TECHNISCHE UNIVERSITÄT CHEMNITZ
FAKULTÄT FÜR MATHEMATIK
DR. RALF HIELSCHER

SOSE15

Übungen zu Numerische Methoden für Ingenieure

http://www.tu-chemnitz.de/~rahi

Übungsblatt 1

Aufgabe 1: Machen Sie sich mit folgenden Zahldarstellungen vertraut. Beispielhaft verwenden wir die Basis $b = 2$ und die Bitlänge $N = 8$.

nicht negative ganze Zahlen: $x = [x_7x_6x_5x_4x_3x_2x_1x_0]_2$,

$$x = x_7 2^7 + x_6 2^6 + x_5 2^5 + x_4 2^4 + x_3 2^3 + x_2 2^2 + x_1 2^1 + x_0 2^0.$$

ganze Zahlen: $x = [x_7x_6x_5x_4x_3x_2x_1x_0]_{\overline{2}}$,

$$x = -x_7 2^7 + x_6 2^6 + x_5 2^5 + x_4 2^4 + x_3 2^3 + x_2 2^2 + x_1 2^1 + x_0 2^0.$$

Fixpunktzahlen: $x = [x_7.x_6x_5x_4x_3x_2x_1x_0]_{\overline{2}}$,

$$x = -x_7 2^0 + x_6 2^{-1} + x_5 2^{-2} + x_4 2^{-3} + x_3 2^{-4} + x_2 2^{-5} + x_1 2^{-6} + x_0 2^{-7}.$$

Gleitpunktzahlen: $x = [\underbrace{x_4x_3x_2x_1x_0}_{\text{Mantisse}} \mid \underbrace{e_2e_1e_0}_{\text{Exponent}}]_{\overline{2}}$,

$$x = [x_4.x_3x_2x_1x_0]_{\overline{2}} \cdot 2^{[e_2e_1e_0]_{\overline{2}}}.$$

Aufgabe 2: Finden Sie für die folgenden Zahlen alle möglichen Darstellungen.

a) 1 b) -1 c) 99 d) -35 e) $\frac{7}{8}$ f) $\frac{3}{64}$

Aufgabe 3: Überzeugen Sie sich, dass man in der Ganzzahlendarstellung mit gewöhlicher schriftlicher Addition addieren kann. Überprüfen Sie dafür

a) 1 + (-1) = 0, b) 13 + (-5) = 8, c) 5 + (-13) = -8.

Aufgabe 4: Geben Sie jeweils die größte, die kleinste und die betragskleinste von Null verschiedene Zahl für alle in Aufgabe 1 aufgeführten Darstellungsarten an.

Aufgabe 5: Skizzieren Sie die Menge aller Gleitpunktzahlen mit Mantissenlänge $M = 3$ und Exponentenlänge 2 auf einem Zahlenstrahl. Vergleichen Sie mit einer geeigneten Fixpunktdarstellung.

Aufgabe 6: Ist x eine beliebige reelle Zahl, so bezeichnen wir mit $[x]$ die Maschinenzahl m für welche $|m - x|$ minimal wird, d.h. $[x]$ ist die Maschinenzahl die x am besten approximiert. Des Weiteren heißt die Differenz $|[x] - x|$ absoluter Fehler und der Quotient $\frac{|[x]-x|}{|x|}$ relativer Fehler der Darstellung $[x]$.

a) Bestimmen sie den maximalen absoluten Fehler und den maximalen relativen Fehler für die Fixpunktdarstellung von $x \neq 0$, falls sich x im darstellbaren Bereich befindet.

b) Bestimmen sie den maximalen absoluten Fehler und den maximalen relativen Fehler für die Gleitpunktdarstellung von $x \neq 0$, falls sich x im darstellbaren Bereich befindet.

Aufgabe 7: Finden Sie die größte Gleitkommazahl ε, so dass $[1 + \varepsilon] = [1]$.