Übungsblatt 6

Aufgabe 1

The number A is positive, so the sign bit is 0. It is true that

$$5_{10} = 101_2 = 1.01 \times 2^2. \tag{1}$$

So for the mantissa we obtain:

$$0100\,0000\,0000\,0000\,0000\,000.$$

Since the exponent is biased, we obtain for the exponent:

$$1000\,0001.$$

All in all:

$$A = 0\;10000001\;0100\,0000\,0000\,0000\,0000\,000.$$

By the same logic we obtain for B:

$$B = 0\;10000011\;0000\,0000\,0000\,0000\,0000\,000.$$

Now we check the correctness of the representation of A:

$$\begin{align}
v(A) &= (-1)^0 \cdot 2^{\sum_{k=0}^{7} 2^k \cdot e_k - (2^7-1)} \cdot \left(1 + \sum_{k=1}^{23} 2^{-k} \cdot m_{23-k}\right) \tag{2}\\
&= 2^{2^0+2^7-2^7+1} \cdot \left(1 + \sum_{k=1}^{23} 2^{-k} \cdot m_{23-k}\right) \tag{3}\\
&= 2^2 \cdot \left(1 + \sum_{k=1}^{23} 2^{-k} \cdot m_{23-k}\right) \tag{4}\\
&= 2^2 \cdot \left(1 + 2^{-2}\right) \tag{5}\\
&= 5. \tag{6}
\end{align}$$

For B we obtain:

$$\begin{align}
v(B) &= (-1)^0 \cdot 2^{\sum_{k=0}^{7} 2^k \cdot e_k - (2^7-1)} \cdot \left(1 + \sum_{k=1}^{23} 2^{-k} \cdot m_{23-k}\right) \tag{7}\\
&= 2^{2^0+2^1+2^7-2^7+1} \cdot \left(1 + \sum_{k=1}^{23} 2^{-k} \cdot m_{23-k}\right) \tag{8}\\
&= 2^4 \tag{9}\\
&= 16. \tag{10}
\end{align}$$

The quotient Q of A divided by B is 3.2:

$$0.3125_{10} = 0\,01111101\,0100\,0000\,0000\,0000\,0000\,000.$$

Now choose $a = 2$ and compute:

$$\begin{aligned}
\text{fdiv}(A, B, a) &= -1^{0-0} \cdot 2^{\sum_{k=0}^{7}(2^k \cdot (e_{1,k} - e_{2,k}))} \cdot \sum_{k=0}^{1} \left(2^{-k} \cdot q_{M'-k}\right) &(11)\\
&= 2^{-2} \cdot \sum_{k=0}^{1} \left(2^{-k} \cdot q_{M'-k}\right) &(12)\\
&= 2^{-2} \cdot \sum_{k=0}^{1} \left(2^{-k} \cdot q_{M'-k}\right) &(13)\\
&= 2^{-2} \cdot (1 + 2^{-2}) &(14)\\
&= 0.3125. &(15)
\end{aligned}$$

This is the case since the mantissa of both A and B are zero after the m_{21}th place, adding nothing to the division. That is why the approximate division gives an accurate result for A and B. The maximum error an a-bit anytime instruction can have is

$$\frac{1 + \sum_{k=a+1}^{M} 2^{-k}}{1} - \frac{1}{1 + \sum_{k=a+1}^{M} 2^{-k}}. \tag{16}$$