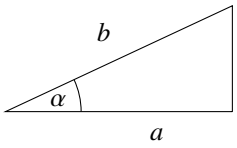
Prof. Dr. Angela Kunoth
Sandra Boschert

Algorithmische Mathematik und Programmieren
WS 2017/2018

Übungsblatt 3

Ausgabe: 26.10.2017 Abgabe: Donnerstag, 2.11.2017 bis 14:00 Uhr im Tutorium/Zentralübung
Besprechung im Tutorium am 2.11.2017 und in der Übung in der Vorlesungswoche vom 6.11. bis 10.11.2017

Aufgabe 12: (5 Punkte, 4+1)
Wir betrachten das folgende rechtwinklige Dreieck mit gegebener Ankathete a, gegebener Hypotenuse b und unbekanntem Winkel $\alpha \in [0, \pi/2)$.



a) Bestimmen Sie die absolute und relative Kondition des Problems „Finde den Winkel $\alpha \in [0, \pi/2)$“ in Abhängigkeit von a und b.

b) Berechnen Sie die absolute und relative Kondition für $a = 3.1$ und $b = 3.8$.

Aufgabe 13: (5 Punkte, 1+1+1+1+1)
Ein – zugegeben etwas primitiver – Rechner stellt reelle Zahlen im *Festkommaformat* mit einem Byte dar. Dabei werden ein Vorzeichen-Bit, vier Bits vor dem Komma und drei Bits hinter dem Komma verwendet. Somit haben die Zahlen im Dualsystem die Form

$$x = (-1)^s \sum_{i=0}^{6} d_i \cdot 2^{i-3} = s d_6 d_5 d_4 d_3 . d_2 d_1 d_0$$

mit $s, d_i \in \{0, 1\}$ für $i = 0, \ldots, 6$. Zum Beispiel hat -2.25 im Dualsystem die Form 10010.010.

a) Welche Darstellungen haben die Zahlen 7.25 und -5.625?

b) Wie viele verschiedene Zahlen können im obigen Format dargestellt werden?

c) Geben Sie die maximal und minimal darstellbare echt positive Zahl $x_{\max} > 0$ und $x_{\min} > 0$ an.

d) Eine nicht darstellbare Zahl x wird auf die nächste darstellbare Zahl $\mathrm{fl}(x) \in [-x_{\max}, x_{\max}]$ gerundet, wobei fl die im Skript in Gleichung (2.2.3) definierte Standardrundung bezeichnet. Dabei tritt ein absoluter Rundungsfehler $\varepsilon_{\mathrm{abs}}$ bzw. relativer Fehler $\varepsilon_{\mathrm{rel}}$ auf. Bestimmen Sie bei der Darstellung von

$$x = \frac{1}{3} \quad \text{die Fehler} \quad \varepsilon_{\mathrm{abs}} := |\mathrm{fl}(x) - x| \quad \text{und} \quad \varepsilon_{\mathrm{rel}} := \frac{|\mathrm{fl}(x) - x|}{|x|}.$$

e) Bestimmen Sie den maximalen absoluten und den maximalen relativen Rundungsfehler für reelle Zahlen im Bereich $[x_{\min}, x_{\max}]$.

Aufgabe 14: (5 Punkte, 2+1+2)
Wir betrachten einen Rechner mit *Gleitkommaarithmetik*, der einen deutlich geringeren maximalen Rundungsfehler aufweisen soll als ein Rechner mit Festkommaarithmetik (vgl. Aufgabe 13). *Gleitkommazahlen* haben die Form $x = \pm M \cdot b^e$ mit *Mantisse* M der Länge m, *Exponent* e und *Basis* b. Zum Beispiel kann die Zahl $x = 0.000031$ in der Basis $b = 10$ dargestellt werden als

$$x = 0.000031 = 0.31 \cdot 10^{-4} = 3.1 \cdot 10^{-5} = M \cdot 10^e.$$

Die durch führende Nullen entstehenden Mehrdeutigkeiten sind unerwünscht, daher einigen wir uns auf die *normalisierte* Darstellung

$$M = d_1.d_2d_3\ldots d_m \text{ mit } d_1 \neq 0.$$

Nur für $x = 0$ erlauben wir, dass alle Ziffern $d_i = 0$ sind. Wir stellen nun die Mantisse in Binärdarstellung $b = 2$ dar. Außerdem verwenden wir $m = 3$, also 2 Ziffern für die Nachkommastellen. Damit ist unsere *Rechner-Gleitkommadarstellung* gegeben durch

$$x = \pm d_1.d_2d_3 \cdot 2^e \quad \text{mit dem Spezialfall} \quad \pm 0 = \pm 0.00 \cdot 2^0.$$

a) Geben Sie für $e \in \{0, 1\}$ alle darstellbaren, nichtnegativen Gleitkommazahlen an.

b) Sie werden feststellen, dass die *Abstände* zwischen aufeinanderfolgenden Gleitkommazahlen stark variieren. Markieren Sie alle darstellbaren nichtnegativen Zahlen auf einem Zahlenstrahl.

c) Zur Darstellung einer beliebigen reellen Zahl x verwenden wir wieder die nächstgelegene Gleitkommazahl. Geben Sie den absoluten und den relativen Rundungsfehler bei der Darstellung der Zahlen $x = \frac{8}{5}$ und $x = \frac{9}{16}$ an.

Aufgabe 15: (5 Punkte,1+2+2)
Sei $f \in C^2(\mathbb{R}, \mathbb{R}^+)$ positiv und zweimal stetig differenzierbar. Um die Ableitung f' von f in einem Punkt $x_0 \in \mathbb{R}$ zu approximieren, betrachten wir den Differenzenquotienten

$$\Delta^h f(x_0) := \frac{f(x_0 + h) - f(x_0)}{h}.$$

Die Ableitung $f'(x_0)$ kann durch $\Delta^h f(x_0)$ beliebig genau approximiert werden.

a) Beweisen Sie, dass für $h \to 0$ gilt: $\Delta^h f(x_0) = f'(x_0) + O(h)$

Hinweis: Taylorentwicklung

b) Wegen der begrenzten Maschinengenauigkeit liegen die Werte für $f(x_0 + h)$ und $f(x_0)$ nur gerundet vor. Wir bezeichnen diese gerundeten Werte mit

$$f_1 := \mathrm{fl}(f(x_0 + h)) \quad \text{und} \quad f_2 := \mathrm{fl}(f(x_0)),$$

wobei fl die im Skript in Gleichung (2.2.3) definierte Standardrundung bezeichnet. Bestimmen Sie diese nicht mit Hilfe von Gleichung (2.2.3), sondern nehmen Sie

$$\left|f_1 - f(x_0 + h)\right| \leq f(x_0)\varepsilon \text{ und } \left|f_2 - f(x_0)\right| \leq f(x_0)\varepsilon \quad \text{für ein } \varepsilon > 0$$

an und folgern Sie daraus die Fehlerabschätzung

$$\left|\frac{f_1 - f_2}{h} - f'(x_0)\right| \leq c\frac{\varepsilon}{|h|} + \tilde{c}|h| \quad \text{für Konstanten } c, \tilde{c} \in \mathbb{R}^+.$$

Wie verhält sich diese obere Schranke asymptotisch für $h \to 0$?

c) Begründen Sie, weshalb die obige Aussage *„Die Ableitung $f'(x_0)$ kann durch $\Delta^h f(x_0)$ beliebig genau approximiert werden“* aus numerischer Sicht falsch ist und finden Sie ein $h \in \mathbb{R}$, sodass die Fehlerabschätzung aus Aufgabenteil b) so scharf wie möglich ist.