Reelle tall på datamaskin, Seksjon 4.2 i komp.

Normalformen til $a \in \mathbb{R}$:

$a = b \cdot 10^n$ ← normalformen til a.

$0.1 \leq |b| < 1$,

$0 = 0 \cdot 10^0$

Eks: $\pi \approx 0.3142 \cdot 10^1$

Binær normalform:

$$\left.\begin{aligned} a &= b \cdot 2^n \\ 0.5 &\leq |b| < 1 \end{aligned}\right\} a \neq 0$$

$0 = 0 \cdot 2^0$

På datamaskin:

32 bit flyttall:

23 bits til b (signifikanden)

9 bits til n (eksponent)

Husk at $\boxed{a = b \cdot 2^n}$, $0.5 \leq |b| < 1$

$F^{+}_{min32} \approx 1,4 \cdot 10^{-45}$ 6-8 desimale

$F^{+}_{max32} \approx 3,4 \cdot 10^{38}$ siffers nøyaktighet.

64 bits flyttall:

53 bits signifikand (b)

11 bits eksponent

$F^{+}_{min64} \approx 5 \cdot 10^{-324}$

$F^{+}_{max64} \approx 1,6 \cdot 10^{308}$

15-17 desimale siffer.

IEEE-standarden beskriver dette.

Egne bitmønstre for Infinity og NaN (Not a Num

Kommer fra ~~…~~ mange

muligheter i $0 \cdot 2^n$

Aritmetikk og avrundingsfeil.

1. Heltall er OK.

2. Reelle tall. Her kan mye rart skje!

Regn ut $(\sqrt{2})^2 - 2$. $\approx 10^{-16}$

Kun tall på formen $a/2^k$, $a,k \in \mathbb{N}$

a,k ikke for store, kan representeres eksakt på datamaskin

3.14159 - avrunding til 4 siffer 3.142

trunkering til 4 siffer 3.141

Forenklet flyttallsmaskin.

Maskinen opererer i 10-tall systemet.

4 desimale siffer til signifikanden
1 siffer til eksponenten
+ fortegn

$\max_+ = 0.9999 \cdot 10^9$

$\min_+ = 0.1000 \cdot 10^{-9}$

Addisjonsalgoritmen.

For å addere a og b:

1. Tallet med størst tallverdi skrives på normalform, f. eks a.

   $a = \alpha \cdot 10^n$

   og $b = \beta \cdot 10^n$

   $a + b = \alpha \cdot 10^n + \beta \cdot 10^n$
   $= (\alpha + \beta) 10^n$

2. Adder signifikandene
3. Konverter resultatet til normalform

Eksempel 5.11.

$a = 10.34, \quad b = -10.27$

Skriver på normalform

$a = 0.1034 \cdot 10^2 \quad , \quad b = -0.1027 \cdot 10^2$

Adder signifikandene

$0.1034 - 0.1027 = 0.0007$

Svar $0.0007 \cdot 10^2 = 0.7000 \cdot 10^{-1}$

Ex. 5.12 $\quad a = 10/7 \quad , \quad b = -1.42$

$\frac{10}{7} \approx a = 0.1429 \cdot 10^1 \quad , \quad b = -0.1420 \cdot 10^1$

Adderer signifikander

$0.1429 - 0.1420 = 0.0009$

Svar $0.0009 \cdot 10^1 = 0.9000 \cdot 10^{-2}$

Riktig svar: $\frac{10}{7} - 1.42 = 0.8571 \cdot 10^{-2}$