Simple Example of Floating Point Numbers and Arithmetic

This example uses 3 digit rounded floating point operations.

We seek to evaluate

$$\frac{12.34-.56789}{.00009876}$$

in floating point.

Step 1: Convert all of the operands to 3 digit rounded floating point numbers:

$12.34 \to 12.3$
$.56789 \to .568$
$.00009876 \to .0000988$

(Notice the decimal point does not move – all we do is drop numbers after the third one.)

Step 2:

a. Do the subtraction **exactly**:

   $12.3-.568=11.732$

b. Convert to three digits:.

   $11.732 \to 11.7$

Step 3:

a. Do the division **exactly** (meaning with enough digits to know if there is rounding or not):

   $\frac{11.7}{.0000988}=118421.05263157...$

b. Convert to three digits:.

   $118421.05263157... \to 118000.$

Thus the result of computing $\frac{12.34-.56789}{.00009876}$ in 3 digit rounded floating point is 118000. The exact answer is $119199.1697043337...$, thus the error is

$$118000.\text{-}119199.1697043337...=\text{-}1199.1697043337...,$$

and the (absolute) relative error is

$$\left|\frac{\text{-}1199.1697043337...}{119199.1697043337...}\right|=0.010060232924217,$$

which is a little over 1%.