Recap

Error Analysis

- True error

→ Approximate error

→ Error bound

- Truncation error

when limiting process truncated

- Data error

$$x \longrightarrow f(x)$$

$$\tilde{x} = x + \Delta x \longrightarrow f(x+\Delta x)$$

$$\Delta f(x) = |\Delta x \, f'(x)|$$

$$\Delta f(x_1, x_2, \ldots x_m) = \left(\sum \left| \Delta x_i \frac{\partial f}{\partial x_i} \right| \right)$$

Quadrature sum

$$\Delta f(x_i, x_2, \ldots x_m) = \sqrt{\sum_{i=1}^{m} \left(\Delta x_i \frac{\partial f}{\partial x_i} \right)^2}$$

Number representation

- Integer — unsigned, signed
- Fixed-point — XXX.XX
- Floating point

$$x = \pm m b^{p}$$

m — mantissa (significand)

b — base (2, 10)

p — exponent

How floating point numbers are stored

Sign	Exponent	Mantissa

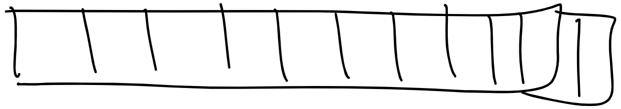
IEEE 754

64-bit

1 bit	11 bits	52 bits

Decimal 64 → -383 to 384 → IEEE 754-r (2008)

11 bits



2^{11} — 0 to 2047

Excess-p $\quad p = 2^{n-1} - 1 \quad$ n-bits

$p = 2^{10} - 1 = 1023$

-1023 to 1024

-1022 to 1024

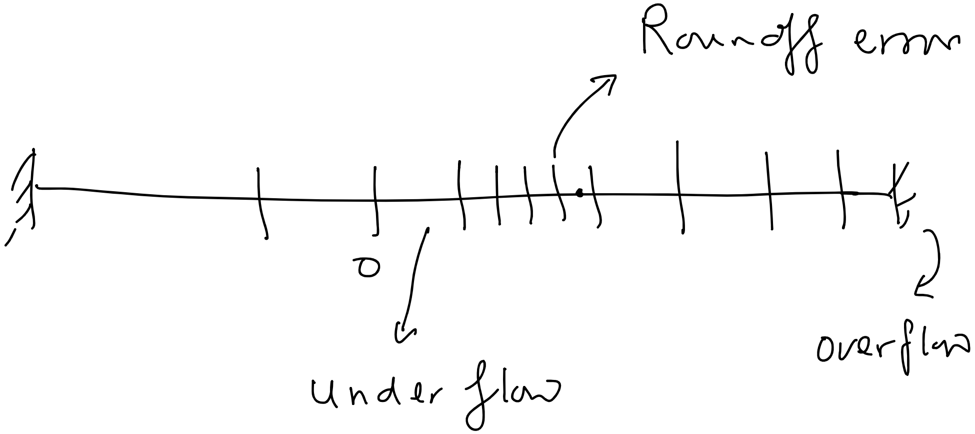
$$m\,2^{1024} = \bar{m}\,10^{a}$$

$$a = \frac{\log(2)}{\log(10)}\,1024$$

$$= 308$$

10^{-308} 10^{308} floating point

characteristics of computer numbers



1. Finite range

2. Hole near zero

3. Non-uniform gaps

Round off errors

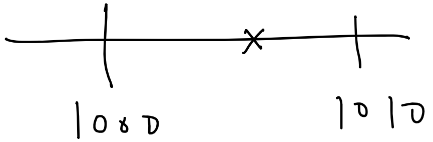
system 3 places for mantissa

1 place for exponent

1000 0.100×10^{4}

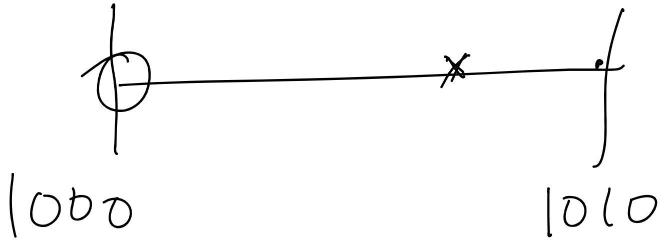
1010 0.101×10^{4}

1007



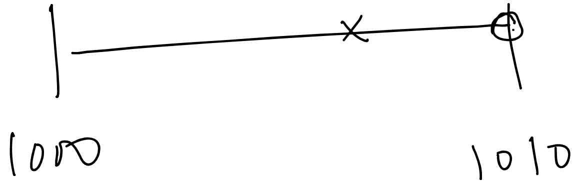
Chopping

1007 → $0.100\cancel{7} \times 10^4$



$|\Delta x| \le 10$

Rounding



$|\Delta x| \le 5$

Relative error

$$\left|\frac{\Delta x}{x}\right| = \frac{7}{1007}$$

Chopping $\left|\frac{\Delta x}{x}\right| \le \frac{10}{1000} = 10^{-2}$

Rounding $\le \frac{1}{2}10^{-2}$

General Round off $\left|\frac{\Delta x}{x}\right| \le \frac{1}{2}10^{1-t}$

t → number of significant digits in mantissa

$\le \frac{1}{2}2^{1-t}$

Relative round off error

$$\left|\frac{\Delta x}{x}\right| \le \frac{1}{2} b^{1-t} = u$$

t → number of significant digits in mantissa

machine precision

epsilon

round off units

64-bit binary

$$u = \frac{1}{2} 2^{1-t} \qquad t = 52$$

$$= 0.5 \times 2^{-51}$$

Addition

$$208.00 + 0.25$$

$$= 208.25$$

$$\begin{array}{r} 0.208 \times 10^{3} \\ + \; 0.25 \times 10^{0} \end{array}$$

$$\begin{array}{r} 0.208 \times 10^{3} \\ 0.00025 \times 10^{3} \\ \hline 0.20825 \times 10^{3} \\ \hline \end{array}$$

$$0.208$$

$$a + 1 - a = \underline{1}$$

Substraction

Two nearly equal numbers

$$x_1 = 0.246 \times 10^3$$

$$x_2 = 0.245 \times 10^3$$

$$= 0.001 \times 10^3$$

$$= 0.100 \times 10^1$$

$\Rightarrow$ Loss of significance

Forward error analysis

$$x \longrightarrow f(x)$$

$$x + \Delta x \longrightarrow f(x + \Delta x)$$

$$\boxed{\Delta f(x) = \left| \Delta x \frac{f'(x)}{f(x)} \right|}$$

Condition number of the problem

$$C_p = \frac{\text{Relative error in function } f(x)}{\text{Relative error in data } x}$$

$$= \frac{\Delta f(x) / f(x)}{\Delta x / x}$$

$$\boxed{C_p = \left| \frac{x f'(x)}{f(x)} \right|}$$

Well-conditioned $C_p < 1$ (attenuated)

ill-cond. $C_p > 1$ (as amplified

Condition number

$$C_p = \left| \frac{x f'(x)}{f(x)} \right|$$

$C_p < 1$ - well

$C_p > 1$ - ill conditn

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

$f'(x) = \frac{1}{2\sqrt{x}}$

$$C_p = \frac{x f'(x)}{f(x)}$$

$$C_p = \frac{1}{2}$$

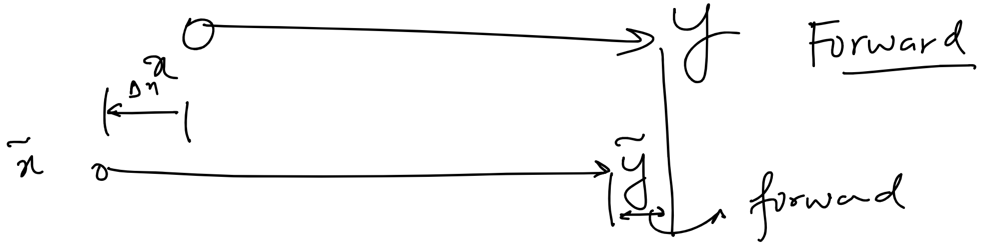
$$f(x) = \frac{10}{1-x^2}$$

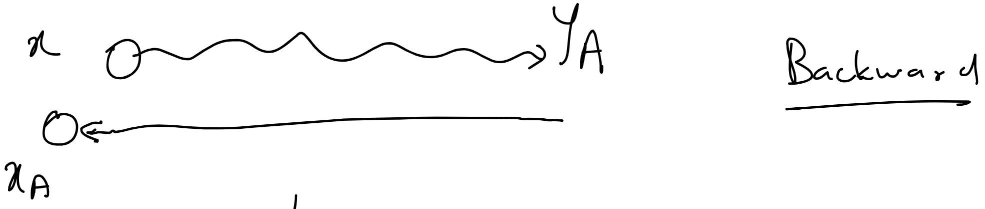
$$f'(x) = \frac{20x}{(1-x^2)^2}$$

$$C_p = \left| \frac{2x^2}{(1-x^2)} \right|$$

$x \simeq 1$ $C_p >> 1$ ill conditioning

Backward error analysis





$$\left|\frac{x - x_A}{x}\right| \le C_A u$$

C_A - condition number of the algorithm

u - epsilon

Example

4 digit decimal

$$u = \frac{1}{2} \times 10^{1-4}$$

$$= 0.5 \times 10^{-3}$$

$$f(x) = \sqrt{1+\sin x} - 1$$

$f(x) = 0.8688 \times 10^{-2}$

$x = 1°$

$fl(\pi/180) = 0.1745 \times 10^{-1}$

$fl(\sin x) = 0.1745 \times 10^{-1}$

$fl(1+\sin x) = 0.1017 \times 10^{1}$

$fl(\sqrt{1+\sin x}) = 0.1008 \times 10^{1}$

$fl(\sqrt{1+\sin x} - 1) = 0.\underline{8000 \times 10^{-2}}$

$$\sqrt{1+\sin x} - 1 = 0.8000 \times 10^{-2}$$

$$x_A = 0.9204 \times 10^{\cancel{-}1}$$

$$\left|\frac{x - x_A}{x}\right| = 0.0796 \overset{\leq}{\cancel{\leq}} C_A\, u$$

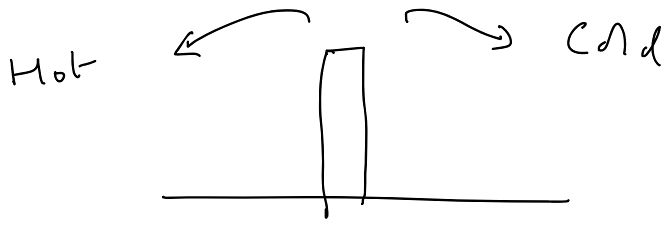
$$\boxed{C_A \approx 160}$$

$$f(x) = \left(\sqrt{1+\sin x} - 1\right) \frac{\left(\sqrt{1+\sin x} + 1\right)}{\left(\sqrt{1+\sin x} + 1\right)}$$

$$= \frac{\sin x}{\sqrt{1+\sin x} + 1}$$

$$C_A \approx 0.4$$

Analogy: Bathroom Shower



$T \qquad \theta$

$$\frac{\Delta\theta \ll \qquad T \gg}{\theta \rightsquigarrow T_A}$$

$$\theta_A \longleftarrow T_A$$

$$\left|\frac{\theta_A - \theta}{\theta}\right|$$