## Kap. 6 i kompendiet: Simulering av differensligninger på datamaskin og avrundingsfeil

**Definisjon 6.7**

En likning på formen

$$(*) \qquad x_{n+k} = g(n) + f_0(n)x_n + f_1(n)x_{n+1} + \cdots + f_{k-1}(n)x_{n+k-1}$$

sies å være lineær, inhomogen, k-ordens differensligning

Kan simuleres på datamaskin, som for 1. orden, 2. orden, når man har initialbetingelser $x_0, x_1, \ldots, x_{k-1}$

For oss opp til nå: $f_0(n), f_1(n)\ldots$ har vært konstanter.

Hva kan gå galt i simuleringer som (*)?

1. Koeffisientene i (*) kan ikke representeres eksakt.
2. Initialbetingelsene i (*) kan ikke representeres eksakt.

### Eksempel med avrundingsfeil

Vi ser på $x_{n+2} - \frac{19}{3}x_{n+1} + 2x_n = 0$, $x_0 = 1$, $x_1 = \frac{1}{3}$.

Hva skjer på datamaskin? Hva er løsningen?

**Løsning:**

Analytisk løsning: kar. ligning:

$$r^2 - \frac{19}{3}r + 2 = 0$$

$$r = \frac{\frac{19}{3} \pm \sqrt{\frac{361}{9} - 8}}{2}$$

$$= \frac{\frac{19}{3} \pm \frac{17}{3}}{2} = \frac{19}{6} \pm \frac{17}{6}$$

$$\Rightarrow r_1 = 6, \quad r_2 = \frac{1}{3}$$

Generell løsning $x_n = C3^{-n} + D6^n$

Init. betingelser:

$x_0 = 1$: $C + D = 1$

$x_1 = \frac{1}{3}$: $\frac{1}{3}C + 6D = \frac{1}{3}$

$\Rightarrow C + D = 1$, $C + 18D = 1$ $\Rightarrow D = 0, C = 1$

$\Rightarrow x_n = 3^{-n}$

Hvorfor? Løsningen er ustabil i den forstand at hvis vi endrer litt på initialbetingelsene, så vil den nye løsningen gå mot $\infty$:

Her $x_0 = 1$

$x_1 = \frac{1}{3} + \varepsilon$ ← siden $\frac{1}{3}$ ikke kan rep. eksakt med 64-bits flyttall.

($\varepsilon \approx 10^{-16}$ for 64-bits flyttall)

Løsning med nye initialbetingelser:

$\tilde{x}_n = C3^{-n} + D6^n$, der $C, D$ løser

$C + D = 1$

$\frac{1}{3}C + 6D = \frac{1}{3} + \varepsilon$

$\Rightarrow D = 1 - C$

$\frac{1}{3}C + 6(1-C) = \frac{1}{3} + \varepsilon \Rightarrow \frac{17}{3}C = \frac{17}{3} - \varepsilon$

$\Rightarrow C = 1 - \frac{3}{17}\varepsilon$

$D = \frac{3}{17}\varepsilon$

Det følger at $\tilde{x}_n = (1 - \frac{3}{17}\varepsilon)3^{-n} + \frac{3}{17}\varepsilon 6^n$

For stor $n$ vil $\tilde{x}_n \approx \frac{3}{17}\varepsilon 6^n$, slik at $x_n \to \infty$.

Anta $x_n$ er første som blir $\infty$ (inf). Neste blir

$x_{n+1} = \frac{19}{3}x_n - 2x_{n-1} = \infty$

Neste: $x_{n+2} = \frac{19}{3}\underset{\infty}{x_{n+1}} - 2x_n$, som er $\infty - \infty = \text{nan}$.

**Merk** Avrundingsfeil i koeffisientene har samme effekt som over: Selv om $x_0, x_1$ kan representeres eksakt, så vil $x_2$ da få avrundingsfeil, som vil forplante seg slik som vi så avrundingsfeilen for $x_1$ i eksemplet over gjorde.

**Noen observasjoner:**

1. Anta $\underbrace{r^2 + br + c = 0}_{\text{kar. ligning}}$ har to forskjellige reelle røtter $r_1$ og $r_2$, og at $|r_1| > |r_2|$

   Hvis $\tilde{x}_n$ er den numeriske løsningen, så vil $\tilde{x}_n \approx Cr_1^n$ for store $n$, hvis det skjer avrundingsfeil.

2. Vi har at

$$\frac{\tilde{x}_n}{\tilde{x}_{n-1}} \approx \frac{Cr_1^n + Dr_2^n}{Cr_1^{n-1} + Dr_2^{n-1}} \approx \frac{Cr_1^n}{Cr_1^{n-1}} = r_1$$

   Dette kan være utgangspunkt for å finne røtter i polynomer.

**Eksempel** (tilsynelatende uten avrundingsfeil)

$$2x_{n+2} - 5x_{n+1} + 2x_n = -1 \qquad x_0 = 2,\ x_1 = \frac{3}{2}$$

Hva skjer når vi simulerer dette på datamaskin?

**Løsning:** Analytisk løsning:

partikulær løsning: prøv $x_n^p = A$:

$$\begin{aligned} 2A - 5A + 2A &= -1 \\ -A &= -1 \\ A &= 1 \end{aligned} \quad \Rightarrow x_n^p = 1$$

Homogen løsning: $2r^2 - 5r + 2 = 0$

$\Rightarrow r = \frac{5 \pm \sqrt{25-16}}{4} = \frac{5 \pm 3}{4} \Rightarrow r_1 = 2, r_2 = \frac{1}{2}$

$\Rightarrow X_n^h = C2^n + D2^{-n}$

Generell løsning: $X_n = X_n^p + X_n^h = 1 + C2^n + D2^{-n}$

Spesiell løsning: $X_0 = 2 \quad 1 + C + D = 2 \quad C + D = 1$

$X_1 = 3/2 \quad 1 + 2C + \frac{1}{2}D = \frac{3}{2} \quad 2C + \frac{1}{2}D = \frac{1}{2}$

$4C + D = 1$

$C + D = 1$
$4C + D = 1$
$C = 0, D = 1 \Rightarrow X_n = 1 + 2^{-n}$

Numerisk løsning: Maskinen regner ut $X_{n+2} = \frac{5}{2}X_{n+1} - X_n - \frac{1}{2}$

Her er det ikke avrundingsfeil, hverken i koeffisientene, eller initialverdiene.
så maskinen skal regne riktig, til å begynne med.

Vi får likevel avrundingsfeil til slutt.

Til slutt får vi overflow, deretter overflow igjen,
-Inf -Inf
til slutt nan.

Oppgave 18, midtveis 2021

Vi ser på $X_{n+1} - 3X_n = 1$, $X_1 = 1$

Hva skjer når vi simulerer dette med 64 bits flyttall?

Løsning: numerisk løsning: Maskinen regner ut $X_{n+1} = 3X_n + 1$,
Ingen avrundingsfeil i starten, vi vil få overflow
uansett.

Analytisk løsning: $X_n^p = A \Rightarrow A - 3A = 1$

$-2A = 1$

$A = -\frac{1}{2}$

$r - 3 = 0 \qquad X_n^n = C3^n$

Generell løsning: $x_n = -\frac{1}{2} + C3^n$

$x_1 = 1$: $\quad 1 = -\frac{1}{2} + 3C \Rightarrow 3C = \frac{3}{2} \Rightarrow C = \frac{1}{2}$

$x_n = -\frac{1}{2} + \frac{1}{2}3^n = \frac{1}{2}(3^n - 1) \longrightarrow \infty$

## Oppgave 19 midtveis 2021

Vi ser på $\quad 9x_{n+2} - 3x_{n+1} - 2x_n = 0 \quad x_0 = 1, x_1 = -\frac{1}{3}$

Hva skjer når vi simulerer med 64 bits flyttall?

**Løsning:** **Analytisk løsning:**

$$9r^2 - 3r - 2 = 0 \Rightarrow \cdots \Rightarrow r_1 = \frac{2}{3}, r_2 = -\frac{1}{3}$$

$$\Rightarrow x_n = C\left(\frac{2}{3}\right)^n + D\left(-\frac{1}{3}\right)^n$$

$$\left.\begin{array}{l} x_0 = 1 \\ x_1 = -\frac{1}{3} \end{array}\right\} \begin{array}{l} C + D = 1 \\ \frac{2}{3}C - \frac{1}{3}D = -\frac{1}{3} \end{array} \left\} \begin{array}{l} C + D = 1 \\ -2C + D = 1 \end{array} \right\} \quad C = 0, D = 1$$

$$\Rightarrow \underline{x_n = \left(-\tfrac{1}{3}\right)^n}$$

**Numerisk:** Maskinen regner ut $x_{n+2} = \frac{1}{3}x_{n+1} + \frac{2}{9}x_n$

Avrundingsfeil i både koeffisienter og initialverdi

Derfor vil maskinen regne ut noe på formen

$$(1 - \varepsilon)\left(-\tfrac{1}{3}\right)^n + \varepsilon\left(\tfrac{2}{3}\right)^n$$

i stedet. Vil uansett gå mot 0.