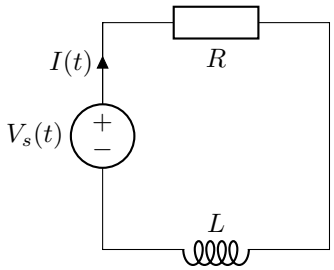
Førsteordens ordinære differensialligninger

Dette kapittelet gir en introduksjon til førsteordens ordinære differensialligninger. For lineære differensialligninger finnes det god teori og universelle løsningsteknikker, og vi kan ofte finne analytiske løsninger for disse. Dette gjelder i mye mindre grad for ikke-lineære differensialligninger, som det ofte er mye vanskeligere å utlede analytiske løsninger til, og som vi er avhengig av kvalitative eller numeriske metoder for å kunne analysere.

Vi begynner med å se på noen eksempler hvor differensialligninger oppstår som modeller for fysiske situasjoner, og kommer deretter tilbake til hvordan vi kan finne løsninger og forstå dynamikken bak modellene.

Eksempel 1. Følgende figur viser en elektrisk krets med en spenningskilde, en motstand og en spole (RL-krets).



Ved Ohms lov er spenningsfallet over motstanden lik RI og spenningsfallet over spolen er $L\dot{I}$, der $I(t)$ er strømmen i kretsen som funksjon av tid. Kirchhoffs spenningslov, som sier at summen av potensialforskjellene over en lukket strømsløyfe må utligne hverandre, gir dermed at

$$\dot{I} + \frac{R}{L} I = \frac{V_s}{L}.$$

Dette er et eksempel på en førsteordens lineær og ordinær differensialligning, hvor vi er på jakt etter den ukjente funksjonen I. △

Eksempel 2. I studiet av populasjonsvekst møter man ofte den logistiske ligningen

$$\dot{x}(t) = Ax(t) - B(x(t))^2,$$

hvor $x(t)$ beskriver en populasjonsstørrelse som funksjon av tid. Dersom vi setter $B = 0$ får vi en malthusiansk modell, etter den britiske samfunnsøkonomen Thomas Robert Malthus. Rundt slutten av 1700-tallet hevdet Malthus at befolkningen i Europa ville øke eksponentielt samtidig som matproduksjonen ville øke lineært, noe som ville føre til befolkningskollaps gjennom sultkatastrofer og epidemier. Senere skal vi se at leddet $-Bx^2$ begrenser veksten i populasjonsmodellen, og bøter på Malthus' dystre fremtidsutsikter. △

En kort oversikt

En førsteordens differensialligning er en relasjon mellom en (ukjent) funksjon og dens første deriverte. Vi skal i dette notatet bare jobbe med ordinære differensialligninger (forkortet som ODE (ordinary differential equation)); ligninger for funksjoner av én variabel. Generelt skal vi skrive førsteordens ODEer på formen

$$\dot{x}(t) = f(t, x(t)), \tag{1}$$

hvor $\dot{x}$ betegner deriverte av funksjonen x med hensyn på variabelen t. Funksjonen f beskriver relasjonen mellom $\dot{x}(t)$, t og $x(t)$. Dersom f ikke eksplisitt er avhengig av t, altså dersom vi kan skrive $f = f(x(t))$, sier vi at ligningen er autonom. Videre, dersom f er en lineær funksjon i x er ligningen lineær.

Definisjon 3. Funksjonen $x(t)$ er en løsning til ODEen (1) på et åpent intervall $I = (a, b)$ dersom den er kontinuerlig deriverbar og tilfredsstiller ligningen på I. △

Vi har altså to krav for at et funksjon $x(t)$ skal være en løsning til (1). Først må funksjonen være deriverbar, og den deriverte må være kontinuerlig. Hvis ikke dette kravet er på plass gir ikke ligningen så mye mening. Videre må funksjonen naturligvis tilfredsstille ligningen. Det åpne intervallet I kan være litt forvirrende, men nytten ligger i at vi nå kan snakke om løsninger som ikke nødvendigvis er definert på hele $\mathbb{R}$ (merk likevel at intervallet I fint kan være hele $\mathbb{R}$).

Eksempel 4. Ligningen

$$\dot{x} = x^2$$

har løsning

$$x(t) = \frac{1}{1-t}$$

på intervallet $(1, \infty)$. Siden $x(t)$ ikke eksisterer for $t = 1$ kan løsningen ikke utvides til å gjelde på hele $\mathbb{R}$. △

Dersom det er mulig å finne et eksakt uttrykk ved hjelp av et endelig antall kjente funksjoner (polynomer, trigonometriske funksjoner, eksponentialfunksjoner, osv.) for løsningen til en differensialligning, sier man ofte at man har en analytisk løsning. Dette står i kontrast til numeriske løsninger, som er tilnærminger.

I modellering kommer differensialligninger sjelden uten betingelser som vi må ta hensyn til dersom vi skal løse dem. Initialverdiproblemer er eksempler på dette. Generelt kan vi skrive initialverdiproblemet for førsteordens ODEer som

$$\begin{cases} \dot{x} = f(t, x), \\ x(t_0) = x_0, \end{cases} \tag{2}$$

hvor $x(t_0) = x_0$ er initialbetingelsen. En vanlig fysisk tolkning av dette er at vi ønsker å finne banen til en partikkel som befinner seg i posisjonen x_0 ved tiden t_0, og hvis fart er bestemt av $\dot{x} = f(t, x)$.

La oss vri og vende litt på initialverdiproblemet gitt i (2). Husk hvordan derivasjon er definert; som grenseverdien av stigningen til en funksjon. På denne måten er ligning (1) grensen til

ligningen

$$\frac{x(t+h)-x(t)}{h} = f(t, x(t)),$$

når $h \to 0$. På den annen side kan vi integrere (2) fra t_0 til t, altså

$$\int_{t_0}^{t} \frac{d}{ds} x(s)\, ds = \int_{t_0}^{t} f(s, x(s))\, ds,$$

og bruke analysens fundamentalteorem på venstre side til å skrive ligningen på integralformen

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds\,. \tag{3}$$

Integrasjonen over er utført over en variabel s. Dette er naturligvis ikke substansielt for utregningen, siden s forsvinner i integrasjonen, men det tillater oss å beholde variabelen t i ligningen.

Integralet på høyre side av (3) kan tilnærmes numerisk, noe som gir opphav til numeriske løsningsteknikker for initialverdiproblemet (2). En annen idé er å prøve fikspunktiterasjon på ligningen (3). Forhåpentligvis gir dette en smakebit på at differensialligninger kan (og bør) ses på i forskjellige lys, og at selv om det kan være vanskelig å utlede analytiske løsninger er det et stort utvalg av teknikker og triks tilgjengelig.

Lineære ligninger

En førsteordens lineær ODE kan skrives på standard form som

$$\dot{x}(t) + p(t)x(t) = q(t), \tag{4}$$

Dersom $q(t) = 0$ sier vi at ligningen er homogen. Ligning (4) kan faktisk løses generelt, som vist i følgende proposisjon.

Proposisjon 5. *Anta at $p(t)$ og $q(t)$ er kontinuerlige funksjoner på et åpent intervall I som inneholder punktet t_0. La C være en konstant, og la $P(t)$ være en antiderivert av $p(t)$, altså*

$$P(t) = \int p(t)\, dt\,.$$

Da er funksjonen

$$x(t) = e^{-P(t)}\left(C + \int_{t_0}^{t} q(s)e^{P(s)}\, ds\right), \tag{5}$$

en generell løsning til ligning (4).

Bevis. Multiplikasjon med funksjonen $e^{P(t)}$ på begge sider av ligning (4) gir

$$\dot{x}(t)e^{P(t)} + x(t)p(t)e^{P(t)} = q(t)e^{P(t)},$$

hvor vi har valgt $P(t)$ lik den antideriverte til $p(t)$. Ved kjerneregelen for derivasjon er venstre side nå lik den deriverte av $x(t)e^{P(t)}$, altså har vi

$$\frac{d}{dt}\big(x(t)e^{P(t)}\big) = q(t)e^{P(t)}.$$

Integrerer vi ligningen med hensyn på t får vi

$$\int_{t_0}^{t} \frac{d}{ds}\big(x(s)e^{P(s)}\big)\,ds = \int_{t_0}^{t} q(s)e^{P(s)}\,ds\,.$$

Analysens fundamentalteorem gir da at

$$x(t)e^{P(t)} - x(t_0)e^{P(t_0)} = \int_{t_0}^{t} q(s)e^{P(s)}\,ds,$$

hvor vi skriver $x(t_0)e^{P(t_0)}$ som en konstant C. Multiplikasjon med $e^{-P(t)}$ på begge sider gir nå den generelle løsningen (5). □

Løsningen kalles generell når den inneholder en ubestemt konstant C. Dersom konstanten C bestemmes av initialbetingelser sier vi at vi har en spesiell løsning. La oss se på noen eksempler.

Eksempel 6. Vi vet at ligningen

$$\dot{x} + ax = 0$$

har generell løsning

$$x(t) = Ce^{-at}$$

på hele $\mathbb{R}$. Dette stemmer overens med formelen (5) dersom vi setter $P(t) = \int a\,dt = at$ og $q(t) = 0$. △

Eksempel 7. Vi løser ligningen

$$\dot{x} - 2x = 3e^{t}$$

med metoden fra Proposisjon 5 på hele $\mathbb{R}$. Her er $P(t) = \int -2\,dt = -2t$, så vi ganger ligningen med e^{-2t}, og får

$$\dot{x}e^{-2t} - 2xe^{-2t} = 3e^{t}e^{-2t} = 3e^{-t}.$$

Dette kan vi forenkle til

$$\frac{d}{dt}\big(xe^{-2t}\big) = 3e^{-t}$$

ved hjelp av kjerneregelen. Siden det ikke er oppgitt en initialbetingelse her kommer løsningen til å inneholde en ukjent konstant. Vi kan for eksempel integrere fra 0 til t, som gir

$$\begin{aligned}\int_{0}^{t} \frac{d}{ds}\big(x(s)e^{-2s}\big)\,ds &= x(t)e^{-2t} - x(0)\\ &= \int_{0}^{t} 3e^{-s}\,ds = -3e^{-t} + 3.\end{aligned}$$

Vi flytter så $x(0)$ over til høyre side og ganger med e^{2t}, og står da igjen med den generelle løsningen

$$x(t) = (3 + x(0))e^{2t} - 3e^{t} = Ce^{2t} - 3e^{t}.$$

Den spesielle løsningen kan finnes ved å bestemme C fra en initialbetingelse for problemet. Vi merker oss at funksjonen e^{2t} løser den homogene ligningen

$$\dot{x} - 2x = 0,$$

og at funksjonen $-3e^{t}$ løser den tilhørende inhomogene ligningen med $3e^{t}$ på høyre side. Disse kalles henholdsvis den homogene og den inhomogene løsningen. △

Eksempel 8. La oss løse ligningen

$$\dot{I} + \frac{R}{L} I = \frac{V_s}{L}$$

fra Eksempel 1, gitt at $V_s(t)$ er en kontinuerlig funksjon på et åpent intervall I. Proposisjon 5 gir da at den generelle løsningen for strømmen i kretsen er

$$I(t) = e^{-\frac{R}{L}t}\left(C + \int_{t_0}^{t} \frac{V_s(s)}{L} e^{\frac{R}{L}s}\, ds\right),$$

for $t_0, t \in I$. Løsningen er naturligvis avhengig av spenningskilden $V_s(t)$. △

Med Proposisjon 5 kan vi løse en ganske stor klasse av førsteordens lineære ODEer. Her kommer et lite frempek på hvorfor vi har valgt å studere lineære ligninger som et eget tema før vi ser på ikke-lineære ligninger: de lineære teknikkene kan i stor grad generaliseres ved hjelp av lineær algebra.

Dersom funksjonen som inngår i ligningen er en vektorfunksjon med n komponenter på formen

$$\boldsymbol{x}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix},$$

kan vi skrive et system av lineære førsteordens homogene ODEer som

$$\dot{\boldsymbol{x}}(t) - A\boldsymbol{x}(t) = 0,$$

der A er en matrise. I prinsippet kan dette problemet løses på samme måte som den skalare ligningen (4). Til dette trenger vi riktignok litt mer matematisk maskineri, som vi ikke skal gå inn på her.

Ikke-lineære ligninger

Ikke-lineære differensialligninger er som regel vanskelige eller umulige å løse analytisk. Vi skal her se på to ulike teknikker: løsning ved separasjon og analyse av faseplan. Senere skal vi utlede numeriske metoder for ligningene.

Noen spesielle ODEer kan løses ved separasjon. Teknikken bygger på at man for ligninger på formen

$$\dot{x}(t) = f(x)g(t),$$

med $f(x(t)) \neq 0$ for $t, t_0 \in I$, og f og g kontinuerlige, kan integrere ligningen

$$\int_{t_0}^{t} \frac{1}{f(x(s))} \dot{x}(s)\, ds = \int_{t_0}^{t} g(t)\, dt,$$

og deretter bruke analysens fundamentalteorem til å finne løsningen $x(t)$. Vi ser på et eksempel.

Eksempel 9. Betrakt ligningen

$$\dot{x} = 1 + x^2.$$

Denne kan vi skrive som

$$\frac{\dot{x}}{1+x^2} = 1.$$

Siden den antideriverte til $\frac{1}{1+x^2}$ er $\arctan x$, kan vi skrive dette som

$$\frac{d}{dt} \arctan x = 1,$$

for så å integrere fra t_0 til t og bruke analysens fundamentalteorem til å regne ut at

$$\arctan x(t) - \arctan x(t_0) = t.$$

Dette gir altså løsningen

$$x(t) = \tan(t + C),$$

der C er konstanten $\arctan x(t_0)$ gitt av en eventuell initialbetingelse for problemet. △

Nå skal vi se på hvordan vi kan argumentere litt mer kvalitativt. Vi ønsker å hente ut informasjon om ligningen og systemet den beskriver uten å faktisk løse den. Dette kan være svært nyttig, både for ligninger vi ikke klarer å løse, og som test på om løsninger vi har funnet gir mening. Som et eksempel på dette skal vi se på populasjonsmodellen fra Eksempel 2.

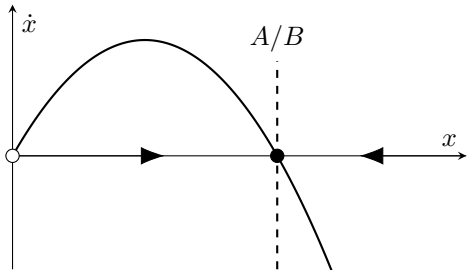
Eksempel 10. Populasjonsmodellen

$$\dot{x} = Ax - Bx^2 \tag{6}$$

kan løses ved separasjon: skriv ligningen som

$$\frac{\dot{x}}{Ax - Bx^2} = 1,$$

og bruk delbrøkoppspalting til å regne ut integralet av venstre side. Det er ikke vanskelig å se at dette blir mye regning. La oss heller skissere hvordan en løsning vil se ut dersom vi plotter $\dot{x}$ mot x, slik som vist i diagrammet under.



Når høyresiden i ligning (6) er positiv, altså for

$$0 < x(t) < \frac{A}{B},$$

er den deriverte også positiv, og populasjonen øker. Dersom populasjonen $x(t)$ overstiger A/B antall individer ser vi fra diagrammet at populasjonen begynner å minke, siden den deriverte nå er negativ (dette er tegnet inn som piler langs den horisontale aksen i diagrammet). Ut fra dette kan vi forvente oss at en hvilken som helst populasjon som følger loven (6) vil bevege seg mot likevektspunktet $x(t) = A/B$. Dette likevektspunktet er stabilt, siden pilene fra begge sider peker inn mot dette punktet. I $x = 0$ har vi et ustabilt likevektspunkt, ettersom et hvert avvik fra null medlemmer vil gjøre at populasjonen øker, helt til den når likevektspunktet i A/B. $\triangle$

Diagrammet som ble brukt i Eksempel 10 er et enkelt eksempel på et fasediagram. I stedet for å plotte løsningen $x(t)$ mot variabelen t plotter vi i fasediagrammet den deriverte $\dot{x}$ mot x selv. Variabelen t er dermed bare implisitt tilstede i fasediagrammet. Analysen består i å identifisere likevektspunktene til systemet, altså verdier for x slik at $\dot{x} = 0$, for så å resonnere seg frem til for hvordan $x(t)$ beveger seg mellom disse.

Eksistens og entydighet

Vi skal nå se på hva som kan sies om eksistens og entydighet av løsninger for det generelle initialverdiproblemet

$$\begin{cases} \dot{x} = f(t, x), \\ x(t_0) = x_0. \end{cases} \tag{7}$$

Dette kan virke litt umotivert for øyeblikket, fordi vi til nå bare har sett et lite utvalg av veldig snille ligninger. I det generelle tilfellet er det ofte svært vanskelig, og noen ganger umulig, å finne entydige løsninger til problemet (7). Derfor inkluderer vi følgende resultat, som gir betingelser for eksistens og entydighet av løsninger.

Teorem 11. *La $f(t, x)$ være en kontinuerlig funksjon på en åpen mengde $U \subset \mathbb{R}^2$ slik at $(t_0, x_0) \in U$, og anta at f er kontinuerlig deriverbar i variabelen x på området U. Da eksisterer det et tall $\delta > 0$ og en entydig løsning $x(t)$ til initialverdiproblemet (7) på intervallet $I = [t_0 - \delta, t_0 + \delta]$.*

Tidligere har vi brukt fikspunktiterasjon til å løse skalare ligninger numerisk. Nå skal vi ta dette et steg videre, ved å løse initialverdiproblemet (7) med fikspunktiterasjon. Beviset her er kun en skisse av de viktigste ideene, mens detaljene er utelatt. Vi må likevel introdusere følgende fakta som argumentasjonen er avhengig av. Dersom $g(x)$ er en kontinuerlig funksjon på et lukket intervall I finnes det en konstant $L < \infty$ slik at

$$\max_{x \in I} |f(x)| \leq L.$$

Videre, dersom g er kontinuerlig deriverbar på et lukket intervall I gjelder midellverditeoremet for funksjonen g, og vi har

$$|g(x) - g(y)| \leq \max_{x \in I} \big|f'(x)\big| |x - y| \leq L|x - y|, \tag{8}$$

for alle $x, y \in I$, siden den deriverte av f er kontinuerlig.

Skisse av bevis for Teorem 11. Som vist ovenfor kan initialverdiproblemet (7) skrives på integralform

$$x(t) = x(t_0) + \int_{t_0}^{t} f(s, x(s))\, ds\,.$$

Med dette som utgangspunkt kan vi definere fikspunktiterasjonen

$$g_{n+1}(t) = x_0 + \int_{t_0}^{t} f(s, g_n(s))\, ds, \tag{9}$$

som gir opphav til en følge funksjoner $\{g_n(t)\}_{n\in\mathbb{N}}$. Merk at dette er en utvidelse av fikspunktalgoritmene vi har sett tidligere: dette er en rekursiv formel for funksjoner, og ikke tall. La oss nå anta at vi begynner med konstanten x_0 som $g_0(t)$, altså vår første gjetning på hva løsningen er. Da blir

$$g_1(t) = x_0 + \int_{t_0}^{t} f(s, x_0)\, ds\,.$$

Hvor stor er forskjellen mellom $g_0(t)$ og $g_1(t)$? Differansen i den første rekursjonen på et intervall I rundt t_0 blir

$$\begin{aligned}
|g_1(t) - g_0(t)| &= \left| x_0 + \int_{t_0}^{t} f(s, x_0)\, ds - x_0 \right| \\
&= \left| \int_{t_0}^{t} f(s, x_0)\, ds \right| \\
&\leq \int_{t_0}^{t} |f(s, x_0)|\, ds \\
&\leq |t - t_0| \max_{t\in I} \big|f(t, x_0)\big|.
\end{aligned}$$

Siden f er kontinuerlig på en mengde U som inneholder (t_0, x_0) kan vi anta at f er kontinuerlig på intervallet $I = [t_0 - \delta, t_0 + \delta]$ (ved å velge δ liten), slik at

$$\max_{t\in I} \big|f(t, x_0)\big| \leq L < \infty.$$

Vi har også at $|t - t_0| \leq \delta$ for $t \in I$. Dermed har vi

$$|g_1(t) - g_0(t)| \leq |t - t_0| \max_{t\in I} \big|f(t, x_0)\big| \leq \delta L. \tag{10}$$

Den neste differansen i rekursjonen blir da

$$\begin{aligned}
|g_2(t) - g_1(t)| &= \left| \int_{t_0}^{t} f(s, g_1(s))\, ds - \int_{t_0}^{t} f(s, g_0(s))\, ds \right| \\
&\leq \int_{t_0}^{t} \big|f(s, g_1(s)) - f(s, g_0(s))\big|\, ds \\
&\leq \int_{t_0}^{t} L|g_1(s) - g_0(s)|\, ds \\
&\leq \delta L^2 \int_{t_0}^{t} ds \\
&\leq (\delta L)^2,
\end{aligned}$$

der vi har brukt at ulikheten (8) siden f er kontinuerlig deriverbar i x, og estimatet (10) for differansen $|g_1 - g_0|$. Dette argumentet kan vi nå iterere i det uendelige, og neste differanse blir

$$\begin{aligned}
|g_3(t) - g_2(t)| &= \left| \int_{t_0}^{t} f(s, g_2(s))\, ds - \int_{t_0}^{t} f(s, g_1(s))\, ds \right| \\
&\leq \int_{t_0}^{t} |f(s, g_2(s)) - f(s, g_1(s))|\, ds \\
&\leq \int_{t_0}^{t} L|g_2(s) - g_1(s)|\, ds \\
&\leq (\delta L)^3.
\end{aligned}$$

Vi kan gjette på at den generelle sammenhengen blir

$$|g_{n+1}(t) - g_n(t)| \leq (\delta L)^n,$$

noe fører oss til kjernen av argumentet: Dersom vi velger δ liten, slik at $\delta L < 1$, kommer differansen $|g_{n+1}(t) - g_n(t)|$ til å bli mindre og mindre for større n. Vi har faktisk at $(\delta L)^n \to 0$ for $n \to \infty$ når $\delta L < 1$. Dette betyr at forskjellen blir mindre og mindre: det kan se ut som om følgen vår kommer til å konvergere!

Matematikere bruker også dette til å vise at følgen $\{g_n(t)\}_{n\in\mathbb{N}}$ konvergerer til en funksjon g, og at denne nødvendigvis løser initialverdiproblemet (2) på intervallet I. Vi skal ikke gjøre dette grundig her, men med litt trening og litt mer kunnskap om konvergens i funksjonsrom kan du kanskje finne ut av detaljene selv. □

Teorem 11 er ofte greit å kjenne til, men kan ha begrenset praktisk betydning for en ingeniør. Som regel bruker vi numeriske metoder for å utforske differensialligninger.

Numeriske metoder

Problemet vi står overfor er å finne en tilnærming til løsningen $x(t)$ av initialverdiproblemet

$$\begin{cases} \dot{x} = f(t, x(t)), \\ x(t_0) = x_0, \end{cases} \tag{11}$$

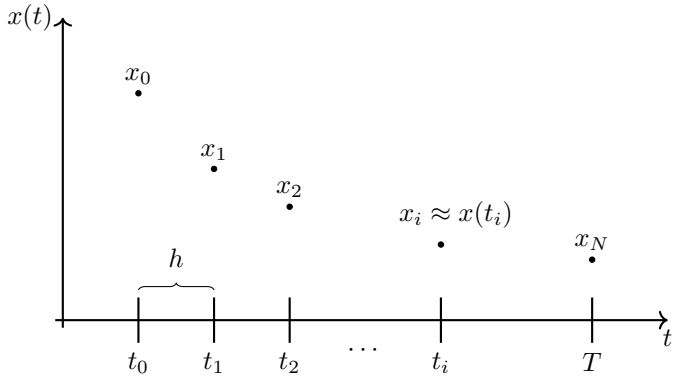
på intervallet $[t_0, T]$, for $T > t_0$. La $\{t_i\}_{i=0,\ldots,N}$ være en jevn partisjon

$$t_i = t_0 + hi, \qquad i = 0, 1, 2, ..., N$$

av intervallet $[t_0, T]$ i N like lange delintervaller $[t_i, t_{i+1})$ av lengde

$$h = \frac{T - t_0}{N}.$$

Vi ønsker å approksimere funksjonen $x(t)$ i punktene t_i, og innfører notasjonen $x_i \approx x(t_i)$ for tilnærmingen. Dette kan illustreres som vist under.



For å utlede formler for tilnærmingen $\{x_i\}_{i=0,\ldots,N}$ for $x(t)$ kan vi integrere initialverdiproblemet (11) over et delintervall $[t_i, t_{i+1})$, noe som gir

$$x(t_{i+1}) - x(t_i) = \int_{t_i}^{t_{i+1}} f(s, x(s))\, ds\,.$$

Dette gir formelen

$$x_{i+1} \approx x_i + \int_{t_i}^{t_{i+1}} f(s, x(s))\, ds,$$

som kan regnes ut ved numerisk approksimasjon av integralet på høyre side. Vi lister opp noen ulike måter å gjøre dette på.

Den enkleste tilnærmingen er å anta at $f(t, x(t)) \approx f(t_i, x_i)$ på $[t_i, t_{i+1})$. Det gir

$$\int_{t_i}^{t_{i+1}} f(s, x(s))\, ds \approx \int_{t_i}^{t_{i+1}} f(t_i, x_i)\, ds = h f(t_i, x_i),$$

som gir opphav til Eulers eksplisitte metode

$$x_{i+1} = x_i + h f(t_i, x_i), \qquad i = 0, \cdots, N-1.$$

Med tilnærmingen

$$\int_{t_i}^{t_{i+1}} f(x(s), s)\, ds \approx h f(t_{i+1}, x_{i+1})$$

får vi Eulers implisitte metode, mens trapesregelen for integraler

$$\int_{t_i}^{t_{i+1}} f(x(s), s)\, ds \approx \frac{h}{2}(f(t_i, x_i) + f(t_{i+1}, x_{i+1}))$$

gir trapesmetoden. Dersom vi i trapesmetoden tilnærmer x_{i+1} med Eulers eksplisitte metode får vi Heuns metode. Til sist, ved approksimasjonen

$$\int_{t_i}^{t_{i+1}} f(x(s), s)\, ds \approx h f\big(\frac{t_{i+1} + t_i}{2}, \frac{x_{i+1} + x_i}{2}\big)$$

får man midtpunktmetoden.

Definisjon 12. Vi har nå definert følgende numeriske metoder for initialverdiproblemet (11) over partisjonen $\{t_i\}_{i=0,\dots,N}$ av $[t_0, T]$.

$$x_{i+1} = x_i + hf(t_i, x_i), \qquad \text{(Euler eksplisitt)}$$
$$x_{i+1} = x_i + hf(t_{i+1}, x_{i+1}), \qquad \text{(Euler implisitt)}$$
$$x_{i+1} = x_i + \frac{h}{2}(f(t_i, x_i) + f(t_{i+1}, x_{i+1})), \qquad \text{(trapesmetoden)}$$
$$x_{i+1} = x_i + hf\left(\frac{t_{i+1} + t_i}{2}, \frac{x_{i+1} + x_i}{2}\right), \qquad \text{(midtpunktmetoden)}$$

og

$$\begin{cases} x^* = x_i + hf(t_i, x_i), \\ x_{i+1} = x_i + \frac{h}{2}(f(t_i, x_i) + f(t_{i+1}, x^*)). \end{cases} \qquad \text{(Hens metode)}$$

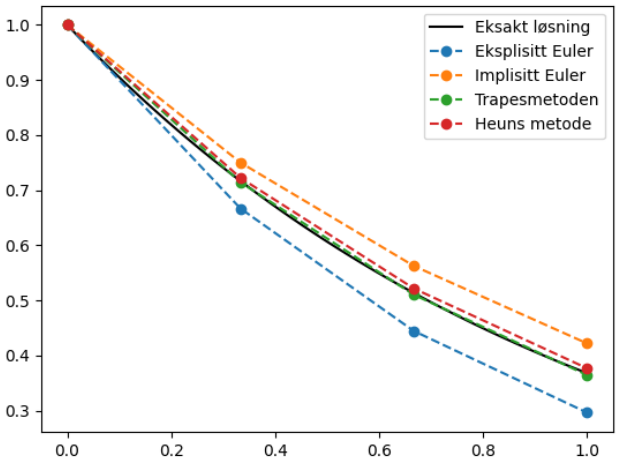
△

Har det noe å si hvilken metode vi bruker? La oss sammenligne metodene på et testproblem.

Eksempel 13. Vi anvender metodene over til å løse initialverdiproblemet

$$\begin{cases} \dot{x} = -x, \\ x(0) = 1, \end{cases}$$

på intervallet $[0, 1]$, med steglengde $h = 0.25$. Resultatet vises i figuren under.



Nå endrer vi på steglengden h, og ser hvordan feilen i forhold til den eksakte løsningen endrer seg. Den lokale feilen, altså differansen $|x(t_1) - x_1|$ for første steg, er vist for de ulike metodene i følgende tabell.

h	Euler eksplisitt	Euler implisitt	Trapesmetoden	Heuns metode
10^{-1}	$5.950e-03$	$5.161e-03$	$1.025e-04$	$2.224e-04$
10^{-2}	$5.084e-05$	$5.017e-05$	$8.502e-08$	$1.713e-07$
10^{-3}	$5.008e-07$	$5.002e-07$	$8.350e-11$	$1.671e-10$
10^{-4}	$5.001e-09$	$5.000e-09$	$8.338e-14$	$1.668e-13$

Her observerer vi at for Eulers eksplisitte og implisitte metode blir feilen omtrent hundre ganger mindre dersom vi gjør steglengden ti ganger mindre, mens for trapesmetoden og Heuns metode kan det se ut som om feilen blir tusen ganger mindre hver gang vi gjør steglengden ti ganger mindre.

Den globale feilen, altså feilen $|x(t_N) - x_N|$ i siste steget, finner vi som gitt i tabellen under.

h	Euler eksplisitt	Euler implisitt	Trapesmetoden	Heuns metode
10^{-1}	$2.144e-02$	$1.954e-02$	$3.790e-04$	$8.237e-04$
10^{-2}	$1.866e-03$	$1.850e-03$	$3.128e-06$	$6.303e-06$
10^{-3}	$1.842e-04$	$1.840e-04$	$3.072e-08$	$6.148e-08$
10^{-4}	$1.840e-05$	$1.840e-05$	$3.069e-10$	$6.133e-10$

Her ser vi at feiler minsker proporsjonalt med steglengden for eksplisitt og implisitt Euler, mens den minsker kvadratisk med steglengden for trapesmetoden og Heuns metode. △

Analyse av numeriske metoder

Målet er nå å forklare mønsteret vi observerte i tabellene fra Eksempel 13. Til dette formål er hovedverktøyet vårt Taylorutviklingen

$$x(t+h) = \sum_{i=0}^{n} \frac{h^n}{n!} x^{(n)}(t) = x(t) + h\dot{x}(t) + \frac{h^2}{2}\ddot{x}(t) + \cdots \tag{12}$$

av løsingen $x(t)$. Merk at denne formelen er en antagelse vi må gjøre, siden vi ikke på forhånd vet at $x(t)$ er en glatt funksjon og derfor kan utvlikles som en rekke i henhold til Taylors teorem. I praksis fungerer likevel dette bra.

Vi utleder den lokale feilen i Eulers eksplisitte metode. Først antar vi at $x_i = x(t_i)$, altså at approksimasjonen er lik den eksakte løsningen i punktet t_i. Så ønsker vi å se på differansen mellom den eksakte løsningen og approksimasjonen i neste steg, altså

$$\varepsilon_i := x(t_i + h) - x_{t_{i+1}}.$$

Ved hjelp av rekkeutviklingen (12) finner vi

$$\begin{aligned}
\varepsilon_i &= x(t_i + h) - x_{t_{i+1}} \\
&= x(t_i) + h\dot{x}(t_i) + \frac{h^2}{2}\ddot{x}(t_i) + O(h^3) - x_i - hf(t_i, x_i) \\
&= x(t_i) - x_i + hf(t_i, x_i) - hf(x_i, t_i) + \frac{h^2}{2}\ddot{x}(t_i) + O(h^3) \\
&= \frac{h^2}{2}\ddot{x}(t_i) + O(h^3)
\end{aligned}$$

der vi har brukt at $x(t_i) = x_i$ og at $\dot{x}(t_i) = f(t_i, x_i)$. For liten steglengde h øker derfor feilen maksimalt som h^2 på dette ene steget, akkurat som observert i Eksempel 13.

Vi kan forvente å få et estimat på den globale feilen dersom vi ser på summen av de lokale feilene. Det gir

$$\mathcal{E} := \sum_{i=0}^{N} \varepsilon_i \leq \sum_{i=1}^{N} \left| \frac{h^2}{2} \ddot{x}(t_i) \right| \leq \frac{Mh^2}{2} N = O(h),$$

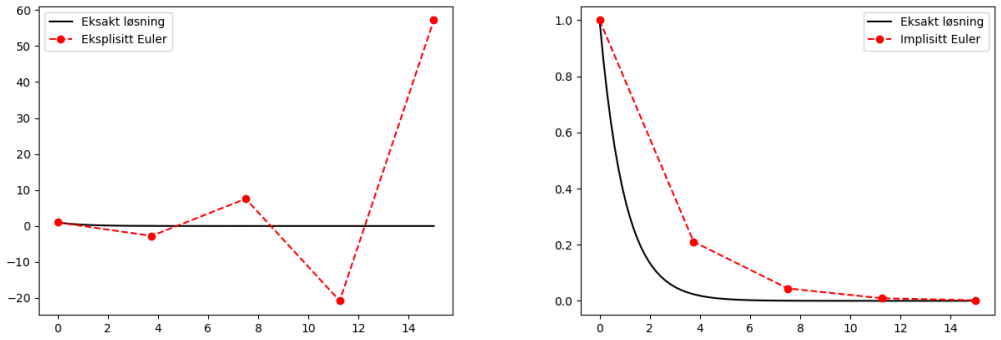
der vi har antatt at $\ddot{x} \leq M$ på $[t_0, T]$ (to ganger kontinuerlig deriverbar). Dette er også i overensstemmelse med Eksempel 13.

Til slutt skal vi se på én fordel og én ulempe med implisitte metoder som kan være greit å vite om.

Eksempel 14. Igjen ser vi på problemet

$$\begin{cases} \dot{x} = -x, \\ x(0) = 1, \end{cases}$$

men denne gangen på intervallet $[0, 15]$. Euler eksplisitt og Euler implisitt på problemet, med steglengde $h = 3$, produserer følgende approksimasjoner til løsningen.



Her kan man tydelig se at den implisitte metoden fungerer mye bedre enn den eksplisitte: den implisitte metoden virker å være mer stabil. Erfaring tilsier at implisitte metoder som regel er mye mer robuste når det kommer til stabilitet. △

Eksempel 15. Vi prøver oss på problemet

$$\begin{cases} \dot{x} = \sin(x), \\ x(0) = \dfrac{\pi}{4}. \end{cases}$$

Eulers implisitte metode med steglengde h blir

$$x_{i+1} = x_i + h \sin(x_{i+1}).$$

Men denne ligningen kan ikke løses eksplisitt for x_{i+1}! Det betyr at for hvert steg må vi beregne x_{i+1} med en numerisk ligningsløser, som f.eks fikspunktiterasjon eller Newtons metode. Dette kan gjøre implisitte metoder mer tidkrevende enn eksplisitte metoder. △