Übungsblatt 1

Aufgabe 1.1

Wir haben ein Gleitkomma-System $\mathbb{F}$ mit 4 Bits für den Exponenten e und $M = 12$ (signifikanten, d.h. 11 eigentlich abgespeicherten) Bits für die Mantisse gegeben.

(a) Was ist der größtmögliche relative Rundungsfehler?
(b) Was ist die kleinste positive (normalisierte) Zahl in $\mathbb{F}$?
(c) Was ist die größte Zahl in $\mathbb{F}$?
(d) Was ist (ungefähr) der größtmögliche absolute Rundungsfehler?
(e) Wieviele Zahlen lassen sich in $\mathbb{F}$ (theoretisch, d.h. ohne Platz für Sonderzahlen wie $\pm\infty$ und `NaN`) darstellen? (In anderen Worten: Was ist $|\mathbb{F}|$?)

Aufgabe 1.2

Zur allgemeinen Approximation der Ableitung einer differenzierbaren Funktion f an der Stelle x_0 wird oft der (rechtsseitige) *Differenzenquotient*

$$\widetilde{f}'_r(x_0, h) := \frac{f(x_0 + h) - f(x_0)}{h} \approx f'(x_0)$$

für ein bestimmtes $h > 0$ verwendet.

Es sei nun $f(x) = x^2$ und $x_0 = 1$. Es gilt bekannterweise $f'(x_0) = 2$.

Plotte den Abfall des Fehlers $|\widetilde{f}'_r(x_0, h) - f'(x_0)|$ für kleiner werdendes $h > 0$ auf logarithmischen Achsen.

Finde eine plausible Erklärung für das Ergebnis.

(Mehr zur numerischen Ableitung gibt es in Kap. 8.)

Aufgabe 1.3

Wir möchten die quadratische Gleichung $x^2 + 2px + q = 0$ numerisch, d.h. auf einem Rechner, lösen. Die aus der Schule bekannte Mitternachtsformel liefert die zwei analytischen Lösungen

$$x_{1/2} = -p \pm \sqrt{p^2 - q}.$$

(a) Welche der beiden Lösungen ist für welche Werte von p und q aus numerischer Sicht von Auslöschung betroffen?

(b) Wie kann man mit dem Satz von Vieta einen Ausweg für den Problemfall der Auslöschung finden?

Hinweis: Der Satz von Vieta besagt, dass die zwei Lösungen folgende Gleichungen erfüllen:

$$x_1 + x_2 = -p$$
$$x_1 \cdot x_2 = q$$

(c) Implementiere die Mitternachtsformel und die Formel aus dem Satz von Vieta. Weise mit mehreren Werten für p und q den Unterschied der beiden Ansätze nach.