WS23/24: Numerische Mathematik
Übungszettel 2

1. Gegeben ein Gleitkommasystem mit darstellbaren Zahlen $\mathcal{F}$ (inklusive 0), in dem die Addition $+_*$ und Multiplikation $\cdot_*$ definiert sind durch $z_1 +_* z_1 = \mathsf{fp}(z_1 + z_2)$ und $z_1 \cdot_* z_1 = \mathsf{fp}(z_1 \cdot z_2)$, wobei $\mathsf{fp} : \mathbb{R} \mapsto \mathcal{F}$ die übliche Rundungsregel ist. Diskutieren Sie, welche Körpereigenschaften $(\mathcal{F}, +_*, \cdot_*)$ noch bzw. nicht hat. Geben Sie gegebenfalls Gegenbeispiele in einem Gleitkommasystem Ihrer Wahl.

2. Sie wollen $y = (a + b) \cdot c$ in einem Gleitkommasystem mit $\beta = 10$ und $t = 5$ berechnen (Exponenten beliebig klein bzw. groß).

   (a) Vergleichen Sie folgende Algorithmen bzgl Ihrer Fehlerverstärkung:
   A1: Schritt 1: Berechne $\eta = a + b$, Schritt 2: Berechne $y_1 = \eta \cdot c$.
   A2: Schritt 1: Berechne $\eta_1 = a \cdot c$, Schritt 2: Berechne $\eta_2 = b \cdot c$, Schritt 3: Berechne $y_2 = \eta_1 + \eta_2$
   Wie hängen die relativen Fehler im Output von den (unvermeidlichen) Rundungsfehlern in jedem Schritt ab?

   (b) Wenden Sie Ihre Analyse auf folgende Zahlen an: $a = 0.1$, $b = -0.099992$ und $c = 90421$. Ermitteln Sie die relativen Fehler in jedem Schritt und im Gesamtergebnis und erklären Sie Ihre Beobachtungen.

   (c) (P) Schreiben Sie ein Programm, welches Ihre Rechnung in (b) automatisiert: Verwenden Sie die Werte von a und c wie in (b), aber verwenden Sie für b viele (zB 1000) Werte zwischen -0.5 und 0.5, jeweils auf 5 signifikante Stellen gerundet. Ihr Programm soll dann:

   - y_1 und y_2 wie in (a) berechnen (mit Runden in jedem Schritt).
   - Die relativen Fehler im Ergebnis ermitteln und als Funktion von b plotten.
   - Plotten Sie außerdem die theoretischen Fehlerschranken aus (a) und diskutieren Sie das Verhalten.

   Anmerkungen 1: Um eine Zahl x auf n signifikante Stellen zu Runden können Sie folgenden Code verwenden:
   `round(x, -int(math.floor(math.log10(abs(x)))) + (n - 1))`
   Anmerkungen 2: Hier nehmen wir an, dass alle anderen Rundungsfehler, die innerhalb des Binärsystems mit double precision entstehen vernachlässigbar sind.

3. For a differentiable, scalar function $f : I \subset \mathbb{R} \mapsto \mathbb{R}$ and two inputs x and $x + \Delta x$ with Δx (very) small, we want to understand the relationship between the relative error in input and the relative error in output.

   (a) The *relative condition number* of f at x is defined as

   $$c_{\mathsf{f}} = \left| \frac{x f'(x)}{f(x)} \right|.$$

   Find a justification why this number is a good measure of how the relative error in input relates to the relative error in output.

(b) Determine the relative condition numbers of $f_1(x) = e^x$ and $f_2(x) = \sin(x)$ and discuss your findings.

4. Finden Sie je ein Beispiel (Gleichung und Skizze) für reelle, stetige Funktionen definiert auf $[0, 1]$ mit (i) keiner, (ii) genau einer, (iii) genau zwei und (iv) genau drei Nullstellen in $[0, 1]$. Diskutieren Sie, welche der Funktionen die Voraussetzungen vom Thm 1.1. erfüllen und was man daraus lernt.

5. Es sei $f(x) = \frac{1}{x} - x^2$ definiert auf $I = [\frac{1}{2}, 2]$.

   (i) Beweisen Sie, dass f eine Nullstelle in I besitzt.

   (ii) Basierend auf f, finden Sie zwei Funktionen $g_1(x)$ und $g_2(x)$ definiert auf I, deren jeweiliger Fixpunkt den Nullstellen von f entspricht. Wählen Sie g_1 und g_2 so, dass nur eine die Voraussetzungen des Brouwer'schen Fixpunktsatzes erfüllt.

   (iii) Skizzieren Sie alle drei Funktionen.

6. We attempt to find all solutions to $f(x) = 0$, where $f(x) = e^x - 3x - 1$.

   (i) Sketch $y = f(x)$ for $-1 \leq x \leq 3$. How many solutions does $f(x) = 0$ have?

   (ii) We now look at the fixed point problem $x = g(x)$ with $g(x) = \ln(3x + 1)$. Show that this is equivalent to finding the roots of f.

   (iii) Plot or sketch $y = g(x)$ and $y = x$ in one plot and find (graphically) two starting values that give a different series behaviour.