Technische Universität Berlin **Wintersemester 2015/2016**
Fakultät II – Mathematik und Naturwissenschaften
Institut für Mathematik
Reinhold Schneider, Benjamin Kutschan

Numerische Mathematik I

1. Übungsblatt: Rechnerarithmetik und Rundungsfehler

Hausaufgaben: (Abgabe 28. Oktober 14:10 - 14:15 in Raum MA 042)

Aufgabe 1: **(10 Punkte)**

Es gelte $x \oplus y = (x+y)(1+\varepsilon)$ mit $|\varepsilon| \leq \varepsilon^*$ für alle Maschinenzahlen x und y. Seien nun Maschinenzahlen $x_1, \ldots, x_n$ gegeben und sei durch

$$\tilde{s}_1 = x_1, \quad \tilde{s}_n = \tilde{s}_{n-1} \oplus x_n$$

ein Algorithmus zur Berechnung der Summe $s_n = \sum_{i=1}^{n} x_i$ definiert.

(a) Berechnen Sie für $n = 1, ..., 4$ den absoluten Fehler $f_n := \tilde{s}_n - s_n$. Was fällt auf? Wie lässt sich dieser Fehler durch einen schönen von n abhängigen Term abschätzen? Sieht dieser so aus, wie in Aufgabe (b)?

(b) Zeigen Sie: zerlegt man $\tilde{s}_n = s_n + f_n$, so gilt für den absoluten Fehler

$$|f_n| \leq [(1+\varepsilon^*)^n - 1] \sum_{i=1}^{n} |x_i|.$$

Diskutieren Sie den relativen Fehler von $\tilde{s}_n$.

(c) In dieser Aufgabe werden wir sehen, dass die berechnete Lösung gleich der exakten Lösung von leicht gestörten Eingabedaten $x_i \cdot (1+\delta_i)$ ist. Zeigen Sie, dass

$$\tilde{s}_n = \sum_{i=1}^{n} x_i(1+\delta_i) \quad \text{mit} \quad (1-\varepsilon^*)^n - 1 \leq \delta_i \leq (1+\varepsilon^*)^n - 1$$

und falls $n\varepsilon^* < 1$ gilt:

$$|\delta_i| \leq \frac{n\varepsilon^*}{1 - n\varepsilon^*}, \quad i = 1, \ldots, n.$$

Aufgabe 2: **(3 Punkte)**

Man betrachte die folgende Tabelle, wobei *eps* die kleinste Gleitkommazahl z ist, für die $1 \oplus z \neq 1$.

	a)	b)	
1.)	$(1+x)^2 - 1$	$x^2 + 2x$	$\lvert x \rvert \approx 10 \cdot eps$
2.)	$\frac{\ln(1+x)}{x}$	$1 - \frac{x}{2} + \frac{x^2}{3}$	$eps < x \leq 10 \cdot eps$

Bei welchen Ausdrücken tritt Auslöschung auf? Auslöschung heißt, dass durch die Subtraktion zweier fast gleich großer Ausdrücke die Präzision des Resultates viel geringer ist, als die Mantissenlänge. Berechnen Sie für alle Ausdrücke den Vorwärtsfehler.

Aufgabe 3: **(5 Punkte)**

Zu lösen sei die quadratische Gleichung

$$x^2 - 2px - q = 0 \quad \text{für} \quad p = 2,\ q = 0.0005$$

in vier- und fünfstelliger Gleitkommaarithmetik im Dezimalsystem. Dabei sollen die folgenden Algorithmen untersucht werden:

(i) $d = p^2 + q$, $x_1 = p + \sqrt{d}$, $x_2 = p - \sqrt{d}$, ($p - q$ Formel)

(ii) $d = p^2 + q$, $x_1 = p + \sqrt{d}$, $x_2 = -q/x_1$. (Vietascher Wurzelsatz)

Vergleichen Sie die Ergebnisse mit der exakten Lösung und erklären Sie die unterschiedlichen Resultate.

Aufgabe 4: **(2 Punkte)**

Bei exakter Arithmetik gilt für das arithmetische Mittel $\bar{x} = \frac{x_1+x_2}{2}$ zweier reeller Zahlen x_1 und x_2 die Ungleichung

$$\min\{x_1, x_2\} \leq \bar{x} \leq \max\{x_1, x_2\}. \tag{1}$$

Für Gleitpunktzahlen und das gemäß $\bar{x} = \mathrm{rd}(\mathrm{rd}(x_1 + x_2)/2)$ berechnete Mittel gilt die Ungleichung (1) im Allgemeinen nicht.

Finde Gleitpunktzahlen $x_1, x_2 \in \mathcal{F}(10, 2, -3, 3)$ derart, dass (1) verletzt ist. Wie ist die Berechnungsvorschrift für das arithmetische Mittel zu modifizieren, damit die Ungleichung (1) auch für das berechnete Mittel gilt?

Programmieraufgabe 1: (Abgabe in den Rechnersprechstunden bis zum 30. Oktober 2015)

Schreiben Sie Matlab-Programme

```
e=meps(),    m=minimum(),    m=maximum()
```

die das Maschienenepsilon `eps` (siehe Aufgabe 2), die kleinste darstellbare positive Zahl x_{min} bzw. die grösste darstellbare Zahl x_{max} berechnen. Dabei ist `eps` als die kleinste Zahl definiert für die $1 \oplus \texttt{eps} > 1$ gilt.

In den Programmen soll nur benutzt werden, dass intern eine Gleitkommadarstellung basierend auf dem Dualsystem verwendet wird. Vergleichen Sie ihre Ergebnisse mit denen der entsprechenden MATLAB Funktionen (`eps,realmin,realmax`). Interpretieren Sie die Unterschiede.