Kapitel 1

Grundlagen der Fehleranalyse

B. 1. 1. (***Fehlerarten***)**:**
Von den Fehlerquellen ausgehend unterscheiden wir drei unterschiedliche Fehlerarten:

(1) *Eingangsfehler*
Diese entstehen durch

a) *Modellierungsfehler*
(Z.B. wenn eine komplizierte Funktion durch eine einfachere ersetzt wird bzw. durch Diskretisierung eines stetigen Modells).

b) *Datenfehler*
Eingangsdaten werden aus physikalischen Messungen, statistischen Erhebungen und Prognosen gewonnen. Sie sind daher zwangsläufig fehlerbehaftet.
(Hierzu werden auch gelegentlich „menschliche" bzw. „maschinelle" Fehler gezählt.)

(2) *Verfahrensfehler*

Diese entstehen durch

a) *Finitisierung* infiniter Prozesse:
Numerische Verfahren können prinzipiell Grenzprozesse nicht nachvollziehen. Daher muss ein Iterationsverfahren nach endlich vielen Schritten abgebrochen werden. Dies hat *Abbruchfehler* (*truncation error*) zur Folge.

b) *Diskretisierung* stetiger Prozesse. Dies hat Dirkretisieungsfehler zur Folge.
(Beispiele: Interpolation, Differentiation, Integration).

(3) *Rundungsfehler*

Bei der Ausführung von Rechenoperationen können Fehler entstehen, da man immer nur in einem begrenzten Zahlenbereich arbeitet. Man behilft sich, indem man Resultate rundet. Eine mögliche Akkumulation solcher Rundungsfehler kann zu einer vollständigen Verfälschung des Endresultats führen.
Vielfach werden die schlimmen Folgen eines sorglosen Umgangs mit numerischem Fehler, speziell Rundungsfehler, unterschätzt. Ein spektakuläres Beispiel hierfür war das Desaster mit Patriot Missile in Dharan, Saudi Arabien, am 25. Februar 1991, bei dem 28 Personen ums Leben kamen. Nachuntersuchungen zeigten, dass die Katastrophe durch unzureichende Behandlung des Rundungsfehlers in der Raketensoftware verursacht wurde.

B. 1. 2.
Alle Fehlerarten, vor allem Rundungsfehler und Verfahrensfehler, sind im Allgemeinen miteinander verflochten; sie überlagern sich in komplizierter Weise und könne bezüglich ihres Einflusses auf das Resultat oft sehr schwer oder gar nicht getrennt werden. Die

Gewinnung quantitativer Sensitivitätsaussagen ist deshalb ebenfalls kompliziert. Meist werden ohne Berücksichtigung der Fehlerüberlagerung Abschätzung für nur eine Fehlerart angeboten.
Im Weiteren wird der gleiche Zugang gewählt.

<u>D. 1. 1.</u> (*Absoluter und relativer Fehler*):

Es seien $x, \tilde{x} \in \mathbb{R}^1$, wobei $\tilde{x}$ eine Näherung für x sein soll.
Sei

$$\delta_x := | x - \tilde{x} |$$

als *absoluter Fehler* und

$$\varepsilon_x := \frac{| x - \tilde{x} |}{| x |}, \quad x \neq 0$$

als *relativer Fehler* genannt.

<u>B. 1. 3.</u>
Zur vergleichenden Beurteilung von verschiedenen Messungen eignet sich der absolute Fehler nicht. Werden z.B. eine Länge $x = 10\ cm$ mit dem absoluten Fehler $|\delta_x| = 1\ mm$ ermittelt und eine Länge $y = 10\ m$ mit demselben absoluten Fehler $|\delta_y| = 1\ mm$, so ist die Qualität der Messung von y offenbar besser als die der Messung von x. Der relative Fehler wurde eingeführt, um die Qualität von Näherungswerten vergleichen zu können.
Im Fall $|x| \approx 1$ gibt es offensichtlich keinen großen Unterschied zwischen dem absoluten und dem relativen Fehler. Dagegen wird diese Differenz im Fall $|x| \gg 1$ ziemlich groß sein.

<u>B. 1. 4</u>
Es gilt

$$\tilde{x} = (1 + \varepsilon_x)\, x \,. \tag{1. 1.}$$

Der Faktor $1 + \varepsilon_x$ wird als *Modifikator* genannt.

<u>D. 1. 2.</u> (*Maximaler absoluter und relativer Fehler*):
Sei

$$|\delta_x| \leq \Delta_x, \qquad |\varepsilon_x| \leq \mathrm{E}_x\,,$$

wobei

$\Delta_x > 0$: eine möglichst kleine obere Schranke für $|\delta_x|$

$\mathrm{E}_x > 0$: eine möglichst kleine obere Schranke für $|\varepsilon_x|$

sind.
Dann heißt Δ_x *maximaler absoluter Fehler* (oder *absoluter Höchstfehler*) von $\bar{x}$ und E_x *maximaler relativer Fehler* (oder *relativer Höchstfehler*) von $\bar{x}$.

<u>B. 1. 5.</u> (*Zur Abschätzung des Eingangsfehlers*):
Hierzu gibt es folgende Möglichkeiten:

1) Anwendung der Analysis (insbesondere der Mittelwertsätze der Differentialrechnung;

2) Anwendung der Theorie der unscharfen Mengen („fuzzy sets“);

3) Anwendung der Intervallmathematik.

<u>B. 1. 6.</u> *(Fehler bei der Ermittlung des Wertes einer Funktion einer Veränderlicher mit Hilfe der Analysis):*
Sei

$x \in \mathbb{R}^1$: ein „exakter“ Wert,

$\bar{x} \in \mathbb{R}^1$: eine Näherung von x.

Bekannt seien $\bar{x}$, Δ_x und E_x.
Soll ein Funktionswert $y = f(x)$ einer (der Einfachheit halber als stetig differenzierbar vorausgesetzten) Funktion einer reellen Veränderlichen bestimmt werden, so wird das Ergebnis durch zwei Faktoren verfälscht:

(1) durch das fehlerhafte Argument $\bar{x}$
(2) durch eine genährte Funktionsberechnung.

Man erhält

$$\bar{y} = \bar{f}(\bar{x}) \text{ statt } y = f(x).$$

Um die Genauigkeit von $\bar{y}$ zu beurteilen, kann man zwei in gewisser Weise entgegengesetzte Wege bestreiten:

<u>1. Weg</u>: (*Rückwärtsanalyse* (*Inverse Aufgabe der Fehlertheorie*)):

Bei der Rückwärtsanalyse des Fehlers interpretiert man den Näherungswert $\bar{y}$ als *exakten* Funktionswert zu einem *gestörten* Argument:

$$\bar{f}(\bar{x}) := f(\bar{x} + \delta) \tag{1. 2.}$$

und versucht, die Störung δ abzuschätzen.
Ist $|\delta| \le \Delta_x$, so liegt $\bar{x} + \delta$ in demselben Intervall $\left[\bar{x} - \Delta_x, \bar{x} + \Delta_x\right]$ wie das unbekannte Argument x. Folglich ist $\bar{y}$ eine <u>akzeptable</u> Näherung von y.

Ist dagegen $|\delta| > \Delta_x$ (oder sogar $|\delta| >> \Delta_x$), so muss $\bar{f}$ durch eine bessere Näherungsfunktion ersetzt werden, wenn ein Genauigkeitsverlust vermieden werden soll.

<u>2. Weg</u>: (*Vorwärtsanalyse (Direkte Aufgabe der Fehlertheorie))*:
Bei der Vorwärtsanalyse des Fehlers schätzt man den *absoluten Gesamtfehler:*

$$\delta_y := \bar{f}(\bar{x}) - f(x) \tag{1. 3.}$$

ab. Analog zu (1. 1.) fasst man die genährte Funktionsberechnung $\bar{f}(\bar{x})$ als eine modifizierte exakte Funktionsberechnung $f(\bar{x})$ auf:

$$\bar{f}(\bar{x}) := \left(1 + \varepsilon_f\right) f(\bar{x}), \qquad \left|\varepsilon_f\right| < \mathrm{E}_f\,, \tag{1. 4.}$$

wobei ε_f von der gewählten Näherungsfunktion $\bar{f}$ und dem Argument $\bar{x}$ abhängt.
Damit folgt:

$$\delta_y := \bar{f}(\bar{x}) - f(x)$$

$$= \varepsilon_f\, f(\bar{x}) + f(\bar{x}) - f(x)\,. \qquad (\because\ (1.\ 4.)) \tag{1. 5.}$$

$\varepsilon_f\, f(\bar{x})$ heißt (*absoluter*) *erzeugter Fehler*; er wird von der genährten Funktionsberechnung verursacht.
$f(\bar{x}) - f(x)$ heißt (*absoluter*) *fortgepflanzter Fehler*; er gibt an, wie sich der Fehler des Arguments $\bar{x}$ bei exakter Funktionsberechnung auf das Ergebnis auswirken würde.

Es gilt nun

$$f(\bar{x}) = f(\bar{x} + x - x)$$

$$= f(x + \delta_x)$$

$$= f(x) + \delta_x f'(x + \alpha\delta_x), \qquad \alpha \in \left]0,\ \ 1\right[\,. \tag{1. 6.}$$

Damit folgt für den absoluten Gesamtfehler

$$\delta_y = \varepsilon_f\, f(\bar{x}) + f(x) + \delta_x f'(x + \alpha\delta_x) - f(x) \qquad (\because\ \ (2.\ \ 5.\ \ ), (2.\ \ 6.))$$

$$\delta_y \approx \varepsilon_f\, f(\bar{x}) + \delta_x f'(\bar{x}), \qquad \left|\varepsilon_f\right| < \mathrm{E}_f\,, \ \ \left|\delta_x\right| < \Delta_x\,. \tag{1. 7.}$$

Verschafft man sich Betragsschranken für $f(x)$ und $f'(x)$:

$$m_u \le \left|f(x)\right| \le m_o, \qquad \left|f'(x)\right| \le m \quad \text{für } x \in \left[\bar{x} - \Delta_x, \bar{x} + \Delta_x\right], \tag{1. 8.}$$

so hat man statt (1. 7.) - ohne Vernachlässigung von Gliedern höherer Ordnung - die Abschätzung

$$\left|\delta_y\right| \le m_o \mathrm{E}_f + m\Delta_x\,. \tag{1. 9.}$$

Es gilt ferner für den relativen Gesamtfehler

$$\frac{\delta_y}{f(\bar{x})} \approx \varepsilon_f + \frac{\bar{x}\, f'(\bar{x})}{f(\bar{x})}\varepsilon_x, \qquad |\varepsilon_f| < \mathrm{E}_f, \quad |\delta_x| < \Delta_x \qquad (\because\ (1.\,1.\,),(1.\,7.)\,) \qquad (1.\,10.)$$

und

$$\left|\frac{\delta_y}{f(\bar{x})}\right| \le \frac{m_o}{m_u}\mathrm{E}_f + \frac{\left(\left|\bar{x}\right| + \Delta_x\right)m}{m_u}\mathrm{E}_x \qquad (\because\ (1.\,9.)\,) \qquad (1.\,11.)$$

<u>BS. 1. 1.</u> :
Gegeben sei die Funktion

$$f(x) := \cos x$$

und die Näherungsfunktion

$$\tilde{f}(x) := 1 - \frac{x^2}{2}.$$

(*Bemerkung*: Es gilt nach MacLaurinscher Formel:

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \ldots + (-1)^k \frac{x^{2k}}{(2k)!} + R_{2k+1},$$

$$R_{2k+1} = \frac{x^{2k+1}}{(2k+1)!}\cos\left[\vartheta x + (2k+1)\frac{\pi}{2}\right], \quad \vartheta \in \,]0,\ 1[\,)$$

Gesucht sind die Funktionswerte für die Argumente

$$\tilde{x}_1 = 0.395 \quad \text{und} \quad \tilde{x}_2 = 0.785$$

mit dem einheitlichen maximalen absoluten Fehler $\Delta_x = 0.005$.

Lösung:

$$\tilde{y}_1 = \tilde{f}\left(\tilde{x}_1\right) = \tilde{f}\left(0.395\right) = 1 - 0.0780125 = 0.9219875 \approx 0.922$$

(gerundet auf drei Stellen nach dem Basispunkt)

$$\tilde{y}_2 = \tilde{f}\left(\tilde{x}_2\right) = \tilde{f}\left(0.785\right) = 1 - 0.3081125 = 0.6918875 \approx 0.692$$

(gerundet auf drei Stellen nach dem Basispunkt).

<u>1. Weg</u>: (*Rückwärtsanalyse*):

Zuerst versuchen wir, Störungen $\delta_i, \quad i=1{,}2,$ zu bestimmen, für die $\tilde{y}_i,\ i=1{,}2,$ exakte Funktionswerte von $\tilde{x}_i+\delta_i,\ i=1{,}2,$ sind:

$$\tilde{y}_i = \cos\left(\tilde{x}_i+\delta_i\right),\ i=1{,}2\,.$$

Es gilt nun

$$\cos\left(\tilde{x}_i+\delta_i\right) = \cos\tilde{x}_i\,.\cos\delta - \sin\tilde{x}_i\,.\sin\delta_i,\ i=1{,}2$$

und für kleine $\delta_i, \quad i=1{,}2$:

$$\cos\delta_i,\approx 1; \qquad \sin\delta_i \approx \delta_i \quad , \quad i=1{,}2\,.$$

Damit hat man

$$\tilde{y}_i \approx \cos\tilde{x}_i\,.1-\delta_i.\sin\tilde{x}_i,\ i=1{,}2\,,$$

$$\delta\tilde{\imath}_i \approx \frac{\cos\bar{x}_i-\bar{y}_i}{\sin\bar{x}_i}, \quad i=1{,}2\,,$$

$$\delta\tilde{\imath}_1 \approx \frac{0.923-0.922}{0.385} = 0.003\,,$$

$$\delta\tilde{\imath}_2 \approx \frac{0.707-0.692}{0.707} = 0.021\,.$$

Wegen $0.003=|\delta_1| < 0.005 = \Delta_x$ liegt also das Argument $\tilde{x}_1+\delta_1$ in demselben Intervall $[0.395-0.005,\quad 0.395+0.005]=[0.390,\quad 0.400]$ wie das Argument $x_1=\frac{\pi}{8}$. Damit ist $\tilde{y}_1$ eine akzeptable Näherung für den gesuchten Funktionswert $y_1=\cos\frac{\pi}{8}$.

Andererseits liegt wegen $0.021=|\delta_2| >> 0.005=\Delta_x$ das gestörte Argument $\tilde{x}_1+\delta_2$ <u>nicht</u> in demselben Fehlerintervall wie das exakte Argument $x_2=\frac{\pi}{4}$. Das bedeutet, $\tilde{y}_2$ kann <u>nicht</u> als Näherung für $y_2=\cos\frac{\pi}{4}$ akzeptiert werden.

<u>2. Weg</u>: (*Vorwärtsanalyse*):
Es gilt:

$$
\begin{aligned}
\delta_y &:= \tilde{f}(\tilde{x}) - f(x) \\
&= 1 - \frac{1}{2}\tilde{x}^2 - \cos x \\
&= 1 - \frac{1}{2}\tilde{x}^2 - \cos(\tilde{x} - \tilde{x} + x) \\
&= 1 - \frac{1}{2}\tilde{x}^2 - \cos(\tilde{x} - \delta_x) \\
&\approx 1 - \frac{1}{2}\tilde{x}^2 - \cos\tilde{x} - \delta_x \sin\tilde{x}\,. \qquad (1.\ 12.)
\end{aligned}
$$

Wegen des Alternierens der Cosinusreihe ist die Differenz von $\tilde{f}(\tilde{x})$ und $f(\tilde{x})$ betragsmäßig kleiner als das erste vernachlässigte Glied der Reihe, d.h.:

$$
\left|\tilde{f}(\tilde{x}) - f(\tilde{x})\right| = \left|1 - \frac{1}{2}\tilde{x}^2 - \cos\tilde{x}\right| < \frac{1}{4!}\tilde{x}^4
$$

$$
\left|(1+\varepsilon_f)\, f(\tilde{x}) - f(\tilde{x})\right| = \left|\varepsilon_f\, f(\tilde{x})\right| \qquad (\because\ (1.\ 5.))
$$

$$
= \left|1 - \frac{1}{2}\tilde{x}^2 - \cos\tilde{x}\right| < \frac{1}{4!}\tilde{x}^4\,, \qquad (1.\ 13.)
$$

$$
\left|\delta_y\right| \approx \left|1 - \frac{1}{2}\tilde{x}^2 - \cos\tilde{x} - \delta_x \sin\tilde{x}\right| \qquad (\because\ (1.\ 12.))
$$

$$
\begin{aligned}
\left|\delta_y\right| &\le \left|1 - \frac{1}{2}\tilde{x}^2 - \cos\tilde{x}\right| + \left|\delta_x \sin\tilde{x}\right| \\
&\le \left|\frac{1}{4!}\tilde{x}^4\right| + \left|\delta_x \sin\tilde{x}\right|\,, \qquad (\because\ (1.\ 13.))
\end{aligned}
$$

d.h.

$$
\left|\delta_y\right| \le \frac{1}{4!}\tilde{x}^4 + 0.005 \qquad (\because \left|\delta_x\right| \le \Delta_x;\ \left|\sin\tilde{x}\right| \le 1)
$$

Mit $\Delta x = 0.0005$ ergibt sich:

(Letzt Aktualisierung: 17.08.09)