MASTERARBEIT | MASTER'S THESIS

Titel | Title

Numerische Methoden zur Lösung von Extremwertaufgaben

verfasst von | submitted by

Julian Muttenthaler BEd

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

Master of Education (MEd)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 199 510 520 02
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Lehramt Sek (AB) Unterrichtsfach Geographie und wirtschaftliche Bildung Unterrichtsfach Mathematik
Betreut von | Supervisor:	Assoz. Prof. Dr. Martin Ehler Privatdoz.

Zusammenfassung

Die vorliegende Masterarbeit untersucht numerische Methoden zur Lösung von Extremwertaufgaben und vergleicht deren Lösungsmöglichkeiten mit den klassischen analytischen Verfahren im Kontext des Mathematikunterrichts. Nach der Darstellung grundlegender Konzepte der Differentialrechnung und einer Variante zur systematischen Bearbeitung von Optimierungsaufgaben werden zentrale numerische Verfahren zur Differenzierung und Nullstellenfindung eingeführt und mithilfe der Programmiersprache *Julia* umgesetzt. Anhand ausgewählter Beispiele wird deren Effizienz analysiert und miteinander verglichen. Abschließend diskutiert die Arbeit die didaktischen Möglichkeiten numerischer Methoden und deren Beitrag zum Verständnis von Optimierungsaufgaben sowie zum sinnvollen Technologieeinsatz im Unterricht.

Abstract

This master's thesis examines numerical methods for solving optimization problems and compares their capabilities with those of classical analytical methods in the context of mathematics education. Following a presentation of fundamental concepts in differential calculus and a systematic approach to solving optimization problems, key numerical methods for differentiation and finding zeros are introduced and implemented using the *Julia* programming language. Their efficiency is analyzed and compared using selected examples. Finally, the thesis discusses the didactic potential of numerical methods and their contribution to understanding optimization problems as well as to the meaningful use of technology in the classroom.

Danksagung

An dieser Stelle möchte ich mich herzlich bei Herrn Assoz. Prof. Dr. Martin Ehler für die Bereitstellung des Themas und die engagierte Betreuung im Rahmen meiner Masterarbeit bedanken. Bei Unklarheiten stand er mir jederzeit mit Rat und Tat zur Seite und schaffte es, mich in mühsamen Phasen zur Fertigstellung der Arbeit zu motivieren.

Mein besonderer Dank gilt meiner Familie, insbesondere meinen Eltern, die mich nicht nur finanziell, sondern auch emotional während des gesamten Studiums unterstützt haben. Hervorheben möchte ich meine Mutter Bernadette, die immer ein offenes Ohr für mich hatte und Verständnis zeigte, dass es neben dem Unterrichten an einer Schule nicht immer einfach war, mit den finalen Prüfungen und der Masterarbeit voranzukommen.

Ebenso danke ich meiner Freundin Annalena für ihre liebevolle Unterstützung, ihr Verständnis und ihre Geduld während der letzten Jahre. Danke auch für das sorgfältige Korrekturlesen und die Unterstützung in den finalen Wochen! Du hast es immer geschafft, mich aufzuheitern, auch wenn der Gedanke an die Masterarbeit mich verzweifeln ließ.

Abschließend danke ich all meinen Freund:innen und Studienkolleg:innen, die mich durch mein Studium begleitet und diese Zeit besonders und unvergesslich gemacht haben. Ihr wart mir stets eine große Stütze, auch wenn nicht selten Wochenenden, die für die Masterarbeit vorgesehen waren, schlussendlich euch zum Opfer fielen.

Inhaltsverzeichnis

1 Einleitung

Der Schwerpunkt der vorliegenden Masterarbeit liegt auf der Entwicklung und Anwendung von numerischen Methoden zur Lösung von Extremwertaufgaben. Ziel ist es, deren Potenzial im Vergleich zu klassischen analytischen Methoden zu untersuchen und im Kontext des Mathematikunterrichts zu bewerten. Eine vertiefte Auseinandersetzung mit dieser Thematik fördert ein umfassenderes Verständnis von Optimierungsaufgaben im Allgemeinen und unterstützt Lernende dabei, die komplexen Vorgänge des Technologieeinsatzes im Mathematikunterricht nachzuvollziehen. Das Lösen von Extremwertaufgaben stellt die Schüler:innen meist vor zwei zentrale mathematische Herausforderungen: das Bestimmen der Ableitungsfunktion sowie das anschließende Berechnen der Nullstelle zur Ermittlung von Extremstellen. Diese beiden Kernbereiche können oftmals mit rein analytischen Methoden aufgrund der Komplexität der Funktionen nicht von Schüler:innen gelöst werden, weshalb sie einen zentralen Fokus der Arbeit bilden.

Da Extremwertaufgaben eine wesentliche Rolle im Mathematikunterricht an allgemeinbildenden höheren Schulen (AHS) spielen, werden zu Beginn der Arbeit die grundlegenden analytischen Kenntnisse und Definitionen, vor allem in Bezug auf die Differentialrechnung und Nullstellenfindung, wiederholt und erläutert. Weiters wird eine strukturierte Schritt-für-Schritt-Anleitung für das Lösen von Extremwertaufgaben vorgestellt und erläutert, weshalb Optimierungsaufgaben einen wichtigen Bestandteil des Mathematikunterrichts darstellen.

Der anschließende Hauptteil der Arbeit ist dem Teilgebiet der numerischen Mathematik zuzuordnen. Nachdem zu Beginn die Grundlagen der numerischen Mathematik anhand der Fachliteratur diskutiert werden, widmen sich die folgenden Kapitel einerseits der numerischen Differenzierung mittels dualer Zahlen und andererseits verschiedenen numerischen Methoden zur näherungsweisen Bestimmung von Nullstellen. Dabei werden zuerst die jeweiligen theoretischen Grundlagen besprochen und anschließend anhand ausgewählter Anwendungsbeispiele veranschaulicht. Im Zentrum dieser Beispiele steht die Programmiersprache *Julia*, mit deren Hilfe die Verfahren implementiert und die Berechnungen durchgeführt werden. Die entwickelten Programme sowie die resultierenden Ergebnisse werden in Form von Abbildungen dargestellt und im Kontext der jeweiligen Beispiele analysiert und diskutiert.

Im weiteren Verlauf der Arbeit werden die zuvor eingeführten numerischen Methoden

auf ausgewählte Extremwertaufgaben angewendet. Ziel dabei ist es, die Effizienz der verschiedenen Methoden zu vergleichen und zu untersuchen, inwiefern eine rein numerische Lösung von Extremwertaufgaben realisierbar ist und welche Verfahren sich am besten eignen. Auch hier kommt überwiegend die Programmiersprache *Julia* zum Einsatz, wobei die zuvor entwickelten Codes integriert und weiterverwendet werden. Das abschließende Kapitel geht der Frage nach, inwiefern numerische Methoden im Mathematikunterricht sinnvoll eingesetzt werden können und welche Vorteile sie für die Lern- und Denkprozesse der Schüler:innen bieten. Damit schließt sich der Kreis zum anfangs dargestellten schulischen Kontext.

Insgesamt liefert die Arbeit einen fundierten Vergleich analytischer und numerischer Lösungsmethoden für Extremwertaufgaben und setzt Impulse für einen verstärkten Einsatz numerischer Methoden im Schulalltag.

2 Optimieren in der Schule

Extremwertaufgaben zählen zu den sogenannten Optimierungsaufgaben, bei denen minimale bzw. maximale Werte eines Problems gefunden werden sollen. Dies spielt sowohl im Alltag als auch in der Schule eine tragende Rolle. Betrachtet man im Alltag zum Beispiel den Weg zum Kühlschrank, wird intuitiv der schnellstmögliche Weg gesucht, und somit versucht, minimale Zeit aufzubringen. Die bekanntesten Anwendungen der Optimierung in der Schule finden sich bei den Extremwertaufgaben in der Sekundarstufe II. Das folgende Kapitel soll zunächst erläutern, welche Rolle diese im Unterricht einnehmen und wie sie in der Regel gelöst werden. Die theoretischen Erklärungen werden dabei am Ende des Kapitels mit praktischen Anwendungsbeispielen unterstützt.

2.1 Bedeutung von Extremwertaufgaben in der Schule

Optimierungsaufgaben stellen fundamentale Ideen der Mathematik dar und sind auch im Alltag weit verbreitet, egal ob in Bereichen der Wirtschaft, Technik oder Gesellschaft. Um diesen Alltagsbezug auch in der Schule sichtbar zu machen, bieten sich diverse Lösungsstrategien an, die im Mathematikunterricht thematisiert werden können. Die bekannteste Methode ist das Optimieren mittels Differentialrechnung, also mittels analytischer Verfahren, auf die im weiteren Verlauf der Fokus gelegt wird. Davor ist es aber noch wichtig klarzustellen, dass dies bei Weitem nicht das einzige Verfahren ist, um optimale Werte zu finden. Das vorliegende Unterkapitel basiert dabei größtenteils auf [11].

Da Kinder und Jugendliche auch schon im frühen Alter dazu drängen, Rekorde zu brechen oder herausfinden möchten, wer oder was „am schnellsten“, „am größten“ oder einfach nur „am besten“ ist, bietet es sich an, schon viel früher als in der Sekundarstufe II mit Optimierungsaufgaben zu beginnen. So besteht die Möglichkeit, unter anderem in der Sekundarstufe I mithilfe von Probieren und Experimentieren erste optimale Lösungen zu finden. Im weiteren Verlauf der Schullaufbahn können Extremwertaufgaben anschließend mittels Dreiecksungleichung oder quadratischem Ergänzen gelöst werden. Je früher Kinder und Jugendliche mit Optimierungsaufgaben in Berührung kommen, desto leichter fällt auch das Verständnis und die Bearbeitung von anspruchsvolleren Aufgaben in der Sekundarstufe II.

Stellen wir uns nun die Frage, warum Optimierungsaufgaben neben dem Alltagsbezug überhaupt in der Schule eingesetzt werden sollten, stoßen wir auf viele weitere Grün-

de. Für die Bearbeitung von Optimierungsaufgaben im Unterricht werden zum Beispiel keine eigenen Kapitel benötigt, da sie immer wieder bei passenden Themen aufgegriffen werden können. Darüber hinaus zeichnen sich optimale Lösungen oftmals durch Einfachheit, Lage oder auch durch Symmetrie aus, was als Motivation für den Unterricht dienen kann. Genauso lernen die Schüler:innen Strategien, die nicht nur bei der Optimierung von Vorteil sein können, sondern auch in anderen Bereichen einsetzbar sind. Ein weiterer Punkt, der für den Einsatz von Optimierungsaufgaben im Unterricht spricht, ist die Tatsache, dass viele Begriffe im Unterricht ohnehin von extremer Natur sind. Zum Beispiel werden in der Sekundarstufe I bereits optimale Werte wie das kleinste gemeinsame Vielfache oder der Abstandsbegriff (minimale Entfernung) eingeführt. Ein letztes Argument für das Bezugnehmen auf Optimierung ist, wie zu Beginn des Kapitels schon erwähnt, das Streben, etwas bestmöglich zu machen. Dies kann sicherlich als gute Motivation für die Schüler:innen dienen, vor allem, wenn dabei noch ein Bezug zum Alltag herrscht.

Bevor nun die detaillierte Erklärung und praktische Aufbereitung der analytischen Methoden folgt, ist es noch von Bedeutung, die Thematik des Optimierens im österreichischen Lehrplan für Mathematik zu verankern. Beim Analysieren des Lehrplans für Mathematik in der Sekundarstufe I und Sekundarstufe II fällt sofort auf, dass der Optimierung nur wenig Aufmerksamkeit geschenkt wird. Einzig allein in der 7. Klasse (1. und 2. Semester) wird von Optimierung in Form von Extremwertaufgaben gesprochen. Im 1. Semester ist dies unter dem Thema „Grundlagen der Differentialrechnung anhand von Polynomfunktionen“, genauer gesagt als Unterpunkt „*Untersuchungen von Polynomfunktionen in inner- und außermathematischen Bereichen durchführen können; einfache Extremwertaufgaben lösen können (Ermittlung von Extremstellen in einem Intervall)*“ verankert. Im 2. Semester ist die Vertiefung im Kapitel „Erweiterungen und Exaktifizierungen der Differentialrechnung“ unter „*Weitere Anwendungen der Differentialrechnung, insbesondere aus Wirtschaft und Naturwissenschaft, durchführen können*“ zu finden. [vgl. [8]]

2.2 Vorgehensweise bei analytischen Extremwertaufgaben in der Schule

Werden unterschiedliche Schulbücher der Sekundarstufe II analysiert, fällt rasch auf, dass so gut wie alle Lehrbücher denselben Weg erklären, um Extremwertaufgaben zu lösen. Diese Vorgehensweise basiert unter anderem auf [6] und lässt sich in folgenden Schritten zusammenfassen:

Schritt 1: Im ersten Schritt wird die Aufgabe analysiert; anschließend werden die vorkommenden Größen bestimmt. Dabei wird die Frage beantwortet, welche Größe minimal bzw. maximal sein soll.

Schritt 2: In diesem Schritt wird die Zielfunktion mit mehreren Variablen aufgestellt. Es wird somit eine Hauptbedingung (HB) ermittelt.

Schritt 3: Ziel des dritten Schrittes ist es, eine Nebenbedingung (NB) anhand der Aufgabenstellung aufzustellen und alle Variablen durch eine einzige auszudrücken.

Schritt 4: Anschließend wird die Nebenbedingung in den Term der Hauptbedingung eingesetzt, wodurch eine Funktion in einer Variablen entsteht. Dabei ist auch der Definitionsbereich der gefundenen Funktion anzugeben.

Schritt 5: Im fünften Schritt wird die Zielfunktion in einer Variablen abgeleitet; anschließend werden mittels Berechnung der Nullstellen der Ableitung die möglichen Extremstellen bestimmt.

Schritt 6: Im letzten Schritt werden die möglichen Extremstellen mit den Werten an den Randstellen verglichen und somit die Lösungen ermittelt.

2.3 Analytische Methoden zum Finden der Ableitungen

Dieses Unterkapitel basiert größtenteils auf den Kapiteln 12 und 13 in [10] und Kapitel 5 in [13].

Grundlage für das Lösen von Extremwertaufgaben in der Sekundarstufe II ist, wie zuvor bereits erwähnt, das Berechnen der Ableitung und anschließend das Finden der Extremstelle. Voraussetzung für das Berechnen der Ableitung sind fundamentale Definitionen und Regeln, die die Schüler:innen beherrschen müssen. In diesem Unterkapitel werden die Vorgehensweisen dafür von Grund auf erläutert.

2.3.1 Differenzenquotient vs. Differentialquotient

Im weiteren Verlauf sei f eine Funktion mit Definitionsmenge $D \subseteq \mathbb{R}$ und $x_0 \in D$, wobei in jeder noch so kleinen Umgebung von x_0 ein weiteres Element liegt, das sich von x_0 unterscheidet.

Definition (z.B. Definition in [10])**. (Differenzenquotient)**
Unter dem Differenzenquotienten von f an der Stelle x_0 bezeichnet man die Funktion

$$\Delta x = \frac{f(x) - f(x_0)}{x - x_0} \quad \forall x \in D \setminus \{x_0\} \tag{2.1}$$

Geometrisch betrachtet versteht man unter dem Differenzenquotienten die durchschnittliche Steigung k zwischen zwei auf der Funktion liegenden Punkten, wie auf Abbildung 1 zu sehen ist.

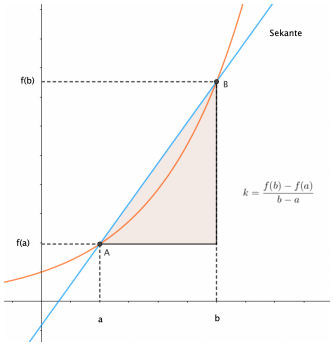


Abbildung 1: Differenzenquotient

Aufbauend auf dem Differenzenquotienten gilt es in der Schule, die Brücke zum Differentialquotienten zu schlagen. Dabei sollen sich die Schüler:innen vorstellen, was passiert, wenn das von x_0 verschiedene Element x immer näher an x_0 rückt. Mittels des Grenzwerts $x \to x_0$ wird schnell der Differentialquotient erreicht.

Definition (z.B. Definition in [10])**. (Differenzierbarkeit)**
Eine Funktion f ist differenzierbar in x_0, wenn der Differentialquotient

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} \tag{2.2}$$

existiert. Der Differentialquotient wird auch als Ableitung f' der Funktion f an der Stelle x_0 bezeichnet.

Geometrisch betrachtet lässt sich der Differentialquotient den Schüler:innen im Vergleich zur durchschnittlichen Änderung bzw. Steigung des Differenzenquotienten als momentane Änderung bzw. Steigung der Funktion f an der Stelle x_0 erklären. Anstatt einer Sekante zwischen zwei Punkten wird eine Tangente an den Punkt $(x_0|f(x_0))$, wie in Abbildung 2 zu sehen ist, gelegt.

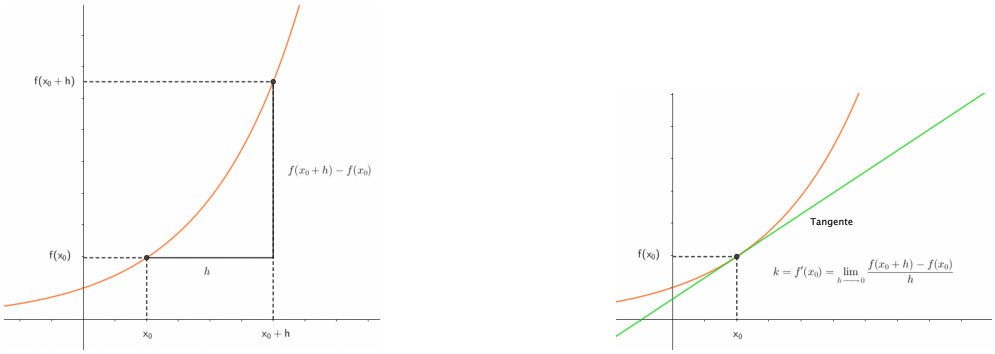


Abbildung 2: Differentialquotient

Bevor die Ableitungsregeln eingeführt werden, können die Schüler:innen anhand folgender Beispiele ihr Verständnis für den Grenzwertbegriff steigern.

Beispiel. Berechne mithilfe des Grenzwerts des Differenzenquotienten die Ableitung folgender Funktionen!

(i) $f(x) = x^2$

(ii) $g(x) = \frac{1}{1+x}, \quad x \neq -1$

Lösung:
Für die Lösung dieser Beispiele verwenden wir die Definition der Differenzierbarkeit und setzen die gegebenen Funktion ein. Nach schrittweiser Umformung und Anwendung der Grenzwertregeln ergeben sich die Ableitungsfunktionen.

(i)

$$\begin{aligned} f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} &= \lim_{h\to 0} \frac{(x+h)^2 - x^2}{h} \\ &= \lim_{h\to 0} \frac{x^2 + 2xh + h^2 - x^2}{h} \\ &= \lim_{h\to 0} 2x + h = 2x \end{aligned}$$

(ii)

$$\begin{aligned} g'(x) = \lim_{h\to 0} \frac{g(x+h)-g(x)}{h} &= \lim_{h\to 0} \frac{\frac{1}{1+x+h} - \frac{1}{1+x}}{h} \\ &= \lim_{h\to 0} \frac{\frac{1+x-1-x-h}{((1+x)+h)\cdot(1+x)}}{h} \\ &= \lim_{h\to 0} \frac{-1}{((1+x)+h)\cdot(1+x)} \\ &= \frac{-1}{(1+x)^2} \end{aligned}$$

2.3.2 Ableitungsregeln

Die Schüler:innen werden schnell merken, dass das Berechnen der Ableitungen mithilfe des Grenzwerts des Differenzenquotienten umständlich und mühsam ist. Deshalb werden in der Folge die wichtigsten Ableitungsregeln eingeführt, die das Finden der Ableitungen wesentlich erleichtern. Dabei werden zuerst grundlegende Regeln bewiesen und anschließend wird der Fokus auf die Ableitungen der elementaren Funktionen gelegt.

Satz 2.1 (z.B. Aufgabe 13.1 in [10])**.** ***(Summenregel)*** *Die Ableitung der Summe von differenzierbaren Funktionen ist gleich der Summe der Ableitung der einzelnen Funktionen. Es gilt:*

$$(f+g)' = f' + g' \tag{2.3}$$

Beweis. Wir definieren eine Funktion F als Summe zweier Funktionen f und g, das heißt, es gilt: $F := f + g$.

$$
\begin{aligned}
F'(x) &= \lim_{h \to 0} \frac{F(x+h) - F(x)}{h} \\
&= \lim_{h \to 0} \frac{(f(x+h) + g(x+h)) - (f(x) + g(x))}{h} \\
&= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} + \lim_{h \to 0} \frac{g(x+h) - g(x)}{h} \\
&= f'(x) + g'(x) = (f+g)'(x)
\end{aligned}
$$

□

Satz 2.2 (z.B. Aufgabe 13.1 in [10]). ***(Regel vom konstanten Faktor)*** *Für Ableitung einer differenzierbaren Funktion f mit einem konstanten Fakotor k gilt:*

$$(k \cdot f)' = k \cdot f' \tag{2.4}$$

Beweis. Wir definieren eine Funktion F als Produkt eines konstanten Faktors k und der differenzierbaren Funktion f. Es gilt somit: $F := k \cdot f$

$$
\begin{aligned}
F'(x) &= \lim_{h \to 0} \frac{F(x+h) - F(x)}{h} \\
&= \lim_{h \to 0} \frac{k \cdot f(x+h) - k \cdot f(x)}{h} \\
&= k \cdot \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = k \cdot f'(x)
\end{aligned}
$$

□

Satz 2.3 (z.B. Aufgabe 13.1 in [10]). ***(Produktregel)*** *Für die Ableitung des Produkts zweier differenzierbarer Funktionen f und g gilt:*

$$(f \cdot g)' = f' \cdot g + f \cdot g' \tag{2.5}$$

Beweis. Wir definieren eine Funktion F als Produkt zweier Funktionen f und g, das heißt, es gilt: $F := f \cdot g$

$$
\begin{aligned}
F'(x) &= \lim_{h \to 0} \frac{F(x+h) - F(x)}{h} \\
&= \lim_{h \to 0} \frac{f(x+h) \cdot g(x+h) - f(x) \cdot g(x)}{h} \\
&= \lim_{h \to 0} \frac{f(x+h) \cdot g(x+h) - f(x) \cdot g(x) + f(x) \cdot g(x+h) - f(x) \cdot g(x+h)}{h} \\
&= \lim_{h \to 0} \frac{g(x+h) \cdot (f(x+h) - f(x)) + f(x) \cdot (g(x+h) - g(x))}{h} \\
&= \lim_{h \to 0} g(x+h) \cdot \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} + f(x) \cdot \lim_{h \to 0} \frac{g(x+h) - g(x)}{h} \\
&= g(x) \cdot f'(x) + f(x) \cdot g'(x)
\end{aligned}
$$

□

Satz 2.4 (z.B. Aufgabe 13.1 in [10])**.** ***(Quotientenregel)*** *Für die Ableitung des Quotienten zweier differenzierbarer Funktionen f und g gilt:*

$$
(\frac{f}{g})' = \frac{f' \cdot g - f \cdot g'}{g^2} \tag{2.6}
$$

Da die Ableitung eines Quotienten durch Umschreiben als Produkt

$$
\frac{f}{g} = f \cdot \frac{1}{g}
$$

und die anschließende Anwendung der Produkt- und Kettenregel hergeleitet werden kann und im weiteren Verlauf der Arbeit keine tragende Rolle spielt, verzichten wir auf die explizite Herleitung.

Satz 2.5 (z.B. Aufgabe 14.1 in [10])**.** ***(Kettenregel)*** *Für die Ableitung zusammengesetzter differenzierbarer Funktionen f und g gilt:*

$$
(f(g(x)))' = f'(g(x)) \cdot g'(x) \tag{2.7}
$$

Beweis. Wir definieren eine Funktion F als Verknüpfung der Funktionen f und g. Es gilt: $F := f \circ g = f(g(x))$. Genauso setzen wir voraus, dass $g(x+h) - g(x) \neq 0$ gilt. Dann zeigt sich Folgendes:

$$
\begin{aligned}
F'(x) &= \lim_{h \to 0} \frac{F(x+h) - F(x)}{h} \\
&= \lim_{h \to 0} \frac{f(g(x+h)) - f(g(x))}{h} \\
&= \lim_{h \to 0} \frac{1}{h} \cdot \left(\frac{f(g(x+h)) - f(g(x))}{g(x+h) - g(x)} \cdot (g(x+h) - g(x)\right) \\
&= \lim_{h \to 0} \frac{f(g(x+h)) - f(g(x))}{g(x+h) - g(x)} \cdot \lim_{h \to 0} \frac{g(x+h) - g(x)}{h}
\end{aligned}
$$

Setzen wir nun $y_0 := g(x)$ und $y := g(x+h)$, erhalten wir:

$$
\begin{aligned}
F'(x) &= \lim_{y \to y_0} \frac{f(y) - f(y_0)}{y - y_0} \cdot g'(x) \\
&= f'(y_0) \cdot g'(x) = f'(g(x)) \cdot g'(x)
\end{aligned}
$$

□

Nachdem nun die grundlegenden Regeln erarbeitet wurden, sind noch die Ableitungsregeln für elementare Funktionen zu klären. Diese erleichtern den Schüler:innen wesentlich das Finden der gesuchten Ableitungen und im weiteren Sinne auch das Lösen von Extremwertaufgaben.

Satz 2.6 (z.B Satz in [13])**.** ***(Potenzregel für natürliche Exponenten)*** *Für eine Potenzfunktion $f(x) = x^n$ mit natürlichem Exponent $n > 1$ gilt:*

$$
f'(x) = n \cdot x^{n-1} \tag{2.8}
$$

Beweis. Mithilfe des Grenzwertes und der Polynomdivision (Regel von Horner) folgt:

$$
\begin{aligned}
f'(x_0) &= \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \\
&= \lim_{x \to x_0} \frac{x^n - x_0^n}{x - x_0} \\
&= \lim_{x \to x_0} \frac{(x - x_0) \cdot (x^{n-1} + x^{n-2}x_0 + x^{n-3}x_0^2 + \cdots + xx_0^{n-2} + x_0^{n-1})}{x - x_0} \\
&= x_0^{n-1} + x_0^{n-1} + x_0^{n-1} + \cdots + x_0^{n-1} + x_0^{n-1} \\
&= n \cdot x_0^{n-1}
\end{aligned}
$$

□

Satz 2.7 (z.B. Satz in [10]). ***(Ableitung von Polynomfunktionen)*** *Sei f eine Summe von Potenzfunktionen mit natürlichem Exponenten. Dann gilt aufgrund von 2.1, 2.2 und 2.6 für die Ableitung:*

$$f'(x) = (\sum_{k=0}^{n} a_k x^k)' = \sum_{k=1}^{n} k a_k x^{k-1} \tag{2.9}$$

Satz 2.8 (z.B. Satz in [10]). ***(Ableitung der Exponentialfunktionen)*** *Für die Ableitung der Exponentialfunktionen $f(x) = e^x$ und $g(x) = a^x$ (wobei $a \in \mathbb{R}^+, a \neq 1$) gelten:*

(i)

$$f'(x) = (e^x)' = e^x \tag{2.10}$$

(ii)

$$g'(x) = a^x \cdot ln(a) \tag{2.11}$$

Beweis. (i)

$$\begin{aligned} f'(x) &= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \to 0} \frac{e^{x+h} - e^x}{h} \\ &= e^x \cdot \lim_{h \to 0} \frac{e^h - 1}{h} \\ &= e^x \cdot 1 = e^x \end{aligned}$$

(ii) Hier folgt aufgrund von 2.8 und 2.5

$$\begin{aligned} g'(x) &= (a^x)' = (e^{ln(a)^x})' = (e^{ln(a)\cdot x})' \\ &= e^{ln(a)\cdot x} \cdot ln(a) = a^x \cdot ln(a) \end{aligned}$$

□

Satz 2.9 (z.B. Satz in [10]). ***(Ableitung von trigonometrischen Funktionen)*** *Für die Ableitungen der elementaren trigonometrischen Funktionen $f(x) = \sin(x)$, $g(x) = \cos(x)$ und $h(x) = \tan(x)$ gelten:*

(i)

$$f'(x) = (\sin(x))' = \cos(x) \tag{2.12}$$

(ii)

$$g'(x) = (\cos(x))' = -\sin(x) \tag{2.13}$$

(iii)

$$h'(x) = (\tan(x))' = \frac{1}{\cos^2(x)} \tag{2.14}$$

Beweis. Mit Hilfe des Grenzwertbegriffs und der Additionstheoreme für Sinus und Kosinus folgt:

(i)

$$\begin{aligned} f'(x) &= \lim_{h \to 0} \frac{\sin(x+h) - \sin(x)}{h} \\ &= \lim_{h \to 0} \frac{\sin(x)\cos(h) + \cos(x)\sin(h) - \sin(x)}{h} \\ &= \lim_{h \to 0} \frac{\sin(x) \cdot (\cos(h) - 1) + \cos(x)\sin(h)}{h} \\ &= \sin(x) \cdot 0 + \cos(x) \cdot 1 = \cos(x) \end{aligned}$$

(ii)

$$\begin{aligned} g'(x) &= \lim_{h \to 0} \frac{\cos(x+h) - \cos(x)}{h} \\ &= \lim_{h \to 0} \frac{\cos(x)\cos(h) - \sin(x)\sin(h) - \cos(x)}{h} \\ &= \lim_{h \to 0} \frac{\cos(x) \cdot (\cos(h) - 1) - \sin(x)\sin(h)}{h} \\ &= \cos(x) \cdot 0 - \sin(x) \cdot 1 = -\sin(x) \end{aligned}$$

(iii) Mithilfe der Quotientenregel 2.4 und der zuvor hergeleiteten Ableitung von Sinus und Kosinus folgt:

$$\begin{aligned} h'(x) = (\tan(x))' &= \left(\frac{\sin(x)}{\cos(x)}\right)' \\ &= \frac{(\sin(x))'\cos(x) - \sin(x)(\cos(x))'}{\cos^2(x)} \\ &= \frac{\cos(x)\cos(x) + \sin(x)\sin(x)}{\cos^2(x)} \\ &= \frac{\cos^2(x) + \sin^2(x)}{\cos^2(x)} = \frac{1}{\cos^2(x)} \end{aligned}$$

□

Satz 2.10 (z.B. Satz in [10])**.** ***(Ableitung von Logarithmusfunktionen)*** *Für die Ableitungen der Logarithmusfunktionen* $f(x) = \ln(x)$ *und* $g(x) = \log_a(x)$ *(wobei* $a > 0, a \neq 1$*) gelten:*

(i)

$$f'(x) = (\ln(x))' = \frac{1}{x} \tag{2.15}$$

(ii)

$$g'(x) = (\log_a(x))' = \frac{1}{x \cdot \ln(a)} \tag{2.16}$$

Beweis. (i)

$$f'(x) = \lim_{h \to 0} \frac{\ln(x+h) - ln(x)}{h} = \lim_{h \to 0} \frac{1}{h} \cdot \ln(\frac{x+h}{x}) = \lim_{h \to 0} \frac{1}{h} \cdot \ln(1 + \frac{h}{x})$$

Setzen wir nun $u = \frac{h}{x}$ und somit $h = u \cdot x$ folgt:

$$= \lim_{u \to 0} \frac{1}{u \cdot x} \cdot \ln(1+u) = \frac{1}{x} \cdot \lim_{u \to 0} \frac{\ln(1+u)}{u} = \frac{1}{x} \cdot 1 = \frac{1}{x}$$

Nun lässt sich die Ableitung von $g(x) = \log_a(x)$ direkt aus der Ableitung von $ln(x)$ herleiten:

$$g'(x) = (\log_a(x))' = (\frac{\ln(x)}{\ln(a)})' = (\ln(x))' \frac{1}{\ln(a)} = \frac{1}{x} \cdot \frac{1}{\ln(a)} = \frac{1}{x \cdot \ln(a)}$$

□

Neben der Erarbeitung der grundlegenden Ableitungsregeln ist auch das Anwenden und Üben dieser von großer Bedeutung für die Schüler:innen der Sekundarstufe II. Vor allem der Verknüpfung der unterschiedlichen Regeln sollte dabei große Aufmerksamkeit geschenkt werden.

Beispiel. Berechne die Ableitung folgender Funktionen mithilfe der Ableitungsregeln!

(i) $f(x) = 3x^2 + 2x + 7$

(ii) $h(x) = \tan(x) \cdot e^{2x}$

Lösung:
Mithilfe der zuvor hergeleiteten Ableitungsregeln können die Beispiele, auch von Schüler:innen, schnell analytisch gelöst werden.

(i)

$$f'(x) = 2 \cdot 3x + 2 \cdot x^0 = 6x + 2$$

(ii)

$$\begin{aligned} h'(x) &= (\tan(x))' \cdot e^{2x} + \tan(x) \cdot (e^{2x})' = \frac{1}{\cos^2(x)} \cdot e^{2x} + \tan(x) \cdot e^{2x} \cdot 2 \\ &= \frac{e^{2x}}{\cos^2(x)} + 2e^{2x} \cdot \tan(x) \end{aligned}$$

2.4 Analytische Methoden zum Finden der Nullstellen

Der zweite wichtige analytische Eckpfeiler zum Lösen von Extremwertaufgaben ist neben dem Finden der ersten Ableitung das Finden der Extremstelle. Dies erfolgt durch Berechnung der Nullstellen der ersten Ableitung. Dieses Unterkapitel soll kurz die wesentlichen Vorgehensweisen zum analytischen Berechnen der Nullstellen erläutern und basiert dabei größtenteils auf Kapitel 3 und 5 in [5] sowie Kapitel 7 in [9]

Definition. (Nullstelle einer Funktion) Unter einer Nullstelle einer Funktion f versteht man jene Stelle, an der f die x-Achse schneidet oder berührt. Das bedeutet, dass an dieser Stelle $f(x) = 0$ gilt.

Je nach Art der Funktion gibt es unterschiedliche Möglichkeiten, die Nullstellen zu berechnen. Grundlage dafür sind die elementaren Äquivalenzumformungen, die bereits in der Sekundarstufe I erlernt werden.

Definition. (Elementarumformungsregeln)

$$A + B = C \Leftrightarrow A = C - B \tag{2.17}$$

$$A \cdot B = C \Leftrightarrow A = C/B \quad (B \neq 0) \tag{2.18}$$

Zum Finden von Nullstellen bei linearen Funktionen sind die Elementarumformungsregeln ausreichend, wie im folgenden Beispiel zu sehen ist:

Beispiel. Berechne die Nullstelle von $f(x) = 2x + 4$!

Lösung:
Mithilfe der Elementarumformungeregeln lässt sich die gewünschte Nullstelle schnell berechnen.

$$\begin{aligned} 2x + 4 &= 0 && | -4 \\ \Leftrightarrow \quad 2x &= -4 && | : 2 \\ \Leftrightarrow \quad x &= -2 \end{aligned}$$

Anspruchsvoller wird das Finden von Nullstellen bei Polynomfunktionen mit höheren Exponenten. Hier gibt es aber einige Tricks, die den Schüler:innen das Finden der gesuchten Stelle erleichtern. So kann zum Beispiel durch gezieltes Umstellen der Funktion (Produkt-Null-Satz), Herausheben von Faktoren oder Polynomdivision der Grad der Polynomfunktion verringert werden. Bei quadratischen Funktionen bietet sich schließlich die Lösungsformel zum Bestimmen der Nullstellen an.

Satz 2.11. ***(Nullstellen bei quadratischen Funktionen)*** *Bei quadratischen Funktionen der Form $f(x) = ax^2 + bx + c$ lassen sich die Nullstellen wie folgt ermitteln:*

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{2.19}$$

Beweis.

$$\begin{aligned} f(x) = ax^2 + bx + c &= 0 \\ \Leftrightarrow x^2 + \frac{b}{a}x + \frac{c}{a} &= 0 \\ \Leftrightarrow (x + \frac{b}{2a})^2 - (\frac{b}{2a})^2 &= -\frac{c}{a} \\ \Leftrightarrow (x + \frac{b}{2a})^2 &= \frac{b^2}{4a^2} - \frac{c}{a} \\ \Leftrightarrow (x + \frac{b}{2a})^2 &= \frac{b^2 - 4ac}{4a^2} \\ \Leftrightarrow (x + \frac{b}{2a}) &= \pm\frac{\sqrt{b^2 - 4ac}}{2a} \\ \Leftrightarrow x_{1,2} &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \end{aligned}$$

□

Anhand folgender Beispiele lassen sich die Möglichkeiten des analytischen Berechnens von Nullstellen demonstrieren.

Beispiel. Berechne die Nullstellen folgender Funktionen!

(i)
$$f(x) = 5x^3 - 5x^2 + 2x - 2$$

(ii)
$$g(x) = 2x^2 + 6x - 56$$

Lösung:

(i) Das erste Beispiel lässt sich durch Umstellen der Funktion bzw. durch Herausheben des Faktors $(x-1)$ und anschließendes Anwenden des Produkt-Null-Satzes lösen.

$$\begin{aligned} f(x) &= 5x^3 - 5x^2 + 2x - 2 = 0 \\ &= (x-1) \cdot (5x^2 + 2) = 0 \\ &\Rightarrow (x-1) = 0 \\ &\Rightarrow x_1 = 1 \end{aligned}$$

Der zweite Term $(5x^2+2)$ liefert uns keine weitere reelle Nullstelle, da nach Durchführung der Elementarumformungen die Wurzel einer negativen Zahl zu ziehen wäre. Somit hat die Funktion f nur eine Nullstelle mit $x_1 = 1$.

(ii) Bei diesem Beispiel lässt sich die Lösungsformel für quadratische Funktionen anwenden.

$$\begin{aligned} g(x) &= 2x^2 + 6x - 56 = 0 \\ &\Rightarrow x_{1,2} = \frac{-6 \pm \sqrt{36 + 448}}{4} = \frac{-6 \pm 22}{4} \\ &\Rightarrow x_1 = \frac{-6 + 22}{4} = 4 \qquad x_2 = \frac{-6 - 22}{4} = -7 \end{aligned}$$

Auch bei weiteren Funktionsarten lassen sich durch einfache Umformungen die Nullstellen berechnen. Wichtig dabei ist, dass den Schüler:innen bewusst ist, welche Umformungen bei unterschiedlichen Funktionstypen zum Ziel führen. So werden zum Beispiel Exponentialfunktionen mithilfe des natürlichen Logarithmus (vice versa) und Wurzeln mittels entsprechenden Exponenten aufgelöst. Bei komplexeren Funktionen stoßen die

analytischen Methoden jedoch an ihre Grenzen und die Nullstellen können nur mittels numerischer Verfahren angenähert werden. Diese werden im Kapitel 5 detailliert erläutert.

2.5 Anwendung der Methoden

In diesem Unterkapitel werden nun die zuvor definierten Vorgehensweisen und Methoden zum analytischen Lösen von Extremwertaufgaben angewandt. Dabei werden zunächst klassische Unterrichtsbeispiele diskutiert und komplexere Aufgaben gelöst.

Beispiel (Aufgabe 3.105 aus [6]). An eine Mauer angrenzend soll mit 20 m Drahtzaun ein rechteckiges Areal so eingezäunt werden, dass dessen Flächeninhalt möglichst groß ist. Ist das Areal quadratisch zu wählen? Wenn nicht, wie lange sind seine Seitenlängen zu wählen?

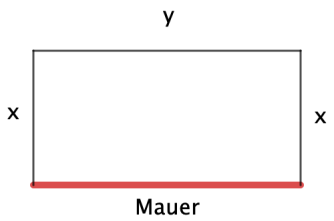


Abbildung 3: Skizze der Extremwertaufgabe

Lösung:

Schritt 1: Zuerst wird die Aufgabe analysiert und anschließend werden die vorkommenden Größen bestimmt. In unserem Fall sollen an eine bestehende Mauer drei Seitenlängen angefügt werden, um den Flächeninhalt des Rechtecks zu maximieren. Die Variablen definieren wir wie folgt:
x $(0 \leq x \leq 10)$: beide Breiten des Zaunes, senkrecht zur Mauer
y $(0 \leq y \leq 20)$: Länge des Zaunes, parallel zur Mauer

Schritt 2: Nun stellen wir die Hauptbedingung HB abhängig von x und y auf:

$$A(x, y) := x \cdot y \rightarrow max$$

Schritt 3: Als nächstes benötigen wir die Nebenbedingung NB. Aus der Angabe

kann entnommen werden, dass der dazukommende Zaun 20 m lang ist. Wir erhalten

$$2x + y = 20 \Rightarrow y = 20 - 2x$$

Schritt 4: Nun kann die Nebenbedingung in die Hauptbedingung eingesetzt werden:

$$A(x) := x \cdot (20 - 2x) = 20x - 2x^2$$

Schritt 5: Im nächsten Schritt ist die Zielfunktion $A(x)$ abzuleiten und anschließend deren Nullstelle zu finden, um die optimale Lösung zu erhalten.

$$A'(x) = 20 - 4x = 0 \Rightarrow 20 = 4x \Rightarrow 5 = x$$

Schritt 6: Nun gilt es zu überprüfen, ob tatsächlich ein Maximum vorliegt, indem wir die Randstellen in die Zielfunktion einsetzen:

$$\begin{aligned} A(0) &= 0 \cdot (20 - 2 \cdot 0) = 0 \\ A(20) &= 20 \cdot (20 - 2 \cdot 20) = -400 \\ A(5) &= 5 \cdot (20 - 2 \cdot 5) = 50 \end{aligned}$$

Wie wir sehen, liegt für $x = 5m$ wirklich ein Maximum vor. Die dazugehörige Länge, welche parallel zur Mauer liegt, ist $y = 20 - 2 \cdot 5 = 10m$ lang. Der maximale Flächeninhalt $A(x, y) = x \cdot y$ beträgt dementsprechend $50m^2$.

Bei der Einführung von Extremwertaufgaben im Unterricht sollte zu Beginn ein Beispiel wie zuvor Schritt-für-Schritt durchgerechnet werden. In der weiteren Übungsphase hingegen wird dies von den Lernenden aus Zeitgründen nicht erwartet, weshalb auch die folgenden Beispiele nicht alle einzelnen Schritte enthalten.

Beispiel. Für die Herstellung einer Getränkedose werden $300cm^2$ Weißblech verwendet. Welche Abmessungen sollte die Dose haben, damit ihr Fassungsvermögen möglichst groß wird? Berechne! [Aufgabe 431 aus [2]]

Lösung:
HB:

$$V(r, h) = r^2 \pi h \rightarrow max$$

NB:

$$O = 300cm^2 \rightarrow 2r^2\pi + 2r\pi h = 300$$
$$2r\pi h = 300 - 2r^2\pi$$
$$h = \frac{300 - 2r^2\pi}{2r\pi} = \frac{150 - r^2\pi}{r\pi}$$

Für den Definitionsbereich müssen $r \geq 0$ sowie $h \geq 0$ gelten. Formen wir die Nebenbedingung nach r um, erhalten wir $0 \leq r \leq \sqrt{\frac{150}{\pi}}$.
Anschließend setzen wir die Nebenbedingung in die Hauptbedingung ein, leiten diese ab und suchen die Nullstelle:

$$V(r) = r^2\pi\frac{150 - r^2\pi}{r\pi} = \frac{150r^2\pi - r^4\pi^2}{r\pi} = 150r - r^3\pi$$

$$V'(r) = 150 - 3r^2\pi = 0$$
$$\Rightarrow 150 = 3r^2\pi \Rightarrow 50 = r^2\pi \Rightarrow r = \pm\sqrt{\frac{50}{\pi}}$$

Da ein negativer Radius für die Dose keinen Sinn ergeben würde und auch außerhalb des Definitionsbereiches liegt, gilt $r = \sqrt{\frac{50}{\pi}} \approx 3,99cm$.
Bevor wir aber vom Optimum ausgehen können, müssen die Randstellen überprüft werden:

$$V(0) = 150 \cdot 0 - 0^3\pi = 0$$
$$V(\sqrt{\frac{150}{\pi}}) = 150 \cdot \sqrt{\frac{150}{\pi}} - (\sqrt{\frac{150}{\pi}})^3\pi = 0$$
$$V(\sqrt{\frac{50}{\pi}}) = 150 \cdot \sqrt{\frac{50}{\pi}} - (\sqrt{\frac{50}{\pi}})^3\pi \approx 398,94$$

Somit sehen wir, dass das maximale Volumen bei einem Radius von $r = \sqrt{\frac{50}{\pi}} \approx 3,99cm$ liegen muss und rund $398,94cm^3$ beträgt. Um abschließend die dazugehörige Höhe berechnen zu können, setzen wir in die Nebenbedingung ein und erhalten

$$h = \frac{150 - r^2\pi}{r\pi} = \frac{150 - \sqrt{\frac{50}{\pi}}^2\pi}{\sqrt{\frac{50}{\pi}}\pi} \approx 7,98cm$$

Das abschließende Beispiel dieses Kapitels soll nun demonstrieren, dass die analytischen Methoden allein oft nicht ausreichen, um eine Extremwertaufgabe zu lösen.

Beispiel (Aufgabe 19 aus [1]). Für den Bau eines Tores stehen für die senkrechten Begrenzungen zwei Pfosten mit $2m$ Länge und für die oben aufgesetzen schrägen Begrenzungen zwei Dachpfosten mit je $4m$ Länge zur Verfügung. Wie muss der Öffnungswinkel der Dachpfosten gewählt werden, wenn das Tor eine möglichst große Durchlassfläche (Querschnittsfläche) haben soll?

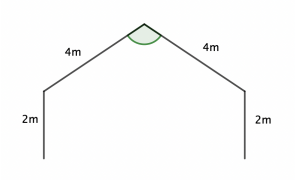


Abbildung 4: Skizze der Extremwertaufgabe [Aufgabe 19 aus [1]]

Lösung:
HB:

$$A(b, h_b) = 2 \cdot b + \frac{b \cdot h_b}{2}$$

NB:

$$sin(\frac{\alpha}{2}) = \frac{\frac{b}{2}}{4} \Rightarrow b = 8 \cdot sin(\frac{\alpha}{2}) \quad (0° \leq \alpha \leq 180°)$$
$$cos(\frac{\alpha}{2}) = \frac{h_b}{4} \Rightarrow h_b = 4 \cdot cos(\frac{\alpha}{2}) \quad (0° \leq \alpha \leq 180°)$$

Setzen wir die beiden Nebenbedingungen nun in die Hauptbedingung ein, erhalten wir eine Funktion in Abhängigkeit des Öffnungswinkels α und infolgedessen die Ableitung $A'(\alpha)$:

$$
\begin{aligned}
A(\alpha) &= 2 \cdot 8 \cdot sin(\frac{\alpha}{2}) + \frac{8 \cdot sin(\frac{\alpha}{2}) \cdot 4 \cdot cos(\frac{\alpha}{2})}{2} = 16 \cdot sin(\frac{\alpha}{2}) + 16 \cdot sin(\frac{\alpha}{2}) \cdot cos(\frac{\alpha}{2}) \\
&= 16 \cdot (sin(\frac{\alpha}{2}) + sin(\frac{\alpha}{2}) \cdot cos(\frac{\alpha}{2})) \\
A'(\alpha) &= 16 \cdot (cos(\frac{\alpha}{2}) \cdot \frac{1}{2} + cos(\frac{\alpha}{2}) \cdot cos(\frac{\alpha}{2}) \cdot \frac{1}{2} + sin(\frac{\alpha}{2}) \cdot -sin(\frac{\alpha}{2}) \cdot \frac{1}{2}) \\
&= 16 \cdot (cos(\frac{\alpha}{2}) \cdot \frac{1}{2} + cos^2(\frac{\alpha}{2}) - sin^2(\frac{\alpha}{2}) \cdot \frac{1}{2}) \\
&= 8 \cdot (cos(\frac{\alpha}{2}) + cos^2(\frac{\alpha}{2}) - sin^2(\frac{\alpha}{2}))
\end{aligned}
$$

Im nächsten Schritt sollte nun die Ableitung $A'(\alpha)$ gleich Null gesetzt werden, was aber mit den zuvor besprochenen analytischen Methoden nicht mehr zu bewältigen wäre. Die Schüler:innen könnten mithilfe eines Computeralgebrasystems oder einer Darstellung des Funktionsgraphens zum Ziel kommen. In unserem Fall werden wir dieses Beispiel aber im weiteren Verlauf der Arbeit noch einmal aufgreifen und mittels numerischer Methoden zu lösen versuchen.

3 Grundlagen der numerischen Mathematik

Dieses Kapitel soll erläutern, was unter numerischer Mathematik verstanden wird und in welchen Bereichen sie Anwendung findet. Darüber hinaus sollen auch die wichtigsten Fehlerarten, die Hand in Hand mit der Anwendung numerischer Verfahren gehen, diskutiert werden. Dieses Kapitel basiert größtenteils auf Kapitel 1 in [4] und Kapitel 1 in [12].

3.1 Was versteht man unter numerischer Mathematik?

Die numerische Mathematik stellt ein Teilgebiet des wissenschaftlichen Rechnens (*scientific computing*) dar, welches Gebiete der Mathematik, Informatik und Naturwissenschaften einschließt. Die Numerik befasst sich dabei mit mathematischen Lösungsverfahren, um verschiedenste Problemstellungen mittels Computertechnologie zu lösen. Neben der Mathematik spielt dementsprechend auch die Wahl der Hard- und Software eine wesentliche Rolle, um die Probleme sinnvoll zu lösen. Mittlerweile sind leistungsstarke Rechner in der Lage, in den unterschiedlichsten Bereichen Lösungen zu eruiren. So findet die Numerik nicht nur als Teilgebiet der Mathematik, sondern auch im Alltag in Natur-, Ingenieur- und Wirtschaftswissenschaften Anwendung. Grundlage des numerischen Rechnens bilden dabei Algorithmen.

Definition (z.B. Definition 1.1 in [4])**.** Unter einem *numerischen Algorithmus* versteht man eine Vorschrift, die endlich vielen Eingabedaten und Problemfunktionen in eindeutiger Weise einen Satz von endlich vielen Ausgabedaten (Resultaten) zuordnet, wobei nur endlich viele arithmetische und logische Operationen und Auswertungen der Problemfunktionen zulässig sind.

Im Folgenden werden die wichtigsten Schritte beim Finden einer numerischen Lösung für ein Anwendungsproblem erläutert.

1. Modellierung: Sobald ein Problem definiert wurde, wird das Ziel gesetzt, ein passendes mathematisches Modell aufzustellen. Da hierbei meistens Annahmen getroffen werden, findet eine erste Annäherung statt.

2. Realisierung: In diesem Schritt muss nun eine Methode zur Lösung des Problems gefunden werden, wobei es oftmals verschiedene Verfahren gibt, die zum Einsatz kommen können. Gleichzeitig sollte geklärt werden, mit welchem Programm das Verfahren

umgesetzt wird oder ob eine neue Software entwickelt werden muss.

3. Validierung: Da beim numerischen Rechnen häufig mit Näherungen gearbeitet wird und zudem Fehler – etwa Rundungs- oder Eingabefehler – auftreten können, ist es notwendig, sowohl das zugrunde liegende Modell auf seine Gültigkeit als auch das Programm und das gewählte Verfahren auf ihre Zuverlässigkeit zu überprüfen. Werden konkrete numerische Daten verwendet, sollte die Berechnung daher idealerweise stets von einer Fehleranalyse unterstützt werden. In der Praxis ist dies jedoch nicht immer möglich.

3.2 Fehlerarten

Wie zuvor schon kurz angeschnitten wurde, spielen Fehler eine tragende Rolle in der numerischen Mathematik. Die Ergebnisse von Aufgaben, die durch numerische Verfahren gelöst werden, können durch verschiedenste Faktoren bzw. Fehler beeinflusst werden und dementsprechend das Resultat verfälschen. Nachfolgend sollen die wichtigsten Fehlerarten beschrieben und erklärt werden.

Eingabefehler: Diese Art von Fehler entsteht bei der Eingabe der gewünschten Daten in den Computer. Benötigt man zum Beispiel eine irrationale Konstante wie π, entsteht durch die Rundung bei der Eingabe ein erster Fehler. Genauso kann es aber passieren, dass mit Daten, die durch Messungen im Labor nicht zu 100 Prozent exakt sind, gearbeitet wird, was zu fehlerhaften Resultaten führt.

Approximationsfehler: Dieser Fehler tritt auf, wenn eine unendliche Vorgehensweise durch eine endliche ersetzt wird. Dies passiert zum Beispiel bei der Berechnung der ersten Ableitung, wenn die Approximation mit einer sehr kleinen Konstante abgeschätzt wird.

Rundungsfehler: Auch wenn dies schon bei den Eingabefehlern thematisiert wurde, ist es wichtig, die Rundungsfehler noch einmal als eigene Fehlerquelle zu nennen. Da mittels Computerzahlen nicht unendlich viele Nachkommastellen angezeigt werden können, muss zu einem gewissen Zeitpunkt das Verfahren abgebrochen und das Ergebnis gerundet werden. Auch wenn die Rundungsfehler meistens sehr klein sind, können sie sich bei aufwendigen Prozessen so summieren, dass erhebliche Verfälschungen entstehen. Wichtig zu erwähnen ist hierbei auch die Darstellung der Zahlen, allen voran die Maschinengenauigkeit. Diese legt den maximalen relativen Rundungsfehler bei Gleitkommazahlen

fest. Moderne Computer verwenden in der Regel die festgelegten Standards des *Institute of Electrical and Electronic Engineers (IEEE)*. Dabei wird für die einfache Genauigkeit $\epsilon = 2^{-24} \approx 10^{-8}$ und für die doppelte Genauigkeit $\epsilon = 2^{-53} \approx 10^{-16}$ verwendet. Im späteren Verlauf der Arbeit wird bei der Programmiersprache Julia mit doppelter Genauigkeit gearbeitet.

Soft- und Hardwarefehler: Auch diese Fehlerquelle lässt sich nicht vermeiden, da es trotz umfangreicher Testungen passieren kann, dass die verwendete Soft- oder Hardware versagt. Die Beseitigung dieser Fehler findet aber nicht im Rahmen der numerischen Mathematik statt.

Wie gerade erläutert, ist es üblich, dass in der Numerik Fehler auftreten. Die Größe der Fehler wird mittels folgender zwei Arten ermittelt:

Definition (z.B. Definition 1.2. in [4])**.** Es sei $\tilde{x} \equiv (\tilde{x}_1, \ldots, \tilde{x}_n)^T \in \mathbb{R}^n$ eine mathematische Näherung für den Vektor der exakten Eingabedaten $x \equiv (x_1, \ldots, x_n)^T \in \mathbb{R}^n$. Man unterscheidet:

(i) *absolute Fehler:* $\delta(\tilde{x}_k) \equiv \tilde{x}_k - x_k, k = 1, \ldots, n$;

(ii) *relative Fehler:* $\epsilon(\tilde{x}_k) \equiv \frac{\tilde{x}_k - x_k}{x_k} = \frac{\delta(\tilde{x}_k)}{x_k}, k = 1, \ldots, n, (x_k \neq 0)$.

4 Differenzieren mittels dualer Zahlen

Nachdem im vorhergehenden Kapitel die Grundlagen der numerischen Mathematik besprochen wurden, wird nun der erste der zwei wesentlichen Bereiche zur Lösung von Extremwertaufgaben, das Differenzieren, mittels numerischer Mathematik erarbeitet. In den folgenden Unterkapiteln wird zuerst die Methode des automatisierten Differenzierens mit Hilfe der komplexen Zahlen hergeleitet und anschließend angewandt. Die Struktur und die Implementierungen der dualen Zahlen und im weiteren Sinne der automatisierten Differenzierung basieren auf Skripten, die von Herrn Prof. Dr. Martin Ehler zur Verfügung gestellt wurden.

4.1 Grundlagen der komplexen Zahlenmenge $\mathbb{C}$

Damit das Prinzip des automatisierten Differenzierens mittels dualer Zahlen nachvollzogen werden kann, wird zunächst der Zahlenraum der komplexen Zahlen detailliert beschrieben, da zwischen den komplexen und dualen Zahlen einige Gemeinsamkeiten bestehen. Dieses Unterkapitel basiert größtenteils auf Kapitel 18 in [7] und Kapitel 15 in [14].

Der komplexe Zahlenraum $\mathbb{C}$ gilt als Erweiterung der reellen Zahlen $\mathbb{R}$. Er wurde eingeführt, da Gleichungen wie $x^2 = -1$ in $\mathbb{R}$ nicht lösbar sind, weil Wurzeln nur für positive reelle Werte definiert sind. Mithilfe der Einführung der imaginären Einheit i wurde diesem Problem entgegengewirkt.

Definition (z.B. Definition in [14])**. (Komplexe Zahl)** Die allgemeine Form einer komplexen Zahl lautet:

$$z = a + i \cdot b \quad \forall a, b \in \mathbb{R}, z \in \mathbb{C} \tag{4.1}$$

Bei der Zahl i handelt es sich um die imaginäre Einheit, für welche $i^2 = -1$ gilt. Die reelle Zahl a wird als Realteil, die reelle Zahl b als Imaginärteil bezeichnet:

$$a = Re(z); \quad b = Im(z) \tag{4.2}$$

Werden nur der Realteil und der Imaginärteil betrachtet, so kann eine komplexe Zahl als Tupel (a, b) dargestellt werden, also als ein geordnetes Paar zweier reeller Zahlen (a, b). Jede komplexe Zahl z kann erzeugt werden, wenn für a und b reelle Werte angenommen werden. Ist $b = 0$, ergibt sich für $z = a$ eine reelle Zahl; ist $a = 0$, ergibt sich für $z = i \cdot b$ eine rein imaginäre Zahl.

Jede komplexe Zahl kann auch mithilfe eines Koordinatensystems (Gauß'sche Zahlenebene) dargestellt werden. Dabei wird die x-Achse für den Realteil und die y-Achse für den Imaginärteil betrachtet, wie in Abbildung 5 zu sehen ist. Auffallend in der Abbildung sind auch die Strecke r (Betrag von z) sowie der Winkel φ, die beide für die Darstellung mittels Polarkoordinaten von großer Bedeutung sind.

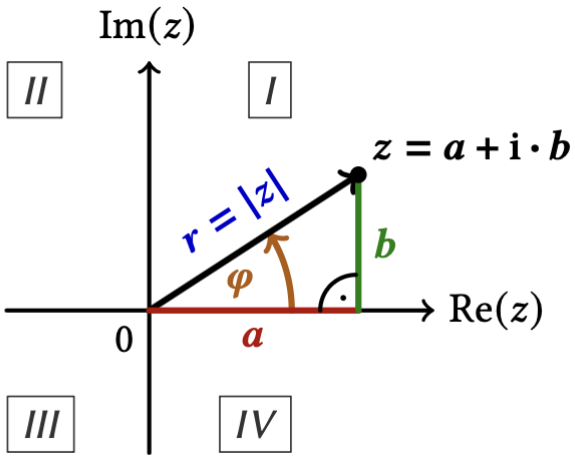


Abbildung 5: Graphische Darstellung einer komplexen Zahl z [Abbildung S. 296 in [7]]

Definition. (Betrag einer komplexen Zahl) Der Betrag $|z|$ einer komplexen Zahl z ist die Länge des Vektors in der Gauß'schen Ebene:

$$|z| = |a + i \cdot b| = \sqrt{a^2 + b^2} \tag{4.3}$$

4.1.1 Rechnen mit komplexen Zahlen

Für die Herleitung der Rechenregeln mit den komplexen Zahlen betrachten wir $z = a+i\cdot b$ als Tupel (a, b).

Definition. (Addition und Subtraktion mit komplexen Zahlen) Für die Addition und Subtraktion zweier komplexer Zahlen $z_1 = (a_1, b_1)$ und $z_2 = (a_2, b_2)$ gilt:

(i) $z_1 + z_2 = (a_1 + a_2, b_1 + b_2)$

(ii) $z_1 - z_2 = (a_1 - a_2, b_1 - b_2)$

Wie in der Definition zu sehen ist, funktionieren die komplexe Addition sowie die komplexe Subtraktion komponentenweise. Es werden dementsprechend zuerst die Realteile

und analog dazu die Imaginärteile addiert bzw. subtrahiert. Damit dies besser nachvollzogen werden kann, werden Addition und Subtraktion in der Form $z = a + i \cdot b$ gezeigt. Für $z_1 = a_1 + i \cdot b_1$ und $z_2 = a_2 + i \cdot b_2$ gilt also:

$$\begin{aligned} z_1 + z_2 &= (a_1 + i \cdot b_1) + (a_2 + i \cdot b_2) = a_1 + i \cdot b_1 + a_2 + i \cdot b_2 \\ &= a_1 + a_2 + i \cdot b_1 + i \cdot b_2 = a_1 + a_2 + i \cdot (b_1 + b_2) = (a_1 + a_2, b_1 + b_2) \\ z_1 - z_2 &= (a_1 + i \cdot b_1) - (a_2 + i \cdot b_2) = a_1 + i \cdot b_1 - a_2 - i \cdot b_2 \\ &= a_1 - a_2 + i \cdot b_1 - i \cdot b_2 = a_1 - a_2 + i \cdot (b_1 - b_2) = (a_1 - a_2, b_1 - b_2) \end{aligned}$$

Definition. (Multiplikation mit komplexen Zahlen) Für die Multiplikation zweier komplexer Zahlen $z_1 = (a_1, b_1)$ und $z_2 = (a_2, b_2)$ gilt:

$$z_1 \cdot z_2 = (a_1 a_2 - b_1 b_2, a_1 b_2 + a_2 b_1)$$

Auch hier wird mithilfe der allgemeinen Darstellung der komplexen Zahlen als $z = a + i \cdot b$ die Rechenregel der Multiplikation erläutert. Für $z_1 = a_1 + i \cdot b_1$ und $z_2 = a_2 + i \cdot b_2$ gilt:

$$\begin{aligned} z_1 \cdot z_2 &= (a_1 + i \cdot b_1) \cdot (a_2 + i \cdot b_2) \\ &= a_1 \cdot a_2 + a_1 \cdot i \cdot b_2 + a_2 \cdot i \cdot b_1 + i \cdot b_1 \cdot i \cdot b_2 \\ &= a_1 \cdot a_2 + a_1 \cdot i \cdot b_2 + a_2 \cdot i \cdot b_2 - b_1 \cdot b_2 \\ &= a_1 \cdot a_2 - b_1 \cdot b_2 + i \cdot (a_1 \cdot b_2 + a_2 \cdot b_1) = (a_1 \cdot a_2 - b_1 \cdot b_2, a_1 \cdot b_2 + a_2 \cdot b_1) \end{aligned}$$

Bei der Multiplikation ist auffallend, dass aufgrund der Vorschreibung von $i^2 = -1$ die Multiplikation beider Imaginärteile zu einem Realteil übergeht, wobei das Vorzeichen gewechselt wird.

Bevor auf die Division näher eingegangen wird, ist es noch wichtig, die konjugiert komplexe Zahl einzuführen. Diese wird benötigt, um den Nenner der komplexen Division in eine rationale Zahl zu überführen.

Definition. (Konjugiert komplexe Zahl) Die konjugiert komplexe Zahl $\overline{z}$ einer komplexen Zahl $z = a + i \cdot b$ ist definiert als:

$$\overline{z} = a - i \cdot b = (a, -b) = Re(z) - i \cdot Im(z)$$

Definition. (Division mit komplexen Zahlen) Für die Division zweier komplexer Zahlen $z_1 = (a_1, b_1)$ und $z_2 = (a_2, b_2)$ werden Zähler und Nenner jeweils mit der konjugiert komplexen Zahl des Nenners erweitert, damit der Imaginärteil des Nenners wegfällt. Es gilt dementsprechend:

$$\frac{z}{w} = \frac{(a_1, b_1)}{(a_2, b_2)} = \frac{(a_1, b_1)}{(a_2, b_2)} \cdot \frac{(a_2, -b_2)}{(a_2, -b_2)} = \frac{(a_1 a_2 + b_1 b_2, a_2 b_1 - a_1 b_2)}{((a_2)^2 + (b_2)^2)}$$

Analog zu den vorherigen Rechenoperationen wird auch die Division mithilfe der Darstellung von $z = a + i \cdot b$ erklärt. Für zwei komplexe Zahlen $z_1 = a_1 + i \cdot b_1$ und $z_2 = a_2 + i \cdot b_2$ gilt also:

$$\begin{aligned}
\frac{z_1}{z_2} &= \frac{a_1 + i \cdot b_2}{a_2 + i \cdot b_2} = \frac{a_1 + i \cdot b_1}{a_2 + i \cdot b_2} \cdot \frac{a_2 - i \cdot b_2}{a_2 - i \cdot b_2} \\
&= \frac{a_1 \cdot a_2 - a_1 \cdot i \cdot b_2 + a_2 \cdot i \cdot b_1 - i \cdot b_1 \cdot i \cdot b_2}{a_2 \cdot a_2 - a_2 \cdot i \cdot b_2 + a_2 \cdot i \cdot b_2 - i \cdot b_2 \cdot i \cdot b_2} \\
&= \frac{a_1 \cdot a_2 - a_1 \cdot i \cdot b_2 + a_2 \cdot i \cdot b_1 + b_1 \cdot b_2}{{a_2}^2 + {b_2}^2} \\
&= \frac{a_1 \cdot a_2 + b_1 \cdot b_2 + i \cdot (a_2 \cdot b_1 - a_1 \cdot b_2)}{{a_2}^2 + {b_2}^2} = \frac{(a_1 \cdot a_2 + b_1 \cdot b_2, a_2 \cdot b_1 - a_1 \cdot b_2)}{({a_2}^2 + {b_2}^2)}
\end{aligned}$$

Beispiel. Für die Zahlen $z = (3, -4) \in \mathbb{C}$ und $w = (-5, 8) \in \mathbb{C}$ gilt:

$$\begin{aligned}
z + w &= (3 + (-5), (-4) + 8) = (-2, 4) \\
z - w &= (3 - (-5), (-4) - 8) = (8, -12) \\
z \cdot w &= (3 \cdot (-5) - (-4) \cdot 8, 3 \cdot 8 + (-4) \cdot (-5)) = (17, 44) \\
\frac{z}{w} &= \frac{(3, -4)}{(-5 + 8)} \cdot \frac{(-5, -8)}{(-5, -8)} = \frac{(3 \cdot (-5) - (-4) \cdot (-8), 3 \cdot (-8) + (-4) \cdot (-5))}{((-5)^2 + 8^2)} = \frac{(-47, -4)}{89}
\end{aligned}$$

4.2 Einführung der dualen Zahlen

Ähnlich wie bei den komplexen Zahlen wird bei den dualen Zahlen im Zuge der automatisierten Differenzierung ein Zahlentupel betrachtet, das aus einem Primärteil und einem Dualteil besteht. Der große Unterschied zu den komplexen Zahlen liegt darin, dass wir keine imaginäre, sondern eine duale Einheit besitzen. Die Rechenoperationen werden wir analog zu den komplexen Zahlen herleiten, wie später zu sehen ist. Zu Beginn definieren wir eine mathematische Struktur, die folgendermaßen aufgebaut ist:

Definition. (Duale Zahl) Die allgemeine Form einer dualen Zahl im Sinne der automatisierten Differenzierung lautet:

$$z = a + \epsilon \cdot b = (a, b) \quad \forall a, b \in \mathbb{R}$$

Für die Zahl ϵ definieren wir eine spezielle duale Einheit, für welche $\epsilon^2 = 0$ gilt.

4.2.1 Rechnen mit dualen Zahlen

Definition. (Addition und Subtraktion mit dualen Zahlen) Für die Addition und Subtraktion zweier dualer Zahlen $z_1 = (a_1, b_1)$ und $z_2 = (a_2, b_2)$ führen wir folgende Rechenvorschriften ein:

(i) $z_1 + z_2 = (a_1 + a_2, b_1 + b_2)$

(ii) $z_1 - z_2 = (a_1 - a_2, b_1 - b_2)$

Bei der Addition und Subtraktion fällt sofort die Ähnlichkeit zu den komplexen Zahlen auf. Diese Rechenoperationen werden analog ausgeführt. Der Unterschied zwischen der imaginären und der dualen Einheit fällt dabei nicht ins Gewicht, da sich dieser erst als Quadrat bemerkbar macht. Nichtsdestotrotz wird für das bessere Verständnis die Addition und Subtraktion mittels der allgemeinen Schreibweise $z = a + \epsilon \cdot b$ hergeleitet. Für $z_1 = a_1 + \epsilon \cdot b_1$ und $z_2 = a_2 + \epsilon \cdot b_2$ gilt:

$$\begin{aligned}
z_1 + z_2 &= (a_1 + \epsilon \cdot b_1) + (a_2 + \epsilon \cdot b_2) = a_1 + \epsilon \cdot b_1 + a_2 + \epsilon \cdot b_2 \\
&= a_1 + a_2 + \epsilon \cdot b_1 + \epsilon \cdot b_2 = a_1 + a_2 + \epsilon \cdot (b_1 + b_2) = (a_1 + a_2, b_1 + b_2) \\
z_1 - z_2 &= (a_1 + \epsilon \cdot b_1) - (a_2 + \epsilon \cdot b_2) = a_1 + \epsilon \cdot b_1 - a_2 - \epsilon \cdot b_2 \\
&= a_1 - a_2 + \epsilon \cdot b_1 - \epsilon \cdot b_2 = a_1 - a_2 + \epsilon \cdot (b_1 - b_2) = (a_1 - a_2, b_1 - b_2)
\end{aligned}$$

Definition. (Multiplikation mit dualen Zahlen) Für die Multiplikation zweier dualer Zahlen $z_1 = (a_1, b_1)$ und $z_2 = (a_2, b_2)$ definieren wir:

$$z_1 \cdot z_2 = (a_1 a_2, a_1 b_2 + a_2 b_1)$$

Um ein besseres Verständnis für die Rechenvorschrift aufzubauen, wird auch die Multiplikation mittels allgemeiner Form der dualen Zahlen hergeleitet. Für zwei duale Zahlen $z_1 = a_1 + \epsilon \cdot b_1$ und $z_2 = a_2 + \epsilon \cdot b_2$ gilt

$$\begin{aligned} z_1 \cdot z_2 = (a_1 + \epsilon \cdot b_1) \cdot (a_2 + \epsilon \cdot b_2) &= a_1 \cdot a_2 + \epsilon \cdot a_1 \cdot b_2 + \epsilon \cdot a_2 \cdot b_2 + \overbrace{\epsilon^2 \cdot b_1 \cdot b_2}^{e^2=0} \\ &= a_1 \cdot a_2 + \epsilon \cdot (a_1 \cdot b_2 + a_2 \cdot b_1) \\ &= (a_1 \cdot a_2, a_1 \cdot b_2 + a_2 \cdot b_1) \end{aligned}$$

Definition. (Division mit dualen Zahlen) Für die Division zweier dualer Zahlen $z_1 = (a_1, b_1)$ und $z_2 = (a_2, b_2)$ definieren wir:

$$\frac{z_1}{z_2} = (\frac{a_1}{a_2}, \frac{a_2 b_1 - a_1 b_2}{{a_2}^2})$$

Die Rechenvorschrift der Division wirkt zwar auf den ersten Blick weit hergeholt, wird aber genau gleich hergeleitet wie die Division der komplexen Zahlen. Damit die duale Einheit im Nenner wegfällt, erweitern wir Zähler und Nenner mit der konjugierten dualen Zahl des Nenners. Für die Herleitung mittels zweier dualer Zahlen z_1 und z_2 der allgemeinen Form $z = a + \epsilon \cdot b$ gilt:

$$\begin{aligned} \frac{z_1}{z_2} = \frac{a_1 + \epsilon \cdot b_1}{a_2 + \epsilon \cdot b_2} &= \frac{a_1 + \epsilon \cdot b_1}{a_2 + \epsilon \cdot b_2} \cdot \frac{a_2 - \epsilon \cdot b_2}{a_2 - \epsilon \cdot b_2} \\ &= \frac{a_1 \cdot a_2 - \epsilon \cdot a_1 \cdot b_2 + \epsilon \cdot a_2 \cdot b_1 - \epsilon^2 \cdot b_1 \cdot b_2}{{a_2}^2 - \epsilon^2 \cdot {b_2}^2} \\ &= \frac{a_1 \cdot a_2 - \epsilon \cdot a_1 \cdot b_2 + \epsilon \cdot a_2 \cdot b_1}{{a_2}^2} \\ &= \frac{a_1 \cdot a_2 + \epsilon \cdot (a_2 \cdot b_1 - a_1 \cdot b_2)}{{a_2}^2} \\ &= (\frac{a_1}{a_2}, \frac{a_2 \cdot b_1 - a_1 \cdot b_2}{{a_2}^2}) \end{aligned}$$

4.3 Übergang zur automatisierten Differenzierung

Um die Vorgehensweise der automatisierten Differenzierung mittels dualer Zahlen zu verstehen, schlagen wir eine Brücke von den dualen Zahlen zu den Funktionswerten und den damit verbundenen Ableitungen. In unserem Kontext setzt sich nun eine duale Zahl aus dem Funktionswert im Primärteil und aus der dazugehörigen Ableitung im dualen Teil zusammen. Dies stellt unsere wesentliche Grundlage für das automatisierte Differenzieren mittels dualer Zahlen dar. Somit erhalten wir für eine duale Zahl

$$z = (a_1, b_1) = (f(x), f'(x)) \tag{4.4}$$

Betrachten wir nun zwei unterschiedliche duale Zahlen, bestehend aus zwei Funktionen f und g, so können wir mithilfe der zuvor hergeleiteten Grundrechnungsarten sofort die wichtigsten Ableitungsregeln erkennen. Nachdem sich der Primärteil aus den Funktionswerten und der Dualteil aus den Ableitungen zusammensetzt, sind die Ableitungsregeln auch jeweils im Dualteil zu finden.

Definition. (Automatisierte Differenzierung) Mithilfe der dualen Zahlen lassen sich wesentliche Ableitungsregeln erklären und anwenden. Seien f und g zwei Funktionen sowie f' und g' die dazugehörigen Ableitungen. Für die dualen Zahlen $z_1 = (f(x), f'(x))$ und $z_2 = (g(x), g'(x))$ gilt:

(i) $z_1 + z_2 = (f(x), f'(x)) + (g(x), g'(x)) = (f(x) + g(x), \underbrace{f'(x) + g'(x)}_{(f+g)'=f'+g'})$

(ii) $z_1 - z_2 = (f(x), f'(x)) - (g(x), g'(x)) = (f(x) - g(x), \underbrace{f'(x) - g'(x)}_{(f-g)'=f'-g'})$

(iii) $z_1 \cdot z_2 = (f(x), f'(x)) \cdot (g(x), g'(x)) = (f(x) \cdot g(x), \underbrace{f'(x) \cdot g(x) + f(x) \cdot g'(x)}_{(f\cdot g)'=f'\cdot g+f\cdot g'})$

(iv) $\frac{z_1}{z_2} = \frac{(f(x), f'(x))}{(g(x), g'(x))} = (\frac{f(x)}{g(x)}, \underbrace{\frac{f'(x) \cdot g(x) - f(x) \cdot g'(x)}{g(x)^2}}_{(\frac{f}{g})'=(\frac{f'\cdot g-f\cdot g'}{g^2})})$

Wie an der vorherigen Definition gut zu erkennen ist, lassen sich mithilfe der dualen Zahlen und der damit verbundenen Grundrechnungsarten wesentliche Regeln der Differentialrechnung erklären und herleiten. Im Primärteil der neu gebildeten Zahlen sind die Funktionswerte der verknüpften Funktionen f und g zu erkennen. Im Dualteil finden wir die Ableitungen davon. Wie anhand des folgenden Satzes zu sehen ist, funktioniert auch

die Herleitung der Ableitungsregel von Polynomfunktionen mithilfe der automatisierten Differenzierung.

Satz 4.1. ***(Ableitung von Polynomfunktionen mithilfe der automatisierten Differenzierung)*** *Sei das Monom $f(x) = x^n$ gegeben. Betrachten wir die Funktion an der dualen Stelle $a + \epsilon$ erhalten wir*

$$\begin{aligned} f(a+\epsilon) = (a+\epsilon)^n &= \sum_{k=0}^{n} \binom{n}{k} a^{n-k}\epsilon^k \\ &= a^n + n \cdot a^{n-1} \cdot \epsilon \\ &= f(a) + \epsilon \cdot f'(a) \end{aligned}$$

Wenn $p(x)$ nun ein beliebiges Polynom darstellt, erhalten wir

$$p(a+\epsilon) = p(a) + \epsilon \cdot p'(a)$$

Die automatisierte Differenzierung stellt somit eine numerische Methode dar, mit der Ableitungen näherungsweise ermittelt werden können. Voraussetzung dafür ist jedoch, dass das ausführende Programm grundlegende Ableitungsregeln kennt bzw. diese im Vorhinein implementiert wurden. Mittels automatisierter Differenzierung können wir anschließend komplexere, verknüpfte Ableitungen bestimmen. Erste Anwendungen dazu sehen wir nun im folgenden Unterkapitel.

4.4 Anwendung der Methode

In diesem Unterkapitel wird das Prinzip der automatisierten Differenzierung anhand der Programmiersprache *Julia* erklärt und angewandt. Die folgenden Abbildungen zeigen dabei die wesentlichen Befehle und Codes, die für einen reibungslosen Ablauf der Methode notwendig sind.

Wie wir anhand von Abbildung 6 sehen können, müssen wir zuerst die grundlegende Struktur der dualen Zahlen festlegen. Den Primärteil nennen wir dabei „val“ für *Value* (Funktionswert) und den Sekundärteil nennen wir „der“ für *Derivative* (Ableitungswert). Anschließend werden dem Programm die grundlegenden Rechenvorschriften übermittelt und damit definiert. Diese leiten sich von der vorhergehenden Einführung (Kapitel 4.2) ab. Im nächsten Schritt muss noch beachtet werden, dass *Julia* nur Zahlentypen im selben Format bearbeiten kann. Mithilfe von „convert“ wird dementsprechend eine reelle Zahl in eine duale Zahl umgewandelt. Ein Beispiel dazu wäre $3 \in \mathbb{R} \rightarrow (3, 0) \in Dual$.

Den Befehl „promote rule“ verwenden wir, damit *Julia* sofort in duale Zahlen umrechnet, falls zum Beispiel eine duale Zahl mit einer reellen Zahl addiert werden soll.

Im nächsten Schritt zur Implementierung der automatisierten Differenzierung in *Julia* müssen die wesentlichen Ableitungen eingespielt werden, wie in Abbildung 7 zu sehen ist. Wie zuvor erwähnt, hilft uns die automatisierte Differenzierung, verknüpfte, komplexe Ableitungen numerisch zu lösen, wobei die grundlegenden Ableitungen jedoch bekannt sein müssen. Anschließend sehen wir erste Berechnungen: Das Programm addiert jeweils die richtigen Werte und ist auch in der Lage, eine duale Zahl mit einer reellen Zahl zu addieren. Zu guter Letzt definieren wir mittels „myAD(f,x::Real)“ die eigentliche Vorgehensweise der automatisierten Differenzierung. Anhand dieses Codes wird mittels dualer Zahlen der Ableitungswert an der reellen Stelle „x::Real“ berechnet.

In den Abbildungen 8 und 9 wird die automatisierte Differenzierung nun erstmals angewandt. Zuerst sehen wir am Beispiel der Sinus-Funktion den Vergleich der exakten Ableitung und der Ableitung mittels „myAD“. Hierbei fällt anhand der abgebildeten Graphen sofort auf, dass die numerische Methode keinerlei Abweichung zur exakten Ableitung aufweist. Daraus können wir schließen, dass die automatisierte Differenzierung eine sehr gute numerische Methode liefert. Abbildung 9 zeigt uns an zwei Funktionen, dass nicht nur die graphische Darstellung funktioniert, sondern auch ganz einfach Funktions- und Ableitungswerte sehr exakt berechnet werden können. Beim Nachrechnen mittels exakter Ableitung kommen wir auf dieselben Werte wie bei der Berechnung mithilfe dualer Zahlen.

[1]:
```
struct Dual <: Number
    val::Float64
    der::Float64
end
```

[3]:
```
X = Dual(2,0)
@show X.val;
@show X.der;
```

```
X.val = 2.0
X.der = 0.0
```

[5]:
```
import Base: +, -, *, /

+(x::Dual, y::Dual) = Dual(x.val+y.val, x.der+y.der)
*(x::Dual, y::Dual)= Dual(x.val*y.val, x.der*y.val + x.val*y.der)

-(x::Dual) = Dual(-x.val,-x.der)
-(x::Dual, y::Dual) = Dual(x.val-y.val, x.der-y.der)
/(x::Dual, y::Dual) = Dual(x.val/y.val, (y.val*x.der - x.val*y.der)/y.val^2)
```

[5]:
```
/ (generic function with 130 methods)
```

[6]:
```
import Base: convert
convert(::Type{Dual}, x::Real) = Dual(x,0)

import Base: promote_rule
promote_rule(::Type{Dual}, ::Type{<:Number}) = Dual
```

[6]:
```
promote_rule (generic function with 135 methods)
```

Abbildung 6: Automatisierte Differenzierung: Festlegen der grundlegenden Funktionen

```
[7]: import Base: sin, cos, exp, log, sqrt, abs
     sin(x::Dual) = Dual(sin(x.val), cos(x.val)*x.der)
     cos(x::Dual) = Dual(cos(x.val), -sin(x.val)*x.der)
     exp(x::Dual) = Dual(exp(x.val), exp(c.val)*x.der)
     log(x::Dual) = Dual(log(x.val), x.der/x.val)
     sqrt(x::Dual) = Dual(sqrt(x.val), 1/(2*sqrt(x.val))*x.der)
     abs(x::Dual) = Dual(abs(x.val), sign(x.val)*x.der)

[7]: abs (generic function with 12 methods)

[8]: import Base: <

     <(x::Dual, y::Dual) = x.val<y.val
     <(x::Dual, y::Real) = x.val<y
     <(x::Real, y::Dual) = x<y.val

[8]: < (generic function with 78 methods)

[9]: @show Dual(1.1,2)+Dual(2.2,0)
     @show Dual(1.1,2)+2.2

     Dual(1.1, 2) + Dual(2.2, 0) = Dual(3.3000000000000003, 2.0)
     Dual(1.1, 2) + 2.2 = Dual(3.3000000000000003, 2.0)
[9]: Dual(3.3000000000000003, 2.0)

[10]: myAD(f, x::Real) = f(Dual(x,1)).der

[10]: myAD (generic function with 1 method)
```

Abbildung 7: Automatisierte Differenzierung: Einspielen der grundlegenden Ableitungen und erste Berechnungen

[13]:
```julia
using Plots: plot, plot!

f(x) = sin(x)
∂f(x) = cos(x)

plot(∂f, 0.1, 10,
     linewidth=8,
     label="exakte Ableitung")

plot!(x -> myAD(f, x),
      0.1, 10,
      linewidth=2,
      label="automatisierte Differenzierung")
```

[13]:

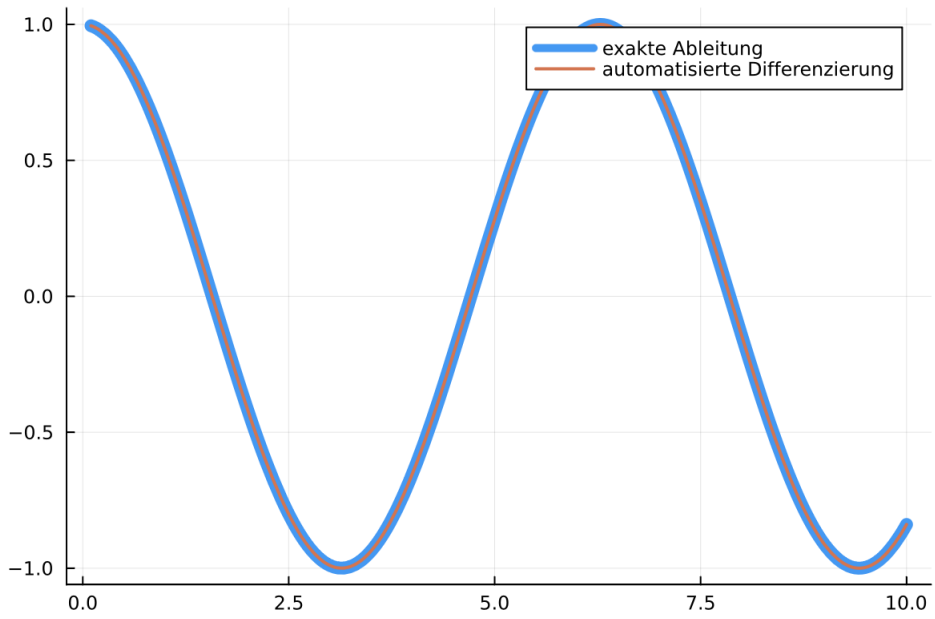


Abbildung 8: Automatisierte Differenzierung: Vergleich der exakten Ableitung und der näherungsweisen Ableitung anhand der Sinusfunktion

[14]:
```
# Funktion definieren
f(x) = x^2 + sin(x)

# Dualzahl erstellen
x_dual = Dual(2.0, 1.0)  # Ableitung nach x markieren

# Funktion auswerten
y = f(x_dual)

# Wert und Ableitung
println("Funktionswert: ", y.val)
println("Ableitung: ", y.der)
```

```
Funktionswert: 4.909297426825682
Ableitung: 3.5838531634528574
```

[15]:
```
# Funktion definieren
f(x) = x^2+3x+7

# Dualzahl erstellen
x_dual = Dual(-5.0, 1.0)  # Ableitung nach x markieren

# Funktion auswerten
y = f(x_dual)

# Wert und Ableitung
println("Funktionswert: ", y.val)
println("Ableitung: ", y.der)
```

```
Funktionswert: 17.0
Ableitung: -7.0
```

Abbildung 9: Automatisierte Differenzierung: näherungsweises Berechnen von Ableitungswerten

4.5 Vergleich der automatisierten Differenzierung und numerischer Differenzierung mittels Differenzenquotienten

Wie im Kapitel 2.3.1 erklärt basiert die Ableitung einer Funktion auf dem Differentialquotienten. Dabei wird mithilfe des Limes die durchschnittliche Änderungsrate zweier Funktionswerte in die momentane Änderungsrate an einer Stelle x überführt. Dieses Unterkapitel soll nun zeigen, warum sich die allgemeine Definition der Differenzierbarkeit nicht als Grundlage für eine zufriedenstellende numerische Ableitungsmethode eignet. Wie wir bereits wissen, ist der Differentialquotient definiert als

$$f'(x) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

Bei einem ausreichend kleinem h sollten wir annehmen können, dass die Ableitung auch damit numerisch gefunden werden kann. Wie anhand der folgenden vier Abbildungen 10, 11, 12 und 13 zu sehen ist, ist dies nicht immer der Fall. Mittels *Julia* wurde im vorherigen Code eine weitere Funktion „numdiff“ definiert, welche die Ableitung mittels des Grenzwerts des Differenzenquotienten darstellen soll. Betrachtet wird dafür die Funktion $f(x) = \cos(x)$, wobei die exakte Ableitung $f'(x) = -\sin(x)$ ist.

Zu Beginn wählen wir $h = \frac{1}{5}$, wobei die dargestellte Funktion hier klar von der exakten Ableitung abweicht. Wählen wir für h kleinere Werte, wäre zu erwarten, dass die Ableitung mit jeder Verkleinerung exakter würde. Bei der nächsten Abbildung scheint dies der Fall zu sein. Bei unserem gewählten $h = 10^{-5}$ ist die Darstellung der Ableitungsfunktion mittels Differentialquotienten identisch mit jener der automatisierten Differenzierung und der exakten Ableitung. Wird unser h im Folgenden aber noch kleiner, in unserem Fall zuerst $h = 10^{-15}$ und anschließend $h = 10^{-20}$, weichen die Ableitungen enorm von der exakten Darstellung ab. Im Fall von $h = 10^{-15}$ liegen wir teilweise noch in der Nähe des gewünschten Ergebnisses, im Fall von $h = 10^{-20}$ erkennen wir, dass hier die Ableitung mittels Differentialquotienten überhaupt nicht mehr funktioniert.

Diesen Abweichungen zufolge können wir schließen, dass die numerische Differenzierung mittels Differentialquotienten bei zu großem, aber auch sehr kleinem h keine gute Alternative zur automatisierten Differenzierung darstellt, da sehr schnell große Fehler auftreten. Dies basiert vor allem auf zwei klassischen Fehlerarten der numerischen Mathematik, die im Kapitel 3 genauer erläutert wurden. Erstens ist es schwierig, auf Anhieb ein passendes h zu definieren, bei dem die Darstellung mittels Differentialquotienten so-

fort ein gutes Ergebnis liefert. Wir sehen hier dementsprechend ein Paradebeispiel des Approximationsfehlers. Erst nach mehreren Versuchen und Vergleichen können wir uns sicher sein, mit dieser Methode ein passendes Ergebnis zu finden. Der Vorteil der automatisierten Differenzierung liegt hier eindeutig darin, dass keine Abhängigkeit eines selbstzuwählenden Parameters gegeben ist. Die automatisierte Differenzierung funktioniert unabhängig und liefert einwandfrei unsere gewünschten Resultate. Zweitens spielt auch die Darstellung der Zahlen beim Rechnen und Arbeiten mit Computersystemen eine wesentliche Rolle. Unser ausführendes Programm *Julia* arbeitet mit einer doppelten Maschinengenauigkeit ($\approx 10^{-16}$) und kann mit kleineren Zahlen nicht mehr rechnen. Auch mit Zahlen, die in der Nähe dieser Maschinengenauigkeit liegen, hat *Julia* in unserem Fall schon große Probleme, was wir gut bei unserem Beispiel mit $h = 10^{-15}$ sehen können. Sobald wir Zahlen wählen, die kleiner als $\approx 10^{-16}$ sind, scheitert das Computerprogramm sofort und liefert kein ernstzunehmendes Ergebnis.

[55…]
```julia
using Plots: plot, plot!

f(x) = cos(x)
∂f(x) = -sin(x)

numdiff(f, x; h=H) = (f(x + h) - f(x)) / h


# Plotbereich
xmin, xmax = 0.1, 10

plot(∂f, xmin, xmax,
     linewidth=8,
     label="exakte Ableitung")

plot!(x -> myAD(f, x),
      xmin, xmax,
      linewidth=2,
      label="automatisierte Differenzierung")
H=1/5

plot!(x -> numdiff(f, x; h=H),
      xmin, xmax,
      linewidth=2,
      linestyle=:dash,
      label="numerische Differenzierung mittels Differenzenquotienten, h=$H)")
```

[55…]

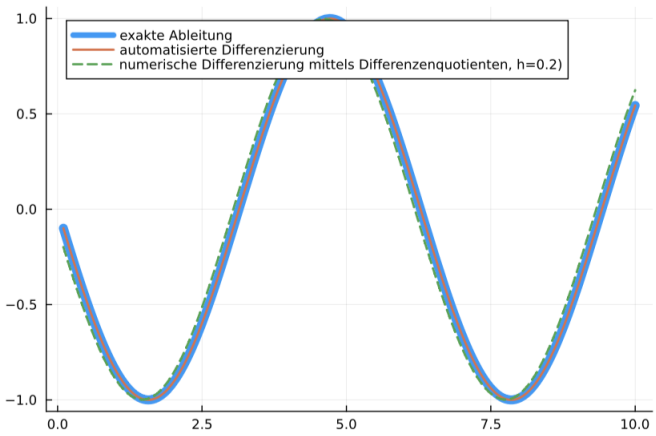


Abbildung 10: Automatisierte Differenzierung vs. numerische Differenzierung ($h = 1/5$)

[57]:
```
using Plots: plot, plot!

f(x) = cos(x)
∂f(x) = -sin(x)

numdiff(f, x; h=H) = (f(x + h) - f(x)) / h


# Plotbereich
xmin, xmax = 0.1, 10

plot(∂f, xmin, xmax,
     linewidth=8,
     label="exakte Ableitung")

plot!(x -> myAD(f, x),
      xmin, xmax,
      linewidth=2,
      label="automatisierte Differenzierung")
H=1e-5

plot!(x -> numdiff(f, x; h=H),
      xmin, xmax,
      linewidth=2,
      linestyle=:dash,
      label="numerische Differenzierung mittels Differenzenquotienten, h=$H)")
```

[57]:

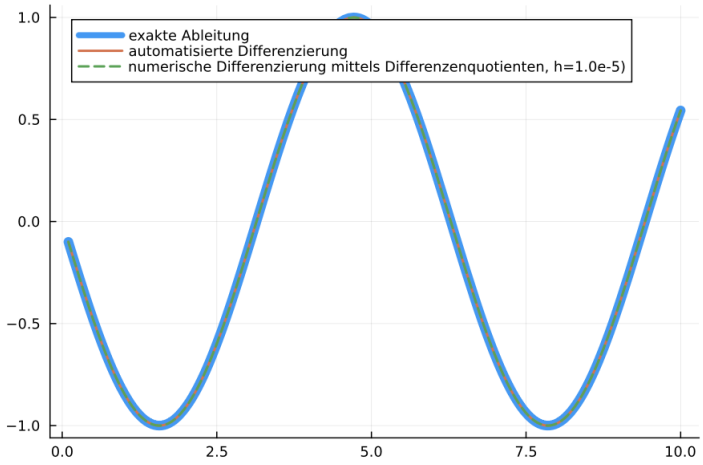


Abbildung 11: Automatisierte Differenzierung vs. numerische Differenzierung ($h = 10^{-5}$)

[26]:
```julia
using Plots: plot, plot!

f(x) = cos(x)
∂f(x) = -sin(x)

numdiff(f, x; h=1e-15) = (f(x + h) - f(x)) / h


# Plotbereich
xmin, xmax = 0.1, 10

plot(∂f, xmin, xmax,
     linewidth=8,
     label="exakte Ableitung")

plot!(x -> myAD(f, x),
      xmin, xmax,
      linewidth=2,
      label="automatisierte Differenzierung")

plot!(x -> numdiff(f, x; h=1e-15),
      xmin, xmax,
      linewidth=2,
      linestyle=:dash,
      label="numerische Differenzierung mittels Differenzenquotienten, h=1e-15)")
```

[26]:

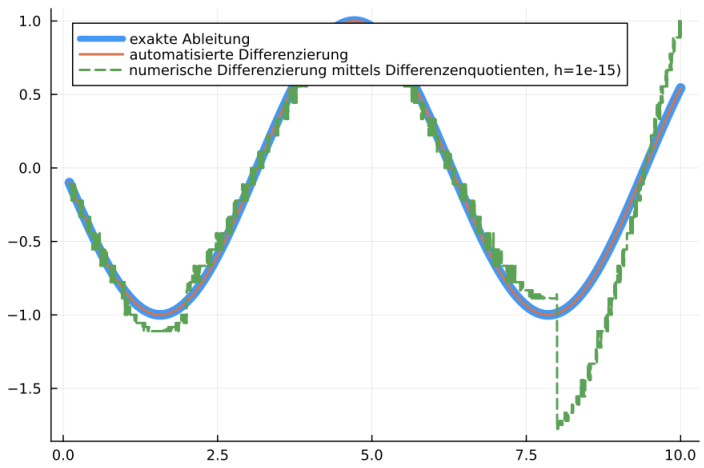


Abbildung 12: Automatisierte Differenzierung vs. numerische Differenzierung ($h = 10^{-15}$)

[28]:

```julia
using Plots: plot, plot!

f(x) = cos(x)
∂f(x) = -sin(x)

numdiff(f, x; h=1e-20) = (f(x + h) - f(x)) / h


# Plotbereich
xmin, xmax = 0.1, 10

plot(∂f, xmin, xmax,
     linewidth=8,
     label="exakte Ableitung")

plot!(x -> myAD(f, x),
      xmin, xmax,
      linewidth=2,
      label="automatisierte Differenzierung")

plot!(x -> numdiff(f, x; h=1e-20),
      xmin, xmax,
      linewidth=2,
      linestyle=:dash,
      label="numerische Differenzierung mittels Differenzenquotienten, h=1e-20)")
```

[28]:

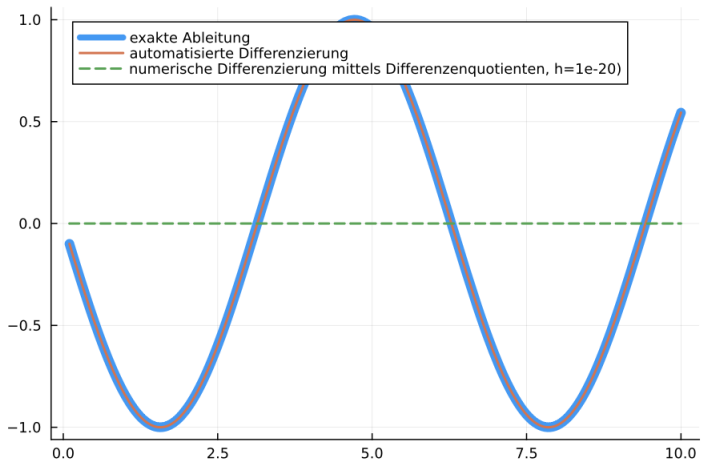


Abbildung 13: Automatisierte Differenzierung vs. numerische Differenzierung ($h = 10^{-20}$)

5 Nullstellenfindung mittels numerischer Methoden

Im folgenden Kapitel werden drei numerische Verfahren (Bisektionsverfahren, Sekantenverfahren und Newtonverfahren) vorgestellt, mit denen näherungsweise Nullstellen gefunden werden können. Diese Verfahren sind besonders dann von Bedeutung, wenn die Nullstellenberechnung der gefundenen Ableitungen analytisch nicht möglich oder zu aufwendig ist. Vor allem bei Polynomfunktionen, die einen Grad ≥ 4 aufweisen, oder bei transzendenten Funktionen wie zum Beispiel den Logarithmusfunktionen oder den Exponentialfunktionen ist es mit Näherungsverfahren vorteilhaft, Nullstellen zu finden. Dieses Kapitel basiert größtenteils auf Kapitel 3 in [3] und Kapitel 4 in [12].

5.1 Definitionen und Sätze

Bevor in den folgenden Unterkapiteln die Verfahren detailliert hergeleitet, erläutert und angewandt werden, ist es fundamental, grundlegende Sätze und Definitionen zu diskutieren, die für die Verfahren wichtige Rollen einnehmen.

Eine notwendige Bedingung für die Existenz von Nullstellen im abgeschlossenen Intervall $[a, b]$ ist, dass zwei Funktionswerte $f(a)$ und $f(b)$ unterschiedliche Vorzeichen aufweisen.

Satz 5.1 (z.B. Satz 3.1.2 in [3]). ***(Zwischenwertsatz)*** *Eine stetige Funktion $y = f(x)$ nimmt im abgeschlossenen Intervall $[a, b]$ jeden Wert zwischen $f(a)$ und $f(b)$ mindestens einmal an.*

Aufgrund des Zwischenwertsatzes folgt, dass im Fall von $f(a) \cdot f(b) < 0$ die Funktion $f(x)$ mindestens eine Nullstelle x^* besitzt.

Für das theoretische Verständnis der unterschiedlichen Iterationsverfahren zur Auffindung von Nullstellen sind auch die Konvergenztheorie und die damit verbundenen Fehlerabschätzungen von wesentlicher Bedeutung. Für diese theoretischen Betrachtungen wird ein Banach-Raum als Grundlage herangezogen.

Definition (z.B. Definition 4.1. in [12]). **(Banach-Raum)** Ein Banach-Raum B ist ein normierter, linearer Vektorraum über einen Zahlenkörper $\mathbb{K}$ ($\mathbb{R}$ oder $\mathbb{C}$), in dem jede Cauchy-Folge x_i von Elementen aus B gegen ein Element in B konvergiert. Dabei hat eine Norm in B die Eigenschaften

- $\|x\| \geq 0 \quad \forall x \in B$,

- $\|x\| = 0 \Leftrightarrow x = 0$,
- $\|yx\| = |y| \cdot \|x\| \quad \forall y \in \mathbb{K}, x \in B$,
- $\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in B$.

Um in der weiteren Folge das Konvergenzverhalten der Methoden zur Nullstellenfindung diskutieren zu können, wird folgende Definition benötigt.

Definition (z.B. Definition 4.2. in [12])**.** Eine Folge x_i mit dem Grenzwert s hat mindestens die Konvergenzordnung $p \geq 1$, wenn es eine von k unabhängige Konstante $K > 0$ gibt, so dass

$$\|x_{(i+1)} - s\| \leq K\|x_{(i)} - s\|^P \quad \forall k \geq 0 \tag{5.1}$$

mit $K < 1$, falls $p = 1$. K wird Konvergenzrate oder Fehlerkonstante genannt. Die Konvergenz heißt *linear*, falls $p = 1$, *quadratisch* , falls $p = 2$ und *kubisch*, falls $p = 3$.

Zwei weitere wichtige Definitionen für die Erklärung der Konvergenztheorie beziehen sich auf die *Lipschitz-Stetigkeit* und im weiteren Sinne auf den *Banachschen Fixpunktsatz*.

Definition (z.B. Definition 4.3. in [12])**. (Lipschitz-Stetigkeit)** Für eine abgeschlossene Teilmenge $A \subset B$ eines Banach-Raumes B sei eine Abbildung $F : A \to A$ von A in A gegeben. Die Abbildung F heißt *Lipschitz-stetig* auf A mit der Lipschitz-Konstanten $0 < L < \infty$, wenn gilt

$$\|F(x) - F(y)\| \leq L\|x - y\| \quad \forall x, y \in A. \tag{5.2}$$

Die Abbildung F nennt man *kontrahierend* auf A, falls die Lipschitz-Konstante in (5.2) $L < 1$ ist.

Definition (z.B. Definition 4.4. in [12])**. (Banachscher Fixpunktsatz)** Sei A eine abgeschlossene Teilmenge eines Banach-Raumes B und sei $F : A \to A$ eine kontrahierende Abbildung. Dann gilt:

(i) Die Abbildung F besitzt genau einen Fixpunkt $s \in A$.

(ii) Für jeden Startwert $x_0 \in A$ konvergiert die durch

$$x_{(k+1)} = F(x_{(k)}), \quad k = 0, 1, 2, \ldots, \tag{5.3}$$

definierte Folge gegen den Fixpunkt $s \in A$.

(iii)

$$\|x_{(k)} - s\| \leq L^k \|x_0 - s\|. \tag{5.4}$$

(iv) A priori Fehlerabschätzung:

$$\|x_{(k)} - s\| \leq \frac{L^k}{1-L} \|x_1 - x_0\|. \tag{5.5}$$

(v) A posteriori Fehlerabschätzung:

$$\|x_{(k)} - s\| \leq \frac{L}{1-L} \|x_k - x_{k-1}\|. \tag{5.6}$$

Beweis. Zu Beginn wird die Differenz zweier Folgenglieder abgeschätzt:

$$\begin{aligned} \|x_{(k+1)} - x_k\| &= \|F(x_{(k)}) - F(x_{(k-1)})\| \\ &\leq L\|x_{(k)} - x_{(k-1)}\| = L\|F(x_{(k-1)} - F(x_{(k-2)}\| \\ &\leq L^2 \|x_{(k-1)} - x_{(k-2)}\| \leq \dots \end{aligned}$$

beziehungsweise allgemein:

$$\|x_{(k+1)} - x_{(k)}\| \leq L^{k-l} \|x_{(l+1)} - x_{(l)}\| \qquad \text{für} \quad 0 \leq l \leq k. \tag{5.7}$$

Aufgrund der Voraussetzung $F : A \to A$ liegt mit $x_{(0)} \in A$ auch jedes $x_{(k)} \in A$. Dementsprechend wird nun gezeigt, dass die in (5.3) definierte Folge $x_{(k)}$ für jedes $x_{(0)} \in A$ konvergiert. Wegen (5.7) mit $l = 0$ bildet $x_{(k)}$ eine Cauchy-Folge, denn es gilt für beliebige $m \geq 1$ und $k \geq 1$

$$\begin{aligned} \|x_{(k+m)} - x_{(k)}\| &= \|x_{(k+m)} - x_{(k+m-1)} + x_{(k+m-1)} - + \dots - x_{(k)}\| \\ &\leq \sum_{\mu=k}^{k+m-1} \|x_{(\mu+1)} - x_{(\mu)}\| \\ &\leq L^k (L^{m-1} + L^{m-2} + \dots + L + 1) \|x_{(1)} - x_{(0)}\| \\ &= L^k \frac{1-L^m}{1-L} \|x_{(1)} - x_{(0)}\|. \end{aligned} \tag{5.8}$$

Aufgrund von $L < 1$ existiert somit zu jedem $\epsilon > 0$ ein $N \in \mathbb{N}^*$, sodass $\|x_{(k+m)} - x_{(k)} < \epsilon$ für $k \geq N$ und $m \geq 1$. Deshalb besitzt die Folge $x_{(k)}$ in der abgeschlossenen Teilmenge

A einen Grenzwert:

$$s := \lim_{k\to\infty} x_{(k)}, \qquad s \in A.$$

Wegen der Stetigkeit der Abbildung F gilt weiter

$$F(s) = F(\lim_{k\to\infty} x_{(k)}) = \lim_{k\to\infty} F(x_{(k)}) = \lim_{k\to\infty} x_{(k+1)} = s.$$

Damit ist nicht nur die Existenz eines Fixpunktes $s \in A$ bewiesen, sondern auch die Konvergenz der Folge $x_{(k)}$ gegen einen Fixpunkt.
Um die Eindeutigkeit des Fixpunktes nachzuweisen, wird indirekt vorgegangen. Es seien $s_1 \in A$ und $s_2 \in A$ zwei Fixpunkte der Abbildung F mit $\|s_1 - s_2\| > 0$. Da $s_1 = F(s_1)$ und $s_2 = F(s_2)$ gelten, führt dies wegen $\|s_1 - s_2\| = \|F(s_1) - F(s_2)\| \leq L\|s_1 - s_2\|$ zu dem Widerspruch $L \geq 1$. In (5.8) halten wir k fest und lassen $m \to \infty$ streben. Wegen $\lim_{m\to\infty} x_{(k+m)} = s$ und $L < 1$ ergibt sich die Fehlerabschätzung (5.5).

(5.4) ergibt sich aus:

$$\begin{aligned}\|x_{(k)} - s\| &= \|F(x_{(k-1)} - F(s)\| \leq L\|x_{(k-1)} - s\| \\ &= L\|F(x_{(k-2)}) - F(s)\| \leq L^2\|x_{(k-2)} - s\| = \dots \\ &\leq L^k\|x_{(0)} - s\|.\end{aligned}$$

(5.6) folgt zu guter Letzt aus (5.5) durch Umnummerierung $k \to 1$. Das bedeutet, es wird $x_{(k-1)}$ als Stratwert $x_{(0)}$ angenommen, dann wird $x_{(k)}$ zu $x_{(1)}$ und die beiden Abschätzungen werden gleich. □

Satz 5.2 (z.B. Satz 4.5. in [12])**.** *Es sei $n = 1$. In der abgeschlossenen Umgebung A des Fixpunktes s sei die Iterationsfunktion F $p+1$ mal stetig differenzierbar.*

Dann ist das Verfahren mindestens linear konvergent, falls

$$|F'(s)| < 1. \tag{5.9}$$

Das Verfahren ist mindestens p-ter Ordunng, $p \geq 2$, falls

$$F^{(k)}(s) = 0 \quad k = 1, 2, \dots, p-1. \tag{5.10}$$

Zu guter Letzt sollte noch geklärt werden, wann und unter welchen Bedingungen die Iterationsverfahren abgebrochen werden sollten. Das heißt, wir legen fest, nach welchen

Kriterien zu entscheiden ist, ob eine näherungsweise gefundene Lösung genau genug ist.

Satz 5.3. *Ein Iterationsverfahren muss durch eine zuvor festgelegte Abbruchbedingung gestoppt werden. Mögliche Vorschriften könnten sein:*

(i) Falls $n \geq N_0$, dann abbrechen.

(ii) Falls $|x^* - x_k| < \epsilon_1$, dann abbrechen.

(iii) Falls $|x_k - x_{(k-1)}| < \epsilon_2$, dann abbrechen.

5.2 Bisektionsverfahren

Die erste einfache Methode zum Finden von Nullstellen nichtlinearer Gleichungen ist das Bisektionsverfahren, das auch Intervallhalbierungsverfahren genannt wird. Dabei wird versucht, durch fortlaufendes Halbieren eines Intervalls die gesuchte Lösung immer weiter einzugrenzen. Sei eine stetige Funktion $f(x)$ und angenommen, es existiert ein Intervall $I = [a, b]$, sodass $f(a) \cdot f(b) < 0$ gilt. Nach dem Zwischenwertsatz hat die Funktion $f(x)$ im Intervall $I = [a, b]$ mindestens eine Lösung x^* mit $f(x^*) = 0$. Im ersten Schritt berechnet man den Mittelwert $\mu = \frac{(a+b)}{2}$ und anschließend den Funktionswert $f(\mu)$. Gilt für den Wert $f(\mu) \neq 0$, entscheidet das Vorzeichen, ob die gesuchte Lösung im Intervall $[a, \mu]$ oder im Intervall $[\mu, b]$ liegt. Nachdem überprüft wurde, in welchem halb so langen Teilintervall die gesuchte Nullstelle liegen muss, wird das Verfahren fortgesetzt und so lange wiederholt, bis die Lösung x^* in einem genügend kleinen Intervall liegt. Dabei wird nach jedem Iterationsschritt eine der zuvor besprochenen Abbruchbedingungen überprüft, damit die Genauigkeit der Lösung eingeschätzt werden kann. Wird nun $L_0 := b-a$ als die Länge des ersten Intervalls betrachtet, so bildet die Folge $L_k := (b-a)/2^k, k = 0, 1, 2, \ldots$ eine Nullfolge. Der Mittelpunkt x_k des Intervalls nach k Halbierungen wird mit der a-priori Fehlerabschätzung

$$\|x_k - s\| \leq \frac{b-a}{2^{k+1}}, \qquad k = 0, 1, 2, ... \tag{5.11}$$

als Näherung angesehen. Wie gut zu erkennen ist, nimmt hierbei die Fehlerschranke mit dem Quotienten $q = \frac{1}{2}$ ab, womit für die Konvergenzordnung $p = 1$ gilt. Das Verfahren konvergiert somit nicht schnell.

[vgl. Kapitel 4.2.1 in [12]]

Beispiel. Abbildung 14 illustriert anhand der Funktion $f(x) = sin(x-2)+ln(x)$ die Einfachheit des Intervallhalbierungsverfahren. Durch geeignetes Halbieren wird sich schrittwiese der Nullstelle genähert. Wie an $f(\mu)$ in der Abbildung zu sehen ist, scheint eine schnelle Konvergenz vorzuliegen, da sich mit nur einer Iteration der Nullstelle merklich genähert wird. Dieser Eindruck täuscht jedoch, wie zuvor erläutert wurde und mit folgendem Beispiel untermalt wird.

Beispiel. Betrachten wir die Funktion $f(z) = \sin(z)-3\cdot\log(z)$. Mithilfe der Programmiersprache *Julia* wurde ein Programm geschrieben, welches mittels des Bisektionsverfahrens näherungsweise die Nullstelle ermittelt.

Abbildung 15 zeigt dabei die Implementierung des Verfahrens in *Julia* mittels „function myBisection“ in Abhängigkeit einer definierten Funktion f und zweier Startwerte a und b. Die Toleranzgrenze liegt bei 10^{-5}. Dies bedeutet, dass das Verfahren so lange durchgeführt werden soll, bis ein Abstand zweier Folgewerte kleiner als 10^{-5} erreicht ist. Die maximale Anzahl an Wiederholungen wird mit „maxiter=50“ festgelegt. Das bedeutet, dass das Verfahren endet, falls nach 50 Wiederholungen das Abbruchkriterium nicht eingetreten ist. Danach startet der eigentliche Code des Verfahrens: Zuerst wird ein Vektor x definiert, der alle Iterationen der Intervallhalbierungen speichert, maximal 50. Die Halbierung des Intervalls wird durch $x[k] = (a+b)/2$ definiert. Anschließend wird mit einer „if-else Schleife“ gearbeitet. Falls der Betrag der Differenz zweier Werte kleiner als die Toleranzgrenze oder der Betrag des Funktionswerts kleiner als 2ϵ ist, wird das Verfahren beendet. Falls dies in einer Iteration nicht der Fall ist, wird mit einer weiteren Schleife entschieden, in welcher Hälfte die Nullstelle liegt und somit ein neues Intervall definiert. Dieser Ablauf wiederholt sich so lange, bis ein Abbruchkriterium eingetreten ist oder die maximale Anzahl an Iterationen erreicht wurde. Für den Fall, dass die festgelegte Anzahl an Wiederholungen nicht ausreicht, schreibt Julia „no convergence“. Die folgende Zeile zeigt uns den Funktionswert der mittels Bisektionverfahren gefunden Nullstelle. Da dieser mit $\approx -4,12 * 10^{-6}$ sehr klein ist, kann davon ausgegangen werden, dass das Verfahren erfolgreich war. Zeile 3 liefert uns als Abschluss noch den Vektor $x[k]$, der alle Ergebnisse der Iterationen zeigt. Wie zu erkennen ist, benötigt das Verfahren insgesamt 18 Iterationen, bis ein geeignetes Ergebnis vorliegt. Die näherungsweise Lösung beträgt $\approx 1,3878746$. Wie wir später noch sehen werden, kann die Nullstelle mit anderen Methoden noch schneller gefunden werden.

Das langsame Konvergenzverhalten wird anschließend noch mit Abbildung 16 verdeutlicht. Hier wird die Differenz der mittels Bisektionsverfahrens gefundenen Nullstelle und

der exakten Nullstelle der Funktion $f(z)$ als Graph dargestellt. Dabei wurde als Toleranzgrenze 10^{-8} gewählt und die Anzahl der maximalen Iterationen auf 4000 erhöht. Wie wir im Code auch sehen können, wurde die y-Achse logarithmisch skaliert, damit die Fehler gut ersichtlich werden. Erst nach über 25 Schritten ist die Differenz beider Werte kleiner als die Toleranzgrenze.

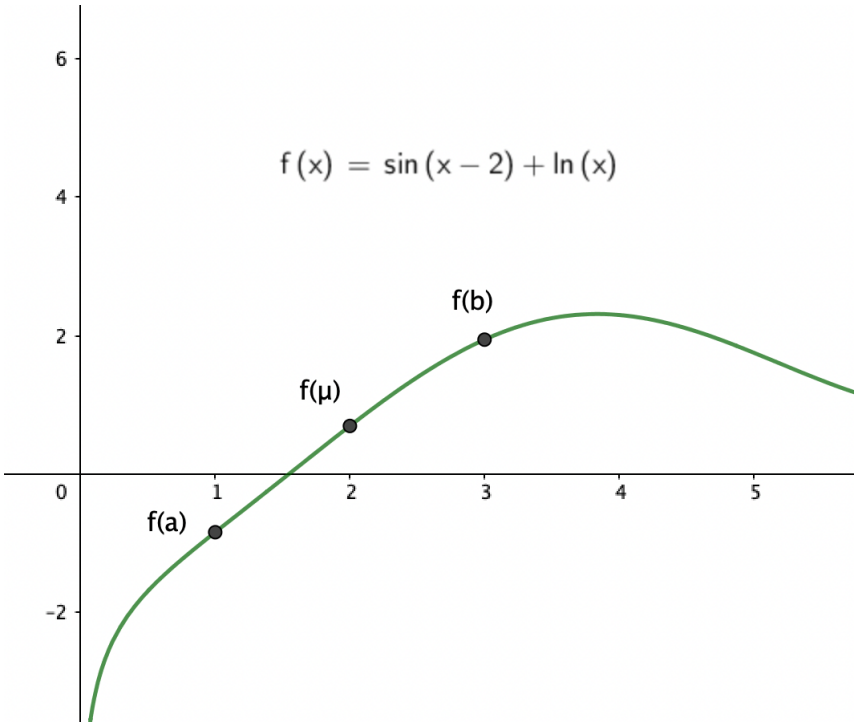


Abbildung 14: Bisektionsverfahren dargestellt mittels GeoGebra

```
[4]: function MyBisection(f,a,b; tol=1e-5, maxiter=50)
         x = Vector{Float64}(undef, maxiter)
         for k in 1: maxiter
             x[k] = (a+b)/2
             if abs(a-b)<tol || abs(f(x[k])) < 2eps()
                 return x[1:k]
             end
             if f(a)*f(x[k])<0
                 b = x[k]
             else
                 a = x[k]
             end
         end
         println("no convergence")
         return x
     end

[4]: MyBisection (generic function with 1 method)

[5]: f(z)=sin(z)-3*log(z)
     Nullstelle = MyBisection(f,1,2)[end]
     f(Nullstelle)

[5]: -4.124273168648607e-6

[6]: MyBisection(f,1,2)

[6]: 18-element Vector{Float64}:
      1.5
      1.25
      1.375
      1.4375
      1.40625
      1.390625
      1.3828125
      1.38671875
      1.388671875
      1.3876953125
      1.38818359375
      1.387939453125
      1.3878173828125
      1.38787841796875
      1.387847900390625
      1.3878631591796875
      1.3878707885742188
      1.3878746032714844
```

Abbildung 15: Implementierung des Bisektionsverfahrens in Julia und näherungsweises Berechnen der Nullstelle von $f(z) = \sin(z) - 3 \cdot \log(z)$

```
In [5]: using Plots:plot,plot!
        X = MyBisection(f,1,2;tol=1e-8, maxiter=4000)
        z= 1.3878725199641680403040239293315416330504330816469 2
        plot(abs.(X.-z),yaxis=:log, title = "Bisection", label = "Bisection")
```

Out[5]:

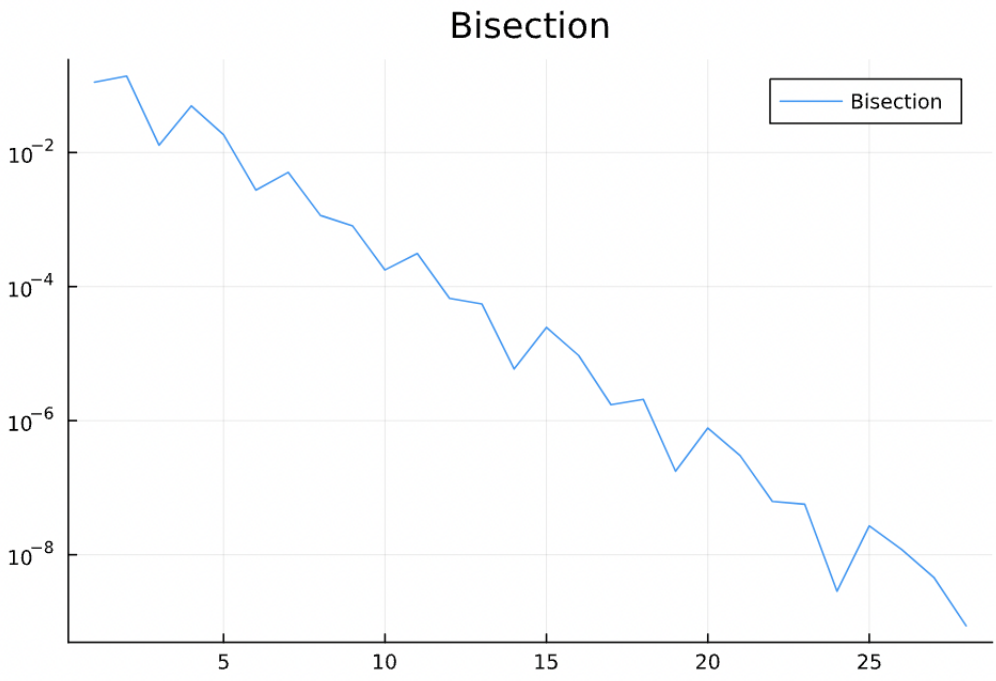


Abbildung 16: Konvergenz des Bisektionsverfahrens

5.3 Sekantenverfahren

Ein weiteres Verfahren zum Auffinden von Nullstellen nichtlinearer Gleichungen ist das Sekantenverfahren. Auch für diese Methode werden keine Ableitungswerte der Funktion benötigt, sondern es wird versucht, mittels Sekanten die Lösung immer weiter einzugrenzen. Um dieses Verfahren ausführen zu können, werden zwei Startwerte x_0 und x_1 herangezogen, die im Gegensatz zum Bisektionsverfahren die zu suchende Lösung nicht beinhalten müssen. Die allgemeinen Folgeglieder x_{k+1} werden anschließend als Nullstelle der Sekante der Funktion bestimmt, die durch die Punkte $(x_k, f(x_k))$ und $(x_{k-1}, f(x_{k-1}))$ verläuft. Damit dies leichter nachzuvollziehen ist, wird die Iterationsfolge der Sekantenmethode Schritt für Schritt hergeleitet:

Schritt 1: Zuerst benötigen wir die allgemeine Gleichung der Sekante durch die Punkte $(x_k, f(x_k))$ und $(x_{k-1}, f(x_{k-1}))$. Diese ist gegeben durch:

$$f(x_k) = k \cdot x_k + d \tag{5.12}$$

Schritt 2: Als nächstes wird die Steigung k der Sekante bestimmt:

$$k = \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}} \tag{5.13}$$

Schritt 3: Anschließend muss die zuvor bestimmte Steigung k in die allgemeine Punktsteigungsform $f(x_{k+1}) - f(x_k) = k \cdot (x_{k+1} - x_k)$ eingesetzt werden. Es folgt:

$$f(x_{k+1}) - f(x_k) = \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}} \cdot (x_{k+1} - x_k) \tag{5.14}$$

Aus 5.14 ergibt sich nun auch mit einem kleinen letzten Umformungsschritt die Sekantenfunktion:

$$f(x_{k+1}) = \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}} \cdot (x_{k+1} - x_k) + f(x_k) \tag{5.15}$$

Schritt 4: Damit wir aus der Sekantengleichung 5.15 nun die Iterationsfolge für das Sekantenverfahren bekommen, muss $f(x_{k+1}) = 0$ gesetzt werden, um den Schnittpunkt der Sekante mit der x-Achse zu berechnen. Anschließend muss die Gleichung nach x_{k+1} umgeformt werden, wonach sich

$$x_{k+1} = x_k - f(x_k) \cdot \frac{x_k - x_{k-1}}{f(x_k) - f(x_{k-1})} \tag{5.16}$$

als Iterationsfolge ergibt. Im Vergleich zum Bisektionsverfahren konvergiert das Sekantenverfahren schneller. Die Konvergenzordnung liegt bei $p = \frac{1}{2}(1+\sqrt{5}) = 1,618\ldots$. Dies entspricht dem Goldenen Schnitt.

[vgl. Kapitel 4.2.3 in [12]]

Beispiel. Wir betrachten wie zuvor bei der Bisektionsmethode die Funktion $f(x) = sin(x-2) + ln(x)$. Anhand der Abbildung 17 sieht man den Ablauf der Sekantenmethode. Nachdem zwei Startwerte x_1 und x_2 gewählt wurden, wird eine Sekante durch die Punkte $(x_1, f(x_1))$ und $(x_2, f(x_2))$ gelegt und diese anschließend mit der x-Achse geschnitten. Dieser Schnittpunkt markiert das nächsten Folgeglied x_3, welches sich schon der Nullstelle angenähert hat. Im weiteren Verlauf werden Sekanten durch die neu gefunden Folgegliedern gelegt, bis man nah genug an der Nullstelle der Funktion ist.

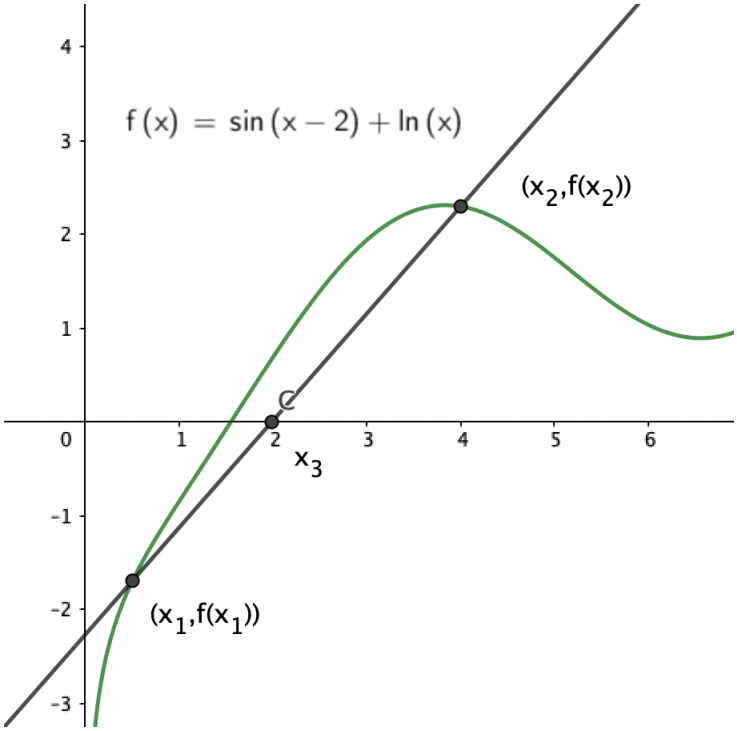


Abbildung 17: Sekantenverfahren dargestellt mittels GeoGebra

Beispiel. Auch für das Sekantenverfahren wurde ein auf *Julia* erstelltes Programm verwendet, um die Nullstelle numerisch zu bestimmen. Im Fokus steht dabei die Funktion $f(z) = sin(z) - 3 \cdot log(z)$.

Abbildung 18 zeigt hier die Implementierung des Sekantenverfahrens mittels „function mySecant“ in Abhängigkeit einer Funktion f und zweier Startwerte x_0 und x_1. Die Toleranz sowie die maximale Anzahl der Iterationen ist analog zum Bisektionsverfahren. Genauso wird wieder ein Vektor x definiert, der die einzelnen Iterationsergebnisse $x[k]$ speichert. Mit „for k in 2: maxiter-1“ werden die Folgeglieder $x[k+1]$ nach der darunterliegenden Iterationsvorschrift berechnet. Dies wird so oft durchgeführt, bis die Abbruchbedingung (Funktionswert des Näherungswertes kleiner als die festgelegte Toleranz) erfüllt ist. Wird nach maximal 50 Iterationen keine näherungsweise Lösung gefunden, bricht der Code ab und es wird „no convergence“ ausgegeben. In der nächsten Zeile sehen wir zuerst die Implementierung der Funktion $f(z)$ und anschließend die Berechnung des Funktionswertes der mittels Sekantenverfahrens gefundenen Lösung. Da das Ergebnis mit $\approx 8,43 * 10^{-6}$ sehr klein ist, können wir davon ausgehen, dass unsere gefundene Lösung eine sehr gute Annäherung an die Nullstelle darstellt. Zuletzt sehen wir in der Abbildung 18 noch die gespeicherten Näherungswerte, indem „mySecant“ mit den Werten 1 und 2 gestartet wird. Nach nur wenigen Schritten bricht das Verfahren ab, da die Abbruchbedingung erfüllt ist. Die näherungsweise Lösung beträgt $1,38787678$.

Abbildung 19 illustriert die Konvergenzgeschwindigkeit des Sekantenverfahrens mithilfe einer graphischen Darstellung. Dabei wurde die Toleranzgrenze auf 10^{-8} erhöht und die maximale Anzahl an Iterationen beträgt 4000. Wie beim Bisektionsverfahren zeigt der Graph die Differenz von gefundener und exakten Nullstelle. Wie anhand der logarithmischen Skalierung der y-Achse klar zu erkennen ist, konvergiert das Verfahren schnell und benötigt nur sechs Wiederholungen bis zum gewünschten Ergebnis.

```
[1]: function mySecant(f, x0, x1; tol=1e-5, maxiter=50)
         x = Vector{Float64}(undef,maxiter)
         x[1] = x0
         x[2] = x1
         for k in 2:maxiter-1
             x[k+1]=x[k]-f(x[k])*(x[k]-x[k-1])/(f(x[k])-f(x[k-1]))
             if abs(f(x[k+1]))<tol
                 return x[1:k+1]
             end
         end
         println("no convergence")
         return x
     end
```

```
[1]: mySecant (generic function with 1 method)
```

```
[2]: f(z)=sin(z)-3*log(z)
     Nullstelle = mySecant(f,1,2)[end]
     f(Nullstelle)
```

```
[2]: -8.43169430786439e-6
```

```
[3]: mySecant(f,1,2)
```

```
[3]: 5-element Vector{Float64}:
      1.0
      2.0
      1.4183061585435817
      1.3868706139820333
      1.3878767790942657
```

Abbildung 18: Implementierung des Sekantenverfahrens mittels Julia und näherungsweises Berechnen der Nullstelle von $f(z) = \sin(z) - 3 \cdot \log(z)$

[9]:
```
using Plots
X = mySecant(f, 1, 2; tol=1e-8, maxiter=4000)
z = 1.3878725199641680403040239293315416330

plot(abs.(X .- z),
     yscale = :log10,
     label = "Secant",
     title = "Secant")
```

[9]:

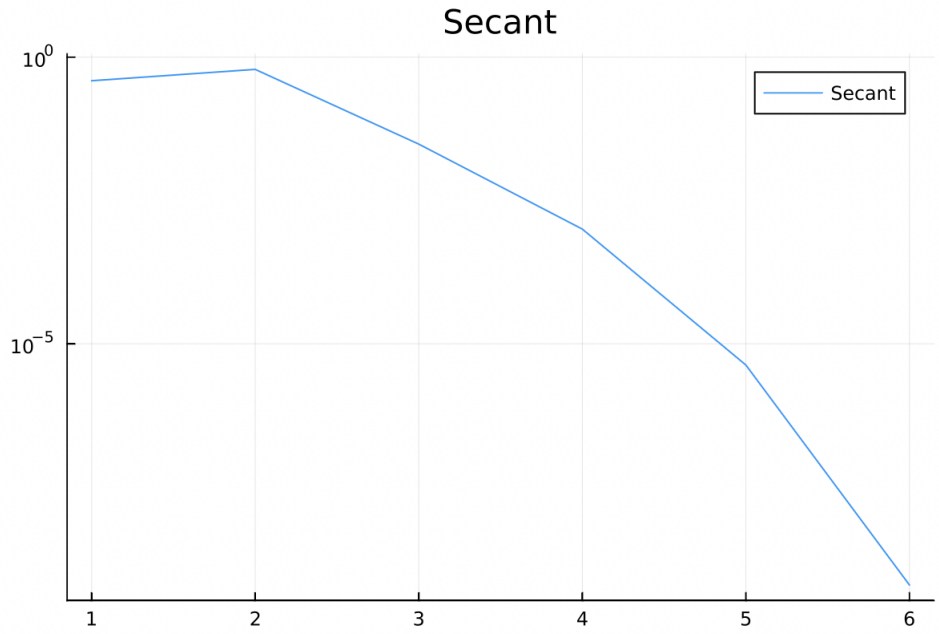


Abbildung 19: Konvergenz des Sekantenverfahrens

5.4 Newton-Verfahren

Das dritte und letzte Verfahren zum Auffinden von Nullstellen nichtlinearer Gleichungen, das im Zuge dieser Arbeit behandelt wird, ist das Newton-Verfahren. Dieses Verfahren ist die am weitesten verbreitete Methode, da sie im Falle eines guten Startwerts sehr schnell konvergiert.

Das Newton-Verfahren basiert auf Linearisierung. Das bedeutet, dass eine nichtlineare Funktion durch eine lineare Funktion angenähert wird und deren Nullstelle als Näherungswert dient. Im Falle des Newton-Verfahrens funktioniert das mittels Tangenten, die, beginnend beim Startwert, an die Funktion gelegt werden. Die Schnittpunkte dieser jeweiligen Tangenten mit der x-Achse markieren die nächsten Folgeglieder, sofern das Verfahren konvergiert. Um die Iterationsvorschrift besser nachvollziehen zu können, wird sie nun Schritt für Schritt hergeleitet.

Schritt 1: Zu Beginn ist es notwendig, die Steigung der Tangente t_0, die am Startwert x_0 an die gegebene Funktion $f(x)$ gelegt wird, zu definieren. Das Steigungsdreieck setzt sich dabei aus den Katheten $\Delta x := |x_0 - x_1|$ sowie $\Delta y := f(x_0)$ zusammen. Für die Steigung k der Tangente ergibt sich daraus:

$$k_{t_0} = \frac{\Delta y}{\Delta x} = \frac{f(x_0)}{|x_0 - x_1|} \qquad x_0 \neq x_1 \tag{5.17}$$

Schritt 2: Als nächstes wird die Steigung der Tangente mit der ersten Ableitung der Funktion gleichgesetzt:

$$k_{t_0} = f'(x_0) \tag{5.18}$$

Schritt 3: Aus 5.17 sowie 5.18 erschließt sich zu guter Letzt mit wenigen Umformungen die Iterationsvorschrift des Newtonverfahrens:

$$\begin{aligned}
\frac{f(x_0)}{|x_0 - x_1|} &= f'(x_0) \\
f(x_0) &= f'(x_0) \cdot |x_0 - x_1| \\
\frac{f(x_0)}{f'(x_0)} &= x_0 - x_1 \\
x_1 &= x_0 - \frac{f(x_0)}{f'(x_0)}
\end{aligned}$$

Wenn wir diese Folge noch verallgemeinern, erhalten wir die Iterationsvorschrift x_k.

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \tag{5.19}$$

Angenommen, der Startwert des Newton-Verfahrens wurde so gewählt, dass das Verfahren konvergiert, so liegt die Konvergenzordnung p bei mindestens 2, d.h. $p \geq 2$.

Satz 5.4 (z.B. Satz 4.6. in [12])**.** *Die Funktion $f(x)$ sei dreimal stetig differenzierbar in einem Intervall $I_1 = [a, b]$ mit $a < s < b$, und es sei $f'(s) = 0$, das heißt s sei eine einfache Nullstelle von $f(x)$. Dann existiert ein Intervall $I = [s - \delta, s + \delta]$ mit $\delta > 0$, für welches $F : I \to I$ eine konrahierende Abbildung ist. Für jeden Startwert $x_0 \in I$ ist die Folge $x_{(k)}$ des Newton-Verfahrens mindestens quadratisch konvergent. Das heißt es gilt $p \geq 2$.*

Beweis. Damit wir die kontrahierende Eigenschaft der Abbildung F in einer Umgebung von s, die nicht leer ist, zeigen können, beweisen wir, dass der Betrag der ersten Ableitung von F in einer Umgebung von s kleiner als 1 ist. Dabei erhalten wir für die erste Ableitung

$$F'(x) = 1 - \frac{f'(x)^2 - f(x)f''(x)}{f'(x)^2} = \frac{f(x)f''(x)}{f'(x)^2}. \tag{5.20}$$

Da $f(x)$ differenzierbar ist, folgt aufgrund von $f(s) = 0$, dass $F'(s) = 0$. Wegen der Stetigkeit muss in diesem Fall $\delta > 0$ existieren, womit

$$-1 < F'(x) < 1 \quad \forall x \in [s - \delta, s + \delta] = I \tag{5.21}$$

gilt. Somit sind für I die Eigenschaften des Banachschen Fixpunktsatzes erfüllt und die Konvergenz der Folge $x_{(k)}$ gegen s gezeigt.
Nach Satz 5.2 liegt die Konvergenzordnung des Newton-Verfahrens mindestens bei $p = 2$.
□

Beispiel. Betrachten wir nun die Funktion $f(x) = x^3 + x + 1$. Anhand der Abbildung 20 ist der Ablauf des Netwon-Verfahrens zu sehen. Nachdem der Startwert x_0 festgelegt wird, wird eine Tangente an den Punkt $(x_0, f(x_0))$ und die Funktion gelegt, welche mit der x-Achse geschnitten wird. Dieser Schnittpunkt markiert das nächste Folgenglied x_1. Diese Vorgehensweise wird so lange wiederholt, bis sich die Folgeglieder nah genug an die Nullstelle angenähert haben.

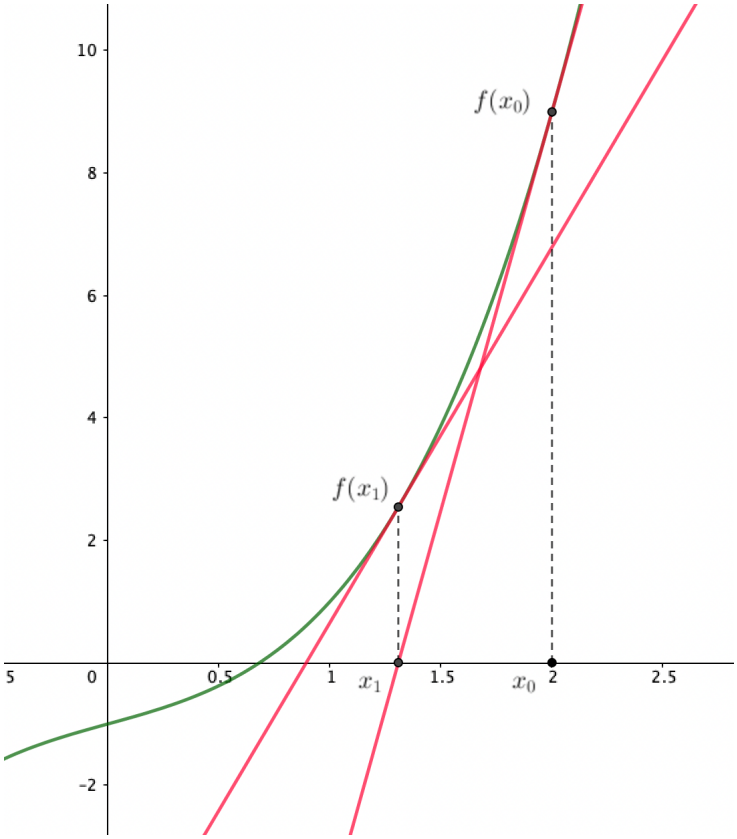


Abbildung 20: Newton-Verfahren dargestellt mittels GeoGebra

Beispiel. Auch beim Newton-Verfahren soll mittels einem *Julia*-Code gezeigt werden, wie schnell die Nullstelle der Funktion $f(z) = sin(z) - 3 \cdot log(z)$ zu finden ist.

Abbildung 21 zeigt wie bei den beiden Verfahren zuvor die Implementierung des Verfahrens in *Julia* mittels „function myNewton“. Das Verfahren benötigt dabei als grundlegende Informationen eine Funktion f, die dazugehörige Ableitungsfunktion und einen Startwert. Die Toleranzgrenze und die maximale Iteration verhalten sich analog zu den beiden zuvor besprochenen Verfahren. Anschließend wird ein Vektor x und der Startwert der Iteration $x[1]$ definiert. Mittels „for k in 1: maxiter-1“ und der anschließenden Iterationsvorschrift des Verfahrens werden im nächsten Schritt solange Näherungswerte berechnet, bis die Toleranzgrenze erreicht ist. Auch bei diesem Code gibt das Verfahren „no convergence“ aus, falls das Verfahren nach 50 Iterationen nicht die gewünschte Näherung erreicht hat und somit nicht konvergiert. In der folgenden Zeile des *Julia*-Codes sehen wir die erste Berechnung mittels Newton-Verfahren. Nachdem dem Verfahren die

Funktion, die Ableitung sowie der Startwert mitgeteilt wurde, wird der Funktionswert der näherungsweise gefundenen Lösung berechnet. Dieser ist mit $\approx 2,2 \cdot 10^{-16}$ verschwindend gering. Am Ende der Abbildung 21 sehen wir noch die gespeicherten Näherungswerte. Nach nur drei Wiederholungen bricht das Verfahren ab, da die Abbruchbedingung erfüllt ist. Die approximierte Lösung des Beispiels beträgt $x \approx 1,3878725$.

Abbildung 22 zeigt als Abschluss die Konvergenzgeschwindigkeit des Newton-Verfahrens mithilfe einer graphischen Darstellung. Wie zuvor beim Bisektions- und Sekantenverfahren wurde die Toleranzgrenze auf 10^{-8} gesetzt und die maximale Anzahl an Wiederholungen auf 4000 erhöht. Der Graph stellt nun die Differenz von näherungsweise gefundenener und exakter Nullstelle dar. Um die Fehler anschaulich erkennen zu können, wurde die y-Achse logarithmisch skaliert. Wie der untenstehenden Graphik entnommen werden kann, ist nach nur fünf Iterationen die gewünschte Lösung gefunden, das Verfahren konvergiert sehr schnell.

```
[17]: function myNewton(f,∂f, x0; tol=1e-5, maxiter=50)
          x= Vector{Float64}(undef, maxiter)
          x[1]=copy(x0)
          for k in 1: maxiter-1
              if abs(f(x[k]))<tol return x[1:k] end
              x[k+1]=x[k]-f(x[k])/∂f(x[k])
          end
          println("no convergence")
          return x
      end

[17]: myNewton (generic function with 1 method)

[18]: f(z)=sin(z)-3*log(z)
      Nullstelle = myNewton(f,z->cos(z)-3/z,1;tol=4eps())[end]
      f(Nullstelle)

[18]: 2.220446049250313e-16

[19]: myNewton(f,z->cos(z)-3/z,1)

[19]: 4-element Vector{Float64}:
       1.0
       1.3421034165358643
       1.3875328689956938
       1.3878725032205583
```

Abbildung 21: Implementierung des Newton-Verfahrens mittels Julia und näherungsweises Berechnen der Nullstelle von $f(z) = \sin(z) - 3 \cdot \log(z)$

[20]:
```
using Calculus:derivative
using Plots: plot, plot!

f(z)=sin(z)-3*log(z)
∂f=derivative(f)
z= 1.387872519964168040304023929331541633050433081646 92

X=myNewton(f,∂f,1,tol=1e-8,maxiter=4000)
plot(abs.(X.-z), yaxis=:log,title = "Newton", label = "Newton")
```

[20]:

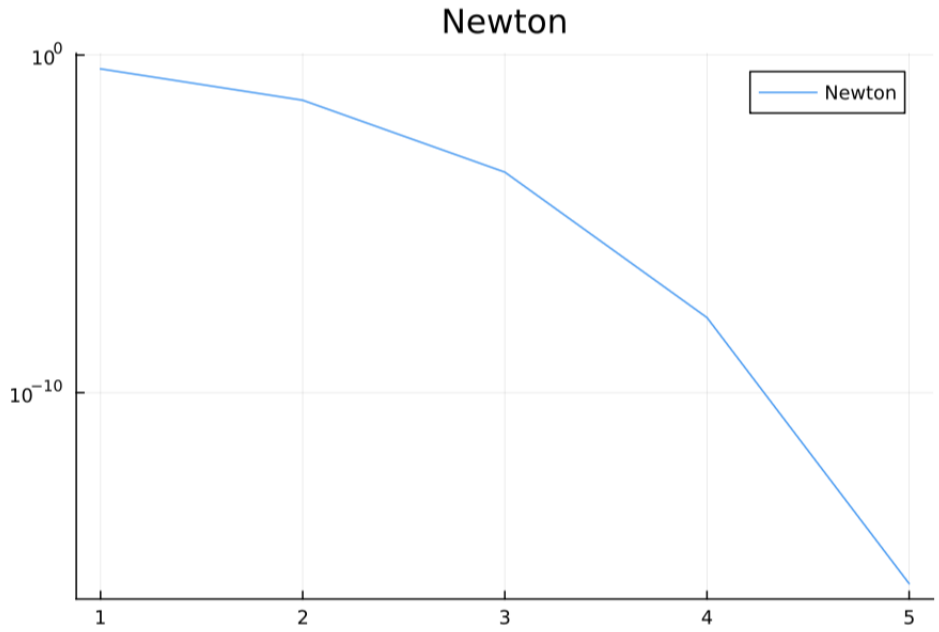


Abbildung 22: Konvergenz des Newton-Verfahrens

5.5 Vergleich der Methoden

Um ein besseres Verständnis für die unterschiedlichen Methoden aufbauen zu können, ist es sinnvoll, diese anhand von Beispielen direkt gegenüberzustellen. Um die Verfahren und dabei auch die Konvergenz dieser zu vergleichen, eignen sich das Berechnen einer Wurzel, die eine irrationale Zahl als Ergebnis liefert, sowie das näherungsweise Finden von π.

Beispiel. Wir definieren eine Funktion $f(x) = x^2 - 3$ und berechnen somit $\sqrt{3}$, indem wir mittels der zuvor besprochenen Verfahren die Nullstelle von $f(x)$ suchen. Weiters definieren wir eine Funktion $g(x) = sin(x)$ und finden mit denselben Methoden näherungsweise die irrationale Kreiszahl π. In den folgenden Abbildungen sieht man die Vorgehensweise zum Finden von $\sqrt{3}$ sowie von π und die Konvergenz der Verfahren. Die einzelnen Programme für die Implementierung der Verfahren werden dabei nicht mehr eingefügt, da diese bereits in den Unterkapiteln davor ersichtlich sind.

Abbildung 23 zeigt zuerst auf der linken Seite die numerische Berechnung von $\sqrt{3}$ mithilfe von allen drei Methoden. Dabei wird zuerst die Differenz der $\sqrt{3}$ und der Ergebnisse der einzelnen Verfahren, die auf die zuvor besprochenen Codes basieren, berechnet. Wie zuvor auch schon zu erkennen war, liefert uns das Bisektionsverfahren das abweichendste und das Newton-Verfahren das exakteste Ergebnis. Die Differenz aus $\sqrt{3}$ und dem mittels Newton-Verfahren gefundenen Ergebnis beträgt nur $\approx -2,44585 \cdot 10^{-9}$. Das Newton-Verfahren ist aber in diesem Beispiel nicht nur die genaueste, sondern auch die schnellste Methode. Benötigt das Bisektionsverfahren über 15 Iterationen, so ist das Newton-Verfahren nach nur vier Wiederholungen nah genug an der Lösung. Das Sekantenverfahren ist nur minimal langsamer und benötigt eine Iteration mehr. Genau das gleiche Bild vermittelt uns auch die die rechte Seite der Abbildung 23. Beim näherungsweisen Berechnen der irrationalen Kreiszahl π ist erneut das Newton-Verfahren am exaktesten und schnellsten, das Bisektionsverfahren am ungenauesten und langsamsten.

Die Abbildungen 24 und 25 zeigen die beiden Beispiele noch einmal in graphischer Darstellung, wobei die Plots aller drei Methoden übereinandergelegt werden. Die einzelnen Graphen stellen sowohl in Abbildung 24, als auch in Abbildung 25 die Differenz der exakten Werte von *pi* und $\sqrt{3}$ mit den näherungsweise gefundenen Lösungen dar. Bei allen drei Verfahren wurde die Toleranzgrenze auf 10^{-8} gesetzt und die maximale Iterationsanzahl auf 4000 erhöht, wie in den *Julia*-Codes ersichtlich ist. Beide Abbildungen liefern ein sehr ähnliches Bild. Auffallend ist nur, dass das Bisektionsverfahren in der Abbildung

23 am schnellsten startet, anschließend aber sofort von den beiden anderen Berechnungen eingeholt wird. Auch hier wird im Endeffekt wieder ersichtlich, dass das Newton Verfahren am schnellsten konvergiert, das Sekantenverfahren nur wenige Schritte mehr benötigt und das Bisektionsverfahren mehr als fünfmal so viele Wiederholungen aufweist.

Zusammenfassend lässt sich sagen, dass das Bisektionsverfahren bezüglich Effizienz und Konvergenzgeschwindigkeit im Vergleich zu den anderen beiden Methoden weit zurückfällt. Der Vorteil liegt aber dafür sicherlich in der Einfachheit der Handhabung. Es basiert weder auf komplizierten Iterationsvorschriften, noch wird eine Ableitung benötigt. Aufgrund der mangelnden Effizienz, wird das Bisektionsverfahren dennoch nicht als ideale Methode zum Auffinden der Nullstellen für den weiteren Verlauf der Arbeit dienen. Ob im Vergleich dazu das Sekantenverfahren oder das Newton-Verfahren die beste Methode ist, wird sich erst im folgenden Kapitel endgültig zeigen. Stand jetzt lässt sich aber vermuten, dass die für das Newton-Verfahren benötigte Ableitung einen großen Nachteil darstellen wird. Auch wenn dieses Verfahren schneller als das Sekantenverfahren konvergiert, wird das Auffinden der Ableitung für die Iterationsorschrift eine numerische Herausforderung darstellen. Im Vergleich dazu konvergiert das Sekantenverfahren nur geringfügig langsamer, ist jedoch aufgrund des Verzichts auf die Berechnung von Ableitungen einfacher anzuwenden. Es bildet damit einen sinnvollen Kompromiss zwischen dem Bisektionsverfahren und dem Newton-Verfahren in Bezug auf Einfachheit und Konvergenzgeschwindigkeit.

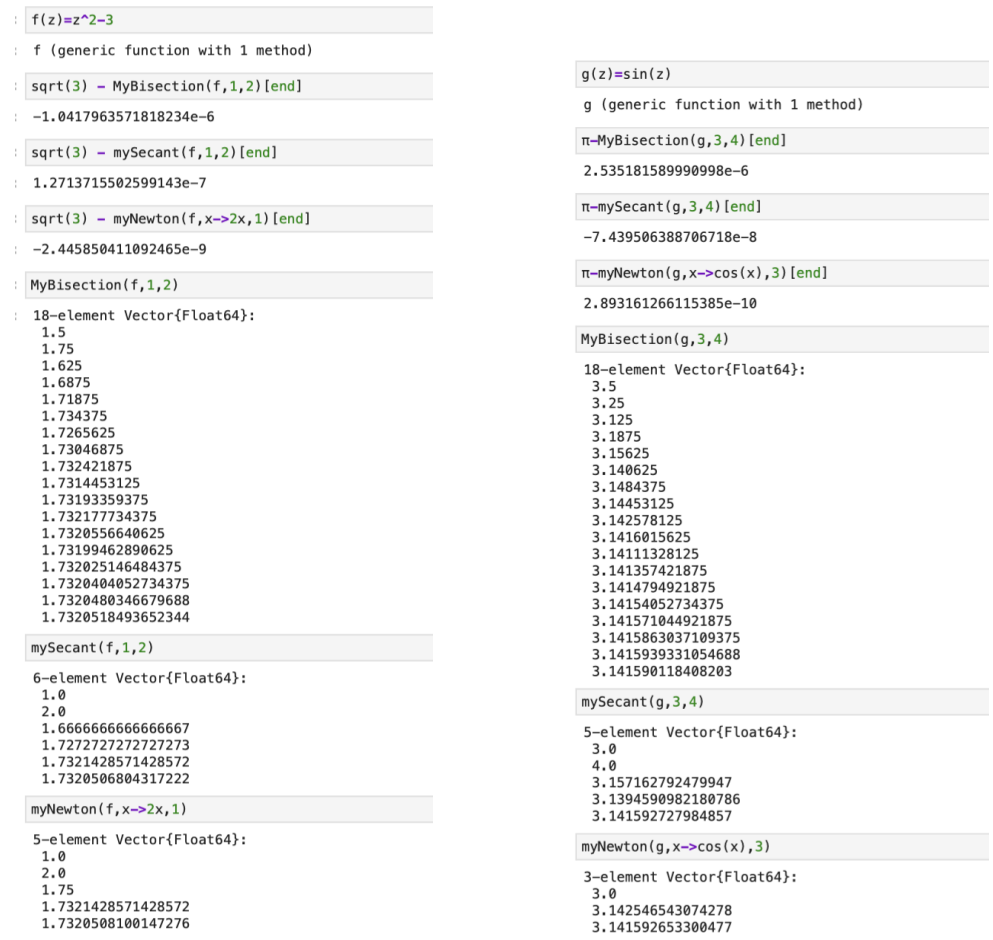


Abbildung 23: Finden von $\sqrt{3}$ sowie π mittels Bisektions-, Sekanten-, und Newton-Verfahrens in Julia

```
using Plots:plot,plot!
using Calculus:derivative
f(x)=x^2-3
z = sqrt(3)
X = MyBisection(f, 1, 2; tol = 1e-8, maxiter = 4000)
Y = mySecant(f, 1, 2; tol = 1e-8, maxiter = 4000)
Z = myNewton(f, derivative(f),1; tol = 1e-8, maxiter = 4000)
plot(replace!(abs.(X.-z),0.0 => 1e-16), yaxis=:log, title = "Berechnen der Wurzel", label = "Bisektionsverfahren")
plot!(replace!(abs.(Y.-z),0.0=> 1e-16), yaxis=:log, label = "Sekantenverfahren")
plot!(replace!(abs.(Z.-z),0.0=> 1e-16), yaxis=:log, label = "Newton-Verfahren")
```

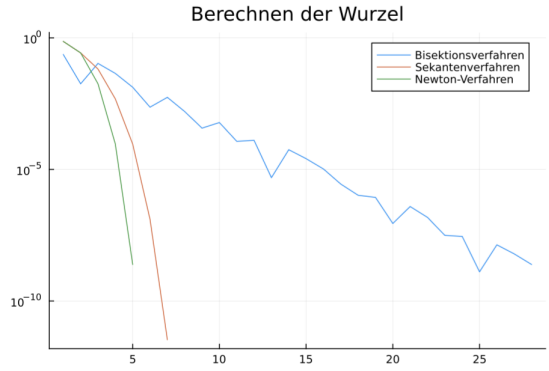


Abbildung 24: Graphischer Vergleich des Bisektionsverfahrens, Sekantenverfahrens und Newton-Verfahrens in Bezug auf die Genauigkeit bei der Berechnung von $\sqrt{3}$

```
using Plots:plot,plot!
using Calculus:derivative
g(x)=sin(x)
z = π
X = MyBisection(g, 3, 4; tol = 1e-8, maxiter = 4000)
Y = mySecant(g, 3, 4; tol = 1e-8, maxiter = 4000)
Z = myNewton(g, derivative(g), 3.; tol = 1e-8, maxiter = 4000)
plot(replace!(abs.(X.-z),0.0 => 1e-16), yaxis=:log, title = "Berechnen von π", label = "Bisektionsverfahren")
plot!(replace!(abs.(Y.-z),0.0=> 1e-16), yaxis=:log, label = "Sekantenverfahren")
plot!(replace!(abs.(Z.-z),0.0=> 1e-16), yaxis=:log, label = "Newton-Verfahren")
```

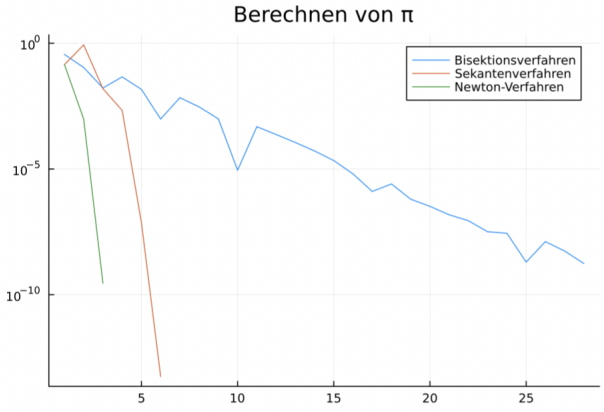


Abbildung 25: Graphischer Vergleich des Bisektionsverfahrens, Sekantenverfahrens und Newton-Verfahrens in Bezug auf die Genauigkeit bei der Berechnung von π

6 Numerisches Lösen von Extremwertaufgaben

Dieses Kapitel soll nun die in den vorhergehenden Kapiteln besprochenen numerischen Methoden zur Lösung von Extremwertaufgaben zusammenführen und anhand von Beispielen die Anwendung dieser thematisieren. Wenn wir uns an den Aufbau von Extremwertaufgaben zurückerinnern, müssen zunächst die grundlegenden Informationen eines Problems herausgearbeitet werden, um damit analytisch die Haupt- und Nebenbedingungen für die Extremwertaufgabe zu finden. Anstelle des analytischen Differenzierens versuchen wir, mithilfe von dualen Zahlen die Ableitungswerte zu berechnen. Beim anschließenden Suchen der maximalen bzw. minimalen Stelle anhand der Nullstellen der Ableitung werden wir in diesem Kapitel zuerst auf das Sekantenverfahren und anschließend auf das Newton-Verfahren zurückgreifen. Obwohl das Newton-Verfahren effizienter ist und schneller konvergiert, bietet das Sekantenverfahren den Vorteil, dass innerhalb des Verfahrens keine Ableitung benötigt wird.

6.1 Anwendung mittels dualer Zahlen und Sekantenverfahren

Als wesentliches Beispiel, anhand dessen die numerischen Methoden angewandt werden, greifen wir auf die letzte Aufgabe aus Kapitel 2 zurück.

Beispiel (Aufgabe 19 aus [1]). Für den Bau eines Tores stehen für die senkrechten Begrenzungen zwei Pfosten mit $2m$ Länge und für die oben aufgesetzen schrägen Begrenzungen zwei Dachpfosten mit je $4m$ Länge zur Verfügung. Wie muss der Öffnungswinkel der Dachpfosten gewählt werden, wenn das Tor eine möglichst große Durchlassfläche (Querschnittsfläche) haben soll?

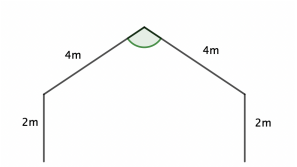


Abbildung 26: Skizze der Extremwertaufgabe [Aufgabe 19 aus [1]]

Numerische Lösung:
Da dieses Beispiel wie bereits erwähnt in Kapitel 2 thematisiert wurde, wird auf die Erklärung der Haupt- und Nebenbedingung verzichtet. Wir übernehmen diese und erhalten

HB:

$$A(b, h_b) = 2 \cdot b + \frac{b \cdot h_b}{2},$$

NB:

$$\sin(\frac{\alpha}{2}) = \frac{\frac{b}{2}}{4} \Rightarrow b = 8 \cdot \sin(\frac{\alpha}{2}) \quad (0° \leq \alpha \leq 180°),$$
$$\cos(\frac{\alpha}{2}) = \frac{h_b}{4} \Rightarrow h_b = 4 \cdot \cos(\frac{\alpha}{2}) \quad (0° \leq \alpha \leq 180°).$$

Setzen wir die beiden Nebenbedingungen nun in die Hauptbedingung ein, erhalten wir eine Funktion in Abhängigkeit des Öffnungswinkels α. Auch wenn die Ableitung für Schüler:innen analytisch auffindbar wäre, wird sie im Zuge der numerischen Lösung mittels Sekantenverfahren nicht benötigt.

$$A(\alpha) = 16 \cdot (\sin(\frac{\alpha}{2}) + \sin(\frac{\alpha}{2}) \cdot \cos(\frac{\alpha}{2}))$$

Da wir nun den Maximalwert der Funktion, dementsprechend also die Nullstelle der Ableitungsfunktion, finden müssen, greifen wir nun auf unsere numerischen Methoden zurück. Die folgenden Abbildungen zeigen dabei, wie schnell und einfach ohne analytische Berechnung der Ableitung und anschließendes kompliziertes (bzw. kaum mögliches) Finden der Nullstelle, die Lösung gefunden werden kann. Voraussetzung für das Berechnen der Lösung mittels automatisierter Differenzierung ist die Implementierung der Basisinformationen. Diese wurden in Kapitel 4 detailliert beschrieben und werden dementsprechend im Zuge der Anwendung nicht weiter erklärt.

Abbildung 27 zeigt den Code des Sekantenverfahrens, ähnlich wie im Kapitel 5 bereits erklärt. Der große Unterschied liegt hier nun aber, dass damit nicht die Nullstelle einer Funktion per se gefunden wird, sondern direkt die Nullstelle der dazugehörigen Ableitungsfunktion. Anstatt die Funktionswerte einer Funktion aufzurufen, greift hier das Verfahren auf die numerisch gefundenen Ableitungswerte von „myAD(f,x[k])“ zurück. Damit sparen wir uns das Berechnen der Ableitung und es wird direkt numerisch die gesuchte Nullstelle gefunden.

Abbildung 28 zeigt uns nun zuerst die Darstellung der Zielfunktion $A(x)$. Zur Vereinfachung der *Julia*-Codes verwenden wir dafür die Variable x anstatt der abhängigen Variable α (Öffnungswinkel der Dachpfosten). Mittels „using Plots: plot, plot!“ wird die

Funktion abgebildet, was für das Finden der benötigten Startwerte eine große Erleichterung darstellt. Näherungsweise können wir mithilfe des Graphens eine erste Schätzung für den Maximalwert ablesen.

$$\begin{aligned} x &\approx 2 \\ \alpha = \frac{x}{\pi} \cdot 180 &\approx \frac{2}{\pi} \cdot 180 \\ \alpha &\approx 114,6^\circ \end{aligned}$$

Zu guter Letzt können wir nun die tatsächliche Genauigkeit unserer Schätzung überprüfen, indem wir das Sekantenverfahren anwenden. Wie in Abbildung 28 zu sehen wählen wir als Startwerte $x_0 = 2$ und $x_1 = 3$. Nach nur wenigen Schritten wird das Verfahren abgebrochen, da eine numerische Lösung gefunden wurde, die um weniger als 10^{-5} von der Nullstelle abweicht. Mit dieser numerischen Lösung erhalten wir folgenden Öffnungswinkel, bei dem die Querschnittsfläche des Tores maximal ist:

$$\begin{aligned} x &\approx 2,094395 \\ \alpha = \frac{x}{\pi} \cdot 180 &\approx \frac{2,094395}{\pi} \cdot 180 \\ \alpha &\approx 120^\circ \end{aligned}$$

Anhand der graphischen Darstellung der Zielfunktion $A(x)$ können wir ausschließen, dass die Randwerte $x = 0$ und $x = \pi$ als maximale Lösungen infrage kommen. Dementsprechend muss unsere numerisch gesuchte Lösung bei $x \approx 2,094395 \approx 120^\circ$ liegen. Wenn wir diese in unsere Zielfunkton $A(\alpha)$ einsetzen, so erhalten wir den maximalen Flächeninhalt des Tores:

$$\begin{aligned} A(\alpha) &= 16 \cdot (\sin(\frac{\alpha}{2}) + \sin(\frac{\alpha}{2}) \cdot \cos(\frac{\alpha}{2})) \\ A(120^\circ) &\approx 20,7861m^2 \end{aligned}$$

```
function mySecant(f, x0, x1; tol=1e-5, maxiter=50)
    x = Vector{Float64}(undef,maxiter)
    x[1] = x0
    x[2] = x1
    for k in 2:maxiter-1
        x[k+1]=x[k]-myAD(f,x[k])*(x[k]-x[k-1])/(myAD(f,x[k])-myAD(f,x[k-1]))
        if abs(myAD(f,x[k+1]))<tol
            return x[1:k+1]
        end
    end
    println("no convergence")
    return x
end
```

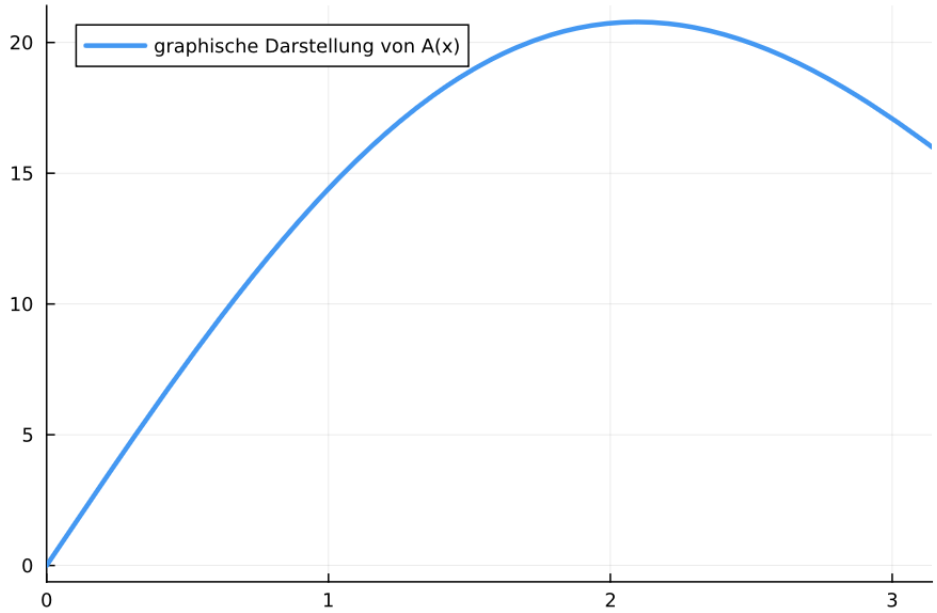
```
mySecant (generic function with 1 method)
```

Abbildung 27: Implementieren des Sekantenverfahrens mithilfe der automatisierten Differenzierung

```
A(x)=16*(sin(x/2)+sin(x/2)*cos(x/2))
```

```
A (generic function with 1 method)
```

```
using Plots: plot, plot!
plot(A, xlimits=[0, π],
     linewidth=3,
     label="graphische Darstellung von A(x)")
```




```
mySecant(A,2,3)
```

```
6-element Vector{Float64}:
 2.0
 3.0
 2.11899002154707
 2.087386106037034
 2.0944204276146507
 2.0943951278331743
```

Abbildung 28: Darstellung der Zielfunktion $A(x)$ und anschließendes Berechnen der Nullstelle von $A'(x)$ mithilfe des Sekantenverfahrens

Damit die numerischen Methoden nicht nur anhand eines Beispiels angewandt werden, betrachten wir nun dasselbe Beispiel noch einmal, wobei die Aufgabenstellung geändert wird. Ziel der neuen Aufgabenstellung ist es, eine neue Zielfunktion zu erhalten, die nach dem Finden der Extremstelle den gleichen maximalen Flächeninhalt der Querschnittsfläche aufweist.

Beispiel. Für den Bau eines Tores stehen für die senkrechten Begrenzungen zwei Pfosten mit $2m$ Länge und für die oben aufgesetzten schrägen Begrenzungen zwei Dachpfosten mit je $4m$ Länge zur Verfügung. Wie groß muss der Abstand zwischen den beiden senkrechten Pfosten gewählt werden, damit die Durchlassfläche des Tores (Querschnittsfläche) möglichst groß wird?

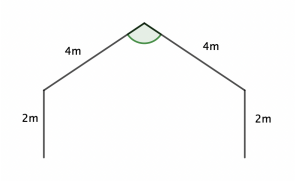


Abbildung 29: Skizze der Extremwertaufgabe [vgl. Aufgabe 19 aus [1]]

Numerische Lösung:

Im Vergleich zum Beispiel davor ist die Hauptbedingung zwar ident, die Nebenbedingung unterscheidet sich hingegen eindeutig. Wir werden somit eine völlig andere Funktionsart erhalten, mit welcher wir ausprobieren können, ob unsere numerischen Methoden erneut zum gewünschten Ziel führen.
HB:

$$A(b, h_b) = 2 \cdot b + \frac{b \cdot h_b}{2}$$

NB:

$$h_b{}^2 + (\frac{b}{2})^2 = 4^2 \Rightarrow h_b = \sqrt{16 - \frac{b^2}{4}} \quad (0 < b < 8)$$

Setzen wir die gefundene Nebenbedingung in die Hauptbedingung ein, erhalten wir unsere Zielfunktion in Abhängigkeit der Entfernung der beiden senkrechten Pfosten b.

$$A(b) = 2 \cdot b + \frac{b}{2} \cdot \sqrt{16 - \frac{b^2}{4}}$$

Auch wenn die Ableitungsfunktion hier analytisch mit den bekannten Regeln auffindbar wäre, ist sie aufgrund der Implementierung der automatisierten Differenzierung innerhalb des Sekantenverfahrens nicht notwendig. Das Beispiel kann dementsprechend rein numerisch gelöst werden.

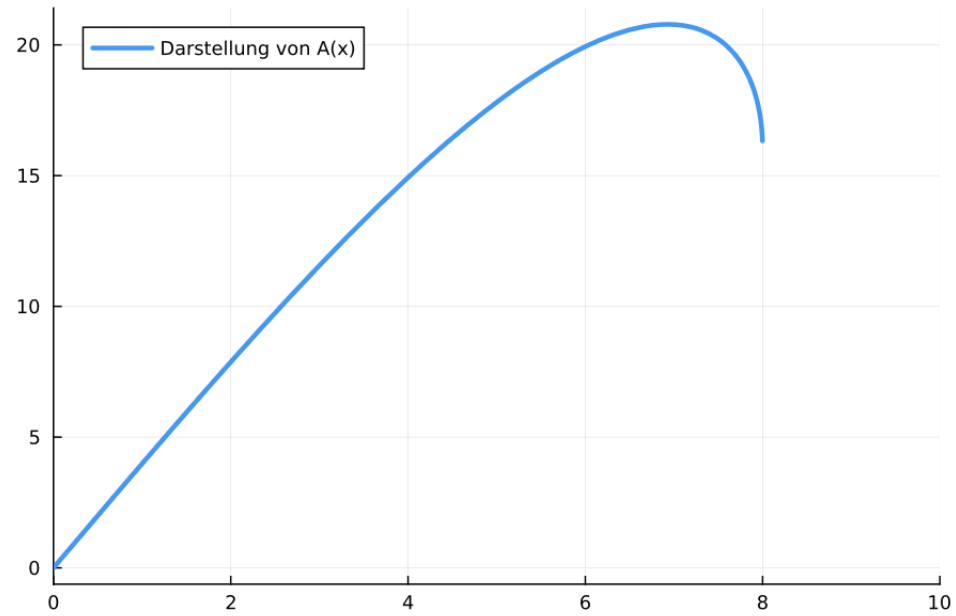
Abbildung 30 zeigt uns zuerst die Darstellung der Zielfunktion, in diesem Beispiel nun in Abhängigkeit der Entfernung der beiden senkrechten Pfosten. Damit der Code vereinfacht wird, verwenden wir die Variable x anstatt b. Mithilfe der graphischen Darstellung können wir den Maximalwert grob eingrenzen und somit geeignete Startwerte für das Sekantenverfahren wählen. Wir sehen, dass das Maximum zwischen $x = 6$ und $x = 7,5$ liegen muss, weshalb wir das Sekantenverfahren mit diesen beiden Werten starten. Daraus ergibt sich nach einigen Schritten eine Lösung, die um weniger als 10^{-5} von der eigentlichen Nullstelle abweicht. Anhand der graphischen Darstellung können wir die beiden Randwerte $x = 0$ sowie $x = 8$ als Optimum ausschließen, womit unsere numerisch gefundene Lösung den Maximalwert darstellen muss. Setzen wir diese in unsere Zielfunktion $A(b)$ ein, erhalten wir den maximalen Flächeninhalt des Tores:

$$
\begin{aligned}
b &\approx 6,92820323m \\
A(b) &= 2 \cdot b + \frac{b}{2} \cdot \sqrt{16 - \frac{b^2}{4}} \\
A(6,92820323) &\approx 20,78461m^2
\end{aligned}
$$

```
A(x)=2*x+x/2*(sqrt(16-x^2/4))
```

```
A (generic function with 1 method)
```

```
using Plots: plot, plot!
plot(A, xlimits=[0, 10],
     linewidth=3,
     label="Darstellung von A(x)")
```




```
mySecant(A,6,7.5)
```

```
8-element Vector{Float64}:
 6.0
 7.5
 6.61170967948869
 6.815105070535082
 6.9477306436738955
 6.926944183542601
 6.92818899775138
 6.9282032406193865
```

Abbildung 30: Darstellung der Zielfunktion $A(x)$ und Berechnen der Nullstelle von A'(x) mithilfe des Sekantenverfahrens

6.2 Anwendung mittels dualer Zahlen und Newton-Verfahren

Neben der Lösung durch automatisierter Differenzierung und Sekantenverfahren besteht auch die Möglichkeit, Nullstellen mithilfe des Newton-Verfahrens zu finden. Wie wir im Kapitel 5 sehen konnten, konvergiert dieses Verfahren quadratisch und dementsprechend schneller als das Sekantenverfahren. Somit wäre es die effizienteste Vorgehensweise, Nullstellen rein numerisch zu finden. Das Problem dabei ist aber, dass das Verfahren mit der ersten Ableitung arbeitet, was eine numerische Herausforderung darstellt. Möchten wir, ähnlich wie beim Sekantenverfahren, direkt die Ableitungswerte mittels automatisierter Differenzierung aufrufen, so würden wir innerhalb des Verfahrens in der Folge die zweite Ableitung benötigen, was mit der automatisierten Differenzierung nicht umsetzbar ist. Da unsere gewählten Beispiele jedoch Zielfunktionen aufweisen, bei denen mit den bekannten Ableitungsregeln die Ableitungsfunktion gefunden werden kann, können wir diese auch mithilfe des Newton-Verfahrens lösen. Dies zeigen wir anhand des folgenden Beispiels:

Beispiel (Aufgabe 19 aus [1]). Für den Bau eines Tores stehen für die senkrechten Begrenzungen zwei Pfosten mit $2m$ Länge und für die oben aufgesetzen schrägen Begrenzungen zwei Dachpfosten mit je $4m$ Länge zur Verfügung. Wie muss der Öffnungswinkel der Dachpfosten gewählt werden, wenn das Tor eine möglichst große Durchlassfläche (Querschnittsfläche) haben soll?

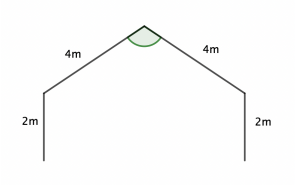


Abbildung 31: Skizze der Extremwertaufgabe [Aufgabe 19 aus [1]]

Numerische Lösung:

Da der Beginn des Beispiels gleich aufgebaut ist wie im Unterkapitel zuvor, verzichten wir auf eine Erklärung der Haupt- und Nebenbedingungen und arbeiten direkt mit der Zielfunktion in Abhängigkeit des Öffnungswinkels α. Da wie bereits erwähnt für das Newton-Verfahren das analytische Finden der Ableitungsfunktion Voraussetzung ist, berechnen wir diese im nächsten Schritt:

$$A(\alpha) = 16 \cdot (\sin(\frac{\alpha}{2}) + \sin(\frac{\alpha}{2}) \cdot \cos(\frac{\alpha}{2}))$$
$$A'(\alpha) = 8 \cdot (\cos(\frac{\alpha}{2}) + \cos^2(\frac{\alpha}{2}) - \sin^2(\frac{\alpha}{2}))$$

Mithilfe dieser Informationen können wir die numerische Lösung mittels dualer Zahlen und Newton-Verfahren beginnen. Die Abbildungen 32 und 33 zeigen uns dabei die wichtigsten Schritte:

Abbildung 32 stellt uns zuerst unser modifiziertes Newton-Verfahren dar. Dies wirkt auf den ersten Blick sehr ähnlich zu dem Code, der im Kapitel 5 schon beschrieben wurde. Der große Unterschied liegt hier dabei, dass innerhalb des Verfahrens die erste Ableitung benötigt wird und diese Werte mithilfe der automatisierten Differenzierung („myAD“) aufgegriffen werden. Das Newton-Verfahren arbeitet somit, sobald eine Funktion und ein Startwert festgelegt werden, rein numerisch. Anschließend sehen wir in dieser Abbildung unsere Funktionsgleichung $A(x)$ (Variable x statt α) und mittels „Using Plots: plot, plot!“ die graphische Darstellung dieser, um einen Startewert für das Newton-Verfahren bestimmen zu können. Die Vorgehensweise ist analog zu jener im vorherigen Unterkapitel.

Abbildung 33 zeigt uns jetzt den primären Unterschied zur Lösung mit dem Sekantenverfahren. Wir definieren in *Julia* zuerst unsere analytisch gefundene Ableitung und stellen diese sowohl mithilfe der automatisierten Differenzierung als auch exakt dar. Wie wir sehen, weicht die numerische Ableitungsfunktion nicht von der exakten Ableitung ab. Im letzten Schritt finden wir mithilfe des zuvor definierten „myNewton“ näherungsweise die Nullstelle der Ableitungsfunktion. Da die Ableitung analytisch gefunden wurde, können zumindest innerhalb des Verfahrens die Folgewerte numerisch mittels dualer Zahlen berechnet werden. Als Startwert wird dabei unsere geschätzte Lösung 2 verwendet, damit wir nicht zu weit vom gewünschten Ergebnis abweichen und das Verfahren überhaupt konvergieren kann. Wie in der untenstehenden Graphik klar ersichtlich, benötigt das Verfahren nur wenige Schritte bis ein Ergebnis, das um weniger als 10^{-5} von der Nullstelle abweicht, erreicht ist. Mit dieser numerischen Lösung erhalten wir folgenden Öffnungswinkel, bei dem die Querschnittsfläche des Tores maximal ist:

$$
\begin{aligned}
x &\approx 2,094395 \\
\alpha &= \frac{x}{\pi} \cdot 180 \approx \frac{2,094395}{\pi} \cdot 180 \\
\alpha &\approx 120^\circ
\end{aligned}
$$

Wie zuvor können die Randwerte ausgeschlossen werden und wir setzen unsere Lösung in die Zielfunktion ein:

$$
\begin{aligned}
A(\alpha) &= 16 \cdot (\sin(\frac{\alpha}{2}) + \sin(\frac{\alpha}{2}) \cdot \cos(\frac{\alpha}{2})) \\
A(120^\circ) &\approx 20,7861 m^2
\end{aligned}
$$

```
function myNewton(f, x0; tol=1e-5, maxiter=50)
    x= Vector{Float64}(undef, maxiter)
    x[1]=copy(x0)
    for k in 1: maxiter-1
        if abs(f(x[k]))<tol return x[1:k] end
        x[k+1]=x[k]-f(x[k])/myAD(f,x[k])
    end
    println("no convergence")
    return x
end
```

```
myNewton (generic function with 2 methods)
```

```
A(x)=16*(sin(x/2)+sin(x/2)*cos(x/2))
```

```
A (generic function with 1 method)
```

```
using Plots: plot, plot!
plot(A, xlimits=[0, π],
     linewidth=3,
     label="graphische Darstellung von A(x)")
```

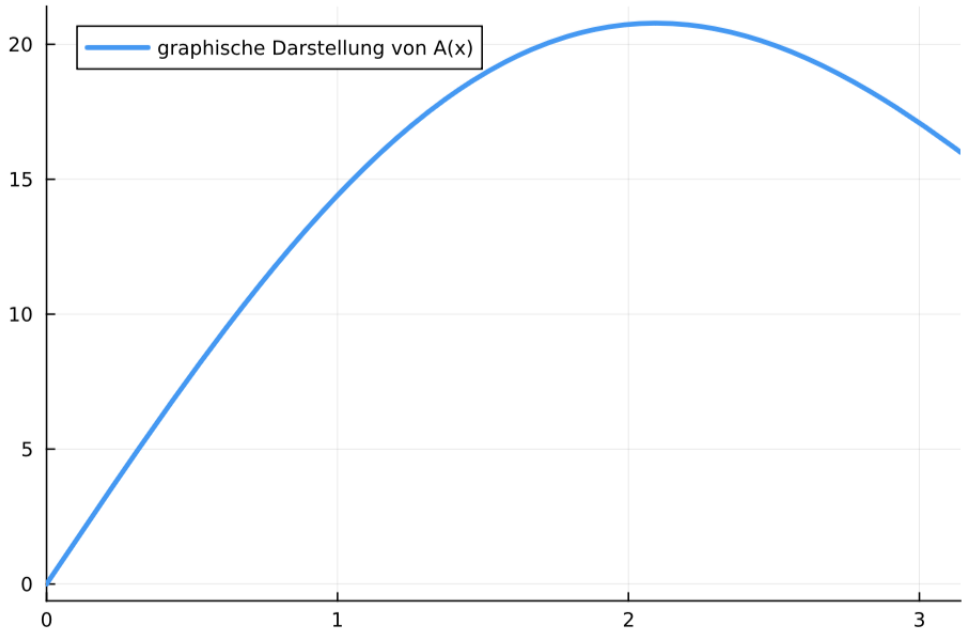


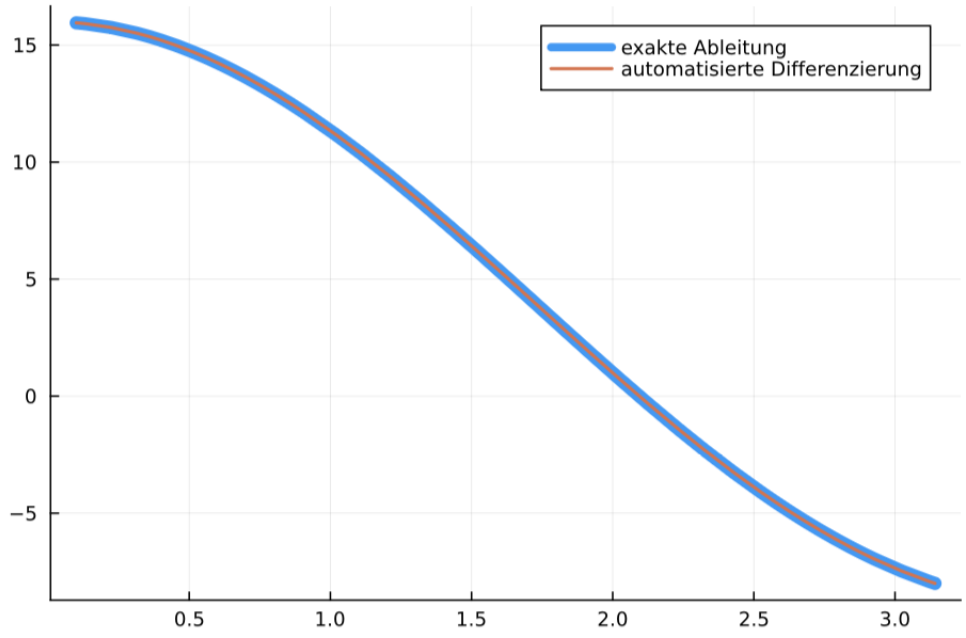
Abbildung 32: Implementieren des Newton-Verfahrens mithilfe der automatisierten Differenzierung und Darstellung der Funktion A(x)

```
A(x)=16*(sin(x/2)+sin(x/2)*cos(x/2))
∂A(x)=8*(cos(x/2)+cos(x/2)^2-sin(x/2)^2)

xmin, xmax = 0.1, π

plot(∂A, xmin, xmax,
     linewidth=8,
     label="exakte Ableitung")

plot!(x -> myAD(A, x),
      xmin, xmax,
      linewidth=2,
      label="automatisierte Differenzierung")
```




```
myNewton(∂A,2)
```

```
3-element Vector{Float64}:
 2.0
 2.09334766645694
 2.0943949443723304
```

Abbildung 33: Darstellung der Ableitung A' und näherungsweises Finden der Nullstelle mithilfe des Newton-Verfahrens und der automatisierten Differenzierung

6.3 Gegenüberstellung der Methoden

Zum Abschluss dieses Kapitels sollen beide Vorgehensweisen miteinander verglichen und die jeweiligen Vor- und Nachteile der Verfahren auf den Punkt gebracht werden.

Wenn wir die Ergebnisse der beiden unterschiedlichen Lösungsmethoden betrachten, fällt sofort auf, dass mit dem Sekanten- sowie dem Newton-Verfahren dieselbe Nullstelle gefunden wurde und somit auch der maximale Flächeninhalt des Tores identisch ist. Beide Verfahren funktionieren demnach und führen in diesem Beispiel zur korrekten Lösung. Wenn wir nun aber der Frage nachgehen, welches der im Rahmen dieser Arbeit besprochenen Verfahren zum Finden von Nullstellen in Kombination mit der automatisierten Differenzierung am besten geeignet ist, erhalten wir eine eindeutige Antwort: Das Sekantenverfahren.

Der Vorteil des Newton-Verfahrens liegt in der Schnelligkeit. In unserem Fall lagen wir mit der dritten Iteration unterhalb unserer Toleranzgrenze, womit die Lösung schneller gefunden wurde als mit dem Sekantenverfahren. Da das Newton-Verfahren jedoch in der Iterationsvorschrift die Ableitung benötigt, stellt es keine optimale Methode für das numerische Lösen von Extremwertaufgaben dar. Damit das Verfahren überhaupt numerisch arbeiten kann, muss in Kombination mit den dualen Zahlen die Ableitungsfunktion analytisch bestimmt werden. Erst dann kann in der Folge mithilfe der automatisierten Differenzierung die Nullstelle gefunden werden. Im Gegensatz dazu ist das Sekantenverfahren nur minimal langsamer und ermöglicht es uns, die Aufgabe rein numerisch zu lösen. Darüber hinaus ist die Iterationsvorschrift des Sekantenverfahrens nicht besonders kompliziert und lässt sich ohne Probleme gemeinsam mit der automatisierten Differenzierung in *Julia* implementieren und anschließend umsetzen. Somit erweist sich das Sekantenverfahren in Kombination mit der automatisierten Differenzierung im Rahmen dieser Arbeit als geeignetste Methode.

7 Anwendung der numerischen Lösungsmethoden in der Schule

Wie bereits in Kapitel 2 zu lesen ist, stellen Optimierungsaufgaben einen grundlegenden Baustein der Schulmathematik dar. Numerische Methoden, mit denen Optimierungsaufgaben anschaulich gelöst werden können, sind jedoch im Lehrplan und dementsprechend im Schulalltag nicht zu finden. Dieses Kapitel soll nun wesentliche Argumente anführen, wie numerische Methoden im Mathematikunterricht integriert werden können und welchen Mehrwert der Umgang mit numerischen Methoden für Schüler:innen bietet.

Bevor auf die Frage eingegangen wird, wie die numerischen Methoden in der Schule Platz finden können, soll an dieser Stelle diskutiert werden, inwiefern diese Verfahren Vorteile für Schüler:innen mit sich bringen. Wurde in den letzten Jahren noch gezielt ein Fokus auf analytische Berechnungen gelegt, so fällt heute auf, dass immer mehr auf die Unterstützung diverser Computerprogramme wie GeoGebra zurückgegriffen wird, um Ableitungen, Nullstellen oder Gleichungssysteme zu lösen. Vor zehn Jahren war es wohl kaum vorstellbar, dass die Ableitungsregeln nicht beherrscht werden. Mittlerweile ist erkennbar, dass sich viele Schüler:innen auf die Berechnungen mit den Computerprogrammen verlassen und nicht mehr in der Lage sind Ableitungsfunktionen geometrisch zu interpretieren. Manche Beispiele, wie das im vorherigen Kapitel behandelte, zielen sogar direkt darauf ab, mittels Technologieeinsatz gelöst zu werden.

Natürlich bringt der technische Fortschritt einige Vorteile mit sich. Meiner Meinung nach wäre es aber durchaus wichtig, dass die Schüler:innen auch grob verstehen, wie ein Computer arbeitet bzw. ob sie sich wirklich auf Ergebnisse verlassen können. Somit könnte mit einem grundlegenden Verständnis von numerischen Methoden eine kritische Denkweise der Jugendlichen gefördert werden, da technologische Programme auch auf näherungsweisen Lösungsmethoden basieren. Die Schüler:innen könnten dementsprechend auch ein Verständnis dafür entwickeln, wie Computer arbeiten und was in der „Black-Box“ zwischen Input und Output passiert. Darüber hinaus wäre es auch durchaus spannend, konkrete Verfahren, zum Beispiel die in der Arbeit vorliegenden Methoden zum Finden von Nullstellen, selbst zu erarbeiten und auszuprobieren. Einerseits sind die Iterationsvorschriften von angemessenem Schwierigkeitsgrad, andererseits beschäftigt man sich während der gesamten Schulzeit in irgendeiner Art und Weise immer wieder mit dem Problem der Nullstellenfindung. Das Anwenden und Verstehen dieser Methoden würde in vielen Bereichen die Denkmuster und -strategien der Schüler:innen fördern und vor

allem im Hinblick auf das Verständnis hinter dem Technologieeinsatz zu Erkenntnisgewinn führen.

Nun ist noch der Frage nachzugehen, wie eine Einbindung der numerischen Methoden, in unserem Fall zur Lösung von Extremwertaufgaben, im Schulalltag sinnvoll umgesetzt werden kann. Da der Lehrplan in allgemeinbildenden höheren Schulen (AHS) und berufsbildenden höheren Schulen (BHS) streng vorgegeben ist und die Zeit auch meist knapp ist, um diese Inhalte vollständig im Unterricht abzudecken, macht es wenig Sinn, im Regelunterricht an numerische Methoden zu denken. Meiner Meinung nach gibt es infolgedessen dennoch zwei Wege, um vor allem an der Mathematik interessierte Schüler:innen die Konzepte rund um die Numerik näherzubringen.

Erstens könnte im Zuge der schulautonomen Wahlpflichtfächer eine Vertiefung in Mathematik angeboten werden. Dies wäre vor allem im Bereich der AHS eine gute Möglichkeit, im Umfang von zwei Wochenstunden weiterführende Inhalte der Mathematik zu vermitteln. Dabei könnte gezielt ein Fokus auf numerische Methoden gelegt werden, sodass auch eine Brücke zum IT-Fachbereich geschlagen wird. Voraussetzung für das Zustandekommen eines Wahlpflichtgegenstandes ist, dass einerseits genügend interessierte Schüler:innen für diese Thematik vorhanden sind und andererseits eine motivierte Lehrperson sich zutraut, auch solche etwas komplexeren Inhalte zu vermitteln. Diese Faktoren erschweren natürlich die Einbindung der numerischen Mathematik im Schulalltag, wären aber dennoch eine Option, falls die Voraussetzungen gegeben sind. Vor allem in Realgymnasien mit naturwissenschaftlichem Zweig ist die Realisierung dieser gut vorstellbar.

Die zweite Möglichkeit ist meiner Ansicht nach leichter umsetzbar, bedarf aber auch wieder der Eigeninitiative der Schüler:innen. Im Zuge einer abschließenden Arbeit an der AHS (ABA) können sich motivierte Schüler:innen mit der Thematik rund um Methoden zum näherungsweisen Finden von Nullstellen oder sogar mit den dualen Zahlen auseinandersetzen. Dabei wäre es im Vorhinein jedoch wichtig, dass die Jugendlichen über diese Themen Bescheid wissen und aufgeklärt werden, dass sich hinter dem im Mathematikunterricht eingesetzten Technologieeinsatz unterschiedlichste Verfahren verbergen. Dabei spielt vor allem die Lehrkraft eine tragende Rolle als Motivator, damit die Schüler:innen sich einerseits motiviert mit dieser Thematik beschäftigen möchten und auch gewillt sind, den Technologieeinsatz kritisch zu hinterfragen.

Zusammenfassend lässt sich behaupten, dass die Einbindung der in der Arbeit vorge-

stellten numerischen Methoden im Schulalltag definitiv Vorteile mit sich bringt. Vor allem wird dabei das kritische Denken gefördert, aber auch ein konkreteres Verständnis für den Technologieeinsatz im Unterricht aufgebaut. Die Umsetzung stellt sich zwar als nicht einfach heraus, da die Thematik nicht im Lehrplan vorgesehen ist, könnte aber von interessierten Schüler:innen und ambitionierten Lehrkräften gemeistert werden.

8 Fazit

Zusammenfassend zeigt die vorliegende Arbeit, dass numerische Methoden eine sinnvolle Ergänzung zu den analytischen Verfahren bei der Lösung von Extremwertaufgaben darstellen. Während analytische Berechnungen weiterhin eine zentrale Position im Mathematikunterricht der Sekundarstufe einnehmen, eröffnen numerische Ansätze neue Perspektiven auf mathematische Problemstellungen und fördern einerseits das Verständnis des Technologieeinsatzes und stärken andererseits die Lern- und Denkprozesse bei Schüler:innen.

Die Ergebnisse verdeutlichen, dass das Zusammenspiel von numerischen Verfahren zum Finden von Nullstellen mit der automatisierten Differenzierung neue Zugänge zu komplexen Problemstellungen ermöglicht. Darüber hinaus können auch scheinbar unlösbare Aufgaben bewältigt werden, da weder eine Ableitungsfunktion noch eine Nullstelle analytisch gefunden werden müssen. Während sich das Newton-Verfahren per se als besonders effizient erweist, liefert das Sekantenverfahren im Zusammenspiel mit der automatisierten Differenzierung die besten Ergebnisse, da explizite Ableitungen innerhalb der Iterationsvorschrift vermieden werden. Weiters zeigt die Arbeit, dass die numerische Differenzierung mithilfe der dualen Zahlen eine hohe Genauigkeit aufweist und sich als geeignete Alternative zur Bestimmung von Ableitungswerten anbietet, wie mit Hilfe der Abbildungen und Beispiele demonstriert wurde. Insgesamt wird deutlich, dass die diskutierten Methoden nicht nur rechnerische Vorteile bieten, sondern auch didaktisch gewinnbringend eingesetzt werden können.

Abschließend lässt sich festhalten, dass die Implementierung numerischer Methoden im Mathematikunterricht vielseitige Vorteile bietet: Es können damit nicht nur komplexere Ableitungen und Nullstellen ohne analytische Vorgehensweise berechnet, sondern auch das Verständnis für den klassischen Technologieeinsatz im Unterricht und das kritische Denken gefördert werden. Weiterführende Forschungen könnten sich mit der praktischen Umsetzung dieser Methoden im Schulalltag auseinandersetzen und sich dabei insbesondere mit den Lernfortschritten der Schüler:innen beschäftigen.

Literatur

[1] Bleier, Gabriele; Lindenberg, Judith; Lindner, Andres; Süss-Stepancik, Evelyn: *Dimensionen – Mathematik 7. Differentialrechnung: Extremwertaufgaben (Arbeitsblatt)*. Wien: Verlag E. Dorner, Westermann, 2016. Online verfügbar unter: `https://c.wgr.de/f/verlage/westermanngruppe-at/dimensionen-mathematik/_dim7/materialien/02_Differentialrechnung/13_Extremwertaufgaben/02_13_extremwertaufgaben.pdf` (Zugriff am 28.03.2026).

[2] Dorfmayr, Anita; Mistlbacher, August; Sator-Wunsch, Katharina; Zillner, Michaela. *Thema Mathematik 7. 1. Semester (Schulbuch)*. Linz: VERITAS. Online verfügbar unter: `https://sbo.veritas.at/sbo/ebook/px/42991_1T/` (Zugriff am 28.03.2026).

[3] Friedrich, Hermann; Pietschmann, Frank. *Numerische Methoden*. Berlin; New York: De Gruyter Verlag, 2010

[4] Hermann, Martin. *Numerische Mathematik. Band 1: Alegbraische Probleme (4. Auflage)*. Berlin; Boston: De Gruyter Verlag, 2020

[5] Karpfinger, Christian. *Höhere Mathematik in Rezepten: Begriffe, Sätze und zahlreiche Beispiele in kurzen Lerneinheiten*. Berlin; Heidelberg: Springer-Verlag, 2014

[6] Koth, Maria; Salzger, Bernhard; Ulovec, Andreas; Woschitz, Helge: *Mathematik verstehen 7. Schulbuch*. Wien: öbv. Online verfügbar unter: `https://www.oebv.at/flippingbook/9783209123633/` (Zugriff am 28.03.2026).

[7] Koß, Rainer; Feiler, Simon P.; Liedtke, Jürgen; Bitzer, Jürgen; Essig, Timo; Marz, Michael; Rapedius, Kevin; Rutka, Vita. *Wegweiser durch die Mathematik - Grundlegende Verfahren*. Berlin: Springer-Verlag, 2024

[8] Lehrplan der allgemeinbildenden höheren Schulen (AHS), BGBl. II Nr. 277/2004 idgF, online verfügbar unter: `https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10008568` (Zugriff am 28.03.2026).

[9] Leuders, Timo. *Erlebnis Algebra: zum aktiven Entdecken und selbständigen Erarbeiten*. Freiburg: Springer-Verlag, 2016.

[10] Marti, Kurt. *Übungsbuch zum Grundkurs Mathematik für Ingineure, Natur- und Wirtschaftswissenschaftler*. Berlin; Heidelberg: Springer-Verlag, 2010

[11] Schupp, Hans. *Optimieren ist fundamental, In: Mathematik lehren, Heft 81.* Hannover: Friedrich Verlag, 1997

[12] Schwarz, Hans Rudolf; Köckler Norbert. *Numerische Mathematik (7. Auflage).* Wiesbaden: Vieweg+Taubner Verlag, 2009

[13] Weltner, Klaus. *Mathematik für Physiker und Ingenieure 1 (17. Auflage).* Berlin; Heidelberg: Springer Verlag, 2017

[14] Ye, Jim. *Mathematik jenseits des Schulstoffs. 16 Themen in 128 Lehr-Lern-Einheiten - für Lehrkräfte, Eltern und interessierte Jugendliche ab etwa 12 Jahren.* Berlin: Springer-Verlag, 2025

Abbildungsverzeichnis