21.3.2012 Bsp (b-adische Darstellung) ④

$b=2$, $x=(13.1)_{10}$ ← Festpunkt-Darstellung

Vorkomma-Anteil (Algorithmus 1.1)

K	0	1	2	3	4
U	13	6	3	1	0
x_K	1	0	1	1	

also: $(13)_{10} = (1101)_2$

Nachkomma-Anteil (Algorithmus 1.2)

i	0	1	2	3	4	5	6	
u im Z3	0.1	0.2	0.4	0.8	1.6	1.2	0.4	...
x_i	–	0	0	0	1	1	1	...
u im Z5	–	0.2	0.4	0.8	0.6	0.2	0.4	

also:

$(0.1)_{10} = (0.0\overline{0011})$

$\Rightarrow x=(1101.0\overline{0011})_2$

$x=(1.101000\overline{11001})_2 \cdot 2^3$

↑ 2-adische Darstellung

Umwandlung Dual -> Dezimal

Vorkomma-Anteil mit Horner-Schema (Algorithmus 1.3)

	$(1$	1	0	$1)_2$
	0	2	6	12
+				
$\cdot 2$	1	3	6	$\underline{13}$ ← Dezimalwert

Nachkomma-Anteil (Algorithmus 1.4)

← Zahl wurde abgeschnitten!

$x =$	$(0.$	0	0	0	1	$1)_2$
		$\frac{3}{16}$	$\frac{3}{8}$	$\frac{3}{4}$	$\frac{1}{2}$	0
+						
		$\frac{3}{16}$	$\frac{3}{8}$	$\frac{3}{4}$	$\frac{3}{2}$	1

$\frac{1}{2}$ ← Start

Zeile 5: $\left(\frac{3}{32}\right)_{10} = x$

Bsp (IM (b=2, p=3, $E_{min}=-1$, $E_{max}=1$)) "Toy System"

Mögliche Mantissen (für $\sigma=0$)	Dezimaler Wert		
	$E=-1$	$E=0$	$E=1$
$(1.00)_2$	0.5 (x_{min})	1	2
$(1.01)_2$	0.625	1.25	2.5
$(1.10)_2$	0.75	1.5	3.0
$(1.11)_2$	0.875	1.75	3.5 (x_{max})
Abstand	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$

Minimale und maximale positive norm. Maschinenzahl

$$x = (\underset{\neq 0}{d_0} \bullet d_1 d_2 \dots d_{p-1}) \cdot b^E$$

$$x_{min} = (1.0\dots00)_b \cdot b^{E_{min}} = b^{E_{min}}$$

$$x_{max} = \left((b-1)_\bullet (b-1)(b-1)\dots(b-1)\right)_b \cdot b^{E_{max}}$$

$$= \left(\sum_{j=0}^{p-1} (b-1) \cdot b^{-j} \right) \cdot b^{E_{max}}$$

$$= \sum_{j=0}^{p-1} b^{-j+1} - \sum_{j=0}^{p-1} b^{-j}$$

Teleskopsumme

$\Rightarrow$ Indexverschiebung

$j = j+1$

$$= \sum_{j=-1}^{p-2} b^{-j} - \sum_{j=0}^{p-1} b^{-j}$$

$$= b^1 + \sum_{j=0}^{p-2} b^{-j} - \left(\sum_{j=0}^{p-2} b^{-j} + b^{-p+1} \right)$$

$$= b - b^{-p+1}$$

$$x_{max} = (b - b^{-p+1}) \cdot b^{E_{max}}$$

$$= (1 - b^{-p}) \cdot b^{E_{max}+1}$$

Gesamtanzahl Maschinenzahlen

	$\pm$	$(d_0$	$.$	d_1	d_2	$\cdots$	$d_{p-1})_b \cdot b^E$	
Wahlmöglichkeiten	2	$(b-1)$		b	b		b	$E_{min} \le E \le E_{max}$
								$(E_{max} - E_{min} + 1)$ Stück

Gesamtanzahl

$$2(b-1) \cdot b^{p-1} \cdot (E_{max} - E_{min} + 1) + 1$$

↑ für 0

Abstand der Maschinenzahlen

Im Intervall $[1, b] = [b^0, b^1]$ liegen die Zahlen $(1.d_1 \dots d_{p-1})_b \cdot b^0$

diese Stelle hat den Wert $b^{-(p-1)}$ ⟶ ↑ erhöhe um 1 (mit Übertrag)

Diese sind gegeben durch

$$1 + K \cdot b^{-p+1} \quad \text{mit} \quad \underbrace{K \in \{0, 1, \dots (b-1) \cdot b^{(p-1)} - 1\}}_{\text{Indexmenge } \Pi}$$

Wieviele Zahlen liegen in $[1, b]$?

$\Rightarrow$ $\underbrace{(b-1)}_{\text{für } d_0} \cdot \underbrace{b^{(p-1)}}_{\text{für } d_1 \text{ bis } d_{p-1}}$

Deshalb haben die Maschinenzahlen den Abstand b^{-p+1}.

- Im Intervall $[b^E, b^{E+1}]$ liegen die Maschinenzahlen $(1 + K \cdot b^{-p+1}) \cdot b^E$, mit $K \in \Pi$.
- " haben den Abstand $b^{-p+1} \cdot b^E = b^{E+1-p}$

Bezeichnung

$$x = \pm (d_0 . d_1 \dots d_{p-1})_b \cdot b^E$$

dann heißt $(0.0 \dots 01)_b \cdot b^E = \underbrace{ulp}(x)$

unit in last place

absoluter Fehler

$$|\tilde{x} - x| \leq \varepsilon_{abs}$$

$$-\varepsilon_{abs} \leq \tilde{x} - x \leq \varepsilon_{abs} \quad | + x$$

$$x - \varepsilon_{abs} \leq \tilde{x} \leq x + \varepsilon_{abs}$$

$|x| \leq a$

$\Leftrightarrow -a \leq x \leq a$

Relativer Fehler

Wenn: $\dfrac{|\tilde{x} - x|}{|x|} \leq \varepsilon_{rel}$

$\left| \dfrac{\tilde{x} - x}{x} \right|$

$$\Leftrightarrow \quad -\varepsilon_{rel} \leq \frac{\tilde{x} - x}{x} \leq \varepsilon_{rel} \quad | \cdot x$$

$$\Leftrightarrow \quad -x \cdot \varepsilon_{rel} \leq x - \tilde{x} \leq x \cdot \varepsilon_{rel} \quad | + x$$

$$\Leftrightarrow \quad \underbrace{x - x \cdot \varepsilon_{rel}}_{x(1-\varepsilon_{rel})} \leq \tilde{x} \leq \underbrace{x + x \cdot \varepsilon_{rel}}_{x(1+\varepsilon_{rel})}$$

minimaler relativer Abstand

Im Intervall $[b^E, b^{E+1}]$ haben die Maschinenzahlen den Abstand b^{E+1-p}

somit:

$$\min_{x \in \mathbb{M}} |x-y| \leq \frac{1}{2} b^{E+1-p} \quad (*) \text{ abs. Fehlerabstand}$$

$y \in [b^E, b^{E+1}]$, o.B.d.A. $y > 0$

$$b^E \leq y \Rightarrow \frac{1}{y} \leq \frac{1}{b^E} \quad (**)$$

$$(*)\,(**) \Longrightarrow \min_{x \in \mathbb{M}} \frac{|x-y|}{|y|} \leq \frac{\frac{1}{2} b^{E+1-p}}{\underbrace{b^E}} = \frac{1}{2} b^{1-p} = u$$

Maschinen Epsilon

$$eps = \varepsilon_M = b^{-p+1}$$

↳ MatLab/Octave

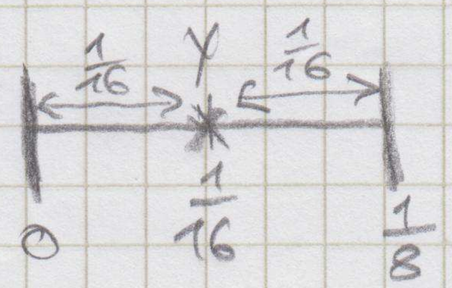
Subnormaler Toy-System

$\mathbb{M}_S(2,3,-1,1)$ haben die Form $(0.d_1d_2) \cdot 2^{-1}$,

mit $d_1, d_2 \in \{0,1\}$

Mantissen	$\cdot 2^{-1}$	Abstand
$(0.01)_2 = \frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
$(0.10)_2 = \frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$
$(0.11)_2 = \frac{3}{4}$	$\frac{3}{8}$	

relativer Abstand zur reeller Zahl y steigt auf Wert 1 an.



relativer Abstand:

$$\min_{x \in \mathbb{M}_s} \frac{|x-y|}{|y|} = \frac{\overset{=\frac{1}{16}}{\frac{1}{16}}}{\frac{1}{16}} = 1$$

Rundung in Toy-Systemen

$$y = (1.7)_{10} = (1.10\overline{1100})_2 \cdot 2^1$$

a) Abrunden $rd(y) = y_- = (1.10)_2 \cdot 2^1$
$= (1.5)_{10}$

b) Aufrunden $rd(y) = y_+ = (1.11)_2 \cdot 2^1$
$= (1.75)_{10}$

c) Abschneiden $rd(y) = y_- = (1.5)_{10}$

d) Optimale Rundung:

Abstände

$|y_- - y| = |1.5 - 1.7| = 0.2$

$|y_+ - y| = |1.75 - 1.71| = \underline{0.05}$

$rd(y) = 1.75$