Kompendiet kap. 4

På en datamaskin representeres tall med 0 og 1 (bits).

1 byte = 8 bits
1 KB = 1000 bytes
1 MB = 1000 KB
1 GB = 1000 MB
1 TB = 1000 GB
1 PB = 1000 TB (petabyte)

Datamaskinen bruker vanligvis 32 eller 64 bits per tall (float/double)

Representasjon av heltall

Med 64 bits kan vi representere:

$$\underbrace{-2^{63}, \dots, -2, -1}_{2^{63} \text{ tall}}, \underbrace{0, 1, 2, \dots, 2^{63}-1}_{2^{63} \text{ tall}}$$

Eks: Fortegnet kan representeres med 1 bit
Talletts absoluttverd kan representeres med resterende 63 bits.

Representasjon av desimaltall

Reelle tall har endelig antall sifre før punktum (heltallsdel), men kan ha uendelig mange sifre etter punktum (desimaldel).

Normal-formen til et reelt tall a:
- i base 10: $a = b \cdot 10^n$, $\frac{1}{10} \leq |b| < 1$, $n \in \mathbb{Z}$
- i base 2: $a = b \cdot 2^n$, $\frac{1}{2} \leq |b| < 1$, $n \in \mathbb{Z}$

b kalles signifikand, n for eksponent.

Eksempel base 10: $a = 8$, $8 = \underbrace{0.8}_{\text{signifikand}} \cdot 10^{\overbrace{1}^{\text{eksponent}}}$

Eksempel base 2: $a = \frac{1}{4}$ $\quad \frac{1}{4} = 1 \cdot 2^{-2} = \frac{1}{2} \cdot 2^{-1}$ (signifikand, eksponent)

64 bits flyttall på en datamaskin bruker 53 bits for signifikandom, 11 bits for eksponenter.

Dette gir 15-17 signifikante 10-talls sifre, siden.

$$2^{52} = 10^{52 \log 2} \approx 4.5 \cdot 10^{15}$$

Hva er det største positive tallet maskinen kan representere?

Svar: største eksponent: $2^{10} - 1 = 1023$

største tall: $1 \cdot 2^{1023} \approx 8.99 \cdot 10^{307}$.

Hva er det minste tallet maskinen kan representere (i absoluttverdi)

Svar: minste eksponent: $-2^{10} = -1024$

minste tall: $\frac{1}{2} \cdot 2^{-1024} = 2^{-1025} \approx 2.8 \cdot 10^{-309}$.

Aritmetikk med flyttall og avrundingsfeil

Regel for å addere to tall a og b på normalform (i 10-tallsystem)

steg 1: Anta a er størst i absoluttverdi, og på normalform:

$$a = \alpha \cdot 10^n, \quad \frac{1}{10} \leq \alpha < 1$$

Skriv b med samme eksponent:

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

steg 2: Adder signifikandene: $\gamma = \alpha + \beta$

steg 3: Konverter $c = \gamma \cdot 10^n$ til normalform: $\hat{\gamma} \cdot 10^m$.

To problemer når vi bruker maskinens representasjon av tall her:

problem 1: Når b bruker n som eksponent også, så mister vi kanskje sifre i representasjon:

Eksempel 1: La $a = 7.923$, $b = 0.4621$

Regn ut $a+b$ på normalform, når vi bruker 4 10-talls sifre for signifanden.

Løsning: steg 1: a er størst, og $a = 0.7923 \cdot 10^1$

$b = 0.04621 \cdot 10^1 \approx 0.0462 \cdot 10^1$

mister 1 siffer (problem 1)

steg 2: Adder signifikand:

$$\begin{array}{r} 0.7923 \\ 0.0462 \\ \hline 0.8385 \end{array}$$

steg 3: $0.8385 \cdot 10^1$ er allerede på normalform.

Egentlig svar: 8.3851, så 4 riktige sifre.

problem 2: Mister kanskje sifre i representasjon når vi regner ut $x+B$.

Eksempel 2: La $a = 7.923$, $b = 4.621$

Regn ut $a+b$ på normalform når vi bruker 4 10-talls sifre for signifikanden.

Løsning: steg 1: a er størst, og $a = 0.7923 \cdot 10^1$

$b = 0.4621 \cdot 10^1$ ← mistet ikke siffer.

steg 2: Adder signifikander:

$$\begin{array}{r} 0.7923 \\ + \ 0.4621 \\ \hline 1.2544 \end{array}$$ ← mister sifferet.

steg 3: Vi får $1.254 \cdot 10^1 = 0.1254 \cdot 10^2 = 12.54$

egentlig svar: 12.544, så 1 siffer ble feil.

Eksempel uten avrundingsfeil

$a = 0.2 \qquad b = 0.1$

a størst: $0.2 \cdot 10^{0} + 0.1 \cdot 10^{0} \longrightarrow 0.3 \cdot 10^{0}$

Eksempel med stor feil:

$a = 10/7$, $b = -1.42$

Adder på normalform, me 4 10-talls sifre for signifikand.

Løsning: $a = 1.42857\ldots \qquad b = -1.42$

a er størst, og har normalform $\underbrace{0.1429}_{\alpha} \cdot 10^{1}$

$b = \underbrace{-0.1420}_{\beta} \cdot 10^{1}$

steg 2: $\gamma = \alpha + \beta = 0.1429 - 0.1420 = 0.0009$

steg 3: $0.0009 \cdot 10^{1} = 0.9 \cdot 10^{-2} = 0.009$

egentlig svar: $\frac{10}{7} - 1.42 = 0.00857142857\ldots$

Kun ett riktig svar i svaret, så stor avrundingsfeil.

Oppgaver kap. 4

4.1 : 1 ac, 2

4.2 : 1, 2, 4, 5 abc

Måling av feil (kap 5 i kompendiet)

Anta $\tilde{a}$ er en tilnærming til a (feks. beste tilnærming med flyttall).

To måter å måle feilen mellom a og $\tilde{a}$ på:

1. Absolutt feil: $|\tilde{a} - a|$
2. Relativ feil: $\left|\frac{\tilde{a} - a}{a}\right|$ (anta her at $a \neq 0$)

Eksempel La a være et tall, og la oss sammenligne a med

$$\tilde{a} = a + 100$$

Absolutt feil: $|(a+100) - a| = 100$ (uavhengig av a)

Relativ feil: $\left|\frac{(a+100)-a}{a}\right| = \left|\frac{100}{a}\right|$, (kan tolkes som andel avkastning når vi tjener 100 kr. på et år

Observasjon: Hvis $\tilde{a}$ er nærmest flyttall til a. Da er

$$\left|\frac{\tilde{a}-a}{a}\right| \lesssim 10^{-16}$$

(antar 64 bits flyttall)