Technische Universität München
Institut für Informatik
Prof. Dr. Daniel Cremers
Dr. Rudolph Triebel
Nikola Tchipev, M.Sc.
Sebastian Rettenberger, M.Sc.

WiSe 2014/2015

Numerisches Programmieren, Übungen

Musterlösung 1. Übungsblatt: Zahlendarstellung, Rundungsfehler

1) Gleitkomma-Zahlen im IEEE-Standard

Bit-Verwendungszweck	gesetztes Bit für $-1.0\overline{00011} \cdot 2^0$
Vorzeichen (1 Bit) ('-' wird zu '1')	1
Exponent (8 Bits) (0 + 127 = 127)	0 1 1 1 1 1 1 1
Mantisse (23 Bits)	0 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 1

Das letzte Bit der Mantisse ergibt sich nach der Rundungsregel.

2) Ganzzahlen, Fest- und Gleitkommazahlen

Nr.	Frage		$A = I$	$A = F$	$A = G$
1	Darstellungsbeispiele	1 0000000	-0	-0,0	$-2^{-15} \cdot 2^0$
2		0 0000000	+0	+0,0	$2^{-15} \cdot 2^0$
3		0 1000000	64	8,0	$2^1 \cdot 2^0$
4		0 1000100	68	8,5	$2^2 \cdot 2^0$
5		0 1000110	70	8,75	$6 = 2^2 \cdot (2^0 + 2^{-1})$
6		0 1000111	71	8,875	$7 = 2^2 \cdot (2^0 + 2^{-1} + 2^{-2})$
7	Allg. Eigenschaften	$\lvert A \rvert$	256	256	256
8		$\max_{a \in A} a$	127	15,875	$114688 = 2^{16} \cdot (2^0 + 2^{-1} + 2^{-2})$
9		$\min_{a \in A} a$	-127	-15,875	-114688
10		$\min_{a \in A, a > 0} a$	1	0,125	$2^{-15} \cdot 2^0$
11		Ist 0 in A?	ja	ja	nein
12	Anzahl Zahlen von A im	$[2^{-15}, 2^{-14})$	0	0	4
13		$[1, 2)$	1	8	4
14		$[2, 4)$	2	16	4
15		$[4, 8)$	4	32	4
16		$[8, 16)$	8	64	4
17		$[16, 32)$	16	0	4
18		$[32, 64)$	32	0	4
19		$[64, 128)$	64	0	4
20		$[2^{16}, 2^{17})$	0	0	4
21		$[0, 1)$	1	8	$60 = (2^4 - 1) \cdot 2^2$

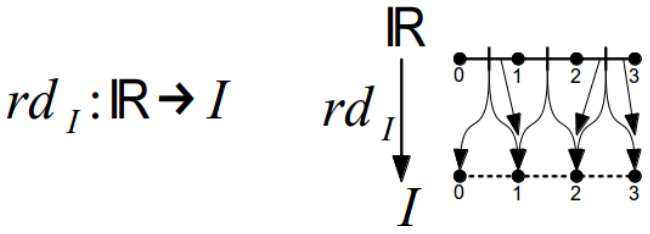

Abbildung 1: Veranschaulichung von rd_I.

Nr.	Frage		$A = I$	$A = F$	$A = G$
22	Rundungen: $\mathrm{rd}_A(x)$	2^{-5}	0	0	2^{-5}
23		3^{-5}	0	0	2^{-8}
24		2,1	2	2,125	2
25		3,1	3	3,125	3
26		4,1	4	4,125	4
27		9	9	9	8
28		18	18	15,875	16
29		1023	127	15,875	1024
30		1151	127	15,875	1024
31	Absoluter Rundungsfehler: $\lvert x - \mathrm{rd}_A(x)\rvert$	2^{-5}	2^{-5}	2^{-5}	0
32		3^{-5}	3^{-5}	3^{-5}	$2{,}09{\cdot}10^{-4}$
33		2,1	0,1	0,025	0,1
34		3,1	0,1	0,025	0,1
35		4,1	0,1	0,025	0,1
36		9	0	0	1
37		18	0	2,125	2
38		1023	...	...	1
39		1151	...	...	127
40	Relativer Rundungsfehler: $\left\lvert\frac{x-\mathrm{rd}_A(x)}{x}\right\rvert$	2^{-5}	1	1	0
41		3^{-5}	1	1	0,051
42		2, 1	0,048	0,012	0,048
43		3, 1	0,032	0,008	0,032
44		4, 1	0,024	0,006	0,024
45		9	0	0	0,111
46		18	0	0,118	0,111
47		1023	...	...	0,001
48		1151	...	...	0,110

Beobachtungen

a) Betrachten Sie Ihre Antwort zu Frage Nr. 27. In manchen Programmiersprachen (z.B. JavaScript) gibt es kein Format für Ganzzahlen. Welche Konsequenzen kann das haben?

- Es kann vorkommen, dass Rundungsfehlern auftreten, wenn man Objekte einfach zählen möchte.

b) Was ist die kleinste Ganzzahl, die bei 32-Bit IEEE Gleitkommazahlen nicht exakt darstellbar ist?

- $2^{24} + 1$. Vergleichswerte: 1 Gigabyte = 2^{30} Bytes,
  1 Mililiter Wasser enthält $\approx 2^{74}$ Moleküle.

c) Betrachten Sie die Antworten zu den Fragen 28.-30. für $A = F$. Welche Konsequenzen kann die Abwesenheit einer Inf-Darstellung haben?

- Man weiß nicht, ob 01111111_F genau 15,875 entspricht, oder etwas größerem.

d) Für eine Mantissendarstellung mit 2 Bits ist die Maschinengenauigkeit $\varepsilon_{Ma} = 2^{-3}$. Vergleichen Sie ihre Antworten zu Fragen Nr. 40-48 mit der Maschinengenauigkeit. Gilt

$$\left| \frac{x - \mathrm{rd}_G(x)}{x} \right| \leq \varepsilon_{Ma} \tag{1}$$

immer?

- Die Rundungsfehler, die durch rd_G entstehen sind immer kleiner oder gleich ε_{Ma} (solange die Zahl im Range ist, natürlich). Man kann Gleichung (1) beweisen (betrachten Sie Ihre Antworten zu Fragen Nr. 12-20.).

e) Gibt es eine Relation zwischen der Maschinengenauigkeit und der Zahl aus Frage Nr. 10?

- Nein. Die kleinste darstellbare positive Zahl ist nur von der Anzahl an Bits in den Exponenten abhängig. Die Maschinengenauigkeit ist nur von der Anzahl an Bits in der Mantisse abhängig.

f) Was ist der Anteil an Zahlen im $[0, 1)$ beim Format G? Bei 32-Bit IEEE?

- Fast ein Viertel für beide. Bei G: $\frac{60}{256}$ =0,23.

g) Wie viele Zahlendarstellungen sind für spezielle Werte (Exponentenkombinationen $00\ldots0$ $11\ldots1$) bei IEEE reserviert?

- 2 (Vorzeichen) $\cdot$ 2 (Exponentenkombinationen) $\cdot$ 2^{23} (Mantissenmöglichkeiten) = 2^{25} = 33554432 Zahlendarstellungen.

Zusätzliches

h) Gibt es wirklich einen Unterschied zwischen I und F? Betrachten Sie Ihre Antworten zu Fragen 12-21 für $A = I$ und $A = F$. Kann man F anhand I implementieren?

- Es gibt kaum Unterschiede. Man kann F durch I und einen Skalierungsfaktor implementieren.