Algorithmische Mathematik I

Wintersemester 2017/18
Prof. Dr. Ira Neitzel
AR. Dr. Tino Ullrich

universitätbonn

Übungsblatt 3.

Abgabe am **30.10.2017 vor der Vorlesung**.

Aufgabe 1. (Gleitkommazahlen, Maschinengenauigkeit)

Gegeben sei ein Zwei-Byte-Rechner (1 Byte = 8 Bit) mit Gleitkomma-Arithmetik und interner Zahldarstellung

$$|s|m_1|m_2|m_3|\dots|m_9|e_1|e_2|e_3|\dots|e_6|,$$

mit einem Vorzeichen-Bit s, 9 Bits m_i für die Mantisse sowie 6 Bits e_j für den Exponenten (in dieser Reihenfolge).
Die Zahldarstellung ist *normalisiert* und arbeitet mit einem *hidden bit*, d.h. die 1 vor dem Komma wird nicht gespeichert. Der Exponent soll im Bereich $-31 \leq e \leq 31$ darstellbar sein, der gespeicherte Exponent ist stets positiv und wird dafür mit einem Bias von 32 versehen. Beachte den Unterschied zur Normalisierung in der Vorlesung $0.m_1m_2\dots$.
Die Rechner-Bits repräsentieren also die Zahl

$$z = (-1)^s \cdot (1.m_1m_2\dots m_9)_2 \cdot 2^{\widetilde{e}-32} \quad \text{mit} \quad \widetilde{e} = \sum_{j=1}^{6} e_j \cdot 2^{6-j}.$$

Für die Zahl 0 wird das Bitmuster $000\dots0$ verwendet, außerdem wird noch die Exponentenbitfolge 000000 zur Kennzeichnung von Sonderfällen verwendet — sie steht also nicht für die Zahldarstellung zur Verfügung.

a. Geben Sie die Rechner-Bitfolgen der Dezimalzahlen 13 und 42.125 an.

b. Wie viele Zahlen können in diesem Gleitkomma–Format dargestellt werden? Die Bitkombinationen für Sonderfälle seien zu vernachlässigen.

c. Geben Sie die *Bitfolgen* der betragsmäßig größten darstellbaren Dezimalzahl $z_{\max}$ sowie der betragsmäßig kleinsten darstellbaren Dezimalzahl $z_{\min} \neq 0$ an.

d. Definieren Sie den Begriff „Maschinengenauigkeit“ und geben Sie die Maschinengenauigkeit des Rechners an.

$(2+2+2+2 = 8$ Punkte)

Aufgabe 2. (Vermeidung von Auslöschung)

Wir haben in der Vorlesung gelernt, dass mit Auslöschung eine inakzeptable Vergrößerung des relativen Eingabefehlers bezeichnet wird. Ferner haben wir gelernt, dass die Hauptquelle für Auslöschung die Subtraktion von betragsmäßig nahezu gleich großen Zahlen ist. Schreiben Sie die folgenden Ausdrücke so um, dass für die angegebenen Argumente Auslöschung vermieden wird:

- $\sqrt[3]{1+x} - 1$ für $x \approx 0$,
- $\sin x - \sin y$ für $x \approx y$,

- $\dfrac{1-\cos x}{\sin x}$ für $x \approx 0$,
- $\dfrac{1}{x-\sqrt{x^2-1}}$ für $x \gg 1$.

(2 + 2 + 2 + 2 = 8 Punkte)

Aufgabe 3. (Rundung)

a. Zeigen Sie, daß die folgenden Ausdrücke mathematisch äquivalent sind:

- $((a+b)(a-b))^2$
- $(a^2+b^2)^2 - 4(ab)^2$
- $(a^2-b^2)^2$

b. Seien nun $a = 10^6 + 1$ und $b = 10^6 - 2$. Multiplizieren Sie damit obige Ausdrücke aus. *Jedes* Zwischenergebnis, das nicht mit 10 Dezimalstellen dargestellt werden kann, soll auf 10 Stellen gerundet werden.

c. Berechnen Sie jeweils den relativen Fehler der Resultate (2 gültige Ziffern genügen). Was ist der Grund für dieses Verhalten?

(4 Punkte)

Programmieraufgabe 1. (Maschinengenauigkeit)

In der Praxis lässt sich die Maschinengenauigkeit ε als die kleinste positive Gleitkommazahl ε ermitteln, so dass $\mathrm{rd}(1+\varepsilon) > 1$. Verwenden Sie dieses Verfahren um je ein C/C++ Programm zu schreiben, dass die Maschinengenauigkeit epsilon des `float`, `double` bzw des `long double` -Gleitkommasystems ermittelt.

Diese Programmieraufgabe ist eine Präsenzaufgabe und wird nicht bewertet.

Programmieraufgabe 2. (Approximation von Reihen)

a. Schreiben Sie ein C/C++ Programm, welches für gegebenes $n \in \mathbb{N}$ den Wert der Summe

$$s_n = \sum_{k=1}^{n} \frac{1}{k^2}$$

ausgibt. Überlegen Sie sich, wie groß n gewählt werden muss, damit alle Summanden für $k > n$ kleiner als 10^{-6} sind.

b. Der Grenzwert für $n \to \infty$ der Summe aus Aufgabenteil (a), also die Reihe $\sum_{k=1}^{\infty} 1/k^2$, ist endlich und hat den Wert $\pi^2/6$. Verändern Sie das Programm aus (a) so, dass eine Approximation an diese Reihe bestimmt wird, die alle Summanden kleiner als eine vorgegebene (eingelesene) Genauigkeit vernachlässigt. Wie groß is der Fehler im Vergleich zur Genauigkeit $\varepsilon = 10^{-6}$ und $\varepsilon = 10^{-10}$?

Diese Programmieraufgabe ist eine Präsenzaufgabe und wird nicht bewertet.