Verbessern, immer wieder verbessern ...
Gleichungen numerisch behandeln

H.R. Schneebeli, T.P. Wihler

Version vom 31. März 2016

Zusammenfassung

Zur näherungsweisen numerischen Lösung von nichtlinearen Gleichungen, werden drei Verfahren erarbeitet und anhand von Beispielen verglichen. Der Zwischenwertsatzes für stetige Funktionen garantiert die Existenz von exakten Lösungen.

- Das *Biselektionsverfahren* steht als Vertreter *allgemeiner numerischer Löser*.
- Das *Heronverfahren* zur Approximation von Quadratwurzeln vertritt die *speziellen Löser,* welche auf eine bestimmte Klasse von Gleichungen zugeschnitten und optimiert sind.
- Der *Störungsansatz* mit Linearisierung ist ein Beispiel eines allgemeineren Lösers. Ohne Differentialrechnung bleibt sein Anwendungsbereich auf Polynomgleichungen beschränkt. Dabei sind Polynome mit komplexen Koeffizienten und einer komplexen Variablen kein Tabu.

Einige typische Verhaltensmuster der drei Löser werden an den Beispielen vorgeführt.

Voraussetzungen: Funktionsbegriff, Gleichungslehre aus der Algebra: lineare und quadratische Gleichungen mit einer Unbekannten, Äquivalenzumformungen, binomische Formel.
Elemente der Programmierung, bedingte Schleifen.
Zahlfolgen.
Differentialrechnung wird *nicht* vorausgesetzt.

Ziele: Die Funktionsweise, Stärken und Schwächen numerischer Gleichungslöser kennen lernen und verstehen. Verschiedene Arten von numerischen Lösern konstruieren und damit Erfahrungen sammeln. Lösung und numerische Näherung verstehen und unterscheiden. Konvergenzrate empirisch bestimmen. Numerische Löser aus einer Toolbox, einer mathematischen Software oder einem Taschenrechner kompetent einsetzen.

1 Gleichungen numerisch bearbeiten oder Gleichungen lösen?

Ein Kerngebiet der Algebra sind Gleichungen und Methoden, um ihre Lösungen zu finden. In einfachen Fällen lassen sich die Lösungen mit einer Formel gewinnen. Dieser Fall wird in der Schule oft so stark betont, dass man glauben könnte, er sei typisch. Tatsache ist: Die überwiegende Mehrzahl aller Gleichungen sind der exakten Bearbeitung mit Mitteln der Algebra nicht zugänglich. Wenn algebraische Methoden zur gewünschten Lösung führen, ist dies ein Glücksfall. Leider sind Glücksfälle nicht die Regel.

Oft beschreiben Gleichungen in Anwendungen Beziehungen zwischen verschiedenen gemessenen Grössen, die also selbst wegen der unvermeidlichen Messfehler nicht absolut genau bekannt sind. Ein Beispiel: Kepler fand im Zusammenhang mit der Berechnung von Planetenpositionen die Gleichung $x - e \cdot \sin(x) = m$. Die Grössen e und m konnte er messen, aber er benötigte x. Die Keplergleichung lässt sich nicht algebraisch und exakt lösen.

Wenn die Gleichungen selbst schon Näherungen sind, so gibt es gute Gründe, sie mit Näherungsmethoden zu behandeln. Entscheidend ist, dass durch vorgegebene Toleranzen kontrollierte Näherungslösungen berechnet werden. Es ist für viele Anwendungen wesentlich, hinreichend genaue Näherungen in Dezimalzahldarstellung zu finden. Exakte Lösungen treten selten als Dezimalzahlen auf. Manchmal ist eine rasche numerische Lösung innerhalb einer gegebenen Toleranz gefragt und ein Umweg über die formal exakte Antwort chancenlos.

Numerik befasst sich mit Näherungsverfahren, die mit endlich vielen Rechenoperationen zum Ziel gelangen, beispielsweise im Zusammenhang mit Gleichungen. Numerik und Algebra müssen sich ergänzen. Näherungslösungen bestehen die Einsetzprobe in der gegebenen Gleichung in der Regel nicht. Die Gewissheit, dass im Toleranzbereich einer Näherungslösung eine wirkliche Lösung existiert, verdanken wir meist abstrakten Resultaten aus der Algebra oder der Analysis.

Erinnern wir an einige Begriffe und Tatsachen aus der Algebra: Jede Gleichung mit einer reellen Unbekannten x lässt sich algebraisch in die Gestalt $f(x) = 0$ überführen ohne Veränderung der Lösungsmenge. Dabei bezeichnet $f : x \mapsto f(x)$ eine Funktion. Die *Gleichung lösen* heisst: wir suchen alle Zahlen x im Definitionsbereich von f, welche die *Einsetzprobe* $f(x) = 0$ bestehen.

Beim Berechnen der Näherungen können prinzipiell nur endlich viele Rechenoperationen mit endlich vielen Zahlen auftreten. Genauer: Wir gehen vorerst von einem Rechner aus, der die vier Grundrechenoperationen auf den Maschinenzahlen ausführen kann. Obwohl schon der primitivste numerische Taschenrechner eine Quadratwurzeltaste aufweist, zählen wir die Berechnung von Quadratwurzeln hier *nicht* zu den arithmetischen Grundoperationen. Mit der Quadratwuzeltaste wird eigentlich ein spezieller Löser aktiviert, der auf die Eingabe von $c > 0$ die positive Lösung der Gleichung $x^2 = c$ annähert. Was dabei im Verborgenen abläuft, wird in diesem Abschnitt erklärt werden. Dabei werden wir nur drei Löser betrachten, die sich in Eigenschaften wie allgemeine Anwendbarkeit, Laufzeit, garantierte Genauigkeit unterscheiden werden. Den Nachweis der *Existenz der Lösungen* leistet oft der

Zwischenwertsatz für stetige Funktionen:
Eine stetige Funktion $f : [a, b] \to \mathbb{R}$, für die $f(a) < 0$ und $f(b) > 0$ gelten, besitzt eine Nullstelle x_0 zwischen a und b.

Der Beweis dieses Satzes wird in der Analysis erbracht. Dabei zeigt sich, dass die reellen Zahlen $\mathbb{R}$ genau die Eigenschaften besitzen, die für den Beweis nötig sind. ($\mathbb{R}$ ist ein vollständiger, total geordneter Körper)

Alle unsere konkreten Rechnungen finden jedoch mit physikalisch realisierbaren Zahldarstellungen und Operationen statt, entweder in unserem Gehirn oder in einem Rechner. Dabei benutzen wir zum Beispiel Dezimalzahlen oder die Maschinenzahlen $\mathbb{R}_\mathbb{M} \subset \mathbb{Q}$ anstelle von $\mathbb{R}$. In der Realisierung als Rechnerzahlen $\mathbb{R}_\mathbb{M}$ müssen alle für die reine Mathematik wesentlichen Eigenschaften von $\mathbb{R}$ geopfert werden. Das zwingt die Numerik dazu, praktikable und konstruktive Antworten zu suchen, die oft Näherungen sind. Zudem muss gezeigt werden, wie unvermeidbare Fehler, zum Beispiel Rundungsfehler, unter Kontrolle gehalten werden können. Dass dies oft gelingt, ist fast ein Wunder.

2 Die Gleichung $x^2 - 5 = 0$

Die Funktion $f : x \mapsto x^2 - 5$ ist stetig. Ihr Graph in Abbildung 1 zeigt zwei Nullstellen. Die Gleichung $x^2 - 5 = 0$ lässt sich algebraisch lösen. Die exakten Lösungen sind $\pm\sqrt{5}$. Diese

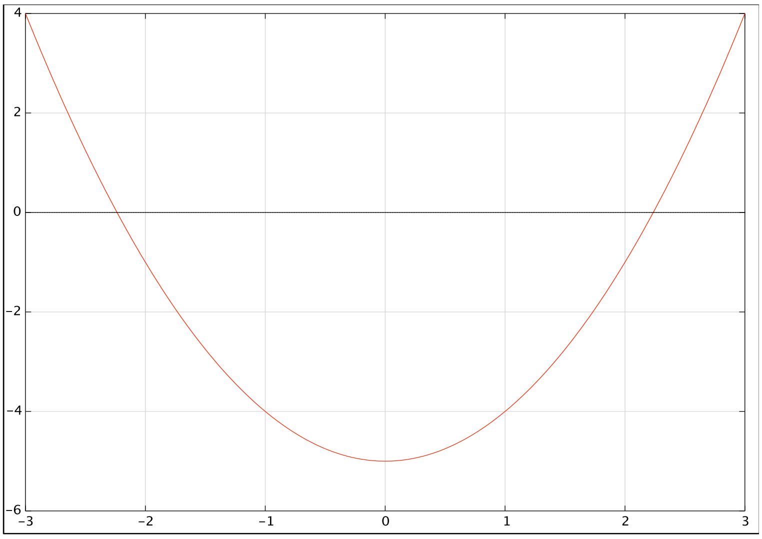

Abbildung 1: Graph der Funktion $f : x \mapsto x^2 - 5$ im Bereich $-3 \leq x \leq 3$

Lösungen sind irrational, aber alle Rechnerzahlen sind Dezimalzahlen und daher rational, $\mathbb{R}_{\mathbb{M}} \subset \mathbb{Q}$. Also gibt es für die Mustergleichung prinzipiell keine echten numerischen Lösungen. Dennoch werden wir diese Gleichung mit drei verschiedenen Methoden numerisch bearbeiten und untersuchen, wie sich die Folgen der Näherungslösungen verhalten.

2.1 Bisektion

Eine kleine Rechnung oder ein Blick auf Abbildung 1 zeigt, dass $f(2) < 0$ und $f(3) > 0$ gelten. Mit dem Zwischenwertsatz schliessen wir, dass es ein x_0 gibt mit $2 < x_0 < 3$ und $f(x_0) = 0$. Die einfache Grundidee des Bisektionsverfahrens ist, das Suchintervall zu halbieren und nachzuweisen, dass die Voraussetzungen für den Existenznachweis auch auf dem kleineren Intervall gelten. Damit lässt sich das Verfahren wiederholt fortsetzen. Es erzeugt eine Folge von abgeschlossenen Suchintervallen $I_0 = [a, b] \supset I_1 \dots I_n \supset I_{n+1}$ deren Länge gegen 0 strebt. Das heisst genau, dass es zu jeder Toleranz $\tau > 0$ einen Index n gibt, so dass die Länge der Suchintervalle I_r für alle $r \geq n$ kleiner sind als τ. Damit ist sichergestellt, dass wiederholte Verbesserung mit der Bisektionsmethode nach endlich vielen Schritten ein so kleines Suchintervall liefert, dass der Mittelpunkt in diesem Intervall die vereinbarte Genauigkeitsforderung zu allen denkbaren Kandidaten, die sich ja in diesm Intervall befinden müssen, erfüllt. Damit ist gezeigt, dass das Bisektionsverfahren zwar im Sinne der reinen Mathematik die Gleichung $f(x) = 0$ nicht löst, aber die abgeschwächte Forderung an eine kontrollierte numerische Näherung erfüllt.

Nach einem Satz der Analysis gilt: Die Suchintervalle I_n, die bei der Bisektion entstehen, haben genau einen Punkt gemeinsam. So lässt sich mit einer ideal unbegrenzt fortgesetzten Folge von Bisektionen eine Lösung der Gleichung nachweisen. Der Zwischenwertsatz ist zwar anschaulich plausibel, aber eben auch in der Analysis beweisbar.

Betrachten wir das Beispiel $x^2 - 5 = 0$ und das Startintervall $[2, 3]$. Die folgende Tabelle zeigt die untere und die obere Intervallgrenze nach n Bisektionsschritten und die Mitte des jeweiligen Intervalles in Dezimaldarstellung D_n, sowie in Binärdarstellung B_n. Die stabilen Ziffern der Dezimaldarstellung sind hervorgehoben. Alle Binärziffern, allenfalls die letzte Stelle ausgenommen, sind stabil.

Tabelle 1: Bisektion mit der Gleichung $x^2 - 5 = 0$

n	a_n	b_n	D_n	B_n
0	2	3	**2**.5	10.1
1	2	2.5	**2**.25	10.01
2	2	2.25	**2**.125	10.001
3	2.125	2.25	**2**.1875	10.0011
...	...	...	...	...
10	...	...	**2.23**58...	10.00111100011
20	...	...	**2.23606**8...	10.001111000110111011111
30	...	...	**2.2360679772**7	10.0011110001101110111100110111001

Wenn wir annehmen, dass die Lösungen im Suchintervall der Bisektion gleichverteilt sind, so wird die Ungewissheit über die Lage der Lösung mit jedem Bisketionsschritt halbiert. Die Information über die Lage einer Lösung, die mit Bisektion gefunden wird, wächst demnach bei jedem Bisektionsschritt um ein Bit. Dieses Bit tritt unmittelbar in Erscheinung, wenn statt Dezimalzahlen Dualzahlen zum Rechnen verwendet werden und wenn das erste Suchintervall, wie in unserem Beispiel die Länge 1 hat. Oder anders als Faustregel ausgedrückt: Entweder stoppt das Bisektionsverfahren, weil einer der Intervallendpunkte eine exakte Lösung ist oder mit je 10 Iterationen werden etwa drei Dezimalziffern an Genauigkeit gewonnen. Beim praktischen Rechner mit dem Computer ist die Zahl der verfügbaren Dezimalstellen begrenzt. Ab einem bestimmten Punkt kann es geschehen, dass die Anfangsziffern der Näherung stabil bleiben und sich nur noch wenige Ziffern am Ende der Zahldarstellung zyklisch wiederholen. Oft tritt Sättigung ein, und alle berechneten Dezimalziffern ändern sich nicht mehr. Das bedeutet aber keineswegs, dass alle stabil angezeigten Ziffern exakt sind.

2.2 Das Verfahren von Heron

Die Streckenrechnung zeigt, wie sich die arithmetischen Grundoperationen und Quadratwurzeln konstruktiv mit Zirkel und Lineal realisieren lassen. Schon sehr früh wurden Verfahren erfunden, um Quadratwurzeln näherungsweise zu berechnen. Eines davon ist unter dem Namen 'Heronverfahren' bekannt.

Folgende Überlegungen begründen die Methode. Angenommen, die zu lösende Gleichung lautet $x^2 = c$. Für $c > 0$ hat diese Gleichung die formalen Lösungen $\pm\sqrt{c}$. Wir wollen die positive Lösung angenähert berechnen und benutzen dazu eine erste geschätzte Näherung s. Gilt $s^2 = c$, so sind wir am Ziel. Sonst ist s entweder zu klein oder zu gross. Die Zahl $t(s) := c/s$ hat die Eigenschaft, dass $s \cdot t(s) = c$ gilt. Folglich kompensiert $t(s)$ den Fehler von s: Die Zahl $t(s)$ ist eine zweite Schätzung für die gesuchte Grösse x, für welche gilt $s \leq x \leq t(s)$ oder $t(s) \leq x \leq s$.

Weil $t(t(s)) = s$ gilt, spielen beide Schätzungen symmetrische Rollen. Wenn wir eine von beiden verbessern, verbesserst sich auch die andere.

Nun kommt eine weitere Einsicht ins Spiel. Die wahre Lösung wird durch das geometrische Mittel von s und t berechnet, $\sqrt{s \cdot t(s)}$, aber diese Zahl lässt sich allein mit den Grundoperationen ohne Wurzeln nicht finden. Wenn jedoch s und $t(s)$ fast gleich sind, so ist das arithmetische Mittel $\frac{1}{2}(s + t(s))$ fast gleich dem geometrischen Mittel von s und $t(s)$. Natürlich liegt das arithmetische Mittel in der Mitte des Intervalles $[s, t(s)]$, das die Lösung einschliesst. Da s und $t(s)$ Randpunkte des Suchintervalles sind, liegt die Mitte näher bei der unbekannten Lösung als die schlechtere der beiden Schätzungen. Weil aber beide Schätzungen dieselbe Rolle spielen, ersetzen wir beide und fahren mit dem Mittelwert als verbesserter Schätzung fort.

In der *Zusammenfassung* umfasst das *Heronverfahren* zum Annähern von Lösungen für $x^2 = c > 0$

1. eine Anfangsschätzung $s_0 > 0$ und die Berechnung einer zweiten Schätzung $t_0 := c/s_0$

2. die Iterationsvorschrift (Heronformel)

$$s_{n+1} := \frac{1}{2}(s_n + \frac{c}{s_n})$$

Das Intervall $[s_n, c/s_n]$ schliesst für jedes $n \geq 0$ die Lösung ein.

Wir probieren das Verfahren aus am Beispiel $x^2 = 5$ mit dem Startwert $s = 2$. Die folgende Tabelle enthält die Ergebnisse dieses Versuchs.

Tabelle 2: Heronverfahren mit der Gleichung $x^2 - 5 = 0$

n	s_n	$t(s_n)$	m_n
0	**2**	2.5	2.25
1	**2.2**5	2.222222...	2.2361111...
2	**2.236**111...	2.236024...	2.23606797792
3	**2.23606797 7**92	2.23606797708...	2.236067977499
4	**2.2360679774997**	2.2360679774997	2.23606797749972
...	...	...	...

In der Spalte 2 von Tabelle 2 sind die stabilen Dezimalziffern hervorgehoben. Die Näherung erfolgt mit einer verblüffenden Effizienz. Als empirische Faustregel zeigt sich, dass beim Heronverfahren die Zahl der gültigen Dezimalziffern mit jedem Schritt verdoppelt wird, soweit dies die Operationen in den Rechnerzahlen $\mathbb{R}_\mathbb{M}$ überhaupt zulassen. Auch das Heronverfahren kann die irrationale Zahl $\sqrt{5}$ nicht in endlich vielen Schritten mit Dezimalzahlen erreichen.

Das Heronverfahren ist auf die Berechnung von Quadratwurzeln zugeschnitten. Besondere Eigenschaften der Aufgabenstellung werden ausgenutzt. Somit ist es möglich, die Rate der Annäherung an die Lösung gegenüber dem Bisektionsverfahren stark zu erhöhen. Das Bisektionsverfahren ist viel allgemeiner anwendbar, in jedem Einzelfall aber nicht besonders effizient. Beide Verfahren lassen sich so formulieren, dass sie nicht nur eine Näherung liefern, sondern auch die Grenzen des Suchintervalles nach jeder Verbesserung angeben. Das ist vorteilhaft, weil sich dann möglicherweise Rundungsfehler oder Sättigung erkennen lassen.

2.3 Störungsansatz und lokale Linearisierung

Angenommen, es ist wiederum die positive Lösung der Gleichung $x^2 - 5 = 0$ anzunähern. Der Störungsansatz geht davon aus, dass sich die gesuchte Lösung darstellen lässt in der Form $x = s + e$. Dabei ist s eine Schätzung für x und $e = x - s$ ist der Schätzfehler. Mit etwas Geschick sollte es gelingen, s so zu bestimmen, dass $|e|$ 'klein' ist. Dann trägt s den Grossteil der Information über x und e enthält eine Korrektur dieser Information. Was aber bedeutet 'klein' im Zusammenhang mit der Gleichung $x^2 = (s + e)^2 = s^2 + 2 \cdot e \cdot s + e^2$? Eigentlich bringt die Einführung von e nur eine zusätzliche Komplikation. Statt einer reinquadratischen Gleichung für x entsteht eine ausgewachsene quadratische Gleichung für e. Wenn wir e 'klein' betrachten, falls $e^2 \approx 0$ gilt, dann ist die neue quadratische Gleichung wenigstens nahe bei der Lösung fast eine lineare Gleichung für e.

$$0 = x^2 - c = s^2 + 2 \cdot e \cdot s + e^2 - c \approx s^2 - c + 2 \cdot e \cdot s$$

Auf der rechten Seite der angenäherten Gleichheit steht eine *lineare* Gleichung für e. Es handelt sich um eine lokal gültige Linearisierung, die ungefähr das Gleiche zum Ausdruck bringt, wie die ursprüngliche nichtlineare Gleichung. Der Vorteil der vereinfachten Gleichung: wir können sie lösen. Zwar erhalten wir nicht den exakten Fehler e sondern bestenfalls eine Näherung

$$\tilde{e} := \frac{c - s^2}{2 \cdot s} \quad \text{oder} \quad \tilde{x} := s + \tilde{e} = s + \frac{c - s^2}{2 \cdot s} = \frac{1}{2}(s + c/s)$$

Es gibt keine Gewähr, dass mit $\tilde{e}$ unter allen Umständen eine Näherung verbessert wird.

Die Rechnung zeigt, dass wir im vorliegenden Beispiel auf einem anderen Weg nochmals zum Heronverfahren gelangt sind. Damit gelten alle Erfahrungen, die das Heronverfahren betreffen, auch für die Mustergleichung und den Störungsansatz. Allerdings hat der Störungsansatz ein viel grösseres Potential für Verallgemeinerungen als die Heronformel, die explizit nur auf die Gleichung $x^2 = c > 0$ passt.

Ist $f : x \mapsto f(x)$ ein beliebiges reellwertiges Polynom und ist s ein Näherungswert für eine reelle Nullstelle, so lässt sich die Gleichung $f(s + e) = 0$ mit der allgemeinen binomischen Formel umformen und auf eine neue Gestalt bringen der Art

$$f(s + e) = f(s) + e \cdot f_1(s) + e^2 \cdot \{\text{Restterme in } e \text{ und } s\}$$

Die lineare Ersatzgleichung für $\tilde{e}$ lautet dann $f(s) + \tilde{e} \cdot f_1(s) = 0$. Damit wird die Schätzung s verbessert zu einer neuen Schätzung

$$\tilde{x} := s - \frac{f(s)}{f_1(s)}$$

Bei Bedarf lässt sich diese Verbesserung beliebig iterieren. Eine *exakte Lösung der Gleichung* $f(x) = 0$ entspricht genau einem *Fixpunkt der Abbildung* $j : x \mapsto x - \frac{f(x)}{f_1(x)}$. Die Iteration $x = j(x)$ kommt gerade auf den Lösungen der Gleichung $f(x) = 0$ zum Stillstand. Allerdings spielt da die noch nicht näher identifizierte Funktion f_1 eine Rolle. Die Algebra verlangt bloss, dass $f1(\tilde{x}) \neq 0$ gilt für alle Näherungen $\tilde{x}$, die in den Iterationen auftreten. Praktisch muss die lineare Gleichung numerisch stabil lösbar sein und bis auf endlich viele Ausnahmen wird $|f_1(\tilde{x})| \gg |f(\tilde{x})|$ notwendig, damit die Iteration konvergiert.

Es ist nun nötig, die Einzelheiten in einem Beispiel zu zeigen. Statt $x^2 - 5 = 0$ wollen wir $f(x) := x^3 - 5 = 0$ mit dem Störungsansatz lösen. Im Intervall $[1, 2]$ erfüllt f die Voraussetzungen des Zwischenwertsatzes, denn $f(1) = 1^3 - 5 < 0$ und $f(2) = 2^3 - 5 > 0$ und f ist stetig. Als erste Schätzung scheint $s_0 := 1.5$ brauchbar. Mit $x = s + e$ liefert die binomische Formel $(s+e)^3 = s^3 + 3 \cdot e \cdot s^2 + 3 \cdot e^2 \cdot s + e^3$. Damit ist $f_1(s) = 3s^2$ und der nichtlineare Rest, der vernachlässigt wird, ist $e^2 \cdot (3 \cdot s + e)$. Die Verbesserungsformel lautet

$$\tilde{s} = s - \frac{s^3 - 5}{3 \cdot s^2} = \frac{1}{3}(s + 5 \cdot s^{-2})$$

Wir starten die Näherung mit $s_0 := 1.5$ und beobachten die von der Iteration erzeugten Näherungen:

$$1.74074074\ldots \quad 1.71051\ldots \quad 1.709976\ldots \quad 1.7099759466767$$

Die Superkonvergenz, die wir im Heronverfahren beobachten konnten, scheint sich auch beim Störungsansatz für die rein kubische Gleichung zu bestätigen. Allerdings ist zu bemerken, dass unsere Version des Heronverfahrens mit jeder Näherung auch gleich ein Intervall lieferte, das die wahre Lösung enthält. Dieses Intervall fehlt beim Störungsansatz. Es müsste speziell konstruiert werden.

3 Übungen

Wenn sich einfache Gleichungen formal exakt lösen lassen, besteht in der Regel kein Bedarf an numerischen Näherungen. In den folgenden Aufgaben sollen aber numerische Verfahren an einfachen Beispielen getestet werden. Wenn eine Gleichung formal exakte Lösungen hat, lassen sie sich zum Überprüfen der Näherungen verwenden. Vorsicht: Sobald eine formal exakte Lösung numerisch ausgewertet wird, kann sie selbst zu einer Näherung werden.

1. Schreiben Sie je ein einfaches Programm, das nach Eingabe von c und allenfalls weiteren benötigter Startdaten Näherungslösungen der Gleichung $x^2 - c = 0$, $c > 0$ berechnet. Wie geht das mit

   (a) dem Bisektionsverfahren?

   (b) dem Heronverfahren?

   (c) dem Störungsansatz?

   Dabei soll die maximale Zahl der Verbesserungsschritte wählbar sein und jede Verbesserung soll zusammen mit den Grenzen des Suchintervalles ausgegeben werden.

   Hinweis: Die Programme sollen auf einem Ihnen verfügbaren Hilfsmittel (Taschenrechner, Laptop, Desktop) lauffähig sein. Für die Aufgabe 2 lassen sich die Programme von Aufgabe 1 mit geringem Aufwand anpassen. Es lohnt sich nicht, einen allgemeinen Löser zu programmieren, der Aufwand wäre für unsere Anwendungen zu gross. Wir benötigen die Programme nur kurzfristig, um erste Erfahrungen mit numerischen Lösern zu sammeln.

2. Bearbeiten Sie die Gleichung $x^5 - 2 = 0$

   (a) mit dem Bisketionsverfahren

   (b) mit einer für Sie naheliegenden Verallgemeinerung der Ideen, die zum Heronverfahren geführt haben.

   (c) mit dem Störungsansatz

Versuchen Sie, in jedem der Fälle zu bestimmen, wie rasch die Lösung angenähert wird. Wie lässt sich ein Mass für die ungefähre Annäherungsrate finden, ohne dass die exakte Lösung bekannt sein muss.

3. Was taugen folgende Methoden bei der Gleichung $2^x - 5 = 0$?

   (a) das Bisketionsverfahren

   (b) der Störungsansatz

   Versuchen Sie, in jedem der Fälle zu bestimmen, ob und wie rasch die Lösung angenähert wird.

4. Wie viele Lösungen hat die Gleichung $x^2 = 2^x$? Berechnen Sie Näherungen für alle Lösungen bei einer Toleranz von 10^{-6} ?

5. Lässt sich Bisektion beschleunigen dank folgender Beobachtung? Das Bisektionsverfahren nützt aus der Einsetzprobe für die Gleichung $f(x) = 0$ nur die Vorzeicheninformation, nicht aber die Grösse des Betrages der einzelnen Funktionswerte.

   Angenommen, $f(a) < 0$, $f(b) > 0$ und $|f(a)|$ ist deutlich kleiner als $|f(b)|$. Wenn der Graph von f zwischen a und b nahezu geradlinig verläuft, könnte es sich lohnen den neuen Testpunkt x_* als Schnitt der Verbindungsstrecke von $(a|f(a))$ und $(b|f(b))$ mit der x-Achse zu ermitteln, statt die Mitte zwischen a und b zu verwenden. Was bringt es, wenn das Bisektionsverfahren mit x_* fortgesetzt? Betrachten Sie die drei Testfälle aus der Aufgabenserie und untersuchen Sie, ob das neue Verfahren eine Folge von Näherungen erzeugt, die deutlich rascher gegen die Lösung strebt als das Bisektionsverfahren.

6. Eine Holzkugel der Dichte $\rho = 800$ kg $\cdot$ m^{-3} schwimmt auf Wasser der Dichte $\rho_0 = 10^3$ kg $\cdot$ m^{-3}. Der Kugelradius beträgt $r = 1$ m. Wie tief sinkt die Kugel ins Wasser ein?

   Hinweis: Wie lautet die Bedingung für die Schwimmlage der Kugel nach dem archimedischen Prinzip? Lässt sich die Aufgabe algebraisch exakt lösen? Versuchen Sie es mit einem Computer-Algebra-System. Was zeigt dessen Antwort?

7. Die einzige Nullstelle von $q : x \mapsto x^2$ liegt bei 0.

   (a) Skizzieren Sie den Verlauf des Graphen von q in der Nähe von 0 qualitativ richtig.

   (b) Warum eignet sich Bisektion nicht zum Lösen von $x^2 = 0$?

   (c) Welche Folge von Näherungen ergibt sich, wenn das Heronverfahren mit Startwert $x_0 = 1$ zum Lösen von $x^2 = 0$ verwendet wird? Welches Verhalten der Näherungen beobachten Sie?

8. Die einzige Nullstelle von $k : x \mapsto x^3$ liegt bei 0.

   (a) Skizzieren Sie den Verlauf des Graphen von k in der Nähe von 0 qualitativ richtig.

   (b) Findet Bisektion die Lösung von $x^3 = 0$ ausgehend von den Startwerten -1 und 2 ? Wie verhält sich das Verfahren, wenn die Startwerte -1 und 1 benutzt werden?

   (c) Welche Folge von Näherungen ergibt sich, wenn der Störungsansatz und Linearisierung mit Startwert $x_0 = 1$ zum Lösen von $x^3 = 0$ verwendet wird? Welches Verhalten der Näherungen beobachten Sie?

9. [Diese Aufgabe setzt Vertrautheit mit den komplexen Zahlen $\mathbb{C}$ voraus]
Das Heronverfahren lässt sich formell auch auf die Gleichung $x^2 + 1 = 0$ anwenden.

   (a) Welche Folgen erzeugt das Verfahren zu verschiedenen reellen Startwerten? [Literatur: Gilbert Strang, Calculus].
   (b) Warum lassen sich die komplexen Lösungen nicht mit reellen Startwerten erreichen?
   (c) Welches Verhalten der Näherungsfolge beobachten Sie, wenn der komplexe Startwert $z_0 := 1 + 0.01 \cdot i$ gewählt wird?
   (d) Welches Verhalten der Näherungsfolge beobachten Sie, wenn der komplexe Startwert $z_0 := 1 + i$ für das Heronverfahren bei der Gleichung $x^2 - 2 = 0$ gewählt wird?

10. Versuchen Sie, die Lösungen der folgenden Gleichungen in der komplexen Ebene $\mathbb{C}$ mit dem Störungsansatz so anzunähern, dass die Fehler im Realteil und Imaginärteil der Näherungslösung je geringer ist als 10^{-4}.

    Wie lautet die zur Gleichung gehörige Iterationsvorschrift, die sich aus dem Störungsansatz ergibt?

   (a) $z^4 - iz - 16 = 0$
   (b) $z^3 + (2 - i)z - 4i = 0$

4 Zusammenfassung und Schlussfolgerungen

Das Bisektionsverfahren ist ein allgemeines Werkzeug, um Gleichungen $f(x) = 0$ zu lösen, wenn $f : [a, b] \longrightarrow \mathbb{R}$ eine stetige Funktion ist und die Vorzeichen von $f(a)$ und $f(b)$ verschieden sind. Das Verfahren setzt bloss voraus, dass sich die die Vorzeichen der Funktionswerte effektiv und exakt bestimmen lassen. Das ist unter anderem wegen der Rundungsfehler beim Rechnen mit Maschinenzahlen keine harmlose Forderung, sobald $f(x) \approx 0$ gilt.

Mit jedem Bisektionsschritt wird ein Bit an Information über die Lage der Nullstelle in $[a, b]$ gewonnen. Beim Abbruch des Verfahrens ist zum Näherungswert für die Nullstelle auch ein Intervall bekannt, in dem sich die Nullstelle befindet (natürlich nur bis zur Genauigkeit der Zahldarstellung des verwendeten Rechners). Das Verfahren ist anspruchslos und daher allgemein einsetzbar, aber es ist nicht besonders effizient, weil es nur die Stetigkeit der Funktion f und eine isolierte Nullstelle voraussetzt, die lokal trennt zwischen positiven und negativen Funktionswerten.

Das Heronverfahren ist ein Beispiel für ein sehr spezielles Verfahren zum angenäherten Berechnen von Quadratwurzeln. Die Konvergenzordnung ist quadratisch. Verallgemeinerungen auf Wurzeln der Ordnung $n > 2$ sind möglich, jedoch ist die Konvergenz der naheliegendsten Verallgemeinerunge nur noch linear. Das Verfahren liefert aber auch zu jeder Näherung ein Intervall, das die wahre Lösung 'sicher' enthält.

Der Störungsansatz zerlegt eine Lösung x einer Gleichung $f(x) = 0$ in $x := s + \varepsilon$. Dabei wird angenommen, dass s nahe bei der Lösung liegt. Das soll bedeuten, dass der Fehler $\varepsilon := x - s$ im folgenden Sinne *klein* sei: Für alle natürlichen Zahlen $n \leq N$ soll die Näherung

$$(s + \varepsilon)^n \approx s^n + n \cdot \varepsilon \cdot s^{n-1}$$

anstelle der allgemeinen binomischen Formel verwendbar sein.

Ist nun f eine Polynomfunktion vom Grad N, so wird x durch $s + \varepsilon$ ersetzt. Die Näherungen $x^n = (s + \varepsilon)^n \approx s^n + n \cdot s^{n-1} \cdot \varepsilon$ für $2 \leq n \leq N$ ergeben beim Einsetzen in der Polynomgleichung $f(x) = 0$ eine *lineare Gleichung* für ε. Für fast jede Wahl von s hat die linearisierte Gleichung genau eine Lösung; sie heisse $\tilde{\varepsilon}$, da sie nicht mit der eigentlich gesuchten Grösse ε übereinzustimmen braucht. Also kann die Schätzung s in der Regel bloss verbessert werden zu einer neuen Schätzung $s_+ := s + \tilde{\varepsilon}$.

Iteration des Verfahrens liefert jedoch bei guten Startwerten s und bei Nullstellen erster Ordnung eine Folge von quadratisch konvergenten Näherungen. Die Methode ist auch für Polynome über $\mathbb{C}$ tauglich.

Der Störungsansatz lässt sich weitgehend verallgemeinern indem die Linearisierung nicht mehr algebraisch, sondern mit Mitteln der analytischen oder numerischen Differentialrechnung gefunden wird. Dabei gelangt man zum Verfahren von Newton-Raphson und seinen Abkömmlingen. Ohne besondere Vorkehrungen liefert die Methode kein Suchintervall, in dem sich eine Näherung und eine Lösung garantiert befinden.

Es kann sinnvoll sein, Bisektion mit dem Störungsansatz zu verbinden, dann nämlich, wenn die Startwerte für den Störungsansatz noch zu grob sein sollten.

Der Störungsansatz lässt sich auch für Funktionen in $\mathbb{R}^n$ oder in $\mathbb{C}^n$ formulieren. Die Störung ist dann ein Vektor $\vec{\varepsilon} \in \mathbb{R}^n$. Die angenäherte Bestimmung von $\vec{\varepsilon}$ führt auf die Lösung eines Systems von n linearen Gleichungen für die n Unbekannten.

5 Hinweise zu den Lösungen

Die folgenden Hinweise sind technisch und knapp formuliert. Es sind keine Musterlösungen für die Lernenden. Gute Lösungen der Aufgaben sollen persönliche Kommentare, Beobachtungen, Folgerungen enthalten, die belegen, dass beim Lösen möglicherweise mehr geschieht als die Reproduktion von Mustern, die vorher im Unterricht aufgetreten sind. Einige Aufgaben zeigen gerade, dass der enge Pfad der einführenden Beispiele verlassen werden kann und dass im erweiterten Umfeld neue Einsichten zur Einschätzung der Verfahren möglich sind.

Programme werden nur in einer Minimalversion in Pseudocode mitgeteilt. Diese Versionen sind auch auf den in der Schule gängigen programmierbaren Taschenrechnern lauffähig.

1. **bisect**(f(x),a,b,itr)
   ```
   sgn(f(a)) → sa
   sgn(f(b)) → sb
   If sa * sb < 0 Then
   (a + b)/2 → m
   else
   return "kein Vorzeichenwechsel"
   EndIf
   For j,1,itr,1
   sgn(f(m)) → sm
   if sm = 0 Then
   return sm
   ElseIf sm * sa < 0 Then m → b
   Else
   m → a
   EndIf
   EndFor
   Return m
   ```

   heron(c,s,itr)
   ```
   For j,1,itr,1
   (s + c/s)/2 → s
   EndFor
   Return s
   ```

2. Das *Bisektionsverfahren* zeigt ein Verhalten analog zum Fall der quadratischen Gleichung.

   Das *Heronverfahren* lässt sich auf die Gleichung $x^3 = c$ verallgemeinern, indem die zweiten Schätzung mit $t(s) := c/s^2$ berechnet wird. Diese Variante zeigt jedoch eine langsame Konvergenz (linear). Eine analoge Verallgemeinerung für die Gleichung $x^5 = c$ mit $t(s) := c/s^4$ konvergiert nicht.

   Der *Störungsansatz* liefert ein Verfahren, das quadratisch gegen die Lösung $\sqrt[5]{2}$ konvergiert. Iteration mit $g(x) = x - (x^5 - 2)/(5 \cdot x^4)$

   Ein Mass zum Beurteilen der Konvergenzeigenschaft einer Folge von Zahlen $a(n)$, die sich einem unbekannten Grenzwert nähern:

   $$q(n) := \frac{a(n+1) - a(n)}{a(n) - a(n-1)}$$

Bei linearer Konvergenz wird $|q| > 0$ asymptotisch konstant. Für das Bisektionsverfahren ist $|q| = 1/2$.

Variante: Graphische Darstellung der Datenpaare $(n|a(n))$.

3. Das *Bisektionsverfahren* liefert Näherungswerte. Es zeigt in der Verallgemeinerung einen Weg auf, um allgemein $\log_2(a)$ für $a > 0$ numerisch zu approximieren.

   Das *Heronverfahren* ist nicht anwendbar.

   Der *Störungsansatz* ist im Rahmen der angenommenen Vorkenntnisse nicht unmittelbar anwendbar. Wer Differentialrechnung kennt und die Funktion $\exp_2 : x \mapsto 2^x$ ableiten kann, findet eine brauchbare Linearisierung

   $$2^{x+\varepsilon} = 2^x \cdot 2^\varepsilon \approx 2^x \cdot (1 + \varepsilon \cdot \ln(2))$$

   Dieser Weg endet in einer etwas absurden Sackgasse: Es ist plausibel anzunehmen, dass, wer über numerische Näherungen für $\ln(2)$ verfügt, auch Zugang zu Näherungen für $\ln(5)$ hat und damit die formal exakte Lösung $\log_2(5) = \ln(5)/\ln(2)$ durch eine einzige Division annähern kann.

4. Die Graphen der beiden Funktionen $q : x \mapsto x^2$ und $e : x \mapsto 2^x$ haben drei Schnittpunkte gemeinsam. Die x-Koordinaten der Schnittpunkte sind $x_1 \approx -0.7666647$ (mit Bisektion), $x_2 = 2$, $x_3 = 4$ (z.B. anhand der Graphik vermutet oder erraten und mit Einsetzprobe geprüft).

5. Die Verbesserung ist bekannt als *regula falsi.* Mittelalterliche Rechenmeister wie Fibonacci und Adam Riese haben sie oft benutzt.

6. Es sei $x > 0$ die Höhe [in Metern] des Kugelsegmentes, das ins Wasser taucht und dabei so viel Wasser verdrängt, dass dessen Gewicht gleich dem Gewicht der Kugel ist. Algebraisch formuliert ist die Aussage äquivalent zu $x^3 - 3 \cdot x^2 + 3.2 = 0$

   Diese Gleichung ist formal exakt lösbar. Das wxMaxima CAS liefert:

   $$\begin{aligned}
   x_1 &= \frac{\frac{\sqrt{3}\,i}{2} - \frac{1}{2}}{\left(\frac{4\,i}{5} - \frac{3}{5}\right)^{\frac{1}{3}}} + \left(\frac{4\,i}{5} - \frac{3}{5}\right)^{\frac{1}{3}} \left(-\frac{\sqrt{3}\,i}{2} - \frac{1}{2}\right) + 1 \\
   x_2 &= \left(\frac{4\,i}{5} - \frac{3}{5}\right)^{\frac{1}{3}} \left(\frac{\sqrt{3}\,i}{2} - \frac{1}{2}\right) + \frac{-\frac{\sqrt{3}\,i}{2} - \frac{1}{2}}{\left(\frac{4\,i}{5} - \frac{3}{5}\right)^{\frac{1}{3}}} + 1 \\
   x_3 &= \left(\frac{4\,i}{5} - \frac{3}{5}\right)^{\frac{1}{3}} + \frac{1}{\left(\frac{4\,i}{5} - \frac{3}{5}\right)^{\frac{1}{3}}} + 1
   \end{aligned}$$

   Was sich mit dem CAS(!) in eine besser verständliche Form bringen lässt

   $$\begin{aligned}
   x_1 &= \sqrt{3}\sin\left(\frac{\pi - \operatorname{atan}\left(\frac{4}{3}\right)}{3}\right) - \cos\left(\frac{\pi - \operatorname{atan}\left(\frac{4}{3}\right)}{3}\right) + 1 \\
   x_2 &= -\sqrt{3}\sin\left(\frac{\pi - \operatorname{atan}\left(\frac{4}{3}\right)}{3}\right) - \cos\left(\frac{\pi - \operatorname{atan}\left(\frac{4}{3}\right)}{3}\right) + 1 \\
   x_3 &= 2\cos\left(\frac{\pi - \operatorname{atan}\left(\frac{4}{3}\right)}{3}\right) + 1
   \end{aligned}$$

Hier stellen sich Fragen: Welche Einsicht vermitteln diese Lösungen? Gibt es physikalisch drei Lösungen?

Weil die Daten approximativ sind, ist eine numerische Näherung zweckmässig.

Die numerische Näherungslösung mit dem Störungsansatz ist abhängig vom Startwert. Sicher ist $1 < x < 2$, womit sich beispielsweise ein plausibler Startwert $x_0 := 1.5$ finden lässt. Er liefert mit dem Störungsansatz nach wenigen Schritten die Näherung $x \approx 1.426$ m.

Weitere Lösungen der Gleichung lauten: $x \approx -0.9052$ und $x \approx 2.4795$. Sie sind geometrisch ausgeschlossen oder physikalisch nicht sinnvoll.

7. Doppelte Nullstelle in 0 mit einseitiger Berührung der Achse.

   Die Voraussetzungen für *Bisektion* sind nicht gegeben, kein Vorzeichenwechsel auf den beiden Seiten der Nullstelle. Dennoch kann Bisektion die Nullstelle zufällig finden, wenn als Anfangsdaten a und $-a$ benutzt werden.

   Das *Heronverfahren* mit Startwert $s \neq 0$ funktioniert, aber es konvergiert nur noch linear, ebenso der *Störungsansatz.*

8. *Bisektion* bewältigt das Beispiel, solange das Vorzeichen von x^3 für $x \approx 0$ noch korrekt erkannt wird. Die symmetrisch zu 0 gelegenen Startwerte 1 und -1 führen sofort zur Nullstelle in 0.

   Modifiziertes *Heronverfahren* und *Störungsansatz* sind möglich, aber die Konvergenz der Näherungen ist linear.

   Mit dem Störungsansatz lautet die Rekursion $x_{n+1} := \frac{2}{3} \cdot x_n$. Der Konvergenzfaktor ist 2/3. Zum Startwert $x_0 := 1$ gehört die geometrische Folge $x_n := (2/3)^n$. Weil $2/3 > 1/2$ ist, konvergiert sie langsamer gegen 0 als Bisektion mit beliebigen Startwerten $x_0 < 0$ und $x_1 > 0$.

   [Allgemein: mehrfache Nullstellen bereiten numerischen Lösern oft Schwierigkeiten. Numerische Löser funktionieren am besten bei regulären Nullstellen von differenzierbaren Funktionen, das sind Nullstellen, wo der Graph die Abszisse transversal schneidet. Formal formuliert: Bei jeder regulären Nullstelle x_0 ist $f(x_0) = 0$ und $f'(x_0) \neq 0$.]

9. Das *Heronverfahren* mit reellen Startwerten liefert nur reelle Antworten, weil es nur die Operationen im Zahlkörper $\mathbb{R}$ benutzt. Das Verfahren kann in $\mathbb{R}$ nicht konvergieren, denn die Fixpunktgleichung wird nur von Lösungen der zugehörigen Gleichung erfüllt. Entweder bricht das Verfahren ab, weil 0 erreicht wird und der nächste Schritt nicht ausführbar ist, oder die Folge wird zyklisch oder chaotisch werden. Chaotisches Verhalten kann kein deterministisches Rechnerexperiment beweisen, weil $\mathbb{R}_{\mathbb{M}}$ endlich ist und jede Iteration zyklisch endet oder abbricht.

   Mit komplexen Startwerten $x_0 \notin \mathbb{R}$ konvergiert das Heronverfahren jedoch gegen eine der komplexen Lösungen $\pm i$.

   Bei der Gleichung $x^2 - 2 = 0$ konvergiert die Folge der Näherungen mit komplexen Startwerten, die nicht rein imaginär sind, gegen die reellen Lösungen.

10. (a) Iterationsvorschrift: $z_+ := z - \dfrac{z^4 - iz - 16}{4z^3 - i}$

    vier Lösungen, zwei davon sind rein imaginär. Die Lösungen liegen in der Nähe der Lösungen der vereinfachten Gleichung $z^4 - 16 = 0$ mit Lösungen $\pm 2,\ \pm 2i$

(b) Iterationsvorschrift: $z_+ := z - \frac{z^3 + (2-i)z - 4i}{3z^2 + 2 - i}$

drei Lösungen, Näherungswerte sind:
$-0.20023 - 2.0041i \qquad -0.91031 + 0.80505i \qquad 1.1105 + 1.1990i$

Hinweis: In Verallgemeinerung von Aufgaben 9 und 10 ist es interessant zu untersuchen, wie die Startwerte der Iterationsfolgen und allfällige Fixpunkte zusammenhängen.