Darstellung Gleitkommazahlen

by

Dr. Günter Kolousek

Zahlen

- ganze Zahlen
  - vorzeichenlos
  - vorzeichenbehaftet
    - mit designierten VZ-Bit
    - Einerkomplement
    - Zweierkomplement
- Kommazahlen
  - Festkommazahlen
  - Gleitkommazahlen

Festkommazahlen

- Position des Komma ist fixiert
- Beispiele
  - 9.87654321 … große Genauigkeit
  - 98765432.1 … großer Wertebereich
  - 9876.54321 … Kompromiss
- Problem: Welche Position soll gewählt werden?
- Genauigkeit im gesamten Bereich gesichert
  - dezimale Festkommazahlen zur Rechnung mit Geldbeträgen!
    - z.B. `decimal` in C#

Überblick Gleitkommazahlen

- Gleitkommazahlen
  - vs. rationale/reelle Zahlen
  - alternative Namen
    - Gleitpunktzahlen, Fließkommazahlen, Fließpunktzahlen
    - floating point numbers
- IEEE 754
  - binäre Formate: 16 Bits, 32 Bits, 64 Bits, 128 Bits
    - nicht standardisiert: 80 Bits (z.B. IA-32 Prozessoren)
  - dezimale Formate: 32 Bits, 64 Bits, 128 Bits

Gleitkommazahlen

- Komma ”gleitet”
- Gleitkommadarstellung
  - $s \cdot m \cdot b^e$
    - s … Vorzeichen
    - m … Mantisse
    - b … Basis
    - e … Exponent
  - Beispiel: $123.456 = +1 \cdot 0.123456 \cdot 10^3$
    - z.B. in C++: `0.123456e3`
  - Beispiel: $123.456 = +1 \cdot 1234.56 \cdot 10^-1$
  - Welche Darstellung soll zur Abspeicherung verwendet werden?

Gleitkommazahlen

- Komma ”gleitet”
- Gleitkommadarstellung
  - $s \cdot m \cdot b^e$
    - s … Vorzeichen
    - m … Mantisse
    - b … Basis
    - e … Exponent
  - Beispiel: $123.456 = +1 \cdot 0.123456 \cdot 10^3$
    - z.B. in C++: `0.123456e3`
  - Beispiel: $123.456 = +1 \cdot 1234.56 \cdot 10^-1$
  - Welche Darstellung soll zur Abspeicherung verwendet werden?
    - $\rightarrow$ Normalisierung!
  - Wie wird Exponent abgespeichert?
    - $\rightarrow$ Charakteristik!

Gleitkommazahlen – 2

- binäre Exponentialdarstellung
  - $11.1_2 = 11.1_2 \cdot 2^0$
  - $11.1_2 = 1.11_2 \cdot 2^1$
  - $11.1_2 = 0.111_2 \cdot 2^2$
  - $11.1_2 = 0.0111_2 \cdot 2^3$
- *Normalisierung*
  - Vorkommastelle ist 1
    - da immer 1 $\rightarrow$ muss nicht mitgespeichert werden!
    - $11.1_2 = 1.11_2 \cdot 2^1$ $\rightarrow$ Speicherdarstellung für Mantisse: 11

IEEE Zahlenformat

IEEE Zahl besteht aus 3 Teilen:

1. Getrenntes Vorzeichenbit S
   - S = 0 $\rightarrow$ positiv, S = 1 $\rightarrow$ negativ
   - d.h.: $s = (-1)^S$
2. Exponent E (auch Charakteristik genannt)
   - $E = e + B$, B … Biaswert
   - d.h. $e = E - B$
3. Mantisse M
   - in normalisierter Darstellung
   - ohne Abspeicherung der führenden 1

IEEE Zahlenformat – 2

- Exponent
  - $E = 0$
    - $M = 0 \rightarrow$ Gleitkommazahl 0
    - $M > 0 \rightarrow$ denormalisierte Zahlen (m nicht mehr normalisiert)
  - E lauter 1er
    - $M = 0 \rightarrow$ `inf`
    - $M > 0 \rightarrow$ nan
- Spezielle Werte
  - 0
    - vorzeichenbehaftet, d.h. +0 und -0
    - $1/(-0)$ liefert `-inf`!
  - `+inf` und `-inf`
  - nan, z.B. `0.0 / 0.0`

IEEE Zahlenformat – 3

- half, single, double, extended
- single
  - Exponent: 8 Bits ($B = 127$), Mantisse: 23 Bits
  - ca. 7.2 Dezimalstellen
- double
  - Exponent: 52 Bits ($B = 1023$), Mantisse: 52 Bits
  - ca. 15.9 Dezimalstellen

Gleitkommadarstellung?

1. Vorkommazahl ermitteln
2. Nachkommazahl ermitteln
3. Normalisieren
4. Exponent ermitteln
5. Vorzeichen bestimmen
6. Gleitkommazahl bilden

Gleitkommadarstellung?

1. Vorkommazahl ermitteln
2. Nachkommazahl ermitteln
3. Normalisieren
4. Exponent ermitteln
5. Vorzeichen bestimmen
6. Gleitkommazahl bilden

19.2 in binärer Darstellung?

Umrechnen – 1

1. Vorkommazahl ermitteln
   - $19 \div 2 = 9R1$
   - $9 \div 2 = 4R1$
   - $4 \div 2 = 2R0$
   - $2 \div 2 = 1R0$
   - $1 \div 2 = 0R1$

   $19_{10} = 10011_2$
2. Nachkommazahl ermitteln
   - $0.2 \cdot 2 = 0.4 \rightarrow 0$
   - $0.4 \cdot 2 = 0.8 \rightarrow 0$
   - $0.8 \cdot 2 = 1.6 \rightarrow 1$
   - $0.6 \cdot 2 = 1.2 \rightarrow 1$
   - $0.2 \cdot 2 = 0.4 \rightarrow 0$
   - $0.4 \cdot 2 = 0.8 \rightarrow 0$
   - $0.8 \cdot 2 = 1.6 \rightarrow 1$
   - $\dots$

   $0.2 = 0.00\overline{1100}$

Umrechnen – 2

3. Normalisieren $19.2_{10} = 1.0011001\overline{1100} \cdot 2^4 = 1.00\overline{1100} \cdot 2^4$
4. Exponent ermitteln: $4 + B$

vs. Festkommadarstellung?

- Nachteile gegenüber Festkomma
  - Rechenaufwand prinzipiell höher
  - Genauigkeit nicht im gesamten Wertebereich gesichert
- Vorteile gegenüber Festkomma
  - Adaption des Wertebereiches und der Genauigkeit
  - Berechnung in FPU (Gleitkommaeinheit)