

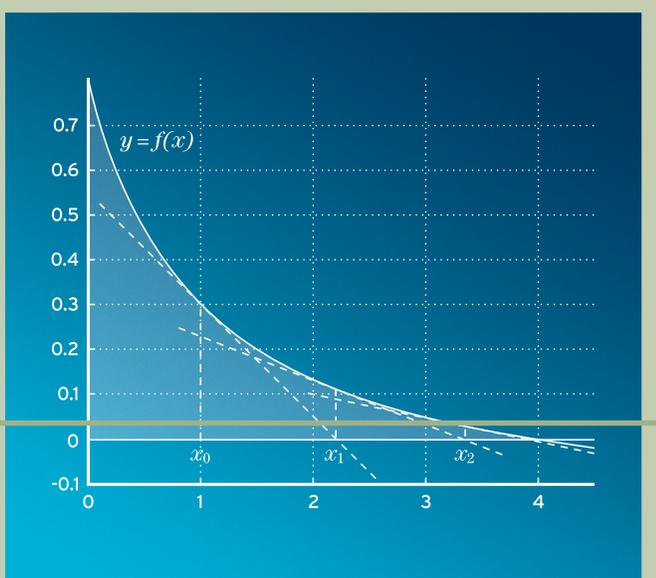
Michael Knorrenschild



Mathematik-Studienhilfen

Numerische Mathematik

Eine beispielorientierte Einführung



8., aktualisierte Auflage

HANSER

Knorrenschild Numerische Mathematik



Bleiben Sie auf dem Laufenden!

Hanser Newsletter informieren Sie regelmäßig über neue Bücher und Termine aus den verschiedenen Bereichen der Technik. Profitieren Sie auch von Gewinnspielen und exklusiven Leseproben. Gleich anmelden unter

www.hanser-fachbuch.de/newsletter



Ihr Plus – digitale Zusatzinhalte!

Auf unserem Download-Portal finden Sie zu diesem Titel kostenloses Zusatzmaterial.

Geben Sie dazu einfach diesen Code ein:

`plus-g7jg9-vdxcf`

plus.hanser-fachbuch.de

Mathematik–Studienhilfen

Herausgegeben von

Prof. Dr. Bernd Engelmann

Hochschule für Technik, Wirtschaft und Kultur Leipzig,

Fachbereich Informatik, Mathematik und Naturwissenschaften

Zu dieser Buchreihe

Die Reihe Mathematik-Studienhilfen richtet sich vor allem an Studierende technischer und wirtschaftswissenschaftlicher Fachrichtungen an Fachhochschulen und Universitäten.

Die mathematische Theorie und die daraus resultierenden Methoden werden korrekt aber knapp dargestellt. Breiten Raum nehmen ausführlich durchgerechnete Beispiele ein, welche die Anwendung der Methoden demonstrieren und zur Übung zumindest teilweise selbstständig bearbeitet werden sollten.

In der Reihe werden neben mehreren Bänden zu den mathematischen Grundlagen auch verschiedene Einzelgebiete behandelt, die je nach Studienrichtung ausgewählt werden können. Die Bände der Reihe können vorlesungsbegleitend oder zum Selbststudium eingesetzt werden.

Lieferbar:

Gramlich, *Lineare Algebra*

Gramlich, *Anwendungen der Linearen Algebra*

Knorrenschild, *Numerische Mathematik*

Knorrenschild, *Vorkurs Mathematik*

Martin, *Finanzmathematik*

Nitschke, *Geometrie*

Preuß, *Funktionaltransformationen*

Sachs, *Wahrscheinlichkeitsrechnung/Statistik*

Tittmann, *Graphentheorie*

Michael Knorrenschild

Numerische Mathematik

Eine beispielorientierte Einführung

8., aktualisierte Auflage

HANSER

Über den Autor:

Prof. Dr. rer. nat. Michael Knorrenschild, Bochum



Print-ISBN: 978-3-446-48024-7

E-Book-ISBN: 978-3-446-48126-8

Alle in diesem Werk enthaltenen Informationen, Verfahren und Darstellungen wurden zum Zeitpunkt der Veröffentlichung nach bestem Wissen zusammengestellt. Dennoch sind Fehler nicht ganz auszuschließen. Aus diesem Grund sind die im vorliegenden Werk enthaltenen Informationen für Autor:innen, Herausgeber:innen und Verlag mit keiner Verpflichtung oder Garantie irgendeiner Art verbunden. Autor:innen, Herausgeber:innen und Verlag übernehmen infolgedessen keine Verantwortung und werden keine daraus folgende oder sonstige Haftung übernehmen, die auf irgendeine Weise aus der Benutzung dieser Informationen – oder Teilen davon – entsteht. Ebenso wenig übernehmen Autor:innen, Herausgeber:innen und Verlag die Gewähr dafür, dass die beschriebenen Verfahren usw. frei von Schutzrechten Dritter sind. Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt also auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benützt werden dürften.

Die endgültige Entscheidung über die Eignung der Informationen für die vorgesehene Verwendung in einer bestimmten Anwendung liegt in der alleinigen Verantwortung des Nutzers.

Bibliografische Information der Deutschen Nationalbibliothek:

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter <http://dnb.d-nb.de> abrufbar.

Dieses Werk ist urheberrechtlich geschützt.

Alle Rechte, auch die der Übersetzung, des Nachdruckes und der Vervielfältigung des Werkes, oder Teilen daraus, vorbehalten. Kein Teil des Werkes darf ohne schriftliche Einwilligung des Verlages in irgendeiner Form (Fotokopie, Mikrofilm oder einem anderen Verfahren), auch nicht für Zwecke der Unterrichtsgestaltung – mit Ausnahme der in den §§ 53, 54 UrhG genannten Sonderfälle –, reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

Wir behalten uns auch eine Nutzung des Werks für Zwecke des Text- und Data Mining nach § 44b UrhG ausdrücklich vor.

© 2024 Carl Hanser Verlag GmbH & Co. KG, München

www.hanser-fachbuch.de

Lektorat: Dipl.-Ing. Natalia Silakova-Herzberg

Coverkonzept: Marc Müller-Bremer, www.rebranding.de, München

Covergestaltung: Max Kostopoulos

Titelmotiv: © Max Kostopoulos, unter Verwendung von Grafiken von © gettyimages.de/filo

Satz: Michael Knorrenschild

Druck: CPI Books GmbH, Leck

Printed in Germany

Inhalt

1	Rechnerarithmetik und Gleitpunktzahlen	1
1.1	Grundbegriffe und Gleitpunktarithmetik	1
1.2	Auslöschung	8
1.3	Fehlerrechnung	9
1.3.1	Fehlerfortpflanzung in arithmetischen Operationen	10
1.3.2	Fehlerfortpflanzung bei Funktionsauswertungen	11
2	Numerische Lösung von Nullstellenproblemen	17
2.1	Problemstellung	17
2.2	Das Bisektionsverfahren	18
2.3	Die Fixpunktiteration	19
2.4	Das Newton-Verfahren und seine Abkömmlinge	24
2.5	Konvergenzgeschwindigkeit	28
2.6	Das Horner-Schema – schnelle Auswertung von Polynomen	29
3	Numerische Lösung linearer Gleichungssysteme	33
3.1	Problemstellung	33
3.2	Der Gauß-Algorithmus	34
3.3	Fehlerfortpflanzung beim Gauß-Algorithmus und Pivottisierung	39
3.4	Dreieckszerlegungen von Matrizen	41
3.4.1	Die LR-Zerlegung	41
3.4.2	Die Cholesky-Zerlegung	44
3.4.3	Die QR-Zerlegung	46

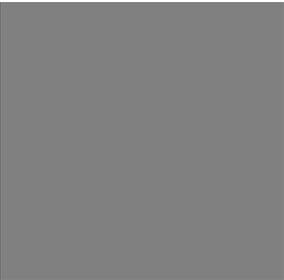
3.5	Fehlerrechnung bei linearen Gleichungssystemen	53
3.6	Iterative Verfahren	58
4	Numerische Lösung nichtlinearer Gleichungssysteme	67
4.1	Problemstellung	67
4.2	Das Newton-Verfahren für Systeme	69
5	Interpolation	73
5.1	Problemstellung	73
5.2	Polynominterpolation	74
5.2.1	Das Neville-Aitken-Schema	78
5.2.2	Der Fehler bei der Polynominterpolation	80
5.3	Splineinterpolation	84
5.3.1	Problemstellung	84
5.3.2	Interpolation mit kubischen Splines	85
6	Ausgleichsrechnung	93
6.1	Problemstellung	93
6.2	Lineare Ausgleichsprobleme	95
6.3	Nichtlineare Ausgleichsprobleme	101
6.4	Das Gauß-Newton-Verfahren	102
7	Numerische Differenziation und Integration	107
7.1	Numerische Differenziation	107
7.1.1	Problemstellung	107
7.1.2	Differenzenformeln für höhere Ableitungen	112
7.1.3	Differenzenformeln für partielle Ableitungen	112
7.1.4	Extrapolation	113
7.2	Numerische Integration	120
7.2.1	Problemstellung	120
7.2.2	Interpolatorische Quadraturformeln	123
7.2.3	Der Quadraturfehler	124
7.2.4	Transformation auf das Intervall $[a, b]$	125
7.2.5	Der Fehler der summierten Quadraturformeln	127
7.2.6	Newton-Cotes-Formeln	128

7.2.7	Gauß-Formeln	129
7.2.8	Extrapolationsquadratur	131
7.2.9	Praktische Aspekte	134
8	Anfangswertprobleme gewöhnlicher Differenzialgleichungen	137
8.1	Problemstellung	137
8.2	Das Euler-Verfahren	139
8.3	Praktische Aspekte	144
8.4	Weitere Einschrittverfahren	145
8.5	Weitere Verfahren	151
	Lösungen	153
	Literatur	171
	Stichwortverzeichnis	173

Zusatzmaterial

MATLAB¹-Programme auf <https://plus.hanser-fachbuch.de/>. Code dazu auf Seite 1.
 Inhalt: m-Files zur Erstellung aller Bilder in diesem Buch sowie zu einigen der hier vorgestellten Algorithmen.

¹ MATLAB ist eingetragenes Warenzeichen von The Mathworks Inc.



Vorwort

Numerische Mathematik gehört zu den Teilgebieten der Mathematik, die von Ingenieuren im beruflichen Alltag verwendet werden. Durch verstärkte Verwendung von Computer-Simulationen in allen Bereichen erhöht sich die Bedeutung dieses Themas, in dem Fragestellungen der Mathematik und der Informatik zusammenkommen, zunehmend.

Der vorliegende Band deckt die wichtigsten Themen der numerischen Mathematik für Studierende der Ingenieurwissenschaften ab und entspricht in etwa dem Umfang einer einsemestrigen Lehrveranstaltung. Das Anliegen ist dabei, die Ideen der wichtigsten numerischen Verfahren zu präsentieren und anhand einer Vielzahl von Beispielen deren charakteristische Eigenschaften zu illustrieren. Auf Beweise und längere Herleitungen wird dabei weitgehend verzichtet. Vorausgesetzt werden Vorkenntnisse zur elementaren Differenzial- und Integralrechnung sowie zur linearen Algebra im Umfang etwa einer Anfängervorlesung zu diesen Themen.

Die Darstellungsweise profitiert von Erfahrungen, die ich in Lehrveranstaltungen zur Numerischen Mathematik für Studierende der Ingenieurwissenschaften an der Rheinisch-Westfälischen Technischen Hochschule Aachen, der Simon Fraser University in Burnaby (Kanada), der Eidgenössischen Technischen Hochschule Zürich und der Hochschule Bochum gesammelt habe. Die Anordnung der Themen folgt der bewährten Reihenfolge von Grundlagen der Gleitpunktarithmetik über die numerische Lösung von eindimensionalen Gleichungen, von linearen und nichtlinearen Gleichungssystemen, die Behandlung von Interpolations- und Ausgleichsproblemen bis hin zu numerischer Differenziation und Integration. Den Abschluss bildet ein Einblick in die numerische Lösung von Anfangswertaufgaben gewöhnlicher Differenzialgleichungen.

Die Entstehung und Weiterentwicklung dieses Buchs wurde im Laufe der Jahre von verschiedener Seite tatkräftig und wohlwollend unterstützt. Zuerst ist das Team des Hanser-Verlags zu nennen, beginnend 2003 mit Frau Christine Fritzsch bis heute mit

Frau Natalia Silakova. Für fachliche Ratschläge danke ich dem Herausgeber Prof. Dr. Bernd Engelmann. Herrn Dr. Thomas Schenk gebührt Dank für die kritische Durchsicht weiter Teile des Manuskripts.

Für die achte Auflage werden zu diesem Buch MATLAB[®]-Programme bereitgestellt. Diese sind auf <https://plus.hanser-fachbuch.de/> zu finden. Der dafür benötigte Code befindet sich vorne auf Seite 1. Darunter sind MATLAB²-Programme, die zur Erstellung der Bilder in diesem Buch benutzt werden, sowie einige in MATLAB implementierte Algorithmen. Die Umsetzung der Algorithmen orientiert sich an den Schreibweisen im Buch und den vorgestellten Beispielen. Die betreffenden Stellen sind mit ↔ plus.hanser gekennzeichnet. Die Programme sind nicht auf besondere Effizienz ausgelegt, sondern sollen Interessierten einen schnellen Einstieg in numerische Experimente ermöglichen und zu Erweiterungen anregen.

Hinweise und Anregungen aus dem Leserkreis sind auch weiterhin jederzeit willkommen.

Bochum, im Mai 2024

Michael Knorrenschild

² MATLAB ist eingetragenes Warenzeichen von The Mathworks Inc.

1

Rechnerarithmetik und Gleitpunktzahlen

In der Numerischen Mathematik geht es in der Regel um die näherungsweise Berechnung von Lösungen von Gleichungen oder anderen Größen wie z. B. Funktionswerte oder Integrale mithilfe von Computern. Dies geschieht aus zwei möglichen Gründen:

- Diese Größen sind auf dem Papier nicht exakt berechenbar, also muss es mit anderen Mitteln geschehen.
- Die Größen sind zwar auf dem Papier exakt bestimmbar, aber die Anwendung erfordert, diese wiederholt und zuverlässig in kurzer Zeit zur Verfügung zu stellen, sodass eine Rechnung von Hand auch wieder nicht infrage kommt.

Der Computer hat jedoch zwei prinzipielle Handicaps:

- Er kann durch die beschränkte Stellenzahl nicht alle Zahlen exakt darstellen.
- Er kann die gewünschten Rechnungen nicht exakt ausführen.

Im Folgenden werden Auswirkungen dieser Handicaps anhand von Beispielen und Aufgaben veranschaulicht.

■ 1.1 Grundbegriffe und Gleitpunktarithmetik

Wir beginnen mit der Frage, wie Zahlen auf dem Rechner dargestellt werden. Vom Taschenrechner kennen wir Formate wie z.B: $1.234 E 12$, was für $1.234 \cdot 10^{12}$ steht, die sogenannte wissenschaftliche Darstellung. Die Verwendung des Exponenten erlaubt eine Kommaverschiebung und damit große Zahlenbereiche. Auf dem Rechner ist es ganz ähnlich.

Definition

Eine n -stellige Gleitpunktzahl zur Basis B hat die Form

$$x = \pm (0.z_1 z_2 \dots z_n)_B \cdot B^E \quad \text{und den Wert} \quad \pm \sum_{i=1}^n z_i \cdot B^{E-i} \quad (1.1)$$

wobei $z_i \in \{0, 1, \dots, B-1\}$ und, falls $x \neq 0$, $z_1 \neq 0$ (**normalisierte Gleitpunktdarstellung**). Den Anteil $(0.z_1 z_2 \dots z_n)_B$ bezeichnet man auch als **Mantisse**. Für den Exponenten $E \in \mathbb{Z}$ gilt: $m \leq E \leq M$. ■

Beispielsweise ist also $x = -(0.2345)_{10} \cdot 10^3$ eine 4-stellige Gleitpunktzahl und hat den Wert -234.5 .

Übliche Basen sind $B = 2$ (Dualzahlen), $B = 8$ (Oktalzahlen), $B = 10$ (Dezimalzahlen) und $B = 16$ (Hexadezimalzahlen). Für letztere benötigt man für eine eindeutige Schreibweise 16 verschiedene Zeichen, man verwendet dabei die Ziffern $0, 1, \dots, 9$ sowie die Buchstaben A, \dots, F , wobei $A \triangleq 10, B \triangleq 11, \dots, F \triangleq 15$. Die Werte n, m, M, B sind maschinenabhängig (wobei unter Maschine der Rechner zusammen mit dem benutzten Compiler zu verstehen ist).

Als Beispiel erwähnen wir die IEC/IEEE-Gleitpunktzahlen. Dabei unterscheidet man zwei Grundformate ($B = 2$):

- **single format** Gesamtlänge der Zahl ist 32 Bit. Dieses teilt sich auf in 1 Bit für das Vorzeichen, 23 Bit für die Mantisse und 8 Bit für den Exponenten.
- **double format** Gesamtlänge der Zahl ist 64 Bit. Dieses teilt sich auf in 1 Bit für das Vorzeichen, 52 Bit für die Mantisse und 11 Bit für den Exponenten.

Das Vorzeichenbit $v \in \{0, 1\}$ erzeugt das Vorzeichen der Zahl über den Faktor $(-1)^v$, d. h. $v = 0$ ergibt positives Vorzeichen, $v = 1$ negatives. Eine umfassende Abhandlung dieser und anderer Formate findet man in [13].

Aufgaben

1.1 Welchen Wert haben die folgenden Gleitpunktzahlen im Dezimalsystem:

$$x_1 = 0.76005 \cdot 10^5, \quad x_2 = 0.571 \cdot 10^{-3} ?$$

1.2 Welchen Wert haben die folgenden Gleitpunktzahlen im Dualsystem:

$$x_1 = 0.111 \cdot 2^3, \quad x_2 = 0.1001 \cdot 2^{-3} ?$$

1.3 Wie viele Stellen n benötigt man, um die folgenden Zahlen als n -stellige Gleitpunktzahlen im Dezimalsystem darzustellen?

$$x_1 = 0.00010001, \quad x_2 = 1230001, \quad x_3 = \frac{4}{5}, \quad x_4 = \frac{1}{3}$$

Bei der letzten Aufgabe haben Sie festgestellt, dass nicht jede reelle Zahl als Gleitpunktzahl dargestellt werden kann. Dies trifft insbesondere auf Zahlen zu, die

unendlich viele Stellen benötigen würden, beispielsweise kann $\frac{1}{7}$ nicht als Gleitpunktzahl im Dezimalsystem dargestellt werden. Ebenso kann z. B. 12345 nicht als 3-stellige Gleitpunktzahl im Dezimalsystem geschrieben werden. Die Lage ist sogar noch ernster, denn es gilt:



Die Menge der auf einem Rechner darstellbaren Zahlen, die sog. **Maschinen-zahlen**, ist endlich.

Aufgaben

- 1.4 Bestimmen Sie alle dualen 3-stelligen Gleitpunktzahlen mit einstelligem Exponenten sowie ihren dezimalen Wert. Hinweis: Sie sollten 9 finden.
- 1.5 Wie viele verschiedene Maschinenzahlen gibt es auf einem Rechner, der 20-stellige Gleitpunktzahlen mit 4-stelligen Exponenten sowie dazugehörige Vorzeichen im Dualsystem verwendet? Wie lautet die kleinste positive und die größte Maschinenzahl?

Auch sind die Maschinenzahlen ungleichmäßig verteilt. Bild 1.1 zeigt alle binären normalisierten Gleitpunktzahlen mit 4-stelliger Mantisse und 2-stelligem Exponenten.

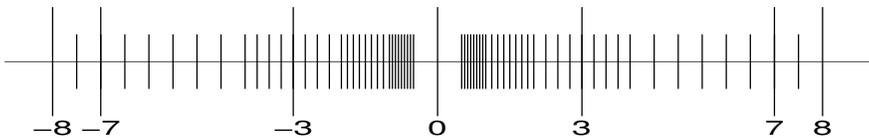


Bild 1.1 Alle binären Maschinenzahlen mit $n = 4$ und $0 \leq E \leq 3$

Unter den endlich vielen Maschinenzahlen gibt es zwangsläufig eine größte und eine kleinste:



- Die größte Maschinenzahl ist $x_{max} = (1 - B^{-n}) B^M$,
- die kleinste positive ist $x_{min} = B^{m-1}$.

x_{min} basiert auf der normalisierten Gleitpunktdarstellung. Sieht man von der Normalisierung $z_1 \neq 0$ in (1.1) ab, führt dies auf die **subnormalen Zahlen**, die bis hinunter zu B^{m-n} reichen (IEEE Standard 754). Führt eine Rechnung in den Zahlenbereich der subnormalen Zahlen, so bezeichnet man dies als **graduellen Unterlauf** (gradual underflow). Ein (echter) **Unterlauf** (underflow) tritt erst unterhalb der subnormalen Zahlen auf. In diesem Fall wird meist mit Null weitergerechnet.

Taucht im Verlauf einer Rechnung eine Zahl auf, die betragsmäßig größer als x_{max} ist, so bezeichnet man dies als **Überlauf** (overflow). Mit IEEE 754 konforme Systeme setzen diese Zahl dann auf eine spezielle Bitsequenz **inf** und geben diese am Ende aus.¹

Jede reelle Zahl, mit der im Rechner gerechnet werden soll und die selbst keine Maschinenzahl ist, muss also durch eine Maschinenzahl ersetzt werden. Idealerweise wählt man diese Maschinenzahl so, dass sie möglichst nahe an der reellen Zahl liegt (Rundung).

Definition

Hat man eine Näherung \tilde{x} zu einem exakten Wert x , so bezeichnet $|\tilde{x} - x|$ den **absoluten Fehler** dieser Näherung. ■

Beispiel 1.1

Gesucht ist eine Näherung \tilde{x} zu $x = \sqrt{2} = 1.414213562\dots$ mit einem absoluten Fehler von höchstens 0.001.

Lösung: $\tilde{x}_1 = 1.414$ erfüllt das Verlangte, denn $|\tilde{x} - x| = 0.000213562\dots \leq 0.001$. Andere Möglichkeiten sind $\tilde{x}_2 = 1.4139$. \tilde{x}_1 stimmt auf 4 Ziffern mit dem exakten Wert überein, \tilde{x}_2 nur auf 3. Eine größere Anzahl an übereinstimmenden Ziffern bedeutet aber keinesfalls immer einen kleineren absoluten Fehler, wie das Beispiel $x = \sqrt{3} = 1.732050808\dots$ und $\tilde{x}_1 = 2.0$, $\tilde{x}_2 = 1.2$ zeigt: \tilde{x}_1 hat keine gültige Ziffer, \tilde{x}_2 hat eine gültige Ziffer, trotzdem besitzt \tilde{x}_1 den kleineren absoluten Fehler. ■



Beim Runden einer Zahl x wird eine Näherung $rd(x)$ unter den Maschinenzahlen gesucht, die einen minimalen absoluten Fehler $|x - rd(x)|$ aufweist. Dabei entsteht ein (unvermeidbarer) Fehler, der sog. **Rundungsfehler**. ■

¹ Achtung: IEEE 754 regelt nicht die Rechnung mit integer-Größen. Ein overflow in einer integer-Variablen kann zu falschen Ergebnissen ohne jede Fehlermeldung führen. Hier ist also die besondere Aufmerksamkeit des Benutzers gefordert.

Eine n -stellige dezimale Gleitpunktzahl $\tilde{x} = \pm(0.z_1z_2\dots z_n)_B \cdot 10^E = \text{rd}(x)$, die durch Rundung eines exakten Wertes x entstand, hat also einen absoluten Fehler von höchstens

$$|x - \text{rd}(x)| \leq \underbrace{0.00\dots 005}_{n \text{ Nullen}} \cdot 10^E = 0.5 \cdot 10^{-n+E}.$$

Rechnet man mit diesen Maschinenzahlen weiter, so werden die entstandenen Rundungsfehler weiter durch die Rechnung getragen. Unter **n -stelliger Gleitpunktarithmetik** versteht man, dass jede einzelne Operation (wie z. B. $+$, $-$, $*$, \dots) auf $n+1$ Stellen genau gerechnet wird und das Ergebnis dann auf n Stellen gerundet wird. Erst dann wird die nächste Operation ausgeführt. Jedes Zwischenergebnis wird also auf n Stellen gerundet, nicht erst das Endergebnis einer Kette von Rechenoperationen. Von nun an werden wir uns, wenn nichts anderes gesagt ist, auf dezimale Gleitpunktarithmetik beziehen.

Aufgabe

1.6 Bekanntlich ist $\lim_{n \rightarrow \infty} (1 + \frac{1}{n})^n = e$. Versuchen Sie damit auf Ihrem Rechner näherungsweise e zu berechnen, indem Sie immer größere Werte für n einsetzen. Erklären Sie Ihre Beobachtung.

Beispiel 1.2

Es soll $2590 + 4 + 4$ in 3-stelliger Gleitpunktarithmetik (im Dezimalsystem) gerechnet werden und zwar zum einen mit Rechnung von links nach rechts und zum anderen von rechts nach links.

Lösung: Alle 3 Summanden sind exakt darstellbar. Als Ergebnis erhält man, bei Rechnung von links nach rechts:

$$2590 + 4 = 2594 \xrightarrow{\text{runden}} 2590, \quad 2590 + 4 = 2594 \xrightarrow{\text{runden}} 2590.$$

Die beiden kleinen Summanden gehen damit gar nicht sichtbar in das Ergebnis ein. Rechnet man jedoch in anderer Reihenfolge

$$4 + 4 = 8 \xrightarrow{\text{runden}} 8, \quad 8 + 2590 = 2598 \xrightarrow{\text{runden}} 2600$$

so erhält man einen genaueren Wert, sogar den in 3-stelliger Gleitpunktarithmetik besten Wert (2598 wird bestmöglich durch die Maschinenzahl 2600 dargestellt). ■

Es kommt also bei n -stelliger Gleitpunktarithmetik auf die Reihenfolge der Operationen an, anders als beim exakten Rechnen. Man sieht, dass in der zweiten Rechnung die kleinen Summanden sich erst zu einem größeren Summanden finden, der sich dann auch in der Gesamtsumme auswirkt. Beginnt man die Rechnung mit dem größten Summanden, so werden die kleinen nacheinander vom größten verschluckt und spielen gar keine Rolle mehr. Als Faustregel kann man daher festhalten:



Beim Addieren sollte man die Summanden in der Reihenfolge aufsteigender Beträge addieren.

Dadurch erreicht man – bei gleicher Rechenzeit! – ein wesentlich genaueres Ergebnis. Ein eindrucksvolles Beispiel ist das folgende.

Beispiel 1.3

Es soll $s_{300} := \sum_{i=1}^{300} \frac{1}{i^2}$ berechnet werden.

Lösung: Mit dezimaler Gleitpunktarithmetik erhält man

$$s_{300} = 1.6416062828976228698\dots \quad \text{bei exakter Rechnung}$$

$$s_{141} = s_{142} = \dots = s_{300} = 1.6390 \quad \text{5-stellig gerechnet, addiert von 1 bis 300}$$

$$s_{300} = 1.6416 \quad \text{5-stellig gerechnet, addiert von 300 bis 1}$$

$$s_{14} = s_{15} = \dots = s_{300} = 1.59 \quad \text{3-stellig gerechnet, addiert von 1 bis 300}$$

$$s_{300} = 1.64 \quad \text{3-stellig gerechnet, addiert von 300 bis 1.}$$

Bei 3- bzw. 5-stelliger Rechnung und geeigneter Wahl der Summationsreihenfolge wird also das auf 3 bzw. 5 Stellen genaue exakte Ergebnis erzielt.

Dagegen wird bei 3- bzw. 5-stelliger Rechnung und ungeschickter Wahl der Summationsreihenfolge das exakte Ergebnis nur auf 1 bzw. 2 Stellen genau erreicht. Dies macht den Einfluss deutlich, den die Summationsreihenfolge bei der Rechnung auf einem Computer besitzt. ■

Aufgabe

1.7 Weisen Sie durch Betrachtung von Rundungsfehler und Stellenzahl nach, dass in obigem Beispiel der Summenwert bei der Summation von 1 bis 300 bei 3- bzw. 5-stelliger Rechnung ab s_{14} bzw. s_{141} stagniert. Ab welchem Index stagniert der Summenwert bei n -stelliger Rechnung?

Man beachte, dass die verschiedenen Möglichkeiten der Berechnung der Summe in obigem Beispiel genau gleiche Gleitpunktoperationen benötigen, die Rechenzeit ist also stets die gleiche. Der einzige Unterschied besteht in der Reihenfolge der Operationen.

Als ein Maß für den Rechenaufwand kann man die Anzahl der durchgeführten Rechenschritte in der Gleitpunktarithmetik heranziehen, d. h. die Anzahl der Gleitpunktoperationen, im Englischen kurz „Flops“ („floating point operations“) genannt. Manchmal bezeichnet man auch eine Operation der Form $a + b \cdot c$, also eine Addition und eine Multiplikation zusammen, als einen Flop. Wir werden hier der Einfachheit

halber aber nicht alle Flops zählen, sondern nur die Punktrechenoperationen, also Multiplikationen und Divisionen. Als Maß für die Rechengeschwindigkeit eines Rechners ist die Einheit „flops per second“, also die Anzahl der möglichen Gleitpunktoperationen pro Sekunde, üblich. Der derzeit (Ende 2023) weltweit schnellste Rechner („Frontier“) steht beim Oak Ridge National Laboratory (USA) und hat eine Leistung von 1194 Petaflops (1.194 Exaflops), also $1194 \cdot 10^{15}$ Operationen pro Sekunde, der schnellste Rechner in Deutschland („JUWELS Booster Module“) steht im Forschungszentrum Jülich und belegt weltweit Platz 18 mit ca. 44.12 Petaflops².

Wir haben bisher nur den absoluten Fehler betrachtet. Dieser für sich allein sagt aber nicht viel aus – man kann z. B. die Qualität eines Messwertes nicht beurteilen, wenn man nur weiß, dass ein Widerstand R auf z. B. $\pm 2\Omega$ genau gemessen wurde. Zur Beurteilung muss man berücksichtigen, wie groß der Wert, den man messen möchte, wirklich ist. Man muss also den absoluten Fehler in Relation zur Größe der zu messenden Werte sehen, und dazu dient der relative Fehler:

Definition

Hat man eine Näherung \tilde{x} zu einem exakten Wert $x \neq 0$, so bezeichnet $\left| \frac{\tilde{x} - x}{x} \right|$ den **relativen Fehler** dieser Näherung. ■

In der Literatur findet man oft auch \tilde{x} im Nenner statt x . Der relative Fehler wird auch gern in % angegeben, d. h. statt von einem relativen Fehler von z. B. 0.15 redet man auch von 15 %.

Der maximal auftretende relative Fehler bei Rundung kann bei n -stelliger Gleitpunktarithmetik als

$$eps := \frac{B}{2} \cdot B^{-n}$$

angegeben werden. eps ist die kleinste positive Zahl, für die auf dem Rechner $1 + eps \neq 1$ gilt. Man bezeichnet eps auch als **Maschinengenauigkeit**. Es gilt dann:

$$\text{rd}(x) = (1 + \varepsilon) x \quad \text{mit } |\varepsilon| \leq eps.$$

Dies besagt, dass ε , also der relative Fehler der Näherung $\text{rd}(x)$ an x , stets durch die Maschinengenauigkeit beschränkt ist.

² Eine aktuelle Liste der 500 schnellsten Rechner findet man unter www.top500.org