

Rundungsfehler und ihre Auswirkungen

Stefan Ritter

04.05.2018



Hochschule Karlsruhe
Technik und Wirtschaft
UNIVERSITY OF APPLIED SCIENCES

Näher dran.

Inhaltsverzeichnis

Einführung

Gleitkommazahlen

Fehlerfortpflanzung

Katastrophen

Zusammenfassung

Vertrauen Sie Ihrem Taschenrechner?

Folgende Ausdrücke sind mathematisch äquivalent:

$$\frac{1}{\sqrt{a+b}-\sqrt{a}} = \frac{\sqrt{a+b}+\sqrt{a}}{(\sqrt{a+b}-\sqrt{a}) \cdot (\sqrt{a+b}+\sqrt{a})} = \frac{\sqrt{a+b}+\sqrt{a}}{b}$$

Also:

$$\frac{1}{\sqrt{a+b}-\sqrt{a}} = \frac{\sqrt{a+b}+\sqrt{a}}{b}$$

$a = 10^5$ und $b = 10^{-4}$: werte beide Terme mit dem Taschenrechner aus

$$6329113,924 = 6324555,322$$

Bereits die 4. Stelle ist falsch bei 10-stelliger Anzeige! Welches Ergebnis ist „richtig“ ?

Auch bei Matlab ist Vorsicht geboten!

Berechnen Sie in Matlab:

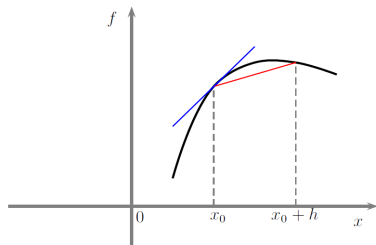
```
>>  $u = 0.3/0.1$ 
```

```
>>  $3 - u$ 
```

```
>>  $ans = 4.440892098500626e - 16$ 
```

Berechnung der Ableitung 1

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



In der Ingenieurmathematik wird gezeigt

$$\left| f'(x_0) - \frac{f(x_0 + h) - f(x_0)}{h} \right| \approx \frac{h}{2} |f''(x_0)|.$$

Diskretisierungsfehler: Abweichung des Differenzenquotienten für h von der Ableitung.

Berechnung der Ableitung 2

$$\left| f'(x_0) - \frac{f(x_0 + h) - f(x_0)}{h} \right| \approx \frac{h}{2} |f''(x_0)|$$

Betrachte $f(x) = \sin(x)$ und $x_0 = 1.2$:

$f'(x_0) = \cos(1.2) = 0.362357754476674$

Es gilt $\frac{1}{2}f''(1.2) \approx -0.466$:

h	Absoluter Fehler
$1 \cdot 10^{-1}$	$0.4716676 \cdot 10^{-1}$
$1 \cdot 10^{-2}$	$0.4666196 \cdot 10^{-2}$
$1 \cdot 10^{-3}$	$0.4660799 \cdot 10^{-3}$
$1 \cdot 10^{-4}$	$0.4660256 \cdot 10^{-4}$
$1 \cdot 10^{-7}$	$0.4619326 \cdot 10^{-7}$
$1 \cdot 10^{-8}$	$0.4361052 \cdot 10^{-8}$

h	Absoluter Fehler
$1 \cdot 10^{-9}$	$0.5594726 \cdot 10^{-7}$
$1 \cdot 10^{-10}$	$0.1669696 \cdot 10^{-6}$
$1 \cdot 10^{-11}$	$0.7938531 \cdot 10^{-5}$
$1 \cdot 10^{-13}$	$0.4250484 \cdot 10^{-3}$
$1 \cdot 10^{-15}$	$0.8173146 \cdot 10^{-1}$
$1 \cdot 10^{-16}$	$0.3623578 \cdot 10^0$

Zahlen auf dem Rechner

Die Menge der exakt darstellbaren Zahlen ist **endlich**

Ganze Zahlen

- ▶ 16 Bit signed integer:

$$\begin{aligned}x &= \pm b_{14} \cdot 2^{14} + \dots + b_1 \cdot 2^1 + b_0 \cdot 2^0, \quad b_k \in \{0, 1\} \\ &= \pm (b_{14} \dots b_1 b_0)_2\end{aligned}$$

- ▶ Wertebereich: $\pm(2^{15} - 1) = \pm 32767$, Überlauf

Reelle Zahlen

- ▶ **Festkommazahlen:** $\pm d_1 d_2 d_3, d_4 d_5 d_6$; z.B. 110,123
- ▶ 000,001, ... 999,999
- ▶ beschränkter Wertebereich, Überlauf
- ▶ **Gleitkommazahlen:** $\pm 0, d_1 d_2 d_3 d_4 d_5 d_6$ bis $\pm d_1 d_2 d_3 d_4 d_5 d_6, 0$

Gleitkommazahlen

Reelle Zahlen: b -adische Darstellung für $x \in \mathbb{R}$

$$\begin{aligned} x &= \pm (d_1 \cdot b^{-1} + d_2 \cdot b^{-2} + \dots) \times b^e, \quad d_j \in \{0, 1, \dots, b-1\} \\ &= \pm \underbrace{0, d_1 d_2 \dots}_{=f} \times b^e \end{aligned}$$

- ▶ Basis: $b = 2$, $b = 10$, Exponent $e \in \mathbb{Z}$
- ▶ Mantisse $f = 0, d_1 d_2 \dots$ mit $0 \leq f < 1$
- ▶ $d_1 \neq 0$ für $x \neq 0$: Eindeutigkeit

Maschinenzahlen: $\mathbb{M}(b, m, r, R)$

$$x = \pm 0, d_1 d_2 \dots d_m \times b^e, \quad e \in [r, R]$$

- ▶ IEEE754, double precision: $\mathbb{M}(2, 52, -1021, 1024)$, 64 bit, ca. 14 Stellen dezimal

Maschinenzahlen $\mathbb{M}(10, 3, -10, 10)$

$\mathbb{M}(10, 3, -10, 10)$: $b = 10$, $m = 3$, $r = -10$, $R = 10$

$$x = \pm 0, \underbrace{d_1 d_2 d_3}_f \times b^e$$

- ▶ endliche Länge der Mantisse f beschränkt die Genauigkeit
- ▶ Endlichkeit von e beschränkt den Bildbereich
- ▶ x_{\min} , x_{\max} kleinste (positive) und größte Maschinenzahl

$$x_{\min} = 0,100 \cdot 10^{-10}$$

$$x_{\max} = 0,999 \cdot 10^{10}$$

$$\begin{aligned}\Delta x_{\min} &= (0,101 - 0,100) \cdot 10^{-10} \\ &= 10^{-13}\end{aligned}$$

$$\begin{aligned}\Delta x_{\max} &= (0,999 - 0,998) \cdot 10^{10} \\ &= 10^7\end{aligned}$$

Verteilung Maschinenzahlen

Maschinenzahlen sind nicht gleichmäßig auf reeller Achse verteilt!

$$x = 0, \underbrace{d_1 \dots d_m}_f \times 10^e$$



$\mathbb{M}(10, 3, \dots)$ und \mathbb{R}

Rundung: $fl : \mathbb{R} \rightarrow \mathbb{M}(10, 3, \dots), \quad x \mapsto fl(x)$

$$\begin{aligned} x &= 0,123\textcolor{red}{4}\dots \cdot 10^e \\ fl(x) &= 0,123 \cdot 10^e \end{aligned}$$

$$\begin{aligned} x &= 0,123\textcolor{red}{5}\dots \cdot 10^e \\ fl(x) &= 0,12\textcolor{blue}{4} \cdot 10^e \end{aligned}$$

Rundungsfehler: $|fl(x) - x| \leq 0,0005 \cdot 10^e$

relativer Rundungsfehler:

$$\frac{|fl(x) - x|}{|x|} = \frac{0,0005 \cdot 10^e}{0,123 \dots \cdot 10^e} \leq \frac{0,0005}{0.1} = 0,005 = 5 \cdot 10^{-3}$$

Maschinengenauigkeit: $\text{eps} := 5 \cdot 10^{-m}$

Bei IEEE754 double prec.: $\text{eps} := 2^{-52} = 2.2204 \cdot 10^{-16}$

Arithmetik in $\mathbb{M}(10, 3, -10, 10)$

Rechengenauigkeit:

$$\begin{aligned} &0,111 \cdot 10^e \\ &+0,111 \cdot 10^e \\ &=0,222 \cdot 10^e \checkmark \end{aligned}$$

$$\begin{aligned} &0,111000 \cdot 10^e \\ &+0,000444 \cdot 10^e \\ &=0,111\textcolor{red}{444} \cdot 10^e \quad \textcolor{red}{\downarrow} \\ &\text{max. Fehler: } 0,0005 \cdot 10^e \end{aligned}$$

$$\text{rel. Fehler} \leq \frac{0,0005 \cdot 10^e}{0,1 \cdot 10^e} \leq 5 \cdot 10^{-3} = 5 \cdot 10^{-m} = \text{eps}$$

- ▶ $\mathbb{M}(10, 3, \dots)$ ist nicht abgeschlossen bezüglich $\{+, -, \cdot, /\}$
- ▶ Assoziativgesetze, Distributivgesetze gelten nicht

Auslöschung in $\mathbb{M}(10, 3, \dots)$

$$x = 44,544$$

$$y = 24,400$$

$$d := x - y = 20,144$$

$$\hat{x} = 44,5$$

$$\hat{y} = 24,4$$

$$\hat{d} := \hat{x} - \hat{y} = 20,1$$

$$\hat{d} = fl(d), \quad \text{rel. Fehler: } \frac{d - \hat{d}}{d} = \frac{0,044}{20,144} = 0,002 < 0,005 = \text{eps}$$

$$x = 44,544$$

$$y = 44,400$$

$$d := x - y = 0,144$$

$$\hat{x} = 44,5$$

$$\hat{y} = 44,4$$

$$\hat{d} := \hat{x} - \hat{y} = 0,100$$

$$\hat{d} \neq fl(d), \quad \text{rel. Fehler: } \frac{d - \hat{d}}{d} = \frac{0,044}{0,144} = 0,306 \gg \text{eps}$$

Quadratische Gleichung

$$x^2 - 2p \cdot x + q = 0 \quad \Rightarrow \quad x_{1,2} = p \pm \sqrt{p^2 - q},$$

$$\text{Vieta: } q = x_1 \cdot x_2 \Rightarrow x_2 = \frac{q}{x_1}$$

Setze $p = 100$, $q = 1$:

$$d := \sqrt{p^2 - q} = \sqrt{9999} = 99,99499987$$

$$x_1 = p + d = 199,994999, \quad x_2 = p - d = 0,5000126 \cdot 10^{-2}.$$

Die Rechnung in $\mathbb{M}(10, 3, \dots)$:

$$\hat{d} := \sqrt{p^2 - q} = \sqrt{\text{fl}(9999)} = \sqrt{10000} = 100$$

$$\hat{x}_1 = p + \hat{d} = 100 + 100 = 200$$

$$\hat{x}_2 = p - \hat{d} = 100 - 100 = 0, \quad \text{Auslöschung, falsch!}$$

$$\text{Vieta: } \hat{x}_2 = \text{fl}\left(\frac{q}{\hat{x}_1}\right) = \frac{1}{200} = 0,005 \quad \text{exakt!}$$

Fehlerfortpflanzung

- ▶ Zahlen, die aus Messungen stammen, sind fehlerbehaftet
- ▶ Gerundete Zahlen sind fehlerbehaftet
- ▶ Ab der ersten arithmetischen Operation wird nur noch mit Näherungen gearbeitet
- ▶ Wie pflanzen sich kleine Fehler fort?

Ein fataler Rundungsfehler

Wir betrachten die „harmlose“ Iteration

$$x_{n+1} = (x_n - 1.0) \cdot (\alpha + 1), \quad x_0 = \frac{\alpha + 1}{\alpha}, \quad \alpha = 9 \text{ oder } 16.$$

Es gilt

$$x_1 = \left(\frac{\alpha + 1}{\alpha} - \frac{\alpha}{\alpha} \right) \cdot (\alpha + 1) = \frac{\alpha + 1}{\alpha} = x_0$$

und weiter

$$x_n = \frac{\alpha + 1}{\alpha}, \quad n = 1, 2, \dots$$

Ein fataler Rundungsfehler, Fortpflanzung

Wie entwickelt sich der Startfehler in jedem Schritt?

$$x_{n+1} = (x_n - 1.0) \cdot (\alpha + 1)$$

$$\hat{x}_0 = fl(x_0) = \frac{\alpha + 1}{\alpha} + \Delta x_0$$

$$\hat{x}_1 = \left(\underbrace{\frac{\alpha + 1}{\alpha} + \Delta x_0}_{\hat{x}_0} - 1.0 \right) \cdot (\alpha + 1) = \frac{\alpha + 1}{\alpha} + \Delta x_0 \cdot (\alpha + 1)$$

$$\hat{x}_2 = \left(\underbrace{\frac{\alpha + 1}{\alpha} + \Delta x_0 \cdot (\alpha + 1)}_{\hat{x}_1} - 1.0 \right) \cdot (\alpha + 1) = \frac{\alpha + 1}{\alpha} + \Delta x_0 \cdot (\alpha + 1)^2$$

Der Rundungsfehler vergrößert sich pro Iteration um den Faktor $\alpha + 1$

Ein fataler Rundungsfehler, Rechnung

$\alpha = 9$: $x_0 = \frac{10}{9}$ nicht exakt darstellbar, $\Delta x_0 \approx 0.2 \cdot \text{eps} \approx 4.441 \cdot 10^{-17}$.

$\alpha = 16$: $x_0 = \frac{17}{16} = (1,0001)_2$

n	$x_0 = 10/9.0, x_{n+1} = (x_n - 1.0) * 10.0$	$x_0 = 17/16.0, x_{n+1} = (x_n - 1.0) * 17.0$
1	1,1111111111111100	1,0625000000000000
2	1,1111111111111100	1,0625000000000000
3	1,1111111111111200	1,0625000000000000
4	1,1111111111111600	1,0625000000000000
5	1,11111111111116000	1,0625000000000000
6	1,11111111111160500	1,0625000000000000
7	1,11111111111604500	1,0625000000000000
8	1,11111111116045400	1,0625000000000000
9	1,1111111160454400	1,0625000000000000
10	1,1111111604543600	1,0625000000000000
11	1,1111116045435700	1,0625000000000000
12	1,1111160454356700	1,0625000000000000
13	1,1111604543566500	1,0625000000000000
14	1,1116045435665000	1,0625000000000000
15	1,1160454356650000	1,0625000000000000
16	1,1604543566500100	1,0625000000000000
17	1,6045435665000700	1,0625000000000000
18	6,0454356650007000	1,0625000000000000
19	50,4543566500070000	1,0625000000000000
20	494,5435665000690000	1,0625000000000000
21	4935,4356650006900000	1,0625000000000000
22	49344,3566500069000000	1,0625000000000000
23	493433,5665000700000000	1,0625000000000000
24	4934325,6650007000000000	1,0625000000000000
25	49343246,6500070000000000	1,0625000000000000

Patriot-Rakete tötet 28 US-Soldaten

- ▶ Im zweiten Golfkrieg verfehlte am 25.2.1991 eine amerikanische Patriot-Rakete in Saudi-Arabien eine nahende irakische Scud-Rakete
- ▶ Die Scud-Rakete schlug in eine Kaserne ein, wobei 28 US-Soldaten ums Leben kamen
- ▶ Ursache war ein Rundungsfehler

Patriot-Rakete II

- ▶ Interne Uhr speichert Zeit seit Systemstart in Zehntelsekunden (24-Bit-Register)
- ▶ 0,1 ist im Binärsystem nicht exakt darstellbar:

$$0.1 = (0,000\overline{1100})_2 \approx 0,00011001100110011001100$$

Rundungsfehler: $\approx 9,5 \cdot 10^{-8}$

- ▶ Nach 100 Betriebsstunden:

$$100 \cdot 60 \cdot 60 \cdot 10 \cdot 9,5 \cdot 10^{-8} \text{ s} \approx 0,34 \text{ s.}$$

- ▶ Geschwindigkeit der Scud-Rakete: $6034 \text{ km/h} = 1676 \text{ m/s}$. In 0,34 Sekunden legte die Scud-Rakete rund 570 Meter zurück

Absturz der Ariane 5-Rakete am 4. Juni 1996

- ▶ Vierzig Sekunden nach dem Start explodierte die Rakete
- ▶ Verlust ca. 7,5 Mrd USD (Rakete, Ladung, Entwicklung)
- ▶ Vermutlich teuerster Computerfehler der Geschichte
- ▶ Ursache: Absturz des Bordcomputers 36.7 Sek. nach Start

Absturz der Ariane 5-Rakete II

(sample)

Absturz der Ariane 5-Rakete III

- ▶ Versuch der Umwandlung einer 64 Bit Gleitkommazahl in 16 Bit signed Integer: $x = \pm b_{14} \dots b_1 b_0$, $b_k \in \{0, 1\}$
- ▶ Die entsprechende Zahl war größer als $2^{15} - 1 = 32767$ und erzeugte einen Overflow
- ▶ Überlauf wurde als valides Flugdatum interpretiert
- ▶ Rakete steuerte dadurch vom Kurs ab und zerstörte sich umgehend selbst
- ▶ Software stammte von Ariane 4, die entsprechende Zahl war die horizontale Geschwindigkeit, und Ariane 5 flog schneller

Börsencrash in Vancouver in 1982

- ▶ Am Vancouver Stock Exchange wurde 1982 ein neuer Aktienindex mit einem Startwert von 1000 Punkten eingeführt
- ▶ Nach jeder Transaktion wurde Index neu berechnet
- ▶ Nach 22 Monaten stand der Index auf 524,811 Punkten, der korrekt ausgewertete Index sollte bei 1098,892 Punkten stehen

Börsencrash in Vancouver II

- ▶ Index wurde nach jeder Transaktion auf drei Nachkommastellen abgeschnitten, d.h. **immer abgerundet**
- ▶ Nach der Neuberechnung mit korrekter Rundung: Verdoppelung des Index!
- ▶ Tausende kleiner Fehler (immer in derselben Richtung) addierten sich zu einem großen Fehler auf
- ▶ Mittlerer Rundungsfehler: 0,0005 Punkte pro Transaktion, bei ca. 2000 Transaktionen pro Tag verliert der Index etwa einen Punkt pro Tag
- ▶ Mit 22 Börsentagen pro Monat verliert der Index in 22 Monaten so 484 Punkte (tatsächlich waren es 475 Punkte)

Zusammenfassung

- ▶ Mathematisch äquivalente Ausdrücke verhalten sich auf dem Rechner nicht „gleich“
- ▶ Alternativen zu IEEE754? 128-Bit Architektur, Intervallarithmetik, Softwarelösungen
- ▶ Präzision oder Schnelligkeit?
- ▶ Aktueller Trend: mehr speichern, schneller rechnen, ...
- ▶ Forderung: genauer rechnen!
- ▶ Markt für Ingenieur Anwendungen ist relativ klein
- ▶ Die Problematik der standardisierten Rechnerarithmetik bleibt den meisten Benutzern verborgen, sie wird unsichtbar durch Gewöhnung!