

Vorlesung Höhere Mathematik 1

Kapitel 2: Rechnerarithmetik

9. September 2024

Zürcher Hochschule
für Angewandte Wissenschaften



Gliederung des Kapitels 2

HM 1,
Kapitel 2

Geschichte
der Zahldarstellung

Maschinenzahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinengenauigkeit

Fehlerfortpflan-
zung /
Konditionierung

1 Geschichte der Zahlendarstellung

2 Maschinenzahlen

3 Approximations- und Rundungsfehler

- Rundungsfehler und Maschinengenauigkeit
- Fehlerfortpflanzung / Konditionierung

Lernziele

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinen-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- Sie verstehen die Definition der maschinendarstellbaren Zahlen.
- Sie können die Fehler, die beim Abbilden von reellen Zahlen auf Maschinenzahlen entstehen, sowie die Maschinengenauigkeit berechnen.
- Sie können die Fortpflanzung von Fehlern bei Funktionsauswertungen abschätzen und die Konditionszahl berechnen.

Geschichte der Zahlendarstellung

HM 1,
Kapitel 2

Geschichte
der Zahldarstellung

Maschinenzahlen

Approximations- und Rundungsfehler

Rundungsfehler und Maschinengenauigkeit

Fehlerfortpflanzung / Konditionierung

- In den frühen Hochkulturen entwickelten sich unterschiedliche Konzepte zur Darstellung von Zahlen, die nach Art der Zusammenstellung und der Anordnung der Ziffern in Additionssysteme und Positionssysteme (auch Stellenwertsysteme genannt) einteilbar sind:
 - Additionssysteme ordnen jeder Ziffer eine bestimmte Zahl zu.
 - Im Gegensatz dazu ordnen Positions- oder Stellenwertsysteme jeder Ziffer aufgrund der relativen Poision zu anderen Ziffern eine Zahl zu: 25 -> 52

Geschichte der Zahlendarstellung

- Alle Zahlensysteme bauen dabei auf einer sogenannten ganzzahligen Grundzahl $B > 1$, auch Basis genannt, auf.
- Vor allem wurden die Zahlen 2, 5, 10, 12, 20 und 60 benutzt. Die wohl wichtigsten Grundzahlen sind 2 und 10. Von besonderem Interesse für die Babylonier war die Zahl 60, da sie zugleich die Zahl 30, also ungefähr die Anzahl Tage in einem Monat, als auch die Zahl 12, die Anzahl Monate in einem Jahr, als Teiler besitzt.

Geschichte der Zahlendarstellung

HM 1,
Kapitel 2

Geschichte
der Zahldarstellung

Maschinenzahlen

Approximations- und Rundungsfehler

Rundungsfehler und Maschinengenauigkeit

Fehlerfortpflanzung / Konditionierung

- Ägyptisches Additionssystem (ca. 3000 v. Chr.)



Abbildung: Symbole zum Darstellen von Zahlen bei den antiken Ägyptern: Ein Strich war ein Einer, ein umgekehrtes U ein Zehner, die Hunderter wurden durch eine Spirale, die Tausender durch die Lotusblüte mit Stil und die Zehntausender durch einen oben leicht angewinkelten Finger dargestellt. Dem Hunderttausender entsprach eine Kaulquappe mit hängendem Schwanz. Ergänzend ohne Bild hier: Die Millionen wird durch einen Genius, der die Arme zum Himmel erhebt, repräsentiert (aus [7]).

Geschichte der Zahlendarstellung

HM 1,
Kapitel 2

Geschichte
der Zahle-
ndarstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinen-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- Babylonisches Positionssystem (ca. 2000 v. Chr.)



Abbildung: Babylonische Form der Zahl

$$46821 = 13 \cdot 60^2 + 0 \cdot 60^1 + 21 \cdot 60^0 \text{ als } 13 | 0 | 21 \text{ (aus [7])}.$$

Geschichte der Zahlendarstellung

- Sollen die Ziffern unabhängig von der Position verwendet werden, kommen wir also um die Darstellung der Ziffer Null nicht herum.
- Wir sind dann z.B. in der Lage, die beiden Zahlen $701 = 7 \cdot 10^2 + 0 \cdot 10^1 + 1 \cdot 10^0$ und $71 = 7 \cdot 10^1 + 1 \cdot 10^0$ zu unterscheiden.
- Die Ziffer 0 deutet das Auslassen einer “Stufenzahl” B^i an und ermöglicht eine übersichtlichere Darstellung in der modernen Nomenklatur

$$z = \sum_{i=0}^n z_i \cdot B^i.$$

Geschichte der Zahlendarstellung

- Indisch-Arabisches Zahlensystem (ca. 3. Jhr.v.Chr. bis 5. Jhr. n.Chr.)



Abbildung: Arabische und indische Symbole zum Darstellen von Zahlen:
In der ersten Zeile sehen wir die indischen Ziffern des 2. Jahrhunderts n.Chr. Diese bildhaften Ziffern wurden erst von den Arabern übernommen (zweite Zeile) und später von den Europäern (dritte bis sechste Zeile: 12., 14., 15. und 16. Jhr.) immer abstrakter dargestellt (aus [7]).

Geschichte der Zahlendarstellung

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinen-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- In Europa galt die Ziffer Null lange als Teufelswerk und findet sich erstmalig in einer Handschrift von 976.
- Bis ins Mittelalter wurden in Europa Zahlen in lateinischen Grossbuchstaben geschrieben.
- Im römischen Zahlensystem standen I, V, X, L, C, D und M für 1, 5, 10, 50, 100, 500 und 1000.
- Beispiel: MMMDCCCLXXVI = 3876.
- Zum Rechnen war dieses Zahlensystem allerdings kaum geeignet und wurde im Laufe des 13. Jahrhunderts von den arabischen Ziffern abgelöst.

Geschichte der Zahlendarstellung

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinen-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- Das neue Zahlensystem mit der Null ermöglichte auch das Automatisieren von Rechenschritten.
- Im Jahre 1671 entwickelte Gottfried Wilhelm Leibniz seine Rechenmaschine, die REPLICA, die bereits alle vier Grundrechenarten beherrschte.
- Kurz darauf beschrieb er das binäre (oder auch duale) Zahlensystem, ohne das die heutige elektronische Datenverarbeitung kaum vorstellbar wäre.
- Auf dieses und weitere Zahlensysteme und die Implikationen auf arithmetische Operationen wollen wir im weiteren näher eingehen.

- Die Menge der reellen Zahlen \mathbb{R} hat unendlich viele Elemente.
- Jede Rechenmaschine ist aber ein endlicher Automat, d.h. er kann aufgrund der beschränkten Stellenzahl nicht alle Zahlen exakt darstellen und nur endlich viele Operationen ausführen.
- Für eine gegebene Basis $B \in \mathbb{N}$ kann jede reelle Zahl $x \in \mathbb{R}$ aber als

$$x = m \cdot B^e$$

dargestellt werden, wobei $m \in \mathbb{R}$ die **Mantisse** und $e \in \mathbb{Z}$ der **Exponent** genannt wird.

- Computerintern wird üblicherweise die Basis $B = 2$ verwendet (als Binär- od. Dualzahlen benannt), dies als direkte Folge der Zustände 'Strom' / 'kein Strom' (bzw. 1 und 0) von mikroelektronischen Schaltungen.
- Man spricht hier von einem *Bit* ('binary digit')¹
- Weitere Basen sind $B = 8$ (Oktalz.), $B = 10$ (Dezimalz.) und $B = 16$ (Hexadez.).
- Für letztere benötigt man 16 verschiedene Zeichen und verwendet die Ziffern 0,1,...,9 sowie A,...,F (wobei $A \triangleq 10$, $B \triangleq 11$ etc., auch Kleinbuchstaben sind erlaubt).

¹Gemäss [7] wurde der Begriff *bit* das erste Mal wahrscheinlich von John Tukey (amerikanischer Mathematiker, 1915 - 2000, Träger der IEEE 'Medal of Honor') verwendet, als kürzere Alternative zu *bigit* oder *binit*. Das Wort *digit* kommt aus dem Lateinischen und bedeutet *Finger*.

- ① Überlegen Sie sich: wie viele verschiedene Möglichkeiten gibt es, mit Binärzahlen ein Byte zu füllen?
- ② Wieviele Ziffern bräuchten Sie im Hexadezimalsystem, um die gleiche Anzahl Möglichkeiten zu erhalten?
- ③ Was folgern Sie daraus bzgl. der Vorteile des Hexadezimalsystems?

Maschinenzahlen

Aufgabe 2.1: Lösungen

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinien-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinien-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

Maschinenzahlen

Aufgabe 2.1: Lösungen

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinien-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinien-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

Maschinenzahlen

Definition 2.1: Maschinenzahlen / Gleitpunktzahlen

- Im Rechner stehen nun nur endlich viele Stellen für m und e zur Verfügung, z.B. n Stellen für m und l Stellen für e .
- Wir schreiben (entsprechend der englischen Schreibweise wird das Komma durch einen Punkt ersetzt):
 - $m = \pm 0.m_1m_2\dots m_n$ und
 - $e = \pm e_1e_2\dots e_l$.
- Unter der zusätzlichen Normalisierungsbedingung $m_1 \neq 0$ (falls $x \neq 0$) ergibt sich eine eindeutige Darstellung der sogenannten **maschinendarstellbaren Zahlen** zur Basis B :

$$M = \{x \in \mathbb{R} \mid x = \pm 0.m_1m_2m_3\dots m_n \cdot B^{\pm e_1e_2\dots e_l}\} \cup \{0\}$$

Dabei gilt $m_i, e_i \in \{0, 1, \dots, B - 1\}$ für $i \neq 0$ und $B \in \mathbb{N}, B > 1$.

- Der **Wert** \hat{w} einer solchen Zahl ist definiert als

$$\hat{w} = \sum_{i=1}^n m_i \cdot B^{\hat{e}-i}$$

und ergibt gerade die (nicht normalisierte) Darstellung der Zahl im Dezimalsystem. Dabei ist \hat{e} hier ebenfalls im Dezimalsystem zu nehmen, also $\hat{e} = \sum_{i=1}^l e_i \cdot B^{l-i}$ und es gilt $\hat{e} \in \mathbb{Z}$, d.h. \hat{e} kann natürlich auch negativ sein. Weiter gibt es eine obere und untere Schranke: $\hat{e}_{\min} \leq \hat{e} \leq \hat{e}_{\max}$.

- Man spricht bei x auch von einer **n -stelligen Gleitpunktzahl zur Basis B** (engl: floating point).
- Zahlen, die nicht in dieser Menge M liegen, müssen durch Rundung in eine maschinendarstellbare Zahl umgewandelt werden.

- Der Exponent $\hat{e} \in \mathbb{Z}$ definiert, wie wir es vom Dezimalsystem kennen, die Position des Dezimalpunktes, also z.B.

$$x = 112.78350 = 112.78350 \cdot 10^0 = 1127835.0 \cdot 10^{-4} = 0.11278350 \cdot 10^3.$$

- Um Missverständnisse zu vermeiden, kann die Basis explizit als Index zu einer Mantisse in Klammern angegeben werden. Wird kein Exponent angegeben, ist das gleichbedeutend mit $\hat{e} = 0$, z.B. $(1011100.111)_2 = 1011100.111 \cdot 2^0 = 0.1011100111 \cdot 2^7 = (0.1011100111)_2 \cdot 2^7$.
- Die in Definition 2.1 gewählte Normalisierungsbedingung, kann auch durch andere ersetzt werden. Was wären weitere Möglichkeiten? Weshalb normalisiert man überhaupt?

Maschinenzahlen

Beispiele 2.1: Normalisierte Gleitpunktzahlen

① Normalisierte Gleitpunktzahlen (gemäss Definition 2.1):

- ① $x_1 = -0.2345 \cdot 10^3$ ist eine vierstellige Gleitpunktzahl im Dezimalsystem mit dem Wert

$$-\sum_{i=1}^4 m_i \cdot 10^{3-i} = -(2 \cdot 10^2 + 3 \cdot 10^1 + 4 \cdot 10^0 + 5 \cdot 10^{-1}) = -234.5 (= -0.2345 \cdot 10^3)$$

- ② $x_2 = 0.111 \cdot 2^3$ ist eine dreistellige Gleitpunktzahl im Binär-/Dualsystem mit dem Wert

$$\sum_{i=1}^3 m_i \cdot 2^{3-i} = 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0 = 7 (= 0.7 \cdot 10^1)$$

- ③ $x_3 = 0.1001 \cdot 2^{-3}$ ist eine vierstellige Gleitpunktzahl im Binär-/Dualsystem mit dem Wert

$$\sum_{i=1}^4 m_i \cdot 2^{-3-i} = 2^{-4} + 2^{-7} = 0.0703125 (= 0.703125 \cdot 10^{-1})$$

Maschinenzahlen

Beispiele 2.1: Normalisierte Gleitpunktzahlen

(d) $x_4 = 0.71537 \cdot 8^2$ ist eine fünfstellige Gleitpunktzahl im Oktalsystem mit dem Wert

$$\begin{aligned}\sum_{i=1}^5 m_i \cdot 8^{2-i} &= 7 \cdot 8^1 + 1 \cdot 8^0 + 5 \cdot 8^{-1} + 3 \cdot 8^{-2} + 7 \cdot 8^{-3} \\ &= 57.685546875 (= 0.57685546875 \cdot 10^2)\end{aligned}$$

(e) $x_5 = 0.AB3C9F \cdot 16^4$ ist eine sechsstellige Gleitpunktzahl im Hexadezimalsystem mit dem Wert

$$\begin{aligned}\sum_{i=1}^6 m_i \cdot 16^{4-i} &= 10 \cdot 16^3 + 11 \cdot 16^2 + 3 \cdot 16^1 + 12 \cdot 16^0 + 9 \cdot 16^{-1} + 15 \cdot 16^{-2} \\ &= 43836.62109375 (= 0.4383662109375 \cdot 10^5)\end{aligned}$$

2. Nicht normalisierte Gleitpunktzahlen:

- Die obigen Beispiele x_1 bis x_5 sind alle gemäss Definition 2.1 normalisiert, d.h. die erste Ziffer vor dem Punkt ist Null, die erste Ziffer nach dem Punkt ist ungleich Null für $x \neq 0$.
- Damit sind ihre Mantisse und der Exponent eindeutig definiert.
- Die folgenden Beispiele geben zur Illustration nicht normalisierte Varianten für x_1 und x_2 . Es ist offensichtlich, dass eine nicht normalisierte Darstellung bzgl. Mantisse und Exponent nicht mehr eindeutig ist (auch wenn der Wert immer gleich bleibt). Dies ist ein Zustand, der bei der Speicherung vermieden werden muss.

$$\begin{aligned} 1 \quad \tilde{x}_1 &= -0.002345 \cdot 10^5 = -23.45 \cdot 10^2 = -234500 \cdot 10^{-3} = \dots \\ 2 \quad \tilde{x}_2 &= 0.0111 \cdot 2^4 = 1.11 \cdot 2^2 = 111 \cdot 2^0 = 11100 \cdot 2^{-2} = \dots \end{aligned}$$

3. IEC/IEEE - Gleitpunktzahlen mit $B = 2$ (vgl. nächste Folie)

- single precision: Gesamtlänge der Zahl ist 32 Bit, wobei 1 Bit für das Vorzeichen, 23 Bit für die Mantisse, und 8 Bit für den Exponenten
- double precision: Gesamtlänge der Zahl ist 64 Bit, wobei 1 Bit für das Vorzeichen, 52 Bit für die Mantisse, und 11 Bit für den Exponenten

single: (32 bit)

V EEEEEEEE MBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

0 1 8 9 31

$$0.10000001101000000000000000000000000000 \equiv +1.101 * 2^{129-127} \equiv 6.5$$

$$0.00000001 \cdot 0000000000000000000000000000 = +1.0 * 2^{1-127} = 2^{-126} = x_{\min}$$

\equiv kleinste darstellbare Zahl

double: (64 bit)

0 1 11 12

63

Maschinenzahlen

Beispiele 2.1: IEC/IEEE

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinen-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- Bei den IEEE-Formaten ist die Mantisse so normalisiert, dass das erste Bit vor dem Punkt ungleich 0 sein muss
- Und da das erste Bit also immer 1 ist, wird es nicht gespeichert (hidden bit). Man gewinnt so ein Bit an Mantissenlänge
- Das Bit V erzeugt das Vorzeichen der Zahl: $V = 0$ entspricht einem positiven, $V = 1$ einem negativen Vorzeichen
- Im Exponenten wird ein von der Anzahl Bits abhängiger Biaswert subtrahiert, womit kein eigenes Vorzeichen-Bit benötigt wird
- Der Wert einer Gleitpunkt-Zahl x im IEEE-Format lautet also:

$$x = (-1)^V \cdot (\underbrace{1}_{\text{hidden bit}} \cdot \underbrace{\text{.} \overbrace{\text{MMM...MMM}}_{\text{fraction}}} \cdot 2^{(E...E)-bias})$$

Mantisse

- Wieso rechnet man eigentlich mit Gleitpunktzahlen? Nimmt man, abgesehen vom Vorzeichen, an, dass 8 dezimale Speichereinheiten zur Verfügung stehen, so liessen sich damit in den folgenden Systemen die folgenden positiven Zahlen darstellen:
 - Ganzzahlsystem:
 - ① kleinste darstellbare positive Zahl: 00000001
 - ② grösste darstellbare positive Zahl: 99999999
- Es lassen sich also im positiven Bereich alle ganzen Zahlen zwischen 1 und 99999999 ($= 10^8 - 1$) hinterlegen. Der Abstand zwischen aufeinanderfolgenden Zahlen ist konstant gleich 1.

Maschinenzahlen

Beispiele 2.1: Weshalb Gleitpunktzahlen

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinen-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- Festpunktsystem mit 4 Dezimalen:

- ① kleinste darstellbare positive Zahl: 0000.0001
- ② grösste darstellbare positive Zahl: 9999.9999

Es lassen sich im positiven Bereich Zahlen zwischen 0.0001 und 9999.9999 ($= 10^4 - 10^{-4}$) darstellen. Der Abstand zwischen aufeinanderfolgenden Zahlen ist konstant gleich 10^{-4} .

- Normalisiertes Gleitpunktsystem mit 6 Mantissen- und 2 Exponentenziffern (mit Bias 50)

- ① kleinste darstellbare positive Zahl: $0.100000 \cdot 10^{-50}$
- ② grösste darstellbare positive Zahl: $0.999999 \cdot 10^{49}$

Es lassen sich im positiven Bereich Zahlen von 10^{-51} bis $10^{49} - 1$ darstellen. Der Abstand zwischen aufeinanderfolgenden Zahlen ist allerdings variabel, wie in Kap. 2.4 gezeigt.

Maschinenzahlen

Aufgaben 2.2 [1]

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinen-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- ① Wie viele Stellen benötigt man, um die folgenden Zahlen als n-stellige Gleitpunktzahlen im Dezimalsystem darzustellen?
 $x_1=0.00010001$, $x_2=1230001$, $x_3=\frac{4}{5}$, $x_4=\frac{1}{3}$

- ② Bestimmen Sie alle dualen 3-stelligen positiven Gleitpunktzahlen mit einstelligem positiven binären Exponenten sowie ihren dezimalen Wert.
- ③ Wie viele verschiedene Maschinenzahlen gibt es auf einem Rechner, der 20-stellige Gleitpunktzahlen mit 4-stelligen binären Exponenten sowie dazugehörige Vorzeichen im Dualsystem verwendet? Wie lautet die kleinste positive und die größte Maschinenzahl?
- ④ Verstehen Sie den folgenden 'Witz'?

Es gibt 10 Gruppen von Menschen: Diejenigen, die das Binärsystem verstehen, und die anderen.

Maschinenzahlen

Aufgaben 2.2: Lösungen [1]

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinien-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinien-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

Maschinenzahlen

Aufgaben 2.2: Lösungen [1]

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinien-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinien-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

Approximations- und Rundungsfehler

Ungleichmässige Verteilung der Maschinenzahlen

- Die Maschinenzahlen sind nicht gleichmäßig verteilt.
- Ein Beispiel für alle binären normalisierten Gleitpunktzahlen mit 4-stelliger Mantisse und 2-stelligem Exponenten ist auf der nächsten Seite dargestellt.
- Zwangsläufig gibt es bei jedem Rechner eine grösste (x_{max}) und kleinste positive Maschinenzahl (x_{min}).
- Dabei gilt für normalisierte Gleitpunktzahlen:

$$x_{max} = B^{e_{max}} - B^{e_{max}-n} = (1 - B^{-n})B^{e_{max}}$$

$$x_{min} = B^{e_{min}-1}$$

- Wird auf die Normalisierung der Mantisse ($m_1 \neq 0$) verzichtet, führt dies zu sog. subnormalen Zahlen, die bis B^{m-n} hinunter reichen (IEEE Standard 754).

Approximations- und Rundungsfehler

Ungleichmässige Verteilung der Maschinenzahlen

HM 1,
Kapitel 2

Geschichte
der Zahldarstellung

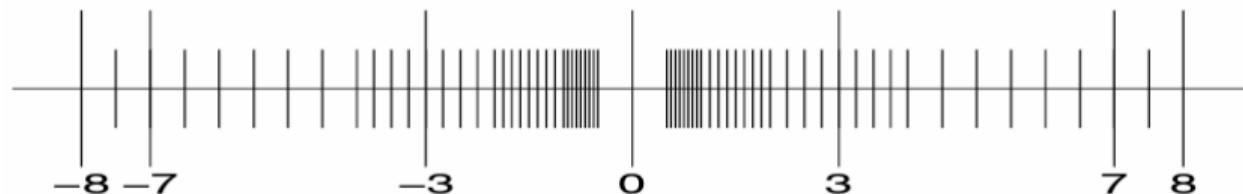
Maschinenzahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinen-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- Alle binären Maschinenzahlen mit $n = 4$ und $0 \leq e \leq 3$
(Abbildung entnommen aus Knorrenschild)



Approximations- und Rundungsfehler

Aufgabe 2.3

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinen-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- Schreiben Sie die kleinste und grösste binäre positive Maschinenzahl für die vorhergehende Abbildung ($n = 4$ und $0 \leq e \leq 3$) explizit auf und berechnen Sie deren Wert.
- Stimmt das mit x_{max} und x_{min} überein?

Approximations- und Rundungsfehler

Aufgabe 2.3: Lösungen

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinen-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

Approximations- und Rundungsfehler

Ungleichmässige Verteilung der Maschinenzahlen

- Zahlen, die ausserhalb des Rechenbereichs $[-x_{max}, x_{max}]$ liegen, sind im *Überlaufbereich (overflow)* und führen zum Abbruch der Rechnung (mit IEEE 754 konforme Systeme geben die Bitsequenz *inf* aus).
- Zahlen ungleich 0, die innerhalb des Bereichs $[-x_{min}, x_{min}]$ liegen, führen zu einem *Unterlauf (underflow)*. Dann ist es sinnvoll, die Rechnung mit 0 weiterzuführen.
- Offensichtlich ist die Anzahl n der Mantissestellen von entscheidender Bedeutung für den Bereich der Zahlen, die abgebildet werden können. Dies wird eindrücklich illustriert in folgendem Beispiel:

Approximations- und Rundungsfehler

Beispiel 2.2

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinen-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

Am 4. Juni 1996 startete zum ersten Mal eine Ariane 5-Rakete der ESA von Französisch Guyana aus. Die unbemannte Rakete hatte vier Satelliten an Bord. 36.7 Sekunden nach dem Start wurde in einem Programm versucht, den gemessenen Wert der horizontalen Geschwindigkeit von 64 Bit Gleitpunktdarstellung in 16 Bit Ganzahldarstellung (signed Integer) umzuwandeln. Da die entsprechende Masszahl grösser war als $2^{15} = 32768$, wurde ein Überlauf erzeugt. Das Lenksystem versagte daraufhin seine Arbeit und gab die Kontrolle an eine zweite, identische Einheit ab. Diese produzierte folgerichtig ebenfalls einen Überlauf. Da der Flug der Rakete instabil wurde und die Triebwerke abzubrechen drohten, zerstörte sich die Rakete selbst. Es entstand ein Schaden von ca. 500 Millionen Dollar durch den Verlust der Rakete und der Satelliten. Die benutzte Software stammte vom Vorgängermodell Ariane 4. Die Ariane 5 flog schneller und offensichtlich wurde dies bei der Repräsentation der Geschwindigkeit nicht beachtet.

Approximations- und Rundungsfehler

Beispiel 2.2

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinen-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- Absturz der Ariane 5 Rakete am 4. Juni 1996 (siehe https://youtu.be/gp_D8r-2hwk)



Rundungsfehler und Maschinengenauigkeit

Definition 2.2: Absoluter / Relativer Fehler

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinengenauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- Jede reelle Zahl, die von einem Rechner verwendet werden soll, aber selber keine Maschinenzahl ist, muss also durch eine solche ersetzt werden.
- Dabei entstehen Fehler, wie wir gesehen haben.

Definition 2.2:

- Hat man eine Näherung \tilde{x} zu einem exakten Wert x , so ist der Betrag der Differenz $|\tilde{x} - x|$ der **absolute Fehler**.
- Falls $x \neq 0$, so ist $|\frac{\tilde{x}-x}{x}|$ bzw. $\frac{|\tilde{x}-x|}{|x|}$ der **relative Fehler** dieser Näherung. In der Numerik ist der relative Fehler der wichtigere. Weshalb?

Rundungsfehler und Maschinengenauigkeit

- Natürlich sollte die Maschinenzahl dabei so gewählt werden, dass sie möglichst nahe bei der reellen Zahl liegt.
- Einfaches Abschneiden ist dazu nicht geeignet. Ein besseres Verfahren ist die Rundung (vgl. Aufg. 2.4).
- Beim Runden einer Zahl x wird eine Näherung unter den Maschinenzahlen gesucht, die einen minimalen absoluten Fehler $|rd(x) - x|$ aufweist.

Rundungsfehler und Maschinengenauigkeit

- Eine n -stellige dezimale Gleitpunktzahl

$\tilde{x} = 0.m_1 m_2 m_3 \dots m_n \cdot 10^e = rd(x)$ die durch die Rundung eines exakten Wertes x entstanden ist, hat also einen absoluten Fehler von höchstens

$$|rd(x) - x| \leq \underbrace{0.00\dots005}_{n} \cdot 10^e = 0.5 \cdot 10^{e-n}, \quad = 5 \cdot 10^{e-n-1}$$
$$= \frac{B}{2} \cdot B^{e-n-1}$$

wobei die 5 an der Stelle $n+1$ nach dem Dezimalpunkt auftritt.

- Für eine beliebige Basis gilt analog

$$|rd(x) - x| \leq \underbrace{0.00\dots00}_{n} \frac{B}{2} \cdot B^e = \frac{B}{2} \cdot B^{e-n-1},$$

Rundungsfehler und Maschinengenauigkeit

Beispiel 2.3

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinengenauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- Sei $x = 180.1234567 = 0.1801234567 \cdot 10^3$. Gerundet auf eine siebenstellige Mantisse ($n = 7$) erhält man $rd(x) = 0.1801235 \cdot 10^3$ und es gilt wegen $e = 3$

$$|rd(x) - x| = \underbrace{0.0000000433}_{n=7} \cdot 10^3 = 0.433 \cdot 10^{-4} \leq 0.5 \cdot 10^{-4}$$

Absoluter Fehler = $\frac{B}{2} \cdot B^{e-n-1}$

Relativer Fehler = $\frac{B}{2} \cdot B^{-n}$

Rundungsfehler und Maschinengenauigkeit

Aufgabe 2.4

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinengenauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- ① Vergewissern Sie sich anhand einfacher Zahlenbeispiele, dass die Rundung ein besseres Verfahren für die Abbildung einer reellen Zahl auf eine Maschinenzahl darstellt als einfaches Abschneiden der überzähligen Ziffern. Was ist der maximale Fehler, der durch das Abschneiden auftreten kann?
- ② Wir kennen die Rundungsregeln für das Dezimalsystem. Verallgemeinern sie diese für eine beliebige Basis B . Runden Sie anschliessend die folgenden Zahlen auf eine vierstellige Mantisse, berechnen Sie den absoluten Fehler der Rundung und vergewissern Sie sich, dass $|rd(x) - x| \leq \frac{B}{2} \cdot B^{e-n-1}$. Gilt diese Relation (bei gleichen Rundungsregeln) auch für ungerade Basen?
 - a) $(11.0100)_2$
 - b) $(11.0110)_2$
 - c) $(11.111)_2$
 - d) $(120.212)_3$
 - e) $(120.222)_3$
 - f) $(0.FFFF)_{16}$

Rundungsfehler und Maschinengenauigkeit

n-stellige Gleitpunktarithmetik

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinengenauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- Für die Berechnungen bedeutet Rundung, dass jede einzelne Operation (+, -, *, ...) auf $n+1$ Stellen genau gerechnet wird und das Ergebnis auf n Stellen gerundet wird (*n-stellige Gleitpunktarithmetik*).
- Jedes Zwischenergebnis wird also gerundet, nicht erst das Endergebnis.
- Das bedeutet auch, dass die einzelnen Rundungsfehler durch die Rechnung weitergetragen werden und allenfalls das Endergebnis verfälschen können.

Rundungsfehler und Maschinengenauigkeit

Definition 2.3: Maschinengenauigkeit

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinengenauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- Für den maximal auftretenden relativen Fehler bei der Rundung ergibt sich bei n-stelliger Gleitpunktarithmetik im Dezimalsystem:

$$\left| \frac{rd(x) - x}{x} \right| \leq 5 \cdot 10^{-n} \text{ (da } x \geq 10^{e-1}).$$

Definition 2.3:

- Die Zahl $\textcolor{red}{eps := 5 \cdot 10^{-n}}$ heisst **Maschinengenauigkeit**. Bei allgemeiner Basis B gilt $\textcolor{red}{eps := \frac{B}{2} \cdot B^{-n} = \frac{1}{2} \cdot B^{1-n}}$.
 - Sie entspricht dem maximalen relativen Fehler, der durch Rundung entstehen kann.

Rundungsfehler und Maschinengenauigkeit

Definition 2.3: Maschinengenauigkeit

Bemerkungen:

- Die Maschinengenauigkeit eps gemäss Def. 2.3 kann auch interpretiert werden als die grösste positive Maschinenzahl, für die auf dem Rechner $1 + \text{eps} = 1$ gerundet wird.
- **ACHTUNG:** Im IEEE-754 Standard und damit auch in Python, Matlab, etc. wird die Maschinengenauigkeit anders definiert: Dort entspricht eps dem Abstand zwischen 1 und der nächstgrösseren Maschinenzahl und ist damit genau doppelt so gross wie eps gemäss Def. 2.3. Falls nicht anders deklariert, gilt in diesen Unterlagen stets eps gemäss Def. 2.3.
- Die Maschinengenauigkeit eps darf nicht mit der (viel kleineren) kleinsten positiven Maschinenzahl x_{\min} verwechselt werden. Ein Rechner kann also auch mit deutlich kleineren Zahlen $x < \text{eps}$ noch 'genau' (nämlich mit einem relativen Rundungsfehler $\leq \text{eps}$) rechnen.

Rundungsfehler und Maschinengenauigkeit

Beispiel 2.4a

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinengenauigkeit

Fehlerfortpflan-
zung /
Konditionierung

Für das Format double precision in IEEE-754 gilt für die Mantisse $n = 53$ (hidden bit!), und damit ist für die Basis $B = 2$ die Maschinengenauigkeit

$$\text{eps} = \frac{1}{2} \cdot B^{1-n} = \frac{1}{2} \cdot 2^{1-53} = 2^{-53} = 1.110223\ldots \cdot 10^{-16}$$

Rundungsfehler und Maschinengenauigkeit

Beispiel 2.4b

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinengenauigkeit

Fehlerfortpflan-
zung /
Konditionierung

Am Freitag, dem 25. November 1983, schloss der Aktienindex von Vancouver bei 524.811 Punkten und eröffnete am folgenden Montag, dem 28. November, bei 1098.892 Punkten. Was war passiert? Seit dem Start im Januar 1982 bei 1000 Punkten war der Aktienindex kontinuierlich gefallen, trotz florierendem Handel und guter Wirtschaftslage. Der Index wurde ca. 3000 mal am Tag neu berechnet, jeweils auf vier Dezimalstellen genau. Doch statt auf drei Dezimalstellen zu runden, wurde die vierte Dezimalstelle einfach abgeschnitten. Der dabei maximal mögliche Fehler von 0.0009 mutet zwar klein an, doch bei 3000 Wiederholungen pro Tag konnte sich dieser Abschneidefehler auf bis zu $0.0009 \cdot 3000 = 2.7$ Punkte pro Tag aufsummieren. Über die Zeitspanne von fast zwei Jahren verlor der Index so fast die Hälfte seines Wertes. Dies wurde am 28. November basierend auf korrekter Rundung korrigiert. Größere Auswirkungen hatte diese Korrektur offenbar nicht, da zum damaligen Zeitpunkt das Volumen an Derivaten gering war.

Rundungsfehler und Maschinengenauigkeit

Aufgabe 2.5

HM 1,
Kapitel 2

Geschichte
der Zahlena-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinengenauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- ① Gesucht ist eine Näherung \tilde{x} zu $x = \sqrt{2} = 1.414213562\dots$ mit einem absoluten Fehler von höchstens 0.001.
- ② Es soll $2590 + 4 + 4$ in 3-stelliger Gleitpunktarithmetik gerechnet werden (im Dezimalsystem), einmal von links nach rechts und einmal von rechts nach links. Wie unterscheiden sich die Resultate?

Was lernen wir daraus? Beim Addieren sollte man die Summanden in der Reihenfolge aufsteigender Beträge sortieren

- ③ Berechnen Sie $s_{300} := \sum_{i=1}^{300} \frac{1}{i^2}$ sowohl auf- als auch absteigend, je einmal mit 3-stelliger und 5-stelliger Gleitpunktarithmetik. In Python müssen Sie dafür zuerst eine Funktion schreiben, die Ihnen eine reelle Zahl (hier $1/r^2$) auf die nächstgelegene Maschinenzahl mit $B = 10$ und $n = 3$ oder $n = 5$ runden.

Rundungsfehler und Maschinengenauigkeit

Aufgabe 2.5 Fortsetzung

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinengenauigkeit

Fehlerfortpflan-
zung /
Konditionierung

4. Es ist $\lim_{n \rightarrow \infty} (1 + \frac{1}{n})^n = e$. Erstellen Sie eine Tabelle mit Ihrem Rechner oder Python für $n = 1, 10, 100, \dots$ für den Ausdruck $(1 + \frac{1}{n})^n$ sowie den absoluten und relativen Fehler. Erklären Sie Ihre Beobachtungen.

n	$(1 + \frac{1}{n})^n$	absoluter Fehler	relativer Fehler
10^0			
10^2			
10^3			
10^4			
10^5			
10^6			
10^8			
10^9			
10^{10}			
10^{15}			
10^{16}			

Rundungsfehler und Maschinengenauigkeit

Aufgabe 2.5 Fortsetzung

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinen-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

5. Überlegen Sie sich einen kurzen iterativen Algorithmus, der die Maschinengenauigkeit Ihres Rechners prüft. Schliessen Sie aus dem Ergebnis, mit welcher Stellenzahl er operiert.

Rundungsfehler und Maschinengenauigkeit

Aufgabe 2.5: Lösungen

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinengenauigkeit

Fehlerfortpflan-
zung /
Konditionierung

Fehlerfortpflanzung bei Funktionsauswertungen

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinen-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- Wir haben gesehen, dass ein Rundungsfehler durch die Abbildung einer reellen Zahl x auf ihre Maschinenzahl \tilde{x} in die Berechnungen einfließt.
- Soll nun eine Funktion f an der Stelle x ausgewertet werden, wird ein zusätzlicher Fehler dadurch generiert, dass nicht $f(x)$ sondern $f(\tilde{x})$ berechnet wird.
- Für den fehlerbehafteten Wert \tilde{x} können wir den Fehler quantifizieren als $\Delta x = \tilde{x} - x$ (vgl. Def. 2.2) oder

$$\tilde{x} = x + \Delta x \quad \frac{|f(\tilde{x}) - f(x)|}{|f(x)|} \stackrel{?}{\sim} \frac{|\tilde{x} - x|}{|x|}$$

- Nun wollen wir den absoluten Fehler $|f(\tilde{x}) - f(x)|$ und den relativen Fehler $\frac{|f(\tilde{x}) - f(x)|}{|f(x)|}$ dieser Funktionsauswertung berechnen.

Fehlerfortpflanzung bei Funktionsauswertungen

- Aus der allg. Taylor-Reihe (bekannt aus der Analysis) einer Funktion $f(x)$ um den Entwicklungspunkt x_0

$$f(x) = \sum_{i=0}^{\infty} \frac{f^{(i)}(x_0)}{i!} (x - x_0)^i$$

erhalten wir für die Entwicklung von $f(\tilde{x})$ um den Entwicklungspunkt x

$$\begin{aligned} f(\tilde{x}) &= f(x + \Delta x) = \sum_{i=0}^{\infty} \frac{f^{(i)}(x)}{i!} (\Delta x)^i \\ &= f(x) + f'(x)\Delta x + \frac{f''(x)}{2}(\Delta x)^2 + \dots \end{aligned}$$

wobei wir in der Taylor-Reihe x durch \tilde{x} und x_0 durch x ersetzt haben.

Fehlerfortpflanzung bei Funktionsauswertungen

- Unter der Annahme $\Delta x \ll 1$ können die höheren Fehlerterme $(\Delta x)^n$ für $n \geq 2$ vernachlässigt werden und es ergibt sich die folgende Näherung

$$\begin{aligned}f(\tilde{x}) - f(x) &\approx f'(x)\Delta x \\&\approx f'(x)(\tilde{x} - x)\end{aligned}$$

bzw. bei beidseitiger Division durch $f(x)$ und rechtseitiger Multiplikation mit $\frac{x}{x}$:

$$\frac{f(\tilde{x}) - f(x)}{f(x)} \approx \frac{f'(x) \cdot x}{f(x)} \cdot \frac{\tilde{x} - x}{x}$$

Fehlerfortpflanzung bei Funktionsauswertungen

Wir erhalten also die folgenden Näherungen:

- Näherung für den **absoluten Fehler bei Funktionsauswertungen**:

$$\underbrace{|f(\tilde{x}) - f(x)|}_{\text{absoluter Fehler von } f(x)} \approx |f'(x)| \cdot \underbrace{|\tilde{x} - x|}_{\text{absoluter Fehler von } x}$$

- Näherung für den **relativen Fehler bei Funktionsauswertungen**:

$$\underbrace{\frac{|f(\tilde{x}) - f(x)|}{|f(x)|}}_{\text{relativer Fehler von } f(x)} \approx \underbrace{\frac{|f'(x)| \cdot |x|}{|f(x)|}}_{\text{Konditionszahl } K} \cdot \underbrace{\frac{|\tilde{x} - x|}{|x|}}_{\text{relativer Fehler von } x}$$

Fehlerfortpflanzung bei Funktionsauswertungen

Definition 2.4: Konditionszahl

- Den Faktor

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

nennt man **Konditionszahl**. Er gibt näherungsweise an, um wieviel grösser der relative Fehler der Funktionsauswertung $f(x)$ wird im Vergleich zum relativen Fehler von x .

- Man unterscheidet **gut konditionierte Probleme**, d.h. die Konditionszahl ist klein, und **schlecht konditionierte Probleme** (ill posed problems) mit grosser Konditionszahl. Bei gut konditionierten Problemen wird der relative Fehler durch die Auswertung der Funktion nicht grösser.

Fehlerfortpflanzung bei Funktionsauswertungen

Bemerkungen:

- \tilde{x} kann generell als fehlerbehafteter Näherungswert für x angesehen werden. Ob der Fehler nun durch Rundung oder andere Effekte verursacht wird (z.B. durch fehlerhafte Messungen) ist hierbei nicht von Belang.
- Bei Funktionsauswertungen pflanzt sich der absolute Fehler in x näherungsweise mit dem Faktor $f'(x)$ fort. Falls $|f'(x)| > 1$ wird der absolute Fehler grösser, falls $|f'(x)| < 1$ kleiner.
- Bei Funktionsauswertungen pflanzt sich der relative Fehler in x näherungsweise mit der Konditionszahl fort.

Fehlerfortpflanzung bei Funktionsauswertungen

Beispiele 2.5

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinen-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- Was lässt sich über die Fehlerfortpflanzung des absoluten Fehlers für die Funktion $f(x) = \sin(x)$ aussagen?
Da $|f'(x)| = |\cos(x)| \leq 1$, folgt dass der absolute Fehler in den Funktionswerten nicht grösser sein kann als in den x -Werten sondern eher kleiner.
- Bei der Funktion $f(x) = 1000 \cdot x$ wird wegen $f'(x) = 1000$ der absolute Fehler in der Funktionsauswertung um den Faktor 1000 grösser.
- Die Konditionszahl für das Quadrieren, also $f(x) = x^2$, ist $K = \frac{|2x| \cdot |x|}{|x^2|} = 2$, d.h. der relative Fehler verdoppelt sich in etwa. Dies ist aber noch keine schlechte Konditionierung.

Fehlerfortpflanzung bei Funktionsauswertungen

Beispiele 2.5

HM 1,
Kapitel 2

Geschichte
der Zahldarstellung

Maschinenzahlen

Approximations- und Rundungsfehler

Rundungsfehler und Maschinengenauigkeit

Fehlerfortpflanzung / Konditionierung

- Als Beispiel für ein schlecht konditioniertes Problem betrachten wir das Wilson-Polynom, definiert als

$$P(x) = \prod_{k=1}^{20} (x - k) = (x - 1)(x - 2) \cdot \dots \cdot (x - 20)$$

mit den Nullstellen

$$x_k \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20\}.$$

Ausmultipliziert erhält man

$$\begin{aligned} P(x) = & x^{20} - 210x^{19} + 20615x^{18} - 1256850x^{17} + 53327946x^{16} - 1672280820x^{15} \\ & + 40171771630x^{14} - 756111184500x^{13} + 11310276995381x^{12} \\ & - 135585182899530x^{11} + 1307535010540395x^{10} - 10142299865511450x^9 \\ & + 63030812099294896x^8 - 311333643161390640x^7 + 1206647803780373360x^6 \\ & - 3599979517947607200x^5 + 8037811822645051776x^4 \\ & - 12870931245150988800x^3 + 13803759753640704000x^2 \\ & - 8752948036761600000x + 2432902008176640000 \end{aligned}$$

Fehlerfortpflanzung bei Funktionsauswertungen

Beispiele 2.5

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinen-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- Wird nun der Koeffizient -210 von x^{19} um nur $2^{-23} (\approx 1.1920928 \cdot 10^{-7})$ verkleinert (resp. "gestört") auf neu -210.0000001192093 , verändert sich der Funktionswert $P(20)$ von ursprünglich 0 auf $-6.24997 \cdot 10^{17}$ und die Nullstellen des gestörten Polynoms werden zu

$$\tilde{x}_k \in \{1.00000, 2.00000, 3.00000, 4.00000, 5.00000, 6.00001, 6.99970, 8.00727, 8.91725, \\ 10.09527 \pm 0.64350i, 11.79363 \pm 1.65233i, 13.99236 \pm 2.51883i, 16.73074 \pm 2.81262i, \\ 19.50244 \pm 1.94033i, 20.84691\},$$

zum Teil also sogar aus dem komplexen Zahlenraum \mathbb{C} (auf die komplexen Zahlen werden wir in Kapitel 4 näher eingehen)

Fehlerfortpflanzung bei Funktionsauswertungen

Beispiele 2.5

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinen-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- Während nur ein Koeffizient von P mit dem absoluten Fehler von $\sim 10^{-7}$ gestört wurde, haben sich die Nullstellen um bis zu ~ 1 verändert, d.h. die Störung wurde um einen Faktor von $\sim 10^7$ verstärkt.
- Die Verstärkung des relativen Fehlers liegt in einem ähnlichen Bereich.
- Die Nullstellen von P sind also schlecht konditioniert. Schaut man sich den Graphen von $P(x)$ und der gestörten Version auf der nächsten Slide, sieht man die Auswirkungen der Störung.

Fehlerfortpflanzung bei Funktionsauswertungen

Beispiele 2.5

HM 1,
Kapitel 2

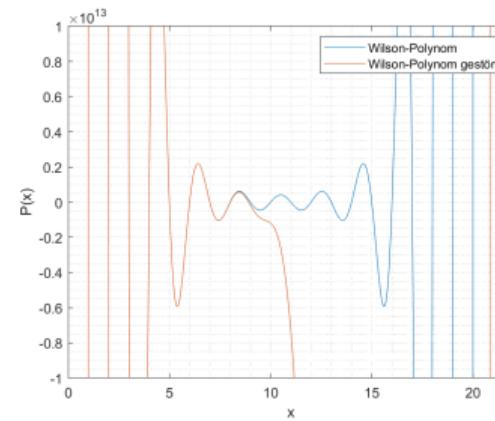
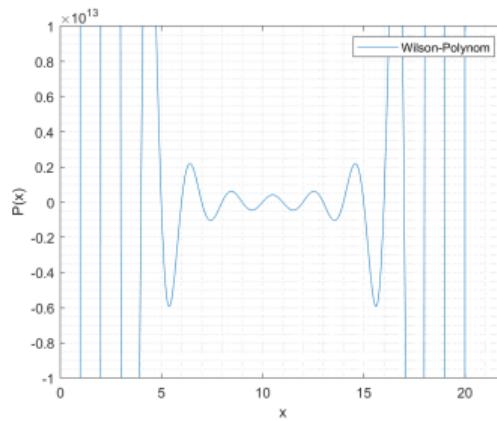
Geschichte der Zahlen- darstellung

Maschinen-
zahlen

Approxima- tions- und Rundungs- fehler

Rundungsfehler und Maschinen- genauigkeit

Fehlerfortpflan- zung / Konditionierung



Fehlerfortpflanzung bei Funktionsauswertungen

Beispiele 2.5

HM 1,
Kapitel 2

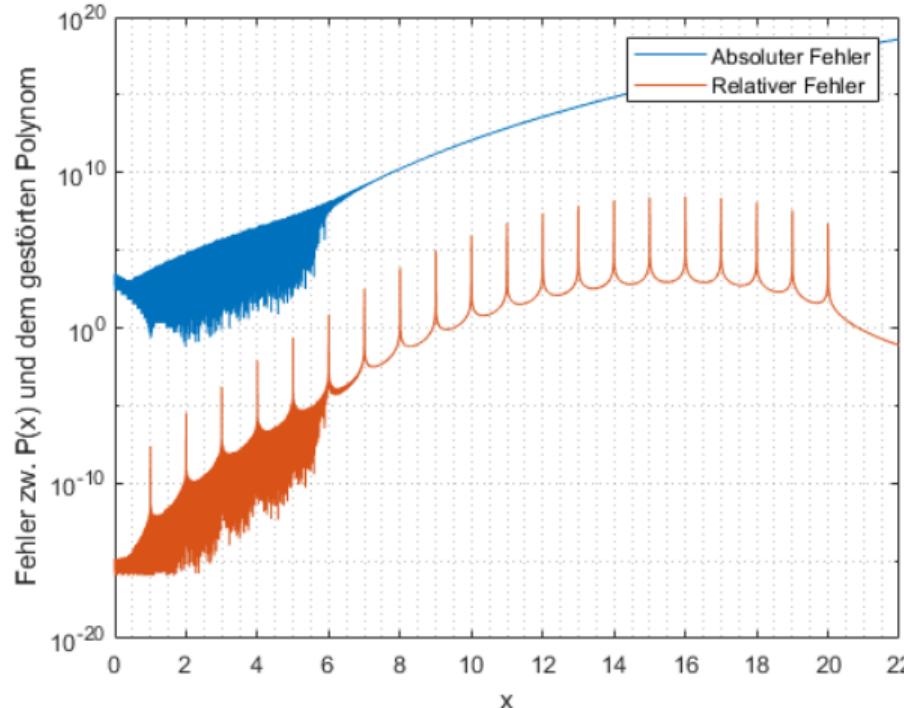
Geschichte der Zahlen- darstellung

Maschinen-
zahlen

Approxima- tions- und Rundungs- fehler

Rundungsfehler und Maschinen- genauigkeit

Fehlerfortpflan- zung / Konditionierung



Fehlerfortpflanzung bei Funktionsauswertungen

Fehlerfortpflanzung der Summation

- Betrachten wir nun die relativen Fortpflanzungsfehler für die grundlegenden arithmetischen Operationen.
- Für

$$f(x) = x + c \quad (c \in \mathbb{R})$$

haben wir für die Ableitung $f'(x) = 1$ und damit

$$\frac{|f(\tilde{x}) - f(x)|}{|f(x)|} \approx \frac{|x|}{|x+c|} \cdot \frac{|\tilde{x}-x|}{|x|}$$

bzw.

$$K = \frac{|x|}{|x+c|}$$

Fehlerfortpflanzung bei Funktionsauswertungen

Fehlerfortpflanzung der Summation

- Wenn x und die Konstante c gleiches Vorzeichen haben, gilt $K \leq 1$ dann haben wir also ein gut konditioniertes Problem.
- Was passiert aber, wenn x und c entgegengesetzte Vorzeichen haben und betragsmässig fast gleich gross sind? Dann wird $|x + c|$ sehr klein und somit K sehr gross, die Addition (bzw. Subtraktion) ist dann schlecht konditioniert.
- Dieses Phänomen nennt man auch **Auslöschung**. Es tritt immer dann auf, wenn ungefähr gleich grosse fehlerbehaftete Zahlen voneinander abgezogen werden und das Resultat anschliessend normalisiert wird.

Fehlerfortpflanzung bei Funktionsauswertungen

Beispiel 2.6

- Für die beiden reellen Zahlen $r = \frac{3}{5}$ und $s = \frac{4}{7}$ mit den normalisierten gerundeten Repräsentationen mit fünfstelliger Mantisse, also $\tilde{r} = (0.10011)_2$ und $\tilde{s} = (0.10010)_2$, berechnen wir die Differenz $r - s = \frac{1}{35}$ näherungsweise als

$$0.10011 \cdot 2^0 - 0.10010 \cdot 2^0 = 0.00001 \cdot 2^0 = 0.10000 \cdot 2^{-4} = \frac{1}{32}.$$

Für den relativen Fehler erhalten wir

$$\frac{\frac{1}{32} - \frac{1}{35}}{\frac{1}{35}} = 0.0938 \approx 9.4\%$$

was viel ist (zum Vergleich, dies ist rund dreimal grösser als die Maschinengenauigkeit $2^{-5} = 0.0313$). Für die Berechnung mit dreistelliger Mantisse erhalten wir

$$0.101 - 0.101 = 0$$

und damit einen Fehler von 100%.

Fehlerfortpflanzung bei Funktionsauswertungen

Beispiel 2.7

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinen-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- Gegeben seien die drei Werte

$$x_1 = 123.454 \cdot 10^9$$

$$x_2 = 123.446 \cdot 10^9$$

$$x_3 = 123.435 \cdot 10^9$$

- Legt man eine 5-stellige dezimale Gleitpunktarithmetik zugrunde, so wird durch Rundung

$$\tilde{x}_1 = 0.12345 \cdot 10^{12}$$

$$\tilde{x}_2 = 0.12345 \cdot 10^{12}$$

$$\tilde{x}_3 = 0.12344 \cdot 10^{12}$$

Fehlerfortpflanzung bei Funktionsauswertungen

Beispiel 2.7

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinen-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- Man erhält statt

$$x_1 - x_2 = x_1 + (-x_2) = 8 \cdot 10^6$$

$$x_1 - x_3 = x_1 + (-x_3) = 19 \cdot 10^6$$

die fehlerhaften Werte

$$\tilde{x}_1 - \tilde{x}_2 = 0$$

$$\tilde{x}_1 - \tilde{x}_3 = 10 \cdot 10^6$$

- Dies zeigt, dass die Subtraktion ein schlecht konditioniertes Problem darstellt, wenn x_1 und x_2 nahe beieinander liegen.

Fehlerfortpflanzung bei Funktionsauswertungen

Aufgabe 2.6

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinen-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

- ① Untersuchen Sie, ob die Multiplikation und die Division zweier Zahlen gut oder schlecht konditionierte Funktionsauswertungen sind.

Fehlerfortpflanzung bei Funktionsauswertungen

Aufgabe 2.6: Lösung

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinen-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

Bei Division und Multiplikation
passieren keine zusätzlichen Rundungsfehler.

Fragen?

HM 1,
Kapitel 2

Geschichte
der Zahlen-
darstellung

Maschinen-
zahlen

Approxima-
tions- und
Rundungs-
fehler

Rundungsfehler
und Maschinen-
genauigkeit

Fehlerfortpflan-
zung /
Konditionierung

