Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

Fachbereich Informatik
Dr. Marco Hülsmann

Numerische Mathematik 1
Übungsblatt 2, WS 2019/20
– Musterlösungen –

Aufgabe 1 (Gleitkommazahlen nach IEEE)

Nach dem standardisierten Format IEEE-754 wird eine Zahl $x \in \mathbb{R}$ im Rechner als (gerundete) Gleitkommazahl

$$\tilde{x} = \mathit{sign}(x) \cdot a \cdot 2^E$$

dargestellt, wobei der Exponent $E = e + B$ gegeben ist durch

$$e = \lfloor \log_2(|x|) \rfloor$$

und die Verzerrung $B = 127$ für einfache und $B = 1023$ für doppelte Genauigkeit gegeben ist. Die Anzahl an Bits, die für den Exponenten verwendet werden, beträgt 8 bei einfacher und 11 bei doppelter Genauigkeit.

Die Mantisse a kann mithilfe der Formel

$$a = \left\lfloor \left(\frac{|x|}{2^e} - 1 \right) \cdot 2^{\tau_a} \right\rfloor$$

in eine natürliche Zahl konvertiert werden. Dabei ist $\tau_a = 23$ bei einfacher und $\tau_a = 52$ bei doppelter Genauigkeit die Anzahl an Bits, die für die Mantisse verwendet werden.

Schreiben Sie ein Computerprogramm, welches die Binärdarstellung einer beliebigen Zahl gemäß IEEE-754 realisiert, sowohl für einfache als auch für doppelte Genauigkeit. Speichern Sie die Bits für das Vorzeichen, den Exponenten und die Mantisse separat, und geben Sie das Ergebnis als String aus! Testen Sie Ihr Programm für die Zahl 26.9!

Musterlösung:

Für die Zahl 26.9 ergibt sich mit dem *python*-Programm `binary.py` als Exponent $e = 4$. Bei einfacher Genauigkeit ergibt sich $a = 5714739$, und die Binärdarstellung nach IEEE-754 ist gegeben durch

0 10000011 10101110011001100110011

und bei doppelter Genauigkeit gilt $a = 3068077246146150$ sowie

0 10000000011 1010111001100110011001100110011001100110011001100110

Aufgabe 2 (Kondition)

a) Die Kantenlänge eines Würfels wird mit einem Mikrometer gemessen. Die Messung ergibt (50 ± 0.01) mm. Die Ungenauigkeit in dieser Messung hat Auswirkungen auf das Volumen des Würfels, $V = x^3$. Berechnen Sie den absoluten Fehler, der sich bei der Berechnung des Volumens ergibt! Ist dieses Problem numerisch gutartig?

b) Zur Lösung der quadratischen Gleichung

$$y^2 + 2py - q = 0$$

wird das Resultat y aus den Eingabeparametern p und q mithilfe der Funktion $\varphi : \mathbb{R}^2 \to \mathbb{R}$ gemäß

$$y = \varphi(p, q) = -p + \sqrt{p^2 + q}$$

ermittelt. In welchen Fällen ist dieses Problem numerisch gut und in welchen Fällen schlecht konditioniert?

Musterlösung:

a) Sei x die Kantenlänge des Würfels. Als Schätzwert für den absoluten Fehler $|\Delta V|$ erhält man mit $V = \varphi(x) = x^3$ und dem absoluten Fehler Δx bei der Messung der Kantenlänge

$$|\Delta V| \doteq |\varphi'(x)||\Delta x| \leq 3x^2 \cdot 10^{-2}$$

Mit $x = 50$ mm ergibt sich

$$|\Delta V| \approx 75 \text{ mm}^3$$

also ein Fehler in der Größenordnung 10^1, wobei Δx in der Größenordnung 10^{-2} lag. Dieses Problem ist daher keinesfalls numerisch gutartig.

b) Es gilt

$$\frac{\partial \varphi}{\partial p} = -1 + \frac{p}{\sqrt{p^2+q}} = \frac{-y}{\sqrt{p^2+q}}$$

sowie

$$\frac{\partial \varphi}{\partial q} = \frac{1}{2\sqrt{p^2+q}}$$

Somit gilt für den relativen Fehler ε_y, da $\frac{q}{y} = p + \sqrt{p^2+q}$ (3. binom. Formel):

$$\begin{aligned} \varepsilon_y &\doteq \frac{p}{y}\frac{\partial \varphi}{\partial p}\varepsilon_p + \frac{q}{y}\frac{\partial \varphi}{\partial q}\varepsilon_q = \frac{-p}{\sqrt{p^2+q}}\varepsilon_p + \frac{q}{2y\sqrt{p^2+q}}\varepsilon_q \\ &= \frac{-p}{\sqrt{p^2+q}}\varepsilon_p + \frac{p+\sqrt{p^2+q}}{2\sqrt{p^2+q}}\varepsilon_q \end{aligned}$$

Falls $q > 0$, so gilt

$$\left|\frac{p}{\sqrt{p^2+q}}\right| \leq 1$$

sowie

$$\left|\frac{p+\sqrt{p^2+q}}{2\sqrt{p^2+q}}\right| \leq 1,$$

und das Problem ist numerisch gut konditioniert. Falls $q < 0$, so kann das Problem schlecht konditioniert werden, insbesondere wenn $q \approx -p^2$.