universität wien

DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

Wie weit ist die Nullstelle noch entfernt?

-Iterationsverfahren zur Auffindung von Nullstellen nichtlinearer Funktionen

verfasst von / submitted by

Alexandra Uhl

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

Magistra der Naturwissenschaften (Mag. rer. nat)

Wien, 2018 / Vienna, 2018

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 190 406 477
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Lehramtsstudium Unterrichtsfach Mathematik Unterrichtsfach Haushaltsökonomie und Ernährung
Betreut von / Supervisor:	ao. Univ.-Prof. Mag. Dr. Peter Raith

Danksagung

Schon während meiner Schulzeit hatte ich ein starkes Interesse an der Mathematik und begeisterte mich für mathematische Probleme und Fragestellungen. Ein großer Motivator war vor allem mein verstorbener Großvater, der sehr stolz auf mein naturwissenschaftliches Interesse und mathematischen Leistungen war sowie meine Entscheidung Mathematik und Ernährung auf Lehramt zu studieren immer unterstützte und mein größter Fan gewesen ist. Ich weiß er würde diese Arbeit mit großer Freude lesen und allen davon erzählen.

Mithilfe meines Betreuers Univ.-Prof. Mag. Dr. Peter Raith hatte ich das Thema meiner Arbeit recht schnell gefunden. Dank seiner Unterstützung und aufmerksamen Betreuung konnte ich eines meiner mathematischen Lieblingsgebiete „Iterationsverfahren im ein- und mehrdimensionalen Raum zur Auffindung von Nullstellen von Funktionen“ erforschen und verstehen. Wenn ich zwischendurch nicht weiterwusste oder nicht ganz durchblickte, war er stets bemüht mir weiterzuhelfen, meine eigenen Gedanken zu ordnen und mich zur Lösung zu führen. Auch hier nochmals ein herzliches Danke für alles.

Am allerwichtigsten für mich war jedoch die Unterstützung meiner Eltern Michael und Susanne. Ohne ihre Hilfe hätte ich nicht die Möglichkeit gehabt mich voll und ganz auf mein Studium zu konzentrieren und die Ausbildung zu meinem Wunschberuf abzuschließen. Mein Vater weckte schon in früher Kindheit mein Interesse für Ernährungswissenschaften, welches bis heute anhält. Hiermit möchte ich mich nicht nur für den finanziellen, sondern vor allem für den emotionalen und familiären Beistand während des Studiums und meines bisherigen Lebens danken. Ich bin sehr froh und glücklich so liebevolle Eltern zu haben.

Ein großer Dank geht auch an meine Studienkollegin Tabea, die zu einer engen Freundin geworden ist. Stundenlang lernten wir gemeinsam, pushten uns zum Weitermachen und bauten uns gegenseitig nach Rückschlägen wieder auf.

Und jetzt viel Spaß beim Lesen!

Inhalt

Vorbereitung 3

Vektor- und Matrizennormen 3

Vektorfunktion 4

Ableitungen einer Vektorfunktion 4

Satz von Taylor 7

1. Allgemeines Iterationsverfahren 8

1.1. Herleitung der allgemeinen Iterationsvorschrift 8

1.2. Existenz und Eindeutigkeit der Lösung 13

1.3. Konvergenz 17

1.4. Abbruchbedingungen 23

1.5. Fehlerabschätzung 23

1.6. Konvergenzordnung 27

1.6. Praktische Anwendung des Algorithmus 29

2. Regula falsi 33

2.1. Herleitung der Formel 33

2.2. Existenz und Eindeutigkeit der Lösung 35

2.3. Konvergenz 35

2.4. Fehlerabschätzung und Abbruchbedingung 38

2.5. Praktische Anwendung des Algorithmus 39

3. Newtonverfahren in $\mathbb{R}$ 42

3.1. Herleitung der Formel 42

3.2. Konvergenz und Konvergenzordnung 44

3.3. Fehlerabschätzung und Abbruchbedingung 47

3.4. Vergleich mit Regula falsi 49

3.5. Praktische Anwendung des Algorithmus 50

3.5.1. Allgemein 50

3.5.2. Wurzelberechnung 53

4. Newtonverfahren in $\mathbb{R}^n$ 61

4.1. Herleitung der Formel und Konvergenz 61

4.2. Praktische Anwendung des Algorithmus 70

4.2.1. Vereinfachtes Newtonverfahren 73

4.2.2. Spezialfall $n = 2$ 76

5. Iteration in der Schule 81

Literaturverzeichnis 82

Abbildungsverzeichnis 84

Zusammenfassung 85

Abstract 86

Vorbereitung

Zur Vorbereitung auf die folgende Arbeit werden hier nun einige Definitionen, Sätze und Instrumente der Analysis in $\mathbb{R}^n$ definiert die bei der Herleitung der Iterationsverfahren benötigt werden.

Vektor- und Matrizennormen

Definition: Sei $\mathbb{R}^n$ ein Vektorraum und $x, y \in \mathbb{R}^n, a \in \mathbb{R}$, dann ist $\|x\|$ eine Norm von x, wenn für alle $x \in \mathbb{R}^n$gilt:

(i) $\|x\| \geq 0,$ wobei $\|x\| = 0$ gilt wenn $x = 0$

(ii) $\|ax\| = |a|\ \|x\|$

(iii) $\|x + y\| \leq \|x\| + \|y\|$

Weiters gilt mit $n \in \mathbb{N}$

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$$

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}$$

sowie die Ungleichung

$$\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1 \leq n\|x\|_\infty$$

Definition: Sei M, eine $(n \times n)$ Matrix mit $M = (m_{ik})$; $m_{ik}, a \in \mathbb{R}$, dann $\|M\|$ eine Matrixnorm von M, wenn

(i) $\|M\| \geq 0,$ wobei $\|M\| = 0$ gilt wenn $M = 0$

(ii) $\|aM\| = |a|\ \|M\|$

(iii) $\|M + N\| \leq \|M\| + \|N\|,$ wobei N ebenfalls eine Matrix ist.

(iv) $\|M\ N\| \leq \|M\|\ \|N\|,$ wobei N ebenfalls eine Matrix ist

Weiters gelten mit $n \in \mathbb{N}$ die Matrixnormen

$$\|M\|_\infty = \max_{1\le i\le n} \sum_{k=1}^{n} |m_{ik}|$$

$$\|M\|_1 = \max_{1\le k\le n} \sum_{i=1}^{n} |m_{ik}|$$

$$\|M\|_2 = \sqrt{\sum_{i,k=1}^{n} |m_{ik}|^2}$$

Vektorfunktion

Ein System von Funktionen $h_j: \mathbb{R}^n \to \mathbb{R}$ in einem Gebiet $A \subset \mathbb{R}^n$

$$\begin{aligned} h_1 &= f_1(x_1, \dots, x_n) \\ &\vdots \\ h_m &= f_m(x_1, \dots, x_n) \end{aligned}$$

kann als Vektorfunktion $f: A \subset \mathbb{R}^n \to \mathbb{R}^m$ durch

$$f(x_1, x_2, \dots, x_n) = \begin{bmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_m(x_1, \dots, x_n) \end{bmatrix}$$

definiert werden, wobei $f_i: A \subset \mathbb{R}^n \to \mathbb{R},\ i = 1,2,3, \dots, m$ jedem Punkt $(x_1, x_2, \dots, x_n) \in A$ eine reelle Zahl $h_i = f_i(x_1, x_2, \dots, x_n) \in \mathbb{R}$ zugeordnet.

Die Vektorfunktion wird im Folgenden abgekürzt durch:

$$f(x) \coloneqq \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix}$$

Ableitungen einer Vektorfunktion

Um die Ableitung von einem System von Funktionen zu erhalten, muss partiell abgeleitet werde. Wir nehmen f als stetig differenzierbar an.

Definition: Sei $f: A \subset \mathbb{R}^n \to \mathbb{R}$. Die partielle Ableitung nach der j Variable an der Stelle $x \in A$ definiert man dadurch, dass man die anderen Variablen fix lässt und nur nach x_j ableitet. Man setzt dann

$$D_j f(x) = \frac{\partial f}{\partial x_j}(x) = f_{x_j}(x) \qquad j = 1, \dots, n$$

Weiteres wird der Vektor von f an der Stelle x als Gradient von f bezeichnet und ist beschrieben durch:

$$gradf(x) \coloneqq \nabla f(x) \coloneqq \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x) e_i = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}(x)$$

Wobei $(e_1, \dots, e_n)$ die Standardbasis von $\mathbb{R}^n$ ist und die einzelnen Komponenten des Vektors die partiellen Ableitungen sind.

Definition: Sei A ein offenes Gebiet mit $A \subseteq \mathbb{R}^n$. Eine Funktion $f: A \to \mathbb{R}^m$ heißt differenzierbar in $x_0 \in A$, falls es eine $m \times n$- Matrix $J(x_0)$ gibt, sodass

$$\lim_{x \to x_0} \frac{\left\| f(x) - f(x_0) - J_f(x_0)(x - x_0) \right\|}{\|x - x_0\|} = 0.$$

Proposition: Sei $f: A \subset \mathbb{R}^n \to \mathbb{R}^m$. Falls f differenzierbar in $x \in A$ ist, dann wird die erste Ableitung durch eine Matrix der Form

$$f'(x) = J_f(x) \coloneqq \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \frac{\partial f_m}{\partial x_2}(x) & \cdots & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix}$$

$$= \left(\frac{\partial f_i}{\partial x_j}(x) \right) \qquad \begin{matrix} i = 1, \dots, m \\ j = 1, \dots, n \end{matrix}$$

dargestellt. $J_f(x)$ wird dabei als Funktionalmatrix oder Jacobi-Matrix von f bezeichnet.

Definition: Sei $A \subseteq \mathbb{R}^n$ ein offenes Gebiet und $f: A \to \mathbb{R}^m$. Dann ist f auf A differenzierbar, wenn f in jedem $x \in A$ differenzierbar ist.

Satz: Sei A offen und ein Teilraum von $\mathbb{R}^n$ ist und alle partiellen Ableitungen von einer Funktion $f: A \to \mathbb{R}^m$ stetig auf A sind[1], dann ist f differenzierbar auf A.

Beweis: Es reicht den Satz für $m = 1$ zu beweisen. Sei $\varepsilon > 0$. Zuerst wird ein $\tilde{x} \in \mathbb{R}^n$ gewählt sodass für ein $\delta > 0$ mit $U_\delta(x) \subseteq A$

[1] Also stetig partiell differenzierbar

$$\|\tilde{x} - x\|_\infty < \delta$$

und aufgrund der Stetigkeit der partiellen Ableitungen von f

$$\left|\frac{\partial f}{\partial x_i}(\tilde{x}) - \frac{\partial f}{\partial x_i}(x)\right| < \frac{\varepsilon}{n}$$

gilt. Sei nun $v \in \mathbb{R}^n,\ \ v \neq 0,\ \ \|v\|_\infty < \delta$

$$f(x+v) - f(x) = \sum_{i=1}^{n}\left[f\left(x + \sum_{j=1}^{i} v_j e_j\right) - f\left(x + \sum_{j=1}^{i-1} v_j e_j\right)\right]$$

Jetzt kann der Mittelwertsatz angewandt werden, demnach existiert für jedes $1 \leq i \leq n$ ein $\xi_i \in (0,1)$, sodass

$$f\left(x + \sum_{j=1}^{i} v_j e_j\right) - f\left(x + \sum_{j=1}^{i-1} v_j e_j\right) = \frac{\partial f}{\partial x_i}\left(x + \sum_{j=1}^{i-1} v_j e_j + \xi_i v_i e_i\right) h v_i$$

$$f(x+v) - f(x) - \sum_{i=1}^{n}\frac{\partial f}{\partial x_i}(x) v_i = \sum_{i=1}^{n}\left[\frac{\partial f}{\partial x_i}\left(x + \sum_{j=1}^{i-1} v_j e_j + \xi_i v_i e_i\right) - \frac{\partial f}{\partial x_i}(x)\right] v_i$$

Für jedes $1 \leq i \leq n$ gilt außerdem wegen

$$\left\|\left(x + \sum_{j=1}^{i-1} v_j e_j + \xi_i v_i e_i\right) - x\right\|_\infty = \max\{|v_1|, |v_2|, \dots, |v_{i-1}|, |v_i|\} \leq \|v\|_\infty < \delta$$

immer

$$\left|\frac{\partial f}{\partial x_i}\left(x + \sum_{j=1}^{i-1} v_j e_j + \xi_i v_i e_i\right) - \frac{\partial f}{\partial x_i}(x)\right| < \frac{\varepsilon}{n}.$$

Damit kann folgende Abschätzung getroffen werden

$$\left|f(x+v) - f(x) - \sum_{i=1}^{n}\frac{\partial f}{\partial x_i}(x) v_i\right| < \sum_{i=1}^{n}\frac{\varepsilon}{n}|v_i| \leq \varepsilon\|h\|_\infty$$

$$\frac{\left|f(x+v) - f(x) - \sum_{i=1}^{n}\frac{\partial f}{\partial x_i}(x) v_i\right|}{\|h\|_\infty} < \varepsilon$$

Somit ist bewiesen, dass f für alle $x \in A$ differenzierbar ist. □

Definition: Falls $f(x)$ den obigen Satz erfüllt, dann nennt man f stetig differenzierbar auf A, beziehungsweise kurz f ist C^1.

Satz von Taylor

Sei $f(x)\colon \mathbb{R}^n \to \mathbb{R}^n$ eine $n+1$ mal stetig differenzierbare Funktion. Weiteres existiert f^{n+1} in einer Umgebung A um $a \in \mathbb{R}^n$. Dann gilt für jedes $x \in A$

$$f(x) = f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2 + \frac{1}{3!}f'''(a)(x-a)^3 + \cdots + \frac{1}{n!}f^n(a)(x-a)^n + R_n(x),$$

wobei für ein b zwischen x und a

$$R_n(x) = \frac{1}{(n+1)!}f^{n+1}(b)(x-a)^{n+1}$$

gilt.

Ist zusätzlich $|f^{n+1}(x)|$ für alle $x \in A$ durch ein M_{n+1} beschränkt, dann gilt außerdem

$$|R_n(x)| \leq \frac{1}{(n+1)!}M_{n+1}|x-a|^{n+1}$$

1. Allgemeines Iterationsverfahren

Angenommen es wird die Nullstelle η einer Polynomfunktion $f(x)$, $f: \mathbb{R}^n \to \mathbb{R}^n$ mit Grad n gesucht, sodass $f(\eta) = 0$

$$f(x) = p(x) \coloneqq \sum_{j=0}^{n} a_j x^j \quad \text{mit } j, n \in \mathbb{N}; \quad a_j, x \in \mathbb{R}^n; \quad a_n \neq 0$$

Nun kann mit Hilfe von Lösungsformeln aus der Algebra für alle $n \leq 4$ das Polynom $p(x) = 0$ aufgelöst werden, allerdings ist dies in der Praxis bereits für Polynome mit Grad 3 schon recht kompliziert.

Ab $n > 4$ gibt es jedoch keine Auflösungsformel[2] mehr (Karpfinger & Meyberg, 2012). Die Lösung kann nur näherungsweise bestimmt werden, indem eine erste grobe Näherungslösung x_0 bis zu einer bestimmten Genauigkeit durch schrittweise Iteration optimiert wird (Friedrich & Pietschmann, 2010).

1.1. Herleitung der allgemeinen Iterationsvorschrift

Zuerst wird die Gleichung $f(x) = 0$ auf eine Gleichung der Form $x = \varphi(x)$ gebracht[3].

Definition: Sei $\varphi(x)$ eine stetige Funktion auf einem abgeschlossenen Intervall $A \subset \mathbb{R}^n$, $\varphi: A \to \mathbb{R}^n$, mit $x \to \varphi(x)$. Dann ist $\eta \in A$ ein Fixpunkt der Funktion $\varphi(x)$, wenn $\varphi(\eta) = \eta$. Die Gleichung

$$x = \varphi(x)$$

wird als Fixpunktgleichung bezeichnet. Die Funktion $\varphi(x)$ ist die Schrittfunktion oder Iterationsfunktion zu $f(x)$.

Diese Fixpunktgleichung kann nun ohne Beschränkung der Allgemeinheit untersucht werden, da $f(x) = 0$ und $x = \varphi(x)$ die gleichen Lösungen besitzen. Anders ausgedrückt ist die Abszisse des Fixpunkts der Funktion $\varphi(x)$ gleich die Nullstelle von $f(x)$. (Freund & Hoppe, 2007) (Engeln-Müllges, Niederdrenk, & Wodicka, Numerik-Algorithmen, 2005)

[2] Dies gilt auch für Nullstellenberechnung von Polynomfunktionen mit Grad >4 und transzendente Gleichungen

[3] Meistens indem $f(x) = 0$ nach x aufgelöst wird

Satz 1.1: Seien f und h stetige Funktionen in einem abgeschlossenen Intervall $A \subset \mathbb{R}^n$, wobei für alle $x \in A$auch $h(x) \neq 0$ gilt. Sei $\varphi(x)$ eine weitere Funktion, definiert durch

$$\varphi(x) \coloneqq x - f(x)h(x).$$

Dann sind die beiden Gleichungen $f(x) = 0$ und $\varphi(x) = x$ äquivalent.

Satz 1.2: Die Gleichung $x = \varphi(x)$ besitzt in einem endlich abgeschlossenen Gebiet $A \subset \mathbb{R}^n$ höchstens eine Lösung, wenn die Funktion $\varphi(x)$: $A \to \mathbb{R}^n$ eine der folgende Bedingungen erfüllt:

(i)	$\|\varphi(x) - \varphi(\tilde{x})\| \leq L\|x - \tilde{x}\|$	$0 \leq L < 1$
(ii)	$\sup_{x \in A} \|\varphi'(x)\| < 1$	falls φ differenzierbar in A

Beweis:

(i) Wenn $\eta \in A$ eine Nullstelle von $f(x)$ ist, dann gilt:

$$f(\eta) = 0 \;\Rightarrow\; \varphi(\eta) = \eta - \underbrace{f(\eta)}_{=0} * h(\eta) = \eta$$

(ii) Wenn $\eta \in A$ eine Lösung für $\varphi(x) = x$ ist, dann gilt

$$\varphi(\eta) = \eta \;\Rightarrow\; \varphi(\eta) = \eta - \underbrace{f(\eta) * h(\eta)}_{=0}$$

Nun muss laut Satz 1.1 $h(x) \neq 0, \forall x \in A$ also ist $f(\eta) = 0$

(Engeln-Müllges, Niederdrenk, & Wodicka, Numerische Verfahren zur Lösung nichtlinearer Gleichungen, 2011)

□

Beispiel 1: Gegeben ist die Gleichung $f(x) = x^2 - x - 1 = 0$. Gesucht ist die Schrittfunktion $\varphi(x)$ der Funktion $f(x)$.

Je nachdem wie $f(x)$ nach x aufgelöst wird, erhält man verschiedene Schrittfunktionen

(i) $x = x^2 - 1 \;\Rightarrow \varphi(x) \coloneqq x^2 - 1$,

(ii) $x = \sqrt{x+1} \;\Rightarrow \varphi(x) \coloneqq \sqrt{x+1}, \qquad x \geq -1$

(iii) $x = 1 + \frac{1}{x} \quad \Rightarrow \varphi(x) \coloneqq 1 + \frac{1}{x}, \qquad x \neq 0$

Der Zusammenhang zwischen der Nullstelle einer Funktion und dem Fixpunkt der dazugehörigen Schrittfunktion lässt sich im zweidimensionalen Raum gut darstellen. In Abbildung 1 ist die Funktion $f(x)$ mit der Schrittfunktion (ii) aus dem vorherigen Beispiel 1 abgebildet. Der Schnittpunkt zwischen $\varphi(x)$ und der ersten Mediane[4] besitzt die selbe x-Koordinate η wie die Nullstelle von $f(x)$. Demnach haben die Gleichungen $\varphi(x) = x$ und $f(x) = 0$ dieselbe Lösung[5].

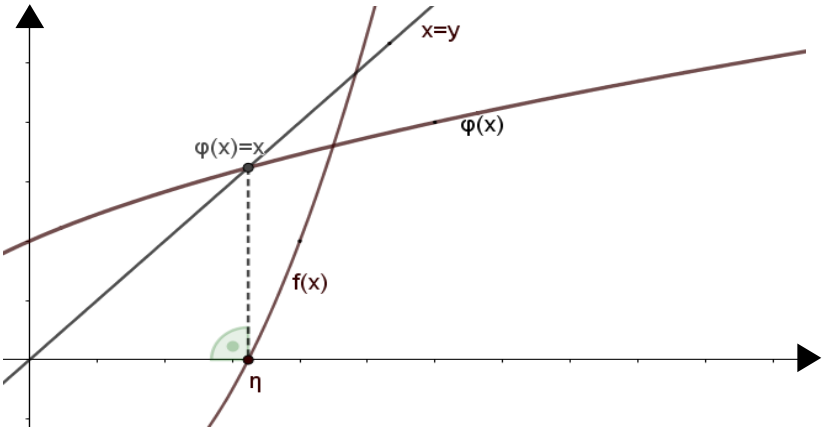

Abbildung 1

Für die numerische Auffindung der Nullstelle der Funktion $f(x)$, beziehungsweise zur Lösung der Gleichung $f(x) = 0$ wird mithilfe eines Startwerts $x_0 \in A$ und der Schrittfunktion eine rekursive Zahlenfolge $\{x_i\}$ konstruiert, nach der Vorschrift

$$x_{i+1} = \varphi(x_i), \qquad i = 0,1,2, \dots$$

Wird in die Schrittfunktion eingesetzt, liefert $\varphi(x_o)$ das nächste Folgenglied x_1, auf das wiederum die Funktion angewendet wird um x_2 zu berechnen. Die so konstruierte Folge $\{x_i\}$ heißt Iterationsfolge und ist nur dann sinnvoll, wenn $\varphi(x_i)$ selbst wieder im Intervall A liegt. Wenn nun $\{x_i\}$ gegen η konvergiert, also sich $x_i$mit größer werdenden i immer näher an η annähern, dann ist η eine Lösung der Fixpunktgleichung und somit auch eine Lösung für $f(x) = 0$.

Satz 1.3: Sei φ eine stetige Funktion. Wenn die Iterationsfolge $\{x_i\} \in A$ gegen η konvergiert dann ist $\eta = \varphi(\eta)$

[4] Definiert durch $x = y$
[5] Vgl. Satz 1.1

Beweis: $\eta = \lim_{i\to\infty} x_i = \lim_{i\to\infty} \underbrace{x_{i+1}}_{\varphi(x_i)} = \lim_{i\to\infty} \varphi(x_i) \xrightarrow[stetig]{\varphi} \varphi\left(\underbrace{\lim_{i\to\infty} x_i}_{\eta}\right) = \varphi(\eta)$ □

(Engeln-Müllges, Niederdrenk, & Wodicka, Numerische Verfahren zur Lösung nichtlinearer Gleichungen, 2011) (Friedrich & Pietschmann, 2010) (Freund & Hoppe, 2007)

Definition: Sei $x_0 \in \mathbb{R}^n$ die erste grobe Näherungslösung für die Gleichung $f(x) = 0$, $f: \mathbb{R}^n \to \mathbb{R}^n$, dann wird x_0 als Startwert bezeichnet.

Zur Auffindung eines geeigneten Startwertes bieten sich Wertetabellen und grafische Darstellungen der Funktion $f(x)$ an.

Beispiel 2: Gegeben ist die Funktion $g(x) = x^3 - \sin(x) + 2$. Gesucht wird ein Startwert x_0.

x	$g(x)$
-2,5	-13,03
-2	-5,09
-1,5	-0,38
-1	1,84
-0,5	2,35

Tabelle 1

Tabelle zur Auffindung: Wichtig für diese Methode ist der Zwischenwertsatz von Bolzano. Dieser sagt aus, dass eine stetige Funktion f auf einem abgeschlossenen Intervall $[a, b]$ jeden Wert zwischen $f(a)$ und $f(b)$ mindestens einmal annimmt. Falls nun $f(a)$ und $f(b)$ verschiedene Vorzeichen hat, so besitzt die Funktion mindestens eine Nullstelle η, sodass $a < \eta < b$ gilt (Arens, Busam, Hettlich, Karpfinger, & Stachel, 2013). Aus Tabelle I folgt nun, dass $g(x)$ im Intervall $[-1{,}5; -1]$ mindestens eine Nullstelle besitzt. Also wählt man den Mittelwert des Intervalls als Startwert.

$$x_0 = \frac{-1{,}5 - 1}{2} = -1{,}25$$

Graph zur Auffindung: Durch plotten der Funktion wird ein geeigneter Startwert grafisch ermittelt. Aus Abbildung 2 kann x_0= -1,45 abgelesen werden.

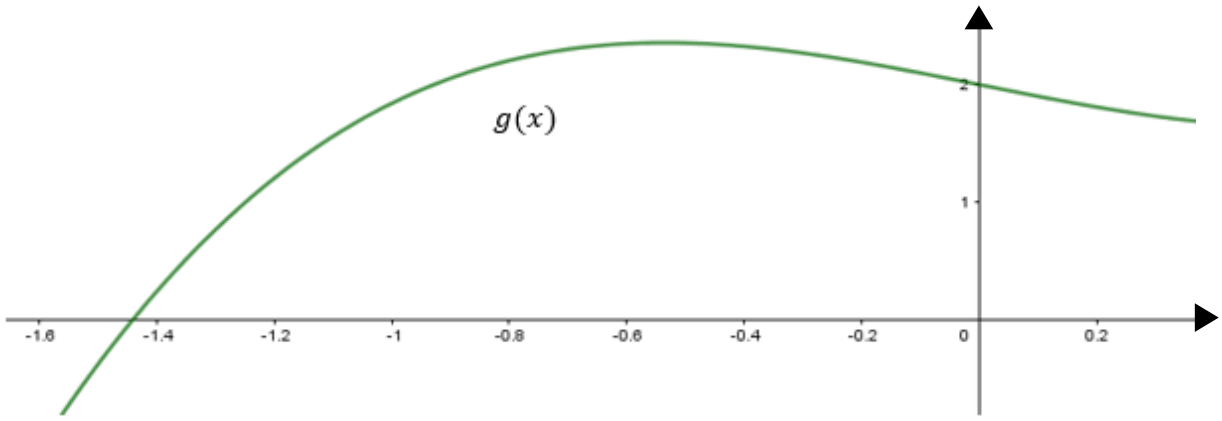

Abbildung 2

Definition: Sei $\varphi(x)$ eine stetige Funktion, $\varphi: A \to A$, $A \subset \mathbb{R}$ und $\varphi(x) = x$ die Fixpunktgleichung für $f(x) = 0$, dann ist die Iterationsvorschrift gegeben durch

$$x_{i+1} = \varphi(x_i), \qquad i = 0,1,2, \ldots$$

Beispiel 3: Gegeben ist die Funktion $f(x) = x^2 - x - 1$. Gesucht ist eine Iterationsvorschrift zur Berechnung der Nullstelle der Funktion $f(x)$, um die ersten 4 Näherungslösungen anzugeben.

Für die Gleichung $x^2 - x - 1 = 0$ sind in Beispiel 1 mögliche Fixpunktgleichungen ausgerechnet. Als Startwert x_0 wird 2 angenommen

(i) $x_{i+1} = \varphi(x_i) = x_i^2 - 1, \qquad x_i \in \mathbb{R}$

(ii) $x_{i+1} = \varphi(x_i) = \sqrt{x_i + 1} \qquad x_i \in \mathbb{R},\ x_i \geq -1$

(iii) $x_{i+1} = \varphi(x_i) = 1 + \frac{1}{x_i} \qquad x_i \in \mathbb{R},\ x_i \neq 0$

	(i)	(ii)	(iii)
x_0	2	2	2
x_1	3	1,7321	1,5
x_2	8	1,6529	1,6667
x_3	63	1,6288	1,6
x_4	3968	1,6213	1,625

Tabelle 2

In Tabelle 2 sind die algorithmischen Berechnungen der drei verschiedenen Iterationsverfahren dargestellt. Die Berechnung von (i) zeigt, dass die konstruierte Iterationsfolge nicht gegen die Lösung 1,62 konvergiert und demnach nicht zur Lösungsberechnung geeignet ist. Bei (ii) und (iii) hingegen deutet die Tabelle 1 an, dass die Folge gegen 1,62 konvergiert und somit sind die Schrittfunktionen gut geeignet.

Die Wahl der Funktion $\varphi(x)$ ist somit von großer Bedeutung, da diese sowohl Einfluss auf die Konvergenzgeschwindigkeit als auch die Konvergenz selbst hat (Friedrich & Pietschmann, 2010). Nun stellt sich die Frage ob es Konvergenzbedingungen für eine Iterationsfolge gibt, damit die passende Schrittfunktion ausgewählt wird. Um das zu beantworten ist zuerst die Frage nach der Eindeutigkeit und Existenz von η zu klären.

1.2. Existenz und Eindeutigkeit der Lösung

Um die Existenz nachzuweisen, wird das Intervall $A \coloneqq [a, b] \subset \mathbb{R}$ betrachtet in dem die Lösung η liegt. Es gilt

$$a < \eta < b$$

Aus dem Zusammenhang zwischen der Funktion $f(x)$ für dessen Gleichung $f(\eta) = 0$ und der dazugehörigen Schrittfunktion $\varphi(x)$ wissen wir, das η die Abszisse des Schnittpunktes des Grafen von φ und der ersten Mediane ist. Deswegen muss das Bild von φ für alle $x \in A$ auch im Intervall A liegen. Da a und b nicht die Lösung η sein können ist $a \neq \varphi(a)$ und $b \neq \varphi(b)$. Weil $\varphi(a), \varphi(b) \in A$ ist, muss $a < \varphi(a)$ und $b > \varphi(b)$ sein. Werden diese Bedingungen in $\mathbb{R}$ betrachtet so bilden die Intervalllängen auf den beiden Koordinatenachsen einen Definitionsbereich der als Quadrat darstellbar ist. Die Diagonale dieses Quadrats liegt wegen der Symmetrie der Intervalllängen genau auf der ersten Mediane. Die Punkte $P\ (a|\varphi(a))$, $Q\ (b|\varphi(b))$ liegen auf den Seiten dieses Quadrats und somit auf verschiedenen Seiten der Mediane. Jetzt erklärt sich auch die Notwendigkeit, dass $\varphi(x)$ im Intervall A stetig ist. In Abbildung 3 ist φ in A stetig und schneidet deswegen mit Sicherheit die erste Mediane. In Abbildung 4 hingegen ist φ in A nicht stetig, so muss dieser Schnittpunkt nicht zwingend existieren und somit ist es nicht sicher ob die Lösung η existiert.

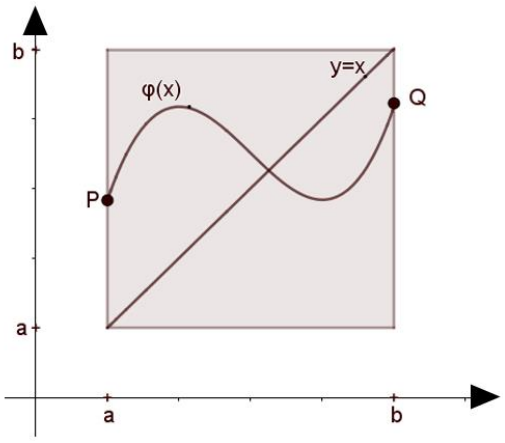

Abbildung 3

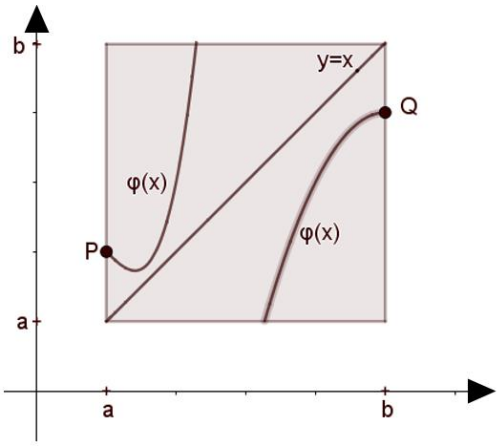

Abbildung 4

Satz 1.4: Die Gleichung $x = \varphi(x)$ besitzt in einem endlich abgeschlossenen Intervall $A \subset \mathbb{R}$ mindestens eine Lösung, wenn die Funktion $\varphi(x) : A \to \mathbb{R}$ folgende Bedingungen erfüllt:

(i) $\varphi(x) \in A \quad \forall x \in A$

(ii) $\varphi(x)$ ist in A stetig

(Engeln-Müllges, Niederdrenk, & Wodicka, Numerische Verfahren zur Lösung nichtlinearer Gleichungen, 2011)

Die Existenz dieser Lösung lässt sich auch analytisch beweisen.

Beweis: Sei $h: A \coloneqq [a; b] \subset \mathbb{R} \to \mathbb{R}$ eine in A stetige Funktion definiert durch

$$h(x) = x - \varphi(x)$$

Weiteres gilt $\varphi(a) \geq a$ und $\varphi(b) \leq b$ und die Lösung η liegt in A. Deswegen auch $h(a) \leq 0$ und $h(b) \geq 0$. Der Zwischenwertsatz besagt nun, dass die Funktion h jeden Wert zwischen $h(a)$ und $h(b)$ in A annimmt. Demnach muss ein η existieren, sodass $h(\eta) = 0$ ist. Daraus folgt:

$$\underbrace{h(\eta)}_{=0} = \eta - \varphi(\eta)$$

$$0 = \eta - \varphi(\eta)$$

$$\eta = \varphi(\eta)$$

Somit existiert mindestens eine Lösung für die Gleichung $x = \varphi(x)$ in A.

□

(Henrici, 1972)

Da wir nun wissen, dass $\eta \in A$ existiert, wenn Satz 1.4 erfüllt ist, bleibt noch zu zeigen, dass das Gleichungssystem eindeutig lösbar ist wenn η auch die einzige Lösung ist. Dazu wird bewiesen, dass $x = \varphi(x)$ genau dann eindeutig lösbar ist, wenn $\varphi(x)$ in A eine kontrahierende Abbildung ist. Die Bezeichnung kontrahierend bedeutet, dass die Norm $\|\varphi(x) - \varphi(\tilde{x})\|, x, \tilde{x} \in A$ kleiner ist als die Norm $\|x - \tilde{x}\|$. $\varphi(x)$ „zieht“ sozusagen die Punkte enger zusammen. (Törnig & Spellucci, 1988) (Werner, 1992)

Definition: Eine Funktion $\varphi(x): A \to A, A \subset \mathbb{R}$ heißt lipschitzstetig[6], wenn es für $x, \tilde{x} \in A$ die Lipschitzbedingung:

$$\|\varphi(x) - \varphi(\tilde{x})\| \leq L\|x - \tilde{x}\| \qquad L \geq 0,$$

gilt. L wird als Lipschitzkonstante bezeichnet und ist definiert durch

$$L = \max_{x \in A} \|\varphi'(x)\|.$$

Wenn $0 \leq L < 1$ ist, so kann L auch als Kontraktionszahl und $\varphi(x)$ als kontrahierende Abbildung bezeichnet werden.

(Locher, Numerische Mathematik für Informtiker, 1992)

Durch Umformen der Lipschitzbedingung erhält man eine neue Ungleichung die mithilfe der ersten Ableitung von $\varphi(x)$ berechnet werden kann

$$\|\varphi(x) - \varphi(\tilde{x})\| \leq L\|x - \tilde{x}\|$$

$$\frac{\|\varphi(x) - \varphi(\tilde{x})\|}{\|x - \tilde{x}\|} \leq L$$

Laut Mittelwertsatz der Differenzialrechnung gilt: Wenn $\varphi(x)$ auf einem Intervall $[a, b]$ stetig ist und im Intervall (a, b) differenzierbar ist, dann exisitiert mindestens ein Punkt $\lambda \in (a, b)$ sodass:

$$\frac{f(b) - f(a)}{b - a} = f'(\lambda)$$

gilt (Heuser, 1990). Durch Anwenden des Mittelwertsatzes auf die Ungleichung und hinzufügen der Norm erhält man für alle $x \in A$

$$\underbrace{\frac{\|\varphi(x) - \varphi(\tilde{x})\|}{\|x - \tilde{x}\|}}_{\|\varphi'(x)\|} \leq L < 1$$

$$\|\varphi'(x)\| < 1$$

(Engeln-Müllges, Niederdrenk, & Wodicka, Numerische Verfahren zur Lösung nichtlinearer Gleichungen, 2011) (Locher, Numerische Mathematik für Informtiker, 1992) (Freund & Hoppe, 2007)

[6] Und somit auch stetig

Proposition: Falls $\sup_{x\in A}\|\varphi'(x)\| < 1$, dann ist φ kontrahierend auf dem Intervall A.

Beweis: Setzte $L \coloneqq \sup_{x\in A}\|\varphi'(x)\|$. Nach den Bemerkungen oben ist

$$\|\varphi(x) - \varphi(\tilde{x})\| < L\|x - \tilde{x}\|$$

□

Satz 1.5: Die Gleichung $x = \varphi(x)$ besitzt in einem endlich abgeschlossenen Gebiet $A \subset \mathbb{R}^n$ höchstens eine Lösung, wenn die Funktion $\varphi(x)\colon A \to \mathbb{R}^n$ eine der folgende Bedingungen erfüllt:

(i)	$\|\varphi(x) - \varphi(\tilde{x})\| \le L\|x - \tilde{x}\|$	$0 \le L < 1$
(ii)	$\sup_{x\in A}\|\varphi'(x)\| < 1$	falls φ differenzierbar in A

Beweis: Angenommen es gibt A zwei Lösungen $\eta_1,\eta_2 \in A$ für die Gleichung $x = \varphi(x)$, sodass (i):

$$\eta_1 = \varphi(\eta_1) \quad \text{und} \quad \eta_2 = \varphi(\eta_2)$$

Nach der Lipschitzbedingung für φ muss nun gelten

$$\underbrace{\|\varphi(\eta_2) - \varphi(\eta_1)\|}_{\|\eta_2-\eta_1\|} \le L\|\eta_2 - \eta_1\|$$

$$\|\eta_2 - \eta_1\| \le L\|\eta_2 - \eta_1\|$$

$$\underbrace{(1 - L)}_{>0}\|\eta_2 - \eta_1\| \le 0$$

Da $1 - L > 0$ ist muss $\|\eta_2 - \eta_1\| \le 0$ sein. Die Norm $\|\cdot\|$bildet jedoch nicht in $\mathbb{R}^{n^-}$ ab, also muss $\|\eta_2 - \eta_1\| = 0$ gelten und daraus folgt, dass $\eta_1 = \eta_2$ ist. (Engeln-Müllges, Niederdrenk, & Wodicka, Numerische Verfahren zur Lösung nichtlinearer Gleichungen, 2011)

(ii) Nach der obigen Proposition ist (i) erfüllt und damit hat $\varphi(x) = x$ höchstens eine Lösung

□

Laut Satz 1.4 besitzt die Gleichung $x = \varphi(x)$ mindestens eine Lösung, laut Satz 1.5 besitzt sie maximal eine Lösung. Fasst man diese beiden Sätze nun zusammen erhält man folgendes:

Satz 1.6: Die Gleichung $x = \varphi(x)$ besitzt in einem endlich abgeschlossenen Gebiet $A \subset \mathbb{R}^n$ genau eine Lösung, wenn die Funktion $\varphi(x) : A \to \mathbb{R}^n$ die folgenden Bedingungen erfüllt:

(i) $\varphi(x) \in A \quad \forall x \in A$

(ii) $\|\varphi(x) - \varphi(\tilde{x})\| \leq L\|x - \tilde{x}\|$, für $0 \leq L < 1$ oder $\sup_{x \in A} \|\varphi'(x)\| < 1,$, falls φ differenzierbar in A.

1.3. Konvergenz

Bei der heuristischen Betrachtung in $\mathbb{R}$ des Konvergenzverhalten von Iterationsverfahren, wird klar wieso die Iterationsfolge $\{x_i\}$ mit der Vorschrift

$$x_{i+1} = \varphi(x_i)$$

genau dann konvergiert, wenn φ die Bedingungen des Satzes 1.6 erfüllen. Angenommen es gilt $\varphi: A \to A, A \subset \mathbb{R}$, also ist (i) aus Satz 1.6 erfüllt. Somit gibt insgesamt vier Fälle die betrachtet werden müssen.

1. Fall: $0 \leq \varphi'(x) < 1$, also φ ist eine konkave[7] Funktion im Intervall A. Bei Betrachtung von Abbildung 5 ist zu sehen, dass $\{x_i\}$ in diesem Fall monoton konvergiert.

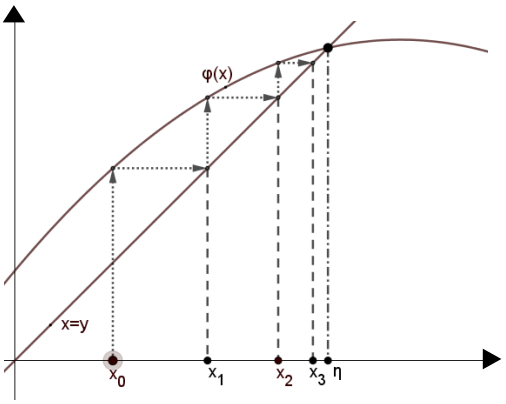

Abbildung 5

[7] Negativ gekrümmte

2. Fall: $-1 < \varphi'(x) \leq 0$.φ ist im Intervall A demnach eine konvexe[8] Funktion. In diesem Fall konvergiert die Iterationsfolge $\{x_i\}$ auch, allerdings alternierend (siehe Abbildung 6).

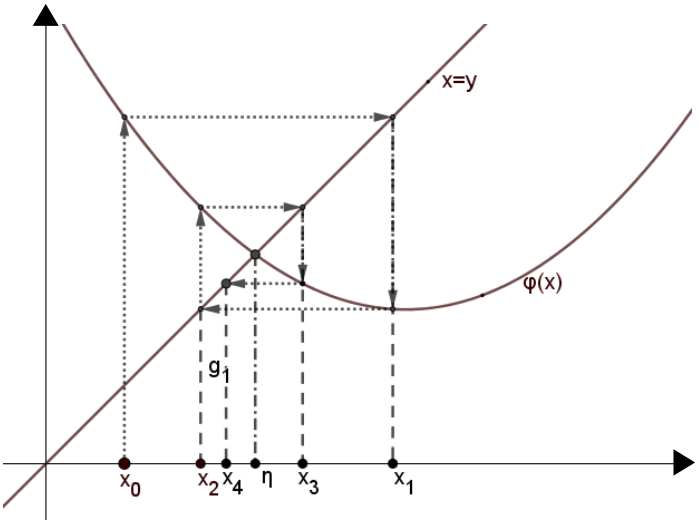

Abbildung 6

3. Fall: $1 < \varphi'(x)$. Die Abbildung 7 zeigt, dass sich die Glieder der Iterationsfolge $\{x_i\}$ bei größer werdendem i immer mehr von der Lösung η entfernen. In diesem Fall ist die Folge divergent.

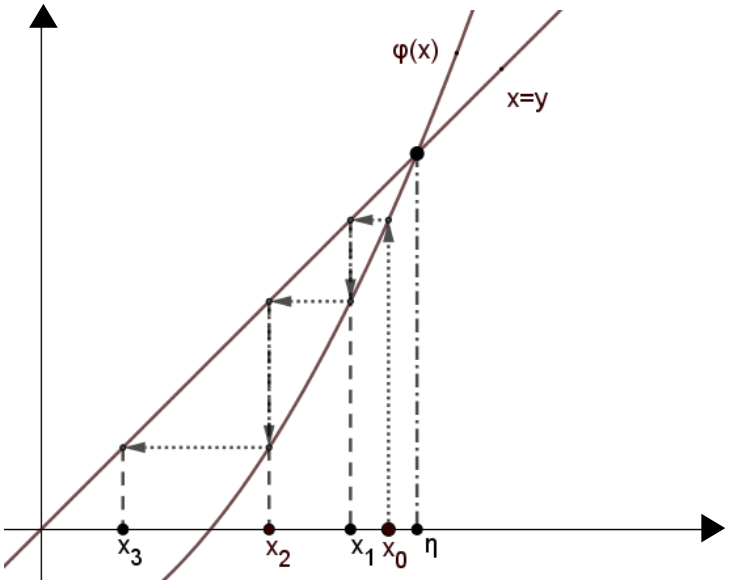

Abbildung 7

4. Fall: Wenn $\varphi' < -1$ im Intervall A gilt, dann ist die Folge divergent, wie in Abbildung 8 zu sehen ist.

[8] Positiv gekrümmt

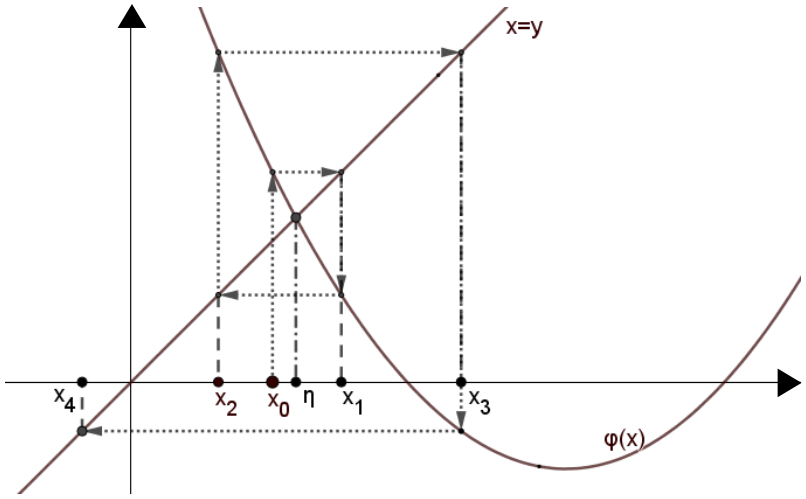

Abbildung 8

Um analytisch zu beweisen, dass $\{x_i\}$ genau dann konvergiert, wenn φ die Bedingungen aus Satz 1.6 erfüllt, zeigen wir, dass $\{x_i\}$ eine Cauchy-Folge ist. Dies ist ausreichend, da jede Cauchy-Folge in vollständigen Räumen konvergent ist (Brosowski & Kreß, 1975). Wir betrachten die Iterationsfolge in einem abgeschlossenen Gebiet A. Da $A \subset \mathbb{R}^n$ ist und $\mathbb{R}^n$ ein vollständiger Raum ist (Neunhäuserer, 2017), ist A durch die Abgeschlossenheit auch vollständig (Brosowski & Kreß, 1975).

Satz 1.7: Die Iterationsfolge $\{x_i\}$ mit der Vorschrift $x_{i+1} = \varphi(x_i)$ ist eine Cauchy-Folge und somit konvergent, wenn $\varphi: A \subset \mathbb{R}^n \to A$ eine kontrahierende Abbildung ist.

Beweis: Aus der Kontraktionseigenschaft

$$\|\varphi(x) - \varphi(\tilde{x})\| \leq L\|x - \tilde{x}\| \quad \text{mit} \quad 0 \leq L < 1;\ x, \tilde{x} \in A$$

und der Eigenschaft[9], das $\varphi(x) \in A$ ist ergibt sich eine Ungleichung der Art

$$\|x_{i+1} - x_i\| = \|\varphi(x_i) - \varphi(x_{i-1})\| \leq L \underbrace{\|x_i - x_{i-1}\|}_{\|\varphi(x_{i-1}) - \varphi(x_{i-2})\|} \leq L^2\|x_{i-1} - x_{i-2}\| \leq \cdots$$

$$\leq L^i\|x_1 - x_0\|.$$

Wendet man nun die Dreiecksungleichung an so ergibt sich für $0 \leq v < u;\ x_u, x_v \in \{x_i\}$

$$\|x_u - x_v\| \leq \|x_u - x_{u-1}\| + \cdots + \|x_{v+1} - x_v\|$$

[9] Selbstabbildungseigenschaft

$$\leq L^{u-1}\|x_1 - x_0\| + \cdots + L^v\|x_1 - x_0\|$$

$$= [L^{u-1} + L^{u-2} + \cdots + L^v]\|x_1 - x_0\|$$

$$= L^v[L^{u-1-v} + L^{u-2-v} + \cdots + 1]\,\|x_1 - x_0\|$$

$$= L^v \sum_{k=0}^{u-1-v} (L^k) \cdot \|x_1 - x_0\|$$

Der letzte Ausdruck beinhaltet die $u - 1 - v$te Partialsumme der geometrischen Reihe und da $0 \leq L < 1$ gilt erhalten wir

$$= L^v \cdot \frac{1 - L^{u-v}}{1 - L}\|x_1 - x_0\|$$

$$= \frac{L^v - L^u}{1 - L}\|x_1 - x_0\|$$

$$\leq \frac{L^v}{1 - L}\|x_1 - x_0\|$$

Nun muss nur noch ein passendes $v(\varepsilon) \in \mathbb{N}\backslash\{0\}$ gewählt werden sodass für alle $\varepsilon > 0$ gilt

$$\|x_u - x_v\| \leq \frac{L^v}{1 - L}\|x_1 - x_0\| < \varepsilon$$

Somit ist $\{x_i\}$ eine Cauchy-Folge und konvergent, wenn φ eine kontrahierende Abbildung ist.

□

(Opfer, 1993) (Locher, Numerische Mathematik für Informtiker, 1992) (Ortega & Rheinboldt, 1970)

Wenn $\varphi: A \subset \mathbb{R}^n \to A$ in A stetig partiell differenzierbar nach $x_j, j = 1, \dots, n$ ist, so kann die erste Ableitung von φ mithilfe der Jacobi-Matrix angegeben werden durch

$$J_\varphi \coloneqq \left(\frac{\partial \varphi_i}{\partial x_j}\right)_{\substack{i=1,\dots,n \\ j=1,\dots,n}} = \begin{pmatrix} \frac{\partial \varphi_1}{\partial x_1} & \frac{\partial \varphi_1}{\partial x_2} & \cdots & \frac{\partial \varphi_1}{\partial x_n} \\ \frac{\partial \varphi_2}{\partial x_1} & \frac{\partial \varphi_2}{\partial x_2} & \cdots & \frac{\partial \varphi_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \varphi_n}{\partial x_1} & \frac{\partial \varphi_n}{\partial x_2} & \cdots & \frac{\partial \varphi_n}{\partial x_n} \end{pmatrix}.$$

Damit φ eine geeignete Iterationsfunktion ist und das Verfahren konvergent ist muss nun

$$\|J_\varphi\| \leq L < 1$$

mit $0 \leq L < 1$ gelten. Mithilfe der Matrixnormen [10] können nun folgende Konvergenzkriterien für φ gefunden werden:

das Zeilensummenkriterium:

$$\|J_\varphi\|_\infty = \max_{\substack{i=1,\dots,n \\ x\in A}} \sum_{j=1}^{n} \left|\frac{\partial \varphi_i}{\partial x_j}\right| \leq L_\infty < 1$$

Das Spaltensummenkriterium

$$\|J_\varphi\|_1 = \max_{\substack{j=1,\dots,n \\ x\in A}} \sum_{i=1}^{n} \left|\frac{\partial \varphi_i}{\partial x_j}\right| \leq L_1 < 1$$

Das Kriterium von E. Schmidt und R. v. Mises:

$$\|J_\varphi\|_1 = \max_{x\in A} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{\partial \varphi_i}{\partial x_j}\right)^2 \right)^{\frac{1}{2}} \leq L_2 < 1$$

(Brosowski & Kreß, 1975) (Engeln-Müllges, Niederdrenk, & Wodicka, Numerische Verfahren zur Lösung nichtlinearer Gleichungen, 2011)

Beispiel 4: Im Beispiel 1 sind drei mögliche Iterationsvorschriften zur Ermittlung der Lösung der Gleichung $f(x) = x^2 - x - 1 = 0$ ermittelt worden. Um nun nicht alle durchrechnen zu müssen um ihr Konvergenzverhalten zu testen, prüfen wir ob $\varphi(x)$ in einem Intervall A eine kontrahierende Abbildung ist[11].

(i) $\quad x_{i+1} = \varphi(x_i) = x_i^2 - 1, \qquad x_i \in \mathbb{R}$

Das Intervall A wird definiert durch $A := [-1; 2]$. Jetzt muss $|\varphi'(x)| < 1$ für alle $x \in A$ sein. Da

$$\varphi'(x) := 2x$$

die erste Ableitung ist und für alle

[10] Vgl. Vorbereitung
[11] Bei der Wahl muss auf den Definitionsbereich von φ Rücksicht genommen werden

$$x \in A, |x| \geq \frac{1}{2} : |\varphi'(x)| \geq 1$$

gilt, ist φ in A keine kontrahierende Abbildung. Die erzeugte Folge müsste somit nicht konvergent sein und es kommt tatsächlich bei der Iteration zu keiner Annäherung an die Nullstelle.

(ii) $\quad x_{i+1} = \varphi(x_i) = \sqrt{x_i + 1} \qquad x_i \in \mathbb{R},\ x_i \geq -1$

$A \coloneqq [1; 2]$,

$$\varphi'(x) \coloneqq \frac{1}{2\sqrt{x+1}}$$

$$x \geq 1$$

$$0 < x$$

$$\frac{4}{4} = 1 < x + 1$$

$$\frac{1}{4} < \frac{1}{4}(x+1)$$

$$\frac{1}{2} < \frac{1}{2}\sqrt{x+1}$$

$$\frac{1}{2}\frac{1}{\underbrace{|\sqrt{x+1}|}_{\leq 0}} < \frac{1}{2}$$

$$\left|\frac{1}{2\sqrt{x+1}}\right| < \frac{1}{2}$$

Somit erfüllen alle $x \in A$ die Bedingung $\sup_{x \in A} \|\varphi'(x)\| < \frac{1}{2} < 1$ und das durch die Folge $\{x_i\}$ erzeugte Iterationsverfahren konvergiert gegen die Lösung in A.

(iii) $\quad x_{i+1} = \varphi(x_i) = 1 + \frac{1}{x_i} \qquad x_i \in \mathbb{R},\ x_i \neq 0$

$A \coloneqq [1{,}5; 2]$,

$$\varphi'(x) \coloneqq -\frac{1}{x^2}$$

$$|x| \geq 1{,}5 \geq \sqrt{2}$$

$$\frac{1}{|x|^2} < \frac{1}{2}$$

$$\left| \underbrace{\frac{1}{x^2}}_{\leq 0} \right| < \frac{1}{2}$$

Die Funktion φ ist demnach kontrahierend und $\{x_i\}$ konvergent auf A da die Bedingung $\sup_{x \in A} \|\varphi'(x)\| < \frac{1}{2} < 1$ für alle $x \in A$ erfüllt ist.

1.4. Abbruchbedingungen

Das Iterationsverfahren kann abgebrochen werden, wenn eine gewisse, manchmal vorgegebene Maximaldistanz zwischen Näherungslösung und der exakten Lösung η unterschritten wird.

Diese Maximaldistanz kann entweder durch den relativen Fehler oder den absoluten Fehler gegeben sein. Allgemein kann man sagen, dass ein Iterationsverfahren nach dem $i$ten Schritt abgebrochen werden kann, wenn für $\varepsilon_1, \varepsilon_2, \varepsilon_3 > 0$ folgendes gilt:

(i) $\|x_i - x_{i-1}\| \leq \varepsilon_1 \|x_i\|$ oder $\|x_i - x_{i-1}\| \leq \varepsilon_2$

(ii) $\|f(x_i)\| \leq \varepsilon_3$

(Engeln-Müllges, Niederdrenk, & Wodicka, Numerik-Algorithmen, 2005)

(Friedrich & Pietschmann, 2010)

1.5. Fehlerabschätzung

Die einzelnen Glieder der Iterationsfolge sind immer exaktere Näherungslösungen für η. Jetzt stellt sich die Frage, wie viele Iterationsschritte benötigt werden um die gewünschte Näherungslösung zu erhalten. Anders gesagt ist es wie nah die Näherungslösung an η ist. Diese Frage kann durch die Fehlerabschätzung geklärt werden. Die Näherungslösung $x_i \in A \subset \mathbb{R}^n$ unterscheidet sich von η um einen absoluten Fehler Δ_i der durch

$$\Delta_i = \eta - x_i$$

definiert wird. Wenn dieser Fehler kleiner als eine vorgegebene Schranke $\varepsilon > 0$ ist, so kann das Iterationsverfahren abgebrochen werden. Die Wahl der Schranke bestimmt

somit die Qualität[12] der Näherungslösung. Je kleiner ε gewählt wird umso kleiner ist auch der Fehler Δ denn es gilt

$$\|\Delta_i\| \leq \varepsilon.$$

Wir erhalten eine Fehlerabschätzung indem wir die Ungleichung aus dem Beweis das $\{x_i\}$ eine Cauchy-Folge ist nochmals betrachten. Sei $u \geq i$, dann gilt

$$\|x_u - x_i\| \leq \frac{L^i}{1-L}\|x_1 - x_0\|$$

Wenn nun i fest ist und u gegen Unendlich strebt ($u \to \infty$) mit $\lim_{u\to\infty} x_u = \eta$, so erhält man

$$\|\Delta_i\| = \|\eta - x_i\| \leq \frac{L^i}{1-L}\|x_1 - x_0\| < \varepsilon_1 \quad \varepsilon_1 > 0.$$

Diese Ungleichung liefert die a priori-Fehlerabschätzung. Mithilfe dieser Formel kann man bei gegebener Schranke $\varepsilon_1 > 0$ abschätzen, wie groß der höchste Fehler der $i$ten Näherung, $i \geq 2$ ist, wenn der Startwert x_0 und die erste Näherungslösung x_1 bekannt sind (Törnig & Spellucci, 1988). Wenn eine bestimmte Qualität der Näherungslösung schon bekannt oder sogar gegeben ist, dann kann mittels dieser Abschätzung auch eine obere Schranke $\varepsilon_1 > 0$ der Anzahl der benötigten Iterationsschritte bestimmt werden. Die Anzahl kann durch umformen der a priori-Fehlerabschätzung erfolgen:

$$\frac{L^i}{1-L}\|x_1 - x_0\| < \varepsilon$$

$$L^i < \frac{\varepsilon(1-L)}{\|x_1 - x_0\|}$$

$$i \ln L < \ln \frac{\varepsilon(1-L)}{\|x_1 - x_0\|}$$

$$i \ln L < \ln \varepsilon + \ln(1-L) - \ln\|x_1 - x_0\|$$

$$i\,(-\ln L) > -\ln \varepsilon - \ln(1-L) + \ln\|x_1 - x_0\|$$

$$i > \frac{-\ln \varepsilon - \ln(1-L) + \ln\|x_1 - x_0\|}{-\ln L}$$

[12] Genauigkeit

Somit müssen mindestens i Iterationsschritte erfolgen um eine vorgegebene Schranke ε zu unterschreiten und der Rechenaufwand kann im Vorhinein bestimmt werden (Brosowski & Kreß, 1975) (Friedrich & Pietschmann, 2010).

Die a priori-Fehlerabschätzung ist in der Regel nicht sehr genau, also eine eher grobe Fehlerabschätzung. Eine bessere Abschätzung des Restfehlers erhält man durch die Differenz der beiden letzten errechneten Näherungswerte während der Iteration.

Wir betrachten die Ungleichung

$$\|x_{i+1} - x_i\| \leq L\|x_i - x_{i-1}\| \quad 0 \leq L < 1,$$

die aus der Kontraktionseigenschaft und Selbstabbildungseigenschaft von φ folgt[13]. Für ein festes i und ein $u > i+1, ; x_i, x_u \in A \subset \mathbb{R}^n$ folgt

$$\|x_u - x_i\| = \|x_u - x_{u-1} + x_{u-1} - x_{u-2} + \cdots + x_{i+1} - x_i\|$$

$$\leq \|x_u - x_{u-1}\| + \|x_{u-1} - x_{u-2}\| + \cdots + \|x_{i+1} - x_i\|.$$

Weiters gilt

$$\|x_u - x_{u-1}\| \leq L\|x_{u-1} - x_{u-2}\| \leq L^2\|x_{u-2} - x_{u-3}\| \leq \cdots \leq L^{u-i}\|x_i - x_{i-1}\|$$

Betrachtet man diese beiden Ungleichungen nun gemeinsam ergibt sich

$$\|x_u - x_i\| \leq \|x_i - x_{i-1}\|\left(L^{u-1} + L^{u-i-1} + \cdots + L\right)$$

$$= L\|x_i - x_{i-1}\| \underbrace{(1 + \cdots + L^{u-i-1})}_{geometrische\ Reihe}$$

$$= L\frac{1 - L^{u-i}}{1 - L}\|x_i - x_{i-1}\|$$

Also gilt die Ungleichung

$$\|x_u - x_i\| \leq \frac{L}{1 - L}\|x_i - x_{i-1}\|.$$

Wenn nun v fest ist und u gegen Unendlich strebt ($u \to \infty$) mit $\lim_{u\to\infty} x_u = \eta$, so erhält man

$$\|\Delta_i\| = \|\eta - x_i\| \leq \frac{L}{1 - L}\|x_i - x_{i-1}\| < \varepsilon_2.$$

[13] Vgl. Beweis von Satz 1.7

Diese Ungleichung wird als a posteriori-Fehlerabschätzung bezeichnet. Wie oben schon erwähnt wird diese Abschätzung während oder nach der Iteration verwendet, da x_i und x_{i-1}bekannt sein müssen.

Da die a posteriori- Fehlerabschätzung genauer ist als die a priori gilt zusätzlich für $\varepsilon_1, \varepsilon_2 > 0$

$$\varepsilon_2 < \varepsilon_1.$$

(Engeln-Müllges, Niederdrenk, & Wodicka, Numerik-Algorithmen, 2005) (Friedrich & Pietschmann, 2010) (Locher, Numeriche Mathematim für Informatiker, 1992)

Diese ganzen Erkenntnisse lassen sich mit einem der wichtigsten Sätze der Analysis zusammenfassen, dem Banach'schen Fixpunktsatz.

Satz 1.8 (Banach'sche Fixpunktsatz): Sei $\mathbb{R}^n$ ein Raum mit gegebener Norm $\|\cdot\|$ und $\varphi: A \subset \mathbb{R}^n \to \mathbb{R}^n$ eine Abbildung mit den Eigenschaften

(i) $A \subset \mathbb{R}^n$ ist nichtleer und abgeschlossen

(ii) φ bildet auf sich selbst ab, also $\varphi(A) \subset A$

(iii) φ ist bezüglich $\|\cdot\|$ kontrahierend, also mit der Lipschitzkonstante $0 \leq L < 1$ oder $\sup_{x \in A} \|\varphi'(x)\| < 1$.

Dann gibt es ein eindeutig bestimmtes $\eta \in A$ mit $\varphi(\eta) = \eta$ und es gilt:

(I) Die Folge $\{x_i\}$ mit der Vorschrift $x_{i+1} = \varphi(x_i), i = 0,1,2, \ldots$ konvergiert für jeden Startwert $x_0 \in A$ gegen $\eta \in A$, welches der einzige Fixpunkt von φ in A ist.

$$\lim_{i \to \infty} x_{i+1} = \eta$$

(II) Die a priori Fehlerabschätzung

$$\|\eta - x_i\| \leq \frac{L^i}{1-L} \|x_1 - x_0\|, \quad i = 0,1,2, \ldots$$

(III) Die a posteriori Fehlerabschätzung

$$\|\eta - x_i\| \leq \frac{L}{1-L} \|x_i - x_{i-1}\|, \quad i = 0,1,2, \ldots$$

1.6. Konvergenzordnung

Die Konvergenzordnung ist eine Größe, mithilfe derer der benötigte Rechenaufwand eines Verfahren abgeschätzt werden kann. Sie gibt die Geschwindigkeit an mit der sich die Glieder einer Iterationsfolge an die gesuchte Lösung η annähern. Deswegen ist die Konvergenzordnung ein gutes Kriterium zur Beurteilung eines Iterationsalgorithmus und kann herangezogen werden, wenn mehrere Algorithmen zur Auswahl stehen.

Satz 1.9: Sei $\varphi \in A \subset \mathbb{R}^n \to \mathbb{R}$ eine kontrahierende Abbildung in A mit dem Fixpunkt $\eta \in A$ und $p \in \mathbb{R}$ eine reelle Zahl. Dann heißt das dazugehörige Iterationsverfahren p-ter Ordnung, wenn es eine Zahl $C \geq 0, C \in \mathbb{R}$ gibt sodass

$$\|\varphi(x) - \eta\| \leq C\|x - \eta\|^p.$$

Durch Ersetzten von x durch x_i erhält man die Ungleichung

$$\left\| \underbrace{\varphi(x_i)}_{x_{i+1}} - \eta \right\| \leq C\|x_i - \eta\|^p$$

$$\|x_{i+1} - \eta\| \leq C\|x_i - \eta\|^p$$

Wenn $p = 1$ ist liegt eine lineare Konvergenz vor, bei $p = 2$ eine quadratische, $p = 3$ eine kubische. Sowohl p als auch L beeinflussen die Konvergenzgeschwindigkeit, denn je kleiner die Lipschitzkonstante und je größer p ist, umso schneller konvergiert das Verfahren. Aus dem Satz von Banach folgt demnach, dass jedes Iterationsverfahren, für das eine kontrahierende Schrittfunktion φ gefunden werden kann, mindestens linear konvergiert. (Freund & Hoppe, 2007).

Ein Verfahren, dass nur innerhalb eines Gebietes $A \subset \mathbb{R}^n$ konvergiert heißt lokal konvergent, wenn allerdings $A = \mathbb{R}^n$ gilt, so heißt es global konvergent.

Im eindimensionalen Raum, also $A \subset \mathbb{R}$ ist die Ordnung eines Iterationsverfahren leicht festzustellen, wenn die dazugehörige Schrittfunktion φ in einer Umgebung um $D(\eta)$ von $x = \eta$ genügend oft differenzierbar ist.

Satz 1.10: Sei $\varphi: A \subset \mathbb{R} \to A$ die Schrittfunktion einer Iteration und p- mal stetig differenzierbar. Weiteres gilt für die Iterationsfunktion $\lim_{i\to\infty} x_i = \eta$ und

$$\varphi(\eta) = \eta$$

$$\varphi'(\eta) = \varphi''(\eta) = \cdots = \varphi^{(p-1)}(\eta) = 0$$

$$\varphi^{(p)}(\eta) \neq 0.$$

Das Iterationsverfahren $x_{i+1} = \varphi(x_i)$ konvergiert dann mit der Ordnung p und es gilt:

$$C = \frac{1}{p!}\left|\varphi^{(p)}(\eta)\right| \leq \frac{1}{p!}\max_{x\in A}\left|\varphi^{(p)}(x)\right| \leq C_1$$

Falls $p = 1$ ist, dann gilt zusätzlich $C = \sup_{x\in A}\|\varphi'(x)\| < 1$

Beweis: Um den Satz 1.10 zu beweisen entwickeln wir φ an der Stelle η.

$$\underbrace{\varphi(x_i)}_{x_{i+1}} = \underbrace{\varphi(\eta)}_{\eta} + (x_i - \eta)\varphi'(\eta) + \frac{1}{2}(x_i - \eta)^2\varphi''(\eta) + \cdots$$

Durch Umformen der Taylorentwicklung erhält man:

$$x_{i+1} - \eta = (x_i - \eta)\underbrace{\varphi'(\eta)}_{=0} + \frac{1}{2}(x_i - \eta)^2\underbrace{\varphi''(\eta)}_{=0} + \cdots$$

$$x_{i+1} - \eta = \frac{(x_i - \eta)^p}{p!}\varphi^{(p)}(\eta) + O((x_i - \eta)^{p+1})$$

$$\frac{x_{i+1} - \eta}{(x_i - \eta)^p} = \frac{1}{p!}\varphi^{(p)}(\eta) + O((x_i - \eta))$$

Lässt man i jetzt gegen Unendlich streben, so erhält man

$$\lim_{i\to\infty}\left|\frac{x_{i+1} - \eta}{(x_i - \eta)^p}\right| = \frac{1}{p!}\left|\varphi^{(p)}(\eta)\right| = C$$

mit

$$C \leq \frac{1}{p!}\max_{x\in A}\left|\varphi^{(p)}(x)\right| \leq C_1$$

Wenn nun $p = 1$ gilt, dann ist das Iterationsverfahren erster Ordnung also linear und wegen der Lipschitzbedingung für φ gilt

$$C = \sup_{x \in A} \|\varphi'(x)\| \leq L < 1$$

□

(Engeln-Müllges, Niederdrenk, & Wodicka, Numerische Verfahren zur Lösung nichtlinearer Gleichungen, 2011) (Freund & Hoppe, 2007) (Törnig & Spellucci, 1988)

1.6. Praktische Anwendung des Algorithmus

Das Ziel ist es die Lösung der Gleichung $f(x) = 0$ mithilfe des allgemeinen Iterationsverfahren zu finden. Dazu benötigen wir sieben Schritte

1. Schritt: Ein Gebiet A bestimmen in dem mindestens eine Nullstelle von f liegt.
2. Schritt: Auffinden einer passenden Schrittfunktion $\varphi(x)$, sodass die Gleichungen $f(x) = 0$ und $x = \varphi(x)$ äquivalent sind
3. Schritt: Prüfen ob die Funktion φ eine kontrahierende Abbildung ist, insbesondere ob $\varphi(A) \in A$.
4. Schritt: Eine Iterationsvorschrift $x_{i+1} = \varphi(x_i)$ bestimmen und einen geeigneten Startwert x_0 wählen.
5. Schritt: Berechnung der Glieder der Iterationsfolge $\{x_i\}$ laut Iterationsvorschrift
6. Schritt: Die Iteration abbrechen, wenn mindestens eine der Abbruchbedingungen erfüllt ist.
7. Schritt: Fehlerabschätzung

(Engeln-Müllges, Niederdrenk, & Wodicka, Numerische Verfahren zur Lösung nichtlinearer Gleichungen, 2011)

Meistens ist Schritt 5 und 6 in einem durchführbar. Eignen sich mehrere Funktionen als Iterationsvorschrift, ist es meistens von Vorteil mithilfe einer Fehlerschranke ε den Rechenaufwand pro Iterationsvorschrift zu bestimmen.

Beispiel 5: Gegeben ist eine Funktion $f: [0; \infty] \to \mathbb{R}, f(x) \coloneqq e^x - \sqrt{x} - 3$. Gesucht ist eine Näherungslösung der Nullstelle von f, die einen absoluten Fehler maximal $\varepsilon = 1 * 10^{-7}$ hat.

1. Schritt:

$$f(1) = -1{,}28 < 0$$
$$f(2) = \quad 2{,}97 > 0$$

Da f stetig ist und zwischen $f(1)$und $f(2)$ ein Vorzeichenwechsel ist muss es ein $x \in [1;2]$ geben mit $f(x) = 0$. Deswegen besitzt die Funktion in [1,2] mindestens eine Nullstelle und das Intervall A ist gefunden.

$$A \coloneqq [1;2]$$

2. Schritt: Die erste mögliche Schrittfunktion erhält man durch Auflösen der Gleichung

$$0 = e^x - \sqrt{x} - 3$$

nach dem x unter der Wurzel.

$$\varphi_1(x) = (e^x - 3)^2$$

3. Schritt: Damit φ_1 eine kontrahierende Abbildung ist muss φ_1 in A selbstabbildend[14] sein und $|\varphi_1(x)'| < 1, x \in A$.
Es muss also $\varphi(1) \in A$ gelten, $\varphi(1) = 0{,}08 \notin A$. Somit ist diese Schrittfunktion nicht geeignet, da die erzeugte Iterationsfolge nicht zwingend konvergiert.

Also zurück zu Schritt 2. Eine andere Möglichkeit wäre die Auflösung nach dem x in der Hochzahl, so ergibt sich

$$\varphi_2 = \ln(\sqrt{x} + 3)$$

Jetzt kann Schritt 3 erneut durchgeführt werden.

$$\varphi_2(1) = 1{,}39 \in A$$
$$\varphi_2(2) = 1{,}48 \in A$$
$$\varphi_2'(x) = \frac{1}{2}\,\frac{1}{\sqrt{x}\,(\sqrt{x} + 3)}$$

Da für alle $x \in A$ $\varphi_2'(x) > 0$ ist, ist die Funktion φ in A streng monoton steigend und somit ist sind alle $\varphi(x) \in A$ und φ selbstabbildend.
Wenn nun auch für $\sup_{x \in A} \|\varphi_2'(x)\| < 1$ gilt, ist φ_2 eine geeignete Schrittfunktion

[14] $\forall x \in A: \varphi(x) \in A$

$$\left|\frac{1}{2}\frac{1}{\sqrt{x}\,(\sqrt{x}+3)}\right| \leq \frac{1}{2}\frac{1}{\min\limits_{x\in A}\sqrt{x}\,\min\limits_{x\in A}(\sqrt{x}+3)} = \frac{1}{2}\frac{1}{\sqrt{1}(\sqrt{1}+3)} = \frac{1}{8} < 1$$

Somit sind alle Bedingungen erfüllt, sodass die erzeugte Folge in A konvergiert und die Schrittfunktion ist definiert durch

$$\varphi: A \to A, \;\; \varphi = \ln(\sqrt{x}+3)$$

4. Schritt: Iterationsvorschrift

$$x_{i+1} = \varphi(x_i) = \ln\left(\sqrt{x_i}+3\right), \qquad i = 0,1,2,\ldots$$

Als Startwert x_0 wird 1,5 gewählt

5. Schritt: Die Glieder werden nun solange berechnet bis $|x_i - x_{i-1}| \leq 1 * 10^{-7}$ gilt

i	x_i	$\lvert x_i - x_{i-1}\rvert$
0	1,5	
1	1,440958874	.
2	1,435179631	.
3	1,434605799	.
4	1,434548741	.
5	1,434543066	0,000018081
6	1,434542502	0,000000564
7	1,434542446	0,000000056

Tabelle 3

Bei $i = 7$ ist der absolute Fehler kleiner als die vorgegebene Schranke ε.

6. Schritt: Nach Berechnung des Gliedes x_7 kann die Iteration abgebrochen werden da

$$0{,}000000056 = 0{,}56 * 10^{-7} \leq 1 * 10^{-7}.$$

7. Schritt: Wir benutzen die a posteriori Fehlerabschätzung, da diese genauer ist als die a priori. Zuerst wird ein L bestimmt mit $0 \leq L = \max\limits_{x\in A}\|\varphi'(x)\| < 1$.

$$L = \max_{x \in A} \left| \frac{1}{2} \frac{1}{\sqrt{x}\,(\sqrt{x}+3)} \right| = \frac{1}{2} \frac{1}{\min\limits_{x \in A} \sqrt{x} \min\limits_{x \in A} (\sqrt{x}+3)} = \frac{1}{8}$$

Jetzt kann durch Einsetzten in die Formel der Fehler geschätzt werden.

$$\begin{aligned} |\eta - x_7| &\leq \frac{L}{1-L} |x_7 - x_6| \leq \frac{\frac{1}{8}}{\frac{7}{8}} |0{,}000000056| \\ &\leq 0{,}142857143 \cdot 0{,}000000056 \leq 0{,}000000008 \\ &\leq 0{,}8 * 10^{-8}, \end{aligned}$$

Somit ist die gesuchte Näherungslösung $\eta \approx 1{,}434542446$ mit einem maximalen Fehler von $0{,}8 * 10^{-8}$, wobei die Näherung auf die siebente Nachkommastelle genau ist. In Abbildung 9 sind die Funktionen f und φ dargestellt, sowie die erste Mediane $x = y$. Das graue Quadrat zeigt das Intervall A an. Die Abbildung 9 verdeutlicht nochmal den Zusammenhang zwischen dem Fixpunkt von φ und der Nullstelle von f.

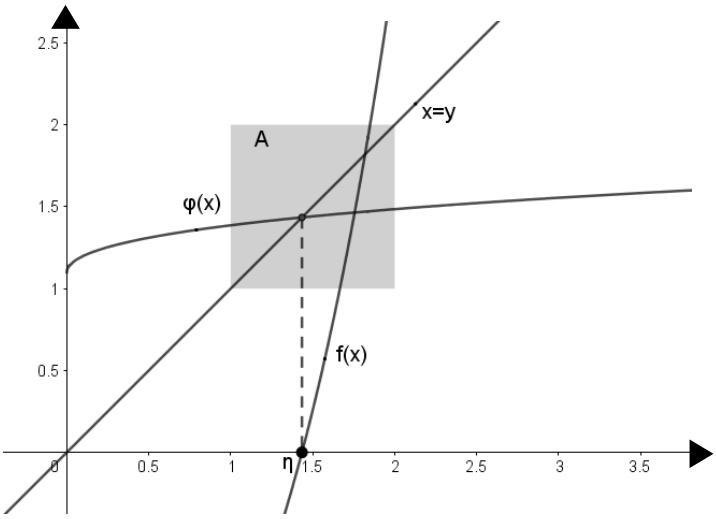

Abbildung 9

2. Regula falsi

Dieses Iterationsverfahren gehört zu den Einschlussverfahren oder Intervallschachtelungsverfahren und ist nur im eindimensionalen Raum anwendbar. Das Ziel ist wie bei jedem Iterationsverfahren die Auffindung der Lösung $\eta \in \mathbb{R}$ der Gleichung $f(x) = 0$, wobei $f: A \subset \mathbb{R} \to \mathbb{R}$ eine stetige Funktion ist und in A monoton. Es wird ein Intervall $A \coloneqq [a; b]$ so angenommen, dass mindestens eine Nullstelle der Funktion $f: A \subset \mathbb{R} \to \mathbb{R}$ in diesem Intervall liegt. Dies ist genau dann der Fall, wenn $f(a)f(b) < 0$ ist. Dieses Intervall A wird als Einschlussintervall bezeichnet. Der x-Wert der Nullstelle der Sekante von f durch die Punkte $P\big(a|f(a)\big)$ und $Q\big(b|f(b)\big)$ sei mit c bezeichnet und ist ein Element von A. Somit kann A in zwei Teilintervalle $[a; c]$ und $[c; b]$ aufgeteilt werden. Eines dieser Teilintervalle ist wiederum ein Einschlussintervall, falls die Intervallgrenzen ungleich η sind. Wiederholt man die Aufspaltung eines Einschlussintervalls immer wieder, so können fortlaufend kleinere Einschlussintervalle gebildet werden, die die Lösung η beinhalten. Wenn $c = \eta$ ist, also $f(c) = 0$ gilt, kann das Verfahren abgebrochen werden, da die Lösung gefunden ist.

2.1. Herleitung der Formel

Sei $f: A[a; b] \subset \mathbb{R} \to \mathbb{R}$ eine stetige Funktion und $f(\eta) = 0$ für $a < \eta < b$ mit $f(a)f(b) < 0$. Jetzt definieren wir $x_1 = a, x_2 = b, x_3 = c, f(x_1) = f(a), f(x_2) = f(b),$ weiteres ist s die Sekante von f in den Punkten $P\big(x_1|f(x_1)\big), Q\big(x_2|f(x_2)\big)$.

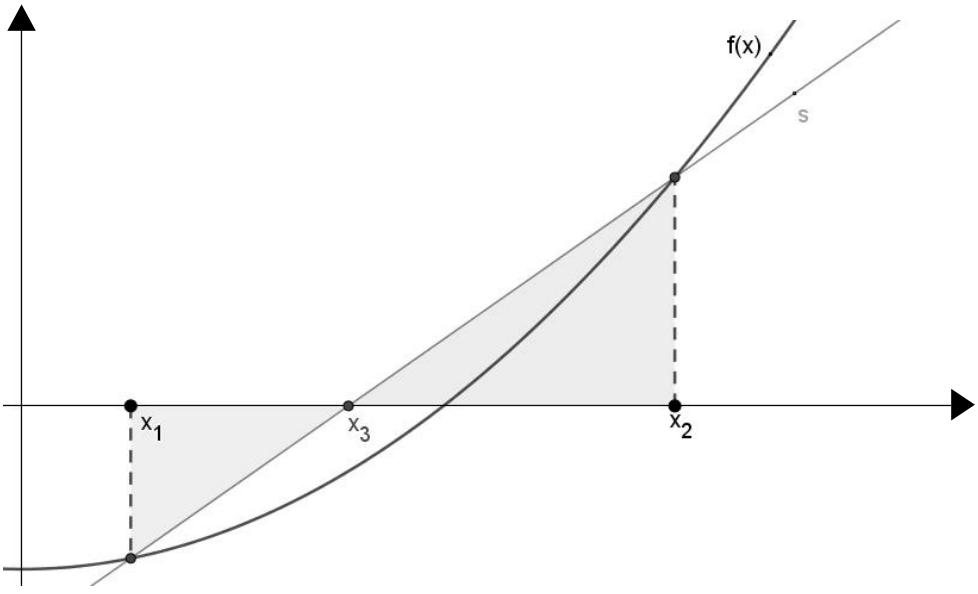

Abbildung 10

Die Sekante $s(x) = kx + d$. Nun gilt $s(x_2) = f(x_2)$ und

$$k = \frac{\Delta y}{\Delta x} = \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

$$f(x_2) = kx_2 + d = \frac{f(x_2) - f(x_1)}{x_2 - x_1} x_2 + d$$

$$d = f(x_2) - \frac{f(x_2) - f(x_1)}{x_2 - x_1} x_2$$

und somit

$$s(x) = \frac{f(x_2) - f(x_1)}{x_2 - x_1} x + f(x_2) - \frac{f(x_2) - f(x_1)}{x_2 - x_1} x_2$$
$$= \frac{f(x_2) - f(x_1)}{x_2 - x_1} (x - x_2) + f(x_2).$$

s besitzt nun im Intervall $A[x_1; x_2]$ eine Nullstelle mit $s(x_3) = 0$ also gilt

$$0 = \frac{f(x_2) - f(x_1)}{x_2 - x_1} (x_3 - x_2) + f(x_2)$$

$$-\frac{f(x_2)(x_2 - x_1)}{f(x_2) - f(x_1)} = (x_3 - x_2)$$

$$x_3 = x_2 - \frac{f(x_2)}{f(x_2) - f(x_1)} (x_2 - x_1)$$

Jetzt kann A in die Teilintervalle $[x_1; x_3]$ und $[x_3; x_2]$ unterteilt werden, wobei eines davon das neue Einschlussintervall ist. Somit erhält man zwei Fälle

1. Fall: $[x_1; x_3]$ ist Einschlussintervall, also $f(x_1)f(x_3) < 0$ und es gilt

$$x_4 = x_3 - \frac{f(x_3)}{f(x_3) - f(x_1)} (x_3 - x_1)$$

2. Fall: $[x_3; x_2]$ ist Einschlussintervall, also $f(x_3)f(x_2) < 0$ und es gilt

$$x_4 = x_2 - \frac{f(x_2)}{f(x_2) - f(x_3)} (x_2 - x_3)$$

Eine allgemeine Formel für dieses Verfahren ist recht schwer anzugeben, da bei jedem Schritt ein neues Intervall um die Lösung η gebildet wird. Allerdings kann man eine allgemeine Formel zur Berechnung einer Näherungslösung x von η in einem geschlossenen Einschlussintervall, welches a als unterer Grenze und b als oberer Grenze hat, angeben

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

x ist hier wohldefiniert da $f(b)f(a) < 0$ ist und es gilt entweder $f(x)f(a) < 0$ oder $f(b)f(x) < 0$.

(Locher, Numerische Mathematik für Informtiker, 1992) (Wille, 1976) (Engeln-Müllges, Niederdrenk, & Wodicka, Numerische Verfahren zur Lösung nichtlinearer Gleichungen, 2011)

Die berechneten x bilden eine Folge $\{x_i\}$, die als Regula falsi-Folge bezeichnet. Jetzt stellt sich wieder die Frage ob diese gebildete Folge auch wirklich immer gegen die Lösung η konvergiert und ob η die einzige Lösung ist.

2.2. Existenz und Eindeutigkeit der Lösung

Wie oben erwähnt, muss das erste Einschlussintervall so gewählt werden, dass die Funktionswerte der Intervallgrenzen unterschiedliche Vorzeichen haben. Der Grund dafür ist der Zwischenwertsatz. Dieser besagt, dass die stetige Funktion f in einem offenen Intervall $(a; b)$ mindestens eine Nullstelle in $[a; b]$ besitzt, wenn $f(a) \cdot f(b) < 0$ gilt. Um nun sicherzustellen, dass η die einzige Lösung in $[a; b]$ ist muss die Funktion in diesem Intervall monoton sein. Demnach muss für alle $x \in [a; b]$ $f'(x) < 0$ oder $f'(x) > 0$ gelten

2.3. Konvergenz

Satz 2.1: Sei $f: A[x_1; x_2] \subset \mathbb{R} \to \mathbb{R}$ eine stetige Funktion mit $f(x_1)f(x_2) < 0$, dann konvergiert die Regula falsi-Folge $\{x_i\}$ in A gegen die Nullstelle der Funktion.

Beweis: O.B.d.A sei $f(x_1) < 0 < f(x_2)$. Der Satz ist bewiesen, wenn für ein Glied x_l der Regula falsi-Folge $f(x_l) = 0$ gilt. Wir nehmen nun an, dass $f(x_i) \neq 0$, für alle $i = 1,2,3, \ldots$. Die Folgenglieder sind nach Konstruktionsvorschrift immer paarweise verschieden und lassen sich so in zwei verschiedenen Mengen einteilen

$$M_- = \{x_n | f(x_n) < 0\}$$

$$M_+ = \{x_k | f(x_k) > 0\}$$

wobei $x_n < x_k$, da M_+ rechts von M_- liegt. Weiters gilt für alle $i < j$

$$x_i < x_j \quad \text{falls } x_i, x_j \in M_-$$

$$x_i > x_j \quad \text{falls } x_i, x_j \in M_+$$

Die Glieder der Menge M_- bilden also eine monoton steigende Folge und die der Menge M_+ eine monoton fallende Folge. Nun gibt es zwei Fälle die betrachtet werden müssen, da die beidem Mengen entweder endlich oder unendlich sein können.

1.Fall: Angenommen M_+ist eine endliche und M_- eine unendliche Menge. Da die Folge der Glieder aus der Menge M_- sowohl monoton steigend als auch beschränkt ist, muss die Folge $\{x_n\}, x_n \in M_-$ gegen einen Grenzwert $\eta < x_K$ konvergieren. Der Fall $x_K = \eta$ kann nicht eintreten, da $f(x_K) > 0$ und $f(\eta) \leq 0$ sein muss.

Die endliche Menge M_+besitzt ein Element x_k, welches den größten Index $k = K$ hat. Setzt man dies jetzt in die Konstruktionsformel der Regula falsi-Folge ein so erhält man für alle $n > K$

$$x_{n+1} = x_n - \frac{f(x_n)}{f(x_n) - f(x_K)}(x_n - x_K)$$

Lässt man nun n gegen Unendlich streben, so erhält ergibt sich die Gleichung

$$\underbrace{\lim_{n\to\infty} x_{n+1}}_{\eta} = \underbrace{\lim_{n\to\infty} x_n}_{\eta} - \lim_{n\to\infty} \frac{f(x_n)}{f(x_n) - f(x_K)}(x_n - x_K)$$

$$\eta = \eta - \frac{f(\eta)}{f(\eta) - f(x_K)}(\eta - x_K)$$

Um Gleichheit zu erlangen muss $f(\eta) = 0$ gelten und somit ist die Konvergenz für diesen Fall gegeben.

2.Fall: Seien sowohl M_+ als auch M_- unendliche Mengen. Da die Regula falsi Folgenglieder immer abwechselnd M_+ und M_- zugeordnet werden kann, entstehen endliche Teilfolgen der Regula falsi-Folge.

$$\underbrace{x_0}_{\in M_-}, \underbrace{x_1, x_2, \dots, x_{k_1}}_{\in M_+}, \underbrace{x_{k_1+1}, \dots, x_{n_1}}_{\in M_-}, \underbrace{x_{n_1+1}, \dots, x_{k_2}}_{\in M_+}, \underbrace{x_{k_2+2}, \dots, x_{n_2}}_{\in M_-}, \dots$$

Innerhalb der Menge M_+ gibt es ein Element a_+, welches das Infimum dieser Menge ist, und analog dazu gibt es Supremum a_- der Menge M_- . Es gilt

$$a_+ := \underbrace{\inf}_{x_k \in M_+} x_k\,, \quad f(a_+) \geq 0$$

$$a_- := \underbrace{\sup}_{x_n \in M_-} x_n\,, \quad f(a_-) \leq 0$$

Weil nun M_- von links gegen a_- und M_+ von rechts gegen a_+ strebt, wäre die Konvergenz der Regula falsi Folge gegen die Lösung der Gleichung $f(\eta) = 0$ bewiesen, wenn $a_- = a_+$ ist, da $f(a_-) \leq 0$ und $f(a_+) \geq 0$ gilt, und somit $f(a_-) = f(a_+) = 0$.

Wir nehmen nun an, dass $a_- \neq a_+$ ist und beweisen die Konvergenz indirekt. Somit muss $a_- < a_+$ sein. Für $x_n \in M_-$ und $x_k \in M_+$ gilt nun laut Konstruktionsvorschrift die Gleichung

$$x_{n_i+1} = x_{n_i} - \frac{f(x_{n_i})}{f(x_{n_i}) - f(x_{k_i})}(x_{n_i} - x_{k_i}).$$

Durch Umformen ergibt sich

$$|f(x_{n_i})| = \frac{|x_{n_i+1} - x_{n_i}|}{|x_{k_i} - x_{n_i+1}|}|f(x_{k_i})|.$$

Da $x_{n_i} < a_- < a_+ < x_{k_i}$ und x_{n_i+1} bereits ein Element der Menge M_+ ist, gilt

$$0 < a_+ - a_- < |x_{n_i+1} - x_{n_i}|.$$

Weil die Elemente in M_+ eine konvergente Folge bilden ergibt sich für groß genug gewählte $i > N$ die Ungleichung

$$|x_{k_i} - x_{n_i+1}| < \frac{a_+ + a_-}{2}.$$

Setzt man nun die beiden obigen Ungleichen in die Gleichung aus der Konstruktionsvorschrift ein so ergibt sich

$$|f(x_{n_i})| = \underbrace{\overbrace{\frac{|x_{n_i+1} - x_{n_i}|}{|x_{k_i} - x_{n_i+1}|}}^{>a_+ - a_-}}_{>\frac{a_+ + a_-}{2}} |f(x_{k_i})|$$

$$|f(x_{n_i})| > 2|f(x_{k_i})| .$$

Mithilfe der Konstruktionsvorschrift erhält man allerdings auch die Gleichung

$$x_{k_i+1} = x_{k_i} - \frac{f(x_{k_i})}{f(x_{k_i}) - f(x_{n_{i-1}})}(x_{k_i} - x_{n_{i-1}}) ,$$

die durch Umformen und Abschätzen die Ungleichung

$$\underbrace{|f(x_{k_i})|}_{>\frac{|f(x_{n_i})|}{2}} > 2|f(x_{n_{i-1}})|$$

liefert und somit für alle $i > N$

$$|f(x_{n_i})| > 4|f(x_{n_{i-1}})|$$

gilt. Dies würde bedeuten, dass die durch $f(x_{n_i})$ gebildete Folge bei wachsendem i gegen Unendlich strebt. Da f allerdings eine stetige Funktion ist und deshalb

$$\lim_{i \to \infty} f(x_{n_i}) = f(a_+)$$

gelten muss, ist die angenommene Aussage $a_- < a_+$ falsch und es muss $a_- = a_+$ sein. Somit ist die Konvergenz der Regula falsi-Iteration bewiesen.

□

(Wille, 1976)

2.4. Fehlerabschätzung und Abbruchbedingung

Anders als das allgemeine Iterationsverfahren wird bei der Regula falsi-Iteration keine Fehlerabschätzung mitgeliefert, das heißt es kann der Rechenaufwand im Vorhinein nicht bestimmt werden kann. Allerdings ist es möglich die Lösung η bis auf einen

vorgegebenen Fehler $\varepsilon > 0$ zu bestimmen, indem die Regula falsi-Folge $\{x_i\}$ bis zu einem $x_i \in \{x_i\}$ berechnet wird, sodass

$$|x_i - x_{i-1}| < \varepsilon$$

gilt. Diese Bedingung kann gleichzeitig als Abbruchbedingung benutzt werden. Ist ein Element der Folge gefunden, welches diese Bedingung erfüllt, wird $f(x_i + \varepsilon)$ und $f(x_i - \varepsilon)$ berechnet. Gilt nun

$$f(x_i + \varepsilon) \cdot f(x_i) < 0 \quad \text{oder} \quad f(x_i - \varepsilon) \cdot f(x_i) < 0 \, .$$

So existiert laut Zwischenwertsatz eine Nullstelle η von f mit

$$|\eta - x_i| < \varepsilon$$

und die Näherungslösung x_i weicht weniger als ε von der exakten Lösung η ab. Man kann auch sagen, dass wenn $f(x_i + \varepsilon) \cdot f(x_i) < 0$ gilt, die Lösung η im Intervall $[x_i, x_i + \varepsilon]$ liegt[15].

2.5. Praktische Anwendung des Algorithmus

Die Regula falsi-Iteration ist in ihrer Durchführung zwar langwierig aber dafür recht einfach. Die Auffindung einer Näherungslösung x kann in vier Schritte unterteilt werden.

1. Schritt: Einschlussintervall $[a, b]$ mit $a = x_1 \, , b = x_2$ wählen, sodass $f(x_1)f(x_2) < 0$ ist.
2. Schritt: Berechnung x_3 mithilfe der Regula falsi-Vorschrift
$$x_3 = b - \frac{f(b)}{f(b) - f(a)}(b - a)$$
3. Schritt: Abbruchbedingung überprüfen
4. Schritt: Ist die Bedingung erfüllt, und $f(x_i + \varepsilon)f(x_i) < 0$ oder $f(x_i - \varepsilon)f(x_i) < 0$ gilt, dann ist x_3 die gewünschte Näherungslösung. Ansonsten wird die Iteration fortgeführt indem neue Intervallgrenzen a und b wie folgt gewählt werden
   i) Wenn $f(x_2)f(x_3) < 0$, dann ist $a = x_3, b = x_2$
   ii) Wenn $f(x_2)f(x_3) > 0$, dann ist $a = x_1, b = x_3$

[15] Wenn $f(x_i - \varepsilon) \cdot f(x_i) < 0$ gilt ist $\eta \in [x_i - \varepsilon, x_i]$

Nun werden der 2.-4. Schritt solange wiederholt, bis die Abbruchbedingung erfüllt ist.

5. Schritt: Fehlerabschätzung

Beispiel 6: (Wir betrachten eine ähnliche Angabe wie im Beispiel 5 verwenden diesmal jedoch die Regula falsi-Iteration). Gegeben ist eine Funktion $f: [0; \infty] \to \mathbb{R}, f(x) \coloneqq e^x - \sqrt{x} - 3$. Gesucht ist eine Näherungslösung der Nullstelle von f, die einen absoluten Fehler maximal $\varepsilon = 1 * 10^{-4}$ hat.

1. Schritt: Geeignete Startwerte für die Intervallgrenzen finden

$$f(1) = -1{,}28 < 0$$
$$f(2) = 2{,}974 > 0$$

Somit ist $a = x_1 = 1$ und $b = x_2 = 2$.

2. Schritt:

$$x_3 = 2 - \frac{f(2)}{f(2) - f(1)}(2 - 1) = 1{,}3$$

3. Schritt:

$$|2 - 1{,}3| = 0{,}7 > 10^{-4}$$

Die Abbruchbedingung ist also nicht erfüllt und die Iteration wird fortgesetzt.

4. Schritt:

$$f(x_3) = -0{,}468$$
$$f(x_2)f(x_3) < 0$$

Somit lautet das neue Einschlussintervall $[x_3, x_2]$ und $a = x_3, b = x_2$.

Die weitere Iteration ist in Tabelle 4 dargestellt.

i	x_i	$f(x_i)$	$\lvert x_i - x_{i-1} \rvert$	Einschlussintervall
1	1	−1,281718172		
2	2	2,974842537	.	$[x_1, x_2]$
3	1,301115915	−0,467271106	.	$[x_3, x_2]$
4	1,395990278	−0,142548027	.	$[x_4, x_2]$
5	1,423609658	−0,041070132	.	$[x_5, x_2]$
6	1,431458833	−0,011636222	0,02943391	$[x_6, x_2]$
7	1,433674041	−0,003281141	0,008355082	$[x_7, x_2]$
8	1,434297989	−0,000923958	0,002357183	$[x_8, x_2]$

9	1,434473636	$-0{,}000260085$	0,000663873	$[x_9, x_2]$
10	1,434523075	$-0{,}000073203$	0,000186881	$[x_{10}, x_2]$
11	1,43453699	$-0{,}000020603$	0,0000526	

Tabelle 4

Das elfte Glied x_{11} der Iterationsfolge erfüllt die Abbruchbedingung, da

$$0{,}0000526 < 10^{-4}.$$

5. Schritt: Nun muss $f(x_{11} + 10^{-4}) \cdot f(x_{11}) < 0$ oder $f(x_{11} - 10^{-4}) \cdot f(x_{11}) < 0$ gelten.

$$\underbrace{f(x_{11} + 10^{-4})}_{>0} \cdot \underbrace{f(x_{11})}_{<0} < 0$$

Somit ist η auf vier Stellen hinter dem Komma genau berechnet und die Lösung lautet $1{,}4345$. Weiteres liegt die exakte Lösung η im Intervall $[x_{11}, x_{11} + 10^{-4}]$.

Es ist sofort erkennbar, dass dieses Iterationsverfahren im Vergleich zum allgemeinen Iterationsverfahren nur sehr langsam gegen die Lösung konvergiert. Im Beispiel 5war die Lösung nach dem 7. Iterationsschritt auf sieben Kommastellen genau mit einem maximalen Fehler von $0{,}8 * 10^{-8}$, hier ist selbst nach dem 11. Schritt die Lösung nur auf vier Kommastellen genau berechnet, bei einem Fehler kleiner als 10^{-4}. Dieses Verfahren hat allerdings den Vorteil der Einfachheit, denn es muss weder abgeleitet noch umgeformt werden. Weiters muss die Konvergenz nicht erst überprüft werden, da die Regula falsi-Folge in einem geeignet gewählten Einschlussintervall für alle stetigen Funktionen konvergent ist.

3. Newtonverfahren in $\mathbb{R}$

Das Newtonverfahren ist ein Standard Iterationsverfahren zur nummerischen Lösung von nichtlinearen Gleichungen. Die Funktion $f: \mathbb{R}^n \longrightarrow \mathbb{R}^n$ Es wurde nach Sir Isaak Newton und Joseph Raphson benannt und wird auch Newton-Raphson Verfahren genannt.

Für dieses Verfahren wird die Funktion f und deren Ableitung f' benötigt. Ein wichtiges Anwendungsgebiet des Newtonverfahren findet sich in der Optimierung, hier wird allerdings zusätzlich auch die zweite Ableitung f'' benötigt da nicht die Nullstellen einer Funktion gefunden werden soll, sondern ein Extremum.

3.1. Herleitung der Formel

Die Idee des Newtonverfahrens ist die sukzessive Linearisierung der Funktion an den Gliedern einer Iterationsfolge$\{x_i\} \in \mathbb{R}$, $i = 0,1,2,3, \ldots$. Diese Linearisierung ist bei einer Funktion $f: \mathbb{R} \longrightarrow \mathbb{R}$ gleichzusetzen mit der Tangente der Funktion in einem Punkt x_i (Freund & Hoppe, 2007).

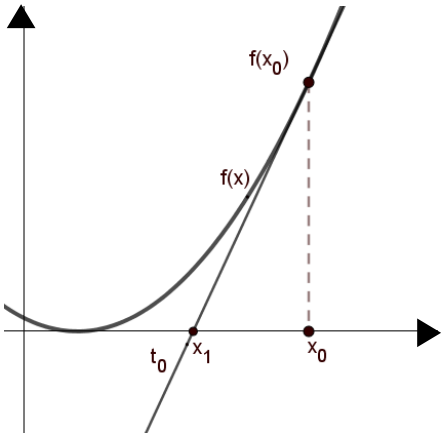

Abbildung 11

Angenommen es ist die Nullstelle $\eta \in [a, b]$ einer Funktion $f: \mathbb{R} \longrightarrow \mathbb{R}$ gesucht, wobei $f(x)$ in dem Intervall $[a, b]$ stetig differenzierbar ist. Nun wird ein Startwert $x_0 \in [a, b]$ passend gewählt[16], dem der Wert $f(x_o)$ zugeordnet wird. Jetzt wird eine Tangente an die Funktion $f(x)$ in x_0 gelegt[17]. Der x-Wert des Schnittpunkts der Tangente mit der x-Achse ist bei Konvergenz die neue bessere Näherung an die Nullstelle η und das nächste Glied der Iterationsfolge $\{x_i\}$, also x_1, welches wiederum Element aus dem Intervall $[a, b]$ ist.

[16] Vgl. Beispiel 2
[17] Siehe Abbildung 11

Dieses Verfahren wird nun so lange wiederholt, bis ein Punkt x_i dicht genug[18] an der Nullstelle liegt. Die Formel kann nun über das Steigungsdreieck und die erste Ableitung der Funktion hergeleitet werden. Das Steigungsdreieck von $t_0$wird von den Eckpunkten x_0, x_1 und $f(x_0)$ begrenzt. Die Steigung $k \in \mathbb{R}$ des Dreiecks kann mithilfe der Katheten, die die Längen $\Delta x := |x_0 - x_1|$ und $\Delta y := f(x_0)$berechnet werden. Es gilt

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

Der Anstieg der Tangente kann jedoch auch analytisch berechnet werden, da der Wert der ersten Ableitung an der Stelle x_0 gleich der Steigung der Funktion in diesem Punkt ist. Demnach gilt auch

$$k_{t_0} = f'(x_0)$$

Setzt man nun die beiden Gleichungen für die Berechnung der Steigung gleich, so erhält man für $x_0 \neq x_1$:

$$\frac{f(x_0)}{|x_0 - x_1|} = f'(x_0)$$

$$f(x_0) = f'(x_0) * |x_0 - x_1|$$

$$\frac{f(x_0)}{f'(x_0)} = x_0 - x_1$$

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

Durch Verallgemeinerung dieser Formel ergibt dies eine Gleichung der Form:

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \qquad i = 0,1,2, \ldots$$

Es ist auch möglich, die Iterationsvorschrift des Newtonverfahren salopp mithilfe der Taylorentwicklung der Funktion f herzuleiten. Wenn eine Funktion $f: A \subset \mathbb{R} \longrightarrow \mathbb{R}$ genügend oft[19] differenzierbar ist, und die Lösung η der Gleichung $f(x) = 0$ in A liegt, erhält man durch die Entwicklung um einen Punkt $x_0 \in A$

[18] Abhängig von der gewünschten Genauigkeit der Näherung an η.
[19] Für dieses Verfahren genügt zwei Mal

$$f(\eta) = 0 = f(x_0) + \frac{(\eta - x_0)}{1!} f'(x_0) + \frac{(\eta - x_0)^2}{2!} f''(x_0) + \cdots$$

$$+ \frac{(\eta - x_0)^k}{k!} f^k(x_0 + \theta(\eta - x_0)), \qquad 0 < \theta < 1$$

Die höheren Potenzen $j > 1$ können für dieses Verfahren vernachlässigt werden, da es sich hier um ein Verfahren ersten Grades handelt. Der Grad eines Verfahrens wird mit der Hochzahl des letzten Terms $(\eta - x_0)$ bestimmt, also dem Term nachdem die Entwicklung abgebrochen wird. Der Grad des Verfahrens bestimmt auch wie oft die Funktion differenzierbar sein muss. Für das Newtonverfahren ergibt sich nun

$$0 \approx f(x_0) + (\eta - x_0) f'(x_0).$$

Jetzt lösen wir nach η auf

$$(\eta - x_0) f'(x_0) \approx -f(x_0)$$

$$\eta - x_0 \approx -\frac{f(x_0)}{f'(x_0)}$$

$$\eta \approx x_0 - \frac{f(x_0)}{f'(x_0)}$$

Der Term $x_0 - \frac{f(x_0)}{f'(x_0)}$ liefert nun eine erste Näherungslösung[20] x_1 an die Lösung η. Um diese zu verfeinern wird eine Iterationsfolge $\{x_i\}$ nach der Vorschrift

$$x_{i+1} = \underbrace{x_i - \frac{f(x_i)}{f'(x_i)}}_{\varphi(x_i)}, \quad i = 0,1, \ldots$$

gebildet. Diese Folge heißt Newtonfolge. (Friedrich & Pietschmann, 2010) (Freund & Hoppe, 2007) (Wille, 1976) (Deuflhard, 2011) (Locher, Numerische Mathematik für Informtiker, 1992).

3.2. Konvergenz und Konvergenzordnung

Um eine Konvergenzbedingung für das Newtonverfahren aufstellen zu können, erinnern wir uns an den Satz 1.1. Wenn nun

[20] Abgesehen von x_0

$$x_{i+1} = \varphi(x_i) = x_i - f(x_i)h(x_i)$$

gilt, und die Funktion h definiert wird durch

$$h(x_i) = \frac{1}{f'(x_i)},$$

so ergibt sich die Iterationsvorschrift des Newtonverfahren

$$x_{i+1} = \varphi(x_i) = x_i - f(x_i)\frac{1}{f'(x_i)} = x_i - \frac{f(x_i)}{f'(x_i)}$$

sowie die Schrittfunktion $\varphi: \mathbb{R} \to \mathbb{R}$

$$\varphi(x) = x - \frac{f(x)}{f'(x)}$$

Laut dem Banach'schen Fixpunktsatz [21] konvergiert die Folge $\{x_i\}$ mit der Iterationsvorschrift $x_{i+1} = \varphi(x_i), \varphi: A \subset \mathbb{R} \to A$ gegen ein $\eta \in A$, wenn für alle $x \in A$ $\sup_{x \in A} |\varphi'(x)|$ gilt. η ist dann gleich dem Fixpunkt von φ und die Lösung der Gleichung $f(x) = 0$, wobei

$$\varphi'(x) = 1 - \frac{f'(x)f'(x) - f(x)f''(x)}{f'(x)^2} = \underbrace{1 - \frac{f'(x)^2}{f'(x)^2}}_{=0} + \frac{f(x)f''(x)}{f'(x)^2} = \frac{f(x)f''(x)}{f'(x)^2}$$

gilt. Erfüllt φ nun die Selbstabbildungs- und die Kontraktionseigenschaft, dann ist das Newtonverfahren lokal konvergent.

Entwickelt man nun $\varphi: A \to \mathbb{R}$ um η für $n = 2$, so erhält man für $x \in A$

$$\varphi(x) = \varphi(\eta) + \varphi'(\eta)(x - \eta) + R_2(x)$$

Nun gilt

$$\varphi(\eta) = \eta$$

$$\varphi'(\eta) = \frac{\overbrace{f(\eta)}^{=0} f''(\eta)}{f'(\eta)^2} = 0$$

$$|R_2(x)| = \frac{|\varphi''(x)|}{2!}(x - \eta)^2 \leq C(x - \eta)^2$$

[21] Vgl. Satz 1.8

$$C = \frac{1}{2}\max_{x\in A}|\varphi''(x)|$$

Setzt man nun all diese Voraussetzungen in die Entwicklung ein so erhält man

$$\varphi(x) = \eta + R_2$$

$$|\varphi(x) - \eta| = |R_2| \leq C(x-\eta)^2$$

Demnach gilt auch

$$|x_{i+1} - \eta| \leq C|x_i - \eta|^2$$

Laut Satz 1.9 ist das Newtonverfahren demnach sogar mindestens quadratisch konvergent, da $p = 2$ ist. Zusammenfassend gilt also:

Satz 3.1: Sei $f: A \to \mathbb{R}$ eine mindestens dreimal stetig differenzierbare Funktion mit $f'(x) \neq 0, x \in A$. Weiteres gäbe es ein eindeutiges $\eta \in A$ mit $f(\eta) = 0$. Zusätzlich gäbe es eine Funktion

$$\varphi(x) = x - \frac{f(x)}{f'(x)},$$

die für alle $x \in A$ folgende Bedingungen erfüllt:

(i) $\sup_{x\in A}|\varphi'(x)| = \sup_{x\in A}\left|\frac{f(x)f''(x)}{f'(x)^2}\right| < 1$

(ii) $\varphi(x) \in A$

Dann ist die Folge $\{x_i\}$ die durch die Iterationsvorschrift

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}, \quad i = 0,1,2,\ldots$$

gegeben ist lokal quadratisch konvergent und es gilt

$$\lim_{i\to\infty} x_i = \eta.$$

Die Bedingung (i) aus diesem Satz kann auch ohne Benutzung der ersten Ableitung formuliert werden, nämlich mithilfe der Lipschitzkonstante L für die gilt $0 \leq L < 1$. Da φ eine kontrahierende Funktion sein muss, gilt nun für $x, y \in A$

$$|\varphi(x) - \varphi(y)| < L|x - y|$$

Wobei die Lipschitzkonstante wie folgt abgeschätzt werden kann

$$\sup_{x \in A} |\varphi'(x)| = \sup_{x \in A} \left| \frac{f(x)f''(x)}{f'(x)^2} \right| \leq L < 1$$

Anders formuliert bedeutet es, dass φ lipschitzstetig ist mit einer Lipschitzkonstanten $L = \sup|\varphi'(x)|$.

Ein Iterationsverfahren, das mit der Iterationsvorschrift aus diesem Satz arbeitet wird als Newtonverfahren bezeichnet. Weil die Newtonfolge nur konvergent ist, wenn der Startwert x_0 „nahe genug“ an der eigentlichen Lösung gewählt wird, ist das Newtonverfahren ein lokal konvergentes Verfahren. (Friedrich & Pietschmann, 2010) (Wille, 1976) (Törnig & Spellucci, 1988) (Locher, Numerische Mathematik für Informtiker, 1992) (Engeln-Müllges, Niederdrenk, & Wodicka, Numerik-Algorithmen, 2005)

3.3. Fehlerabschätzung und Abbruchbedingung

Wie schon bei dem Konvergenzverhalten, verwenden wir wiederum den Banach‘schen Fixpunktsatz um festzustellen, um welchen Wert die Näherungslösung maximal von der exakten Lösung abweicht. Die a priori und a posteriori Fehlerabschätzungen können direkt aus dem allgemeinen Iterationsverfahren übernommen werden, da wir nun schon wissen, dass das Newtonverfahren ein Spezialfall des allgemeinen Iterationsverfahren ist. Demnach gilt mit $0 < L < 1$:

(I) Die a priori Fehlerabschätzung

$$|\eta - x_i| \leq \frac{L^i}{1 - L} |x_1 - x_0|, \quad i = 0,1,2, \dots$$

(II) Die a posteriori Fehlerabschätzung

$$|\eta - x_i| \leq \frac{L}{1 - L} |x_i - x_{i-1}|, \quad i = 0,1,2, \dots$$

(Törnig & Spellucci, 1988)

Da $\varphi(x) = x - \frac{f(x)}{f'(x)}$ gilt kann die a posteriori Fehlerabschätzung verbessert werden. Angenommen die Newtonfolge $\{\tilde{x}_i\}$ konvergiert im Intervall A, dann gilt $x_i \in A$ und $f'(x_i) \neq 0$. Laut der Taylor'schen Formel gilt nun

$$0 = f(\eta) = f(x_i) + f'(x_i)(\eta - x_i) + 1/2 f''(\tilde{x}_i)(\eta - \tilde{x}_i)^2$$

mit $i = 1,2,3, \dots$, dabei liegt $\tilde{x}_i$ zwischen η und x_i. Durch Umformen ergibt sich

$$\underbrace{x_i - \frac{f(x_i)}{f'(x_i)}}_{\substack{\varphi(x_i) \\ x_{i+1}}} - \eta = \frac{1}{2}\frac{f''(\tilde{x}_i)}{f'(x_i)}(\eta - x_i)^2$$

$$x_{i+1} - \eta = \frac{1}{2}\frac{f''(\tilde{x}_i)}{f'(x_i)}(\eta - x_i)^2$$

Durch den Übergang zu den Beträgen ergibt sich jetzt

$$|x_{i+1} - \eta| \leq \frac{1}{2}\frac{\max\limits_{x \in A}|f''(x)|}{\min\limits_{x \in A}|f'(x)|}|\eta - x_i|^2$$

Jetzt wird ein K definiert, sodass

$$\frac{1}{2}\frac{\max\limits_{x \in A}|f''(x)|}{\min\limits_{x \in A}|f'(x)|} \leq K$$

gilt. Demnach ergibt sich die Ungleichung

$$|x_{i+1} - \eta| \leq K|\eta - x_i|^2$$

Für $|\eta - x_i|$ kann die Abschätzung aus der a posteriori Fehlerabschätzung angewendet werden

$$|\eta - x_i| \leq \frac{L}{1 - L}|x_i - x_{i-1}|$$

Somit ergibt sich eine neue a posteriori Fehlerabschätzung:

$$|x_{i+1} - \eta| \leq K\left(\frac{L}{1 - L}\right)^2 |x_i - x_{i-1}|^2$$

Wenn der Betrag der Differenz zweier Folgenglieder x_i und x_{i-1} kleiner ist als eine Schranke $\varepsilon > 0$, so ist auch der Betrag der Differenz von x_i und η kleiner als die Schranke. Aus dieser Eigenschaft kann eine Abbruchbedingung formuliert werden. Die

Iteration kann nach dem i -ten Schritt abgebrochen werden, wenn x_i folgende Bedingungen erfüllt.

(i) $|f(x_i)| < \varepsilon_1$

(ii) $|x_i - x_{i-1}| < \varepsilon_2 \qquad i \in \mathbb{N}\,;\, \varepsilon_1, \varepsilon_2 \in \mathbb{R}^+$

ε_1 und ε_2 geben die Qualität der Nullstelle an und $|x_{i+1} - x_i|$ den absoluten Fehler. Je kleiner ε_1 und ε_2 gewählt werden, desto höher ist die Qualität, also desto kleiner ist der Maximalfehler. (Engeln-Müllges, Niederdrenk, & Wodicka, Numerik-Algorithmen, 2005) (Törnig & Spellucci, 1988)

3.4. Vergleich mit Regula falsi

Wird die allgemeine Formel der Regula falsi nochmals betrachtet, wobei $[a, b]$ das Einschlussintervall der Lösung $f(x) = 0$ ist, so kann diese auf folgende Form gebracht werden

$$x = b - \frac{f(b)}{f(b) - f(a)}(b - a)$$

$$x = b - \frac{f(b)}{\frac{f(b) - f(a)}{b - a}}$$

Der Nenner des Hauptbruchs beschreibt einen Differenzenquotienten im Intervall $[a, b]$, was bei genauer Überlegung gar nicht so abwegig ist, denn schließlich wird die Nullstelle durch Sekantenlegung angenähert[22]. Der Differenzenquotient

$$\frac{f(b) - f(a)}{b - a}$$

beschreibt die Steigung der Sekante die durch die Punkte $(a|f(a))$ und $(b|f(b))$ geht. Jetzt wird die Vorschrift des Newtonverfahren auf eine ähnliche Eigenschaft untersucht.

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

Hier steht die erste Ableitung von $f(x)$ im Nenner, diese kann jedoch auch als Differentialquotient aufgefasst werden, denn es gilt für ein $h \in \mathbb{R}$

[22] Vgl. Abbildung 10

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

wobei h als die Differenz $x-y$ für $x,y\in A$ gedeutet werden kann. A ist hier ein Einschlussintervall um die Lösung η. Der Differentialquotient gibt die Steigung der Tangente an f in x an. Bei grafischer Betrachtung des Newtonverfahren zeigt sich, dass der nächste Näherungswert der Lösung η immer durch Legung der Tangente ermittelt wird[23].

3.5. Praktische Anwendung des Algorithmus

3.5.1. Allgemein

Das Newtonverfahren kann in 5 Schritte unterteilt werden.

1. Schritt: $f(x), f'(x)$ und $f''(x)$ bestimmen[24]
2. Schritt: Konvergenzintervall $A[a,b]$ so wählen, dass $f(a)f(b)<0$ und

   (i) $\varphi(x)=\left(x-\frac{f(x)}{f'(x)}\right)\in A$

   (ii) $|\varphi'(x)|=\left|\frac{f(x)f''(x)}{f'(x)^2}\right|<1$
3. Schritt: Einen Startwert $x_0\in A$ wählen.
4. Schritt: Iterationsfolge $\{x_i\}$ bilden, bis $|x_i-x_{i-1}|$ oder $|f(x_i)|$ eine gewählte Schranke $\varepsilon>0$ unterschreitet.
5. Schritt: Fehlerabschätzung

Beispiel 7: Gesucht ist die Lösung der Gleichung: $e^{2x}-\sin(x)-1=0$ mithilfe des Newtonschen Näherungsverfahren. Der Fehler soll maximal 10^{-9}

1. Schritt: Zuerst bestimmen wir $f(x)$ und die ersten beiden Ableitungen.

$$f(x)=e^{2x}-\sin(x)-3$$

$$f'(x)=2e^{2x}-\cos(x)$$

$$f''(x)=4e^{2x}+\sin(x)$$

2. Schritt

$$f(0)=e^0-\sin(0)-3=-2$$

$$f(1)=3{,}55$$

[23] Vgl. Abbildung 11
[24] Meist ist $f(x)$ gegeben

$$f(0)f(1) < 0$$

Jetzt wird das gefundene Intervall [0,1] soweit eingeengt, dass es die Konvergenzbedingungen (i) und (ii) erfüllt.

x	$f(x)$	$f'(x)$	$f''(x)$	$\varphi(x)$	$\varphi'(x)$
0	−2,00	1,00	4,00	2,00	−8,00
0,1	−1,88	1,45	4,99	1,40	−4,47
0,2	−1,71	2,00	6,17	1,05	−2,62
0,3	−1,47	2,69	7,58	0,85	−1,55
0,4	−1,16	3,53	9,29	0,73	−0,87
0,5	−0,76	4,56	11,35	0,67	−0,42
0,6	−0,24	5,81	13,85	0,64	−0,10
0,7	0,41	7,35	16,87	0,64	0,13
0,8	1,24	9,21	20,53	0,67	0,30
0,9	2,27	11,48	24,98	0,70	0,43
1	3,55	14,24	30,40	0,75	0,53

Tabelle 5

Wie in Tabelle 5 ersichtlich erfüllt das Teilintervall [0,4; 1] die Bedingungen und ist somit A.

3. Schritt: $x_0 = 1$
4. Schritt: Die Berechnung der Iterationsfolge $\{x_i\}$ mit der Vorschrift

$$x_i = x_i - \frac{f(x)}{f'(x)} = x_i - \frac{e^{2x} - \sin(x) - 3}{2e^{2x} - \cos(x)}$$

i	x_i	$f(x_i)$	$f'(x_i)$	$\lvert x_i - x_{i-1}\rvert$
0	1	3,54758511	14,2378099	
1	0,75083351	0,80691797	8,24721227	5,99059762
2	0,65299221	0,08375537	6,58837283	1,65883944
3	0,64027961	0,00123188	6,3953744	0,19299843

4	0,64008699	$2{,}7809 * 10^{-7}$	6,39248717	0,00288723
5	0,64008694	$1{,}4211 * 10^{-14}$	6,39248652	$6{,}5194 * 10^{-7}$
6	0,64008694	0	6,39248652	$3{,}2863 * 10^{-14}$

Tabelle 6

Bei x_6 ist die Fehlerschranke unterschritten und das Verfahren kann abgebrochen werden.

5. Schritt: Um den Fehler abzuschätzen wird die Lipschitzkonstante L und die Konstante K berechnet, dazu wird Tabelle 5 benützt.

$$\frac{1}{2}\frac{\max\limits_{x\in A}|f''(x)|}{\min\limits_{x\in A}|f'(x)|} \leq \frac{30{,}4}{2 * 3{,}53} = 4{,}31 = K$$

$$|\varphi'(x)| \leq 0{,}75 = L$$

Jetzt kann mithilfe dieser Konstanten die a posteriori Fehlerabschätzung [25] berechnet werden.

$$|x_6 - \eta| \leq 4{,}31\left(\frac{0{,}75}{1-0{,}75}\right)^2 |x_5 - x_4|^2 \leq 51{,}72 * (6{,}5194 * 10^{-7})^2$$

$$\leq 2{,}19951 * 10^{-11} \leq 10^{-10}$$

$$|x_6 - \eta| \leq 10^{-10}$$

Das bedeutet, dass die Näherungslösung x_6 sogar auf 10 Dezimalstellen genau ist und somit die Lösung für die Gleichung $e^{2x} - \sin(x) - 1 = 0$ ist.

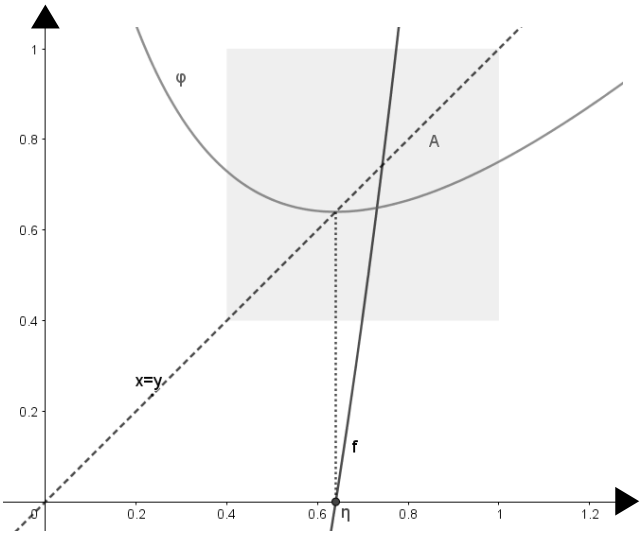

Abbildung 12

[25] Vgl Kapitel 3.3

3.5.2. Wurzelberechnung

Das Newtonverfahren im eindimensionalen Raum eignet sich besonders gut zur Berechnung von Wurzeln. Heutzutage ist Wurzelberechnung von Hand eher unüblich, da technologische Hilfsmittel dies übernehmen. Allerdings ist es spannend, dass Taschenrechner und Computerprogramme zum Großteil einen Algorithmus basierend auf dem Newtonverfahren nutzen um den numerischen Wert einer Wurzel zu berechnen. Dieser Wert ist allerdings wiederum nur ein Näherungswert. Ein Taschenrechner mit einer Anzeigemöglichkeit von beispielsweise neun Nachkommastellen kann demnach einen maximalen Fehler von 10^{-9} (Rießinger, 2001). Betrachten wir nun die allgemeine Wurzel der Form

$$\sqrt[b]{a} = x \qquad a \in \mathbb{R}^+, b \in \mathbb{N}\backslash\{0\}, x \in \mathbb{R}$$

Diese Gleichung kann umgeformt werden zu

$$0 = x^b - a$$

Jetzt kann die rechte Seite der Gleichung als ein Funktionsterm für $f(x)$ aufgefasst werden und die gesamte Gleichung beschreibt die Berechnung der Nullstelle dieser Funktion. Also definierten wir für $f: \mathbb{R} \to \mathbb{R}$

$$f(x) = x^b - a$$

$$f'(x) = bx^{b-1}$$

Diese Funktion ist dreimal stetig differenzierbar mit $f'(x) \neq 0$. Jetzt kann die Newtonfolge $\{x_i\}$ mithilfe der Iterationsvorschrift[26] gebildet werden

$$x_{i+1} = x_i - \frac{x_i^b - a}{bx_i^{b-1}}$$

Aus dieser Vorschrift ist ersichtlich, dass $x_i \neq 0$ gelten muss, demnach darf 0 auch nicht als Startwert gewählt werden.

Beispiel 8: Gesucht ist der numerische Wert von $\sqrt[4]{5}$ mit einem absoluten Fehler von 10^{-7}.

.

[26] Vgl. Satz 3.1

Mit $a = 5$ und $b = 4$ lautet die Vorschrift für die Iteration:

$$x_{i+1} = x_i - \frac{x_i^4 - 5}{4x_i^3}$$

Um einen geeigneten Startwert zu finden, wird $\sqrt[4]{5}$ abgeschätzt

$$1 = \sqrt[4]{1} < \sqrt[4]{5} < \sqrt[4]{16} = 2$$

Demnach kann der Startwert x_0 aus dem Intervall $[1,2]$ gewählt werden. Um nun zu testen ob die Konvergenz in diesem Intervall gegeben ist, wird getestet ob die Funktion $f: [1,2] \to \mathbb{R},\ f(x) = x^4 - 5$ die Konvergenzbedingungen für das Newtonverfahren erfüllt.

(i) $\left|\frac{f(x)f''(x)}{f'(x)^2}\right| < 1$

(ii) $x - \frac{x^4-5}{4x^3} = \varphi(x) \in [1,2]$

$$f'(x) = 4x^3$$

$$f''(x) = 12x^2$$

$$\frac{f(x)f''(x)}{f'(x)^2} = \frac{3x^4 - 15}{4x^4} = \varphi'(x)$$

x	$\varphi(x)$	$\varphi'(x)$
1	2	-3
1,2	1,76	$-1,81$
1,2	1,62	$-1,06$
1,3	1,54	$-0,56$
1,4	1,51	$-0,23$
1,5	1,50	0,0093
1,6	1,51	0,18
1,7	1,53	0,30
1,8	1,56	0,39
1,9	1,61	0,46
2	1,66	0,52

Tabelle 7

Die Bedingung (i) ist erfüllt, da $\varphi(x)$ wieder in das Intervall [1,2] abbildet. Die (ii) Bedingung hingegen ist nicht für das ganze Intervall erfüllt, sondern nur für das Teilintervall $[1{,}3; 2]$, weiteres bildet φ auf sich selbst ab, also $\varphi([1{,}3; 2]) \subseteq [1{,}3; 2]$. Somit ist die Konvergenz des Verfahren nur gewährleistet, wenn $x_i \in [1{,}3; 2]$ ist und somit auch $1{,}3 \leq x_0 \leq 2$ gilt.

Jetzt wissen wir, dass die Iterationsfolge für alle $x_i \in [1{,}3; 2]$ konvergiert und können nun den Näherungswert berechnen

i	x_i	$\lvert x_i - x_{i-1}\rvert$
0	1,3	
1	1,54395767 …	0,24395767
2	1,49759653 …	$-0{,}046361138$
3	1,49535384 …	$-0{,}002242695$
4	1,49534878 …	$-0{,}00000506$
5	1,49534878 …	$-2{,}5 * 10^{-11}$

Tabelle 8

Als Lipschitzkonstante auf dem Intervall $[1{,}3; 2]$ für φ wählen betrachten wir zwei Fälle

1. Fall: $x \geq \sqrt[4]{5}$

$$|\varphi'(x)| = \frac{3x^4 - 15}{4x^4} = \frac{3}{4} - \frac{15}{\underbrace{4x^4}_{>\frac{3}{20}}} < \frac{3}{4} - \frac{3}{20} = \frac{3}{5}$$

2. Fall: $1{,}3 \leq x \leq \sqrt[4]{5}$

$$|\varphi'(x)| = \frac{15 - 3x^4}{4x^4} = \frac{15}{\underbrace{4x^4}_{<\frac{27}{20}}} - \frac{3}{4} < \frac{27}{20} - \frac{3}{4} = \frac{3}{5}$$

Jetzt kann $L = \frac{3}{5}$ als Konstante gewählt werden.

Nun gilt für

$$\frac{L}{1 - L}|x_i - x_{i-1}|$$

Und weiters für $i = 5$

$$\frac{\frac{3}{5}}{1-\frac{3}{5}}(2{,}5 * 10^{-11}) \leq 3{,}75 * 10^{-11} \leq 4 * 10^{-11}$$

Und somit gilt

$$|\eta - x_5| \leq 4 * 10^{-11}$$

Der numerische Wert von $\sqrt[4]{5}$ ist demnach sogar auf einen absoluten Fehler von maximal elf Stellen (10 Nachkommastellen) genau berechnet

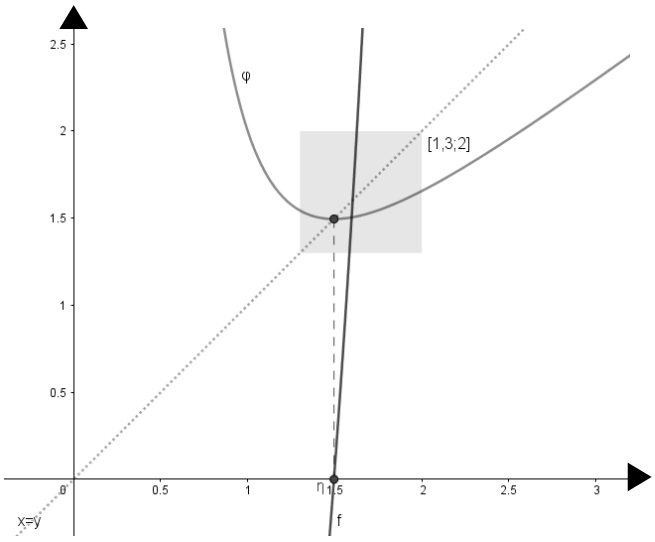

Abbildung 13

Ein Spezialfall tritt ein, wenn $b = 2$ ist. Das Quadratwurzelziehen wurde schon vor langer Zeit in Mesopotamien durchgeführt mit einem sehr ähnlichen Vorgehen, wie jenes mithilfe des Newtonverfahren. Diese damalige systematische Annäherung an den numerischen Wert einer Quadratwurzel ist auch heute noch als babylonisches Wurzelziehen bekannt. Früher, vor der Einführung von technologischen Hilfsmitteln, wurde diese Technik sogar noch in Schulen gelehrt. Später hat Heron von Alexandria das babylonische Wurzelziehen weiterentwickelt und beschrieb das Heron‘sche Wurzelziehen auch Heron-Verfahren genannt (Alten, et al., 2003). Die Herleitung des Verfahren betrachten wir jetzt genauer. Die Quadratwurzel aus einer Zahl $a > 0$ kann als Nullstelle der Funktion

$$f(x) = x^2 - a$$

aufgefasst werden. Wird in die allgemeine Iterationsvorschrift für Wurzeln eingesetzt so ergibt sich für $i = 0,1,2, \ldots$

$$x_{i+1} = x_i - \frac{x_i^2 - a}{2x_i} = \frac{2x_i^2 - x_i^2 + a}{2x_i} = \frac{x_i^2 - a}{2x_i} = \frac{1}{2}\left(x_i + \frac{a}{x_i}\right)$$

Satz 3.2: Sei $a \in \mathbb{R}^+$, so konvergiert die Folge $\{x_i\}$ mit der Vorschrift

$$x_{i+1} = \frac{1}{2}\left(x_i + \frac{a}{x_i}\right)$$

gegen $\sqrt{a}$. Der Startwert $x_0 > 0$ muss folgende Bedingungen erfüllen

(i) $f(x_0) > 0$

(ii) $x_0^2 > a$

Dadurch ist automatisch $x_0 > \sqrt{a}$ und es gilt $x_i > \sqrt{a}$.

Beweis: Der Beweis wird in zwei Teilschritten durchgeführt zuerst wird $x_0 > \sqrt{a}$ nachgewiesen und mithilfe dieser Erkenntnis wird $x_i > \sqrt{a}$ bewiesen. Es gilt: $f: \mathbb{R} \to \mathbb{R}, f(x) = x^2 - a, f(x_0) > 0, f\left(\sqrt{a}\right) = 0, a > 0, x \in \mathbb{R}, x_0 > 0$

$$f'(x) = 2x$$

$$f''(x) = 2$$

Da $f'(0) = 0$ und $f''(0) > 0$ hat die Funktion für alle $a > 0$ an der Stelle 0 ein Minimum. Weiters ist f auf dem gesamten reellen Raum links gekrümmt, da $f''(x) = 2 > 0$ für alle x und für alle $x \in \mathbb{R}^+$ ist $f'(x) = 2 \underbrace{x}_{>0} > 0$. Die Funktion ist also für alle $x \in \mathbb{R}^+$ monoton steigend und $x_0 > \sqrt{a}$ beziehungsweise $x_0^2 > a$.

Jetzt wird mittels vollständiger Induktion $x_i > \sqrt{a}$ bewiesen. Der Induktionsanfang ist mit dem ersten Teil des Beweises geltend.

Weil $x_i > \sqrt{a} > 0$ gilt, ist $x_i - \sqrt{a} > 0$ und $\frac{1}{2x_i} > 0$. Daher ist

$$0 < \frac{1}{2x_i}\left(x_i - \sqrt{a}\right)^2 = \frac{1}{2x_i}\left(x_i^2 - 2x_i\sqrt{a} + a\right) = \frac{1}{2x_i}\left(x_i^2 + a\right) - \sqrt{a}$$
$$= \frac{1}{2}\left(x_i + \frac{a}{x_i}\right) - \sqrt{a}$$

Werden nun beide Seiten der Ungleichung mit $\sqrt{a}$ addiert, so ergibt sich

$$\sqrt{a} < \underbrace{\frac{1}{2}\left(x_i + \frac{a}{x_i}\right)}_{x_{i+1}}.$$

Damit gilt $x_{i+1} > \sqrt{a}$, falls $x_i > \sqrt{a}$

□

Satz 3.3 (Heron Verfahren): Sei $a > 0$, $x_0 > 0$ und $x_0^2 > a$, dann gibt es eine konvergente Iterationsfolge $\{x_i\}$ mit der Vorschrift

$$x_{i+1} = \frac{1}{2}\left(x_i + \frac{a}{x_i}\right), \qquad i = 0,1,2,\dots$$

die quadratisch gegen den numerischen Wert der Quadratwurzel aus a strebt. Es gilt die a posteriori Fehlerabschätzung

$$\left|x_i - \sqrt{a}\right| \leq \frac{x_i^2 - a}{2K}$$

mit $K \coloneqq \min\{1, a\}$

(Wille, 1976) (Werner, 1992)

Beweis: Der erste Teil des Satzes ist oben schon bewiesen worden. Es gilt noch zu zeigen, dass $\{x_i\}$ monoton fallend und für alle $x_i \in (0, \infty)$ quadratisch konvergent ist, sowie die a posteriori Fehlerabschätzung.

$$x_i - x_{i+1} = x_i - \underbrace{x_i + \frac{x_i^2 - a}{2x_i}}_{Newtonverfahren} = \frac{\overbrace{x_i^2}^{>a} - a}{2\underbrace{x_i}_{>0}} > 0$$

Demnach ist $x_i > x_{i+1}$ und $\{x_i\}$ monoton fallend und konvergent. Das heißt es muss ein Grenzwert $s \in \mathbb{R}$ existieren sodass $\lim_{i\to\infty}(x_i) = s$. Betrachtet man nun die Konvergenz der Iterationsvorschrift für das Newtonverfahren so erhält man

$$\lim_{i\to\infty}(x_{i+1}) = \lim_{i\to\infty}\left(x_i - \frac{x_i^2 - a}{2x_i}\right)$$

$$s = s - \frac{s^2 - a}{2s}$$

$$0 = a - s^2$$

$$s = \sqrt{a}$$

Daraus folgt, dass $\lim_{i\to\infty}(x_i) = \sqrt{a}$. Für die Konvergenzordnung verwenden wir aus dem Beweis für Satz 3.2 die Umformung des Terms $x_{i+1} - \sqrt{a}$.

$$x_{i+1} - \sqrt{a} = \frac{1}{2x_i}\left(x_i - \sqrt{a}\right)^2$$

Beim Übergang zu den Beträgen und unter der Beachtung, dass $x_i > \sqrt{a}$ ist ergibt sich

$$\left|x_{i+1} - \sqrt{a}\right| < \frac{1}{2\sqrt{a}}\left|x_i - \sqrt{a}\right|^2$$

Die Hochzahl bei $\left|x_i - \sqrt{a}\right|$ gibt die Konvergenzordnung an. Hiermit ist die quadratische Konvergenz bewiesen. Für die Fehlerabschätzung wird folgende Gleichung betrachtet.

$$\left(x_i + \sqrt{a}\right)\left(x_i - \sqrt{a}\right) = x_i^2 - a$$

Da $x_i > \sqrt{a}$ ist gilt für eine Konstante $K \coloneqq \min\{1, a\}$

$$\left|x_i - \sqrt{a}\right| = \frac{x_i^2 - a}{x_i + \sqrt{a}} \leq \frac{x_i^2 - a}{2\sqrt{a}} \leq \frac{x_i^2 - a}{2K}$$

und es ergibt sich die a posteriori Fehlerabschätzung

$$\left|x_i - \sqrt{a}\right| \leq \frac{x_i^2 - a}{2K}$$

□

(Wille, 1976)

Beispiel 9: Gesucht ist ein numerischer Näherungswert von $\sqrt{7}$.

Weil $3^2 > 7 > 0$ ist, liefert $x_0 = 3$ ein geeigneter Startwert. Die Tabelle 9 zeigt die Iterationsfolge erzeugt durch die Vorschrift

$$x_{i+1} = \frac{1}{2}\left(x_i + \frac{7}{x_i}\right), \qquad i = 0,1,2, \ldots$$

i	x_i
0	3
1	2,66666666666667
2	2,64583333333333
3	2,64575131233596
4	2,64575131106459

Tabelle 9

$$\sqrt{7} = 2,64575131106459$$

Um den Fehler abzuschätzen wird die Konstante $K = 1$ gewählt.

$$\left|x_4 - \sqrt{a}\right| \leq \left|x_3 - \sqrt{a}\right| \leq \frac{7,00000000672744 - 7}{2} = 0,0000000033637$$
$$= 3,33637 \cdot 10^{-9} < 0,5 \cdot 10^{-9}$$

$$\left|x_i - \sqrt{a}\right| \leq \frac{x_i^2 - a}{2K}$$

Demnach unterscheidet sich die Lösung von $\sqrt{7}$ um maximal $0,5 \cdot 10^{-9}$

4. Newtonverfahren in $\mathbb{R}^n$

Zu Beginn dieses Kapitel klären wir kurz die Bedeutung von vier allgemeinen Abbildungen.

$$f(x): \mathbb{R} \to \mathbb{R}$$

$f(x)$ ist hier eine reelle Funktion in einer Variablen die in den eindimensionalen Raum abbildet.

$$g(x): \mathbb{R}^n \to \mathbb{R}$$

$g(x)$ ist eine Funktion in mehreren Variablen die in den eindimensionalen Raum abbildet.

$$s(x): \mathbb{R} \to \mathbb{R}^m$$

$s(x)$ ist ein Beispiel für die Darstellung einer Raumkurve

$$t(x): \mathbb{R}^n \to \mathbb{R}^m$$

$t(x)$ kann hier beispielsweise ein nichtlineares Gleichungssystem ein System von Funktionen oder Flächen darstellen.

4.1. Herleitung der Formel und Konvergenz

Der folgende Absatz bezieht sich auf das Skript *„Einführung in die Numerische Mathematik, Uni Heidelberg“* der Uni Heidelberg, verfasst von Thomas Richter und Thomas Wick und *„Numerische Mathematik. Band 1: Lineare und nichtlineare Gleichungssysteme, Interpolation, numerische Integration“,* geschrieben von Jochen Werner.

Betrachtet man ein System von Funktionen

$$f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{pmatrix}, \qquad n \in \mathbb{N}^*$$

zunächst grafisch, so die Lösung der Gleichung $f(x) = 0$ gleich den Schnittpunkten der Funktionen $f_1, \dots, f_n$. Um die Iterationsvorschrift des Newtonverfahren in $\mathbb{R}^n$ zu erhalten, übertragen wir die Vorschrift aus eindimensionalen in den mehrdimensionalen Raum.

Die Funktion $f(x) \in C^p(A), p \geq 2, A \subseteq \mathbb{R}^n$ beschreibt nun eine $p-$ mal stetig differenzierbare Vektorfunktion und $J_f(x)$ die erste Ableitung dieser Vektorfunktion.

Satz 4.1: Sei $f: \mathbb{R}^n \to \mathbb{R}^n$ eine Abbildung, die eine Nullstelle $\eta \in \mathbb{R}^n$ besitzt, demnach gilt also $f(\eta) = 0$. Weiteres gilt:

(i) f ist auf $A \subseteq \mathbb{R}^n$ stetig partiell differenzierbar

(ii) $J_f(\eta) \in \mathbb{R}^{n\times n}$ ist nichtsingulär, also $\det(J_f) \neq 0$.

Sei nun $\|\cdot\|$ eine zugeordnete Matrizennorm in $\mathbb{R}^n$, dann existiert ein $\sigma > 0$ sodass für jeden Startwert $x_0 \in A_\sigma(\eta) := \{x \in \mathbb{R}^n : \|x - \eta\| \leq \sigma\}$ die durch das Newtonverfahren erzeugte Folge $\{x_i\}$ mit der Vorschrift

$$x_{i+1} = x_i - J_f(x_i)^{-1} f(x_i), \qquad i = 0,1,2,\ldots$$

gegen η konvergiert. Es gilt sogar:

(iii) $J_f(\cdot)$ ist auf einer Kugel um η in η lipschitzstetig, das heißt es existieren ein $\varepsilon > 0$ und ein $L > 0$ mit

$$\|J_f(x) - J_f(\eta)\| \leq L\|x - \eta\|, \quad \text{falls } \|x - \eta\| \leq \varepsilon$$

Demnach konvergiert die Folge für jedes x_0 sogar quadratisch gegen η.

Die Iteration kann abgebrochen werden, wenn eine der folgenden Abbruchbedingungen erfüllt ist.

$$\|x_{i+1} - x_i\| \leq \|x_{i+1}\|\varepsilon_1, \qquad \varepsilon_1 > 0$$

$$\|x_{i+1} - x_i\| < \varepsilon_2, \qquad \varepsilon_2 > 0$$

$$\|f(x_{i+1})\| \leq \varepsilon_3, \qquad \varepsilon_3 > 0$$

Um diesen Satz zu beweisen, benötigen wir zunächst jedoch noch ein Lemma, das Störungslemma.

Lemma (Störungslemma): Sei $M \in \mathbb{R}^{n\times n}$ nichtsingulär. Weiteres sei $N \in \mathbb{R}^{n\times n}$ eine Störung von M mit $\|M^{-1}\|\|N\| < 1$, so ist auch $M + N$ nichtsingulär und

$$\|(M + N)^{-1}\| \leq \frac{\|M^{-1}\|}{1 - \|M^{-1}\|\|N\|}$$

Beweis des Störungslemmas: Sei I die Einheitsmatrix, dann gilt:

$$M + N = M(I + M^{-1}N)$$

Für ein beliebiges $y \in \mathbb{R}^n$ gilt nun:

$$\|(I + M^{-1}N)y\| \geq \|y\| - \|M^{-1}Ny\| \geq \underbrace{(1 - \|M^{-1}\|\|N\|)}_{>0} \|y\|$$

Wenn nun $(I + M^{-1}N)y = 0$ ist, ist $y = 0$. Demnach ist $I + M^{-1}N$ nichtsingulär und $M + N$ folglich auch. Jetzt wird y definiert durch

$$y \coloneqq (I + M^{-1}N)x, \qquad x \in \mathbb{R}$$

Wird nun y in der obigen Ungleichung durch die Definition ersetzt so ergibt sich

$$\|(I + M^{-1}N)^{-1}x\| \leq \frac{1}{1 - \|M^{-1}\|\|N\|} \|x\|$$

Durch Umformung erhält man

$$\|(I + M^{-1}N)^{-1}\| \leq \frac{1}{1 - \|M^{-1}\|\|N\|}$$

$$\|(M + N)^{-1}\| = \|(I + M^{-1}N)^{-1}M^{-1}\| \leq \frac{\|M^{-1}\|}{1 - \|M^{-1}\|\|N\|}$$

und somit ist das Störungslemma bewiesen.

Beweis von Satz 4.1: Zunächst wird etwas mehr Vorbereitung benötigt. Dazu wird ein $\sigma_1 > 0$ so klein gewählt, dass die Jacobi-Matrix J_f auf $A_{\sigma_1}(\eta)$ existiert und es gilt die Ungleichung

$$\|J_f(x) - J_f(\eta)\| \leq \frac{1}{2\|J_f(\eta)^{-1}\|}, \qquad x \in A_{\sigma_1}(\eta)$$

Für alle $x \in A_{\sigma_1}(\eta)$ gilt nun:

(1) Da $J_f(x)$ nichtsingulär ist und wegen der obigen Ungleichung und $J_f(x) = J_f(\eta) + [J_f(x) - J_f(\eta)]$ gilt:

$$\|J_f(\eta)^{-1}\|\|J_f(x) - J_f(\eta)\| \leq \frac{1}{2}$$

Durch die Anwendung des Störungslemmas erhält man nun

$$\|J_f(\eta)^{-1}\| \leq 2\|J_f(\eta)^{-1}\|$$

(2) Betrachtet man

$$J_f(x)^{-1} - J_f(\eta)^{-1} = -J_f(x)^{-1}[J_f(x) - J_f(\eta)]\, J_f(\eta)^{-1}$$

unter der Berücksichtigung von (1) so folgt

$$\left\|J_f(x)^{-1} - J_f(\eta)^{-1}\right\| \leq 2\left\|J_f(\eta)^{-1}\right\|^2 \left\|J_f(x) - J_f(\eta)\right\|$$

Jetzt folgt der eigentliche Beweis. Sei nun $\varphi: A_{\sigma_1}(\eta) \subseteq \mathbb{R}^n \to \mathbb{R}^n$ die Iterationsfunktion des Newtonverfahren.

$$\varphi(x) \coloneqq x - J_f(x)^{-1} f(x)$$

Jetzt kann die Konvergenz wieder über den Banach`schen Fixpunktsatz bewiesen werden. Wenn $\varphi(x)$ in die Kugel A abbildet ist die Selbstabbildungseigenschaft erfüllt und die Iterationsfolge $\{x_i\}$ mit $x_{i+1} = \varphi(x_i)$ ist für jeden Startwert $x_0 \in A$ definiert. Wenn φ nun auch noch kontrahierend ist, konvergiert die Folge $\{x_i\}$ sogar gegen die Lösung η. Sei $x \in A_{\sigma_1}(\eta)$, dann gilt

$$\varphi(x) - \eta = x - \eta - J_f(x)^{-1}[f(x) - f(\eta)]$$

$$= \left[J_f(\eta)^{-1} - J_f(x)^{-1}\right] J_f(\eta)(x - \eta) - J_f(x)^{-1}\left[f(x) - f(\eta) - J_f(\eta)(x - \eta)\right]$$

$$= \left[J_f(\eta)^{-1} - J_f(x)^{-1}\right] J_f(\eta)(x - \eta) - J_f(x)^{-1} \int_0^1 \left[J_f\left(\eta + t(x - \eta)\right) - J_f(\eta)\right](x - \eta)dt$$

Mit den Erkenntnissen aus (1) und (2) gilt demnach also:

$$\|\varphi(x) - \eta\| \leq \omega(x)\|x - \eta\|$$

mit

$$\omega(x) \coloneqq 2\left\|J_f(\eta)^{-1}\right\| \left\{\left\|J_f(\eta)\right\|\left\|J_f(\eta)^{-1}\right\|\left\|J_f(x) - J_f(\eta)\right\| + \int_0^1 \left\|J_f\left(\eta + t(x - \eta)\right) - J_f(\eta)\right\| dt\right\}.$$

Weil $J_f(\cdot)$ in η stetig ist, existiert nun ein $\sigma \in (0, \sigma_1]$, sodass für alle $x \in A_{\sigma_1}(\eta)$

$$\|\varphi(x) - \eta\| \leq \frac{1}{2}\|x - \eta\|$$

gilt. Damit ist bewiesen, dass φ selbstabbildend ist, also $\varphi: A \to A$ und die Folge $\{x_i\}$ gegen η strebt. $\{x_i\}$ konvergiert sogar superlinear gegen η, da

$$\|x_{i+1} - \eta\| \leq \omega(x_i)\|x_i - \eta\| \text{ und } \lim_{i \to \infty} \omega(x_i) = 0.$$

Ist J_f in η lipschitzstetig, so gilt für hinreichend große i

$$\omega(x_i) \coloneqq 2\|J_f(\eta)^{-1}\| \left\{ \|J_f(\eta)\| \|J_f(\eta)^{-1}\| \|J_f(x_i) - J_f(\eta)\| \right.$$
$$\left. + \int_0^1 \|J_f(\eta + t(x_i - \eta)) - J_f(\eta)\| dt \right\}$$
$$\leq 2L\|J_f(\eta)^{-1}\| \left\{ \|J_f(\eta)\| \|J_f(\eta)^{-1}\| + \frac{1}{2} \right\} \|x_i - \eta\|.$$

Hiermit zeigt sich, dass die Iterationsfolge quadratisch konvergent ist. Der Satz 4.1 ist somit bewiesen.

□

(Werner, 1992)

In der Praxis ist es für $\mathbb{R}^n, n > 2$ problematisch die inverse Funktionalmatrix von f zu berechnen. Deshalb wird die Vorschrift des Newtonverfahren durch einen Zwischenschritt so vereinfacht, dass diese Berechnung nicht mehr notwendig ist.

$$x_{i+1} = x_i \underbrace{- J_f(x_i)^{-1} f(x_i)}_{= \Delta x_{i+1}}$$

$$x_{i+1} = x_i + \Delta x_{i+1}$$

$$\Delta x_{i+1=} - J_f(x_i)^{-1} f(x_i)$$

$$J_f(x_i)\Delta x_{i+1} = -f(x_i)$$

Jetzt muss stattdessen ein lineares Gleichungssystem gelöst werden. Die nächste Näherungslösung x_{i+1} wird nun in zwei Schritten (1) und (2) berechnet[27].

(1) Berechnung von Δx_{i+1} durch Lösen des Gleichungssystems

$$J_f(x_i)\Delta x_{i+1} = -f(x_i)$$

(2) Berechnung der Näherungslösung x_{i+1}

$$x_{i+1} = x_i + \Delta x_{i+1}$$

[27] Bis jetzt war es immer nur ein Schritt.

Dies führt zu einem wesentlich praxisorientierteren Satz, den Newton-Kantorovich Satz. Er ist allgemein auch unter dem Kantorovich Theorem bekannt und garantiert die Konvergenz des Verfahren.

Satz 4.2 (Newton-Kantorovich): Sei U eine konvexe, offene Teilmenge des $\mathbb{R}^n$ und $f: U \to \mathbb{R}^n$ eine stetig differenzierbare Funktion. Weiteres gelten folgende vier Punkte

(i) Die Jacobimatrix $J_f(x)$ ist gleichmäßig lipschitzstetig für alle $x, \tilde{x} \in U$

$$\left\|J_f(x) - J_f(\tilde{x})\right\| \leq L\|x - \tilde{x}\|$$

(ii) $J_f(x)$ besitzt eine gleichmäßig beschränkte Inverse für alle $x \in U$

$$\left\|J_f(x)^{-1}\right\| \leq \alpha$$

(iii) Für den Startwert $x_0 \in U$ gilt

$$\beta \coloneqq \left\|J_f(x_0)^{-1} f(x_0)\right\|, \qquad q \coloneqq \alpha\beta L < \frac{1}{2}$$

(iv) Weiteres ist in U eine abgeschlossene Kugel

$$A_r(x_0) \coloneqq \{x \in \mathbb{R}^n : \|x - x_0\| \leq r\}, \qquad r = 2\beta$$

enthalten.

Dann ist in der Kugel $A_r(x_0)$ die Nullstelle η der Funktion f enthalten und die Newtoniteration mit der Vorschrift für $i = 0,1,2, \ldots$

$$J_f(x_i)\Delta x_{i+1} = -f(x_i)$$
$$x_{i+1} = x_i + \Delta x_{i+1}$$

konvergiert quadratisch gegen η. Zusätzlich gilt die a priori Fehlerabschätzung

$$\|x_i - \eta\| \leq 2\beta q^{2^i - 1}$$

Beweis (Newton-Kantorovich): Der Beweis wird in fünf Schritten durchgeführt.

1. Herleitung von Abschätzungen

Sei $x, \tilde{x}, \bar{x} \in U \subseteq \mathbb{R}^n$. Weil U konvex ist, gilt nun für $x, \tilde{x}$

$$f_m(x) - f_m(\tilde{x}) = \int_0^1 \frac{d}{ds} f_m(sx + (1-s)\tilde{x}) ds\,, \qquad m = 1,2,3, \ldots$$

$$\frac{d}{ds} f_m(sx + (1-s)\tilde{x}) = \sum_{i=1}^{n} \frac{\partial f_m}{\partial x_i}\big(sx + (1-s)\big)(x_i - \tilde{x}_i)$$

$$f_m(x) - f_m(\tilde{x}) = \int_0^1 \sum_{i=1}^{n} \frac{\partial f_m}{\partial x_i}(sx + (1-s)\tilde{x})(x_i - \tilde{x}_i)\, ds$$

$$f(x) - f(\tilde{x}) = \int_0^1 J_f\,(sx + (1-s)\tilde{x})(x - \tilde{x}) ds$$

Da $J_f(\bar{x})(\tilde{x} - x)$ ist kann die obige Gleichung nun zu

$$f(x) - f(\tilde{x}) - J_f(x_c)(\tilde{x} - x) = \int_0^1 \Big(J_f(sx + (1-s)\tilde{x}) - J_f(\bar{x})\Big)(x - \tilde{x}) ds$$

umgeformt werden. Da laut der (i) Bedingung J_f lipschitzstetig ist, kann der linke Ausdruck abgeschätzt werden durch

$$\big\|f(\tilde{x}) - f(x) - J_f(\bar{x})(\tilde{x} - x)\big\| \le L\|\tilde{x} - x\| \int_0^1 \|s(x - \bar{x}) + (1-s)(\tilde{x} - \bar{x})\| ds$$

$$\le \frac{L}{2}\|\tilde{x} - x\|\,(\|x - \bar{x}\| + \|\tilde{x} - \bar{x}\|).$$

Wird angenommen, dass $\bar{x} = x$ ist, ergibt sich für alle $x, \tilde{x} \in U$

$$\big\|f(\tilde{x}) - f(x) - J_f(x)(\tilde{x} - x)\big\| \le \frac{L}{2}\|\tilde{x} - x\|^2$$

Wird angenommen, dass $\bar{x} = x_0$ ist, ergibt sich für alle $x, \tilde{x} \in A_r(x_0)$

$$\big\|f(\tilde{x}) - f(x) - J_f(x_0)(\tilde{x} - x)\big\| \le rL\|\tilde{x} - x\|$$

2. Zu zeigen ist, dass alle Folgenglieder der Iterationsfolge $\{x_i\}$ in der Kugel $A_r(x_0)$ liegen, also

$$\|x_i - x_0\| \le r, \qquad \|x_i - x_{i-1}\| \le \beta q^{2^i - 1}, \qquad i = 1,2,3, \ldots$$

Der Beweis wird durch Induktion geführt.

Induktionsanfang: $i = 1$. Durch den (ii) und (iii) Punkt aus Satz 4.2 gilt

$$\|x_1 - x_0\| = \big\|J_f(x_0)^{-1} f(x_0)\big\| = \beta = \frac{r}{2} < r\,,$$

daher ist die Aussage wahr.

Induktionsschritt: $i \to i+1$. x_{i+1} ist wohldefiniert, da die Jacobimatrix eine gleichmäßig beschränkte Inverse besitzt und $x_i \in A_r(x_0) \subset U$ ist. Unter der Verwendung der Induktionsvoraussetzung, der Newtoniteration, der Abschätzung für $\bar{x} = x$ der Definition für q und dem Punkt (ii) aus Satz 4.2 kann folgende Abschätzung gefunden werden:

$$\begin{aligned}
\|x_{i+1} - x_i\| &= \left\|J_f(x_i)^{-1} f(x_i)\right\| \\
&\leq \alpha \|f(x_i)\| \\
&= \alpha \left\|f(x_i) - f(x_i) - J_f(x_{i-1})(x_i - x_{i-1})\right\| \\
&\leq \frac{\alpha L}{2} \|x_i - x_{i-1}\|^2 \\
&\leq \frac{\alpha L}{2} \left[\beta q^{2^{i-1}-1}\right] \\
&= \beta q^{2^{i-1}-1}.
\end{aligned}$$

Zusätzlich gilt

$$\begin{aligned}
\|x_{i+1} - x_0\| &\leq \|x_{i+1} - x_i\| + \|x_i - x_{i-1}\| + \cdots + \|x_1 - x_0\| \\
&\leq \beta \left(1 + q + q^3 + q^7 + \cdots + q^{2^i-1}\right) \\
&\leq \beta \frac{1}{1-q} \\
&\leq 2\beta = r.
\end{aligned}$$

Die Induktion ist somit fertig und es ist bewiesen, dass alle Folgenglieder Elemente der Kugel sind.

3. Existenz eines Grenzwerts η der Folge $\{x_i\}$

Dazu benutzen wir die Cauchy Eigenschaft. Sei $k > 0$ und $q < \frac{1}{2}$, dann gilt

$$\begin{aligned}
\|x_i - x_{i+k}\| &\leq \|x_i - x_{i+1}\| + \cdots + \|x_{i+k-1} - x_{i+k}\| \\
&\leq \beta \left(q^{2^i-1} + q^{2^{i+1}-1} + \cdots + q^{2^{k+i-1}-1}\right) \\
&= \beta q^{2^i-1} \left(1 + q^{2^i} + \cdots + \left(q^{2^i}\right)^{2^{k-1}-1}\right) \\
&\leq 2\beta q^{2^i-1}
\end{aligned}$$

Demnach ist $\{x_i\}$ genauso wie im eindimensionalen Fall eine Cauchy Folge und jede Cauchy Folge konvergiert. Zusätzlich ist $\{x_i\} \subset U_0$ und da U endlich dimensional[28] ist, muss der Grenzwert

$$\eta = \lim_{i\to\infty} x_i$$

existieren. Beim Grenzübergang ergibt sich somit

$$\|\eta - x_0\| \leq r.$$

sodass $\eta \in A_r(x_0)$. Weiteres gilt für $i = 0,1,2, \ldots$

$$\lim_{k\to\infty} \|x_i - x_{i+k}\| = \|x_i - \eta\| \leq 2\beta q^{2^i - 1}$$

und somit ist die a priori Fehlerabschätzung gefunden.

4. Der Grenzwert η ist eine Nullstelle von f.

Der Punkt (i) aus Satz 4.2 und die Newtonvorschrift führen zu folgender Abschätzung

$$\|f(x_i)\| = \left\|J_f(x_i)(x_i - x_{i-1})\right\|$$

$$\leq \left\|J_f(x_i) - J_f(x_0) + J_f(x_0)\right\| \|x_{i+1} - x_i\|$$

$$\leq \left(L\|x_i - x_0\| + \left\|J_f(x_0)\right\|\right) \|x_{i+1} - x_i\|$$

Wenn nun i gegen unendlich strebt, dann strebt der rechte Ausdruck gegen 0 und somit gilt auch

$$\lim_{i\to\infty} f(x_i) = 0.$$

Da f laut Voraussetzung stetig ist gilt

$$f(\eta) = 0$$

und die Folge $\{x_i\}$ konvergiert gegen die Nullstelle η von f.

5. Eindeutigkeit von η innerhalb der Kugel $A_r(x_0)$

Hier benutzen wir den Banach'schen Fixpunktsatz aus Satz 1.8. Demnach ist die Nullstelle von f gleich dem Fixpunkt einer Funktion φ die definiert ist durch

[28] Also vollständig

$$\varphi(x) \coloneqq x - J(x_0)^{-1} f(x).$$

Weil nun

$$\begin{aligned}\varphi(x) - \varphi(\tilde{x}) &= x - \tilde{x} - J_f(x_0)^{-1} f(x) + J_f(x_0)^{-1} f(\tilde{x}) \\ &= J_f(x_0)^{-1}\big(f(\tilde{x}) - f(x) - J_f(x_0)(\tilde{x} - x)\big)\end{aligned}$$

Gilt, folgt nun mithilfe der Punkte aus von Satz 4.2 für alle $x, \tilde{x} \in A_r(x_0)$

$$\|\varphi(x) - \varphi(\tilde{x})\| \leq \alpha\alpha Lr\|\tilde{x} - x\| \leq 2q\|\tilde{x} - x\|.$$

Jetzt ist bewiesen, dass φ eine Kontraktion ist da $q < \frac{1}{2}$ gilt. Weil eine Kontraktion höchstens einen Fixpunkt besitzt, ist η die einzige Nullstelle der Funktion innerhalb der Kugel A.

□

Im Gegensatz zum eindimensional Newtonverfahren muss die Existenz der Nullstelle hier nicht vorausgesetzt werden, sondern ihre Existenz folgt automatisch aus dem Satz. Ein weiterer wesentlicher Unterschied ist die Regularität der Funktion f, die in $\mathbb{R}$ genau dann gegeben ist, wenn f zweimal stetig differenzierbar ist. In $\mathbb{R}^n$ ist f regulär, wenn J_f lipschitzstetig ist. Natürlich kann wie im eindimensionalen Fall das Newtonverfahren im mehrdimensionalen auch als allgemeines Iterationsverfahren formuliert werden, und den ganzen Beweis für das Verfahren über den Banach'schen Fixpunktsatz führen.

(Richter & Wick, 2012)

4.2. Praktische Anwendung des Algorithmus

Um die praktische Durchführung des Newtonverfahren aufzuzeigen, betrachten wir das folgende Beispiel

Beispiel 10: Sei f eine Funktion definiert durch

$$f(x, y) = \begin{pmatrix} 3x^2 - 3y \\ 3y^2 - 3x \end{pmatrix}$$

Gesucht ist die Nullstelle η von f, sodass $f(\eta) = 0$ gilt. Die Lösung soll einen maximalen Fehler von 10^{-5}

$$J_f(x, y) = \begin{pmatrix} 6x & -3 \\ -3 & 6y \end{pmatrix}$$

$$J_f^{-1}(x,y) = \begin{pmatrix} \frac{2y}{12xy-3} & \frac{1}{12xy-3} \\ \frac{1}{12xy-3} & \frac{2x}{12xy-3} \end{pmatrix}$$

$$\det(J) = 36xy - 9$$

Die Jacobi-Matrix von f ist nichtsingulär, wenn $y \neq \frac{1}{4x}$ gilt. Damit sind Bedingung (i) und (ii) aus Satz 4.1 erfüllt. Um jetzt die Bedingung (iii) aus Satz 4.1 zu prüfen, wird eine Konstante L gesucht, sodass

$$\|J_f(x) - J_f(\eta)\| \leq L\|x - \eta\|$$

gilt.

$$J_f(x) - J_f(\eta) = \begin{pmatrix} 6x_1 - 6\eta_1 & -3+3 \\ -3+3 & 6x_2 - 6\eta_2 \end{pmatrix} = \begin{pmatrix} 6(x_1 - \eta_1) & 0 \\ 0 & 6(x_2 - \eta_2) \end{pmatrix}$$

Es gilt

$$\|D\| = \left\| \begin{pmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_2 \end{pmatrix} \right\| = \max_{|x|=1} |Dx| = 6\|x - \eta\|_\infty \leq 6\|x - \eta\|_2$$

Die Abschätzung kann mit Hilfe der Ungleichung $\|x\|_\infty \leq \|x\|_2$ erfolgen[29]. Damit ist eine Konstante $L = 6$ gefunden und die Bedingung (iii) ist mit

$$\|J_f(x) - J_f(\eta)\| \leq 6\|x - \eta\|$$

erfüllt.

Als Startwert wird wählen wir

$$x_0 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

$$-f(x_0) = \begin{pmatrix} -6 \\ -6 \end{pmatrix}, \qquad J_f(x_0) = \begin{pmatrix} 12 & -3 \\ -3 & 12 \end{pmatrix}$$

$$\begin{pmatrix} 12 & -3 \\ -3 & 12 \end{pmatrix} \Delta x_1 = \begin{pmatrix} -6 \\ -6 \end{pmatrix}$$

[29] Siehe Vektor- und Matrizennormen

$$\rightarrow \Delta x_1 = \begin{pmatrix} -\frac{2}{3} \\ -\frac{2}{3} \end{pmatrix}$$

$$x_1 = x_0 + \Delta x_1 = \begin{pmatrix} \frac{4}{3} \\ \frac{4}{3} \end{pmatrix}$$

$$-f(x_1) = \begin{pmatrix} -\frac{4}{3} \\ -\frac{4}{3} \end{pmatrix}, \qquad J_f(x_1) = \begin{pmatrix} 8 & -3 \\ -3 & 8 \end{pmatrix}$$

$$\begin{pmatrix} 8 & -3 \\ -3 & 8 \end{pmatrix} \Delta x_2 = \begin{pmatrix} -\frac{4}{3} \\ -\frac{4}{3} \end{pmatrix}$$

$$\rightarrow \Delta x_2 = \begin{pmatrix} -\frac{4}{15} \\ -\frac{4}{15} \end{pmatrix}$$

$$x_2 = x_1 + \Delta x_2 = \begin{pmatrix} \frac{16}{15} \\ \frac{16}{15} \end{pmatrix}$$

Die Iteration lässt vermuten, dass sich die Werte der x_i an $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ annähern und die nächsten Iterationsschritte verstärken diese Vermutung

$$x_3 = \begin{pmatrix} \frac{256}{255} \\ \frac{256}{255} \end{pmatrix} = \begin{pmatrix} 1{,}00392157 \\ 1{,}00392157 \end{pmatrix}$$

$$x_4 = \begin{pmatrix} \frac{65536}{65535} \\ \frac{65536}{65535} \end{pmatrix} = \begin{pmatrix} 1{,}00001526 \\ 1{,}00001526 \end{pmatrix}$$

$$x_5 = \begin{pmatrix} \frac{9118215569408}{9118128187285} \\ \frac{9118215569408}{9118128187285} \end{pmatrix} = \begin{pmatrix} 1{,}00000958 \\ 1{,}00000958 \end{pmatrix}$$

$$0{,}00000568 = \|x_5 - x_4\| < \varepsilon = 10^{-5}$$

Somit kann x_5 als Lösung angenommen werden und ist auf fünf Nachkommastellen genau, demnach ist

$$\eta = 1{,}00000$$

4.2.1. Vereinfachtes Newtonverfahren

Da es recht aufwändig ist die erste Ableitung neu zu berechnen, kann das Newtonverfahren vereinfacht werden indem bei jedem Iterationsschritt $J_f(x_i)$ durch $J_f(x_0)$ ersetzt wird.

Satz 4.3: Sei $f(x)\colon \mathbb{R}^n \to \mathbb{R}^n$ eine stetig differenzierbare Funktion, deren Jacobimatrix $J_f(x)$ nicht singulär ist. Weiteres existiert ein $\eta \in \mathbb{R}^n$ mit $f(\eta) = 0$ und $\det J_f(\eta) \neq 0$. Dann gibt es ein $\delta > 0$,sodass für alle $x_0 \in \mathbb{R}^n$ mit $|x_0 - \eta| < \delta$ die Folge $\{x_i\}$ mit der Vorschrift

$$x_{i+1} = x_i - \frac{f(x)}{J_f(x_0)}$$

gegen η konvergiert.

(Törnig & Spellucci, 1988)

Beispiel 11: Sei $f(x, y)$ gegeben durch

$$f(x,y) = \begin{pmatrix} f_1(x,y) \\ f_2(x,y) \end{pmatrix} = \begin{pmatrix} x^3 + y^2 \\ \mathrm{x}^2 + \mathrm{y}^2 - 5 \end{pmatrix}.$$

Es soll die Nullstelle η des Systems[30] mithilfe des vereinfachten Newtonverfahren annäherungsweise berechnet werden mit einem maximalen Fehler von $5 * 10^{-2}$.

Die Abbildungen 14-16 zeigen die grafische Lösung des Beispiels. Abbildung 14 zeigt f_1, Abbildung 15 zeigt f_2 und Abbildung 16 zeigt die beiden Funktionen in der xy −Ebene.

30

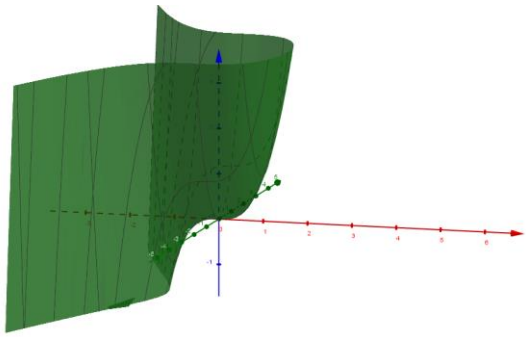

Abbildung 14

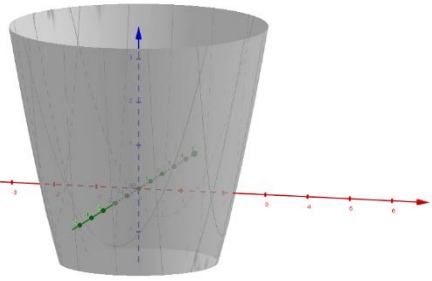

Abbildung 15

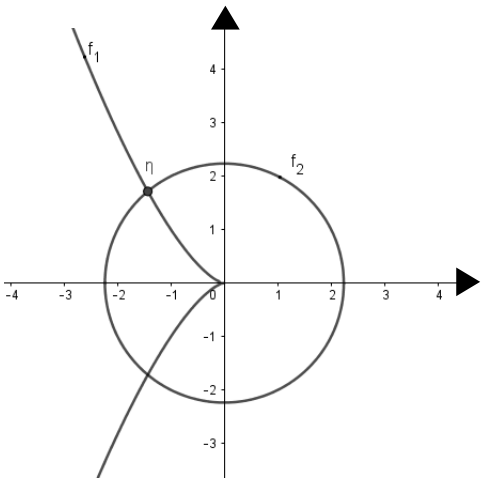

Abbildung 16

Die Ableitung von f ist

$$J_f(x,y) = \begin{pmatrix} 3x^2 & 2y \\ 2x & 2y \end{pmatrix}$$

$$\det(J_f) = 3x^2 * 2y - 2y * 2y = 6x^2y - 4y^2$$

Um nun die Konvergenz des Verfahren zu sichern, wird eine Umgebung A bestimmt, sodass $\det(J_f) \neq 0$ ist. Dies ist sicher der Fall, wenn $y \neq 0$ und $y \neq \frac{3}{2}x^2$ ist.

Als Startwert wird der Vektor $x_0 = (-1 \quad 2)^T$ gewählt und die neue Näherungslösung wird mithilfe der Vorschrift aus Satz 4.3 berechnet.

$$f(x_0) = \begin{pmatrix} 3 \\ 0 \end{pmatrix}, \qquad J_f(x_0) = \begin{pmatrix} 3 & 4 \\ -2 & 4 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 4 \\ -2 & 4 \end{pmatrix} \Delta x_1 = -\begin{pmatrix} 3 \\ 0 \end{pmatrix},$$

$$\rightarrow \Delta x_1 = \begin{pmatrix} -\frac{3}{5} \\ -\frac{3}{10} \end{pmatrix}$$

$$x_1 = x_0 + \Delta x_1$$

$$x_1 = \begin{pmatrix} -1 \\ 2 \end{pmatrix} + \begin{pmatrix} -\frac{3}{5} \\ -\frac{3}{10} \end{pmatrix} = \begin{pmatrix} -\frac{8}{5} \\ \frac{17}{10} \end{pmatrix}$$

$$J_f(x_0)\Delta x_2 = -f(x_1)$$

$$\begin{pmatrix} 3 & 4 \\ -2 & 4 \end{pmatrix} \Delta x_2 = -\begin{pmatrix} -\frac{603}{500} \\ \frac{9}{20} \end{pmatrix}$$

$$\rightarrow \Delta x_2 = \begin{pmatrix} \frac{207}{625} \\ \frac{531}{10000} \end{pmatrix}$$

$$x_2 = x_1 + \Delta x_2$$

$$x_2 = \begin{pmatrix} -\frac{8}{5} \\ \frac{17}{10} \end{pmatrix} + \begin{pmatrix} \frac{207}{625} \\ \frac{531}{10000} \end{pmatrix} = \begin{pmatrix} -\frac{793}{625} \\ \frac{17531}{10000} \end{pmatrix}$$

$$x_3 = \begin{pmatrix} -\frac{1877823368}{1220703125} \\ \frac{2121925357459}{1250000000000} \end{pmatrix}$$

Ab jetzt rechnen wir der Einfachheit halber mit Werten auf vier Dezimalstellen gerundet

$$x_3 = \begin{pmatrix} -1{,}538 \\ 1{,}698 \end{pmatrix}$$

$$x_4 = \begin{pmatrix} -1{,}337 \\ 1{,}736 \end{pmatrix}$$

$$x_5 = \begin{pmatrix} -1{,}501 \\ 1{,}703 \end{pmatrix}$$

$$x_6 = \begin{pmatrix} -1{,}374 \\ 1{,}728 \end{pmatrix}$$

$$x_7 = \begin{pmatrix} -1{,}478 \\ 1{,}708 \end{pmatrix}$$

$$x_8 = \begin{pmatrix} -1{,}395 \\ 1{,}724 \end{pmatrix}$$

$$x_9 = \begin{pmatrix} -1{,}463 \\ 1{,}711 \end{pmatrix}$$

$$x_{10} = \begin{pmatrix} -1{,}409 \\ 1{,}721 \end{pmatrix}$$

$$x_{11} = \begin{pmatrix} -1{,}452 \\ 1{,}712 \end{pmatrix}$$

$$x_{12} = \begin{pmatrix} -1{,}418 \\ 1{,}720 \end{pmatrix}$$

$$x_{13} = \begin{pmatrix} -1{,}446 \\ 1{,}714 \end{pmatrix}$$

$$x_{14} = \begin{pmatrix} -1{,}423 \\ 1{,}718 \end{pmatrix}$$

Somit kann die Iteration abgebrochen werden, da

$$\|x_{14} - x_{13}\| = \left\| \begin{matrix} -0{,}023 \\ -0{,}004 \end{matrix} \right\|$$

$$0{,}023 < 5 * 10^{-2}$$

ist und die x_{14} kann als Näherungslösung angenommen werden.

4.2.2. Spezialfall $\boldsymbol{n = 2}$

Wenn J_f eine 2×2 Matrix ist, kann eine allgemeine vereinfachte Iteration hergeleitet werden. Sei

$$J_f \coloneqq \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}.$$

Um die Inverse zu berechnen, wird ein Gleichungssystem $J_f x = a$ für einen Vektor $a \in \mathbb{R}^2$ gelöst. Der Vektor x ist definiert durch $(x_1 \quad x_2)^T$ somit ergeben sich folgende zwei Gleichungen

(i) $f_{11}x_1 + f_{12}x_2 = a_1$

(ii) $f_{21}x_1 + f_{22}x_2 = a_2$

Wird nun das Eliminationsverfahren angewendet und je ein Mal x_1 beziehungsweise x_2 eliminiert, so ergeben sich

(iii) $(f_{11}f_{22} - f_{21}f_{12})x_1 = f_{22}a_1 - f_{12}a_2$

(iv) $(f_{11}f_{22} - f_{21}f_{12})x_2 = -f_{21}a_1 + f_{11}a_2$.

Da J_f invertierbar ist, muss $(f_{11}f_{22} - f_{21}f_{12}) \neq 0$ gelten, da es aufgrund der Invertierbarkeit für jeden beliebigen Vektor a eine Lösung x gibt. Wenn $(f_{11}f_{22} - f_{21}f_{12}) = 0$ gelten würde, wäre die Existenz nämlich von x an Bedingungen für a geknüpft und somit nicht gesichert. Werden die Gleichungen (iii) und (iv) nach x gelöst erhält man

(v) $x_1 = \frac{1}{(f_{11}f_{22} - f_{21}f_{12})}(f_{22}a_1 - f_{12}a_2)$

(vi) $x_2 = \frac{1}{(f_{11}f_{22} - f_{21}f_{12})}(-f_{21}a_1 + f_{11}a_2)$

als Lösung des Gleichungssystems. Weil

$$J_f x = a$$

$$x = J_f^{-1} a$$

gilt, und mit Betrachtung der Gleichungen (v) und (vi), können die Komponenten der Matrix von J_f^{-1} abgelesen werden.

$$J_f^{-1} = \frac{1}{(f_{11}f_{22} - f_{21}f_{12})}\begin{pmatrix} f_{22} & -f_{12} \\ -f_{21} & f_{11} \end{pmatrix}$$

Wird nun diese Erkenntnis in die Vorschrift des Newtonverfahren eingesetzt so ergibt sich

$$x_{i+1} = x_i - \frac{1}{(f_{11}(x_i)f_{22}(x_i) - f_{21}(x_i)f_{12}(x_i))}\begin{pmatrix} f_{22}(x_i) & -f_{12}(x_i) \\ -f_{21}(x_i) & f_{11}(x_i) \end{pmatrix} f(x_i)$$

$$x_{i+1} = x_i - \frac{1}{(f_{11}f_{22} - f_{21}f_{12})}\begin{pmatrix} f_{22}f_1 - f_{12}f_2 \\ -f_{21}f_1 + f_{11}f_2 \end{pmatrix}$$

wobei

$$x_{i+1_1} = x_{i_1} - \frac{f_{22}f_1 - f_{12}f_2}{f_{11}f_{22} - f_{21}f_{12}}$$

$$x_{i+1_2} = x_{i_2} - \frac{-f_{21}f_1 + f_{11}f_2}{f_{11}f_{22} - f_{21}f_{12}}.$$

(Engeln-Müllges, Niederdrenk, & Wodicka, Numerische Verfahren zur Lösung nichtlinearer Gleichungen, 2011)

Daraus kann nun folgender Satz geschlossen werden.

Satz 4.4: Sei $f: \mathbb{R}^2 \to \mathbb{R}^2$ ein System von Funktionen mit

$$f(x_1, x_2) = \begin{pmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{pmatrix},$$

dann konvergiert die Iterationsfolge $\{x_i\}$ gegen die Lösung η des Systems. Die Vorschrift für die Ermittlung der Komponenten x_{i+1_1} und x_{i+1_2}, der nächsten Näherungslösung x_{i+1} lautet

$$x_{i+1_1} = x_{i_1} - \frac{f_{22}f_1 - f_{12}f_2}{f_{11}f_{22} - f_{21}f_{12}}$$

$$x_{i+1_2} = x_{i_2} - \frac{-f_{21}f_1 + f_{11}f_2}{f_{11}f_{22} - f_{21}f_{12}},$$

wobei

$$J_f(x_1, x_2) = \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}$$

ist. Es gelten die Fehlerabschätzungen sowie die Abbruchbedingen wie in $\mathbb{R}^n$.

Beispiel 12: Sei $f(x_1, x_2)$ ein System von Funktionen definiert durch

$$f(x_1, x_2) = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} = \begin{pmatrix} x_1^2 + x_2 - 10 \\ x_1 + x_2^2 - 6 \end{pmatrix}$$

Gesucht ist die Nullstelle des Systems, also $f(x_1, x_2) = 0$ mit einem Maximalfehler von $\varepsilon \leq 10^{-7}$.

Zuerst wird die Jacobimatrix von f berechnet. J_f ist eine 2×2 Matrix, da f aus zwei Funktionen f_1, f_2 besteht und zwei Variablen besitzt.

$$J_f(x_1, x_2) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix} = \begin{pmatrix} 2x_1 & 1 \\ 1 & 2x_2 \end{pmatrix}$$

Die Matrix ist nichtsingulär, da

$$\det J_f = 4x_1x_2 - 1 \neq 0, \text{ für } x_1 \neq \frac{1}{4x_2}, x_2 \neq \frac{1}{4x_1}$$

gilt. Weil $f(x_1, x_2)$ die Bedingungen aus Satz 4.4 erfüllt, lautet die Iterationsvorschrift für die Komponenten

$$x_{i+1_1} = x_{i_1} - \frac{2x_{i_1}^2 x_{i_2} + x_{i_2}^2 - x_{i_1} - 20x_{i_2} + 6}{4x_{i_1}x_{i_2} - 1}$$

$$x_{i+1_2} = x_{i_2} - \frac{2x_{i_1}x_{i_2}^2 - x_{i_2} + x_{i_1}^2 - 12x_{i_1} + 10}{4x_{i_1}x_{i_2} - 1}$$

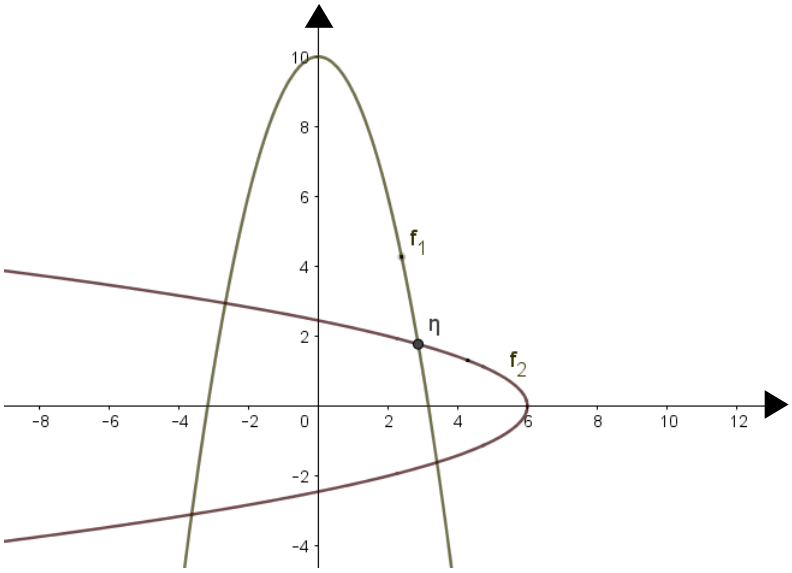

Abbildung 17

Die Abbildung 17 zeigt die grafische Lösung des Systems. Die Funktionen f_1 und f_2 sind hier implizit dargestellt und es ist ersichtlich, dass es vier mögliche Lösungen gibt. In diesem Bespiel wird die Lösung $x_1 > 0, x_2 > 0$ gesucht. In der Abbildung ist diese durch η gekennzeichnet. Als Startvektor wird $x_0 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$ gewählt.

$$x_1 = 2 - \frac{16 + 4 - 2 - 40 + 6}{16 - 1} = 2 + \frac{16}{15} = 3{,}067$$

$$x_2 = 2 - \frac{16 - 2 + 4 - 24 + 10}{16 - 1} = 2 - \frac{4}{15} = 1{,}768$$

x_i	x_{i_1}	x_{i_2}
x_0	2	2
x_1	3,06666667	1,73333333
x_2	2,87551364	1,76796081
x_3	2,86889349	1,76949396
x_4	2,86888558	1,76949553
x_5	2,86888558	1,76949553

Tabelle 10

Tabelle 10 zeigt die Folgenglieder der Iterationsfolge. Bereits nach dem vierten Schritt ist die Näherungslösung auf die achte Kommastelle genau, das zeigt wie schnell das Newtonverfahren konvergiert. Die Abbruchbedingung $\|x_{i+1} - x_i\| \leq \varepsilon$ ist bei x_5 erfüllt, somit kann die Iteration abgebrochen werden.

$$\|x_5 - x_4\| = \left\| \begin{pmatrix} 2{,}868885580346 \\ 1{,}769495526882 \end{pmatrix} - \begin{pmatrix} 2{,}868885580358 \\ 1{,}769495526879 \end{pmatrix} \right\|$$
$$= \left\| \begin{pmatrix} -0{,}000000000011 \\ 0{,}000000000003 \end{pmatrix} \right\| = 0{,}11 * 10^{-10} \leq 10^{-7}$$

Somit lautet der Lösungsvektor des Systems $\begin{pmatrix} 2{,}86888558 \\ 1{,}76949553 \end{pmatrix}$.

5. Iteration in der Schule

Iterationsverfahren sind im Lehrplan der AHS wörtlich nicht enthalten, und sind am ehesten in der 6. Klasse AHS im Kapitel Folgen angeschnitten. Eine rekursive beschränkte Folge kann als Iterationsverfahren zur Auffindung des Grenzwerts gesehen werden. Allerdings kann das Thema Iteration bei einigen anderen Themen des Lehrplans erwähnt und aufgezeigt werden. Ein Beispiel dafür wäre die Zahl π bei der Einführung der irrationalen Zahlen. Den Schülerinnen und Schülern kann hier bewusst gemacht werden, dass der Wert den Taschenrechner angibt nur eine auf die Anzeigemöglichkeit des Taschenrechners angepasste Näherungslösung ist[31] und mittels Iteration berechnet wird. Da in der neuen Zentralmatura ab 2019 die Berechnungen der Beispiele fast zur Gänze mithilfe technologische Unterstützung erfolgt, sollten die Schülerinnen und Schüler verstehen, wie diese Berechnungen im Hintergrund möglich sind. Dies kann das Verständnis für mathematischen Abläufen und Denkweisen fördern. Eine andere Möglichkeit bietet das Thema Tangenten. Die Iteration kann als Anwendungsgebiet für Tangenten angeführt werden, indem das Newtonverfahren grafisch durchgeführt wird. So kann gezeigt werden, dass die Nullstelle von einer Funktion mithilfe von Tangentenlegung[32] gefunden werden kann. Eine besonders gute Kombination bietet das Thema Wurzel ziehen, denn hier kann ein Ausschnitt aus der Geschichte der Mathematik durch die Vorstellung des Heron-Verfahren mit einfließen. Dieses Verfahren bedarf nur einer kurzen und einfachen Bestimmung eines geeigneten Startwerts und benötigt im Zuge der Iteration[33] keine Ableitungen. Diese Einfachheit erlaubt es, das Verfahren zur Wurzelberechnung gemeinsam mit den Schülerinnen und Schülern durchzuführen und aufzuzeigen, dass es auch damals, ohne technologische Hilfsmittel möglich war die Wurzel aus einer reellen Zahl zu ziehen.

Die praktische Durchführung des Newton Algorithmus für die Auffindung von Nullstellen von Funktionen ist im regulären Mathematikunterricht wahrscheinlich nicht realisierbar, da es doch ein etwas umfangreiches Thema ist und mehr Zeit benötigt als der Lehrplan zulassen würde. Allerdings ist das Newtonverfahren sowie das allgemeine Iterationsverfahren, beide im eindimensionalen Raum, für das Wahlpflichtfach oder einen Fortgeschrittenenkurs durchaus interessant. (Bundesministerium Bildung)

[31] Siehe Kapitel 3.5.2
[32] Einem bestimmten Muster entsprechend, vgl Kapitel 3.1
[33] Vgl.Satz 3.3

Literaturverzeichnis

Alten, H.-W., Djafari, A., Folkerts, M., Schlosser, H., Schlote, K.-H., & Wußing, H. (2003). *4000 Jahre Algebra: Geschicht. Kulturen. Menschen.* Berlin Heidelberg: Springer-Verlag.

Arens, T., Busam, R., Hettlich, F., Karpfinger, C., & Stachel, H. (2013). *Grundwissen Mathematikstudium.Analysis und Lineare Algebra mit Querverbindungen.* Berlin Heidelberg: Springer Spektrum.

Brosowski, B., & Kreß, R. (1975). Grundbegriffe der Funktionalanalysis. In *Einführung in die Numerische Mathematik* (S. 13). Mannheim: B. I.-Wissenschaftsverlag.

Bundesministerium Bildung, W. u. (kein Datum). *Lehrplan Mathematik AHS.* Abgerufen am 27. November 2018 von https://bildung.bmbwf.gv.at/schulen/unterricht/lp/lp_neu_ahs_07_11859.pdf

Deuflhard, P. (2011). *Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms* (softcover Ausg.). Berlin/Heidelberg: Springer-Verlag.

Engeln-Müllges, G., Niederdrenk, K., & Wodicka, R. (2005). *Numerik-Algorithmen.* Berlin Heidelberg: Springer-Verlag.

Engeln-Müllges, G., Niederdrenk, K., & Wodicka, R. (2011). Numerische Verfahren zur Lösung nichtlinearer Gleichungen. In *Numerik-Algorithmen* (S. 27-90). Berlin Heidelberg: Springer-Verlag.

Freund, R. W., & Hoppe, R. H. (2007). *Stoer/Bulirsch:Numerische Mathematik 1* (10 Ausg.). Berlin Heidelberg: Springer.

Friedrich, H., & Pietschmann, F. (2010). *Numerische Methoden.* Berlin/New York: de Gruyter.

Henrici, P. (1972). *Elemente der numerischen Analysis.* Mannheim: B.I. Wissenschaftsverlag .

Heuser, H. (1990). *Lehrbuch der Analysis* (10 Ausg.). Wiesbaden: Springer Fachmedien.

Karpfinger, C., & Meyberg, K. (2012). In *Algebra.Gruppen-Ringe-Körper* (S. 359). Berlin Heidelberg: Springer Spektrum.

Locher, F. (1992). *Numeriche Mathematim für Informatiker.* Berlin Heidelberg: Springer-Verlag.

Locher, F. (1992). *Numerische Mathematik für Informtiker.* Berlin Heidelberg: Springer-Verlag.

Neunhäuserer, J. (2017). Topologie. In *Schöne Sätze der Mathematik: Ein Überblick mit kurzen Beweisen* (S. 86). Berlin: Springer Spektrum.

Opfer, G. (1993). Nichtlineare Gleichungen und Gleichungssysteme . In *Numerische Mathematik für Anfänger* (S. 242-272). Braunschweig/Wiesbaden: Vieweg.

Ortega, J. M., & Rheinboldt, W. C. (1970). *Iterative solution of nonlinear equations in severeal variables.* New York, London: Academic Press.

Richter, T., & Wick, T. (30. Oktober 2012). *Einführung in die Numerische Mathematik, Uni Heidelberg.* Abgerufen am 11. November 2018 von https://ganymed.math.uni-heidelberg.de/~lehre/SS12/numerik0/15-fixpunkt-1.pdf

Rießinger, T. (2001). *Mathematik für Ingenieure: Eine anschauliche Einführung für das praxisorientierte Studium* (3 Ausg.). Berlin Heidelberg: Springer- Verlag.

Törnig, W., & Spellucci, P. (1988). *Numerische Mathematik für Ingenieure und Physiker. Numerische Methoden der Algebra* (2 Ausg.). Berlin, Heidelberg: Springer-Verlag.

Werner, J. (1992). *Numerische Mathematik. Band 1: Lineare und nichtlineare Gleichungssysteme, Interpolation, numerische Integration.* Wiesbaden: Springer Fachmedien.

Wille, F. (1976). *Analysis: eine anwendungsbezogene Einführung.* Stuttgart: B.G. Teubner.

Abbildungsverzeichnis

Alle Abbildungen wurden selbst mittels Geogebra erstellt.

Abbildung 1-Zusammenhang $f(\eta) = 0$ und $\varphi(\eta) = \eta$……..10

Abbildung 2- Grafisches Auffinden eines Startwerts……..12

Abbildung 3- stetige Funktion im Intervall $[a, b]$……...13

Abbildung 4- nicht stetige Funktion im Intervall $[a, b]$ ……..……...……13

Abbildung 5- grafische konvergente Iteration…….17

Abbildung 6- grafische konvergente Iteration…….18

Abbildung 7- grafische divergente Iteration……18

Abbildung 8- grafische divergente Iteration……19

Abbildung 9- Zusammenhang Fixpunkt von φ und Nullstelle von f…….32

Abbildung 10- Funktion f mit Sekante……...33

Abbildung 11- Funktion f mit Tangent in (x_0)……...……42

Abbildung 12- grafische Lösung von Beispiel 7…….52

Abbildung 13- grafische Lösung von Beispiel 8…….56

Abbildung 14- dreidimensionale Darstellung von $f_1(x, y) = x^3 + y^2$……...74

Abbildung 15- dreidimensionale Darstellung von $f_2\ (x, y) = x^2 + y^2 - 5$……...74

Abbildung 16- grafische zweidimensionale Lösung von Beispiel 11……..…....74

Abbildung 17- grafische Lösung von Beispiel 12……...……79

Zusammenfassung

Zu Beginn der Arbeit werden Hilfsmittel der mehrdimensionalen Analysis und grundlegende mathematische Formulierungen erklärt die im Laufe der Arbeit verwendet werden. Jedes Kapitel, abgesehen von dem Letzten, beschäftigt sich mit einem Iterationsverfahren, welches theoretisch beschrieben und bewiesen wird und anschließend an einem Beispiel angewendet wird. Zuerst wird das allgemeine Iterationsverfahren beschrieben, da sich alle anderen Iterationsverfahren auf dieses Verfahren zurückführen lassen. Es wird Schritt für Schritt zunächst in der Theorie erklärt wie sich dieses Verfahren verhält und welche Bedingungen die Konvergenz garantieren, wobei zwischen den theoretischen Erklärungen die Anwendung einzelner Schritte mithilfe von Beispielen und Abbildungen aufgezeigt wird. In dem ersten Kapitel spielen die Begriffe lipschitzstetig, kontrahierend und Fixpunktgleichung eine wichtige Rolle und es wird aufgezeigt wie der Fixpunkt einer passenden Schrittfunktion mit der Nullstelle einer Funktion zusammenhängt. Das zweite Kapitel behandelt die Regula falsi Iteration, ein Vertreter der Einschlussverfahren. Auch hier wird zunächst die Formel hergeleitet und anschließend die Konvergenz sowie Abbruchbedingung und Fehlerabschätzung für dieses Verfahren beschrieben und bewiesen. Die folgenden zwei Kapitel behandeln das Newtonverfahren, indem zunächst im dritten Kapitel nur der eindimensionale Fall behandelt wird. Nach der Herleitung und den Voraussetzungen für Konvergenz sowie Fehlerabschätzung und Abbruchbedingung wird auch auf die Ähnlichkeiten zwischen Regula falsi und dem Newtonverfahren eingegangen. Die praktische Anwendung des Newtonverfahren wird in zwei Unterkapitel erklärt, wobei eines ein Exkurs in die Wurzelberechnung ist. Hier findet sich ein Anwendungsgebiet für die Schule, welches im letzten Kapitel nochmal beschrieben wird. Das vierte Kapitel beschäftigt sich mit dem Newtonverfahren im mehrdimensionalen Raum. Danach wird die Formel mittels zweier Sätze beschrieben, wobei einer die praxisorientierte Iteration aufzeigt. Nach der praktischen Anwendung des Algorithmus wird zusätzlich eine, aus dem Newtonverfahren abgeleitete, Iteration für den zweidimensionalen Raum theoretisch und praktisch beschrieben. Da Iteration kein Teil des Lehrplans der AHS ist, wird im letzten Kapitel einige Möglichkeiten aufgezeigt, wie Iteration im Mathematikunterricht integriert werden kann.

Abstract

At the start of this study, tools of multidimensional analyses and fundamental mathematical formulations used in this paper are explained. Every chapter, except for the very last one, deals with an iteration method that is described, proven theoretically, and then applied to an example. Initially, the common iteration method is depicted because all other iteration methods can be traced back to this common one. To begin with, it is explained theoretically step by step how this method works and which conditions guarantee the convergence, whereby between the theoretical explanations the applications of individual steps are demonstrated by examples and mappings. In the first chapter, the terms Lipschitz-continuous, contractive, and fixed point equation play an important role and it is shown how the fixed point of an appropriate iteration function is connected to the zero of a function. The second chapter deals with the Regula falsi algorithm, a representative of containment methods. Here too, at first the formula is derivated and then the convergence as well as conditions of truncation and error estimations for this method are described and proven. The subsequent two chapters deal with the Newton algorithm, as at first, in the third chapter, only the one-dimensional case is covered. After the derivation and requirements for convergence as well as error estimation and conditions of truncations, the similarities between Regula falsi and Newton algorithm are elaborated. The practical application of the Newton algorithm is explained in two subchapters, one of which is an excursus on radical reckoning. Here there is a field of application for schools, which is again described in the last chapter. The fourth chapter addresses the Newton algorithm in multidimensional space. Subsequently, the formula is described employing two propositions/theorems, one of which demonstrates the practice-oriented iteration. Additionally, following the practical application of the algorithm an iteration for two-dimensional space derivated from the Newton algorithm is described in theory as well as in practice. Since iteration is not part of the curriculum of AHS, some possibilities of how iteration can be integrated into math class are shown in the final chapter.