Algorithmische Mathematik

Wintersemester 2013
Prof. Dr. Marc Alexander Schweitzer und
Dr. Einar Smith
Patrick Diehl und Daniel Wissel

universitätbonn

Übungsblatt 8.

Abgabe am **16.12.2013**.

Aufgabe 1. (Rundung)

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

   (i) $((a+b)(a-b))^2$
   (ii) $(a^2-b^2)^2$
   (iii) $(a^2+b^2)^2-4(ab)^2$

b. Seien nun $a = 10^6 + 2$ sowie $b = 10^6 - 1$. Berechnen Sie nun den Wert der drei Ausdrücke. Führen Sie die Berechnung in Gleitkommadarstellung mit einer Mantissenlänge von 10 durch.

c. Berechnen Sie jeweils den relativen Fehler der Resultate (2 gültige Ziffern genügen). Was bemerken Sie?

(1 + 2 + 2 = 5 Punkte)

Aufgabe 2. (Vermeidung von Auslöschung)

Schreiben Sie die folgenden Ausdrücke so um, dass für die angegebenen Argumente Auslöschung vermieden wird:

a. $\sqrt[3]{1+x} - 1$ für $x \approx 0$

b. $\dfrac{1-\cos x}{\sin x}$ für $x \approx 0$

c. $\dfrac{1}{x - \sqrt{x^2-1}}$ für $x \gg 1$

d. $x^3\left(\dfrac{x}{x^2-1} - \dfrac{1}{x}\right)$ für $x \gg 1$

(1 + 1 + 1 + 1 = 4 Punkte)

Hinweis: Beachten Sie die Identität $\cos^2 x + \sin^2 x = 1$.

Aufgabe 3. (IEEE 754)

Die Norm IEEE 754 definiert Standarddarstellungen für binäre Gleitkommazahlen in Computern und legt genaue Verfahren für die Durchführung mathematischer Operationen fest. Wir betrachten die Darstellung einer Gleitkommazahl im Binärsystem mit einfacher Genauigkeit (single precision, 32 Bits) in der Form

$$x = v \cdot m \cdot \beta^e.$$

Dabei ist v das Vorzeichen, m die Mantisse, $\beta = 2$ sowie e der Exponent. Die 32 Bit teilen sich dabei auf als 1 Vorzeichen-Bit, 8 Exponenten-Bits, sowie 23 Mantissen-Bits: $V|E_1|E_2|\cdots|E_8|M_1|M_2|\cdots|M_{23}$.

Die Exponenten-Bits speichern die nichtnegative Binärzahl E, aus welcher sich der Exponent $e = E - B$ durch Subtraktion des Biaswertes $B = 2^7 - 1 = 127$ (dezimal) errechnet.
Die Mantissen-Bits speichern die Mantisse in normalisierter Form, die führende 1 wird nicht gespeichert, d.h. es gilt $m = 1 + M/2^{23}$.

a. Stellen Sie folgende Zahlen (in Dezimaldarstellung) als IEEE 754 Gleitkommazahl dar: 3.5, -8.0625, 0.2.

b. Berechnen Sie die Dezimaldarstellung der folgenden IEEE 754 Gleitkommazahlen: 0|10000101|11111100000000000000000, 1|10000011|10010100000000000000000, 0|11111111|00000000000000000000000.

(3 + 3 = 6 Punkte)

Aufgabe 4. (Was macht dieses Programm? (schriftlich bearbeiten!))

Gegeben seien n Gleitkommazahlen $a_1, \ldots, a_n$. Damit wird folgendes Programm ausgeführt (dabei bezeichne := die Zuweisung):

$s := a_1$
$c := 0$
for $i := 2, \ldots, n$:
    $t_1 := a_i - c$
    $t_2 := s + t_1$
    $t_3 := t_2 - s$
    $s := t_2$
    $c := t_3 - t_1$
return s

a. Was berechnet das Programm, wenn alle Operationen ohne Rundungsfehler ausgeführt werden?

b. Was berechnet das Programm, wenn bei der Operation $t_2 := s + t_1$ Rundungsfehler auftreten, alle anderen Operationen aber exakt berechnet werden?

c. In welchen Situationen könnte dieses Programm nützlich sein?

Hinweis: Mehr Informationen zur sog. *Kahan summation* findet man beispielsweise in

D. Goldberg: *What Every Computer Scientist Should Know About Floating-Point Arithmetic*, ACM Computing Surveys Vol. 23, No. 1, 1991.

(1 + 2 + 2 = 5 Punkte)

Programmieraufgabe 1. (Rundungsfehler bei der Berechnung von e)

Berechnen Sie mit Fließkommazahlen doppelter Genauigkeit (`double`) Näherungen für

$$\mathrm{e} = \lim_{n\to\infty} e_n \text{ mit } e_n := \left(1 + \frac{1}{n}\right)^n,$$

mit $n = 10^k, k = 1, 2, \ldots$ und fertigen Sie ein Diagramm $\log_{10}(|\mathrm{e} - e_n|)$ in Abängigkeit von k an. Was beobachten Sie? Der Datentyp `double` sollte etwa 15 Dezimalstellen liefern – erreichen Sie diese Genauigkeit? Berechnen Sie auch $\sqrt{\epsilon_{\text{machine}}}$. Steht dieser Wert im Zusammenhang mit Ihrer Beobachtung zur Genauigkeit?
Hinweis: Die Maschinengenauigkeit Ihres Rechners bekommen Sie über die $DBL\_\times$ Makros aus dem `float.h` Header

(6 Punkte)

Abgabe am 16.12.2013 zwischen Vorlesung A und Vorlesung B

Programmieraufgabe 2. (Zahldarstellung)

Schreiben Sie ein Programm mit drei Eingabeparametern x, b_1, b_2 (mit $b_1, b_2 \leq 10$). Dabei ist x in der Basis b_1 dargestellt und soll in der Basis b_2 ausgegeben werden.

(5 Punkte)

Abgabe am 16.12.2013 zwischen Vorlesung A und Vorlesung B

Programmieraufgabe 3. (Binärzahlen)

Modifizieren Sie Ihr Programm von Blatt 4 (Programmieraufgabe 1), so dass Sie mit Hilfe der verketteten Listen binäre Integer-Zahlen beliebiger Länge darstellen können. Implementieren Sie Funktionen zur Addition und Multiplikation dieser Zahlen. Schreiben sie außerdem eine Funktion zur Ausgabe der Binärzahl in der Dezimaldarstellung. Überprüfen Sie Ihr Programm, indem Sie einige Additionen / Multiplikationen in der Binärdarstellung durchführen und das Ergebnis in der Dezimaldarstellung verifizieren.

(9 Punkte)

Abgabe am 16.12.2013 zwischen Vorlesung A und Vorlesung B