 f[x_] := Cos[x]
2 a := -1
3 b := 1
4
5 exakt := Integrate[f[x], {x, a, b}]
6
7 B[0] := 1
8 B[1] := 2
9 B[2] := 3
10 B[n_] := 2*B[n - 2]
11
12 h[n_, 1] := (b - a)/2^n
13 h[n_, 2] := (b - a)/B[n]
14 h[n_, 3] := (b - a)/(n + 1)
15
16 R[k_, j_, v_] := If[j == 0, h[k, v]*(f[a]/2 + Sum[f[a + i*h[k, v]], {i, 1, h
       [0, v]/h[k, v] - 1}] + f[b]/2), ((h[k - j, v])^2*R[k, j - 1] - (h[k, v])
       ^2*R[k - 1, j - 1])/((h[k - j, v])^2 - (h[k, v])^2)]
17
18 ERG[k_, j_, v_] := {N[R[k, j, v], 20], N[Abs[exakt - R[k, j, v]], 20]}
```

In den Zeilen 1 bis 3 werden die zu integrierende Funktion und die Integralgrenzen definiert und in Zeile 5 das exakte Ergebnis des Integrals zwischengespeichert (für die spätere Berechnung des Fehlers).

In den Zeilen 7 bis 10 wird die Bulirsch-Folge rekursiv definiert. In den Zeilen 12 bis 14 werden Funktionen zur Berechnung der Schrittweiten der n-ten Stufe des Romberg-Verfahrens definiert. Dazu wird jeweils im zweiten Argument der Funktion `h` der entsprechende Wert verwendet: 1 für die Romberg-Folge, 2 für die Bulirsch-Folge und 3 für die harmonische Folge.

In Zeile 16 wird schließlich eine Funktion für die Berechnung der Näherung $R_{k,j}$ gemäß Definition 8.1 erstellt, wobei das dritte Argument, also `v`, die verwendete Folge angibt und somit die Werte 1, 2 und 3 annehmen kann. Die in Zeile 18 definierte Funktion gibt den gerundeten Näherungswert sowie die Abweichung vom exakten Wert des Integrals aus.

11.4 Gauß-Verfahren

Dieser Abschnitt beschreibt einen Code zur Durchführung des Gauß-Verfahrens mit einer grundsätzlich beliebig wählbaren Gewichtsfunktion.

```
1 alpha[x_] := 1
2 untergrenze := -1
3 obergrenze := 1
4 grad := 3
5 f[x_] := Cos[x]
6 g[x_] := f[x]/alpha[x]
7
8 a[k_] := Integrate[alpha[x]*x*p[k - 1, x]*p[k - 1, x], {x, untergrenze,
      obergrenze}]/Integrate[alpha[x]*p[k - 1, x]*p[k - 1, x], {x, untergrenze,
      obergrenze}]
```

```
9  b[k_] := Integrate[alpha[x]*x*p[k - 1, x]*p[k - 2, x], {x, untergrenze,
       obergrenze}]/Integrate[alpha[x]*p[k - 2, x]*p[k - 2, x], {x, untergrenze,
       obergrenze}]
10 p[0, x_] := 1
11 p[1, x_] := x - a[1]
12 p[k_, x_] := Expand[x*p[k - 1, x] - a[k]*p[k - 1, x] - b[k]*p[k - 2, x]]
13
14 Nullstellen := N[Solve[p[grad, x] == 0, x], 20]
15 NS[n_] := Nullstellen[[n + 1, 1, 2]]
16
17 W[n_] := Integrate[alpha[x]*Product[If[k != n, (x - NS[k])/(NS[n] - NS[k]),
       1], {k, 0, grad - 1}], {x, untergrenze, obergrenze}]
18
19 approx := N[Sum[W[k]*g[NS[k]], {k, 0, grad - 1}], 20]
20 exakt := N[Integrate[f[x], {x, untergrenze, obergrenze}], 20]
21
22 {approx, Abs[exakt - approx]}
```

In den Zeilen 1 bis 6 werden die Gewichtsfunktion $\alpha : [a, b] \to \mathbb{R}$, der Grad $n + 1$ des später zu konstruierenden Polynoms, das gemäß Definition 9.3 zu allen x^k mit $k \in \{0, ..., n\}$ orthogonal bezüglich α ist, die zu integrierende Funktion f und eine Hilfsfunktion definiert.

In den Zeilen 8 bis 12 werden die für die Rekursion aus Satz 5.13 benötigten Funktionen definiert. In Zeile 14 werden die Nullstellen des ausgewählten Polynoms berechnet und in der darauffolgenden Zeile wird eine Funktion definiert, welche die k-te Nullstelle zur weiteren Verarbeitung auswählt.

In Zeile 17 wird eine Funktion zur Berechnung der Gewichte gemäß Satz 9.8 definiert. In den Zeilen 19 und 20 werden schließlich der Näherungswert und der exakte Wert des Integrals berechnet und in Zeile 22 wird die Approximation und deren Fehler ausgegeben.

11.5 Monte-Carlo-Methode

Mit dem folgenden MATHEMATICA-Code kann die Monte-Carlo-Methode auf eine Funktion $f : [a, b] \times [c, d] \to \mathbb{R}$ angewendet werden.

```
1  a:=0
2  b:=2
3  c:=0
4  d:=2
5  n:=1000
6
7  f[x_,y_]:=x*y
8
9  Plot3D[f[x,y],{x,a,b},{y,c,d},PlotRange -> All]
10
11 N[Integrate[f[x,y],{x,a,b},{y,c,d}]]
12
13 zufallX:=RandomReal[{a,b}]
14 zufallY:=RandomReal[{c,d}]
15
16 MC := N[(b-a)*(d-c)/n*Sum[f[zufallX,zufallY],{k,1,n}],10]
17 List[MC,MC,MC,MC,MC]
```

In Zeile 1 bis 4 werden die Intervallgrenzen festgelegt und Zeile 5 legt die Stichprobengröße fest. In Zeile 7 wird die zu integrierende Funktion definiert. Zeile 8 zeigt den Funktionsgraphen. In Zeile 11 wird der tatsächliche Wert des Integrals als Dezimalzahl ausgegeben. In den Zeilen 13 und 14 werden aus den jeweiligen Intervallen zufällige Werte für die beiden Variablen ausgewählt. Der in Zeile 16 definierte Ausdruck wertet die gegebene Funktion n-mal an zufälligen Stellen aus und bildet den arithmetischen Mittelwert. Die letzte Zeile gibt schließlich eine Liste von fünf Durchführungen der Monte-Carlo-Methode aus.

12 Conclusio

In diesem abschließenden Kapitel der Diplomarbeit werden die wichtigsten Erkenntnisse zusammengefasst, Einsatzszenarien erwähnt und gegenübergestellt sowie mögliche weiterführende Untersuchungen aufgezeigt. Die in dieser Arbeit vorgestellten Verfahren haben jeweils individuelle Vorteile, aus denen sich verschiedene Anwendungsmöglichkeiten ergeben.

Die Newton-Cotes-Formeln sind die vermutlich am weitesten verbreiteten und am besten bekannten Verfahren zur numerischen Integration. Sie sind auch diejenigen, die am ehesten in der Schule gelehrt werden. Der große Vorteil dieser Formeln, insbesondere der abgeschlossenen Formeln, ist die unkomplizierte Anwendung: Die Stützstellen sind einfach zu berechnen und die Gewichte sind für die relevanten Vertreter dieser Klasse von Quadraturformeln schnell zu ermitteln, da sie für mehrere Teilintervalle verwendet werden können und daher eine überschaubare Anzahl aufweisen. Für viele technische und physikalische Anwendungen sind die Newton-Cotes-Formeln auch deshalb gut geeignet, weil oftmals der Abstand zwischen zwei Messungen konstant ist. Mit den Gauß-Formeln, die in vielen Fällen unregelmäßige Stützstellenabstände aufweisen, wäre die Auswertung eines solchen Experiments nicht möglich. Das Anwendungsbeispiel mit der Formel-1-Runde (siehe Seite 62) hat auch gezeigt, dass die zusammengesetzten Newton-Cotes-Formeln niedriger Ordnung bereits ausreichend genau sind, da das Ergebnis ohnehin von anderen Unsicherheitsfaktoren beeinflusst wird. Selbst wenn das Integral genauer berechnet wird, wäre das Ergebnis trotzdem nicht „richtiger“, weil ein gewisser Fehler durch die anderen Einflüsse noch immer vorhanden wäre.

Der entscheidende Vorteil des Romberg-Verfahrens und diverser Variationen ist, dass man die gewünschte Genauigkeit vorgeben kann und so lange die Iteration fortführen kann, bis diese Genauigkeit erreicht ist. Bei anderen Verfahren (insbesondere beim Gauß-Verfahren) müssten jedes Mal neue Stützstellen herangezogen werden und die zugehörigen Funktionswerte ermittelt werden.

Die Besonderheit bei den diversen Variationen des Gauß-Verfahrens ist, dass sie die maximale Fehlerordnung aufweisen. Bei $n+1$ gegebenen Stützstellen werden alle Polynome mit Höchstgrad $2n+1$ exakt integriert. Nach Satz 9.1 gibt es keine Quadraturformel mit $n+1$ Stützstellen, die eine noch höhere Fehlerordnung besitzt. Diesen Vorteil nutzt man nach [26] u. a. bei der Finite-Elemente-Methode. Es handelt sich dabei um ein numerisches Verfahren zur Modellierung von technischen bzw. physikalischen Vorgängen. Ein Beispiel wäre die Berechnung der Deformation eines Autos bei einem Zusammenstoß. Da in diesen Modellen häufig Polynomfunktionen als Ansatz dienen, ist die numerische Berechnung der Integrale aufgrund der hohen Fehlerordnung des Gauß-Verfahrens oftmals sogar exakt.

Für Mehrfachintegrale wird die Monte-Carlo-Methode aufgrund ihrer einfachen Anwendbarkeit relativ häufig benutzt. Es ist jedoch bereits ein auf wenige Nachkommastellen exakter Näherungswert mit einem erheblichen Rechenaufwand verbunden.

Bezüglich der Konvergenz lässt sich sagen, dass bei den Newton-Cotes-Formeln steigender Ordnung der Fehler bei „gewöhnlichen“ Funktionen zwar tendenziell kleiner wird, aber nicht konvergiert. Beispiele haben ausführlich gezeigt, welche Probleme durch die negativen Gewichte und die unbeschränkte Summe der Beträge der Gewichte entstehen können. Beim Romberg-Verfahren wurde hingegen gezeigt, dass der Näherungswert des Integrals gegen den tatsächlichen Wert konvergiert. Auch die Gauß-Verfahren konvergieren für orthogonale Polynome, deren Grad gegen unendlich geht.

In dieser Arbeit wurde meist die Anzahl an Funktionsauswertungen als Maß für den Rechenaufwand herangezogen, da davon ausgegangen wurde, dass bei bereits bekannten Gewichten und Stützstellen die Ermittlung der Stützwerte am meisten Rechenzeit beansprucht. Hier wurde jedoch außer Acht gelassen, dass die Ermittlung der Gewichte und Stützstellen je nach Verfahren ebenfalls unterschiedlich anspruchsvoll ist: Bei den Newton-Cotes-Formeln muss für die Ermittlung der Stützstellen lediglich das Intervall in gleiche Abschnitte geteilt werden, während bei den Gauß-Formeln oftmals keine expliziten Formeln für die Berechnung der Nullstellen der orthogonalen Polynome vorliegen. Eine interessante Fragestellung für weiterführende Untersuchungen wäre es, ein System zu entwickeln, welches die diversen Verfahren der numerischen Integration bezüglich ihrer Effizienz vergleicht.

Literaturverzeichnis

[1] **Brokate, Martin u. a.** (2015): Grundwissen Mathematikstudium: Höhere Analysis, Numerik und Stochastik. Springer-Verlag.

[2] **Chu, Moody T.** (o. J.): Newton's Interpolation Formula.
http://www4.ncsu.edu/~mtchu/Teaching/Lectures/MA427/By_subjects/10_newton_formula.pdf [11. September 2017]

[3] **Cigler, Johann**: Konkrete Analysis. Sommersemester 2004. Universität Wien: Fakultät für Mathematik.
https://homepage.univie.ac.at/johann.cigler/preprints/KonkreteAnalysis.pdf [16. September 2017]

[4] **Erdös, Laszlo** (2005): Kapitel 5: Numerische Mathematik.
http://www.mathematik.uni-muenchen.de/~lerdos/SS05/Num/5.pdf [6. September 2017]

[5] **Ferus, Dirk** (2007): Integraltransformationen und partielle Differentialgleichungen für Ingenieure.
http://page.math.tu-berlin.de/~ferus/ING/ITPDG.pdf [15. September 2017]

[6] **Gerdts, Matthias** (2011): Numerische Mathematik I.
https://www.unibw.de/lrt1/gerdts/lehre/numerik_1.pdf [7. Juli 2017]

[7] **Haas, Markus** (2011): Taylor-Formel.
https://www.empiwifo.uni-freiburg.de/lehre-teaching-1/winter-term-11-12/materialien-wirtschaftsmathematik/taylor [10. September 2017]

[8] **Hämmerlin, Günther; Hoffmann, Karl-Heinz** (1990): Numerische Mathematik (4. Auflage). Springer-Verlag.

[9] **Hermann, Martin** (2004): Numerik gewöhnlicher Differentialgleichungen: Anfangs- und Randwertprobleme. De Gruyter Oldenbourg.

[10] **Isenhardt, Nikola** (2008): Numerische Integration.
https://www.mathi.uni-heidelberg.de/~thaeter/anasem08/Isenhardt.pdf [16. September 2017]

[11] **Iske, Armin** (2007): Analysis II – Kapitel 8: Interpolation.
https://www.math.uni-hamburg.de/teaching/export/tuhh/cm/a2/07/vorl03_ana.pdf [3. September 2017]

[12] **Iske, Armin** (2007): Analysis II – Kapitel 12: Numerische Quadratur.
https://www.math.uni-hamburg.de/teaching/export/tuhh/cm/a2/07/vorl13_ana.pdf [20. September 2017]

[13] **Kaplan, Michael; Wehrheim, Jan; Mendl, Christian** (2011): Mathematik für Physiker 1 – Übungsblatt 10.
https://www.ma.tum.de/foswiki/pub/HM/MA9201_2011W/Uebung/Zentral_Loes_10.pdf [3. September 2017]

[14] **Kiekhäfer, Mathias** (2010): Tschebyscheff-Polynome, allgemeine orthogonale Polynome sowie deren Anwendung in der Gauß-Quadratur.
https://www.math.uni-bielefeld.de/~emmrich/studenten/mathias.pdf [15. September 2017]

[15] **Klug, R.** Hrsg. (1908): Neue Stereometrie der Fässer, besonders der in der Form am meisten geeigneten österreichischen, und Gebrauch der kubischen Visierrute. Leipzig: Verlag von Wilhelm Engelmann.

[16] **Köster, Jan** (2007): Der Approximationssatz von Weierstraß.
http://www.math.kit.edu/ianmip/lehre/anganalysis2007w/media/weierstrass.pdf [3. September 2017]

[17] **Lube, Gert** (2004): Interpolationsfehlerabschätzungen.
https://lp.uni-goettingen.de/get/text/1207 [3. September 2017]

[18] **Lube, Gert** (2004): Fehlerabschätzungen.
https://lp.uni-goettingen.de/get/text/1271 [6. September 2017]

[19] **Motschmann, U. u. a.**: Rechenmethoden der Physik. Sommersemester 2015. Technisches Universität Braunschweig: Institut für theoretische Physik.
https://www.tu-braunschweig.de/Medien-DB/theophys/rechenmethodenskript.pdf [16. September 2017]

[20] **Raith, Peter**: Einführung in die Analysis. Sommersemester 2014. Wien: Fakultät für Mathematik, Oskar-Morgenstern-Platz 1.

[21] **Reis, Timo; Rothe, Kai**: Analysis II für Studierende der Ingenieurwissenschaften. Sommersemester 2015. Universität Hamburg: Fachbereich Mathematik.
https://www.math.uni-hamburg.de/teaching/export/tuhh/cm/a2/15/vorlesungen/vl6_dv.pdf [14. September 2017]

[22] **Thiel, Pablo** (o. J.): Numerische Integration durch Extrapolation.
http://www.math.uni-hamburg.de/home/struckmeier/Proseminar/extrapolation.pdf [6. September 2017]

[23] **Unbekannt**: Numerische Quadraturverfahren, Integrationsverfahren.
http://mathematikalpha.de/quadraturverfahren-integrationsverfahren [3. September 2017]

[24] **Unbekannt**: Die Keplersche Fassregel.
http://www.kepler-gesellschaft.de/index.php/kepler-preis/180-kepler-preis/preisgekroente-arbeiten/kepler-preise-2006/johannes-kepler-faecheruebergreifend/309-mathematik.html [6. September 2017]

[25] **Unbekannt**: Formel von Faà di Bruno.
http://matheplanet.com/matheplanet/nuke/html/viewtopic.php?topic=10277&post_id=59270 [18. September 2017]

[26] **Unbekannt**: Gaußsche Quadraturformeln.
http://www.tm-mathe.de/Themen/html/numintgauss.html [25. September 2017]

[27] **Voss, Heinrich** (2004): Grundlagen der Numerischen Mathematik.
https://www.mat.tuhh.de/lehre/material/grnummath.pdf [10. September 2017]

[28] **Weiss, Daniel**: Numerik für Informatiker und Bioinformatiker. Sommersemester 2012. Tübingen: Eberhard Karls Universität.
https://na.uni-tuebingen.de/ex/numinf_ss12/Vorlesung12_SS2012.pdf [12. September 2017]

[29] **Wengenroth, Jochen**: Elemente der Analysis. Sommersemester 2009. Universität

Trier: Fachbereich IV - Mathematik.
https://www.math.uni-trier.de/~wengenroth/ElaScript8.pdf [24. September 2017]

[30] **Zern, Artjom** (2009): Bernoullipolynome und Bernoullizahlen.
https://www.mathi.uni-heidelberg.de/~theiders/PS-Analysis/Bernoullische_Polynome.pdf
[3. September 2017]

Abbildungen: Mit Ausnahme des Logos der Universität Wien am Deckblatt dieser Diplomarbeit wurden alle Abbildungen selbst erstellt.

Abstract (deutsch)

Es gibt zwei typische Situationen, in denen numerische Integrationsverfahren eingesetzt werden. Einerseits werden sie verwendet, um den Wert des Integrals zu approximieren, falls die Stammfunktion nicht durch elementare Funktionen dargestellt werden kann. Andererseits können mit numerischen Verfahren auch Integrale berechnet werden, wenn die Werte der zu integrierenden Funktion nur an einigen Stellen bekannt sind. Das letztgenannte Einsatzszenario ist speziell im Zusammenhang mit diversen Experimenten relevant, da hier oftmals nur zu bestimmten Zeitpunkten ein Messwert vorliegt.

In dieser Arbeit werden drei Klassen von numerischen Integrationsverfahren ausführlich behandelt, nämlich die Newton-Cotes-Formeln, das Romberg-Verfahren und das Gauß-Verfahren. Dabei stellt sich heraus, dass jedes dieser Verfahren unterschiedliche Vor- und Nachteile hat und die Auswahl eines geeigneten Verfahrens somit abhängig von den konkreten Anforderungen ist.

Die bekanntesten numerischen Integrationsverfahren sind die Newton-Cotes-Formeln. Dazu gehören u. a. die Mittelpunktregel, die Sehnentrapezregel und die Simpsonregel (auch als Keplersche Fassregel bekannt). Diese Verfahren sind auch diejenigen, die am ehesten im Schulunterricht behandelt werden. Das gesamte Integrationsintervall wird bei diesen Verfahren jeweils in eine bestimmte Anzahl äquidistanter Teilintervalle zerlegt. Bei den abgeschlossenen Newton-Cotes-Formeln befinden sich die verwendeten Stützstellen, d. h. jene Stellen, an denen der Funktionswert der zu integrierenden Funktion bekannt sein muss, genau an den Teilintervallgrenzen. Daher eignet sich dieses numerische Integrationsverfahren auch besonders gut zur Auswertung von Experimenten.

Die Funktion wird nun durch ein Interpolationspolynom ersetzt, welches an den Stützstellen mit der tatsächlichen Funktion übereinstimmt. Dieses Polynom kann exakt integriert werden. In der Praxis müssen Interpolation und anschließende Integration jedoch nicht durchgeführt werden, sondern man erhält für jede Stützstelle ein sogenanntes Gewicht, mit welchem der entsprechende Stützwert multipliziert wird. Die Summe all dieser Produkte ergibt schließlich den Näherungswert des Integrals. Bei den abgeschlossenen Newton-Cotes-Formeln ergibt es sich, dass für eine ungerade Anzahl an Stützstellen (also eine gerade Anzahl an Teilintervallen) alle Polynome exakt integriert werden, deren Grad höchstens der Anzahl an Stützstellen entspricht. Beispielsweise werden durch die Simpsonregel, welche drei Stützstellen verwendet, alle Polynome mit Höchstgrad 3 exakt integriert. Die Verwendung von Newton-Cotes-Formeln mit einer großen Anzahl an Stützstellen ist jedoch nicht zu empfehlen, da hier aufgrund von negativen Gewichten absurde Näherungswerte entstehen können. So ist es etwa möglich, dass das Integral einer positiven Funktion einen negativen Näherungswert hat. Stattdessen ist es besser, das Intervall in mehrere Teile zu unterteilen, und auf diese Teile jeweils eine Newton-Cotes-Formel niederer Ordnung anzuwenden.

Neben den abgeschlossenen Newton-Cotes-Formeln existieren auch die sogenannten offenen Newton-Cotes-Formeln, welche sich dadurch auszeichnen, dass sich die Stützstellen im Inneren der Teilintervalle befinden. Dadurch ist es beispielsweise möglich, Funktionen zu integrieren, die an einer Teilintervallgrenze nicht definiert sind. Die offenen Newton-Cotes-

Formeln spielen jedoch mit Ausnahme der Mittelpunktregel in der Praxis keine Rolle, da es effizientere Verfahren gibt, deren Stützstellen ebenfalls nicht mit den Intervallgrenzen zusammenfallen, nämlich die Vertreter des sogenannten Gauß-Verfahrens.

Beim Gauß-Verfahren werden als Stützstellen die Nullstellen bestimmter Polynome verwendet. Diese Stützstellen sind im Allgemeinen nicht gleichmäßig über das Integrationsintervall verteilt, weshalb dieses Verfahren für die Auswertung von Experimenten eher ungeeignet ist. Es zeichnet sich jedoch durch eine sehr hohe Fehlerordnung aus. Konkret bedeutet dies, dass bei der Verwendung von n Stützstellen alle Polynome mit Höchstgrad $2n - 1$ exakt integriert werden. Es lässt sich außerdem zeigen, dass es sich hierbei um die maximale Fehlerordnung eines numerischen Integrationsverfahrens handelt, d. h. es gibt kein Verfahren, welches bei der Verwendung von n Stützstellen alle Polynome mit Grad $2n$ exakt integriert. Eingesetzt wird dieses Verfahren u. a. bei der Berechnung uneigentlicher Integrale und aufgrund der hohen Fehlerordnung bei anderen numerischen Verfahren, wie beispielsweise der Finite-Elemente-Methode, um die dort vorkommenden Interpolationspolynome exakt zu integrieren.

Das dritte in dieser Arbeit behandelte Verfahren ist das sogenannte Romberg-Verfahren. Hier wird dasselbe Integral mittels Sehnentrapezregel mehrmals mit jeweils unterschiedlichen Teilintervallbreiten approximiert. Man erhält dadurch eine Abhängigkeit des Näherungswertes vom Quadrat der Teilintervallbreite. Durch den Algorithmus von Aitken und Neville wird eine Extrapolation des Näherungswertes für eine verschwindende Teilintervallbreite durchgeführt. Es ergibt sich auf diese Weise, dass durch Kombination mehrerer Näherungswerte eine weitaus bessere Näherung des Integrals gefunden werden kann. Ein großer Vorteil des Romberg-Verfahrens ist es, dass man solange weitere Näherungswerte in die Berechnung miteinbeziehen kann, bis die gewünschte Genauigkeit erreicht ist, ohne dabei die bisherigen Berechnungen verwerfen zu müssen.

Neben diesen drei Klassen von numerischen Integrationsverfahren wird außerdem oberflächlich auf die Schrittweitensteuerung, d. h. die Optimierung der Stützstellenverteilung, und auf die Monte-Carlo-Methode, welche u. a. zur numerischen Näherung von Mehrfachintegralen verwendet wird, eingegangen.

Abstract (english)

There are two typical situations where numerical integration methods are used. On the one hand they are applied to approximate an integral's value for functions whose antiderivatives can't be expressed in terms of elementary functions. On the other hand they can also be applied if the function value is known only for a few function arguments. The last case is especially relevant in connection with the evaluation of experiments because there are only a few measured quantities available.

This thesis mainly deals with three categories of numerical integration methods, the Newton-Cotes formulas, the Romberg method and the Gaussian quadrature. Each of these methods has different benefits and disadvantages. The choice of the right method depends on the specific requirements.

The most popular numerical integration methods are the Newton-Cotes formulas. These include the rectangle rule, the trapezoidal rule and Simpson's rule (also known as Kepler's rule), which are often taught at school. The whole integration interval is divided into a certain number of equally spaced subintervals. The data points of the closed Newton-Cotes fomulas coincide with the subinterval limits. Therefore this numerical integration method is well suited for the evaluation of experiments.

The integrand is replaced by an interpolation polynomial which corresponds with the initial function at the data points. This polynomial can be integrated exactly. In practice, the interpolation and the following integration are not necessary. For each data point there is a certain weight which is multiplied by the function value of this data point. The sum of all these products corresponds to the approximation of the integral. It turns out that the closed Newton-Cotes formulas produce, for an odd number of data points (which is equivalent to an even number of subintervals), exact integral values for all polynomials whose degree is less or equal to the number of data points. For example, Simpson's rule, which uses three data points, produces exact integral values for all polynomials with a degree less or equal to 3. Due to negative weights the usage of Newton-Cotes formulas with a high number of data points is not recommended because it is possible that the formula leads to an absurd approximation. For example, it can occur that the integral of a positive function has a negative approximation. Instead, it is better to divide the integration interval into a certain number of subintervals. A Newton-Cotes formula with lower degree is now applied to each of these subintervals.

In addition to the closed Newton-Cotes formulas, there are also open Newton-Cotes formulas. The characteristic of these formulas is that the data points are inside of the subintervals. Therefore, it is possible to receive an approximation for functions that are not defined at the subinterval limits. Apart from the rectangle rule the open Newton-Cotes formulas are unimportant in practice because there are more efficient methods that also use data points which do not coincide with the subinterval limits, the so-called Gaussian quadrature rules.

The Gaussian quadrature uses the zeros of certain polynomials as data points. These zeros are generally not equally spaced. Therefore, this method is rather inappropriate for the evaluation of experiments. The special feature of this method is that all polynomials with a degree less or equal to $2n - 1$ are integrated exactly by using only n data points. It can be shown that there exists no numerical integration method which can integrate all polynomials with degree $2n$ by using only n data points. The Gaussian quadrature is, for example, used to approximate improper integrals. Due to its high order of accuracy it is also applied for other numerical methods, such as the finite element method, to integrate the occuring interpolation polynomials exactly.

The third described method is the Romberg method. In this method the integral is approximated by using the trapezoidal rule multiple times with different stepsizes. This results in a dependence between the approximation and the square of the stepsize. Neville's algorithm is now applied to extrapolate the approximation for a vanishing stepsize. It turns out that the combination of several approximations results in a much better approximation. An advantage of the Romberg method is that approximations can be added until the desired accuracy is achieved without rejecting the previous calculations.

Besides these three categories of numerical integration methods, there is also a short introduction to methods with adaptive stepsizes and the so-called Monte Carlo integration, which could be used for approximating multiple integrals.