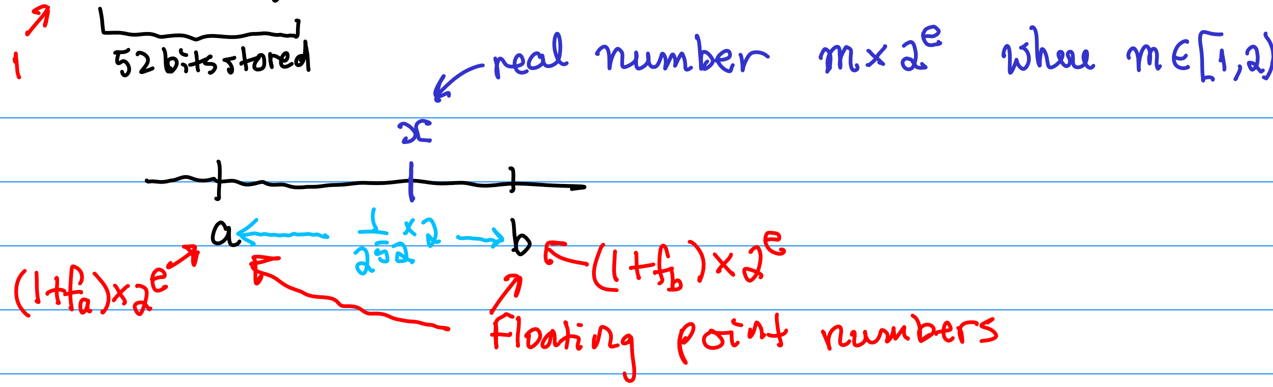
$1+f = d_0 . d_1 d_2 \cdots d_{52}$

$\uparrow$ 1 $\quad \underbrace{\hphantom{d_1 d_2 \cdots d_{52}}}_{52 \text{ bits stored}}$



$x^* = fl(x) = b$

round to the nearest break ties by choosing $d_{52} = 0$

$\varepsilon = x^* - x$ ← absolute error

$$|\varepsilon| \leq \frac{1}{2}|b-a| = \frac{1}{2} \cdot \frac{1}{2^{52}} \cdot 2 = \frac{1}{2^{53}} \cdot 2^e$$

$\delta = \dfrac{x^* - x}{x}$ ← relative error

$$|\delta| = \frac{|\varepsilon|}{|x|} \leq \frac{\frac{1}{2^{53}} \cdot 2^e}{m \times 2^e} \leq \frac{\frac{1}{2^{53}}}{1} = \frac{1}{2^{53}}$$

- Note the relative error does not depend on the scale.

Propagation of error:

Idea: want to compute some output given some input.

$$y = f(x)$$

output (y), input (x)

First problem is we have to approximate the input.

Plug in the approximation x^* and now $y^* = f(x^*)$

$$f(x^*) = f(x^* - x + x) = f(x + \varepsilon)$$
$$= f(x) + \varepsilon f'(x) + \frac{1}{2}\varepsilon^2 f''(x) + \cdots$$

Assuming f has derivatives and a convergent Taylor series

To first order

$$f(x^*) \approx f(x) + \varepsilon \underbrace{f'(x)}_{\text{how its propagated...}}$$

Example $f(x) = x^2$

ignore higher order terms

$$f(x^*) = (x^*)^2 = (x+\varepsilon)^2 = x^2 + 2\varepsilon x + \varepsilon^2$$

$$\approx x^2 + \underbrace{2\varepsilon x}_{\text{absolute propagated error depended on the scale in } x}$$

we're more interested in relative error

$$\delta = \frac{x^* - x}{x} \qquad x\delta = x^* - x \qquad x(1+\delta) = x^*$$

$$f(x^*) = (x^*)^2 = \big(x(1+\delta)\big)^2 = x^2(1 + 2\delta + \delta^2)$$

$$\approx x^2(1 + \underbrace{2\delta}_{})$$

relative propagated error is again independent of scale...

The relative error is now 2δ in the output.

Another example

$$f(x) = \sqrt{x}$$

expand this

$$f(x^*) = \sqrt{x^*} = \sqrt{x(1+\delta)} = \sqrt{x}\sqrt{1+\delta}$$

$$= \sqrt{x}\left(1 + \frac{1}{2}\delta - \frac{1}{4}\delta^2 + \cdots\right)$$

$$\approx \sqrt{x}\left(1 + \frac{1}{2}\delta\right)$$

The relative error is now $\frac{1}{2}\delta$ in the output.

$$\sqrt{x}\Big|_{x=1} = 1$$

$$\frac{1}{2}x^{-1/2}\Big|_{x=1} = \frac{1}{2}$$

$$\left(\frac{1}{2}\right)\left(\frac{-1}{2}\right)x^{-3/2}\Big|_{x=1} = \frac{-1}{4}$$

$\sqrt{x}$

Figuring out the propagated error like above is called forward error analysis...

Plug in the answer to find a residual error, is backwards analysis...

Overflow and underflow...

mathematically $x = \sqrt{x^2}$ for $x > 0$

but overflow could spoil this

practically we want to comput:

$$|x+iy| = \sqrt{x^2+y^2}$$

intermediate part of the calc. may overflow...