Differensligninger og avrundingsfeil

Knut Mørken

1. oktober 2016

Innledning

På forelesningene 28/9 og 30/9 forsøkte Martin og jeg å få fram hvordan avrundingsfeil kan få katastrofale følger under simulering av differensligninger. Jeg vet dette kan være vanskelig å få tak i og vil her forsøke å presentere tankene kort og forhåpentligvis klart.

1 Differensligning i Kalkulus

I Kalkulus er den mest generelle differensligningen på formen

$$x_{n+2} + bx_{n+1} + cx_n = f(n), \quad x_0 \text{ og } x_1 \text{ gitt.} \tag{1}$$

Vi vet at for slike ligninger kan vi finne løsningen som en formel for x_n som funksjon av n, så lenge høyre side $f(n)$ har passende form.

2 Mer generelle differensligninger

Nøkkelegenskapen ved en andreordens differensligning som (1) er at vi alltid kan regne ut neste ledd i følgen fra de to foregående så lenge vi kjenner de to første leddene,

$$x_{n+2} = f(n) - bx_{n+1} - cx_n, \quad \text{for } n \geq 0. \tag{2}$$

Denne egenskapen kan vi fastholde selv om vi tillater langt mer generelle ligninger. Høyresiden må bare være på en slik form at vi kan regne ut x_{n+2} fra x_{n+1} og x_n (og n). Med andre ord kan vi godt generalisere (2) til

$$x_{n+2} = F(n, x_n, x_{n+1}), \quad \text{for } n \geq 0, \tag{3}$$

der $F(n, x_n, x_{n+1})$ er et eller annet uttrykk som involverer x_n, x_{n+1} og n. Et eksempel kan for eksempel være

$$x_{n+2} = \sin x_n + \log x_{n+1}, \quad x_0 = 1, \quad x_1 = \pi.$$

3 Hvorfor trenger vi mer generelle differensligninger?

Jo mer generelle ligninger vi har, jo fler fenomener kan vi beskrive. For eksempel kan vi ikke beskrive utviklingen av et bankinnskudd ved hjelp av (1) om renta varierer, men det kan vi med (3) om vi velger funksjonen F på en passende måte.

Differensligninger brukes ofte til å beskrive utviklingen av en populasjon av dyr og skal slike beskrivelser bli realistiske må de nesten alltid være mer generelle enn (1).

4 To måter å løse differensligninger

Skal vi løse en differensligning, altså finne verdien til de ulike leddene i den ukjente følgen $\{x_n\}$, har vi to grunnleggende forskjellige alternativer.

4.1 Løsning ved formel

Den ene måten er å løse differensligningen ved å finne en formel for x_n som ikke involverer x_{n-1} eller x_{n-2}. Løsningsmetodene i Kalkulus er av denne typen. Ulempen er at slike metoder bare fungerer for veldig spesielle typer av ligninger.

4.2 Løsning ved simulering

Den andre måten er å bruke differensligningen ((2) eller (3)) til å generere verdiene i følgen i rekkefølgen x_2, x_3, x_4 og så videre, noe vi alltid kan gjøre når x_0 og x_1 er gitt. Skal vi regne ut mange verdier må vi skrive et dataprogram som gjør dette for oss. Dette kalles å simulere differensligningen.

Fordelen med denne måten er at den fungerer for alle typer differensligninger og ikke bare for spesielle klasser.

5 Hva med avrundingsfeil?

Avrundingsfeil er ikke vanligvis et problem med løsningsmetoden i seksjon 4.1, det som er begrensningen der er at ligningene må være så spesielle.

Når vi gjør beregninger på datamaskin kan alltid avrundingsfeil skape problemer, og simulering slik som beskrevet i seksjon 4.2 er intet unntak. Slikt kan være vanskelig å oppdage, men for ligninger av typen i (1) kan vi jo bruke begge de to løsningsmetodene og på den måten lære mer om avrundingsfeil. *Vi kan altså finne en formel for løsningen ved hjelp av den første metoden, og så regne ut løsningen ved simulering ved hjelp av den andre metoden. Hvis vi får forskjellig svar er det rimelig* *å tenke at dette må være på grunn av avrundingsfeil i den andre metoden (hvis vi ikke har gjort regnefeil i den første metoden).*

6 Sammenligning i et konkret tilfelle

I eksempel 6.25 i kompendiet har vi gjort nettopp dette, og vi ser at de to metodene gir svært forskjellige svar når n blir større enn 20. Forklaringen finner du i seksjon 6.5.1.

Kort oppsummert: Ligningen er

$$x_{n+2} - \frac{19}{3} x_{n+1} + 2x_n = -10, \quad x_0 = 2, x_1 = 8/3$$

som har løsningen $x_n = 3 - 3^{-n}$ når vi løser den med metoden i seksjon 4.1 over. Hvis vi i steden simulerer ligningen får vi ganske raskt at x_n vokser og blir langt større enn 3.

Forklaringen ligger i at den generelle løsningen av ligningen er $x_n = 3 + C3^{-n} + D6^n$. Startverdiene fører til at $C = -1$ og $D = 0$. Kunne vi brukt eksakt aritmetikk er det denne løsningen vi skulle fått også med simulering. Men startverdien $x_1 = 8/3$ kan ikke representeres eksakt med flyttall slik at vi på datamaskin isteden får som startverdi flyttallet som ligger nærmest $8/3$. Dette er et tall vi kan skrive $x_1 = 8/3 + \epsilon$ der ϵ er et tall vi ikke kjenner eksakt, men som maksimalt er omtrent 10^{-16} (dette vet vi fra kapitlene 4 og 5 i kompendiet). Effekten av dette er at når vi simulerer løsningen med metoden i seksjon 4.2 så svarer dette til at vi følger en eksakt løsning der $C = -1 + 3\epsilon/17$ og $D = 3\epsilon/17$ (se side 144 i kompendiet), altså løsningen

$$\hat{x}_n = 3 + (-1 + 3\epsilon/17)3^{-n} + (3\epsilon/17)6^n.$$

Fra dette ser vi at når n bare blir stor nok så vil det siste leddet dominere selv om koeffisienten foran er liten.

7 Hva kan vi lære av dette?

Fenomenet over skjer selvsagt med andre ligninger også. Den viktige observasjonen er at på grunn av avrundingsfeil er det sjelden startverdier kan representeres eksakt. Dette fører til at simulering av løsningen svarer til en løsning der ingen koeffisienter er eksakt null, noe som enkelte ganger kan ha katastrofale følger, slik som i eksempelet over.

For mer generelle ligninger av typen (3) kan vi vanligvis ikke finne den eksakte løsningen ved formel, vi er avhengige av å simulere, noe som er et svært slagkraftig verktøy i veldig mange sammenhenger. Erfaringene fra de enkle ligningene av typen (1) viser bare at vi må være på vakt, vi bør ikke akseptere resultatene av simuleringene uten et kritisk blikk.