# Homework 1: Problem 1

David Denberg

March 11, 2019

Let

$$x = \pm(b_0.b_1b_2...)2^e$$

and

$$\mathrm{rd}(x) = \pm(b_0.b_1b_2...b_{p-1})2^e$$

where $b_{p-1} = 1$ if $b_p = 1$. We take the two cases separately.
Case 1 ($b_p = 0$):

$$\begin{aligned}\frac{|x - \mathrm{rd}(x)|}{|x|} &= \frac{|(b_0.b_1b_2...)_2 \times 2^e - (b_0.b_1b_2...b_{p-1})_2 \times 2^e|}{|(b_0.b_1b_2...)_2 \times 2^e|} \\ &= \frac{|(0.b_{p+1}b_{p+2}...)_2 \times 2^{e-p}|}{|(b_0.b_1b_2...)_2 \times 2^e|} \\ &= \frac{(0.b_{p+1}b_{p+2}...)_2}{(b_0.b_1b_2...)_2} \times 2^{-p}\end{aligned}$$

To maximize this quantity, the denominator must be minimized and the numerator, maximized. $b_0$ in the denominator must be 1 as the quantity is normalized so the minimum value in the denominator is 1. If we choose $b_i = 1$ for $i = p+1, p+2, ...$ then the numerator must equal:

$$\sum_{i=1}^{\infty} 2^{-i} = 1$$

as it is a geometric series. Then the maximum absolute relative error when $b_p = 0$ is $2^{-p}$.
Case 2 ($b_p = 1$):

$$\begin{aligned}\frac{|x - \mathrm{rd}(x)|}{|x|} &= \frac{|(b_0.b_1b_2...)_2 \times 2^e - (b_0.b_1b_2...b_{p-2}1)_2 \times 2^e|}{|(b_0.b_1b_2...)_2 \times 2^e|} \\ &= \frac{|((b_{p-1} - 1).b_pb_{p+1}...)_2 \times 2^{e-(p-1)}|}{|(b_0.b_1b_2...)_2 \times 2^e|} \\ &= \frac{|((b_{p-1} - 1).b_pb_{p+1}...)_2|}{(b_0.b_1b_2...)_2} \times 2^{-(p-1)}\end{aligned}$$

To maximize this quantity let $b_{p-1} = 0$ and choose $b_i = 1$ for $i = p, p+1, ....$ The numerator must then equal:

$$\sum_{i=0}^{\infty} 2^{-i} = 2$$

The denominator is the same as in Case 1, so the maximum absolute relative error when $b_p = 1$ is $2^{-p}$.