Übung 1 - Numerisches Programmieren

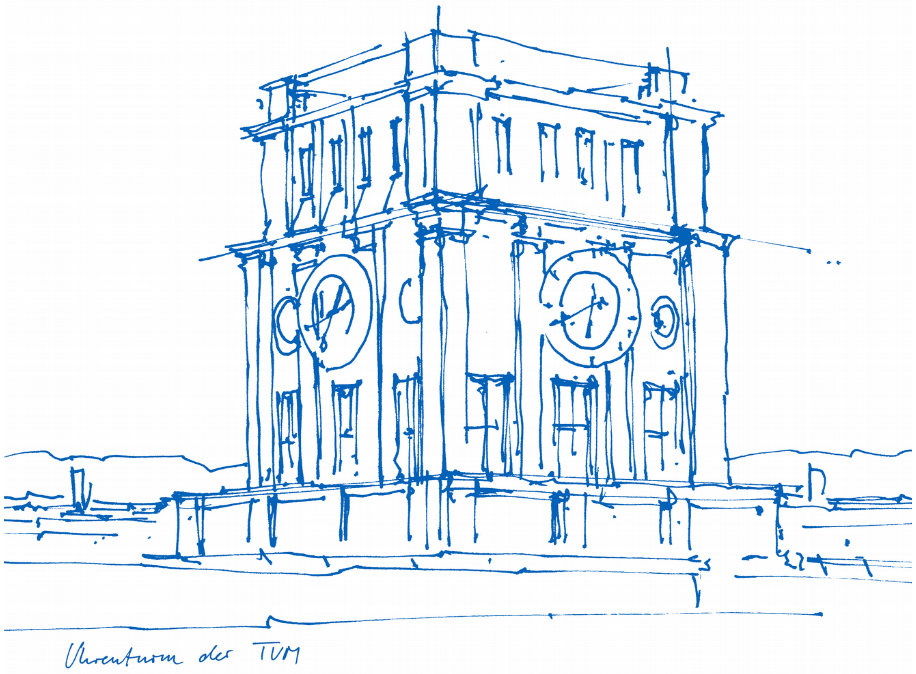
Michael Obersteiner

Technische Universität München

Fakultät für Informatik

Lehrstuhl für Wissenschaftliches Rechnen

BigBlueButton, 11. November 2020



Vorstellung

Name: Michael Obersteiner

Email: oberstei@in.tum.de

Raum: 02.05.043

Doktorarbeit:

- Hochdimensionale Numerik
- Numerische Simulationen
- High Performance Computing
- Plasma Physik
- Machine Learning

Notenbonus

- 4 Programmieraufgaben mit je 100 Punkten
- 280 Punkte nötig für Notenbonus
- Gruppen mit je 3 Studenten (Anmeldung auf Moodle)
- Webinterface zum Testen der eigenen Implementierung
- Abgabe über Moodle
- Ergebnisse über Moodle
- Kontakt: Steffen Seckler (seckler@in.tum.de)

Übung 1 – Zahldarstellung: Ganzzahl

Verscheidene Zahlensysteme:

- Basis B
- Stellenwertigkeit für Position i: B^i
- Codierung an Stelle i: $x_i \in [0, B-1]$
- Berechnung: $x = \sum_{i=0}^{n-1} x_i B^i$
- Beispiel für Umwandlung: B = 2; **x = 43**

Stelle	$64 = 2^6$	$32 = 2^5$	$16 = 2^4$	$8 = 2^3$	$4 = 2^2$	$2 = 2^1$	$1 = 2^0$
Codierung	0	1	0	1	0	1	1

Übung 1 – Zahldarstellung: Ganzzahl 2

Andere Berechnungsmöglichkeit:

- Teile durch 2
- Beispiel für Umwandlung: **x = 43**

43 / 2	=	21 Rest 1
21 / 2	=	10 Rest 1
10 / 2	=	5 Rest 0
5 / 2	=	2 Rest 1
2 / 2	=	1 Rest 0
1 / 2	=	0 Rest 1

- **Ergebnis:** 101011

Übung 1 – Zahldarstellung: Gleitkomma

- Negative Indexe sind jetzt erlaubt!
- Berechnung: $x = \sum_{i=-k}^{n-1} x_i B^i$
- Beispiel für Umwandlung: **x = 1/3**

$$1/3 * 2 = 2/3 + 0$$
$$2/3 * 2 = 1/3 + 1$$

- Ergebnis: $0,\overline{01}$
- Andere Möglichkeit: Schriftliches Dividieren (deutlich Fehleranfälliger!)

Übung 1 – Zahldarstellung

Bearbeitung Aufgabe 1

Umwandlung in Binär

a) 19 =

47 =

511 =

b) 1/7 =

1/10 =

Übung 1 – Zahldarstellung: Zweierkomplement

- Vorderstes Bit hat negativen Vorfaktor!
- Beispiel: n = 8; x = -67 = -128 + 61

10111101

- Alternativ:
  - Erstelle Zahlendarstellung für 67 01000011
  - Invertiere Bitfolge 10111100
  - Addiere 1 10111101

Übung 1 – Zahldarstellung

Bearbeitung Aufgabe 2

Zahlenbereich bei Ganzzahlen mit 2er Komplement:

1 Byte :

2 Byte :

n Byte :

Übung 1 – Zahldarstellung: einfaches Runden

- Abrunden: X|0... → X
- Aufrunden: X|1... → X + 1

- Beispiele:
  1,001|101… → 1,010
  1,001|01… → 1,001

Wie wirkt sich dies auf die Assoziativität aus?

Übung 1 – Zahldarstellung

Bearbeitung Aufgabe 3

Assoziativgesetz bei Rudung auf 4 binäre gültige Stellen:

(-8 + 11) + 0,75 ?=? -8 + (11 + 0,75)

Übung 1 – Zahldarstellung: einfaches Runden

- Abrunden: X|0... → X
- Aufrunden: X|1... → X + 1

- Beispiele:
  1,001|101… → 1,010
  1,001|01… → 1,001

Wie wirkt sich dies auf die Assoziativität aus?
→ keine Assoziativität mehr!

- Beispiel (4 gültige Stellen):
  - (-1000 + 1011) + 0.11 = 11 + 0.11 = 11.11 → 3.75
  - -1000 + (1011 + 0.11) = -1000 + 1100 = 100 → 4

Übung 1 – IEEE Gleitkomma: Definition

- Darstellung einer Gleitkommazahl durch 3 Komponenten:
  - Mantisse
  - Exponent
  - Vorzeichen
- Beispiel: $-1{,}00101 * 2^{5}$
- Normalisierung: Immer genau eine 1 vor dem Komma!
  - $101{,}01 * 2^{5} \rightarrow 1{,}0101 * 2^{7}$
  - $0{,}10101 * 2^{5} \rightarrow 1{,}0101 * 2^{4}$
- Speicherlayout:

Vorzeichen	Exponent	Mantisse

  - Vorzeichen (1 Bit): 1 für negativ, 0 für positiv
  - Exponent (8 Bit): Exponent + Offset 127 (Vermeidet 2er Komplement)
  - Mantisse (23 Bit): Nachkommastellen

Übung 1 – IEEE Gleitkomma: Rundung

- Wie bisher:
  - Abrunden: X|0... → X
- Neu:
  - X|1Y
    - Aufrunden falls irgendwann eine 1 folgt: Y > 0 → X + 1
    - Uneindeutiger Fall mit nur 0en danach: Runden zu gerader Mantisse
      - Y = 0 und X gerade → X
      - Y = 0 und X ungerade → X+1
- Beispiele:

  1,001|101… → 1,010

  1,001|01… → 1,001

  1,001|10000…0 → 1,010

  1,000|10000…0 → 1,000

Übung 1 – Zahldarstellung

Bearbeitung Aufgabe 4 a+b

Umwandlung von 11/10 in IEEE Darstellung:

a) Umwandlung in Binär und Normalisierung

b) Eintragen in IEEE Speicherformat:

Übung 1 – Gleitkomma: Maschinenengenauigkeit

- Maß für die relative Genauigkeit der Gleitkommazahlen
- Definition: $\epsilon_{MA} = \max\{x \in \mathbb{R}^+ | rd(1+x) = 1\}$
- Faustregel: bei m Mantissenbits $2^{-(m+1)}$
- Aussage: $\epsilon_{rel} = \frac{rd(x) - x}{x} \leq \epsilon_{MA}$

- Andere Definitionen in der Literatur:
  - Kleinste Schrittweite → 2^{-m}

Übung 1 – Zahldarstellung

Bearbeitung Aufgabe 4 c+d+e

c) Nennen einer Zahl die nicht exakt dargestellt wird (welcher Fehler?)

d) Welche Maschinengenauigkeit?

e) Erste Ganzzahl (>0) die nicht mehr exakt dargestellt werden kann?

Übung 1 – Numerische Effekte: Auslöschung

- Subtraktion zweier ähnlicher Zahlen in Gleitkommaarithmetik
- **Problem:** Verlust an gültigen Stellen → ungenaue Ergebnisse

- Beispiel:
  - rd(1,00001…..) – rd(1,00000….) = 1,00001 – 1,00000 = **0,00001**
  - Aus 6 gültigen Stellen wird 1!
  - Würden wir nur 5 gültige Stellen verwenden so wäre das Ergebnis 0!

  → Weitere Rechnungen eventuell signifikant verfälscht (Teilen durch 0!)

- Manchmal vermeidbar durch Umformung (siehe Aufgabe 6):
- s → 0

$$s_{new} = \sqrt{2-\sqrt{4-s^2}} = \frac{\sqrt{2-\sqrt{4-s^2}}\cdot\sqrt{2+\sqrt{4-s^2}}}{\sqrt{2+\sqrt{4-s^2}}} = \frac{|s|}{\sqrt{2+\sqrt{4-s^2}}}$$

Übung 1 – Numerische Effekte: Absorption

- Addieren zweier Zahlen unterschiedlicher Größenordnung
- **Problem:** Minimale oder keine Änderung der größeren Zahl
  → Genauigkeitsverlust

- Beispiel:
  - 100000000 + 1 = rd(100000001) = 100000000

- Besonders problematisch bei for loops! (JavaScript kennt nur float!)
- Manchmal lösbar durch Änderung der Additionsreihenfolge:
  - -1000 + (1011 + 0.11) = -1000 + 1011 = 11 → 3 (ungenau!)
  - (-1000 + 1011) + 0.11 = 11 + 0.11 = 11.11 → 3.75 (exakt!)
- Typischerweise weniger problematisch als Auslöschung