Rundungsfehler und Gleitkommaarithmetik Vorlesung vom 6.11.15

Runden und Rundungsfehler:

Der absolute Rundungsfehler ist nicht gleichmäßig beschränkt.

Der relative Rundungsfehler ist gleichmäßig beschränkt.

Obere Schranke: Maschinengenauigkeit $eps = eps(q, \ell)$.

Praktische Realisierung von Gleitkommazahlen:

Endlicher Exponentenbereich bewirkt endlichen Zahlenvorrat. Datentypen: `double`, `float`.

Zahlenmengen statt Zahlen:

Menge aller Gleitkomma-Approximationen von $x \in \mathbb{R}$ mit relativem Fehler $eps(q, \ell)$.

Menge aller reellen Zahlen, die auf $\tilde{x} \in \mathbb{G}(q, \ell)$ gerundet werden.

Folgerung: Gleichheitsabfragen von Gleitkommazahlen verboten.

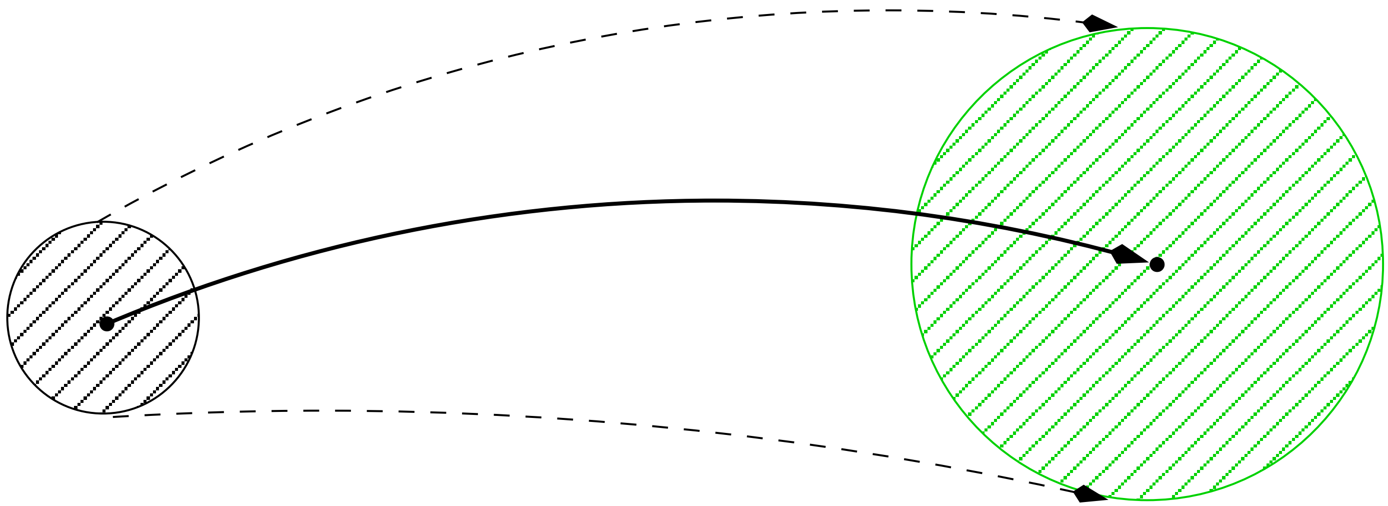
Algebraische Eigenschaften:

Gleitkommaarithmetik, Verlust von Assoziativität, Distributivität, Invertierbarkeit.

Folgerung: Übliche Umformungen sind nicht mehr äquivalent.

Kondition

Auswirkung von Eingabefehlern auf das Ergebnis



Das Landau-Symbol o

Definition

Sei $f : I \to \mathbb{R}$ mit $I = (-a, a)$ eine Funktion.

Wir verabreden die Schreibweise

$$\lim_{\varepsilon \to 0} \frac{f(\varepsilon)}{\varepsilon} = 0 \quad \Longleftrightarrow \quad f(\varepsilon) = o(\varepsilon) \quad (\text{für } \varepsilon \to 0)\ .$$

Beispiele:

$$\varepsilon^2 = o(\varepsilon), \qquad \varepsilon\sqrt{\varepsilon} + \varepsilon \sum_{i=1}^{28} (\sin(\varepsilon))^i = o(\varepsilon), \qquad \ldots$$

Relative Kondition der Multiplikation

gegeben: $x, y \in \mathbb{R}$, x, $y \neq 0$.

Approximationen mit relativem Fehler ε:

$$\tilde{x} = x(1+\varepsilon_x), \quad \tilde{y} = y(1+\varepsilon_y), \qquad \varepsilon = \max\{|\varepsilon_x|, |\varepsilon_y|\}$$

Satz: Es gilt
$$\frac{|(x \cdot y) - (\tilde{x} \cdot \tilde{y})|}{|x \cdot y|} \leq 2\,\varepsilon + \varepsilon^2 \ .$$

Relative Kondition der Multiplikation

gegeben: $x, y \in \mathbb{R}$, x, $y \neq 0$.

Approximationen mit relativem Fehler ε:

$$\tilde{x} = x(1+\varepsilon_x), \quad \tilde{y} = y(1+\varepsilon_y), \qquad \varepsilon = \max\{|\varepsilon_x|, |\varepsilon_y|\}$$

Satz: Es gilt

$$\frac{|(x \cdot y) - (\tilde{x} \cdot \tilde{y})|}{|x \cdot y|} \leq 2\,\varepsilon + \varepsilon^2 \,.$$

Dominierender Fehleranteil: 2ε

Relative Kondition der Multiplikation

gegeben: $x, y \in \mathbb{R}$, x, $y \neq 0$.

Approximationen mit relativem Fehler ε:

$$\tilde{x} = x(1+\varepsilon_x), \quad \tilde{y} = y(1+\varepsilon_y), \qquad \varepsilon = \max\{|\varepsilon_x|, |\varepsilon_y|\}$$

Satz: Es gilt

$$\frac{|(x \cdot y) - (\tilde{x} \cdot \tilde{y})|}{|x \cdot y|} \leq 2\,\varepsilon + \varepsilon^2\ .$$

Dominierender Fehleranteil: 2ε

Vernachlässigung des *Terms höherer Ordnung* $o(\varepsilon) = \varepsilon^2$

Relative Kondition der Multiplikation

gegeben: $x, y \in \mathbb{R}$, x, $y \neq 0$.

Approximationen mit relativem Fehler ε:

$$\tilde{x} = x(1+\varepsilon_x), \quad \tilde{y} = y(1+\varepsilon_y), \qquad \varepsilon = \max\{|\varepsilon_x|, |\varepsilon_y|\}$$

Satz: Es gilt

$$\frac{|(x \cdot y) - (\tilde{x} \cdot \tilde{y})|}{|x \cdot y|} \leq 2\,\varepsilon + \varepsilon^2 \,.$$

Dominierender Fehleranteil: 2ε

Vernachlässigung des *Terms höherer Ordnung* $o(\varepsilon) = \varepsilon^2$

Die relative Kondition ist der Verstärkungsfaktor κ von ε: $\kappa = 2$

Relative Kondition der Division und Addition

Satz: (Division)

Es gilt
$$\frac{|(x/y)-(\tilde{x}/\tilde{y})|}{|x/y|} \leq 2\,\varepsilon + o(\varepsilon)\ .$$

relative Kondition der Division: $\kappa = 2$.

Relative Kondition der Division und Addition

Satz: (Division)

Es gilt $$\frac{|(x/y) - (\tilde{x}/\tilde{y})|}{|x/y|} \leq 2\,\varepsilon + o(\varepsilon)\ .$$

relative Kondition der Division: $\kappa = 2$.

Satz: (Addition)

Es sei $x, y > 0$. Dann gilt $$\frac{|(x+y) - (\tilde{x}+\tilde{y})|}{|x+y|} \leq 1\varepsilon\ .$$

relative Kondition der Addition: $\kappa = 1$.

Relative Kondition der Subtraktion

Satz: (Subtraktion)

Es sei $x, y > 0$. Dann gilt $$\frac{|(x-y)-(\tilde{x}-\tilde{y})|}{|x-y|} \leq \left(\frac{|x|+|y|}{|x-y|}\right) \varepsilon \ .$$

relative Kondition der Subtraktion: $\kappa = \frac{|x|+|y|}{|x-y|}$

Auslöschung: Ist $x \approx y$, so wird $\kappa = \frac{|x|+|y|}{|x-y|}$ beliebig groß!!!

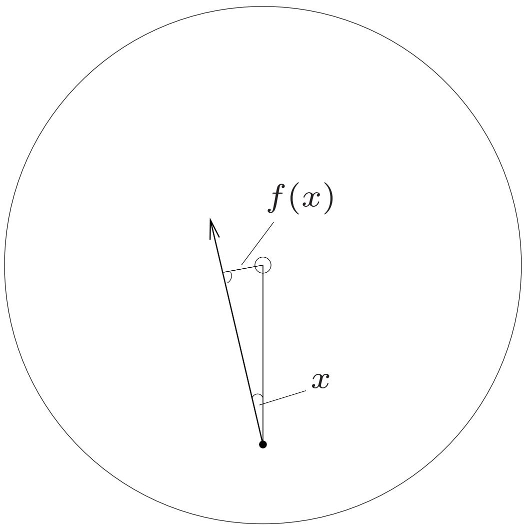
MATLAB – Beispiel

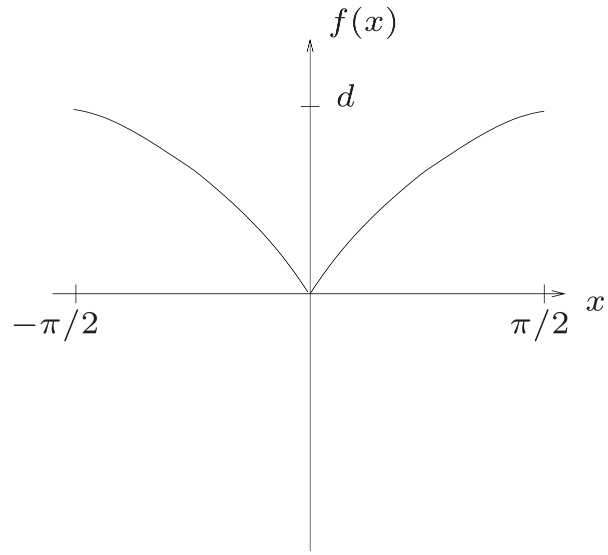
```
>> format long;

x = double(pi)

x = 3.14159265358979

>> y=double(pi+1e-14)

y = 3.14159265358980

>> y-x

ans = 1.021405182655144e-14
```

Nieder mit der Auslöschung

$\boxed{\text{Subtraktion fast gleich großer Zahlen vermeiden!}}$

Einlochen eines Golfballs

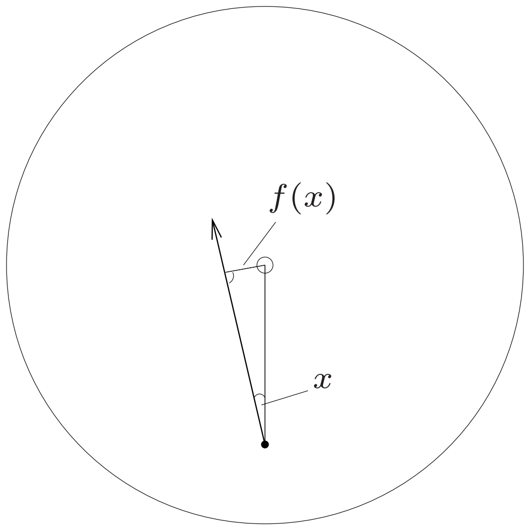


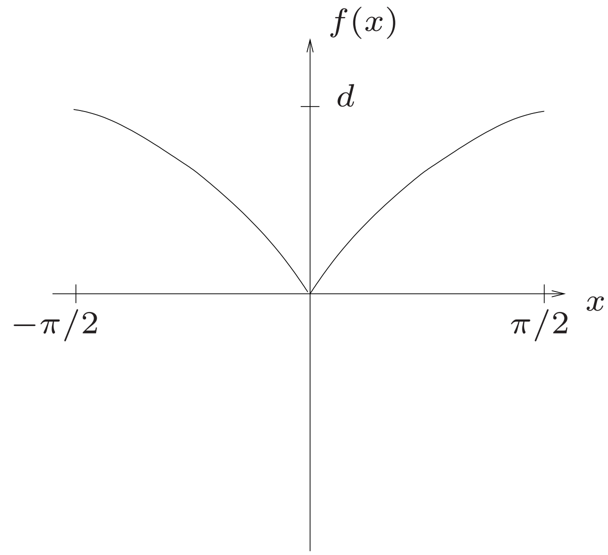


Distanz zum Loch: d, Radius des Lochs: r_L, Abschlagswinkel: x

minimaler Abstand zum Lochmittelpunkt: $f(x) = d|\sin(x)|$

Einlochen eines Golfballs





Distanz zum Loch: d, Radius des Lochs: r_L, Abschlagswinkel: x

minimaler Abstand zum Lochmittelpunkt: $f(x) = d|\sin(x)|$

optimal: $x_0 = 0$, erlaubte Toleranz: $|x - x_0| < |\arcsin(r_L/d)|$

Kondition der Funktionsauswertung

gegeben: Intervall $I \subset \mathbb{R}$, $\quad f : I \mapsto \mathbb{R}$, $\quad x_0 \in I$

Problem: $\quad (*) \quad$ $\boxed{\text{Auswertung von } f \text{ an der Stelle } x_0}$

Kondition der Funktionsauswertung

gegeben: Intervall $I \subset \mathbb{R}$, $\quad f : I \mapsto \mathbb{R}$, $\quad x_0 \in I$

Problem: $(*)$ $\quad \boxed{\text{Auswertung von } f \text{ an der Stelle } x_0}$

Definition (Absolute Kondition)

Die absolute Kondition κ_{abs} von $(*)$ ist die kleinste Zahl mit der Eigenschaft

$$|f(x_0) - f(x)| \leq \kappa_{\text{abs}} |x_0 - x| + o(|x_0 - x|) \ .$$

Liegt dies für keine reelle Zahl κ_{abs} vor, so wird $\kappa_{\text{abs}} = \infty$ gesetzt.

Absolute Kondition und Ableitung

Satz: Ist f differenzierbar in x_0, so gilt $\kappa_{\text{abs}} = |f'(x_0)|$.

Absolute Kondition und Ableitung

Satz: Ist f differenzierbar in x_0, so gilt $\kappa_{\text{abs}} = |f'(x_0)|$.

Beispiel:

Sei $f(x) = x^2$, $x_0 \in \mathbb{R}$. Dann ist $\kappa_{\text{abs}} = |f'(x_0)| = 2|x_0|$.

Kondition und Lipschitz-Stetigkeit

Definition: Die Funktion $f : I \to \mathbb{R}$ heißt Lipschitz-stetig mit Lipschitz-Konstante L, falls

$$|f(x) - f(y)| \leq L|x - y| \qquad \forall x, y \in I \ .$$

Beispiel: $f(x) = |x|$ ist Lipschitz-stetig mit Lipschitz-Konstante $L = 1$.

Satz: Ist $f : I \to \mathbb{R}$ Lipschitz-stetig mit Lipschitz-Konstante L, so genügt die absolute Kondition κ_{abs} von $(*)$ der Abschätzung

$$\kappa_{\text{abs}} \leq L \ .$$

Geschachtelte Funktionen

Satz: Geschachtelte Funktionsauswertung: $f(x) = g \circ h(x) = g(h(x))$.

$\kappa_{\text{abs}}(h, x_0)$: abs. Kondition der Auswertung von h an der Stelle x_0.

$\kappa_{\text{abs}}(g, y_0)$: abs. Kondition der Auswertung von g an der Stelle $y_0 = h(x_0)$.

Dann gilt

$$\kappa_{\text{abs}} \leq \kappa_{\text{abs}}(g, y_0)\ \kappa_{\text{abs}}(h, x_0)\ .$$

Ist h differenzierbar in x_0 und g differenzierbar in y_0, so liegt Gleichheit vor.

Geschachtelte Funktionen

Satz: Geschachtelte Funktionsauswertung: $f(x) = g \circ h(x) = g(h(x))$.

$\kappa_{\text{abs}}(h, x_0)$: abs. Kondition der Auswertung von h an der Stelle x_0.

$\kappa_{\text{abs}}(g, y_0)$: abs. Kondition der Auswertung von g an der Stelle $y_0 = h(x_0)$.

Dann gilt

$$\kappa_{\text{abs}} \leq \kappa_{\text{abs}}(g, y_0)\ \kappa_{\text{abs}}(h, x_0)\ .$$

Ist h differenzierbar in x_0 und g differenzierbar in y_0, so liegt Gleichheit vor.

Beispiel: Das Golfproblem

$$f(x) = |d \cdot \sin(x)| = g(h(x)),\ g(y) = |y|,\ \ h(x) = d \cdot \sin(x), \qquad x_0 = 0$$

$$\kappa_{\text{abs}}(g, h(x_0)) = d, \quad \kappa_{\text{abs}}(h, x_0) \leq 1 \quad \Longrightarrow \quad \kappa_{\text{abs}} \leq d$$

Relative Kondition von Funktionsauswertungen

gegeben: Intervall $I \subset \mathbb{R}$, $\quad f : I \to \mathbb{R}$, $\quad 0 \neq x_0 \in I$, $f(x_0) \neq 0$

Problem: $\quad (*)$ $\quad \boxed{\text{Auswertung von } f \text{ an der Stelle } x_0}$

Definition 3.6 (Relative Kondition)

Die relative Kondition κ_{rel} von $(*)$ ist die kleinste Zahl mit der Eigenschaft

$$\frac{|f(x_0) - f(x)|}{|f(x_0)|} \leq \kappa_{\text{rel}} \frac{|x_0 - x|}{|x_0|} + o(|x_0 - x|) \ .$$

Liegt dies für keine reelle Zahl κ_{rel} vor, so wird $\kappa_{\text{rel}} = \infty$ gesetzt.

Relative versus absolute Kondition

absolute Kondition

$$|f(x_0) - f(x)| \le \kappa_{\text{abs}}|x_0 - x| + o(|x_0 - x|)$$

relative Kondition

$$\frac{|f(x_0) - f(x)|}{|f(x_0)|} \le \kappa_{\text{rel}}\frac{|x_0 - x|}{|x_0|} + o(|x_0 - x|)$$

Relative versus absolute Kondition

absolute Kondition

$$|f(x_0) - f(x)| \leq \kappa_{\text{abs}}|x_0 - x| + o(|x_0 - x|)$$

relative Kondition

$$\frac{|f(x_0) - f(x)|}{|f(x_0)|} \leq \kappa_{\text{rel}}\frac{|x_0 - x|}{|x_0|} + o(|x_0 - x|)$$

Satz: Es gilt

$$\kappa_{\text{rel}} = \frac{|x_0|}{|f(x_0)|}\kappa_{\text{abs}}.$$

Relative versus absolute Kondition

Beispiel: $f(x) = ax$,

absolute Kondition:

$$\kappa_{\text{abs}} = |f'(x_0)| = |a|$$

Relative versus absolute Kondition

Beispiel: $f(x) = ax$,

absolute Kondition:

$$\kappa_{\text{abs}} = |f'(x_0)| = |a|$$

relative Kondition:

$$\kappa_{\text{rel}} = \frac{|x_0|}{|f(x_0)|}\kappa_{\text{abs}} = \frac{|x_0|}{|ax_0|}|a| = 1$$

Relative versus absolute Kondition

Beispiel: $f(x) = ax$,

absolute Kondition:

$$\kappa_{\text{abs}} = |f'(x_0)| = |a|$$

relative Kondition:

$$\kappa_{\text{rel}} = \frac{|x_0|}{|f(x_0)|}\kappa_{\text{abs}} = \frac{|x_0|}{|ax_0|}|a| = 1$$

Folgerung: Relative und absolute Kondition können sich beliebig stark unterscheiden.

aus $|a| \gg 1$ folgt $\kappa_{\text{abs}} \gg \kappa_{\text{rel}}$ aus $|a| \ll 1$ folgt $\kappa_{\text{abs}} \ll \kappa_{\text{rel}}$

weiteres Beispiel: absolute Kondition der Subtraktion