

■ Normalisierte Gleitkommadarstellung reeller Zahlen $x \neq 0$ auf dem Computer:

(1) $x = \pm 0, \alpha_1 \alpha_2 \dots \alpha_t \cdot B^e$
 mit $\alpha_i \in \{0, 1, \dots, B-1\}$, $\alpha_1 \neq 0$,
 $B \in \mathbb{N} - \{1\}$ Basis
 $m = 0, \alpha_1 \alpha_2 \dots \alpha_t$ Mantisse
 $e \in \mathbb{Z}$: $U \leq e \leq O$ Exponent

Beispiel:

1. Datentyp float (einfache Genauigkeit) in $\mathbb{C}$:

$B = 2$, $t = 24$, $U = -125$, $O = 128$: 24 Dualstellen $\hat{=}$ 7 Dezimalst.

2. Datentyp double (doppelte Genauigkeit) in $\mathbb{C}$:

$B = 2$, $t = 53$, $U = -1024$, $O = 1024$: 53 Dualstellen $\hat{=}$ 15 Dezimalst.

Überschreiten (overflow) des Exponentenber. $\Rightarrow$ Programmabbr.

Unterschreiten (underflow) des Exponentenber. $\Rightarrow 0$ v. $\pm$ Kl. Zahl

$$\mathcal{M} := \{0\} \cup \left\{ \pm 0, \underset{(1)}{\alpha_1 \dots \alpha_t} B^e \right\} = \text{Menge aller Maschinenzahlen}$$

■ Runden und Rundungsfehler:

$\mathcal{M}$ besteht nur aus endlich vielen Zahlen !!!

Damit kann natürlich nicht jede reelle Zahl $x \in \mathbb{R}$ auf dem Rechner exakt dargestellt werden !!!

Man muss runden, d.h.

$$\mathbb{R} \ni x \rightsquigarrow \text{rd}(x) \in \mathcal{M}$$

Die bestmögliche Approximation $\text{rd}(x) \in \mathcal{M}$ einer reellen Zahl $x \in \mathbb{R}$ erfüllt die Bedingung:

(2) $|x - \text{rd}(x)| \leq |x - y| \quad \forall y \in \mathcal{M}$

↗ Kaufmännischen Runden: Für $x = \pm 0, \alpha_1 \dots \alpha_t \alpha_{t+1} \dots B^e$

mit $\alpha_1 \neq 0$ rundet man auf, falls $\alpha_{t+1} \geq B/2$ bzw.
 rundet man ab, falls $\alpha_{t+1} < B/2$