Einführung in Computational Engineering

Grundlagen der Modellierung und Simulation

6. Vorlesung: Zeitkontinuierliche Modellierung und Simulation

18. November 2013

Prof. Dr. Jan Peters

MOODLE TEST IN ZWEI FOLIEN

produziert vom

Bitte von der Meise…

- Vergesst nicht: Jeder Prof hat ‘ne Meise! Meine dürft Ihr füttern!
- Habt ein Herz für Meisen !!! [Zitat aus dem Meisenhaus]
- Meine Meise ist immer noch hungrig … gebt doch bitte mehr Feedback!!!

Meisenantworten

- Gleichgewichtslösung in Vorlesung 4 nicht erklärt? *Doch! Sektion 3.3.1,Folien 25-27*
- Moodle Fragezeiten besser an Fragen anpassen? *Werde mich bemühen...*
- Die Fragen nicht so lange erklären? *OK...*
- Bitte korrekte Lösungen in Moodle? *Werde mich bemühen...*
- Nachträglich korri~~e~~gieren? *OK*
- Mehr Beispiele, Geschichten und Anekdoten? *Werde mich bemühen...*
- Pause! Vorlesungszeit nicht mehr als 90min! *Könnt Ihr mich drann errinern?*
- Verschiedene Stiftfarben? *Fällt mir ein Wenig schwer...*
- Die Übungen sind schwer... *Das ist gut für Euch und erlaubt es uns „netter“ zu benoten!*
- *Fragen an Chris und Herke werden von Beiden beantwortet…*

Überblick der Vorlesungsinhalte

1. Einführung
2. Diskrete Modellierung und Simulation
3. Zeitkontinuierliche Modellierung und Simulation
4. Teilschritte einer Simulationsstudie
5. Interpretation und Validierung
6. Modulare und objektorientierte Modellierung und Simulation
7. Parameteridentifikation von Modellen

MOODLE FRAGE

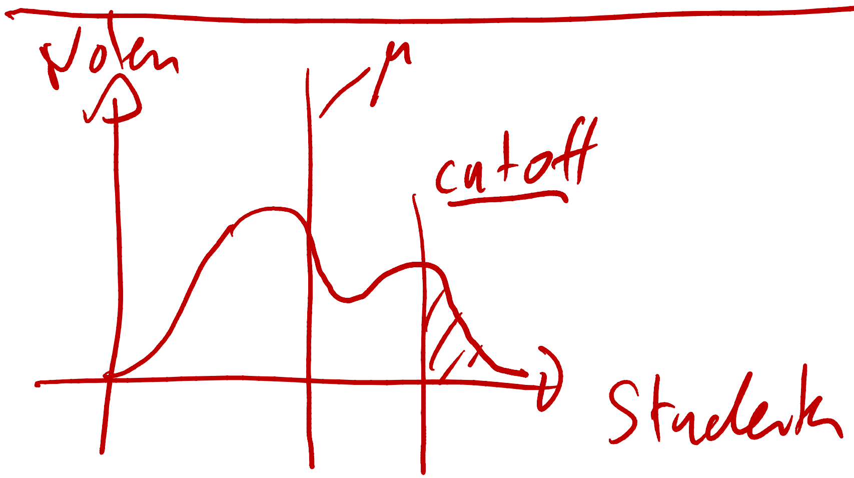
$$\left.\begin{array}{l}\dot{x}_1 = -x_2 \\ \dot{x}_2 = +x_1\end{array}\right\} \dot{\underline{x}} = \underbrace{\begin{bmatrix}0 & -1 \\ 1 & 0\end{bmatrix}}_{A} \underline{x}$$

$$\det(A-\lambda \underline{I}) = \det\begin{bmatrix}-\lambda & -1 \\ 1 & -\lambda\end{bmatrix} = \lambda^2 + 1 = 0 \Rightarrow \lambda = \pm\sqrt{-1} = \underline{\pm i}$$

Bitte jetzt auf Moodle eine Frage beantworten!

$\mathrm{Re}\,\lambda = 0 \leftarrow$

$\mathrm{Im}\,\lambda = \underline{1} \leftarrow$ Schwingt!



Grundlagen der Modellierung und Simulation

ZEITKONTINUIERLICHE MODELLIERUNG UND SIMULATION

Rundungsfehler

- 4. Juni 1996
- Ariane 5 mit 4 Satelliten

http://www.youtube.com/watch?v=gp_D8r-2hwk

Rundungsfehler

- Verlust 500 Millionen U$D
- Entwicklungskosten 7 Milliarden U$D
- Durch heftige Aktivitäten der Schubdüsen vom Kurs abgekommen
- Selbstzerstörung, weil durch aerodynamische Kräfte die Triebwerke abzubrechen drohten

Rundungsfehler

- Der eingebettete Prozessor versuchte, eine 64-Bit Gleitkommazahl (die horizontale Geschwindigkeit) in eine 16-Bit ganze Zahl zu konvertieren

- Die Gleitkommazahl war zu groß, weil Ariane 5 schneller flog als Ariane 4
  Überlauf!

- Backup-Prozessor mit exakt derselben Software stürzte ebenfalls ab

- Viele spannende Beispiele auf http://www.dradio.de/aktuell/791580/

A7

3.4.1 Zahlendarstellung

- Reelle Zahlen werden auf dem Computer als normalisierte Gleitpunktzahlen dargestellt

- Festpunktdarstellung

John von Neumann

  - Beispiel $z = 30.01109$ (Dezimalzahl)

- Normalisierte Dezimalzahl

Konrad Zuse

  - Erste Stelle vor dem Dezimalpunkt P ungleich Null
  - Alle Stellen vor P gleich Null
  - Beispiel $z = 3.001109 \cdot 10^1$

Folie 10

A7 Mit 3.4.7 weitermachen!!!

Andrej; 11.11.2013

3.4.1 Zahlendarstellung: Normalisierte Gleitpunktzahlen

- allgemeine Darstellung

$$z = \pm\overbrace{(\underbrace{d_1}_{\text{Ziffern ("digit")}} \cdot B^0 + d_2 \cdot B^{-1} + \cdots + d_t \cdot B^{-t+1})}^{\text{Mantisse}} \cdot \underbrace{B}_{\text{Basis}}^{\underbrace{E}_{\text{Exponent}}}$$

- Basis B
- Exponent E ganze Zahl und beschränkt $E_{min} \leq E \leq E_{max}$
- Ziffern $d_i \in \{0,1,2,\dots,B-1\}$
    - z normalisiert $\Longrightarrow d_1 \neq 0$ falls $z \neq 0$
- Länge der Mantisse t
- Mantisse M: $\mathrm{d}_1\mathrm{d}_2 \dots d_t$

Beispiel:

$$z = \underbrace{3.001109}_{\text{Mantisse}} \cdot \underbrace{10}_{\text{Basis}}^{\underbrace{1}_{\text{Exponent}}}$$

3.4.1 Zahlendarstellung: Normalisierte Gleitpunktzahlen

- Reelle Zahlen werden auf heutigen Computern (fast immer) als Binärzahlen dargestellt

- Bei bisher gängigen Rechnern mit 32-Bit Speicherwortlänge wird dieses häufig folgendermaßen interpretiert

255 Wak ≠ 0

1 Bit	8 Bits	23 Bits
S	Exponent E	Mantisse M

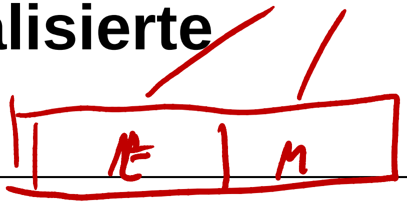
Vorzeichen

$2^E =$

$E_{min} = 0$

- Wert der Zahl: $(-1)^S \times M \times 2^E$

3.4.1 Zahlendarstellung: Normalisierte Gleitpunktzahlen



- Mehr Bits für Mantisse ergibt höhere Genauigkeit
- Mehr Bits für Exponent ergibt größeren Bereich

- IEEE 754 Gleitpunktstandard (1985)
  - Einfache Genauigkeit (single): 8 Bit Exponent, 23 Bit Mantisse 32 bit
  - Doppelte Genauigkeit (double): 11 Bit Exponent, 52 Bit Mantisse 64 bit

  - Oberstes Bit der Mantisse ist implizit, d.h. muss wegen normalisierter Mantisse nicht gespeichert werden

$1{,}\underbrace{11101101}$

↑ $\neq 0$

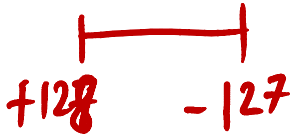
Nur nach dem Komma speichern, da erste Ziffer immer 1 bei Zahlen $\neq 0$

3.4.1 Zahlendarstellung: Normalisierte Gleitpunktzahlen

- Wert der Zahl: $(-1)^S \times (1 + M) \times 2^{E-bias}$
  - $bias = 127$ (single)
  - $bias = 1023$ (double)

bias = 0 ⇒ nur positive Zahlen

bias = 127



- Beispiel (single)
  - Dezimal: $-0.75 = \cancel{-3 \times 2^{-2}}$
  - Binär: $-0.11 = -1.1 \times \cancel{2}10^{-1}$
  - Gleitpunkt: Exponent $E = 126$
  - IEEE einfache Genauigkeit

$$1 \;\; 01111110 \; 10000000000000000000000$$

3.4.1 Zahlendarstellung: Sonderfälle

- Gleitpunktdarstellung von 0 und 1
  - 0: Alle 32 Bits sind 0 (reservierter Sonderfall!)
  - $1 = +1.0 \times 2^0 = (-1)^0 \times (1 + 0.0) \times 2^{127-127}$
  - IEEE einfache Genauigkeit

$$0\ \underbrace{01111111}_{127-127}\ 00000000000000000000000 = 1$$

$= 0$

MOODLE
TEST IN
ZWEI
FOLIEN

3.4.1 Zahlendarstellung: Alle Werte des IEEE 754

Exponent	Mantisse	dargestelltes Objekt
00...00	**00...00**	0
1–254	beliebig	± normalisierte Gleitpunktzahl
00...00	beliebig, von Null verschieden	± sub-normale Zahl (ohne Exponent)
255	**00...00**	± ∞ Unendlich (infinity)
255	beliebig, von Null verschieden	NaN (not a number)

±Inf

NaN

- Mit sub-normalen Zahlen können noch kleinere, positive Zahlen dargestellt werden, bis zu 2^{-149}
- Normalerweise wird bei Unterlauf Ausnahmebehandlung angestoßen

$\frac{1}{\text{sub-normal Zahl}} = \text{Inf}$

3.4.1 Zahlendarstellung: Alle Werte des IEEE 754

TECHNISCHE
UNIVERSITÄT
DARMSTADT

$\frac{1}{0}$

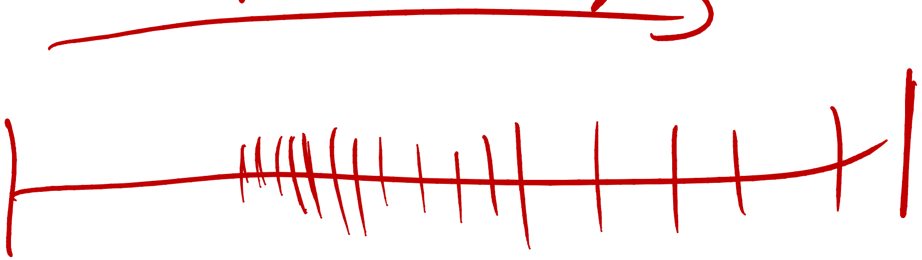
- Überlauf und Division durch Null wird zu $\pm\infty$, dadurch ist Ausnahmebehandlung möglich

- „$0/0$“ und „$\infty - \infty$“ ergeben „NaN“, dadurch ist Ausnahmebehandlung möglich

MOODLE FRAGE

Bitte jetzt auf Moodle eine Frage beantworten!

exponentiell steigende Abstände!

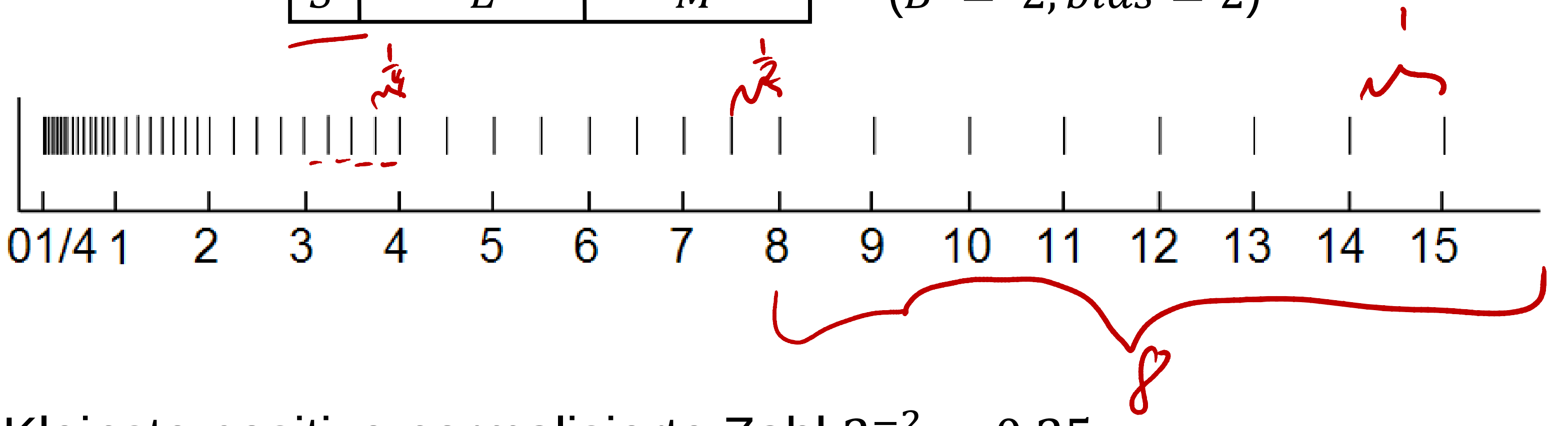


3.4.1 Zahlendarstellung: Bemerkungen

- Beispiel:

1 Bit	3 Bits	3 Bits
S	E	M

$(B = 2, bias = 2)$



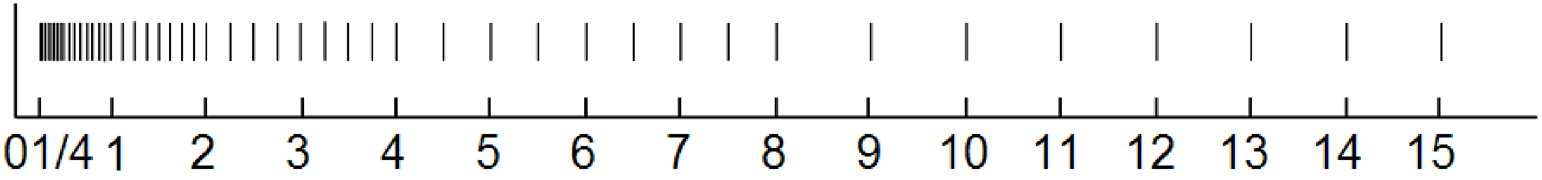
- Kleinste positive normalisierte Zahl $2^{-2} = 0.25$
- Größte positive normalisierte Zahl 15

3.4.1 Zahlendarstellung: Bemerkungen

- Beispiel:

1 Bit	3 Bits	3 Bits
S	E	M

$(B = 2, bias = 2)$



- Die Gleitpunktzahlen liegen zwischen $\underline{2^e}$ und $\underline{2^{e+1}}$
- Abstand $2^{\underline{e-3}}$ mit $e = \underline{-2}, \dots, \underline{+3}$

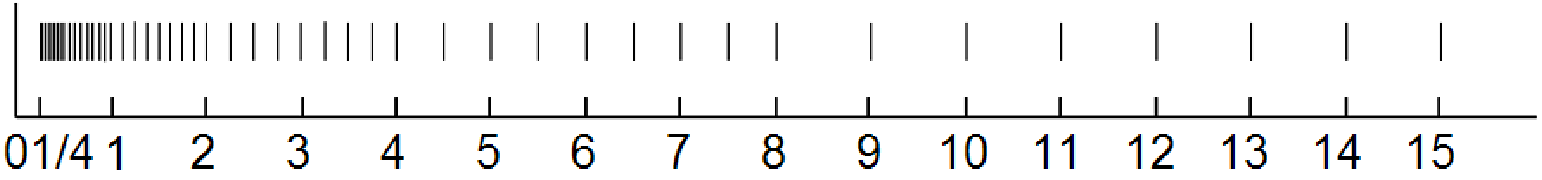
- Größter Abstand bei den größten Zahlen: $\underline{2^0 = 1}$
- Kleinster Abstand bei den kleinsten Zahlen: $\underline{2^{-5}} = \underline{0.03125}$

3.4.1 Zahlendarstellung: Bemerkungen

- Beispiel:

1 Bit	3 Bits	3 Bits
S	E	M

$(B = 2, bias = 2)$



- Abstand zwischen 1 und 2 ist 2^{-3}
- Relative Maschinengenauigkeit: $\varepsilon_{\text{mach}} = 2^{-3}$

- IEEE single precision: $\varepsilon_{\text{mach}} = 2^{-23} \approx 1.19 \times 10^{-7}$
- IEEE double precision: $\varepsilon_{\text{mach}} = 2^{-52} \approx 2.2 \times 10^{-16}$

3.4.2 Zahlendarstellung: Eigenschaften

- Grundlegende Eigenschaften bei endlicher Mantisse und endlichem Exponenten
  - Es gibt nur endlich viele Gleitpunktzahlen
  - Es gibt keine beliebig großen und keine beliebig kleinen (positiven) Zahlen
  - Es gibt keine beliebig nahe benachbarten Zahlen
  - Die Gleitpunktzahlen sind ungleichmäßig verteilt
  - Summe/Differenz, Produkt/Quotient von Gleitpunktzahlen müssen i.Allg. gerundet werden

3.4.3 Rundungsfehler: Beispiel Addition

- Addition zweier Gleitkommazahlen mit vierstelliger Mantisse

$$\underline{9.999} \times \underline{\underline{10^1}} + \underline{1.611} \times \underline{\underline{10^{-1}}}$$

- 1.Schritt: Angleichen der Dezimalkomma-Position durch Rechts-Shift der Zahl mit dem kleineren Exponenten
  - Durch Verschieben verliert man bei fester Stellenzahl für die Mantisse an Genauigkeit!

$$9.999 \times 10^1 + \overset{\longrightarrow\cancel{11}}{\underline{0.016}} \times \underline{10^1}$$

3.4.3 Rundungsfehler: Beispiel Addition

- 2.Schritt: Durchführen der Addition
  - Summe kann zwei Stellen vor dem Komma enthalten

$$\begin{array}{rl} & 9.999 \times 10^{\boxed{1}} \\ + & 0.016 \times 10^{\boxed{1}} \\ \hline & \underline{10.015} \times 10^{1} \end{array}$$

3.4.3 Rundungsfehler: Beispiel Addition

- 3.Schritt: Normalisieren des Ergebnisses durch Verschieben
  - Beim Rechts-Shift Exponenten anpassen und Über-/Unterlauf für Exponenten beachten!

$$\begin{array}{rr} & 9.999 \times 10^{1} \\ + & 0.016 \times 10^{1} \\ \hline & \overrightarrow{1.001}\boxed{5} \times 10^{\boxed{2}} \end{array}$$

$$0 \longrightarrow$$

3.4.3 Rundungsfehler: Beispiel Addition

- 4. Schritt: Anpassen der Mantisse an verfügbare Stellenzahl
  - Runden!
  - Bei Rundung müssen auch aus der Darstellung geschobene Ziffern berücksichtigt werden!

$$1.002\cancel{4} \times 10^2$$

3.4.3 Rundungsfehler: Beispiel Multiplikation

- Multiplikation von zwei Dezimal-Gleitkommazahlen

$$\underline{1.110 \times 10^{\boxed{10}}} \times \underline{9.200 \times 10^{\boxed{-5}}}$$

- 1.Schritt: Berechnung des Produkt-Exponenten durch Addition der Exponenten

$$[(10 - \underline{\text{bias}}) + (-5 - \underline{\text{bias}}) + \text{bias}]$$

$$\underline{10 + (-5)} = 5$$

- Achtung!
  - Bei Exzess-Darstellung muss der bias-Wert von der Summe abgezogen werden

3.4.3 Rundungsfehler: Beispiel Multiplikation

- 2.Schritt: Multiplikation der Mantissen

$$\begin{array}{r} 1.110 \times 9.200 \\ \hline \underline{2220} \\ .9990 \\ \hline \underline{10.212000} \end{array}$$

- Zwischenergebnis also $\underline{10.212 \times 10^5}$
- Achtung!
  - Bei fester Länge der Mantissendarstellung müssen einige Stellen vom Produkt geopfert werden

3.4.3 Rundungsfehler: Beispiel Multiplikation

- 3.Schritt: Normalisierung der Zahlendarstellung

$$10.212 \times 10^5 = 1.0212 \times 10^{\boxed{6}}$$

- Achtung!
  - Bei Exponentenanpassung Über-/Unterlauf beachten! → Ausnahmebehandlung!

- 4.Schritt: Rundung

$$1.021\cancel{2} \times 10^6 = 1.021 \times 10^6$$

$$s_1 \times s_2 = \begin{array}{l} \overset{+1}{(-1)}\,\overset{+1}{(-1)} \ \}\ +1 \\ (-1)\,(+1) \ \}\ -1 \end{array}$$

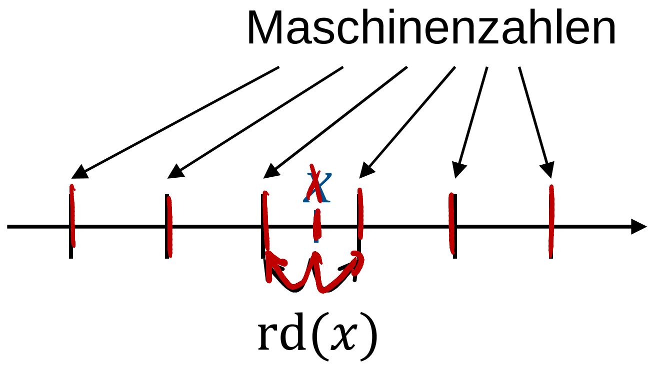
- 5.Schritt:Vorzeichen ermitteln

$$(\underline{+}1.110 \times 10^{10}) \times (\underline{+}9.200 \times 10^{-5}) = \underline{+}1.021 \times 10^6$$

3.4.3 Rundungsfehler

- Nur endliche Menge von Gleitpunktzahlen g ist verfügbar
- Jede reelle Variable x muss auf die „am nächsten liegende“ Maschinenzahl abgebildet werden
- Diese Abbildung nennt man Rundung

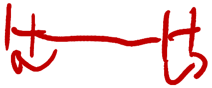
$$|x - \mathrm{rd}(x)| \leq |x - g|, \qquad \forall g$$



3.4.3 Rundungsfehler

- Nach IEEE 754 gibt es 4 Rundungsarten
  - (R1) immer aufrunden („nach rechts“) Ceil
  - (R2) immer abrunden („nach links“) floor
  - (R3) Runden durch Abschneiden („nach Null“) trunck
  - (R4) Runden zur nächsten geraden Gleitpunktzahl („zum Nächsten“) round
    - Bei Unentschieden wird Maschinenzahl mit letztem Bit=0 genommen

  - (R4) ist der Standard-Rundungsmodus, (R1,2) wichtig für Intervallarithmetik



3.4.4 Relativer Rundungsfehler

- Relativer Rundungsfehler
  - Für alle reellen x, außer bei Exponentenüber-/unterlauf

$$\varepsilon(x) \coloneqq x - \frac{\mathrm{rd}(x)}{\underline{x}} \Leftrightarrow \mathrm{rd}(x) = x\big(1 - \varepsilon(x)\big), \qquad \text{mit } \underline{|\varepsilon(x)| \leq \varepsilon_{\mathrm{mach}}}$$

- Das Ergebnis einer arithmetischen Operation ist im Allgemeinen keine Maschinenzahl

3.4.4 Relativer Rundungsfehler

- Für die Implementierung der elementaren, arithmetischen Gleitpunktoperatoren +, −, ×, / gilt in IEEE-Arithmetik
  - $\underline{\mathrm{gl}}(x+y) = (x+y)(1+\varepsilon_1)$
  - $\mathrm{gl}(x - y) = (x-y)(1+\varepsilon_2)$
  - $\mathrm{gl}(x \times y) = (x\times y)(1+\varepsilon_3)$
  - $\mathrm{gl}(x/y) = (x/y)(1+\varepsilon_4)$
  - $|\varepsilon_i| \leq \varepsilon_{\mathrm{mach}}, \quad \varepsilon_i = \varepsilon_i(x,y)$

$$gl(x \circ y) = (x \circ y)(1+\varepsilon_0)$$

$$|\varepsilon_0| \leq \varepsilon_{mach}$$

3.4.5 Fortpflanzung von Rundungsfehlern

- Ein (mathematischer) Algorithmus f sei aus einer endlichen Anzahl von Elementaroperationen zusammengesetzt

$$f = f^{(r)} \circ f^{(r-1)} \circ \dots \circ f^{(1)} \circ f^{(0)} \quad mit\ f^{(i)}: D_i \to D_{i+1}, \qquad D_i \subseteq R^n$$

← Fehlerausbreitung = Fehlerfortpflanzung

- Wie wirken sich Rundungsfehler auf das Ergebnis aus?

3.4.5 Fortpflanzung von Rundungsfehlern

- Erste Erkenntnis: Assoziativ- und Distributivgesetz gelten nicht für Gleitpunktarithmetik!
- Beispiel

$(x+y)-z \quad = x+(y-z)$

$$\mathrm{gl}(\mathrm{gl}(x+y)-z) \neq \mathrm{gl}(x+\mathrm{gl}(y-z))$$

- $\mathrm{gl}(a \circ b)$ sei die Gleitpunktimplementierung der entsprechenden arithmetischen Operation $a \circ b$

MOODLE TEST IN ZWEI FOLIEN

3.4.5 Fortpflanzung von Rundungsfehlern

- Möglichkeit zur Fehleranalyse
- In allen Ausdrücken $\mathrm{gl}(x \circ y) = (x \circ y)(1 - \varepsilon)$ sukzessive ersetzen.
- Daraus Ermittlung einer Darstellung

$$\mathrm{gl}(f) = f \cdot (1 + (\dots)\varepsilon_1 + (\dots)\varepsilon_2 + \cdots)$$

- Eingangsdaten abhängige Verstärkungsfaktoren $(\dots)$

MOODLE FRAGE

$C = 0{,}0005$

Bitte jetzt auf Moodle eine Frage beantworten!

$$524.811 + 22 \text{ Monate} \times 20 \frac{\text{Handelstage}}{\text{Monat}} \times 3000 \frac{\text{Operationen}}{\text{Handelstag}} \times 0.0005 \frac{\text{Fehler}}{\text{Operationen}}$$

$$= 660$$

$$+524.811$$

$$= 1184.811$$

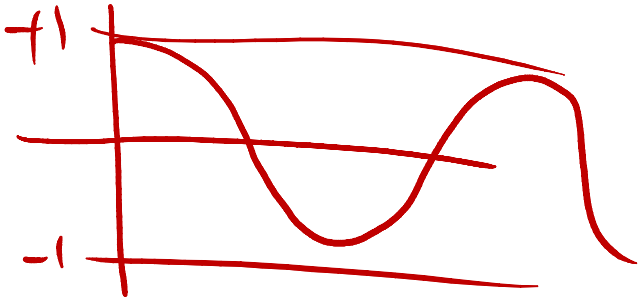
Realer Wert

1098. ...

3.4.6 Auslöschung: Beispiel des Kosinus

- Kosinus mit Taylor-Reihe berechnen

$$\cos x = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$$



- Berechnung mit einfacher Genauigkeit
  - Aufruf mit $x = 3.1415926\ (\approx \pi) \Longrightarrow -1.000000\ \mathrm{E}{+}00$
  - Aufruf mit $x = 31.4159265\ (\approx 10\pi) \Longrightarrow -8.59 \ldots \mathrm{E}{+}04$

$-8.\ldots \times 10^4$

- Ursache: Auslöschung! (cancellation)

Mehr Erläuterungen

$$\sum_{k=0} (-1)^k \underbrace{\frac{x^{2k}}{(2k)!}}_{a_k}$$

k	$(2k)!$	3.14^{2k}	$(2k)!$	31.4^{2k}
0	1.0000e+00	1.0000e+00	1.0000e+00	1.0000e+00
1	2.0000e+00	9.8596e+00	2.0000e+00	9.8596e+02
2	2.4000e+01	9.7212e+01	2.4000e+01	9.7212e+05
3	7.2000e+02	9.5847e+02	7.2000e+02	9.5847e+08
4	4.0320e+04	9.4501e+03	4.0320e+04	9.4501e+11
5	3.6288e+06	9.3174e+04	3.6288e+06	9.3174e+14
6	4.7900e+08	9.1866e+05	4.7900e+08	9.1866e+17
7	8.7178e+10	9.0576e+06	8.7178e+10	9.0576e+20
8	2.0923e+13	8.9305e+07	2.0923e+13	8.9305e+23
9	6.4024e+15	8.8051e+08	6.4024e+15	8.8051e+26
10	2.4329e+18	8.6815e+09	2.4329e+18	8.6815e+29

1

1

3,5

10^{-8}

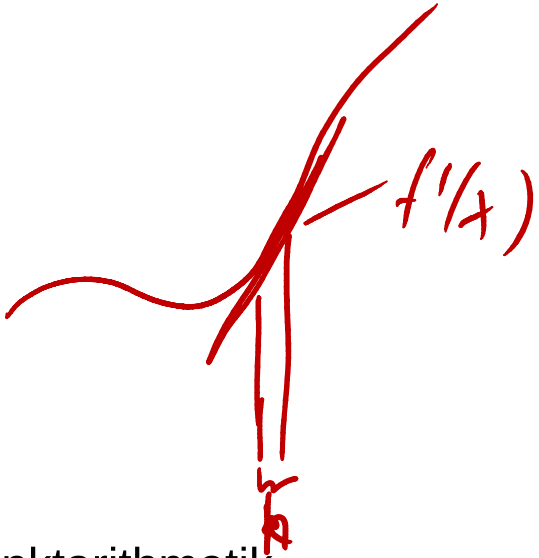
10^4

10^{11}

3.4.6 Auslöschung: Beispiel Approximation der Ableitung

- Approximation einer Funktion $f(x)$ an der Stelle x
- Vorwärtsdifferenzenquotient mit Schrittweite $h > 0$

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}$$



- Problem A
  - h ist eine so kleine Gleitpunktzahl, so dass in Gleitpunktarithmetik $x + h = x$ gilt
  - $f(x+h) = f(x)$ immer für $h <$ Maschinengenauigkeit $\varepsilon_{\text{mach}}$

$f'(x) = 0$

3.4.6 Auslöschung: Beispiel Approximation der Ableitung

- Problem B
  - $\underline{h}$ ist größer als in Problem A aber es gilt $f(x+h) \approx f(x)$
  - $f(x+h) - f(x)$ hat weniger signifikante Ziffern in Gleitpunktdarstellung als $f(x)$ und $f(x+h)$

$$\begin{array}{rl} & 3.1234\underline{666} \times 10^{0} \\ - & 3.1234\underline{555} \times 10^{0} \\ \hline & 0.0000\underline{111} \times 10^{0} \\ & 1.11\underline{?????} \times 10^{-5} \text{ (Normalisierung)} \end{array}$$

  - Auslöschung!

3.4.7 Kondition

- Wie wirken sich Rundungsfehler in Eingabedaten x einer Berechnung $\underline{y = f(x)}$ auf das Ergebnis aus?

- Sei $\underline{f}: D \subseteq \mathbb{R}^n \to \mathbb{R}^n$ mindestens einmal stetig differenzierbar

- Absoluter Fehler in x: $\Delta x_j = \left|\underline{\mathrm{rd}(x_j)} - \underline{x_j}\right|$

- Relativer Fehler in x: $\underline{\varepsilon_{x_j} = \frac{\Delta x_j}{x_j}}$

3.4.7 Kondition

- Absoluter Fehler in y: $\underline{\Delta y_j} = |\underline{\mathrm{rd}(y_j)} - \underline{y_j}| = \underline{|f_i(\mathrm{rd}(x)) - f_i(y)|}$
- Taylor-Entwicklung liefert für y

$$\Delta y_i = f_i(\mathrm{rd}(x)) - f_i(x) = \sum_{j=1}^{n} \underline{\Delta x_j} \cdot \frac{\partial f_i(x)}{\underline{\partial x_j}} + \underline{O(\|\Delta x\|^2)}$$

- Relativer Fehler in y

$\frac{x_j}{x_j} = 1$

$$\varepsilon_{y_i} = \frac{\Delta y_i}{y_i} \approx \sum_{j=1}^{n} \left(\frac{x_j}{f_i(x)} \frac{\Delta x}{x_j} \cdot \frac{\partial f_i(x)}{\partial x_j} \right) = \overbrace{\sum_{j=1}^{n} \left(\frac{x_j}{f_i(x)} \cdot \frac{\partial f_{i(x)}}{\partial x_j} \right)}^{\text{Verstärkung}} \cdot \underline{\overset{\frac{\Delta x}{x_j}}{\varepsilon_{x_j}}}$$

$f_i(x)$

3.4.7 Kondition

$$\varepsilon_{y_i} \approx \sum_{j=1}^{n} \left(\underbrace{\frac{x_j}{f_i(x)} \cdot \frac{\partial f_i(x)}{\partial x_j}}_{\text{Verstärkungsfaktoren}} \right) \cdot \varepsilon_{x_j}$$

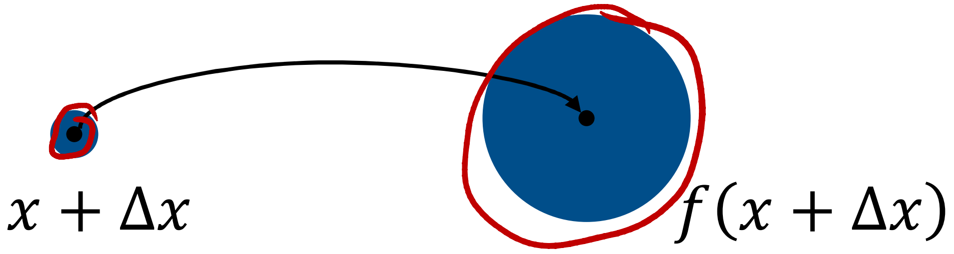
$\mathrm{cond}(A)$

- Die Beträge der Verstärkungsfaktoren des relativen Fehlers in den Eingabedaten nennt man „Konditionszahlen“

- Sind diese „groß“, ist das Problem „schlecht konditioniert“
- Sind diese „klein“, ist das Problem „gut konditioniert“

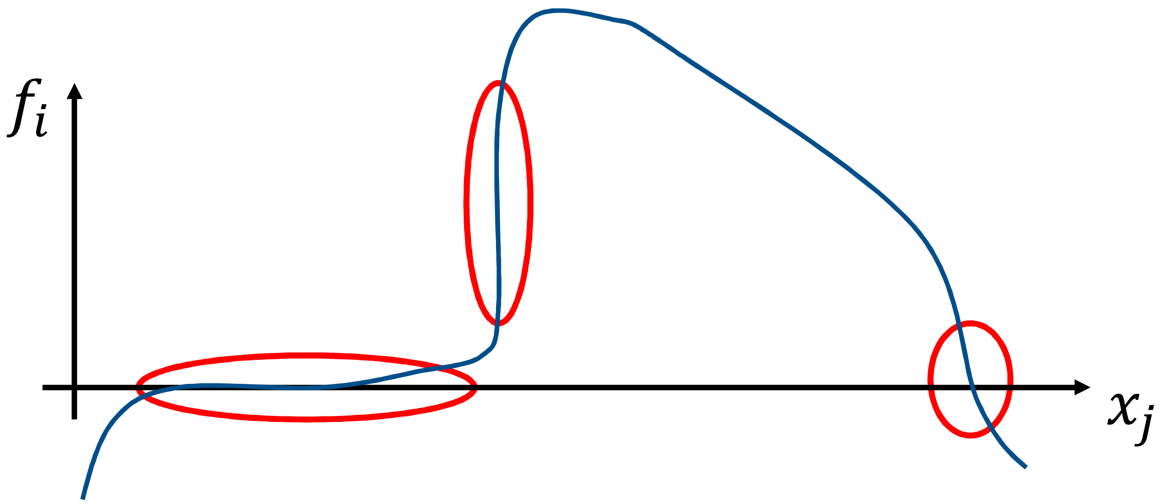
3.4.7 Kondition

- „Schlecht konditioniert“ bedeutet, dass kleine Änderungen Δx in x immer große Änderungen in $f(x)$ bewirken



3.4.7 Kondition

- Die Konditionszahlen hängen nur von f ab, d.h. nur von deren Eigenschaften
  - Hängt **nicht** davon ab wie f ausgewertet wird (Algorithmus)
  - Hängt **nicht** von der Rechnerarithmetik ab



3.4.7 Kondition: Beispiel

- Konditionszahlen $\left|\frac{x_j}{f_i(x)} \cdot \frac{\partial f_i(x)}{\partial x_j}\right|$

- Arithmetische Operationen
  - $f_1(x_1, x_2) = \underline{x_1 \cdot x_2} \Longrightarrow \underline{\varepsilon_{f_1} = \varepsilon_{x_1} + \varepsilon_{x_2}}$
  - $f_2(x_1, x_2) = \underline{x_1/x_2} \Longrightarrow \varepsilon_{f_2} = \underline{\varepsilon_{x_1} - \varepsilon_{x_2}}$
  - $f_3(x_1, x_2) = x_1 \pm x_2 \Longrightarrow e_{f_3} = \underline{\frac{x_1}{x_1 \pm x_2} \varepsilon_{x_1} \pm \frac{x_2}{x_1 \pm x_2} \varepsilon_{x_2}}$
  - $f_4(x_1, x_2) = \sqrt{x_1} \Longrightarrow \varepsilon_{f_4} = \underline{\varepsilon_{x_1}}/2$

3.4.7 Kondition: Beispiel

- Multiplikation, Division, Quadratwurzel gut konditioniert
    - relative Fehler in den Eingabedaten werden im Ergebnis nicht wesentlich verstärkt

- Addition bei Zahlen mit *gleichem* Vorzeichen gut konditioniert

3.4.7 Kondition: Beispiel

- Subtraktion bei Zahlen mit gleichen Vorzeichen schlecht konditioniert

$$\left|\frac{x_1}{\underline{x_1 \pm x_2}}\right|, \left|\frac{x_2}{x_1 \pm x_2}\right| > 1$$

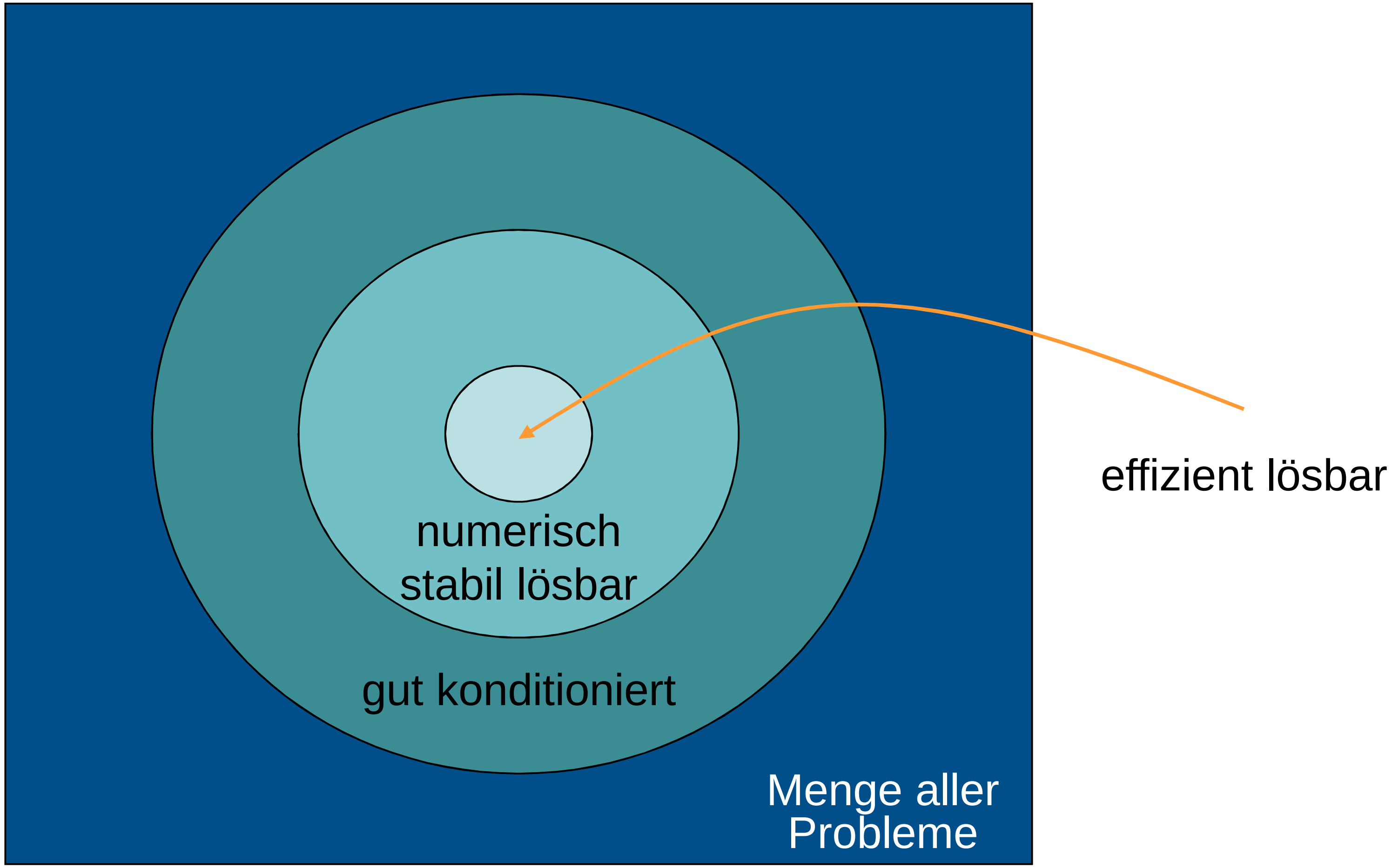
30

- Extreme Verstärkung (Auslöschung) bei $\underline{x_1 \approx x_2}$

3.4.8 Numerische Stabilität

- Gegeben sei eine gut konditionierte Aufgabenstellung $y = f(x)$
- Ein Berechnungsverfahren für f nennt man *numerisch stabil*, falls die relevanten Eingabefehler nicht verstärkt werden
- Ein Berechnungsverfahren, nennt man *numerisch instabil*, falls große relative Fehler im Ergebnis erzeugt werden

3.4.8 Numerische Stabilität



3.4.8 Numerische Stabilität: Beispiel

- Gegeben sei $\underline{f(x) = 1 - \sqrt{1 - x^2}}$
- Konditionszahl

$$\mathrm{cond}_f(x) = \left| \frac{x}{f(x)} \cdot \frac{\partial f(x)}{\partial x} \right| = \left| \frac{x}{1 - \sqrt{1 - x^2}} \cdot \frac{x}{\sqrt{1 - x^2}} \right|$$

- Für kleine Werte $\underline{x \approx 0}$ gut konditioniert, weil

$$\underline{\mathrm{cond}_f(x) \to 2} \qquad \text{mit} \quad \underline{x \to 0}$$

3.4.8 Numerische Stabilität: Beispiel

- Trotzdem erhält man für Werte $x \approx 0$ numerische Ergebnisse mit extrem großen Rundungsfehlern

- Ursache: Das Berechnungsverfahren mit direkter Implementierung der Formel ist *numerisch instabil*

3.4.8 Numerische Stabilität: Beispiel

- Auslöschung im letzten Berechnungsschritt hat große Wirkung

$$1 - \sqrt{\ldots}$$

- Mathematisch äquivalente Umformung, die numerisch stabil ist

$$1^2 + \left(\sqrt{1-x^2}\right)^2 = x^2$$

$$f(x) = \underline{1 - \sqrt{1 - x^2}} = \frac{\left(1 - \sqrt{1 - x^2}\right)\left(1 + \sqrt{1 - x^2}\right)}{1 + \sqrt{1 - x^2}} = \frac{\underline{x^2}}{1 \underline{+} \sqrt{1 - x^2}} = \frac{0^2}{1+\sqrt{1-0}}$$

3.4.8 Numerische Stabilität: 2.Beispiel

MOODLE TEST IN ZWEI FOLIEN

- Gegeben sei $\underline{f(x) = e^x}$
- Konditionszahl

$$\mathrm{cond}_f(x) = \left|\frac{x}{f(x)} \cdot \frac{\partial f(x)}{\partial x}\right| = \left|\frac{x}{\cancel{e^x}}\cancel{e^x}\right| = \underline{\underline{|x|}}$$

- Numerische Berechnung mittels Reihenentwicklung

$$\underline{e^x = 1 \overset{-}{\cancel{+}} x + \frac{x^2}{2!} \overset{-}{\cancel{+}} \frac{x^3}{3!} + \overset{-}{\cdots}}$$

- Numerisch stabil für (nicht zu große) $\underline{x > 0}$

$x = 0.001 \rightarrow x = -0{,}001$

numerisch stabil

numerisch instabil

3.4.8 Numerische Stabilität: 2.Beispiel

- Für $x < 0$ trotz gleicher Konditionszahl instabil!

- Für $x < 0$ äquivalente numerisch stabile Umformung

$$e^x = \frac{1}{e^{-x}} = \frac{1}{1 \overset{+}{-} |x| + \frac{|x|^2}{2!} \overset{+}{-} \frac{|x|^3}{3!} + \cdots}$$

- Mit Fallunterscheidung für $x > 0,\ x < 0$ numerisch stabile Berechnung

MOODLE FRAGE

$$f(x) = 5x \qquad \frac{\partial f}{\partial x} = 5$$

$$\kappa = \mathrm{cond}(f) = \left| \frac{x}{f(x)} \frac{\partial f}{\partial x} \right| = \left| \frac{\cancel{x}}{\cancel{5x}} \cancel{5} \right| = 1$$

Bitte jetzt auf Moodle eine Frage beantworten!

$$f(x) = x^2 \qquad \frac{\partial f}{\partial x} = 2x \qquad x \approx 1$$

$$\kappa_{x\approx 1} = \mathrm{cond}(f(x_{\approx})) = \left| \frac{\cancel{x}}{\cancel{x}^{\cancel{2}}} 2\cancel{x} \right| = 2$$

Moodle Lösung

$$\left.\begin{array}{l}\frac{\partial y_1}{\partial p} + \frac{\partial y_2}{\partial p} = 1 \\ \frac{\partial y_1}{\partial p} y_2 + y_1 \frac{\partial y_2}{\partial p} = 0\end{array}\right\} \Rightarrow \qquad \frac{\partial y_2}{\partial p} = \frac{y_2}{y_2 - y_1}, \quad \frac{\partial y_1}{\partial p} = \frac{y_1}{y_2 - y_1}$$

$$\left.\begin{array}{l}\frac{\partial y_1}{\partial q} + \frac{\partial y_2}{\partial q} = 0 \\ \frac{\partial y_1}{\partial q} y_2 + y_1 \frac{\partial y_2}{\partial q} = 1\end{array}\right\} \Rightarrow \qquad \frac{\partial y_1}{\partial q} = \frac{1}{y_1 - y_2} = -\frac{\partial y_2}{\partial q}$$

$$k_{11} = \frac{\partial y_1}{\partial p}\frac{p}{y_1} = \frac{y_1}{y_2 - y_1}\frac{p}{y_1} = \frac{y_1 + y_2}{y_1 - y_2} = \frac{1 + y_2/y_1}{1 - y_2/y_1}$$

$$k_{12} = \frac{\partial y_1}{\partial q}\frac{q}{y_1} = \frac{1}{y_1 - y_2}\frac{q}{y_1} = \frac{y_2}{y_1 - y_2} = \frac{1}{1 - y_2/y_1}$$

Moodle Lösung

$$p = 4\,, \quad q = 3.999\,, \quad y_{1,2} = 2 \pm 0.01\,,$$

$$k_{12} = \frac{1}{1 - y_1/y_2} = 99.5 \quad \Rightarrow \quad \textit{fast 100-fache Fehlerverstärkung.}$$

3.4.9 Zusammenfassung

- Numerisches Simulationsproblem kann durch mathematisch äquivalente Formulierungen gelöst werden
  - Äquivalente Umformung, die Rundungsfehler möglichst wenig verstärkt

- Vorgehen
  - Analyse der Kondition
  - Auswahl eines mathematisch äquivalenten und numerisch stabilen Berechnungsalgorithmus, der die über die Kondition verursachten Rundungsfehlerverstärkungen nicht verschlechtert

3.4.9 Zusammenfassung

$1-\sqrt{1+x}$

- Vorsicht
  - Bei Subtraktion von nahezu gleich großen Zahlen
  - Bei Berechnung mit relativ großen Zwischenwerten, wobei das Endergebnis relativ klein ist

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

- Was sollte man bei einem schlecht konditionierten Problem machen?
  - Durchführung der Berechnung in höherer Genauigkeit (d.h. double statt single, quadruple statt double)
  - Modifikation des „schlecht gestellten“ Ausgangsproblems, so dass die Konditionszahl kleiner wird („Regularisierung“)