*Richard Huber*
*richard.huber@uni-graz.at*

***Numerische Mathematik für LAK***
*5.Übungsblatt – 5ten November, 2019*

## 5. Aufgabenblatt

**Definition** (Fehlerfortpflanzung und differenzielle Fehleranalyse)**.** *Es sei* $f(x) = \varphi_r \circ \cdots \circ \varphi_0(x)$ *für eine Funktion* $f\colon Dom(f) \subset \mathbb{R}^m \to \mathbb{R}^n$ *und* $\varphi_i\colon Dom(\varphi_i) \subset \mathbb{R}^{m_i} \to \mathbb{R}^{m_{i+1}}$, $x^{(i)} = \varphi_{i-1} \circ \cdots \circ \varphi_0(x) \in \mathbb{R}^{m_i}$ *bezeichne die Zwischenergebnisse und* $\psi_i = \varphi_r \circ \cdots \circ \varphi_i$ *die auf* $x^{(i)}$ *noch anzuwenden Operationen für* $i = 1, \dots, r$*. Dann kann man mittels differenzieller Analyse den Fehler (der durch Datenfehler, Rundungsfehler und deren Fortpflanzung) entsteht wie folgt darstellen:*

$$\delta_{f(x)} = D[f](x)\delta_x + \sum_{i=1}^{r} D[\psi_i](x^{(i)})E_i x^{(i)} + E_{r+1}f(x) + \text{ Terme höherer Ordnung}, \qquad (1)$$

*wobei die Rundungsfehler-Faktoren* $E_i \in \mathbb{R}^{m_i \times m_i}$ *Diagonalmatrizen sind deren Einträge Betrag kleiner als* $\tau$ *haben. Man beachte, dass* $\delta_{f(x)}$ *implizit von* $\delta_x$ *und* $E = (E_i)_{i=1,\dots,r+1}$ *abhängt, wir schreiben auch* $\delta_{f(x)}(\delta_x, E)$ *wenn wir die Abhängigkeit von* $\delta_x$ *und* $E$ *hervorheben wollen. Es bezeichne* $\overline{\delta_{f(x)}}$ *die Linearisierung des Fehlers, d.h. das Ergebnis der differenziellen Analyse* (1) *unter Vernachlässigung der Terme höherer Ordnung, und* $\overline{\varepsilon_{f(x)}} = \overline{\delta_{f(x)}}/f(x)$ *(diese Division ist im Fall* $n > 1$ *komponentenweise zu verstehen).*

**Aufgabe 1** (Differenzielle Analyse der Fehlerfortpflanzung)**.** *Gegeben sei die Funktion* $f : (-1, 1) \to \mathbb{R}$, $x \mapsto 1 - \sqrt{1 - x^2}$*. Es sei* $\tau > 0$ *die Rechengenauigkeit und* $|\delta_x| \leq \tau|x|$*. Im Folgenden wollen wir das Verhalten der Funktion* $f$ *unter Fehlerfortpflanzung für* $x \to 0$ *untersuchen.*

***a)*** *Berechnen Sie die Konditionszahl* $K(x) = \frac{\partial f}{\partial x}(x)\frac{x}{f(x)}$ *für* $x \in (-1, 1)$ *und zeigen Sie dass* $\lim_{x\to 0} K(x) = 2$*.*

***b)*** *Zerlegen Sie* $f$ *in Funktionen* $\varphi_r, \dots, \varphi_0$ *sodass* $f = \varphi_r \circ \cdots \circ \varphi_0$ *und* $\varphi_i(x)$ *in jeder Komponente aus nur einer elementaren Operation* $(+, -, \cdot, \div, \sqrt{\ })$ *besteht. Bestimmen Sie des Weiteren* $\psi_i$ *und* $D[\psi_i](x^{(i)})$ *für* $i = 1, \dots, r$*.*

***c)*** *Bestimmen Sie* $x^{(i)}(x)$ *und* $D[\psi_i](x^{(i)}(x))$ *für* $i = 1, \dots, r$*. Bestimmen Sie des Weiteren* $\overline{\delta_{f(x)}}$ *und* $\overline{\varepsilon_{f(x)}}$ *als Funktion von* $E_i$ *und* $\delta_x$*.*

***d)*** *Zeigen Sie, dass es Rundungsfaktoren* $E_i$ *und eine Konstante* $c > 0$ *gibt, sodass* $\overline{\delta_{f(x)}}(\delta_x, E) \geq c > 0$ *für* $x \to 0$ *und* $|\delta_x| \leq \tau|x|$ *beliebig. Folgern Sie damit* $\overline{\varepsilon_{f(x)}} \to \infty$*.*

***e)*** *Zeigen Sie, dass mit* $g(x) = x^2/(1 + \sqrt{1 - x^2})$ *die äquivalente Formulierung* $g(x) = f(x)$ *gilt und berechnen Sie des Weiteren* $\overline{\varepsilon_{g(x)}}$ *(analog zu b) und c)).*

***f)*** *Zeigen Sie, dass für* $x \to 0$ *der Fehler* $\overline{\varepsilon_{g(x)}}$ *– unabhängig von* $\delta_x$ *und* $E$ *– beschränkt ist.*

**Hinweis.** *Für a) kann de l' Hospital Regel genutzt werden. Für b) überlegen Sie sich wie Sie von Hand die Funktion ausrechnen würden, und jeder Grundrechenschritt hängt mit einem der* $\varphi_i$ *zusammen, wobei Sie mehrere Rechnungen vom selben Input vektorwertig schreiben können (siehe auch die Zerlegung* (2) *in Aufgabe 2). Für d) beachten Sie, dass es Terme in* $\overline{\delta_{f(x)}}$ *gibt, die für* $x \to 0$ *nicht verschwinden, und* $E$ *kann so gewählt werden dass diese nach unten beschränkt sind. Für f) nutzen Sie Dreiecksungleichungen für* $|\overline{\varepsilon_{g(x)}}|$ *und beachten dass alle Terme abhängig von* $x$ *beschränkt bleiben,* $E$ *beschränkt ist und* $|\delta_x| \leq \tau|x|$*.*

*Richard Huber*
*richard.huber@uni-graz.at*



**Bemerkung.** *Dieses Beispiel zeigt, dass die Operation $f$ in der Nähe von 0 stabil ist, wenn man jedoch interne Rundungsfehler berücksichtigt die Berechnung potentiell instabil wird. Punkte d) e) und f) zeigen des Weiteren, dass zwei unterschiedliche Vorgangsweisen zur Berechnung einer Operation unterschiedliche Stabilitätseigenschaften besitzen können.*

**Aufgabe 2** (Rechengenauigkeit des Heronschen Verfahrens)**.** *Wir betrachten das Heronsche Verfahren zur approximativen Berechnung der Quadratwurzel einer positiven Zahl $a > 1$. Dieses iterative Verfahren startet mit $x_0 > \sqrt{a}$ und folgt der Iterationsvorschrift $x_{k+1} = \frac{x_k^2+a}{2x_k} = g(x_k)$, wobei Sie verwenden können, dass dieses Verfahren stets gegen $\sqrt{a} > 0$ konvergiert, $g\colon [\sqrt{a}, \infty) \to [\sqrt{a}, \infty)$ streng monoton fallend und $g(x) \leq x$ ist und daher die Folge $(x_n)_n$ beschränkt und streng monoton fallend ist. Im Folgenden bezeichne $x_k$ die exakte Iterierte und $\widetilde{x_k}$ die Berechnung von $x_k$ via Iteration von $x_0$ startend mit allen dabei auftretenden (linearisierten) Rundungsfehlern und Fehlerfortpflanzung:*

$$\widetilde{x_0} = x_0, \qquad \widetilde{x_{k+1}} = g(x_k) + \overline{\delta_{g(x_k)}}(\delta_{x_k}, E^k), \qquad \text{mit } g = \varphi_2 \circ \varphi_1 \circ \varphi_0 \tag{2}$$
$$\varphi_0(x) = \begin{pmatrix} 2x \\ x^2 \end{pmatrix}, \qquad \varphi_1 \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 + a \end{pmatrix}, \qquad \varphi_2 \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \frac{z_2}{z_1}$$

*wobei es eine zugehörige Folge von Rundungsfaktoren $E^k = (E_i^k)_{i=1,\dots,r+1}$ gibt und $\delta_{x_k} = \widetilde{x_k} - x_k$. Wir betrachten das Verfahren in Anbetracht der Fehler in abgewandelter Form, in der das Verfahren abbricht sobald $\widetilde{x_k}^2 < a$. Des Weiteren nehmen wir an, dass $a$ exakt gegeben ist und $\tau x_0 \ll \sqrt{a}$.*

***a)*** *Stellen Sie $\delta_{x_{k+1}}$ als einer Funktion von $\delta_{x_k}$ und $(E_i^k)_{i=1,\dots,r+1}$ dar.*

***b)*** *Nutzen Sie a), um für $x_k > \sqrt{a}$ zu zeigen, dass*

$$|\delta_{x_{k+1}}| \leq \frac{1}{2}|\delta_{x_k}| + 5\tau x_k. \tag{3}$$

*Nutzen diese Abschätzung, und vollständige Induktion, um zu zeigen dass*

$$|\delta_{x_{k+1}}| \leq 10\tau x_0. \tag{4}$$

***c)*** *Zeigen Sie, im Fall $\widetilde{x_{k+1}}^2 < a$ für ein $k > 0$ (d.h. der Algorithmus terminiert), dass $\sqrt{a} \in [\widetilde{x_{k+1}}, \widetilde{x_{k+1}} + 10\tau\widetilde{x_0}]$ gilt.*

***d)*** *Zeigen Sie im Fall $\widetilde{x_k}^2 > a$ für alle $k > 0$, dass für jedes $\varepsilon > 0$ ein $K > 0$ existiert sodass für $k > K$ die Abschätzung $\sqrt{a} \in [\widetilde{x_k} - 10\tau x_0 - \varepsilon, \widetilde{x_k}]$ gilt.*

**Hinweis.** *Für b) nutzen Sie zunächst die Dreiecksungleichung sowie $|x_k^2 + a| \leq 2x_k^2$, $|1 - \frac{a}{x_k^2}| < 1$ und $|E_i x| \leq \tau|x|$ (komponentenweise) um (3) zu zeigen. Für (4) betrachten Sie zunächst $|\delta_{x_{k+1}}|$ gemäß (3) und setzen dort die Abschätzung für $\delta_{x_k}$ ein. Iteratives anwenden dieses Prozesses liefert die Abschätzung da $\sum_{k=0}^{\infty} 2^{-k} = 2$. Für c) nutzen Sie die Lage von $x_k$, $\widetilde{x_k}$ und $\sqrt{a}$ sowie (4). Für d) nutzen Sie, dass $x_k$ gegen $\sqrt{a}$ konvergiert sowie Abschätzung (4).*

**Bemerkung.** *Man beachte, dass $\widetilde{x_{k+1}} \approx g(\widetilde{x_k})$ (abgesehen von Termen höherer Ordnung) und damit grob der Iteration mit Fehlerfortpflanzung entspricht. Diese Aufgabe zeigt damit, dass die Fehleranalyse einer Operation genutzt werden kann, um die Genauigkeit eines iterativen Algorithmus zu bestimmen. So konnte festgestellt werden, dass das Resultat des Algorithmus trotz der Fehler korrekt ist (abgesehen von Fehlern deren Größe beschränkt ist).*