

Vorlesung

## **Statistische Methoden der Datenanalyse**

Prof. Dr. Dr. Wolfgang Rhode

# Numerische Grundlagen

## Überblick

- Arithmetische Ausdrücke
- Darstellung von Zahlen auf dem Computer
- Operationen und Funktionen
- Rundungsfehler und Fehlerfortpflanzung
- Stabilität
- Kondition

## Arithmetische Ausdrücke

Beispiele:

$$|\vec{v}| = \sqrt{x^2 + y^2} \quad \text{Bahnradius}$$

$$E = E_0(1 + \varepsilon)^n \quad \text{Energiegewinn bei stochastischer Beschleunigung}$$

$$\hat{\sigma} = \sqrt{\frac{\sum x_i^2 - (\sum x_i)^2 / n}{n}} \quad \text{Standardabweichung}$$

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad \text{Lösung einer quadratischen Gleichung}$$

$$h = \arcsin(\sin \psi \sin \delta + \cos \psi \cos \delta \cos t) \quad \text{Höhe eines Sterns}$$

$$\theta = \frac{1}{2\pi} \int_{-x}^x e^{-\frac{t^2}{2}} dt \quad \text{Fläche unter der Normalverteilungskurve}$$

## Arithmetische Ausdrücke

- Definitionen:

Variable:  $x_1, x_2, \dots, x_n \subset \mathcal{R}$ .

Zweistellige Operationen:  $\mathcal{O} = \{+, -, *, /, **\}$ .

Elementare Funktionen:  $\mathcal{F} = \{\sin, \cos, \exp, \ln, \sqrt{}, \text{abs}, \dots\}$ .

## Arithmetische Ausdrücke

Die Menge  $\mathcal{A} = \mathcal{A}(x_1, x_2, \dots, x_n)$  der arithmetischen Ausdrücke in  $x_1, x_2, \dots, x_n$  ist definiert durch

- i)  $\mathcal{R} \subseteq \mathcal{A}$
- ii)  $x_l \in \mathcal{A}$ , für  $l = 1, 2, \dots, n$
- iii)  $g \in \mathcal{A} \Rightarrow (-g) \in \mathcal{A}$
- iv)  $g, h \in A, \cdot \in \mathcal{O} \Rightarrow (g \cdot h) \in A$
- v)  $g \in A, \phi \in \mathcal{F} \Rightarrow \phi(g) \in A$
- vi)  $A(x_1, x_2, \dots, x_n)$  ist minimal unter den Mengen A, die (i) – (v) erfüllen.

## Beispiel: Quadratische Gleichung

$$y = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \in A(a, b, c)$$

## Berechnung von Polynomen

- gegeben:

$$f(x) = a_0x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n$$

- Naive Vorgehensweise:
  - Bildung aller Potenzen  $x^k$
  - Multiplikation mit den Koeffizienten  $a_i$
  - Addition
- Horner-Schema

## Das Horner-Schema

$$f_i := a_0 x^i + a_1 x^{i-1} + \dots + a_{i-1} x + a_i \quad \text{für } i = 1, 2, \dots, n$$

$$f_n = f(x)$$

Horner – Schema : 
$$\begin{cases} f_0 = & a_0 ; \\ f_i = & f_{i-1} \cdot x + a_i, \quad i = 1, 2, \dots, n; \\ f(x) = & f_n. \end{cases}$$

- Vorteile:
  - Rekursive Definition möglich
  - Rekursive Ableitung ebenfalls möglich

## Erste Ableitung des Horner-Schemas

$$\text{Horner-Ableitung} = \begin{cases} f'(x) &= f'_n \\ f'_i &= f'_{i-1} \cdot x + f_{i-1}, \quad i = 1, 2, \dots, n \\ f'_0 &= 0 \end{cases}$$

## Beispiel: Horner-Schema

$$f(x) = 4x^2 + 2x + 3$$

$$f_0 = 4$$

$$f'_0 = 0$$

$$f_1 = 4 \cdot x + 2$$

$$f'_1 = (0 \cdot x) + 4$$

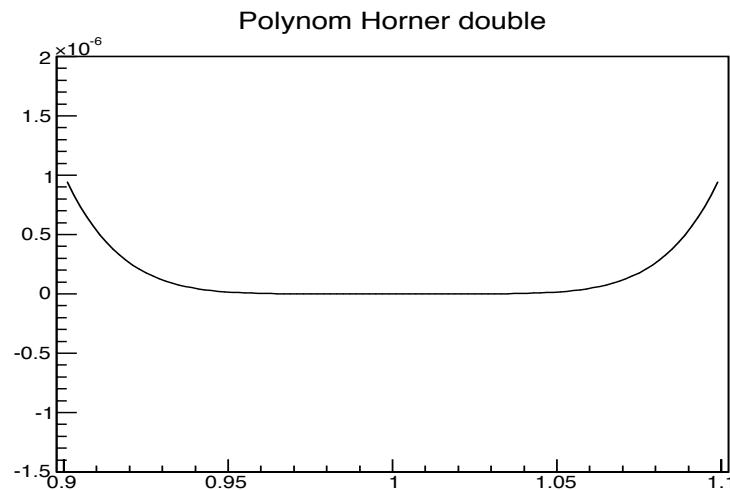
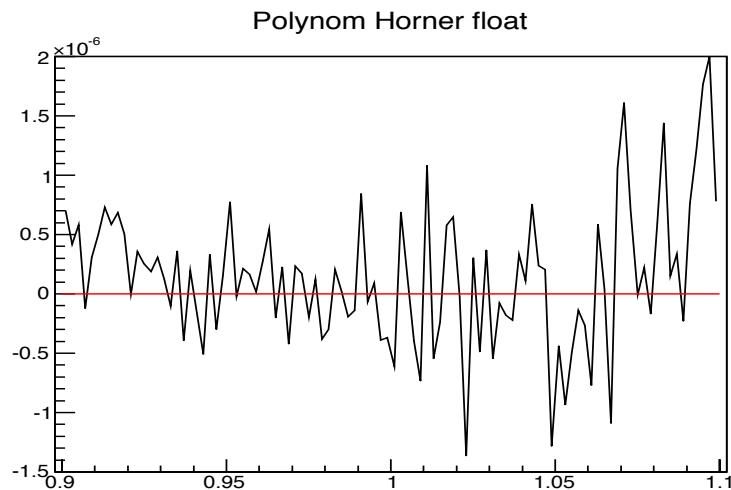
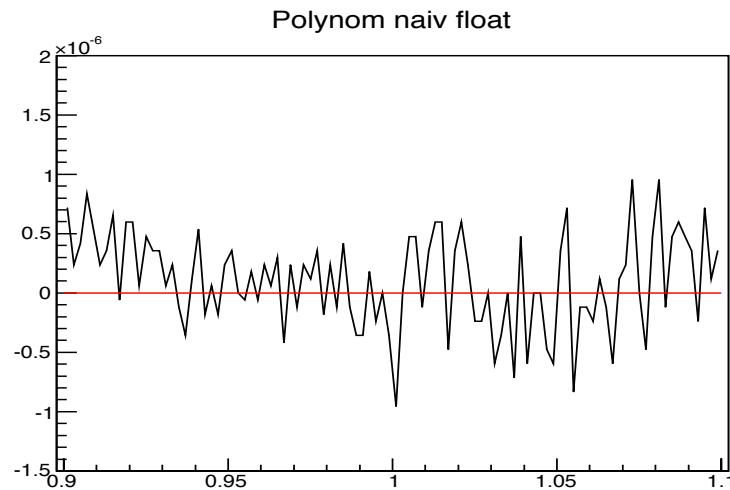
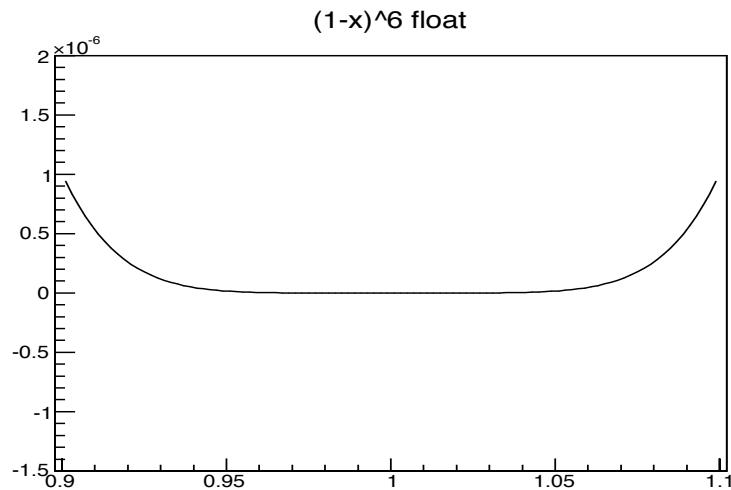
$$f_2 = (4 \cdot x + 2) \cdot x + 3$$

$$f'_2 = 4 \cdot x + (4 \cdot x + 2)$$

## Beispiel: Numerische Unterschiede

- Berechne  $(1 - x)^6$

- Einfach genau
- Doppelt genau
- Naiv
- Horner-Schema



## Darstellung von Zahlen im Computer

0	0000	-8	1000
1	0001	-7	1001
2	0010	-6	1010
3	0011	-5	1011
4	0100	-4	1100
5	0101	-3	1101
6	0110	-2	1110
7	0111	-1	1111

## Ganze Zahlen

Bits	Zahlenbereich	Bezeichnung des Zahlentyps
8	-128 ... 127	Byte
16	-32768 ... 32767	short integer
32	$-2^{31} \dots 2^{31} - 1$	integer
64	$-2^{63} \dots 2^{63} - 1$	long integer
8	0 ... 255	unsigned byte
16	0 ... 65535	word

## Exkurs: Analog-Digital-Converter (ADC)

- Elektronisch registrierte Messdaten werden ganzzahlig im Maschinenformat gespeichert
- ADC misst in einem bestimmten Spannungsintervall und hat  $n$  Speicherbits zur Verfügung
- Unterteilung des Spannungsintervalls in  $2^n$  Intervalle
- Auflösung von 8 Bit entspricht  $1/2^8 = 1/256 \approx 0.39\%$

## Gleitpunktzahlen

	Ziffern		Ziffern	
$\pm$	$\overbrace{xxxxxx}$	.	$\overbrace{xxxx}$	$10^{\pm}$
$\uparrow$		$\uparrow$		$\uparrow$
Vor-	Dezimal-		Vor-	Expo-
zeichen	punkt		zeichen	nent

## Gleitpunktzahlen

- Darstellung:

$$x = s \cdot m \cdot b^e$$

- s: Vorzeichen
- m: Mantisse
- b: Basis
- e: Exponent

Bits	Vorzeichen	Exponent	Mantisse	Zahlenbereich	Zahlentyp
32	1	8	23	$1.4 \cdot 10^{-45} \dots 3.40 \dots \cdot 10^{38}$	float
64	1	11	52	$4.9 \cdot 10^{-324} \dots 1.79 \dots \cdot 10^{308}$	double

## Überlauf und Unterlauf

- Verhalten abhängig von Programmiersprache und Compiler
  - Unterlauf: Null
  - Überlauf: NaN (Not a Number) oder Inf (Infinity)
- Führt oft zu Problemen!
- Möglichst vorher versuchen Problem numerisch zu beheben!

## Rundung

Betrachte Maschinenzahlen der Form:

$$z = 0.x_1 \dots x_L B^e$$

mit  $L$  als Mantissenlänge,  $B$  als Basis und  $e$  als Exponent.

Einer solchen Zahl  $z$  wird als Wert zugeordnet:

$$\begin{aligned} z &= \sum_{i=1}^L x_i B^{e-i} = x_1 B^{e-1} + x_2 B^{e-2} + \dots + x_L B^{e-L} \\ &= B^{e-L} (x_1 B^{L-1} + x_2 B^{L-2} + \dots + x_L) \end{aligned}$$

## Rundung

- Auswertung eines Polynoms an der Stelle  $B$  notwendig
- Konvertierung zwischen Zahlen zur Basis  $B$  und Dezimalzahlen notwendig
  - Horner-Schema!
- Zwischen reellen darstellbaren Maschinenzahlen gibt es nicht-darstellbare reelle Zahlen!
  - Suche nach einer nahegelegenen Maschinenzahl = Rundung!

## Beispiel: Rundungsfehler

- Darstellung der Zahl 0,1

V	E(8 Bit)	M(23 Bit)	
V	e <sub>1</sub> e <sub>2</sub> e <sub>3</sub> ...e <sub>8</sub>	m <sub>1</sub> m <sub>2</sub> m <sub>3</sub> .....m <sub>21</sub> m <sub>22</sub> m <sub>23</sub>	
0	0111.1011	1001.1001.1001.1001.1001.100	= 0,0999999940
0	0111.1011	1001.1001.1001.1001.1001.101	= 0,1000000015

## Rundung

- Definition: Eine Rundung heißt korrekt, wenn zwischen einer reellen Zahl  $x$  und ihrer gerundeten Zahl  $\tilde{x}$  keine Maschinenzahl liegt
- Optimale Rundung: nächstgelegene Maschinenzahl. Sind zwei Zahlen gleich weit entfernt, wird aus statistischen Gründen diejenige mit  $x_L$  gerade genommen
- Rundung durch Abschneiden: Die Ziffern nach der  $L$ -ten Stelle werden weggelassen
- Erfüllen beide die obige Definition!

## Beispiel: Rundung

Beispiel	Optimal	Abschneiden
$x$	$\tilde{x}$	$\tilde{x}$
$x_1 = .123456_{10}5$	$.123_{10}5$	$.123_{10}5$
$x_2 = .56789_{10}5$	$.568_{10}5$	$.567_{10}5$
$x_3 = .123500_{10}5$	$.124_{10}5$	$.123_{10}5$
$x_4 = .234500_{10}5$	$.234_{10}5$	$.234_{10}5$

## Rundung

- Abschätzung des relativen Fehlers:

$$\frac{|x - \tilde{x}|}{|x|} \leq \varepsilon = B^{1-L} \quad x \neq 0$$

- Rundungsfehler ist beschränkt durch maschinenabhängige Zahl  $\varepsilon$

## Numerische Fehler bei verschiedenen Operationen

Für die zweistelligen Operationen  $\circ \in \{+, -, *, /\}$  gilt bei korrekter Rundung:

$$\frac{|x \circ y - \widetilde{x \circ y}|}{|x \circ y|} \leq \varepsilon = B^{1-L} \quad x \neq 0$$

$B$  sei hier die Basis,  $L$  sei die Mantissenlänge, es liege kein Über- oder Unterlauf vor.

## Numerische Fehler bei verschiedenen Operationen

Bei Berechnung einer Potenz  $x^y$  gilt:

Ist  $y$  klein und ganzzahlig (=2, 3, ...) kann der Wert durch Ausmultiplizieren bestimmt werden. Ansonsten wird

$$x^y = \exp(y \cdot \ln(x))$$

berechnet. Der relative Fehler ist i.A. größer als bei zweistelligen Operationen

$$\frac{\Delta f}{f} = c \cdot \varepsilon \quad \text{mit} \quad c > 1.$$

## Numerische Fehler bei verschiedenen Operationen

- Zur Bestimmung von Funktionswerten anderer elementarer Funktionen werden Approximationsverfahren eingesetzt
- Schema des Approximationsverfahrens:  
 $x \rightarrow$  Argumentreduktion  $\rightarrow$  Approximation  $\rightarrow$  Ergebnisanpassung  $\rightarrow f(x)$

## Numerische Fehler bei verschiedenen Operationen

- Argumentreduktion: Zurückführung der Funktionswertberechnung auf kleinen Argumentbereich mithilfe von Hilfsformeln
- Approximation: Mithilfe von verschiedenen Verfahren
  - Kettenbruchdarstellung
  - Polynomapproximation
  - Potenzreihenentwicklung
  - Iterationsverfahren
- Ergebnisanpassung: Rückgängig machen der Argumentreduktion

## Beispiel: Approximationsverfahren

- Betrachtung der Wurzelfunktion:

$$\sqrt{x} = \text{sqrt}(x) \text{ für } x = mB^e$$

mit der Basis  $B$ , dem Exponenten  $e$  und der Mantisse  $m \in [1/B, 1]$

- Argumentreduktion und Ergebnisanpassung:

$$\sqrt{x} = \sqrt{x_0} \cdot B^S \text{ mit } \begin{cases} x_0 = m, & S = e/2, \\ x_0 = \frac{m}{B}, & S = (e + 1)/2, \end{cases} \begin{array}{ll} \text{für } e \text{ gerade} \\ \text{für } e \text{ ungerade} \end{array}$$

Dabei ist

$$x_0 \in [1/B^2, 1]$$

## Beispiel: Approximationsverfahren

- Kettenbruchdarstellung mit optimalen Koeffizienten für Intervall [0.01, 1]
- Ansatz:

$$w^*(x) = t_2 x + t_1 + \frac{t_0}{x + s_0}$$

- Bestimmung der Koeffizienten, so dass

$$\sup_{x \in [0.01, 1]} |\sqrt{x} - w^*(x)| \stackrel{!}{=} \min.$$

- Ergebnis:  
 $t_2 = 0.5881229, t_1 = 0.467975327625$   
 $t_0 = -0.0409162391674, s_0 = 0.099998$
- Relativer Fehler:  $\forall x_0 \in [0.01, 1] : w^* < 0.02$

## Beispiel: Approximationsverfahren

- Potenzreihenentwicklung, z.B. Potenzreihe um 1 von folgender Funktion:

$$\sqrt{1-z} = 1 - \frac{1}{2}z - \frac{1}{8}z^2 - \frac{1}{16}z^3 - \frac{5}{128}z^4 - \dots$$

- Konvergenzradius von 1
- Konvergiert nur für  $x \approx 1$  ( $z \approx 0$ ) genügend schnell
- Sehr langsam für kleine  $x$ 
  - Für praktische Zwecke untauglich

## Beispiel: Approximationsverfahren

- Iterationsverfahren:

Anfangsnäherung  $w_0 > 0$

$$\overline{w_i} = x/w_i$$

$$w_{i+1} = (\overline{w_i} + w_i)/2$$

- Abbruch bei Erreichen gewünschter Genauigkeit:

$$|w_0 + \overline{w_0}| < 10^{-x}$$

## Beispiel: Approximationsverfahren

- Iterationsverfahren auf Taschenrechner mit  $B=10$  und  $L=12$
- Beispiel:  $x=0.01$

$i$	0	1	2	3	4	...	7
$w_i$	1	0.505...	0.2624...	0.1502...	0.1084	...	0.1

- Quadratische Konvergenz!

## Beispiel: Approximationsverfahren

- Vergleich relativer Fehler
  - Kettenbruchentwicklung:  $< 0.02$
  - Potenzreihenentwicklung mit 14 Rechenoperationen:  $< 10^{-5}$
  - Iterationsverfahren mit 10 Iterationen:  $< 10^{-5}$

## Numerische Fehler bei verschiedenen Operationen

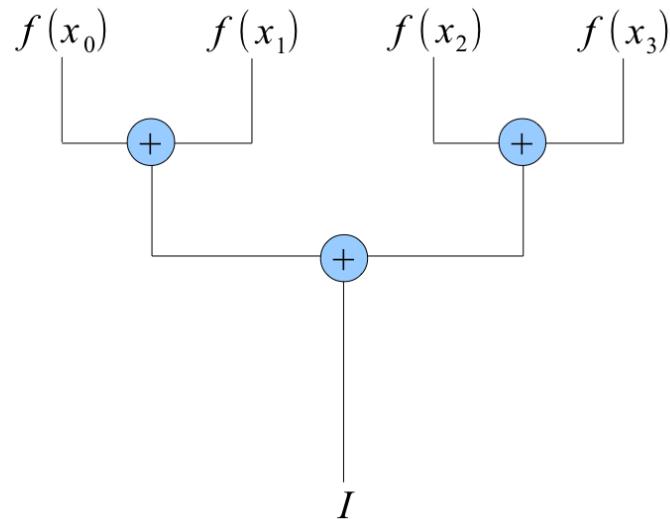
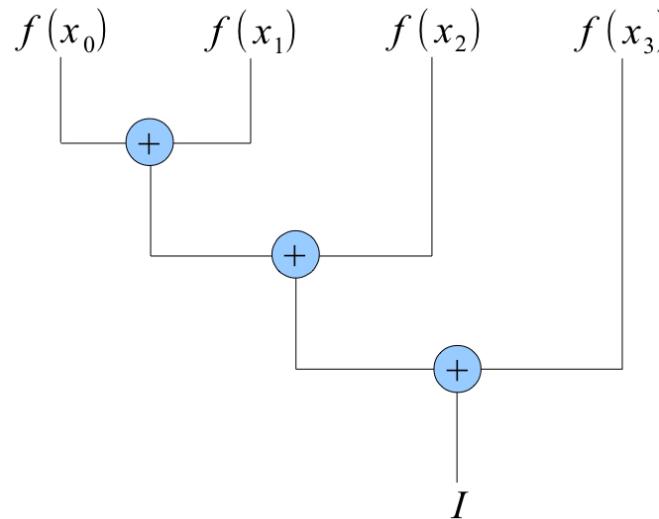
Außer in der Nähe von Nullstellen und Polen gilt für die relative Abweichung:

$$\frac{f(x) - \widetilde{f(x)}}{f(x)} \leq c_f \cdot \varepsilon$$

mit  $\varepsilon = B^{1-L}$ ,  $c_f$  hängt von der Approximation und der Argumentreduktion ab.

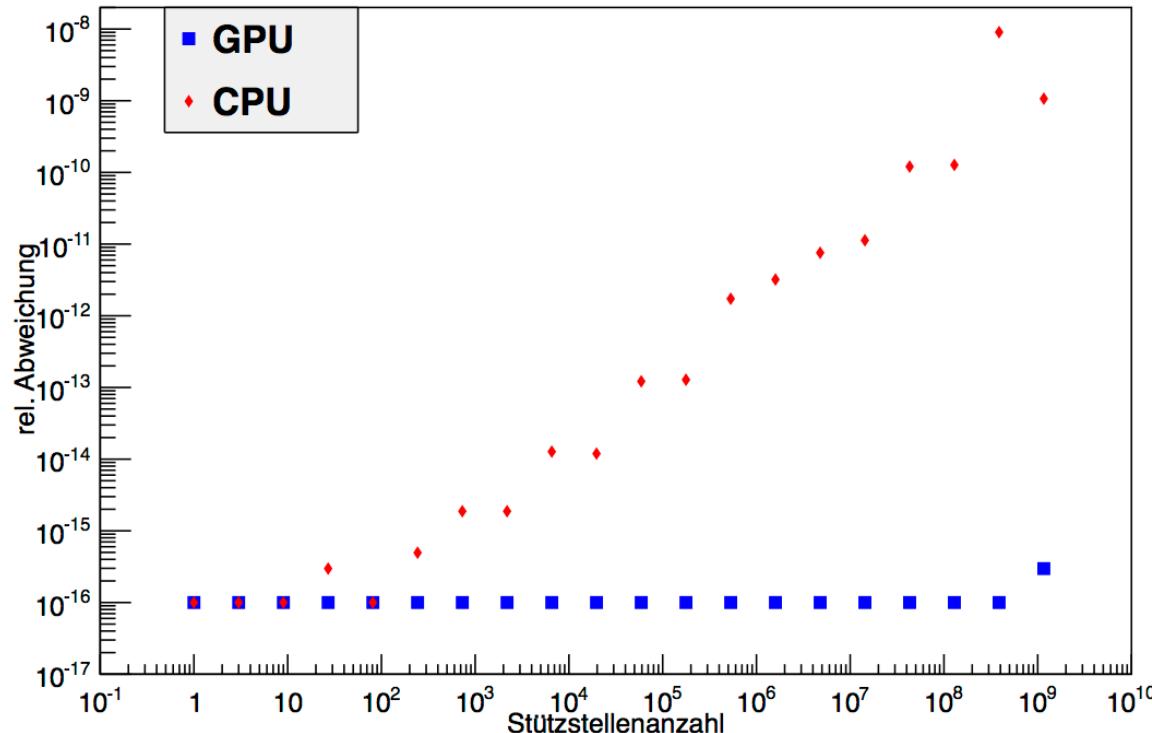
## Beispiel: Fortpflanzung numerischer Fehler

- Verschiedene Ausführungen von Summationen möglich  
(z.B. unterschiedlich für numerische Interpolation bei CPUs und GPUs)



## Beispiel: Fortpflanzung numerischer Fehler

- Reihenfolge bei Ausführung der Summation ist nicht egal...



## Beispiel: Fortpflanzung numerischer Fehler

- Warum?
  - GPU: es werden immer 2 gleich große Zahlen addiert
  - CPU: es werden kleine Zahlen auf eine immer größer werdende Zahl addiert
- Rundungsfehler wirken sich unterschiedlich stark aus!
- Rechnungen stets überprüfen auf Größenordnungen der auftretenden Zahlen und Zwischenschritte
- Sinnvoller Aufbau von Funktionen in Programmen überdenken

## Stabilität und Kondition

- **Stabilität:**

Aussage über Einfluss von Rundungsfehlern bei ungenauer Rechnung

- **Kondition:**

Fortpflanzung von Anfangsfehlern bei genauer Rechnung

## Motivation: Numerische Stabilität

- Betrachte altbekannte Funktion in unterschiedlichen Schreibweisen

a)  $f(x) = (1 - x)^6$

b)  $f(x) = 1 - 6x + 15x^2 - 20x^3 + 15x^4 - 6x^5 + x^6$

- Fall a): „eine“ Operation → stabil
- Fall b): Folge von Operationen → numerisch instabil

## Numerische Stabilität an Beispielen

- Abnahme der Genauigkeit für größer werdendes  $x$   
→ Abnahme der Genauigkeit bei Differenzbildung großer Zahlen

$$f(x) = (x^3 + \frac{1}{3}) - (x^3 - \frac{1}{3}), \Rightarrow \forall x : f(x) = \frac{2}{3}$$

$x$	$\widetilde{f(x)}$
1	0,666.666.666...
$10^3$	0,666.666.663...
$10^9$	0,663...
$10^{11}$	0

## Numerische Stabilität an Beispielen

- Abnahme der Genauigkeit für kleiner werdendes  $x$   
→ Auslöschung führender Ziffern und Verstärkung der relativen Fehler bei Summen- bzw. Differenzbildung zweier gleich großer Zahlen

$$f(x) = ((3 + \frac{x^3}{3}) - (3 - \frac{x^3}{3}))/x^3, \Rightarrow \forall x : f(x) = \frac{2}{3}$$

$x$	$\widetilde{f(x)}$
$10^{-1}$	0,666.666.667...
$10^{-3}$	0,666.67...
$10^{-6}$	0

## Numerische Stabilität an Beispielen

- Abnahme der Genauigkeit für  $x \rightarrow 0$   
→ Division durch kleine Zahl aus Subtraktion gleich großer Zahlen

$$f(x) = \frac{\sin^2(x)}{1 - \cos^2(x)}, \quad \forall x : f(x) = 1.$$

## Numerische Stabilität an Beispielen

- Instabilität in der Nähe des Pols bei  $90^\circ$

$$f(x) = \frac{\sin(x)}{\sqrt{1-\sin^2(x)}}, \quad \forall x : f(x) = \tan(x).$$

## Numerische Stabilität an Beispielen

- Abnahme der Genauigkeit für größer werdendes  $x$   
→ Vergleich mit Anfang der Reihenentwicklung

$$f(x) = e^{\frac{x^2}{3}} - 1.$$

$$(= \frac{x^2}{3} + \frac{x^4}{8} \dots)$$

## Numerische Stabilität an Beispielen

- Programmierung der Formel für die Standardabweichung einer Gaußverteilung und deren Test mit konstanten Messwerten

$$\sigma_n = \sqrt{\frac{1}{n} \left( \sum_{i=1}^n x_i^2 - \frac{1}{n} \left( \sum_{i=1}^n x_i \right)^2 \right)}$$

Für  $x_i = x = \text{const.}$  gilt  $\sigma_n = 0$ .

x	$\tilde{\sigma}_{10}$	$\tilde{\sigma}_{20}$
100/3	$1.204 \cdot 10^{-3}$	$1.131 \cdot 10^3$
1000/29	$8.062 \cdot 10^{-4}$	0 ↑ Negative Wurzel wird 0 gesetzt

## Was sollte also vermieden werden?

- Subtraktion gleich großer Zahlen
  - Auslöschung führender Ziffern
  - Verstärkung relativer Fehler
- Division durch kleine Zahl
  - Verstärkung absoluter Fehler
- Multiplikation mit großen Zahlen
  - Verstärkung absoluter Fehler

## Fehler bei Auslöschung

- Im Allgemeinen fehlerfrei
- Instabilität von Vergrößerung vorher akkumulierter Fehler

$B=10, L=7$

	$0.2789014 \cdot 10^3$	
	$-0.2788876 \cdot 10^3$	
Differenz:	<hr/> $0.0000138 \cdot 10^3$	keine Rundungsfehler
normalisiert:	$1.38 \cdot 10^{-2}$	

## Fehler bei Auslöschung

- Relativer Fehler rund 25.000 mal so groß wie Eingangsfehler

$B=10, L=6$

		<u>Rel. Fehler</u>
opt. gerundet	$0.278901 \cdot 10^3$	$< 2 \cdot 10^{-6}$
opt. gerundet	$-0.278888 \cdot 10^3$	$< 2 \cdot 10^{-6}$
Differenz:	$0.000013 \cdot 10^3$	
normalisiert:	$1.3 \cdot 10^{-2}$	$> 5 \cdot 10^{-2}$

## Division durch kleine Zahl

- Trotz Division gibt es auch stabile Ausdrücke...
  - Stabilität folgt aus Gestalt der Schranken für den relativen Fehler

$$f(x) = \begin{cases} 1, & \text{für } x = 0 \\ \frac{\sin(x)}{x}, & \text{für } x \neq 0 \end{cases} \quad \text{bei } x \rightarrow 0$$

und

$$f(x) = \begin{cases} 1, & \text{für } x = 1 \\ \frac{x-1}{\ln(x)}, & \text{für } x \neq 1 \end{cases} \quad \text{bei } x \rightarrow 1$$

## Stabilisierung

- Umformung der Differenz durch Erweitern

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

↑                              ↑

Instabil für grosse  $x$     Stabil für grosse  $x$ .

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

$$b) 1 - \cos(x) = \frac{\sin^2(x)}{1+\cos(x)} = 2 \sin\left(\frac{x}{2}\right)$$

↑                              ↑

Instabil für  $x \rightarrow 0$     Stabil für  $x \rightarrow 0$ .

## Stabilisierung von Mittelwert und Standardabweichung

- Unpraktisch: Bei direkter Berechnung müssen alle Messwerte gespeichert werden

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\sigma_n = \sqrt{t_n/n} \text{ oder } \sigma_{n-1} = \sqrt{t_n/(n-1)}$$

$$t_n = \sum_{i=1}^n (x_i - \bar{x}_n)^2$$

## Stabilisierung von Mittelwert und Standardabweichung

- Naive Umrechnung:
  - Vorteil: Berechnung ohne Speicherung
  - Nachteil: Instabilität wegen Auslöschung

$$\begin{aligned} t_n &= \sum_{i=1}^n x_i^2 - 2\bar{x}_n \sum_{i=1}^n x_i + \bar{x}_n^2 \sum_{i=1}^n 1 \\ &= \sum_{i=1}^n x_i^2 - n\bar{x}_n^2 \\ &= \sum_{i=1}^n x_i^2 - \frac{1}{n} \left( \sum_{i=1}^n x_i \right)^2 \end{aligned}$$

## Stabilisierung von Mittelwert und Standardabweichung

- Betrachte Differenzen

$\bar{x}_n - \bar{x}_{n-1}$  und  $t_n - t_{n_1}$ :

$$\bar{x}_n - \bar{x}_{n-1} = \frac{(n-1)\bar{x}_{n-1} + x_n}{n} - \bar{x}_{n-1}$$

$$= \frac{x_n - \bar{x}_{n-1}}{n} = \frac{\delta_n}{n},$$

$$\delta_n := x_n - \bar{x}_{n-1},$$

## Stabilisierung von Mittelwert und Standardabweichung

$$\begin{aligned}t_n - t_{n-1} &= \left( \sum_{i=1}^n x_i^2 - n\bar{x}_n^2 \right) - \left( \sum_{i=1}^{n-1} x_i^2 - (n-1)\bar{x}_n^2 \right) \\&= x_n^2 - n\bar{x}_n^2 + (n-1)\bar{x}_{n-1}^2 \\&= (\delta_n + \bar{x}_{n-1})^2 - n \left( \bar{x}_{n-1} + \frac{\delta_n}{n} \right)^2 + (n-1)\bar{x}_{n-1}^2 \\&= \delta_n \left( \delta_n + \frac{\delta_n}{n} \right) \\&= \delta_n [(x_n - \bar{x}_{n-1}) - (\bar{x}_n - \bar{x}_{n-1})] \\&= \delta_n (x_n - \bar{x}_n).\end{aligned}$$

## Stabilisierung von Mittelwert und Standardabweichung

- Erhalte Rekursionsformeln

$$\begin{aligned}\bar{x}_1 &= x_1, \quad t_1 = 0 \\ \delta_i &= x_i - \bar{x}_{i-1}, \quad i \geq 2 \\ \bar{x}_i &= \bar{x}_{i-1} + \frac{\delta_i}{i}, \quad i \geq 2 \\ t_i &= t_{i-1} + \delta_i(x_i - \bar{x}_i), \quad i \geq 2\end{aligned}$$

mit  $\sigma_n = \sqrt{t_n/n}$  bzw.  $\sigma_n = \sqrt{t_n/(n-1)}$ .

- Differenzen nun harmlos, da mögliche Auslöschung keine große Verstärkung des relativen Fehlers bewirken kann, da Differenz mit kleiner Zahl multipliziert wird und dann zu größerer Zahl addiert wird

## Stabilisierung der Lösung einer quadratischen Gleichung

- Instabil für  $b^2 \gg 4ac$ , wenn Wurzel und b das gleiche Vorzeichen haben

$$ax^2 + bx + c = 0$$

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

## Stabilisierung der Lösung einer quadratischen Gleichung

- Umformung ebenfalls instabil, nun wenn Wurzel und b entgegengesetztes Vorzeichen haben

$$x_{1,2} = \frac{2c}{-b \mp \sqrt{b^2 - 4ac}}.$$

## Stabilisierung der Lösung einer quadratischen Gleichung

- Sinnvolle Kombination beider Schreibweisen

$$q := - \left( b \cdot \text{sign}(b) \sqrt{b^2 - 4ac} \right) / 2$$

$$x_1 = \frac{q}{a}, \quad x_2 = \frac{c}{q}.$$

## Motivation: Kondition

$$f(x) = \frac{1}{1-x} \quad x = 0,999 \Rightarrow f(x) = 1000$$

Fehleranalyse für  $\tilde{x} = 0,999 + \varepsilon$  mit  $\varepsilon$  klein

$$f(\tilde{x}) = \frac{1000}{1 - 1000\varepsilon} = 1000(1 + 10^3\varepsilon + 10^6\varepsilon^2 + \dots)$$

Relativer Fehler:

$$\frac{|x - \tilde{x}|}{x} < 1.1\varepsilon \text{ und } \frac{|f(x) - f(\tilde{x})|}{f(x)} = 10^3\varepsilon + \mathcal{O}(\varepsilon^2)$$

- Problem ist **schlecht konditioniert!**

## Konditionszahl $K$

$\tilde{x}$  sei Näherung von  $x$  mit relativem Fehler

$$\varepsilon = \frac{\tilde{x} - x}{x} \text{ bzw. } \tilde{x} = x(1 + \varepsilon)$$

Entwicklung von  $f(\tilde{x})$  in Taylor-Reihe:

$$\begin{aligned} f(\tilde{x}) &= f(x + \varepsilon x) = f(x) + \varepsilon x f'(x) + \mathcal{O}(\varepsilon^2) \\ f(x) &= \left(1 + x \frac{f'(x)}{f(x)} \varepsilon + \mathcal{O}(\varepsilon^2)\right) \end{aligned}$$

Relativer Fehler von  $f$ :

$$\frac{|f(x) - f(\tilde{x})|}{|f(x)|} = \left| x \frac{f'(x)}{f(x)} \right| \cdot |\varepsilon| + \mathcal{O}(\varepsilon^2) = K \cdot |\varepsilon| + \mathcal{O}(\varepsilon^2)$$

Konditionszahl:

$$K := \left| x \frac{f'(x)}{f(x)} \right|$$

## Konditionszahl

Es gilt bei Vernachlässigung der höheren Glieder:

$$\frac{|f(x) - f(\tilde{x})|}{|f(x)|} = K \frac{|x - \tilde{x}|}{x},$$

wobei man folgende Fälle für  $K$  unterscheidet:

- $K < 1$  Fehlerdämpfung;
- $K > 1$  Fehlerverstärkung;
- $K \gg 1$  Problem schlecht konditioniert.

## Eindimensionale Konditionsanalyse

- Es gelten folgende Zusammenhänge:

a) Falls an einer Stelle  $f'(x^*) \neq 0$   
die Funktion  $f(x) \rightarrow 0$  für  $x \rightarrow x^* \neq 0$  geht,  
dann strebt  $K \rightarrow \infty$  für  $x \rightarrow x^*$ .

Mit anderen Worten:

$f$  ist in der Nähe von einfachen Nullstellen  $\neq 0$   
schlecht konditioniert .

## Eindimensionale Konditionsanalyse

b) Sei  $f(x) = (x - x^*)^m g(x)$  bei  $g(x^*) \neq 0$  und  $m \neq 0$ . Dann ist für  $m > 0$  bei  $x^*$  eine Nullstelle  $m$ -ter Ordnung und für  $m < 0$  ein  $x^*$  Pol  $m$ -ter Ordnung.

Es gilt weiter:

$$f'(x) = m(x - x^*)^{m-1}g(x) + (x - x^*)^m g'(x), \text{ und wir erhalten}$$

$$K = x \frac{|f'(x)|}{|f(x)|} = |x| \cdot \left| \frac{m}{x - x^*} + \frac{g'(x)}{g(x)} \right| = |m| \cdot \left| \frac{x - x^*}{x} \right|^{-1} + \dots$$

Für  $x \rightarrow x^*$  ist

$$K = \begin{cases} \infty & \text{falls } x^* \neq 0 \\ |m| & \text{falls } x^* = 0 \end{cases}$$

## Eindimensionale Konditionsanalyse

c) Falls  $f'(x)$  einen Pol bei  $x^*$  hat, ist die Kondition bei  $x^*$  ebenfalls schlecht.

Betrachte z.B.  $f(x) = 1 + \sqrt{x - 1}$ . Diese Funktion hat die Konditionszahl

$$K = \left| \frac{x}{2} \left( 1 + \frac{1}{\sqrt{x-1}} \right) \right|,$$

und  $K \rightarrow \infty$  für  $x \rightarrow 1$ .

## Zusammenhang Stabilität und Kondition

- Korrelation von Stabilität und Kondition?
- Betrachte:

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

für  $0 < x < 1$

## Zusammenhang Stabilität und Kondition

- Stabilität:

$x \rightarrow 0$  : Auslöschung, die zur Instabilität führt

$x \rightarrow 1$  : Stabiles Verhalten des Ausdrucks

## Zusammenhang Stabilität und Kondition

- Kondition:

$$f'(x) = \frac{-1/x^2}{2\sqrt{\frac{1}{x}-1}} - \frac{-1/x^2}{2\sqrt{\frac{1}{x}+1}} = \frac{\sqrt{\frac{1}{x}-1} - \sqrt{\frac{1}{x}+1}}{2x^2\sqrt{\frac{1}{x}-1}\sqrt{\frac{1}{x}+1}}$$

$$\text{und } K = x \frac{|f'(x)|}{|f(x)|} = \frac{1}{2\sqrt{1-x^2}}$$

$x \rightarrow 0$  : Gut konditioniert, da  $K = \frac{1}{2}$

$x \rightarrow 1$  : Schlecht konditioniert, da  $K = \infty$

Erinnerung: Die physikalische Fragestellung:

$$g(y) = \int_c^d A(y, x)f(x)dx + b(y),$$

**g** = alle gemessenen Zahlen

**b** = gemessen, andere Ursache

**A** = Übersetzung: gemessene Zahlen → physikalische Bedeutung

**f** = gesuchte physikalische Funktion

i. A.

---

*Kann man schlecht konditionierte Probleme  
exakt lösen?*

*Nein !*

*Aber: Man könnte ja vielleicht zusätzliche Annahmen  
machen, die eine exakte mathematische Lösung möglich  
machen.*

*Dann löst man ein anderes Problem....*

*Was bedeutet das für die Lösung ?*