

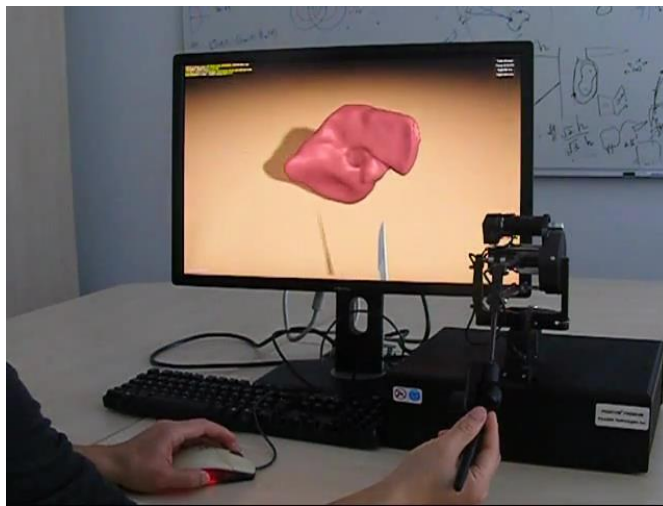
# Angewandte Mathematik

## Numerische Berechnungen

Univ.-Prof. Dr. Matthias Harders  
Dr. Marcel Ritter  
Sommersemester 2022



## Einführungsfilm



## Inhalt

- Einführung
- Zahlendarstellung und Fehlertheorie
- Auswertung von Funktionen
- Numerische Differentiation und Integration



## Annäherung der Kreiszahl $\pi$

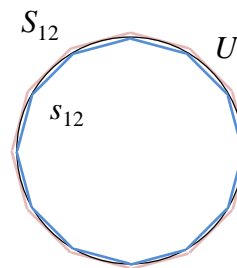
- Methode nach Archimedes: Approximation über ein- und umgeschriebene  $n$ -seitige reguläre Polygone
- Startpunkt: Kreiszahl  $\pi$  als Verhältnis von Kreisumfang zu Kreisdurchmesser

$$\pi = \frac{U}{d} = \frac{2\pi r}{2r}$$

$$s_{12} < U < S_{12}$$

- Allgemein:

$$s_n < s_{2n} < U < S_{2n} < S_n$$



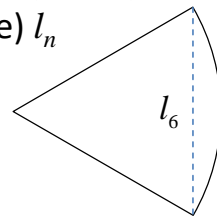
## Annäherung der Kreiszahl $\pi$

- Durch Verdoppelung der Seiten nähern sich die Umfänge der ein- und umgeschriebenen Polygone dem Umfang; bei einem Kreis mit  $d = 1$  somit  $\pi$

$$\lim_{n \rightarrow \infty} s_n = \lim_{n \rightarrow \infty} S_n = \frac{U}{1} = \pi$$

- Umfang eines eingeschriebenen  $n$ -seitigen regulären Polygons, mit Seitenlänge (Kreissehne)  $l_n$

$$s_n = n \cdot l_n$$



## Annäherung der Kreiszahl $\pi$

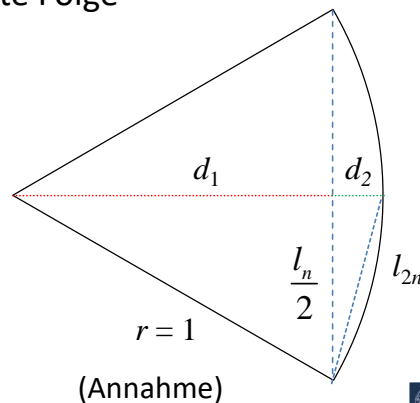
- Exhaustionsmethode: rekursive Berechnung der Sehnenlänge als konvergente Folge

$$\left(\frac{l_n}{2}\right)^2 + d_1^2 = 1$$

$$\Rightarrow d_1 = \frac{1}{2} \sqrt{4 - l_n^2}$$

$$l_{2n}^2 = (1 - d_1)^2 + \left(\frac{l_n}{2}\right)^2$$

$$\Rightarrow l_{2n} = \sqrt{2 - \sqrt{4 - l_n^2}}$$



## Annäherung der Kreiszahl $\pi$

- Start: eingeschriebenes Hexagon, für  $r = 1$  gilt  $l_6 = 1$

Iteration $i$	Seiten $n$	Sehnenlänge $l_n$	Umfang $s_n = n \cdot l_n$
1	6	1	6
2	12	0.51763809	6.211657082
3	24	0.261052384	6.265257227
4	48	0.130806258	6.278700406
5	96	0.065438166	6.282063902
6	192	0.032723463	6.282904945
7	384	0.016362279	6.283115216

$$(s_{n=7} \approx 2\pi \Rightarrow \pi \approx 3.141557608)$$

- Vorgehen für umschriebene Polygone analog



## Weitere Näherungen der Kreiszahl $\pi$

- Archimedes:  $3\frac{10}{71} < \pi < 3\frac{1}{7} = \frac{22}{7}$
- Weitere (grobe) Näherungen:  $\pi \approx \sqrt{2} + \sqrt{3} \approx \frac{333}{106} \approx \frac{355}{113}$
- Möglich als Reihendarstellung, z.B. „Basler-Problem“

$$\frac{\pi^2}{6} = \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \dots$$

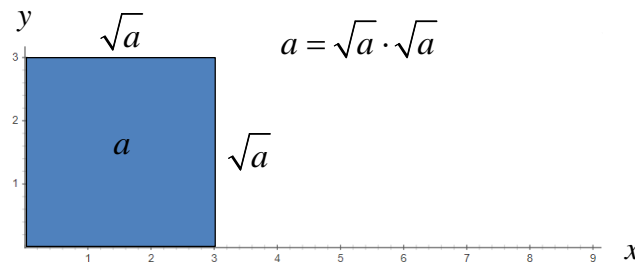
- Chudnovsky-Formel (aktuell ~31 Billionen Stellen für  $\pi$ )

$$\frac{1}{\pi} = 12 \sum_{k=0}^{\infty} \frac{(-1)^k (6k)! (13591409 + 545140134k)}{(3k)! (k!)^3 640320^{3k+3/2}}$$



## Berechnung von Wurzeln

- Methode nach Heron: Berechnung Quadratwurzel durch iterative Annäherung an ein Quadrat (als Form)
- Gesucht:  $b = \sqrt{a}$ ,  $a \geq 0$
- Konstruktion Quadrat mit Flächeninhalt  $a$

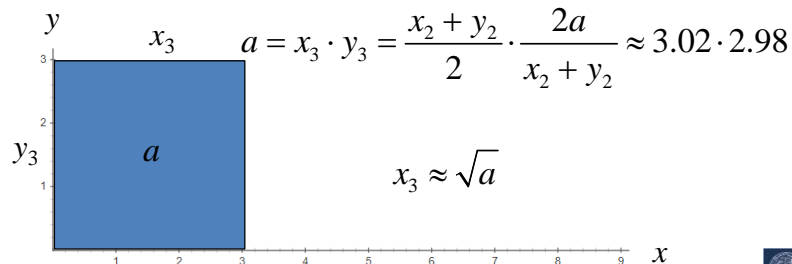


$$a = \sqrt{a} \cdot \sqrt{a}$$



## Berechnung von Wurzeln

- Methode nach Heron: Berechnung Quadratwurzel durch iterative Annäherung an ein Quadrat (als Form)
- Gesucht:  $b = \sqrt{a}$ ,  $a \geq 0$
- Konstruktion Quadrat mit Flächeninhalt  $a$



$$a = x_3 \cdot y_3 = \frac{x_2 + y_2}{2} \cdot \frac{2a}{x_2 + y_2} \approx 3.02 \cdot 2.98$$

$$x_3 \approx \sqrt{a}$$



## Berechnung von Wurzeln

- Seitenlängen für Schritt  $n$

$$x_n = \frac{x_{n-1} + y_{n-1}}{2}, \quad y_n = \frac{2a}{x_{n-1} + y_{n-1}} = \frac{a}{x_n}$$

- Iterationsvorschrift

$$x_n = \frac{x_{n-1} + a/x_{n-1}}{2} = \frac{1}{2} \left( x_{n-1} + \frac{a}{x_{n-1}} \right)$$

- Die Folge  $(x_n)$  konvergiert für  $x_0 > 0$

$$\lim_{n \rightarrow \infty} x_n = \sqrt{a}$$



## Berechnung von Wurzeln

- Alternative Herleitung mittels Newton-Verfahren

- Gesucht:  $b = \sqrt{a}, \quad a \geq 0 \quad \Rightarrow b^2 - a = 0$

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

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad f'(x) = 2x$$

$$\Rightarrow x_{n+1} = x_n - \frac{x_n^2 - a}{2x_n} = x_n - \frac{x_n}{2} + \frac{a}{2x_n} = \frac{1}{2} \left( x_n + \frac{a}{x_n} \right)$$

- Äquivalent kann eine Vorschrift zur Berechnung der  $k$ -ten Wurzel bestimmt werden



## Numerische Mathematik (Numerik)

- Entwicklung sowie Analyse von Algorithmen zur Lösung mathematischer Problemstellungen
- Häufig durch Bestimmung von Näherungslösungen
- In der Informatik insbesondere Fokus auf praktischer Umsetzung auf Rechnersystemen
- Einsatz zur Lösung kontinuierlicher Probleme u.a. aus Natur-, Wirtschafts-, Ingenieurwissenschaften
- Beispiele: Wetterprognosen, Baustatik, Entwicklung von Fahr-/Flugzeugen, elektromagnetische Strahlung, Molekulardynamik, Astrophysik, etc.



## Zentrale Themen der Numerik

- Zahlendarstellungen, Fehleranalyse
- Numerik von Differential- und Integralgleichungen
- Numerisches Lösen von Gleichungssystemen
- Konvergenzanalyse und -beschleunigung
- Approximation und Regression
- Interpolation und Extrapolation
- Lösung von Optimierungsproblemen
- High Performance Computing, Parallelisierung



## Typische Schritte einer Problemlösung

- 1) Reale Problemstellung aus Anwendungsgebiet
- 2) Mathematische Modellierung und Analyse
- 3) Entwicklung eines numerischen Verfahrens
- 4) Implementierung in Code
- 5) Praktische Berechnung, teils auf spezieller Hardware
- 6) Interpretation der Ergebnisse (u.a. Visualisierung)
- 7) Vergleich Simulation und Realität



## Mögliche Fehlerquellen

- **Modellierungsfehler** (z.B. vereinfachende Annahmen)
- **Datenfehler** (z.B. Messungenauigkeit bei Parametern, Verwendung falscher Maßeinheiten)
- **Quantisierungsfehler** (z.B. A/D Wandlung)
- **Rundungsfehler** (z.B. Gleitkommaarithmetik)
- **Verfahrensfehler** (z.B. Abbruch unendlicher Reihen)
- **Implementierungsfehler** (z.B. Fehler im Code)
- **Rechnerfehler** (z.B. Soft Errors, kosmische Strahlung)
- **Interpretationsfehler** (z.B. fehlerhafte oder fehlende statistische Analyse der Resultate)





## Gleitkommazahlen

- Eine Gleitkommazahl wird dargestellt als  $x = \pm m \cdot b^e$ , mit Mantisse  $m$ , Basis  $b$ , und Exponent  $e$ ; wobei

$$m = \sum_{i=1}^k m_i \cdot b^{-i} \quad e = \pm \sum_{i=0}^{l-1} e_i \cdot b^i \quad e_{\min} \leq e \leq e_{\max}$$

$$b, k, l \in \mathbb{N}, \quad b \geq 2, l > 0 \quad m_i, e_i \in \{0, 1, \dots, b-1\}$$

- In Rechnersystemen wird in der Regel  $b = 2$  verwendet, sowie dann  $m_1 \neq 0$  verlangt (Normalisierung, für  $x \neq 0$ )
- Für Zahl Null:  $m = e = 0$ ; sowie Bit  $m_1 = 1$  oft implizit
- Die somit endliche Menge der Maschinenzahlen  $M(b, k, l)$  stellt nur eine Teilmenge von  $\mathbb{R}$  dar

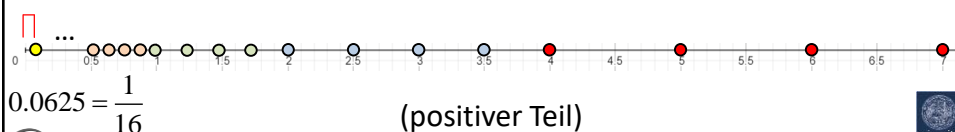


## Gleitkommazahlen

- Die vorgestellten Gleitkommazahlen liegen nicht gleichmäßig auf der Zahlengeraden
- Beispiel:  $M(2, 3, 2)$

Mantissen (binär)	Werte (dezimal)
100	$\frac{1}{2}$
101	$\frac{1}{2} + \frac{1}{8} = \frac{5}{8}$
110	$\frac{1}{2} + \frac{1}{4} = \frac{3}{4}$
111	$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8}$

Exponenten (binär)	Exponenten (dezimal)	Werte $2^e$ (dezimal)
$\pm 00, \pm 01, \pm 10, \pm 11$	$-3, -2, -1, \pm 0, 1, 2, 3$	$\frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8$



## Rundungsfehler

- Umrechnung dezimale Nachkommazahl zu Binärzahl

Nachkommazahl $x$	$2 \cdot x$	Bits
0.6875	1.375	1
0.375	0.75	0
0.75	1.5	1
0.5	1.0	1

$$0.6875_{(10)} = 0.1011_{(2)}$$

Nachkommazahl $x$	$2 \cdot x$	Bits
0.2	0.4	0
0.4	0.8	0
0.8	1.6	1
0.6	1.2	1
0.2	0.4	0

$$0.2_{(10)} = 0.\overline{0011}_{(2)}$$



## Rundungsfehler – IEEE 754 Format

- Im IEEE 754 Standard, Typ *Single*, ist  $e_{\min} = -126$  und  $e_{\max} = 127$  (realisiert über Subtraktion von Bias  $B = 127$ )
- Denormalisierte Werte mit  $e = 0$ ,  $m \neq 0$  sind erlaubt, sowie  $e = e_{\max} + 1$ ,  $m = 0$ :  $\pm\infty$  und  $e = e_{\max} + 1$ ,  $m > 0$ : NaN

Zahl	$e$	Binär (Mantisse, 23 Bit)	Maschinenwert (dezimal)
0.1	-4	10011001100110011001101	0.100000001490116119384765625
0.2	-3	10011001100110011001101	0.20000000298023223876953125
0.3	-3	00110011001100110011010	0.300000011920928955078125
0.4	-2	10011001100110011001101	0.4000000059604644775390625
0.5	-1	00000000000000000000000	0.5
0.7	-1	01100110011001100110011	0.699999988079071044921875



## Rundung Gleitkommazahlen

- Da  $M(b,k,l) \subset \mathbb{R}$ , wird eine Rundungsabbildung benötigt für Zahlen  $\xi \in \mathbb{R}$ , falls  $\xi \notin M(b,k,l)$  (unten kurz  $M$ )

$$rd : \mathbb{R} \rightarrow M, \quad rd(\xi) = x, \quad x \in M$$

- Für die Abbildung soll auch gelten:  $rd(x) = x, \quad \forall x \in M$
- Weiters muss eine Zahl  $\xi$  im darstellbaren Bereich von  $M$  liegen; z.B. gilt für IEEE 754, Typ *Single*
  - Kleinste Zahl (betragsmäßig, normalisiert):  $x_{\min} = 2^{-126} \approx 10^{-38}$
  - Größte Zahl (normalisiert):  $x_{\max} = (2 - 2^{-23}) \cdot 2^{127} \approx 3.4 \cdot 10^{38}$
- Es ergibt sich ein Underflow, wenn  $|\xi| < x_{\min} \rightarrow rd(\xi) = 0$
- Analog ein Overflow, wenn  $|\xi| > x_{\max} \rightarrow rd(\xi) = \pm\infty$



## Rundung Gleitkommazahlen

- Eine Option zur Rundung ist das Abschneiden der Mantisse (nächste Maschinenzahl in Richtung Null)
- Vorzugsweise wird zur nächstgelegenen Maschinenzahl gerundet, so dass gilt

$$|\xi - rd(\xi)| = \min_{x \in M} |\xi - x|, \quad \forall \xi \in \mathbb{R}$$

- Problem: Vorgehen falls Zahl genau zwischen zwei Maschinenzahlen liegt
- Mögliche Optionen für diese Rundung:
  - Zur nächstgelegenen geraden Maschinenzahl (Standard)
  - Andernfalls, z.B. fix Richtung  $\pm\infty$  oder zur Null



## Allgemeine Fehlermaße

- Für eine Näherung  $x' \in \mathbb{R}$  an einen exakten Wert  $x \in \mathbb{R}$  ist:
  - Der absolute Fehler (mit gleicher Einheit):  $\Delta x = x' - x$
  - Der relative Fehler (dimensionslos), mit  $x \neq 0$ :  $\delta_x = \frac{x' - x}{x}$
- Häufig werden diese auch als Betrag betrachtet
- Des Weiteren gilt auch:
 
$$x' = x + \Delta x = x \left( 1 + \frac{\Delta x}{x} \right) = x \cdot (1 + \delta_x)$$
- Hinweis: der exakte Wert  $x$  ist oft nicht bekannt (in iterativen Verfahren z.B. frühere Werte verwenden)



## Fehlerabschätzung Rundung

- Zahl  $\xi = m \cdot 2^e \in \mathbb{R}$ ,  $0 \leq m < 1$ , eingeschlossen zwischen nächsten normalisierten Maschinenzahlen (mit  $k$ -Bit Mantisse)  $x_{li} < \xi < x_{re}$ ; für diese gilt  $x_{re} - x_{li} = 2^{e-k}$
- Absoluter Fehler bei Rundung zur nächsten Zahl  $x_{li}$ ,  $x_{re}$ :
 
$$|\Delta \xi| = |rd(\xi) - \xi| \leq \frac{1}{2} (x_{re} - x_{li}) = 2^{e-k-1}$$
- Somit relativer Fehler:
 
$$\left| \frac{\Delta \xi}{\xi} \right| \leq \frac{2^{e-k-1}}{m \cdot 2^e} \leq \frac{1}{2} 2^{-k} = \varepsilon \quad (\text{Maschinengenauigkeit})$$
- Für IEEE 754, Typ Single:  $\varepsilon = 2^{-24} \approx 5.96 \cdot 10^{-8}$
- Insbesondere gilt:  $rd(\xi) = \xi \cdot (1 + \delta_\xi) \leq \xi \cdot (1 + \varepsilon)$



## Gleitkommaarithmetik

- Für die Operationen  $\circ \in \{+, -, \cdot, /\}$  in  $\mathbb{R}$  haben wir bei Gleitkommazahlen  $x, y \in M$  häufig  $(x \circ y) \notin M$
- Aufgrund der Rundung betrachten wir eine Pseudoarithmetik mit Gleitkommaoperationen  $\bullet \in \{\oplus, \ominus, \odot, \oslash\}$ 

$$x \bullet y = rd(x \circ y) \in M \quad x, y \in M$$
- Es ergeben sich fehlerbehaftete Resultate mit  $|\delta| \leq \varepsilon$ 

$$x \bullet y = (x \circ y)(1 + \delta)$$
- Achtung: übliche Regeln der Algebra nicht mehr gültig
- Reihenfolge muss bei Verknüpfung beachtet werden



## Gleitkommaarithmetik

- Beispiel in  $M(10, 3, 2)$  mit Zahlen  $x, y$  aus  $\mathbb{R}$ 

$$x = 0.63865, \quad y = 0.63742, \quad x - y = 0.0123$$
- Pseudoarithmetik (in diesem Fall ist  $\varepsilon = 0.5 \cdot 10^{-3}$ )

$$rd(x) = 0.639, \quad |\delta_x| \approx 0.54 \cdot 10^{-3}$$

$$rd(y) = 0.637, \quad |\delta_y| \approx 0.65 \cdot 10^{-3}$$

$$\left| \frac{(rd(x) - rd(y)) - (x - y)}{(x - y)} \right| = \left| \frac{0.002 - 0.0123}{0.0123} \right| \approx 0.83$$

(großer relativer Fehler durch Auslöschung)



## Gleitkommaarithmetik

- Assoziativ- und Distributivgesetz gelten generell nicht
- Gefahr der Auslöschung bei Addition/Subtraktion
- Fehler durch Rundungen können sich in längeren Rechnungen stark verstärken
- Vermeidung der Probleme in CPUs z.B. möglich durch Verwendung zweier zusätzlicher Stellen („guard digits“)
- Gleitkommaproblematik auch bei Vergleichen beachten, aus  $f(x) = g(y)$  für  $x, y \in \mathbb{R}$  folgt nicht  $f(rd(x)) = g(rd(y))$
- Insbesondere Vorsicht bei Vergleichen, z.B.  $x == 0.0$ ; weiters zu beachten z.B.  $\text{float}(0.1) \neq \text{double}(0.1)$



## Verwandtes Beispiel

- In 1991 verfehlte eine Patriot-Rakete eine einfliegende Scud-Rakete, ein Fehler mit Todesfolge
- Messung der Zeit in interner Uhr in Zehntelsekunden
- Gespeichert in 24-bit Format; Rundungsfehler ca.  $6 \cdot 10^{-8}$
- Zeitfehler akkumulierte über 100 Betriebsstunden, insgesamt somit ca. 0.302 Sekunden
- Scud-Rakete fliegt mit ca. 1676 m/s, in der Zeitspanne also ca. 500 m



## Rechenbeispiel

- Umformung von Termen in günstigere Formate
- Algebraisch gilt

$$\sqrt{x+1} - \sqrt{x} = \frac{(\sqrt{x+1} - \sqrt{x})\sqrt{x+1} + \sqrt{x}}{\sqrt{x+1} + \sqrt{x}} = \frac{1}{\sqrt{x+1} + \sqrt{x}}$$

- Im ersten Term besteht die Gefahr der Auslöschung
- Demgegenüber erfordert der letzte Term zusätzlich eine Division, ist für  $x \gg 0$  aber vorzuziehen
- Weiteres Beispiel: Bei Summierung in Reihen mit den betragsmäßig kleinsten Summanden beginnen



## Horner-Schema

- Ein Polynom in  $\mathbb{R}$  ist gegeben als

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_2 x^2 + a_1 x + a_0$$

- Auswertung erfordert  $n-1$  Multiplikationen (iterativ) für die Potenzen, sowie  $n$  weitere für die Konstanten
- Durch Umformen erhalten wir äquivalent

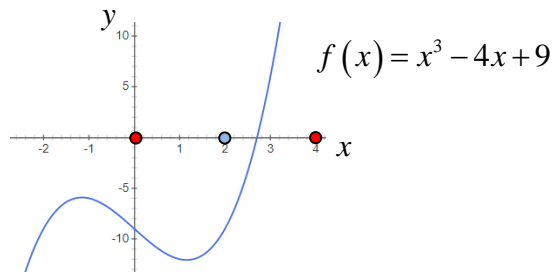
$$p(x) = \left( \dots \left( (a_n x + a_{n-1}) x + a_{n-2} \right) x + \dots + a_1 \right) x + a_0$$

- Diese Form benötigt nur noch  $n$  Multiplikationen; des Weiteren werden Rundungsfehler reduziert
- Schema kann z.B. auch angewandt werden für die Polynomdivision oder Auswertung der Ableitung



## Nullstellenfindung – Bisektionsverfahren

- Einschließung einer Nullstelle einer stetigen Funktion  $f: [a,b] \rightarrow \mathbb{R}$  über Intervallhalbierung;  $f(a) \cdot f(b) < 0$
- Gemäß Zwischenwertsatz mindestens eine Nullstelle
- Verfahren: Berechne  $m = \frac{1}{2}(a+b)$ , falls  $f(a) \cdot f(m) < 0$  setze  $b = m$ , sonst  $a = m$ ; iteriere bis  $b-a < \varepsilon$

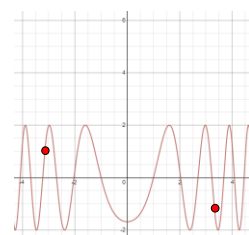
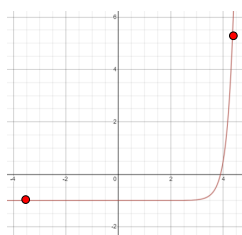
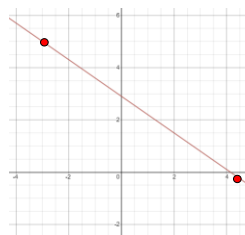
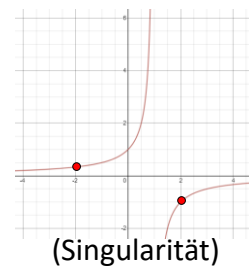
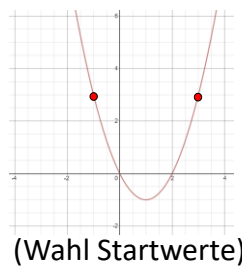
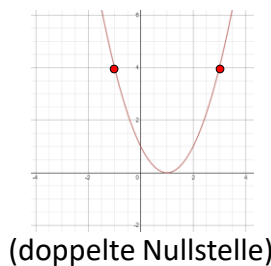


Angewandte Mathematik für die Informatik – SS2022

30



## Nullstellenfindung – Spezielle Fälle



Angewandte Mathematik für die Informatik – SS2022

31





## Nullstellenfindung – Sekantenverfahren

- Iteratives Newton-Verfahren für Nullstellensuche

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

- Alternative, z.B. falls Ableitung nicht bekannt oder aufwändig zu berechnen: Sekantenverfahren
- Verwendet Annäherung des Differentialquotienten  $f'(x_n)$  (mit gewählten Startwerten  $x_0$  und  $x_1$ ):

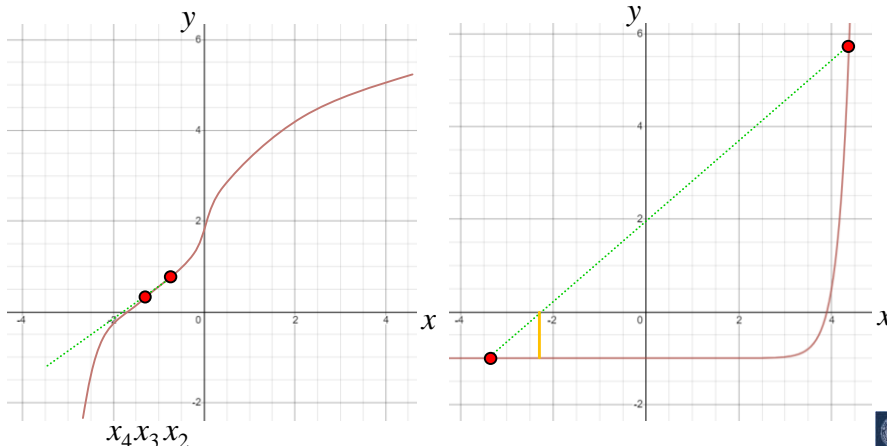
$$f'(x_n) \approx \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}} \quad f(x_n) \neq f(x_{n-1})$$

- Problem: möglicherweise Auslöschung



## Nullstellenfindung – Sekantenverfahren

- Beispiele



## Vergleichsbeispiel

- Anwendung Verfahren auf Funktion  $f(x) = x^2 - 4$ , mit Startwert(en)  $x_0 = 1$  (bzw.  $x_1 = 4$ )

Iteration	Bisektion	Sekantenverfahren	Newton-Verfahren
1	2.5	1.6	2.5
2	1.75	1.8571428571428572	2.04999999999999982
3	2.125	2.0165289256198347	2.00060975609756086
4	1.9375	1.9993904297470284	2.00000009292229474
5	2.03125	1.99997491099676	2.00000000000000222
6	1.984375	2.000000003823963	2
7	2.0078125	2	2
8	1.99609375	2	2



## Konvergenzgeschwindigkeit

- Eine Folge  $(x_n)$  konvergiere gegen Grenzwert  $x^*$
- Diese hat (mindestens) die Konvergenzordnung  $p \geq 1$ , wenn es eine Konstante  $c > 0$  gibt, so dass

$$|x_{k+1} - x^*| \leq c |x_k - x^*|^p \quad \forall k \geq 0 \quad (c < 1, \text{ falls } p = 1)$$

- Konvergenz ist linear, falls  $p = 1$ , quadratisch für  $p = 2$ , sowie kubisch falls  $p = 3$
- Je kleiner Konvergenzrate  $c$  und je größer  $p$  desto besser die Konvergenz
- Konvergenzordnung  $p$ : Anzahl korrekte Stellen „ver- $p$ -fachen“ sich je Iteration



## Konvergenzbetrachtungen

- Bisektionsverfahren: konvergiert linear, somit  $p = 1$ ; garantiert gegen eine Nullstelle einer stetigen Funktion  $f: [a, b] \rightarrow \mathbb{R}$ , falls  $f(a) \cdot f(b) < 0$
- Sekantenverfahren: konvergiert für eine zweimal stetig differenzierbare Funktion, im Falle einer einfachen Nullstelle mit  $f'(x^*) \neq 0, f''(x^*) \neq 0$ , mit der Konvergenzordnung  $p = (1 + \sqrt{5})/2$
- Newton-Verfahren: konvergiert lokal quadratisch,  $p = 2$ ; Konvergenz allgemein nicht garantiert; aber gegeben z.B. wenn  $f' > 0, f'' < 0$  in  $]a, b[$  sowie  $x_0 < x^*$  gewählt



## Finite Differenzen

- Numerische Annäherung von Ableitungen
- Herleitung:  $n$ -tes Taylorpolynom für  $x = x_0 + h$ 

$$f(x_0 + h) \approx f(x_0) + hf'(x_0) + \frac{h^2}{2} f''(x_0) + \dots + \frac{h^n}{n!} f^{(n)}(x_0)$$

$$f(x_0 + h) = f(x_0) + hf'(x_0) + O(h^2)$$

$$\Rightarrow f'(x_0) = \frac{f(x_0 + h) - f(x_0)}{h} + O(h) \quad (\text{Vorwärtsdifferenz})$$
- Äquivalent für  $x = x_0 - h$ 

$$f(x_0 - h) = f(x_0) - hf'(x_0) + O(h^2)$$

$$\Rightarrow f'(x_0) = \frac{f(x_0) - f(x_0 - h)}{h} + O(h) \quad (\text{Rückwärtsdifferenz})$$

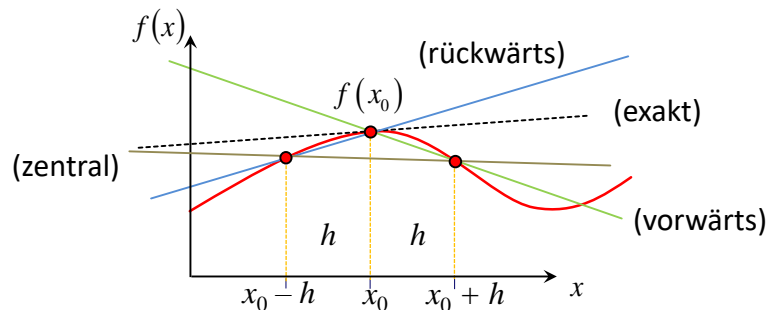


## Finite Differenzen

- Weiters durch Subtraktion der beiden Polynome

$$f(x_0 + h) - f(x_0 - h) = 2hf'(x_0) + O(h^3)$$

$$\Rightarrow f'(x_0) = \frac{f(x_0 + h) - f(x_0 - h)}{2h} + O(h^2) \quad (\text{Zentrale Differenz})$$



## Finite Differenzen

- Annäherung höherer Ableitungen, z.B. über die Addition der beiden Polynome

$$f(x_0 + h) = f(x_0) + hf'(x_0) + \frac{h^2}{2} f''(x_0) + \frac{h^3}{6} f'''(x_0) + O(h^4)$$

$$f(x_0 - h) = f(x_0) - hf'(x_0) + \frac{h^2}{2} f''(x_0) - \frac{h^3}{6} f'''(x_0) + O(h^4)$$

$$\Rightarrow f(x_0 + h) + f(x_0 - h) = 2f(x_0) + h^2 f''(x_0) + O(h^4)$$

$$\Rightarrow f''(x_0) = \frac{f(x_0 + h) - 2f(x_0) + f(x_0 - h)}{h^2} + O(h^2)$$

## Finite Differenzen – Multivariate Funktionen

- Erweiterung auf partiell differenzierbare, skalare Funktionen, z.B.  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$  mit Schrittweiten  $h_x, h_y$

$$f_x(x, y) \approx (f(x + h_x, y) - f(x, y)) / h_x$$

$$f_y(x, y) \approx (f(x, y + h_y) - f(x, y)) / h_y$$

$$f_{xx}(x, y) \approx (f(x + h_x, y) - 2f(x, y) + f(x - h_x, y)) / h_x^2$$

$$f_{yy}(x, y) \approx (f(x, y + h_y) - 2f(x, y) + f(x, y - h_y)) / h_y^2$$

$$f_{xy}(x, y) \approx (f(x + h_x, y + h_y) - f(x + h_x, y - h_y) - f(x - h_x, y + h_y) + f(x - h_x, y - h_y)) / (4h_x h_y)$$



## Numerische Lösungsmethode – Beispiel

- Wärmeleitungsgleichung in 1D (Position  $x$ , Zeit  $t$ )

$$\frac{\partial T(x, t)}{\partial t} = \mu \cdot \frac{\partial^2 T(x, t)}{\partial x^2}$$

- Annäherung linker Term mit Vorwärtsdifferenz

$$\frac{\partial T(x, t)}{\partial t} \approx \frac{T(x, t + h_t) - T(x, t)}{h_t} \quad (\text{Euler Vorwärtsverfahren})$$

- Einsetzen und Umformen

$$T(x, t + h_t) \approx T(x, t) + h_t \mu \cdot \frac{\partial^2 T(x, t)}{\partial x^2}$$



## Numerische Lösungsmethode – Beispiel

- Annäherung rechter Term mit Finiter Differenz

$$\frac{\partial^2 T(x, t)}{\partial x^2} \approx \frac{T(x + h_x, t) - 2T(x, t) + T(x - h_x, t)}{h_x^2}$$

- Einsetzen und Umformen

$$T(x, t + h_t) \approx T(x, t) + \frac{2h_t\mu}{h_x^2} \cdot \left( \frac{T(x + h_x, t) + T(x - h_x, t)}{2} - T(x, t) \right)$$

- Iterative Lösung mit gegebenen Anfangs- und Randwerten, Parameter  $\mu$ , sowie Diskretisierungen  $h_t, h_x$



## Numerische Integration (Quadratur)

- Approximation der Werte bestimmter Integrale (hier Fokus auf stetigen Funktionen  $f: [a, b] \rightarrow \mathbb{R}$ ) mittels Quadraturformeln (mit Fehler/Restglied  $\varepsilon$ )

$$I = \int_a^b f(x) dx = Q_n[f] + \varepsilon$$

- Erforderlich z.B. für nicht elementar integrierbare Funktionen oder falls nur einige Werte (evtl. aus Messungen) bekannt sind
- Eine Quadraturformel hat mindestens Ordnung  $m$ , wenn Polynome vom Grad  $m-1$  exakt integriert werden



## Numerische Integration (Quadratur)

- Generelle Idee: Unterteilen des Intervalls  $[a, b]$  in  $n$  Teilstücke, vorerst gleich groß mit Länge  $h = (b-a)/n$
- Dies ergibt  $n+1$  Stützstellen:  $a = x_0 < x_1 < \dots < x_n = b$
- Annäherung des Integrals (mit Gewichten  $w_i$ )

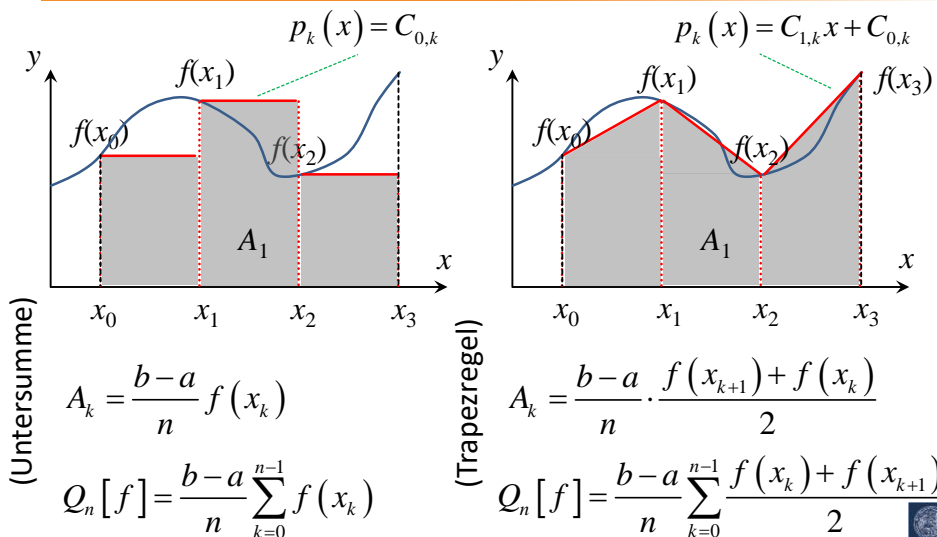
$$I \approx Q_n[f] = \sum_{i=0}^n w_i f(x_i)$$

- Option: Annäherung in äquidistanten Teilintervallen durch Polynome (geschlossene Newton-Cotes Formeln)

$$\sum_{i=0}^n w_i = \sum_{i=0}^n w_i \cdot 1 = \int_a^b 1 dx = b - a$$



## Beispiele – Quadraturverfahren



## Simpsonregel

- Annäherung durch quadratische Terme

$$q(x) = C_2 x^2 + C_1 x + C_0$$

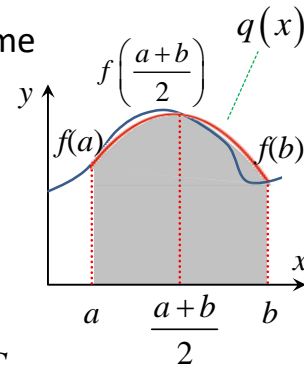
- Ausgewertet an drei Stützstellen

$$q(a) = C_2 a^2 + C_1 a + C_0 = f(a)$$

$$q(b) = C_2 b^2 + C_1 b + C_0 = f(b)$$

$$q\left(\frac{a+b}{2}\right) = C_2 \left(\frac{a+b}{2}\right)^2 + C_1 \frac{a+b}{2} + C_0$$

$$= \frac{1}{4} \left( C_2 (a+b)^2 + 2C_1 (a+b) + 4C_0 \right) = f\left(\frac{a+b}{2}\right)$$



## Simpsonregel

- Bestimmung der drei unbekannten Gewichte für die Quadratur über die drei Gleichungen

$$\begin{aligned} \int_a^b q(x) dx &= \int_a^b C_2 x^2 + C_1 x + C_0 dx = \frac{C_2}{3} x^3 + \frac{C_1}{2} x^2 + C_0 x \Big|_a^b \\ &= \frac{C_2}{3} (b^3 - a^3) + \frac{C_1}{2} (b^2 - a^2) + C_0 (b - a) \\ &= \frac{b-a}{6} \left( 2C_2 (b^2 + ab + a^2) + 3C_1 (b+a) + 6C_0 \right) \\ &= \frac{b-a}{6} \left( f(a) + 4f\left(\frac{a+b}{2}\right) + f(b) \right) \end{aligned}$$





## Simpsonregel

- Annäherung eines Integrals mit Simpsonregel, an  $n+1$  Stützstellen:  $a = x_0 < x_1 < \dots < x_n = b$

$$Q_n[f] = \frac{b-a}{n \cdot 6} \left( f(x_0) + 4f\left(\frac{x_0+x_1}{2}\right) + f(x_1) + \right. \\ \left. f(x_1) + 4f\left(\frac{x_1+x_2}{2}\right) + f(x_2) + \dots + \right. \\ \left. f(x_{n-1}) + 4f\left(\frac{x_{n-1}+x_n}{2}\right) + f(x_n) \right) \\ = \frac{b-a}{n \cdot 6} \left( f(x_0) + f(x_n) + 4 \sum_{k=0}^{n-1} \frac{f(x_k) + f(x_{k+1})}{2} + 2 \sum_{k=0}^{n-2} f(x_{k+1}) \right)$$



## Fehlerabschätzung

- Trapezregel: für das Restglied erhalten wir (u.a. mittels Taylor-Entwicklung um den Punkt  $x_0 = 0$ )

$$\varepsilon = I - Q_n[f] = -\frac{(b-a)}{12} h^2 f''(\xi), \quad \xi \in [a, b]$$

- Der Fehler ist  $O(h^2)$ ; die zusammengesetzte Formel hat die Ordnung 2, Polynome bis Grad 1 werden exakt integriert (für diese gilt  $f''(x) = 0$ )
- Simpsonregel: Ordnung 4, Polynome bis Grad 3 werden exakt integriert

$$\varepsilon = I - Q_n[f] = -\frac{(b-a)}{2880} h^4 f^{(4)}(\xi), \quad \xi \in [a, b]$$



## Weitere Aspekte – Numerische Quadratur

- Bei mehr als 6 Auswertungsstellen im Teilintervall weisen Newton-Cotes Formeln negative Gewichte auf
- Option: Gauss-Quadratur; erlaubt optimale Wahl (falls möglich) der Stützstellen  $x_k$  (nicht mehr äquidistant)
- Dadurch optimale Annäherung des Integrals mit den gegebenen Freiheitsgraden (bei  $n+1$  Stützstellen exakte Integration von Polynomen bis Grad  $2n+1$ )
- In höheren Dimensionen sind die vorgestellten Methoden nicht mehr brauchbar; als Alternative kann die Monte-Carlo Integration verwendet werden



## Einige Hilfreiche Weblinks

- *Numerical Recipes – The Art of Scientific Computing*, wichtiges Fachbuch zu numerischen Methoden, inkl. Implementierungen u.a. in C/C++  
<http://numerical.recipes/>
- *What every computer scientist should know about floating-point arithmetic*, D. Goldberg  
Wissenschaftliches Paper als Tutorial zur Verwendung von Gleitkommazahlen  
<https://dl.acm.org/doi/10.1145/103162.103163>



## Vorlesungsplan

Datum	Thema	Proseminar
11.03.22	Einführung, Grundlagen, Funktionen	<i>(Beginn zuvor am 8.3.)</i>
18.03.22	Differentialrechnung	
25.03.22	Integralrechnung	
01.04.22	Differentialgleichungen	
08.04.22	Weitere Funktionen	
<i>Osterferien</i>		
29.04.22	Reihen und Folgen	
06.05.22	Numerische Auswertung von Funktionen	
13.05.22	Lösung von Gleichungssystemen	
20.05.22	Interpolation	
27.05.22	Zufallszahlen	
03.06.22	Komplexe Zahlen	
10.06.22	Klausurvorbereitung	
<b>01.07.22</b>	<b>Klausur</b>	