Universität Heidelberg

Numerik Zusammenfassung

by Charles Barbret

5 Mai, 2018

Inhaltsverzeichnis

1 Fehleranalyse

1.1

Zahlendarstellung und Rundungsfehler Schema für Gleitkomma Zahl:

$$x = \pm m * b^{\pm e} \tag{1}$$

Basis $b \in \mathbb{N}$ b ≥ 2 [1]
Mantisse $m = m_1 b^{-1} + m_2 b^{-2} + ... \in \mathbb{R}$ [2]
Exponent e = $e_1 b^1 + e_2 b^2 + ... \in \mathbb{N}_0$
$\forall m_i \in m$ und $\forall e_i \in e$ gilt $e_i, m_i \in \{0, ..., b - 1\}$
Sollte b = 10 sein befinden wir uns im Dezimalsystem
⇒ Es gibt keine Ziffer ¿ 9 = 10 - 1 ⇔ b - 1 *q.e.d*
Jede Ziffer $(m_i, e_i, b, \pm, \pm)$ braucht man eine Speicherzelle, wobei b im Computer bereits eingespeichert ist
⇒ muss nicht explizit angegeben werden
X wird gespeichert als: $(\pm)[m_1, ..., m_r](\pm)[e_{s-1}, ..., e_0]$
======= [12pt,a4paper]article graphicx [utf8]inputenc [ngerman]babel datetime
breqn
hyperref
amssymb
soul

[1] Beispiel: 2^e oder 10^e

[2] Beispiel: $m_1 = 3, m_2 = 1, m_3 = 4 \Rightarrow m = 314$ normal kommt nach m_1 ein Komma, also m = 3,14

Universität Heidelberg

Numerik Zusammenfassung

by Charles Barbret

5 Mai, 2018

Inhaltsverzeichnis

2 Fehleranalyse

2.1 Zahlendarstellung und Rundungsfehler Schema für Gleitkomma Zahl

$$x = \pm m * b^{\pm e}$$

Basis $b \in \mathbb{N}$ b ≥ 2 [3]
Mantisse $m = m_1 b^{-1} + m_2 b^{-2} + ... \in \mathbb{R}$ [4]
Exponent e = $e_1 b^1 + e_2 b^2 + ... \in \mathbb{N}_0$
$\forall m_i \in m$ und $\forall e_i \in e$ gilt $e_i, m_i \in \{0, ..., b - 1\}$
Sollte b = 10 sein befinden wir uns im Dezimalsystem
    $\Rightarrow$ Es gibt keine Ziffer ¿ 9 = 10 - 1 $\Leftrightarrow$ b - 1 *q.e.d*
Jede Ziffer $(m_i, e_i, b, \pm, \pm)$ braucht man eine Speicherzelle, wobei b im Computer bereits eingespeichert ist
    $\Rightarrow$ muss nicht explizit angegeben werden
X wird gespeichert als: $(\pm)[m_1, \ldots, m_r](\pm)[e_{s-1}, \ldots, e_0]$
r + 1 Einträge um m zu speichern ($\pm$ ist ein Eintrag)
s + 1 Einträge um s zu speichern ($\pm$ ist ein Eintrag)
$X_{max/min} = \pm(1 - b^{-}r) * b^{b^s - 1}$ [5]
$X_{posmin/negmax} = \pm b^{-b^s}$ [6]

2.1.1 Rundungsoperation

$\|x - rd(x)\| = min_{y \in A}\|x - y\|$ [7]
Dies verläuft von D $\rightarrow$ A, wobei D:=$[X_{min}, x_{negmax}]\{0\}[X_{posmin}, X_{max}]$ ist und A die Menge der darstellbaren Zahlen [8]

$$rd(x) = \pm \begin{cases} m_1 \ldots m_{53} * 2^e & für m_{54} = 0 \\ (m_1 \ldots m_{53} + 2^{-53}) * 2^{e9} & für m_{54} = 1 \end{cases}$$

Der absolute Rundungsfehler sieht aus:

$$|x - rd(x)| \leq \frac{1}{2} b^{-s} b^e$$

[3] Beispiel: 2^e oder 10^e

[4] Beispiel: $m_1 = 3, m_2 = 1, m_3 = 4 \Rightarrow m = 314$ normal kommt nach m_1 ein Komma, also m = 3,14

[5] 1 oder 0 für erste Speicherzelle, sonst nur 1en

[6] 1 oder 0 für erste Speicherzelle, eine 1, sonst nur 0en

[7] Wobei diese Operation x als nächst darstellbare Zahl zurück gibt

[8] D gibt die Theoretisch minimalen Zahlen bis Theoretisch maximalen Zahlen an

TODO: insert image
absolut weil er noch vom Exponenten abhängt
Der relative Fehler:

$$|\frac{x - rd(x)}{x} \leq \frac{1}{2}\frac{b^{-r}b^e}{|m|b^e}$$