as we know, representing a number in Computer with Floating point representation will Cause some round off errors.

For example in previously discussed 6 bit Floating number we don't have number 11.75. instead we have 8 10 11.75 12 14

So if we want to store 11.75 in Computer with in a 6 bit Floating point number representation, we will have some round off errors (as previously disscussed, the relative error will be always less than ε_m (relative machine percision). ε_m is equal to $2^{-(t)}$ in which t is number of bits used to represent mantisa (23 For 32 bit, and 2 for my 6bit Floating point number)

$x \xrightarrow[\text{will turn it to}]{\text{Saving in Computer}} x(1+\varepsilon) \quad ; |\varepsilon| \leq \varepsilon_m$

Note that the relative error is less than ε_m

$$\left| \frac{x - x(1+\varepsilon)}{x} \right| = \varepsilon$$

So when ever we want to save a value in computer (be it an input value or the result of a calculation), we will have the roundoff error

---

Error of addition/substraction

$x_1 \longrightarrow rd(x_1) = x_1(1+\varepsilon_1) = \tilde{x}_1$

$x_2 \longrightarrow rd(x_2) = x_2(1+\varepsilon_2) = \tilde{x}_2$

$\tilde{x}_1 + \tilde{x}_2 \longrightarrow rd(\tilde{x}_1 + \tilde{x}_2) = (\tilde{x}_1 + \tilde{x}_2)(1+\alpha)$

$(x_1(1+\varepsilon_1) + x_2(1+\varepsilon_2))(1+\alpha)$

$$= (x_1 + x_2 + x_1\varepsilon_1 + x_2\varepsilon_2)(1+\alpha)$$

$$= x_1 + x_2 + x_1\varepsilon_1 + x_2\varepsilon_2 + \alpha(x_1+x_2)$$

$$\text{relative Error} = \frac{rd(\tilde{x}_1 + \tilde{x}_2) - (x_1 + x_2)}{(x_1 + x_2)}$$

$$= \frac{x_1\varepsilon_1 + x_2\varepsilon_2 + \alpha(x_1+x_2)}{(x_1+x_2)}$$

$$\text{Relative E} = \frac{x_1}{x_1+x_2}\varepsilon_1 + \frac{x_2}{x_1+x_2}\varepsilon_2 + \alpha$$

if $x_1 \simeq -x_2$ then we will have large relative errors

Error of multiplication/devision

$$x_1 \longrightarrow \tilde{x}_1 = rd(x_1) = x_1(1+\varepsilon_1)$$
$$x_2 \longrightarrow \tilde{x}_2 = rd(x_2) = x_2(1+\varepsilon_2)$$

$$x_1 x_2 \rightarrow rd(\tilde{x}_1\tilde{x}_2) = x_1x_2(1+\varepsilon_1)(1+\varepsilon_2)(1+\alpha)$$

$$(x_1 + x_1\varepsilon_1)(x_2 + x_2\varepsilon_2)(1+\alpha)$$

$$= x_1x_2 + x_1x_2\varepsilon_1 + x_1x_2\varepsilon_2 + \alpha x_1x_2$$

$$= x_1x_2(1 + \varepsilon_1 + \varepsilon_2 + \alpha)$$

$$\text{relative error} = \frac{rd(x_1x_2) - x_1x_2}{x_1x_2} = 1 + \varepsilon_1 + \varepsilon_2 + \alpha$$

For devision

$$x_1 \rightsquigarrow \tilde{x}_1 = rd(x_1) = x_1(1+\varepsilon_1)$$

$$x_2 \rightsquigarrow \tilde{x}_2 = rd(x_2) = x_2(1+\varepsilon_2)$$

$$\frac{x_1}{x_2} \rightarrow rd\left(\frac{\tilde{x}_1}{\tilde{x}_2}\right) = \left(\frac{x_1(1+\varepsilon_1)}{x_2(1+\varepsilon_2)}\right)(1+\alpha)$$

$$\text{relative error} = \frac{\frac{x_1}{x_2} - rd\left(\frac{\tilde{x}_1}{\tilde{x}_2}\right)}{\frac{x_1}{x_2}} = \frac{\frac{x_1}{x_2} - \frac{1+\varepsilon_1}{1+\varepsilon_2}(1+\alpha)\left(\frac{x_1}{x_2}\right)}{\frac{x_1}{x_2}}$$

$$= 1 - \frac{1+\varepsilon_1}{1+\varepsilon_2}(1+\alpha) = 1 - \frac{1+\varepsilon_1}{1+\varepsilon_2} - \alpha\frac{1+\varepsilon_1}{1+\varepsilon_2}$$