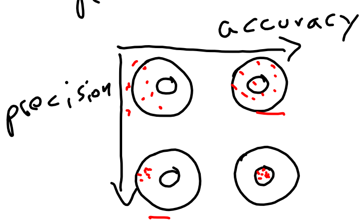
Errors — (Numerical)

percent error — normalize




error = true val − approx.

t.v. = approx + err.

$E_t$ = t.v. − approx.

$\varepsilon_t \equiv$ rel. error = $\dfrac{\text{t.v.} - \text{approx}}{\text{t.v.}}$

approx. error

$\varepsilon_a = \dfrac{\text{present approx} - \text{prev. approx}}{\text{present approx}}$

$\varepsilon_a \to 0$

↑ precision

Jan 19-9:46 AM

2 errors - Numerical

rounding → incr./decr. val

cutoff

digital computer representation of a number (floatingpt) is rarely accurate

base 10 — 1 2 3 4 . . .

10, 11, 12 . . .

base 2 — 1 — 1 = $2^0$

10 — 2 = $2^1 + 0 \cdot 2^0$

11 — 3 = $2^1 + 2^0$

100 — 4 = $2^2$

101 — 5 = $2^2 + 2^0$

Jan 19-9:58 AM

fractions (binary) b-10

1.0 — 1

1.1 — $1.5 = 2^0 + 2^{-1}$

1.01 — $1.25 = 2^0 + 2^{-2}$

$1.\underbrace{0001}$ — $1.0625 = 2^0 + 2^{-4}$

Jan 19-10:04 AM

2-bit → 00
01
10
11

{-2 -1 0 1}

$4 = 2^2$

4-bit → 0000
0001
0010
. . . .
1111

$16 = 2^4$

{-8, -7, -6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7}

Jan 19-10:08 AM

floating point

64-bit

| 1 | 11 bits | 52 bits |
|---|---|---|
| sign | exponent | mantissa |

$$\underbrace{(1+f)}_{\text{mantissa}} \times 2^{\overset{\text{exponent}}{e}}$$

$$\text{max 64bit \#} = 1.11\ldots11 \times 2^{1023} \approx 2^{1024}$$

$$\text{min \#} = -1.11\ldots11 \times 2^{1023} \approx -2^{1024}$$

$$\text{smallest \#} = 1.0000\ldots \times 2^{-1022} \approx 2.2251E{-308}$$

$$\text{max rel err} = \text{eps} = 2^{-52} \approx 2.2204E{-16}$$

Jan 19-10:14 AM